About the configuration guides

The H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series configuration guides describe the software features for the H3C S5590-HI[S5590-EI][S5500V3-HI] switch series in Release 8336P05 and later, and guide you through the software configuration procedures. These configuration guides also provide configuration examples to help you apply the software features to different network scenarios.

Configuration guide	Content
Fundamentals Configuration Guide	Covers information about using the command line interface, logging in to and setting up the device, and using the basic management features. This guide includes: • CLI. • RBAC. • Logging in to the switch. • FTP and TFTP. • File system management. • Configuration file management. • Multichassis configuration sync. • Software upgrade. • ISSU. • GIR. • Automatic configuration. • Device management. • Target configuration management. • Tcl. • Python. • License management.
Virtual Technologies Configuration Guide	Covers the IRF technology, which provides data center class availability and scalability.
Layer 2—LAN Switching Configuration Guide	Covers Layer 2 technologies and features used on a LAN switched network, such as VLAN technology and port isolation. You can use these features to divide broadcast domains, remove Layer 2 loops, isolate users within a VLAN, re-mark VLAN tags, or implement VLAN VPNs over the Internet. This guide includes: • MAC address table. • Bulk interface configuration. • M-LAG. • Port isolation. • VLAN. • MVRP. • Loopback, null, and inloopback interfaces. • QinQ. • VLAN mapping. • Loopback detection. • Spanning tree. • LLDP. • L2PT. • Layer 2 forwarding configuration.

Configuration guide	Content
	• Ethernet interface. • Ethernet link aggregation.
Layer 3—IP Services Configuration Guide	Covers IP addressing (including static and dynamic IPv4 and IPv6 address assignment), network performance optimization, ARP, and interoperation between IPv4 and IPv6. This guide includes: • ARP. • IP addressing. • DHCP. • DNS. • mDNS gateway. • mDNS relay. • IP forwarding basics. • Fast forwarding. • Adjacency table. • IRDP. • IP performance optimization. • UDP helper. • IPv6 basics. • DHCPv6. • IPv6 fast forwarding. • Tunneling. • GRE. • HTTP redirect.
Layer 3—IP Routing Configuration Guide	Covers the routing technologies for IPv4 and IPv6 networks of different sizes, route filtering, route control, and policy based routing. This guide includes: • Basic IP routing. • Static routing. • RIP. • OSPF. • IS-IS. • BGP. • Policy-based routing. • IPv6 static routing. • RIPng. • OSPFv3. • IPv6 policy-based routing. • Routing policy. • DCN. • Dual-stack PBR.
IP Multicast Configuration Guide	Covers Layer 2 IPv4 multicast protocols (including IGMP snooping, and multicast VLAN), Layer 3 IPv4 multicast protocols (including IGMP and PIM), Layer 2 IPv6 multicast protocols (including MLD snooping and IPv6 multicast VLAN), and Layer 3 IPv6 multicast protocols (including MLD, IPv6 PIM, and MSDP). This guide includes: • Multicast overview. • IGMP snooping. • PIM snooping. • Multicast VLAN.

Configuration guide	Content
	• PIM. • MSDP. • MLD snooping. • IPv6 PIM snooping. • IPv6 multicast VLAN. • IPv6 PIM. • Multicast VPN. • IGMP. • MLD. • Multicast routing and forwarding. • IPv6 multicast routing and forwarding.
MPLS Configuration Guide	Covers MPLS features. This guide includes: • Basic MPLS. • Static LSP. • LDP. • Tunnel policy. • MPLS L3VPN. • MPLS L2VPN. • VPLS. • MPLS OAM. • MCE.
Segment Routing Configuration Guide	Covers the *Segment Routing* fundamentals and configuration procedures.
ACL and QoS Configuration Guide	Covers information about classifying traffic with ACLs, and allocating network resources and managing congestions with QoS technologies to improve network performance and network use efficiency. This guide includes: • ACL. • QoS. • Data buffer. • Time range.
Security Configuration Guide	Covers security features. Available security features include identity authentication (AAA and PKI), access security (802.1X, MAC authentication, portal, and port security), secure management (SSH), SSL, and attack protection (IP source guard, ARP attack protection, MFF, and uRPF). This guide includes: • AAA. • 802.1X. • MAC authentication. • Portal. • Web authentication. • Triple authentication. • Port security. • User profile. • Password control. • Keychain. • Public key management. • PKI.

Configuration guide	Content
	• IPsec. • SSH. • SSL. • Attack detection and prevention. • TCP attack prevention. • IP source guard. • ARP attack protection. • ND attack defense. • uRPF. • MFF. • FIPS. • MACsec. • 802.1X client. • CloudSec. • Microsegmentation. • SAVI. • Object group.
High Availability Configuration Guide	Covers high availability technologies and features available on the switch for failure detection and failover. Failure detection technologies focus on fault detection and isolation. Failover technologies focus on network recovery. This guide includes: • Ethernet OAM. • DLDP. • RRPP. • ERPS. • Smart Link. • Monitor Link. • Error code detection. • VRRP. • BFD. • Track. • Process placement. • Interface collaboration.
Network Management and Monitoring Configuration Guide	Covers features that help you manage and monitor your network, for example, manage system events, collect traffic statistics, sample packets, assess network performance, and test network connectivity. This guide includes: • System maintenance and debugging. • NQA. • iNQA. • NTP. • PTP. • PoE. • SNMP. • RMON. • Ansible configuration. • EPA. • CWMP. • EAA. • Process monitoring and maintenance.

Configuration guide	Content
	• Sampler. • Mirroring. • sFlow. • Information center. • Packet capture. • VCF. • NetStream. • Cloud connection. • IPv6 NetStream. • Performance management. • eMDI. • SQA. • TCP connection trace. • SmartMC. • AON Ethernet system • NETCONF. • Quicknet.
Telemetry Configuration Guide	Covers the telemetry fundamentals and configuration procedures. You can use the telemetry technologies to remotely collect data from network devices, so as to monitor device performances and network operations. This guide includes: • gRPC. • Flow group. • MOD.
OpenFlow Configuration Guide	Covers the *OpenFlow* fundamentals and configuration procedures. OpenFlow allows a controller to directly access and operate the forwarding plane of the switch, separating the control plane from the data plane. The switch matches and forwards packets according to the flow tables deployed by the controller.
VXLAN Configuration Guide	Covers the *VXLAN* fundamentals and configuration procedures.
EVPN Configuration Guide	Covers the *EVPN* fundamentals and configuration procedures.
TAP Configuration Guide	Covers the *TAP* fundamentals and configuration procedures.
Network Slicing Configuration Guide	Covers the *Network Slicing* fundamentals and configuration procedures.
Intelligent Lossless Network Configuration Guide	Covers the *Intelligent Lossless Network* fundamentals and configuration procedures.

Preface

This configuration guide describes features and tasks that help you get started with the switch.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Using the CLI 1

About the CLI 1
Using CLI views 1
About CLI views 1
Entering common system view from user view 2
Entering real-time-mode system view from user view 2
Entering non-real-time-mode system view from user view 2
Entering system view for multichassis configuration sync from user view 3
Specifying the default mode of the system view 3
Returning to the upper-level view from a view 4
Returning to user view 4
Accessing the CLI online help 4
Using the undo form of a command 5
Entering a command 6
Editing a command line 6
Entering a text or string type value for an argument 7
Entering an interface type 7
Abbreviating commands 8
Configuring and using command aliases 8
Configuring and using hotkeys 9
Enabling redisplaying entered-but-not-submitted commands 11
Understanding command-line syntax error messages 11
Using the command history feature 11
About command history buffers 11
Command buffering rules 12
Managing and using the command history buffers 12
Repeating commands in the command history buffer for a user line 13
Controlling the CLI output 13
Pausing between screens of output 13
Numbering each output line from a display command 14
Filtering the output from a display command 14
Displaying information for the peer member device 18
Saving the output from a display command to a file 19
Viewing and managing the output from a display command effectively 20
Enabling reauthentication for disruptive commands 21
Entering MMI mode 21

Using the CLI

About the CLI

At the command-line interface (CLI), you can enter text commands to configure, manage, and monitor the device.

You can use different methods to log in to the CLI. For example, you can log in through the console port or Telnet. For more information about login methods, see "Login overview."

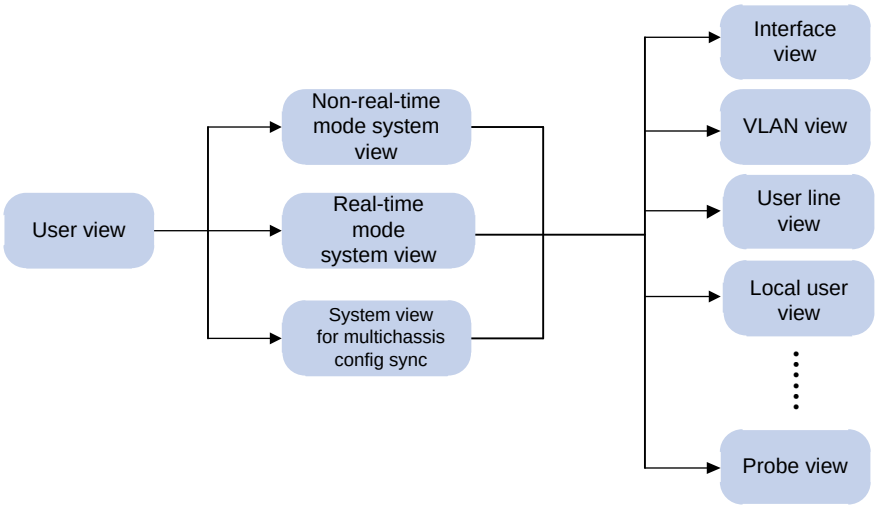
Using CLI views

About CLI views

Commands are grouped in different views by feature. To use a command, you must enter its view.

CLI views are hierarchically organized, as shown in Figure 1. Each view has a unique prompt, from which you can identify where you are and what you can do. For example, the prompt [Sysname-vlan100] shows that you are in VLAN 100 view and can configure attributes for that VLAN.

Figure 1 CLI views



You are placed in user view immediately after you log in to the CLI.

In user view, you can perform the following tasks:

- Perform basic operations including display, debug, file management, FTP, Telnet, clock setting, and reboot.
- Enter system view.

In system view, you can perform the following tasks:

- Configure settings that affect the device as a whole, such as the daylight saving time, banners, and hotkeys.
- Enter feature views.

  For example, you can perform the following tasks:
  - Enter interface view to configure interface parameters.

- Enter VLAN view to add ports to the VLAN.
- Enter user line view to configure login user attributes.

A feature view might have child views. For example, NQA operation view has the child view HTTP operation view.

- Enter probe view by using the **probe** command.

  The probe view provides display, debugging, and maintenance commands, which are mainly used by developers and testers for system fault diagnosis and system operation monitoring.

CAUTION:

Use the commands in probe view under the guidance of engineers to avoid system exceptions caused by incorrect operations.

For more information about the commands in probe view, see the probe commands manual for each feature.

To display all commands available in a view, enter a question mark (?) at the view prompt.

Entering common system view (commit command not supported)

Entering common system view from user view

About this task

In common system view and its lower-level views, settings are written to the running configuration and take effect immediately.

Procedure

To enter common system view from user view, execute the following command:

system-view

Entering real-time-mode system view from user view

About this task

In real-time-mode system view and its lower-level views, settings are written to the running configuration and take effect immediately.

Available commands in the system views might vary. For more information, enter a question mark (?) at the system view prompts.

Unless otherwise stated, the term "system view" refers to real-time-mode system view.

Procedure

To enter real-time-mode system view from user view, execute the following command:

system-view [**immediate**]

Entering non-real-time-mode system view from user view

About this task

For non-real-time-mode system view, you can enter exclusive-mode system view and private-mode system view.

In exclusive-mode system view, private-mode system view, and their lower-level views, settings are saved to the target configuration, which takes effect after being committed by using the **commit** command. For more information, see "Managing target configuration."

Only one user can enter exclusive-mode system view or its lower-level views.

- When a user is in an exclusive-mode view, users in private-mode views cannot commit their target configuration command lines.
- When no user is in an exclusive-mode view, users in private-mode views can commit their respective target configuration command lines. However, only one user in a private-mode view can commit target configuration command lines at a time.

Available commands in the system views might vary. For more information, enter a question mark (?) at the system view prompts.

Private-mode views and exclusive-mode views have a leading tilde sign (~) in their prompts as compared to their counterpart real-time-mode views. After you execute a command that does not take effect in real time, the prompt will have a leading asterisk sign (*) to indicate that the command has not taken effect in the current mode. After the **commit** command is executed to commit the configuration, the prompt will have a leading tilde sign (~).

Procedure

To enter non-real-time-mode system view from user view, execute the following command:

system-view { **exclusive** | **private** }

Entering system view for multichassis configuration sync from user view

About this task

In a network configured with multichassis configuration synchronization, such as an M-LAG network, some commands must be executed on both member devices. For these commands, you have to execute them separately on the local and peer member devices. This is cumbersome and might cause configuration errors.

To resolve this issue, you can execute the **system-view mc-config** command to enter system view for multichassis configuration sync. In system view for multichassis configuration sync and its lower-level views, settings on the local member device are synchronized to the peer member device. This ensures configuration consistency on the local and peer member devices and avoids repeated configuration. For more information about multichassis configuration synchronization, see "Configuring multichassis configuration synchronization."

Procedure

To enter system view for multichassis configuration sync from user view, execute the following command:

system-view mc-config

Specifying the default mode of the system view

About this task

By default, the system enters real-time-mode system view after the **system-view** command is executed. To configure the settings in non-real-time-mode, you must specify the **exclusive** or **private** keyword each time the system enters system view. To simplify the configuration, you can execute the **configuration system-view-mode** command to specify the default mode of the system view that the system will enter after the **system-view** command is executed.

Procedure

To specify the default mode of the system view that the system will enter after the **system-view** command is executed, execute the following command:

configuration system-view-mode { **immediate** | **private** }

By default, the system enters real-time-mode system view after the **system-view** command is executed.

Returning to the upper-level view from a view

Restrictions and guidelines

Executing the **quit** command in user view terminates your connection to the device.

To return from public key view to system view, you must use the **peer-public-key end** command.

Procedure

To return to the upper-level view from a view, execute the following command:

quit

Returning to user view

About this task

This feature enables you to return to user view from any other view by performing a single operation. You do not need to execute the **quit** command multiple times.

To return directly to user view, you can execute the **abort** or **return** command or press **Ctrl+Z**.

- If you execute **return** or press **Ctrl+Z** to return from private-mode or exclusive-mode view to user view, the system displays a message to notify you of uncommitted target configuration command lines (if any).
- If you execute **abort** to return from private-mode or exclusive-mode view to user view, the system discards uncommitted target configuration command lines (if any) without any prompt.

Procedure

To return directly to user view from any other view except Tcl configuration view and the Python shell, use one of the following methods:

- Execute the **return** command.
- Press **Ctrl+Z**.
- Execute the **abort** command in any private-mode/exclusive-mode view.

To return to user view from Tcl configuration view, use the **tclquit** command.

To return to user view from the Python shell, use the **exit()** command.

Accessing the CLI online help

The CLI online help is context sensitive. Enter a question mark at any prompt or in any position of a command to display all available options.

To access the CLI online help, use one of the following methods:

- Enter a question mark at a view prompt to display the first keyword of every command available in the view. For example:

```
<Sysname> ?
User view commands:
  archive             Archive configuration
  arp                 Address Resolution Protocol (ARP) module
  backup              Backup the startup configuration file to a TFTP server
  boot-loader         Software image file management
...
```

- Enter a space and a question mark after a command keyword to display all available keywords and arguments.
  - If the question mark is in the place of a keyword, the CLI displays all possible keywords, each with a brief description. For example:

    ```
    <Sysname> terminal ?
      debugging  Enable to display debugging logs on the current terminal
      logging    Display logs on the current terminal
      monitor    Enable to display logs on the current terminal
    ```

  - If the question mark is in the place of an argument, the CLI displays the description for the argument. For example:

    ```
    <Sysname> system-view
    [Sysname] interface vlan-interface ?
      <1-4094>  Vlan-interface interface number
    [Sysname] interface vlan-interface 1 ?
      <cr>
    [Sysname] interface vlan-interface 1
    ```

    <1-4094> is the value range for the argument. **<cr>** indicates that the command is complete and you can press **Enter** to execute the command.
- Enter an incomplete keyword string followed by a question mark to display all keywords starting with that string. The CLI also displays the descriptions for the keywords. For example:

  ```
  <Sysname> f?
    fdisk    Partition a storage medium
    fixdisk   Check and repair a storage medium
    format    Format a storage medium
    free      Release a connection
    ftp       Open an FTP connection
  <Sysname> display ftp?
    ftp         FTP module
    ftp-server  FTP server information
    ftp-user    FTP user information
  ```

Using the undo form of a command

Most configuration commands have an **undo** form for the following tasks:

- Canceling a configuration.
- Restoring the default.
- Disabling a feature.

For example, the **`info-center enable`** command enables the information center. The **undo `info-center enable`** command disables the information center.

Entering a command

When you enter a command, you can perform the following tasks:

- Use keys or hotkeys to edit the command line.
- Use abbreviated keywords or keyword aliases.

Editing a command line

To edit a command line, use the keys listed in Table 1 or the hotkeys listed in "Configuring and using hotkeys." When you are finished, you can press **Enter** to execute the command. If the command entered by the user cannot be executed in the current view, the system will try to execute the command in the upper level view of the current view. The execution fails until the command cannot be executed in system view.

The command edit buffer can contain a maximum of 511 characters. If the total length of a command line exceeds the limit after you press **Tab** to complete the last keyword or argument, the system does not complete the keyword.

IMPORTANT:

If the amount of free memory space reaches a free-memory threshold set by the **`memory-threshold`** command, the device displays memory usage information every time you execute a command. For more information about the **`memory-threshold`** command, see *Fundamentals Command Reference.*

Table 1 Command line editing keys

Keys	Function
Common keys	If the edit buffer is not full, pressing a common key inserts a character at the cursor and moves the cursor to the right. The edit buffer can store up to 511 characters. Unless the buffer is full, all common characters that you enter before pressing **Enter** are saved in the edit buffer.
Backspace	Deletes the character to the left of the cursor and moves the cursor back one character.
Left arrow key (←)	Moves the cursor one character to the left.
Right arrow key (→)	Moves the cursor one character to the right.
Up arrow key (↑)	Displays the previous command in the command history buffer.
Down arrow key (↓)	Displays the next command in the command history buffer.
Tab	If you press **Tab** after typing part of a keyword, the system automatically completes the keyword. • If a unique match is found, the system displays the complete keyword. • If there is more than one match, press **Tab** multiple times to pick the keyword you want to enter. • If there is no match, the system does not modify what you entered but displays it again in the next line.

The device supports the following special commands:

- #—Used by the system in a configuration file as separators for adjacent sections.
- **`version`**—Used by the system in a configuration file to indicate the software version information. For example, **`version 7.1. xxx`**, **`Release xxx`**.

These commands are special because of the following reasons:

- These commands are not intended for you to use at the CLI.
- You can enter the **#** command in any view or the **version** command in system view, or enter any values for them. For example, you can enter **# abc** or **version abc**. However, the settings do not take effect.
- The device does not provide any online help information for these commands.

Entering a text or string type value for an argument

A text type argument value can contain any characters except question marks (?).

A string type argument value can contain any printable characters except question marks (?).

- To include a quotation mark (") or backward slash (\) in a string type argument value, prefix the character with an escape key (\), for example, \" and \\.
- To include a blank space in a string type argument value, enclose the value in quotation marks, for example, "my device".

A specific argument might have more requirements. For more information, see the relevant command reference.

To enter a printable character, you can enter the character or its ASCII code in the range of 32 to 126.

Entering an interface type

You can enter an interface type in one of the following formats:

- Full spelling of the interface type.
- An abbreviation that uniquely identifies the interface type.
- Acronym of the interface type.

For a command line, all interface types are case insensitive. The following table shows the full spellings and acronyms of interface types.

For example, to use the **interface** command to enter the view of interface GigabitEthernet 1/0/1, you can enter the command line in the following formats:

- **interface gigabitethernet 1/0/1**
- **interface g 1/0/1**
- **interface ge 1/0/1**

Spaces between the interface types and interfaces are not required.

Table 2 Full spellings and acronyms of interface types

Full spelling	Acronym
Bridge-Aggregation	BAGG
GigabitEthernet	GE
InLoopBack	InLoop
LoopBack	Loop
M-GigabitEthernet	MGE
NULL	NULL
Register-Tunnel	REG
Route-Aggregation	RAGG

Ten-GigabitEthernet	XGE
Tunnel	Tun
Twenty-FiveGigE	WGE
Vsi-interface	Vsi
Vlan-interface	Vlan-int

Abbreviating commands

You can enter a command line quickly by entering incomplete keywords that uniquely identify the complete command. In user view, for example, commands starting with an **s** include **startup saved-configuration** and **system-view**. To enter the command **system-view**, you need to type only **sy**. To enter the command **startup saved-configuration**, type **st s**.

You can also press **Tab** to complete an incomplete keyword.

Configuring and using command aliases

About this task

You can configure one or more aliases for a command or the starting keywords of commands. Then, you can use the aliases to execute the command or commands. If the command or commands have **undo** forms, you can also use the aliases to execute the **undo** command or commands.

For example, if you configure the **shiprt** alias for **display ip routing-table**, you can enter **shiprt** to execute the **display ip routing-table** command. If you configure the **ship** alias for **display ip**, you can use **ship** to execute all commands starting with **display ip**, including:

- Enter **ship routing-table** to execute the **display ip routing-table** command.
- Enter **ship interface** to execute the **display ip interface** command.

The device provides a set of system-defined command aliases, as listed in Table 3.

Table 3 System-defined command aliases

Command alias	Command or command keyword
access-list	**acl**
end	**return**
erase	**delete**
exit	**quit**
hostname	**sysname**
logging	**info-center**
no	**undo**
show	**display**
write	**save**

Restrictions and guidelines

A command alias can be used only as the first keyword of a command or the second keyword of the **undo** form of a command.

After you successfully execute a command by using an alias, the system saves the command, instead of the alias, to the running configuration.

The command string can include up to nine parameters. Each parameter starts with the dollar sign ($) and a sequence number in the range of 1 to 9. For example, you can configure the alias **shinc** for the **display ip $1 | include $2** command. Then, to execute the **display ip routing-table | include Static** command, you need to enter only **shinc routing-table Static**.

To use an alias for a command that has parameters, you must specify a value for each parameter. If you fail to do so, the system informs you that the command is incomplete and displays the command string represented by the alias.

System-defined command aliases cannot be deleted.

Procedure

1. Enter system view.

   system-view
2. Configure a command alias.

   alias *alias command*

   By default, the device has a set of command aliases, as listed in Table 3.
3. (Optional.) Display command aliases.

   display alias [*alias*]

   This command is available in any view.

Configuring and using hotkeys

About this task

The device supports a set of hotkeys. Pressing a hotkey executes the command or function assigned to the hotkey. Table 4 shows the hotkeys and their default definitions. You can configure all the hotkeys except **Ctrl+]**.

If a hotkey is also defined by the terminal software you are using to interact with the device, you can reconfigure the hotkey or remove the hotkey.

Restrictions and guidelines

A hotkey can correspond to only one command or function. If you assign multiple commands or functions to the same hotkey, the most recently assigned command or function takes effect.

A command or function can be assigned to multiple hotkeys. You can use any of the hotkeys to execute the command or function.

If a hotkey is also defined by the terminal software you are using to interact with the device, the terminal software definition takes effect.

Procedure

1. Enter system view.

   system-view
2. Assign a command to a hotkey.

   hotkey *hotkey* { *command* | **function** *function* | **none** }

   Table 4 shows the default definitions for the hotkeys.
3. (Optional.) Display hotkeys.

   display hotkey

   This command is available in any view.

Table 4 Default definitions for hotkeys

Hotkey	Function or command
Ctrl+A	**move_the_cursor_to_the_beginning_of_the_line**: Moves the cursor to the beginning of a line.
Ctrl+B	**move_the_cursor_one_character_to_the_left**: Moves the cursor one character to the left.
Ctrl+C	**stop_the_current_command**: Stops the current command.
Ctrl+D	**erase_the_character_at_the_cursor**: Deletes the character at the cursor.
Ctrl+E	**move_the_cursor_to_the_end_of_the_line**: Moves the cursor to the end of a line.
Ctrl+F	**move_the_cursor_one_character_to_the_right**: Moves the cursor one character to the right.
Ctrl+G	**`display current-configuration`**: Display the running configuration.
Ctrl+H	**erase_the_character_to_the_left_of_the_cursor**: Deletes the character to the left of the cursor.
Ctrl+L	**`display ip routing-table`**: Display the IPv4 routing table information.
Ctrl+N	**display_the_next_command_in_the_history_buffer**: Displays the next command in the history buffer. Password configuration commands, if any, are skipped.
Ctrl+O	**`undo debugging all`**: Disable all debugging functions.
Ctrl+P	**display_the_previous_command_in_the_history_buffer**: Displays the previous command in the history buffer. Password configuration commands, if any, are skipped.
Ctrl+R	**redisplay_the_current_line**: Redisplays the current line.
Ctrl+T	N/A
Ctrl+U	N/A
Ctrl+W	**delete_the_word_to_the_left_of_the_cursor**: Deletes the word to the left of the cursor.
Ctrl+X	**delete_all_characters_from_the_beginning_of_the_line_to_the_cursor**: Deletes all characters to the left of the cursor.
Ctrl+Y	**delete_all_characters_from_the_cursor_to_the_end_of_the_line**: Deletes all characters from the cursor to the end of the line.
Ctrl+Z	**return_to_the_User_View**: Returns to user view.
Ctrl+]	**kill_incoming_connection_or_redirect_connection**: Terminates the current connection.
Esc+B	**move_the_cursor_back_one_word**: Moves the cursor back one word.
Esc+D	**delete_all_characters_from_the_cursor_to_the_end_of_the_word**: Deletes all characters from the cursor to the end of the word.
Esc+F	**move_the_cursor_forward_one_word**: Moves the cursor forward one word.

Enabling redisplaying entered-but-not-submitted commands

About this task

Your input might be interrupted by system information output. If redisplaying entered-but-not-submitted commands is enabled, the system redisplays your input after finishing the output. You can then continue entering the command line.

Procedure

1. Enter system view.

   `system-view`

2. Enable redisplaying entered-but-not-submitted commands.

   `info-center synchronous`

   By default, the system does not redisplay entered-but-not-submitted commands.

   For more information about this command, see *Network Management and Monitoring Command Reference*.

Understanding command-line syntax error messages

After you press **Enter** to submit a command, the command line interpreter examines the command syntax.

- If the command passes syntax check, the CLI executes the command.
- If the command fails syntax check, the CLI displays an error message.

Table 5 Common command-line syntax error messages

Syntax error message	Cause
% Unrecognized command found at '^' position.	The keyword in the marked position is invalid.
% Incomplete command found at '^' position.	One or more required keywords or arguments are missing.
% Ambiguous command found at '^' position.	The entered character sequence matches more than one command.
% Too many parameters found at '^' position.	The entered character sequence contains excessive keywords or arguments.
% Wrong parameter found at '^' position.	The argument in the marked position is invalid.

Using the command history feature

About command history buffers

The system automatically saves commands successfully executed by a login user to the following two command history buffers:

- Command history buffer for the user line.
- Command history buffer for all user lines.

Table 6 Comparison between the two types of command history buffers

Item	Command history buffer for a user line	Command history buffer for all user lines
Which commands are saved in the buffer?	Commands successfully executed by the current user of the user line.	Commands successfully executed by all login users.
Can commands in the buffer be displayed?	Yes.	Yes.
Can commands in the buffer be recalled?	Yes.	No.
Are buffered commands cleared when the user logs out?	Yes.	No.
Is the buffer size adjustable?	Yes.	No. The buffer size is fixed at 1024.

Command buffering rules

The system follows these rules when buffering commands:

- If you use incomplete keywords when entering a command, the system buffers the command in the exact form that you used.
- If you use an alias when entering a command, the system transforms the alias to the represented command or command keywords before buffering the command.
- If you enter a command in the same format multiple times in succession, the system buffers the command only once. If you enter a command in different formats multiple times, the system buffers each command format. For example, **`display cu`** and **`display current-configuration`** are buffered as two entries but successive repetitions of **`display cu`** create only one entry.
- To buffer a new command when a buffer is full, the system deletes the oldest command entry in the buffer.

Managing and using the command history buffers

Displaying the commands in command history buffers

To display the commands in command history buffers, execute the following commands in any view:

- Display the commands in command history buffers for a user line.
  `display history-command`
- Display the commands in command history buffers for all user lines.
  `display history-command all`

Recalling commands in the command history buffer for a user line

IMPORTANT:
Password configuration commands cannot be recalled.

Use up and down arrow keys to navigate to a command and press **Enter**.

Setting the size of the command history buffer for a user line

Use the **history-command max-size** command in user line or user line class view. For more information, see *Fundamentals Command Reference*.

Repeating commands in the command history buffer for a user line

About this task

You can recall and execute commands in the command history buffer for the current user line multiple times.

Restrictions and guidelines

The **repeat** command is available in any view. However, to repeat a command, you must first enter the view for the command. To repeat multiple commands, you must first enter the view for the first command.

The **repeat** command executes commands in the order they were executed.

The system starts a timer and waits for your interaction when it repeats an interactive command. If you do not provide the required information at prompt before the timer expires, the system skips the interactive command.

The system skips all password configuration commands.

Procedure

To repeat commands in the command history buffer for the current user line, execute the following command:

repeat [*number*] [**count** *times*] [**delay** *seconds*]

Controlling the CLI output

This section describes the CLI output control features that help you identify the desired output.

Pausing between screens of output

About this task

The device can automatically pause after displaying a specific number of lines if the output is too long to fit on one screen. At a pause, the device displays **----more----**. You can use the keys described in Table 7 to display more information or stop the display.

You can also disable pausing between screens of output for the current session. Then, all output is displayed at one time and the screen is refreshed continuously until the final screen is displayed.

Table 7 Output controlling keys

Keys	Function
Space	Displays the next screen.
Enter	Displays the next line.
Ctrl+C	Stops the display and cancels the command execution.
<PageUp>	Displays the previous page.
<PageDown>	Displays the next page.

Disabling pausing between screens of output

To disable pausing between screens of output, execute the following command in user view:

screen-length disable

The default depends on the settings of the **screen-length** command in user line view. The following are the default settings for the **screen-length** command:

- Pausing between screens of output is enabled.
- The maximum number of lines to be displayed at a time is 24.

For more information about the **screen-length** command, see *Fundamentals Command Reference*.

This command is a one-time command and takes effect only for the current CLI session.

Numbering each output line from a display command

About this task

For easy identification, you can use the **| by-linenum** option to display a number for each output line from a **display** command.

Each line number is displayed as a 5-character string and might be followed by a colon (:) or hyphen (-). If you specify both **| by-linenum** and **| begin** *regular-expression* for a **display** command, a hyphen is displayed for all lines that do not match the regular expression.

Procedure

To number each output line from a **display** command, execute the following command in any view:

display *command* **| by-linenum**

Example

Display information about VLAN 999, numbering each output line.

```
<Sysname> display vlan 999 | by-linenum
    1:  VLAN ID: 999
    2:  VLAN type: Static
    3:  Route interface: Configured
    4:  IPv4 address: 192.168.2.1
    5:  IPv4 subnet mask: 255.255.255.0
    6:  Description: For LAN Access
    7:  Name: VLAN 0999
    8:  Tagged ports:   None
    9:  Untagged ports: None
```

Filtering the output from a display command

About this task

You can use the [**|** { **count** | [**by-linenum**] { **begin** | **exclude** | **include** } *regular-expression* }]&<1-128> option to filter the output from a **display** command.

- You can use the option to specify a maximum of 128 filter conditions. The system displays only output lines that meet all the conditions.
- **count**—Displays the number of lines that match the specified regular expressions.

- **by-linenum**—Displays a number before each output line. You need to specify this keyword in only one filter condition.
- **begin**—Displays the first line matching the specified regular expression and all subsequent lines.
- **exclude**—Displays all lines not matching the specified regular expression.
- **include**—Displays all lines matching the specified regular expression.
- *regular-expression*—A case-sensitive string of 1 to 256 characters, which can contain the special characters described in Table 8.

Table 8 Special characters supported in a regular expression

Characters	Meaning	Examples
^	Matches the beginning of a line.	"^u" matches all lines beginning with "u". A line beginning with "Au" is not matched.
$	Matches the end of a line.	"u$" matches all lines ending with "u". A line ending with "uA" is not matched.
. (period)	Matches any single character.	".s" matches "as" and "bs".
*	Matches the preceding character or string zero, one, or multiple times.	"zo*" matches "z" and "zoo", and "(zo)*" matches "zo" and "zozo".
+	Matches the preceding character or string one or multiple times.	"zo+" matches "zo" and "zoo", but not "z".
|	Matches the preceding or succeeding string.	"def|int" matches a line containing "def" or "int".
( )	Matches the string in the parentheses, usually used together with the plus sign (+) or asterisk sign (*).	"(123A)" matches "123A". "408(12)+" matches "40812" and "408121212", but not "408".
\N	Matches the preceding strings in parentheses, with the *Nth* string repeated once.	"(string)\1" matches a string containing "stringstring". "(string1)(string2)\2" matches a string containing "string1string2string2". "(string1)(string2)\1\2" matches a string containing " string1string2string1string2".
[ ]	Matches a single character in the brackets.	"[16A]" matches a string containing 1, 6, or A; "[1-36A]" matches a string containing 1, 2, 3, 6, or A (- is a hyphen). To match the character "]", put it immediately after "[", for example, []abc]. There is no such limit on "[".
[^]	Matches a single character that is not in the brackets.	"[^16A]" matches a string that contains one or more characters except for 1, 6, or A, such as "abc". A match can also contain 1, 6, or A (such as "m16"), but it cannot contain these three characters only (such as 1, 16, or 16A).
{n}	Matches the preceding character *n* times. The number *n* must be a nonnegative integer.	"o{2}" matches "food", but not "Bob".
{n,}	Matches the preceding character *n* times or more. The number *n* must be a nonnegative integer.	"o{2,}" matches "foooood", but not "Bob".

Characters	Meaning	Examples
{n,m}	Matches the preceding character *n* to *m* times or more. The numbers *n* and *m* must be nonnegative integers and *n* cannot be greater than *m*.	" o{1,3}" matches "fod", "food", and "foooood", but not "fd".
\<	Matches a string that starts with the pattern following \<. A string that contains the pattern is also a match if the characters preceding the pattern are not digits, letters, or underscores.	"\<do" matches "domain" and "doa".
\>	Matches a string that ends with the pattern preceding \>. A string that contains the pattern is also a match if the characters following the pattern are not digits, letters, or underscores.	"do\>" matches "undo" and "cdo".
\b	Matches a word that starts with the pattern following \b or ends with the pattern preceding \b.	"er\b" matches "never", but not "verb" or "erase". "\ber" matches "erase", but not "verb" or "never".
\B	Matches a word that contains the pattern but does not start or end with the pattern.	"er\B" matches "verb", but not "never" or "erase".
\w	Same as [A-Za-z0-9_], matches a digit, letter, or underscore.	"v\w" matches "vlan" and "service".
\W	Same as [^A-Za-z0-9_], matches a character that is not a digit, letter, or underscore.	"\Wa" matches "-a", but not "2a" or "ba".
\	Escape character. If a special character listed in this table follows \, the specific meaning of the character is removed.	"\\" matches a string containing "\", "\^" matches a string containing "^", and "\\b" matches a string containing "\b".

Restrictions and guidelines

The required filtering time increases with the complexity of the regular expression. To abort the filtering process, press **Ctrl+C**.

If a regular expression contains spaces, you can select whether to enclose the regular expression with double quotation marks, for example, **display current-configuration | include interface A** or **display current-configuration | include "interface A"**.

The following rules apply when you specify regular expressions with or without double quotation marks:

- If a regular expression has double quotation marks, the string enclosed by double quotation marks is the filter condition.
- If a regular expression does not have double quotation marks, the string from the first non-space character to the last non-space character is the filter condition. The device uses this string as a filter condition for exact match of strings in the command output from a **display** command.

For example, *regular-expression*=**"interface A "** and *regular-expression*=**interface A** are different filter conditions.

If a regular expression contains **>**, **>**, or **> |**, enclose the regular expression with double quotation marks (") in case the device mistake these special characters as keywords.

Strings with single quotation marks (') are not supported.

If you specify multiple filter conditions, the system displays the output lines that meet all the conditions.

Examples

Display the running configuration, starting from the first configuration line that contains **line**. (The output depends on the device model and the current configuration.)

```
<Sysname> display current-configuration | begin line
line class console
 user-role network-admin
#
line class vty
 user-role network-operator
#
line con 0
 user-role network-admin
#
line vty 0 63
 authentication-mode none
 user-role network-admin
 user-role network-operator
#
...
```

Display brief information about interfaces in up state.

```
<Sysname> display interface brief | exclude DOWN
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP      Description
InLoop0              UP   UP(s)    --
NULL0                UP   UP(s)    --
Vlan1                UP   UP       192.168.1.83

Brief information on interfaces in bridge mode:
Link: ADM - administratively down; Stby - standby
Speed: (a) - auto
Duplex: (a)/A - auto; H - half; F - full
Type: A - access; T - trunk; H - hybrid
Interface            Link Speed    Duplex Type PVID Description
GE1/0/1              UP   1G(a) F(a)  A    1
```

Display SNMP-related running configuration lines.

```
<Sysname> display current-configuration | include snmp
snmp-agent
 snmp-agent community write private
 snmp-agent community read public
 snmp-agent sys-info version all
 snmp-agent target-host trap address udp-domain 192.168.1.26 params securityname public
```

Display log entries in the log buffer that contain both **SHELL** and **VTY**.

```
<Sysname> display logbuffer | include SHELL | include VTY
%Sep  6 10:38:12:320 2018 Sysname SHELL/5/SHELL_LOGIN: VTY logged in from 169.254.100.171.
```

```
%Sep  6 10:52:32:576 2018 Sysname SHELL/5/SHELL_LOGOUT: VTY logged out from
169.254.100.171.
%Sep 6 16:03:27:100 2018 Sysname SHELL/5/SHELL_LOGIN: VTY logged in from 169.254.100.171.
%Sep  6 16:44:18:113 2018 Sysname SHELL/5/SHELL_LOGOUT: VTY logged out from
169.254.100.171.
```

Displaying information for the peer member device

About this task

In a network configured with multichassis configuration synchronization, perform this task on the local member device to view information for the peer member device as follows:

- To view information for only the peer member device, execute the **display** *command* **peer-device** command on the local device. This facilitates device management and maintenance.
- To view information for both member devices at the same time, execute the **display** *command* **with-peer-device** command. This is convenient for comparison between two member devices.

For more information about multichassis configuration synchronization, see "Configuring multichassis configuration synchronization."

Procedure

To view information for only the peer member device or information for both local and peer member devices, execute the following command in any view:

display { **peer-device** | **with-peer-device** }

Configuration examples

Display the running configuration in user view on both the local and peer member devices.

```
<Sysname> display this with-peer-device
#
sysname sysname
#
system-working-mode standard
xbar load-single
password-recovery enable
lpu-type f-series
#
m-lag mad exclude interface GigabitEthernet1/0/1
m-lag system-mac 0001-0001-0001
m-lag system-number 1
m-lag system-priority 123
m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
#
scheduler logfile size 16
#
domain default enable system
#
security-enhanced level 2
#
return
```

```
-----------------------  peer information  -------------------------
#
sysname sysname
#
system-working-mode standard
xbar load-single
password-recovery enable
lpu-type f-series
#
m-lag mad exclude interface GigabitEthernet1/0/1
m-lag system-mac 0001-0001-0001
m-lag system-number 2
m-lag system-priority 123
m-lag keepalive ip destination 1.1.1.2 source 1.1.1.1
#
scheduler logfile size 16
#
domain default enable system
#
security-enhanced level 2
#
return
```

Saving the output from a display command to a file

About this task

A **display** command shows certain configuration and operation information for the device. Its output might vary over time or with user configuration or operation. You can save the output to a file for future retrieval or troubleshooting.

Use one of the following methods to save the output from a **display** command:

- Save the output to a separate file. Use this method if you want to use one file for a single **display** command.
- Append the output to the end of a file. Use this method if you want to use one file for multiple **display** commands.

Procedure

To save the output from a **display** command to a file, use one of the following commands in any view:

- Save the output from a **display** command to a separate file.

  display *command* **>** *filename*
- Append the output from a **display** command to the end of a file.

  display *command* **>>** *filename*

Examples

Save the VLAN 1 settings to a separate file named **vlan.txt**.

```
<Sysname> display vlan 1 > vlan.txt
```

Verify that the VLAN 1 settings are saved to the file **vlan.txt**.

```
<Sysname> more vlan.txt
```

```
VLAN ID: 1
 VLAN type: Static
 Route interface: Not configured
 Description: VLAN 0001
 Name: VLAN 0001
 Tagged ports:   None
 Untagged ports: None
```

Append the VLAN 999 settings to the end of the file **vlan.txt**.

```
<Sysname> display vlan 999 >> vlan.txt
```

Verify that the VLAN 999 settings are appended to the end of the file **vlan.txt**.

```
<Sysname> more vlan.txt
VLAN ID: 1
 VLAN type: Static
 Route interface: Not configured
 Description: VLAN 0001
 Name: VLAN 0001
 Tagged ports:   None
 Untagged ports: None
 VLAN ID: 999
 VLAN type: Static
 Route interface: Configured
 IP address: 192.168.2.1
 Subnet mask: 255.255.255.0
 Description: For LAN Access
 Name: VLAN 0999
 Tagged ports:   None
 Untagged ports: None
```

Viewing and managing the output from a display command effectively

You can use the following methods in combination to filter and manage the output from a **display** command:

- Numbering each output line from a display command
- Filtering the output from a display command
- Saving the output from a display command to a file

Procedure

To use multiple measures to view and manage the output from a **display** command effectively, execute the following command in any view:

display *command* [**|** [**by-linenum**] { **begin** | **exclude** | **include** } *regular-expression*]&<1-128> [**>** *filename* | **>>** *filename*]

Examples

Save the running configuration to a separate file named **test.txt**, with each line numbered.

```
<Sysname> display current-configuration | by-linenum > test.txt
```

Append lines including **snmp** in the running configuration to the file **test.txt**.

```
<Sysname> display current-configuration | include snmp >> test.txt
```

Display the first line that begins with **user-group** in the running configuration and all the following lines.

```
<Sysname> display current-configuration | by-linenum begin user-group
  114:  user-group system
  115-  #
  116-  return
```

// The colon (:) following a line number indicates that the line contains the string user-group. The hyphen (-) following a line number indicates that the line does not contain the string **user-group**.

Enabling reauthentication for disruptive commands

About this task

Disruptive commands might power off or restart the device, interrupt services, cause service anomaly, delete key files, clear configuration, cause user login failure, or make users go offline, for example, the **reboot** command. Only the qualified professionals can execute disruptive commands.

As a best practice to avoid service failure due to misoperation, enable reauthentication for disruptive commands.

The following rules apply if you enable reauthentication for disruptive commands:

- Authentication-free users can execute disruptive commands without reauthentication.
- Non-authentication-free users must enter the login password for re-authentication when they execute disruptive commands.

Procedure

1. Enter system view.

   system-view

2. Enable reauthentication for disruptive commands.

   configuration reauthentication enable

   By default, reauthentication for disruptive commands is disabled.

Entering MMI mode

About this task

In machine-machine interaction (MMI) mode, interactive commands that require confirmation by pressing Y/N, Y/N/C, or Enter in human-machine interaction mode do not require any confirmation. For example, when the **reboot** command is executed in human-machine interaction mode, you must press Y/N to proceed with the reboot operation. In MMI mode, the reboot operation is completed without requiring any confirmation after you execute the **reboot** command.

Restrictions and guidelines

If pausing between screens of output for the current CLI session has been enabled before you enter MMI mode, all output is displayed at a time and the screen is refreshed continuously until the final screen is displayed.

After you configure the **configuration reauthentication enable** command, disruptive commands that require reauthentication do not need to be confirmed again.

Commands that require strings to be input for execution still require manual input and confirmation.

Procedure

Execute the following command in user view.

`terminal mmi-mode enable`

By default, the user is in human-machine interaction mode.

Contents

Configuring RBAC 1

About RBAC 1
Permission assignment 1
User role assignment 4
FIPS compliance 4
RBAC tasks at a glance 4
Creating a user role 5
Configuring user role rules 6
Configuring a feature group 7
Configuring resource access policies 8
About resource access policies 8
Restrictions and guidelines for resource access policy configuration 8
Configuring the interface policy for a user role 8
Configuring the VLAN policy for a user role 8
Configuring the VPN instance policy for a user role 9
Assigning user roles 9
Restrictions and guidelines for user role assignment 9
Enabling the default user role feature 9
Assigning user roles to remote AAA authentication users 10
Assigning user roles to local AAA authentication users 10
Assigning user roles to non-AAA authentication users on user lines 11
Configuring temporary user role authorization 11
About temporary user role authorization 11
Restrictions and guidelines for temporary user role authorization 12
Setting the authentication mode for temporary user role authorization 13
Specifying the default target user role for temporary user role authorization 13
Setting an authentication password for temporary user role authorization 13
Automatically obtaining the login username for temporary user role authorization 14
Obtaining temporary user role authorization 14
Enabling SNMP notification for user role changes 15
Display and maintenance commands for RBAC 15
RBAC configuration examples 16
Example: Configuring RBAC for local AAA authentication users 16
Example: Configuring RBAC for RADIUS authentication users 17
Example: Configuring RBAC temporary user role authorization (HWTACACS authentication) 20
Example: Configuring RBAC temporary user role authorization (RADIUS authentication) 24
Troubleshooting RBAC 27
Local users have more access permissions than intended 27
Login attempts by RADIUS users always fail 27

Configuring RBAC

About RBAC

Role-based access control (RBAC) controls access permissions of users based on user roles.

RBAC assigns access permissions to user roles that are created for different job functions. Users are given permission to access a set of items and resources based on the users' user roles. Separating permissions from users enables simple permission authorization management.

Permission assignment

Use the following methods to assign permissions to a user role:

- Define a set of rules to determine accessible or inaccessible items for the user role. (See "User role rules.")
- Configure resource access policies to specify which resources are accessible to the user role. (See "Resource access policies.")

To use a command related to a system resource, a user role must have access to both the command and the resource.

For example, a user role has access to the **vlan** command and access only to VLAN 10. When the user role is assigned, you can use the **vlan** command to create VLAN 10 and enter its view. However, you cannot create any other VLANs. If the user role has access to VLAN 10 but does not have access to the **vlan** command, you cannot use the command to enter the view of VLAN 10.

When a user logs in to the device with any user role and enters **<?>** in a view, help information is displayed for the system-defined command aliases in the view. However, the user might not have the permission to access the command aliases. Whether the user can access the command aliases depends on the user role's permission to the commands corresponding to the aliases. For information about command aliases, see "Using the CLI."

A user that logs in to the device with any user role has access to the **system-view**, **quit**, and **exit** commands.

User role rules

User role rules permit or deny access to the items, including commands, XML elements, or MIB nodes. You can define the following types of rules for different access control granularities:

- **Command rule**—Controls access to a command or a set of commands that match a regular expression.
- **Feature rule**—Controls access to the commands of a feature by command type.
- **Feature group rule**—Controls access to the commands of features in a feature group by command type.
- **XML element rule**—Controls access to XML elements used for configuring the device.
- **OID rule**—Controls SNMP access to a MIB node and its child nodes. An OID is a dotted numeric string that uniquely identifies the path from the root node to a leaf node.

The items (commands, XML elements, and MIB nodes) are controlled based on the following types:

- **Read**—Items that display configuration and maintenance information. For example, the **display** commands and the **dir** command.
- **Write**—Items that configure the features in the system. For example, the **info-center enable** command and the **debugging** command.

- **Execute**—Items that execute specific functions. For example, the **ping** command and the **ftp** command.

A user role can access the set of permitted items specified in the user role rules. The user role rules include predefined (identified by sys-*n*) and user-defined user role rules. For more information about the user role rule priority, see "Configuring user role rules."

Resource access policies

Resource access policies control access of a user role to system resources and include the following types:

- **Interface policy**—Controls access to interfaces.
- **VLAN policy**—Controls access to VLANs.
- **VPN instance policy**—Controls access to VPN instances.

Resource access policies do not control access to the interface, VPN instance, or VLAN options in the **display** commands. You can specify these options in the **display** commands if the options are permitted by any user role rule.

Predefined user roles

The system provides predefined user roles. These user roles have access to all system resources. However, their access permissions differ, as shown in Table 1.

Among all of the predefined user roles, only network-admin, system-admin, and level-15 allow users to create, modify, and delete local users and local user groups. The security-admin user role allows users to change their own passwords and create, modify, and delete network access users. The level-0 to level-14 user roles and user-defined user roles only allow users to modify their own passwords and create, modify, and delete network access users if they have permissions to configure local users and local user groups. They cannot create, modify, or delete device management users or user groups. The other user roles only allow users to modify their own passwords if they have permissions to configure local users and local user groups.

The access permissions of the level-0 to level-14 user roles can be modified through user role rules and resource access policies. However, you cannot make changes on the predefined access permissions of these user roles. For example, you cannot change the access permission of these user roles to the **display history-command all** command.

Table 1 Predefined roles and permissions matrix

User role name	Permissions
network-admin	Accesses all features and resources in the system, except for the **display security-logfile summary**, **info-center security-logfile directory**, and **security-logfile save** commands.
network-operator	• Accesses the **display** commands for features and resources in the system. To display all accessible commands of the user role, use the **display role** command. • Enables local authentication login users to change their own passwords. • Accesses the command used for entering XML view. • Accesses all read-type XML elements. • Accesses all read-type MIB nodes. • Cannot execute the **delete**, **reset**, **logbuffer**, **more**, or **display logbuffer** commands.
level-*n* (*n* = 0 to 15)	• **level-0**—Has access to commands including **ping**, **tracert**, **ssh2**, **telnet**, **mtrace**, and **super**. Level-0 access rights are configurable. • **level-1**—Has access to the **display** commands of features and

User role name	Permissions
	resources in the system. The **level-1** user role also has all access rights of the level-0 user role. Level-1 access rights are configurable. • **level-2 to level-8, and level-10 to level-14**—Have no access rights by default. Access rights are configurable. • **level-9**—Has access to most of the features and resources in the system. If you are logged in with a local user account that has a level-9 user role, you can change the password in the local user account. The following are the major features and commands that the level-9 user role cannot access: ○ RBAC non-debugging commands. ○ Local users. ○ File management. ○ Device management. ○ The **`display history-command all`** command. • **level-15**—Has the same rights as network-admin.
security-audit	Security log manager. The user role has the following access rights to security log files: • Accesses the commands for displaying and maintaining security log files (for example, the **`dir`**, **`display security-logfile summary`**, and **`more`** commands). • Accesses the commands for managing security log files and security log file system (for example, the **`info-center security-logfile directory`**, **`mkdir`**, and **`security-logfile save`** commands). For more information about security log management, see *Network Management and Monitoring Configuration Guide*. For more information about file system management, see "Managing file systems." **IMPORTANT:** Only the security-audit user role has access to security log files. You cannot assign the security-audit user role to non-AAA authentication users on a user line.
system-admin	System administrator. The user role has the following access permissions: • Displays and configures user roles. • Displays, deletes, and configures local users. • Configures VTY user lines. • Enables or disables the Stelnet server. • Executes the **`save`** command. • Executes the **`reboot`** command. **IMPORTANT:** You cannot assign the system-admin user role to non-AAA authentication users on a user line.
security-admin	Security administrator. The user role has access to most of the commands in the system except for the following commands: • Commands used to display local users. • Commands used to display and configure user roles. • Commands used to configure VTY user lines. • Command used to enable the Stelnet server. • The **`save`** and **`reboot`** commands. • Information center-related commands. To display the detailed access permissions of this user role, use the

User role name	Permissions
	display role command. **IMPORTANT:** You cannot assign the security-admin user role to non-AAA authentication users on a user line.
audit-admin	Audit administrator. he user role has access to information center-related commands. To display the detailed access permissions of this user role, use the **display role** command. **IMPORTANT:** You cannot assign the audit-admin user role to non-AAA authentication users on a user line.

User role assignment

You assign access rights to a user by assigning a minimum of one user role. The user can use the collection of items and resources accessible to all user roles assigned to the user. For example, you can access any interface to use the **qos apply policy** command if you are assigned the following user roles:

- User role A denies access to the **qos apply policy** command and permits access only to interface GigabitEthernet 1/0/1.
- User role B permits access to the **qos apply policy** command and all interfaces.

Depending on the authentication method, user role assignment has the following methods:

- **AAA authorization**—If scheme authentication is used, the AAA module handles user role assignment.
  - If the user passes local authorization, the device assigns the user roles specified in the local user account.
  - If the user passes remote authorization, the remote AAA server assigns the user roles specified on the server. The AAA server can be a RADIUS or HWTACACS server.
- **Non-AAA authorization**—When the user accesses the device without authentication or by passing password authentication on a user line, the device assigns user roles specified on the user line. This method also applies to SSH clients that use publickey or password-publickey authentication. User roles assigned to these SSH clients are specified in their respective device management user accounts.

For more information about AAA and SSH, see *Security Configuration Guide*. For more information about user lines, see "Login overview" and "Configuring CLI login."

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

RBAC tasks at a glance

To configure RBAC, perform the following tasks:

1. Creating a user role

2. Configuring user role rules
3. (Optional.) Configuring a feature group
4. Configuring resource access policies
   - Configuring the interface policy for a user role
   - Configuring the VLAN policy for a user role
   - Configuring the VPN instance policy for a user role
5. Assigning user roles
   - Enabling the default user role feature
   - Assigning user roles to remote AAA authentication users
   - Assigning user roles to local AAA authentication users
   - Assigning user roles to non-AAA authentication users on user lines
6. Configuring temporary user role authorization
   a. Setting the authentication mode for temporary user role authorization
   b. Specifying the default target user role for temporary user role authorization
   c. Setting an authentication password for temporary user role authorization
   d. (Optional.) Automatically obtaining the login username for temporary user role authorization
   e. Obtaining temporary user role authorization
   f. (Optional.) Enabling SNMP notification for user role changes

Creating a user role

About this task

In addition to the predefined user roles, you can create a maximum of 64 custom user roles for granular access control.

Procedure

1. Enter system view.

   `system-view`

2. Create a user role and enter its view.

   `role name` *`role-name`*

   By default, the system has the following predefined user roles:
   - network-admin.
   - network-operator.
   - level-*n* (where *n* equals an integer in the range of 0 to 15).
   - security-audit.
   - system-admin.
   - security-admin.
   - audit-admin.

   Among these user roles, only the permissions and descriptions of the level-0 to level-14 user roles are configurable.

3. (Optional.) Configure a description for the user role.

   `description` *`text`*

   By default, a user role does not have a description.

Configuring user role rules

About this task

You can configure user role rules to permit or deny the access of a user role to specific commands, XML elements, and MIB nodes.

The following guidelines apply to non-OID rules:

- If two user-defined rules of the same type conflict, the rule with the higher ID takes effect. For example, a user role can use the **tracert** command but not the **ping** command if the user role contains rules configured by using the following commands:
  - **rule** *1* **permit command** *ping*
  - **rule** *2* **permit command** *tracert*
  - **rule** *3* **deny command** *ping*
- If a predefined user role rule and a user-defined user role rule conflict, the user-defined user role rule takes effect.

The following guidelines apply to OID rules:

- The system compares an OID with the OIDs specified in user role rules, and it uses the longest match principle to select a rule for the OID. For example, a user role cannot access the MIB node with OID 1.3.6.1.4.1.25506.141.3.0.1 if the user role contains rules configured by using the following commands:
  - **rule** *1* **permit read write oid** *1.3.6*
  - **rule** *2* **deny read write oid** *1.3.6.1.4.1*
  - **rule** *3* **permit read write oid** *1.3.6.1.4*
- If the same OID is specified in multiple rules, the rule with the higher ID takes effect. For example, a user role can access the MIB node with OID 1.3.6.1.4.1.25506.141.3.0.1 if the user role contains rules configured by using the following commands:
  - **rule** *1* **permit read write oid** *1.3.6*
  - **rule** *2* **deny read write oid** *1.3.6.1.4.1*
  - **rule** *3* **permit read write oid** *1.3.6.1.4.1*

Restrictions and guidelines

- Only the network-admin, and level-15 user roles have access to all of the following commands:

IMPORTANT:

The other predefined user roles, except for those mentioned above, have access to only some of the following commands. For more information, see the command output of the **display role** command.

User-defined user roles cannot access any of the following commands.

  - The **display history-command all** command.
  - All commands that start with the **display role**, **reboot**, **startup saved-configuration**, and **undo startup saved-configuration** keywords.
  - All commands that start with the **role**, **undo role**, **super**, **undo super**, **password-recovery**, and **undo password-recovery** keywords in system view.
  - All commands that start with the **snmp-agent community**, **undo snmp-agent community**, **snmp-agent usm-user**, **undo snmp-agent usm-user**, **snmp-agent group**, and **undo snmp-agent group** keywords in system view.

- All commands that start with the **user-role**, **undo user-role**, **authentication-mode**, **undo authentication-mode**, **set authentication password**, and **undo set authentication password** keywords in user line view or user line class view.
- All commands that start with the **user-role** and **undo user-role** keywords in schedule view or in CLI-defined policy view.
- All commands of the event MIB feature.

- You can configure a maximum of 256 user-defined rules for a user role. The total number of user-defined user role rules cannot exceed 1024.
- Any rule modification, addition, or removal for a user role takes effect only on users who are logged in with the user role after the change.

Procedure

1. Enter system view.

   system-view
2. Enter user role view.

   role name *role-name*
3. Configure rules for the user role. Choose the options to configure as needed:
   - Configure a command rule.

     rule *number* { **deny** | **permit** } **command** *command-string*
   - Configure a feature rule.

     rule *number* { **deny** | **permit** } { **execute** | **read** | **write** } * **feature** [*feature-name*]
   - Configure a feature group rule.

     rule *number* { **deny** | **permit** } { **execute** | **read** | **write** } * **feature-group** *feature-group-name*

     A feature group rule takes effect only after the feature group is created.
   - Configure an XML element rule.

     rule *number* { **deny** | **permit** } { **execute** | **read** | **write** } * **xml-element** [*xml-string*]
   - Configure an OID rule.

     rule *number* { **deny** | **permit** } { **execute** | **read** | **write** } * **oid** *oid-string*

Configuring a feature group

About this task

Use feature groups to bulk assign command access permissions to sets of features. In addition to the predefined feature groups, you can create a maximum of 64 custom feature groups and assign a feature to multiple feature groups.

Procedure

1. Enter system view.

   system-view
2. Create a feature group and enter its view.

   role feature-group name *feature-group-name*

   By default, the system has the following predefined feature groups, which cannot be deleted or modified:

- **L2**—Includes all Layer 2 commands.
- **L3**—Includes all Layer 3 commands.

3. Add a feature to the feature group.

   feature *feature-name*

   By default, a feature group does not have any feature.

Configuring resource access policies

About resource access policies

Every user role has one interface policy, VPN instance policy, and VLAN policy. By default, these policies permit a user role to access any system resources. You can configure the policies of a user-defined user role or a predefined level-*n* user role to limit its access to any resources.

Restrictions and guidelines for resource access policy configuration

The policy configuration takes effect only on users who are logged in with the user role after the configuration.

Configuring the interface policy for a user role

1. Enter system view.

   system-view

2. Enter user role view.

   role name *role-name*

3. Enter user role interface policy view.

   interface policy deny

   By default, the interface policy for a user role permits access to all interfaces.

CAUTION:

This command denies the access of the user role to any interfaces if you do not specify accessible interfaces by using the **permit interface** command.

4. (Optional.) Specify a list of interfaces accessible to the user role.

   permit interface *interface-list*

   By default, no accessible interfaces are configured in user role interface policy view.

   Repeat this step to add additional accessible interfaces.

Configuring the VLAN policy for a user role

1. Enter system view.

   system-view

2. Enter user role view.

   role name *role-name*

3. Enter user role VLAN policy view.

vlan policy deny

By default, the VLAN policy for a user role permits access to all VLANs.

CAUTION:

This command denies the access of the user role to any VLANs if you do not specify accessible VLANs by using the **permit vlan** command.

4. (Optional.) Specify a list of VLANs accessible to the user role.

   permit vlan *vlan-id-list*

   By default, no accessible VLANs are configured in user role VLAN policy view.

   Repeat this step to add additional accessible VLANs.

Configuring the VPN instance policy for a user role

1. Enter system view.

   system-view

2. Enter user role view.

   role name *role-name*

3. Enter user role VPN instance policy view.

   vpn-instance policy deny

   By default, the VPN instance policy for a user role permits access to all VPN instances.

CAUTION:

This command denies the access of the user role to any VPN instances if you do not specify accessible VPN instances by using the **permit vpn-instance** command.

4. (Optional.) Specify a list of VPN instances accessible to the user role.

   permit vpn-instance *vpn-instance-name*&<1-10>

   By default, no accessible VPN instances are configured in user role VPN instance policy view.

   Repeat this step to add additional accessible VPN instances.

Assigning user roles

Restrictions and guidelines for user role assignment

To control user access to the system, you must assign a minimum of one user role. Make sure a minimum of one user role among the user roles assigned by the server exists on the device.

Enabling the default user role feature

About this task

The default user role feature assigns the default user role to AAA-authenticated users if the authentication server (local or remote) does not assign any user roles to the users. These users are allowed to access the system with the default user role.

You can specify any user role existing in the system as the default user role.

Procedure

1. Enter system view.

system-view

2. Enable the default user role feature.

   role default-role enable [*role-name*]

   By default, the default user role feature is disabled.

   If you do not use the **authorization-attribute user role** command to assign user roles to local users, you must enable the default user role feature. For information about the **authorization-attribute user role** command, see AAA commands in *Security Command Reference*.

Assigning user roles to remote AAA authentication users

For remote AAA authentication users, user roles are configured on the remote authentication server. For information about configuring user roles for RADIUS users, see the RADIUS server documentation. For HWTACACS users, the role configuration must use the **roles="***role-1 role-2 … role-n***"** format, where user roles are space separated. For example, configure **roles="level-0 level-1 level-2"** to assign level-0, level-1, and level-2 to an HWTACACS user.

If the AAA server assigns the security-audit user role and other user roles to the same user, only the security-audit user role takes effect.

Assigning user roles to local AAA authentication users

About this task

Configure user roles for local AAA authentication users in their local user accounts. For information about AAA and local user configuration, see AAA configuration in *Security Configuration Guide*.

Restrictions and guidelines

- Every local user has a default user role. If this default user role is not suitable, remove it.
- If a local user is the only user with the security-audit user role, the user cannot be deleted.
- The security-audit user role is mutually exclusive with other user roles.
  - When you assign the security-audit user role to a local user, the system requests confirmation to remove all the other user roles from the user.
  - When you assign the other user roles to a local user who has the security-audit user role, the system requests confirmation to remove the security-audit role from the user.
- The system-admin, security-admin, and audit-admin user roles are mutually exclusive in a user account. In addition, these user roles are mutually exclusive with other user roles in a user account. When you assign user roles to a user, the system prompts you to confirm the deletion of the user roles that are mutually exclusive with the new user roles.
- You can assign a maximum of 64 user roles to a local user.

Procedure

1. Enter system view.

   system-view

2. Create a local user and enter its view.

   local-user *user-name* **class** { **manage** | **network** }

3. Assign a user role to the local user.

   authorization-attribute user-role *role-name*

   By default, the network-operator user role is assigned to local users created by a network-admin or level-15 user.

Assigning user roles to non-AAA authentication users on user lines

About this task

Specify user roles for the following two types of login users on the user lines:

- Non-SSH users that use password authentication or no authentication.
- SSH clients that use publickey or password-publickey authentication. User roles assigned to these SSH clients are specified in their respective device management user accounts.

For more information about user lines, see "Login overview" and "Configuring CLI login." For more information about SSH, see *Security Configuration Guide*.

Restrictions and guidelines

- You can assign a maximum of 64 user roles to a non-AAA authentication user on a user line.
- You cannot assign the security-audit user role to non-AAA authentication users on user lines.
- The device cannot assign the following user roles to non-AAA authentication users on user lines:
  - system-admin.
  - security-admin.
  - audit-admin.

Procedure

1. Enter system view.

   system-view
2. Enter user line view or user line class view.
   - Enter user line view.

     line { *first-num1* [*last-num1*] | { **console** | **vty** } *first-num2* [*last-num2*] }
   - Enter user line class view.

     line class { **console** | **vty** }

   For information about the priority order and application scope of the settings in user line view and user line class view, see "Configuring CLI login."
3. Specify a user role on the user line.

   user-role *role-name*

   By default, the network-admin user role is specified on the console user line, and the network-operator user role is specified on any other user line.

Configuring temporary user role authorization

About temporary user role authorization

Temporary user role authorization allows you to obtain another user role without reconnecting to the device. This feature is useful when you want to use a user role temporarily to configure a feature.

Temporary user role authorization is effective only on the current login. This feature does not change the user role settings in the user account that you have been logged in with. The next time you are logged in with the user account, the original user role settings take effect.

To enable a user to obtain another user role without reconnecting to the device, you must configure user role authentication. Table 2 describes the available authentication modes and configuration requirements.

Table 2 User role authentication modes

Keywords	Authentication mode	Description
local	Local password authentication only (local-only)	The device uses the locally configured password for authentication. If no local password is configured for a user role in this mode, a console user can obtain the user role by either entering a string or not entering anything.
scheme	Remote AAA authentication through HWTACACS or RADIUS (remote-only)	The device sends the username and password to the HWTACACS or RADIUS server for remote authentication. To use this mode, you must perform the following configuration tasks: • Configure the required HWTACACS or RADIUS scheme, and configure the ISP domain to use the scheme for the user. For more information, see *Security Configuration Guide*. • Add the user account and password on the HWTACACS or RADIUS server.
local scheme	Local password authentication first, and then remote AAA authentication (local-then-remote)	Local password authentication is performed first. If no local password is configured for the user role in this mode: • The device performs remote AAA authentication for VTY users. • A console user can obtain another user role by either entering a string or not entering anything.
scheme local	Remote AAA authentication first, and then local password authentication (remote-then-local)	Remote AAA authentication is performed first. Local password authentication is performed in either of the following situations: • The HWTACACS or RADIUS server does not respond. • The remote AAA configuration on the device is invalid.

Restrictions and guidelines for temporary user role authorization

If HWTACACS authentication is used, the following rules apply:

- If the device is not enabled to automatically obtain the login username as the authentication username, you must enter a username to request role authentication.
- The device sends the username to the server in the *username or username@domain-name* format. Whether the domain name is included in the username depends on the **user-name-format** command in the HWTACACS scheme.
- To obtain a level-*n* user role, the user account on the server must have the target user role level or a level higher than the target user role. A user account that obtains the level-*n* user role can obtain any user role among level-0 through level-*n*.
- To obtain a non-level-*n* user role, make sure the user account on the server meets the following requirements:
  - The account has a user privilege level.

- The HWTACACS custom attribute is configured for the account in the form of **allowed-roles="***role***"**. The variable *role* represents the target user role.

If RADIUS authentication is used, the following rules apply:

- The device does not use the username you enter or the automatically obtained login username to request user role authentication. It uses a username in the **$enab*n*$** format. The variable *n* represents a user role level, and a domain name is not included in the username. You can always pass user role authentication when the password is correct.
- To obtain a level-*n* user role, you must create a user account for the level-*n* user role in the **$enabn$** format on the RADIUS server. The variable *n* represents the target user role level. For example, to obtain the level-3 user role, you can enter any username. The device uses the username **$enab3$** to request user role authentication from the server.
- To obtain a non-level-*n* user role, you must perform the following tasks:
  - Create a user account named **$enab0$** on the server.
  - Configure the cisco-av-pair attribute for the account in the form of **allowed-roles="***role***"**. The variable *role* represents the target user role.

The device selects an authentication domain for user role authentication in the following order:

1. The ISP domain included in the entered username.
2. The default ISP domain.

If you execute the **quit** command after obtaining user role authorization, you are logged out of the device.

Setting the authentication mode for temporary user role authorization

1. Enter system view.
   system-view
2. Set the authentication mode.
   super authentication-mode { **local** | **scheme** } *
   By default, local-only authentication applies.

Specifying the default target user role for temporary user role authorization

1. Enter system view.
   system-view
2. Specify the default target user role for temporary user role authorization.
   super default role *role-name*
   By default, the default target user role is network-admin.

Setting an authentication password for temporary user role authorization

About this task

Authentication passwords are required only for local password authentication.

Procedure

1. Enter system view.

   system-view

2. Set a local authentication password for a user role.

   In non-FIPS mode:

   super password [**role** *role-name*] [{ **hash** | **simple** } *string*]

   In FIPS mode:

   super password [**role** *role-name*]

   By default, no password is set.

   If you do not specify the **role** *role-name* option, the command sets a password for the default target user role.

Automatically obtaining the login username for temporary user role authorization

About this task

This feature is applicable only to the login from a user line that uses scheme authentication, which requires a username for login. This feature enables the device to automatically obtain the login username when the login user requests a temporary user role authorization from a remote authentication server.

Restrictions and guidelines

If the user was logged in from a user line that uses password authentication or no authentication, the device cannot obtain the login username. The request for temporary user role authorization from a remote authentication server will fail.

This feature does not take effect on local password authentication for temporary user role authorization.

Procedure

1. Enter system view.

   system-view

2. Enable the device to automatically obtain the login username when a login user requests temporary user role authorization from a remote authentication server.

   super use-login-username

   By default, the device prompts for a username when a login user requests temporary user role authorization from a remote authentication server.

Obtaining temporary user role authorization

Restrictions and guidelines

The operation of obtaining temporary user role authorization fails after three consecutive unsuccessful authentication attempts.

You might fail to switch to a non-level-*n* user role if both of the following conditions exist:

- User role switching authentication is performed in the same ISP domain as the current login user.
- User role switching authentication uses a different AAA method than the login authorization method configured for the ISP domain.

To resolve this issue, make sure the AAA methods configured by using the **authentication super** command are consistent with those configured by using the **authorization login** command for the ISP domain.

For more information about AAA, see *Security Configuration Guide*.

Prerequisites

Before you obtain temporary user role authorization, make sure the current user account has the permission to execute the **super** command to obtain temporary user role authorization.

Procedure

To obtain the temporary authorization to use a user role, execute the following command in user view:

super [*role-name*]

If you do not specify the *role-name* argument, you obtain the default target user role for temporary user role authorization.

Enabling SNMP notification for user role changes

About this task

After you enable SNMP notification for user role changes, the following types of SNMP notifications will be generated:

- An SNMP notification to report a successful user role change if you use the **super** command to change a user role successfully.
- An SNMP notification to report a failed user role change if you fail to use the **super** command to change a user role. Typically, you must enter the user role change password for changing a user role. The user role change operation fails if the authentication server does not respond or no user role change password is configured.

Procedure

1. Enter system view.

   system-view
2. Enable SNMP notification for user role changes.

   snmp-agent trap enable super [**change-failure-warning** | **change-succesful-warning**] *

   By default, SNMP notification for user role changes is disabled.

Display and maintenance commands for RBAC

Execute **display** commands in any view.

Task	Command
Display user role information.	**display role** [**name** *role-name*]
Display user role feature information.	**display role feature** [**name** *feature-name* | **verbose**]
Display user role feature group information.	**display role feature-group** [**name** *feature-group-name*] [**verbose**]

RBAC configuration examples

Example: Configuring RBAC for local AAA authentication users

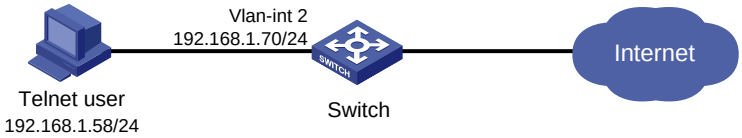
Network configuration

As shown in Figure 1, the switch performs local AAA authentication for the Telnet user. The user account for the Telnet user is **user1@bbb**, which is assigned user role **role1**.

Configure **role1** to have the following permissions:

- Execute the read commands of any feature.
- Access VLANs 10 to 20. Access to any other VLANs is denied.

Figure 1 Network diagram



Procedure

Assign an IP address to VLAN-interface 2 (the interface connected to the Telnet user).

```
<Switch> system-view
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.1.70 255.255.255.0
[Switch-Vlan-interface2] quit
```

Enable the Telnet server.

```
[Switch] telnet server enable
```

Enable scheme authentication on the user lines for Telnet users.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Enable local authentication and authorization for ISP domain **bbb**.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login local
[Switch-isp-bbb] authorization login local
[Switch-isp-bbb] quit
```

Create a user role named **role1**.

```
[Switch] role name role1
```

Configure rule 1 to permit the user role to access the read commands of all features.

```
[Switch-role-role1] rule 1 permit read feature
```

Configure rule 2 to permit the user role to create VLANs and access commands in VLAN view.

```
[Switch-role-role1] rule 2 permit command system-view ; vlan *
```

Change the VLAN policy to permit the user role to configure only VLANs 10 to 20.

```
[Switch-role-role1] vlan policy deny
[Switch-role-role1-vlanpolicy] permit vlan 10 to 20
[Switch-role-role1-vlanpolicy] quit
```

```
[Switch-role-role1] quit
```

Create a device management user named **user1** and enter local user view.

```
[Switch] local-user user1 class manage
```

Set a plaintext password of **123456TESTplat&!** for the user.

```
[Switch-luser-manage-user1] password simple 123456TESTplat&!
```

Set the service type to **Telnet**.

```
[Switch-luser-manage-user1] service-type telnet
```

Assign **role1** to the user.

```
[Switch-luser-manage-user1] authorization-attribute user-role role1
```

Remove the default user role (**network-operator**) from the user. This operation ensures that the user has only the permissions of **role1**.

```
[Switch-luser-manage-user1] undo authorization-attribute user-role network-operator
[Switch-luser-manage-user1] quit
```

Verifying the configuration

Telnet to the switch, and enter the username and password to access the switch. (Details not shown.)

Verify that you can create VLANs 10 to 20. This example uses VLAN 10.

```
<Switch> system-view
[Switch] vlan 10
[Switch-vlan10] quit
```

Verify that you cannot create any VLAN other than VLANs 10 to 20. This example uses VLAN 30.

```
[Switch] vlan 30
Permission denied.
```

Verify that you can use all read commands of any feature. This example uses **`display clock`**.

```
[Switch] display clock
09:31:56.258 UTC Sun 01/01/2017
[Switch] quit
```

Verify that you cannot use the write or execute commands of any feature.

```
<Switch> debugging role all
Permission denied.
<Switch> ping 192.168.1.58
Permission denied.
```

Example: Configuring RBAC for RADIUS authentication users

Network configuration

As shown in Figure 2, the switch uses the FreeRADIUS server to provide AAA service for login users, including the Telnet user. The user account for the Telnet user is **hello@bbb**, which is assigned user role **role2**.

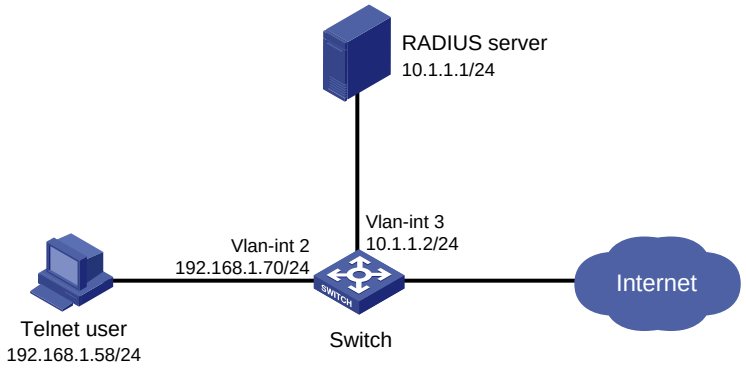
User role **role2** has the following permissions:

- Use all commands in ISP domain view.
- Use the read and write commands of the **arp** and **radius** features.
- Cannot access the read commands of the **acl** feature.

- Configure VLANs 1 to 20 and interfaces GigabitEthernet 1/0/1 to GigabitEthernet 1/0/4. Access to any other VLANs and interfaces is denied.

The switch and the FreeRADIUS server use a shared key of **expert** and authentication port **1812**. The switch delivers usernames with their domain names to the server.

Figure 2 Network diagram



Procedure

Make sure the settings on the switch and the RADIUS server match.

1. Configure the switch:

Assign VLAN-interface 2 an IP address from the same subnet as the Telnet user.

```
<Switch> system-view
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.1.70 255.255.255.0
[Switch-Vlan-interface2] quit
```

Assign VLAN-interface 3 an IP address from the same subnet as the RADIUS server.

```
[Switch] interface vlan-interface 3
[Switch-Vlan-interface3] ip address 10.1.1.2 255.255.255.0
[Switch-Vlan-interface3] quit
```

Enable the Telnet server.

```
[Switch] telnet server enable
```

Enable scheme authentication on the user lines for Telnet users.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Create RADIUS scheme **rad** and enter RADIUS scheme view.

```
[Switch] radius scheme rad
```

Specify the primary server address and the service port in the scheme.

```
[Switch-radius-rad] primary authentication 10.1.1.1 1812
```

Set the shared key to **expert** in the scheme for the switch to authenticate to the server.

```
[Switch-radius-rad] key authentication simple expert
[Switch-radius-rad] quit
```

Specify scheme **rad** as the authentication and authorization schemes for ISP domain **bbb**, and configure the ISP domain to not perform accounting for login users.

IMPORTANT:

Because RADIUS user authorization information is piggybacked in authentication responses, the authentication and authorization methods must use the same RADIUS scheme.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login radius-scheme rad
[Switch-isp-bbb] authorization login radius-scheme rad
[Switch-isp-bbb] accounting login none
[Switch-isp-bbb] quit
```

Create feature group **fgroup1**.

```
[Switch] role feature-group name fgroup1
```

Add the **arp** and **radius** features to the feature group.

```
[Switch-featuregrp-fgroup1] feature arp
[Switch-featuregrp-fgroup1] feature radius
[Switch-featuregrp-fgroup1] quit
```

Create user role **role2**.

```
[Switch] role name role2
```

Configure rule 1 to permit the user role to use all commands available in ISP domain view.

```
[Switch-role-role2] rule 1 permit command system-view ; domain *
```

Configure rule 2 to permit the user role to use the read and write commands of all features in **fgroup1**.

```
[Switch-role-role2] rule 2 permit read write feature-group fgroup1
```

Configure rule 3 to disable access to the read commands of the **acl** feature.

```
[Switch-role-role2] rule 3 deny read feature acl
```

Configure rule 4 to permit the user role to create VLANs and use all commands available in VLAN view.

```
[Switch-role-role2] rule 4 permit command system-view ; vlan *
```

Configure rule 5 to permit the user role to enter interface view and use all commands available in interface view.

```
[Switch-role-role2] rule 5 permit command system-view ; interface *
```

Configure the user role VLAN policy to disable configuration of any VLAN except VLANs 1 to 20.

```
[Switch-role-role2] vlan policy deny
[Switch-role-role2-vlanpolicy] permit vlan 1 to 20
[Switch-role-role2-vlanpolicy] quit
```

Configure the user role interface policy to disable configuration of any interface except GigabitEthernet 1/0/1 to GigabitEthernet 1/0/4.

```
[Switch-role-role2] interface policy deny
[Switch-role-role2-ifpolicy] permit interface gigabitethernet 1/0/1 to
gigabitethernet 1/0/4
[Switch-role-role2-ifpolicy] quit
[Switch-role-role2] quit
```

2. Configure the RADIUS server:

Add either of the user role attributes to the dictionary file of the FreeRADIUS server.

```
Cisco-AVPair = "shell:roles=\"role2\""
Cisco-AVPair = "shell:roles*\"role2\""
```

Configure the settings required for the FreeRADIUS server to communicate with the switch. (Details not shown.)

Verifying the configuration

Telnet to the switch, and enter the username and password to access the switch. (Details not shown.)

Verify that you can use all commands available in ISP domain view.

```
<Switch> system-view
[Switch] domain abc
[Switch-isp-abc] authentication login radius-scheme abc
[Switch-isp-abc] quit
```

Verify that you can use all read and write commands of the **radius** and **arp** features. This example uses **radius**.

```
[Switch] radius scheme rad
[Switch-radius-rad] primary authentication 2.2.2.2
[Switch-radius-rad] display radius scheme rad
...
```

Verify that you cannot configure any VLAN except VLANs 1 to 20. This example uses VLAN 10 and VLAN 30.

```
[Switch] vlan 10
[Switch-vlan10] quit
[Switch] vlan 30
Permission denied.
```

Verify that you cannot configure any interface except GigabitEthernet 1/0/1 to GigabitEthernet 1/0/4. This example uses GigabitEthernet 1/0/2 and GigabitEthernet 1/0/5.

```
[Switch] vlan 10
[Switch-vlan10] port gigabitethernet 1/0/2
[Switch-vlan10] port gigabitethernet 1/0/5
Permission denied.
```

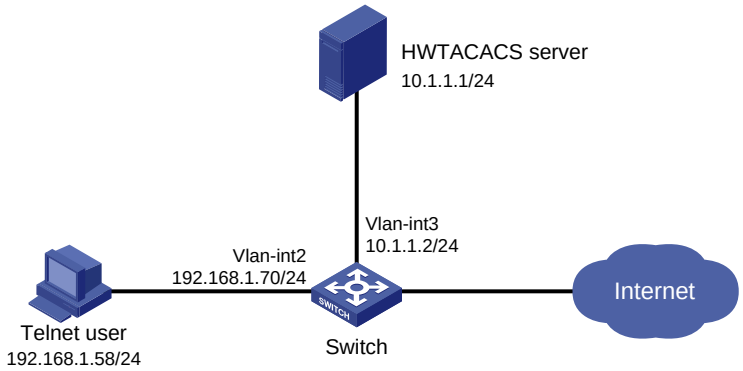
Example: Configuring RBAC temporary user role authorization (HWTACACS authentication)

Network configuration

As shown in Figure 3, the switch uses HWTACACS authentication for login users, including the Telnet user. The user account for the Telnet user is **test@bbb**, which is assigned user role **level-0**.

Configure the remote-then-local authentication mode for temporary user role authorization. The switch uses the HWTACACS server to provide authentication for changing the user role among **level-0** through **level-3** or changing the user role to **network-admin**. If the AAA configuration is invalid or the HWTACACS server does not respond, the switch performs local authentication.

Figure 3 Network diagram



Procedure

1. Configure the switch:

   # Assign an IP address to VLAN-interface 2 (the interface connected to the Telnet user).

   ```
   <Switch> system-view
   [Switch] interface vlan-interface 2
   [Switch-Vlan-interface2] ip address 192.168.1.70 255.255.255.0
   [Switch-Vlan-interface2] quit
   ```

   # Assign an IP address to VLAN-interface 3 (the interface connected to the HWTACACS server).

   ```
   [Switch] interface vlan-interface 3
   [Switch-Vlan-interface3] ip address 10.1.1.2 255.255.255.0
   [Switch-Vlan-interface3] quit
   ```

   # Enable the Telnet server.

   ```
   [Switch] telnet server enable
   ```

   # Enable scheme authentication on the user lines for Telnet users.

   ```
   [Switch] line vty 0 63
   [Switch-line-vty0-63] authentication-mode scheme
   [Switch-line-vty0-63] quit
   ```

   # Enable remote-then-local authentication for temporary user role authorization.

   ```
   [Switch] super authentication-mode scheme local
   ```

   # Create an HWTACACS scheme named **hwtac** and enter HWTACACS scheme view.

   ```
   [Switch] hwtacacs scheme hwtac
   ```

   # Specify the primary authentication server address and the service port in the scheme.

   ```
   [Switch-hwtacacs-hwtac] primary authentication 10.1.1.1 49
   ```

   # Specify the primary authorization server address and the service port in the scheme.

   ```
   [Switch-hwtacacs-hwtac] primary authorization 10.1.1.1 49
   ```

   # Set the shared key to **expert** in the scheme for the switch to authenticate to the authentication server.

   ```
   [Switch-hwtacacs-hwtac] key authentication simple expert
   ```

   # Set the shared key to **expert** in the scheme for the switch to authenticate to the authorization server.

   ```
   [Switch-hwtacacs-hwtac] key authorization simple expert
   ```

   # Exclude ISP domain names from the usernames sent to the HWTACACS server.

   ```
   [Switch-hwtacacs-hwtac] user-name-format without-domain
   [Switch-hwtacacs-hwtac] quit
   ```

   # Create ISP domain **bbb** and enter ISP domain view.

   ```
   [Switch] domain bbb
   ```

   # Configure ISP domain **bbb** to use HWTACACS scheme **hwtac** for login user authentication.

   ```
   [Switch-isp-bbb] authentication login hwtacacs-scheme hwtac
   ```

   # Configure ISP domain **bbb** to use HWTACACS scheme **hwtac** for login user authorization.

   ```
   [Switch-isp-bbb] authorization login hwtacacs-scheme hwtac
   ```

   # Configure ISP domain **bbb** to not perform accounting for login users.

   ```
   [Switch-isp-bbb] accounting login none
   ```

   # Apply HWTACACS scheme **hwtac** to the ISP domain for user role authentication.

   ```
   [Switch-isp-bbb] authentication super hwtacacs-scheme hwtac
   [Switch-isp-bbb] quit
   ```

   # Set the local authentication password to **654321TESTplat&!** for user role **level-3**.

```
[Switch] super password role level-3 simple 654321TESTplat&!
```

Set the local authentication password to **654321TESTplat&!** for user role **network-admin**.

```
[Switch] super password role network-admin simple 654321TESTplat&!
[Switch] quit
```

2. Configure the HWTACACS server:

   This example uses ACSv4.0. This example provides only the key configuration steps. For more information about configuring a HWTACACS server, see the server documentation.

   a. Access the **User Setup** page.

   b. Add a user account named **test** and set its password to **123456TESTplat&!**. (Details not shown.)

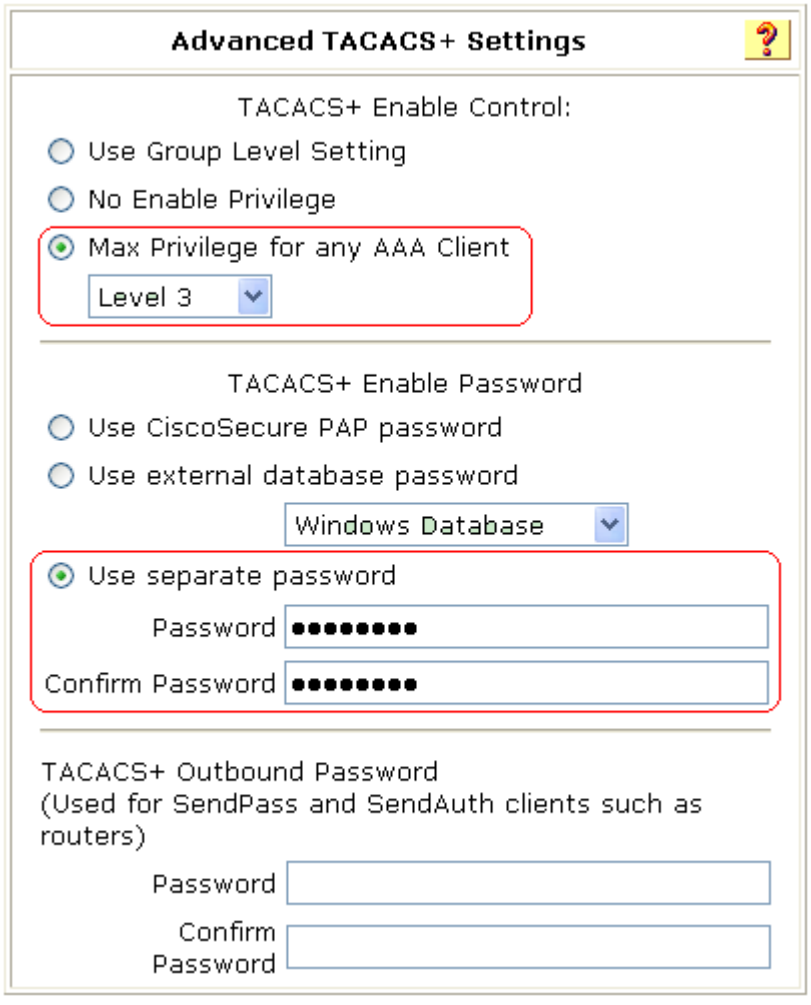
   c. In the **Advanced TACACS+ Settings** area, configure the following parameters:

      – Select **Level 3** for the **Max Privilege for any AAA Client** option.

        If the target user role is only **network-admin** for temporary user role authorization, you can select any level for the option.

      – Select the **Use separate password** option, and specify **enabpass** as the password.

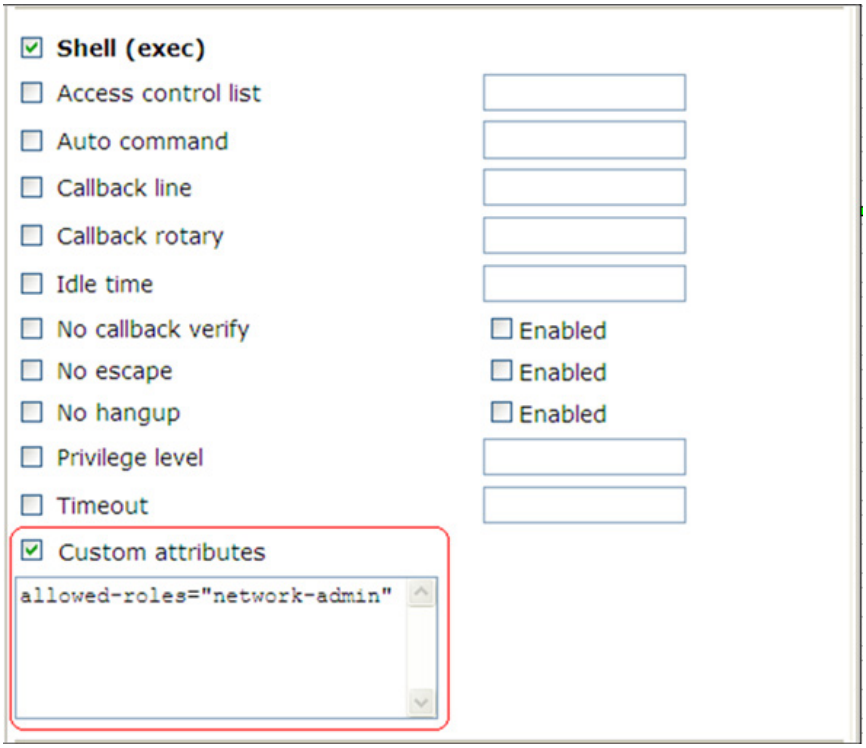
   Figure 4 Configuring advanced TACACS+ settings



   d. Select **Shell (exec)** and **Custom attributes**, and enter **allowed-roles="network-admin"** in the **Custom attributes** field.

      Use a blank space to separate the allowed roles.

Figure 5 Configuring custom attributes for the Telnet user



Verifying the configuration

1. Telnet to the switch, and enter username **test@bbb** and password **123456TESTplat&!** to access the switch. Verify that you have access to diagnostic commands.

```
<Switch> telnet 192.168.1.70
Trying 192.168.1.70 ...
Press CTRL+K to abort
Connected to 192.168.1.70 ...
login: test@bbb
Password:
******************************************************************************
* Copyright (c) 2004-2017 New H3C Technologies Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                                 *
* no decompiling or reverse-engineering shall be allowed.                    *
******************************************************************************

<Switch>?
User view commands:
  erase        Alias for 'delete'
  exit         Alias for 'quit'
  mtrace       Configure the multicast traceroute
  no           Alias for 'undo'
  ping         Ping function
  quit         Exit from current command view
  show         Alias for 'display'
  ssh2         Establish an Stelnet connection to an Stelnet server
  super        Switch to a user role
```

```
  system-view  Enter the System View
  telnet       Establish a telnet connection
  tracert      Tracert function
  write        Alias for 'save'
<Switch>
```

2. Verify that you can obtain the level-3 user role:

   # Use the super password to obtain the level-3 user role. When the system prompts for a username and password, enter username **test@bbb** and password **enabpass**.

   ```
   <Switch> super level-3
   Username: test@bbb
   Password:
   ```

   The following output shows that you have obtained the level-3 user role.

   ```
   User privilege role is level-3, and only those commands that authorized to the role
   can be used.
   ```

   # If the ACS server does not respond, enter local authentication password **654321TESTplat&!** at the prompt.

   ```
   Invalid configuration or no response from the authentication server.
   Change authentication mode to local.
   Password:
   User privilege role is level-3, and only those commands that authorized to the role
   can be used.
   ```

   The output shows that you have obtained the level-3 user role.

3. Use the method in step 2 to verify that you can obtain the level-0, level-1, level-2, and network-admin user roles. (Details not shown.)

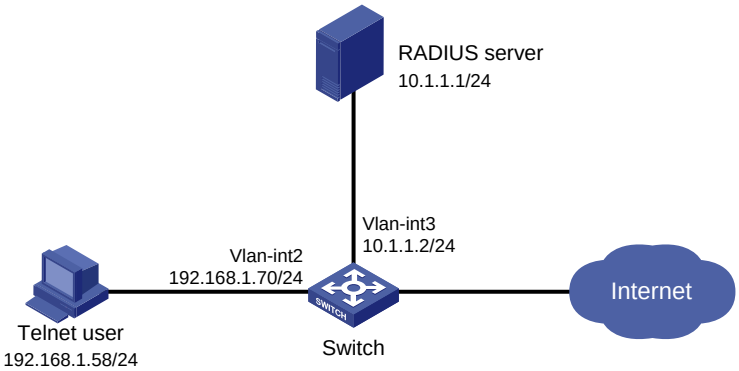
Example: Configuring RBAC temporary user role authorization (RADIUS authentication)

Network configuration

As shown in Figure 6, the switch uses RADIUS authentication for login users, including the Telnet user. The user account for the Telnet user is **test@bbb**, which is assigned user role **level-0**.

Configure the remote-then-local authentication mode for temporary user role authorization. The switch uses the RADIUS server to provide authentication for the **network-admin** user role. If the AAA configuration is invalid or the RADIUS server does not respond, the switch performs local authentication.

Figure 6 Network diagram



Procedure

1. Configure the switch:

   # Assign an IP address to VLAN-interface 2 (the interface connected to the Telnet user).

   ```
   <Switch> system-view
   [Switch] interface vlan-interface 2
   [Switch-Vlan-interface2] ip address 192.168.1.70 255.255.255.0
   [Switch-Vlan-interface2] quit
   ```

   # Assign an IP address to VLAN-interface 3 (the interface connected to the RADIUS server).

   ```
   [Switch] interface vlan-interface 3
   [Switch-Vlan-interface3] ip address 10.1.1.2 255.255.255.0
   [Switch-Vlan-interface3] quit
   ```

   # Enable the Telnet server.

   ```
   [Switch] telnet server enable
   ```

   # Enable scheme authentication on the user lines for Telnet users.

   ```
   [Switch] line vty 0 15
   [Switch-line-vty0-15] authentication-mode scheme
   [Switch-line-vty0-15] quit
   ```

   # Enable remote-then-local authentication for temporary user role authorization.

   ```
   [Switch] super authentication-mode scheme local
   ```

   # Create RADIUS scheme **radius** and enter RADIUS scheme view.

   ```
   [Switch] radius scheme radius
   ```

   # Specify the primary authentication server address and the shared key in the scheme for secure communication between the switch and the server.

   ```
   [Switch-radius-radius] primary authentication 10.1.1.1 key simple expert
   ```

   # Exclude ISP domain names from the usernames sent to the RADIUS server.

   ```
   [Switch-radius-radius] user-name-format without-domain
   [Switch-radius-radius] quit
   ```

   # Create ISP domain **bbb** and enter ISP domain view.

   ```
   [Switch] domain bbb
   ```

   # Configure ISP domain **bbb** to use RADIUS scheme **radius** for login user authentication.

   ```
   [Switch-isp-bbb] authentication login radius-scheme radius
   ```

   # Configure ISP domain **bbb** to use RADIUS scheme **radius** for login user authorization.

   ```
   [Switch-isp-bbb] authorization login radius-scheme radius
   ```

   # Configure ISP domain **bbb** to not perform accounting for login users.

   ```
   [Switch-isp-bbb] accounting login none
   ```

   # Apply RADIUS scheme **radius** to the ISP domain for user role authentication.

   ```
   [Switch-isp-bbb] authentication super radius-scheme radius
   [Switch-isp-bbb] quit
   ```

   # Set the local authentication password to **abcdef654321TESTplat&!** for user role **network-admin**.

   ```
   [Switch] super password role network-admin simple abcdef654321TESTplat&!
   [Switch] quit
   ```

2. Configure the RADIUS server:

   This example uses ACSv4.2. This example provides only the key configuration steps. For more information about configuring a RADIUS server, see the server documentation.

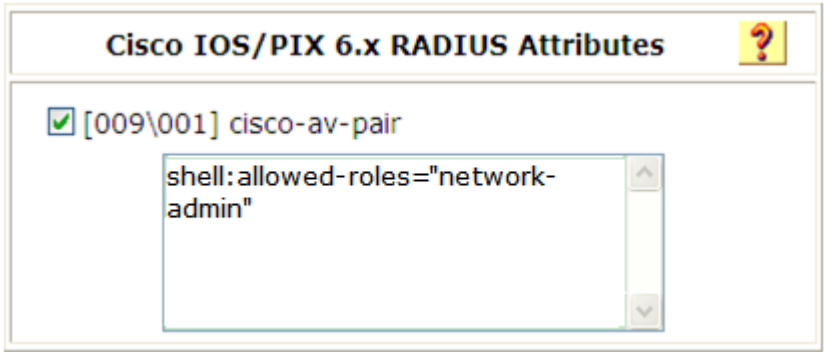
   a. Add a Telnet user account for login authentication. Set the username to **test@bbb** and the password to **123456TESTplat&!**. (Details not shown.)

b. Add a user account for temporary user role authorization. Set the username to **$enab0$** and the password to **123456**. (Details not shown.)

c. Access the **Cisco IOS/PIX 6.x RADIUS Attributes** page.

d. Configure the **cisco-av-pair** attribute, as shown in Figure 7.

Figure 7 Configuring the cisco-av-pair attribute



Verifying the configuration

1. Telnet to the switch, and enter username **test@bbb** and password **123456TESTplat&!** to access the switch. Verify that you have access to diagnostic commands.

```
<Switch> telnet 192.168.1.70
Trying 192.168.1.70 ...
Press CTRL+K to abort
Connected to 192.168.1.70 ...
login: test@bbb
Password:
******************************************************************************
* Copyright (c) 2004-2017 New H3C Technologies Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                                 *
* no decompiling or reverse-engineering shall be allowed.                    *
******************************************************************************

<Switch>?
User view commands:
  erase        Alias for 'delete'
  exit         Alias for 'quit'
  mtrace       Configure the multicast traceroute
  no           Alias for 'undo'
  ping         Ping function
  quit         Exit from current command view
  show         Alias for 'display'
  ssh2         Establish an Stelnet connection to an Stelnet server
  super        Switch to a user role
  system-view  Enter the System View
  telnet       Establish a telnet connection
  tracert      Tracert function
  write        Alias for 'save'
<switch>
```

2. Verify that you can obtain the network-admin user role:

Use the super password to obtain the network-admin user role. When the system prompts for a username and password, enter username **test@bbb** and password **123456**.

```
<Switch> super network-admin
Username: test@bbb
Password:
```

The following output shows that you have obtained the network-admin user role.

```
User privilege role is network-admin, and only those commands that authorized to the role can be used.
```

If the ACS server does not respond, enter local authentication password **abcdef654321TESTplat&!** at the prompt.

```
Invalid configuration or no response from the authentication server.
Change authentication mode to local.
Password:
User privilege role is network-admin, and only those commands that authorized to the role can be used.
```

The output shows that you have obtained the network-admin user role.

Troubleshooting RBAC

This section describes several typical RBAC issues and their solutions.

Local users have more access permissions than intended

Symptom

A local user can use more commands than should be permitted by the assigned user roles.

Analysis

The local user might have been assigned to user roles without your knowledge. For example, the local user is automatically assigned the default user role when you create the user.

Solution

To resolve the issue:

1. Use the **`display local-user`** command to examine the local user accounts for undesirable user roles, and remove them.
2. If the issue persists, contact H3C Support.

Login attempts by RADIUS users always fail

Symptom

Attempts by a RADIUS user to log in to the network access device always fail, even though the following conditions exist:

- The network access device and the RADIUS server can communicate with one another.
- All AAA settings are correct.

Analysis

RBAC requires that a login user have a minimum of one user role. If the RADIUS server does not authorize the login user to use any user role, the user cannot log in to the device.

Solution

To resolve the issue:

1. Use one of the following methods:
   - Configure the **`role default-role enable`** command. A RADIUS user can log in with the default user role when no user role is assigned by the RADIUS server.
   - Add the user role authorization attributes on the RADIUS server.
2. If the issue persists, contact H3C Support.

Contents

Login overview 1

Using the console port for the first device access 2

Configuring CLI login 3

About CLI login 3
User lines 3
Login authentication modes 3
User roles 4
FIPS compliance 4
Restrictions and guidelines: CLI login configuration 4
Configuring console login 5
About console, AUX, and USB login 5
Restrictions and guidelines 5
Console login configuration tasks at a glance 5
Configuring console login authentication 5
Configuring common console login settings 7
Configuring Telnet login 9
About Telnet login 9
Restrictions and guidelines 9
Configuring the device as a Telnet server 9
Using the device to log in to a Telnet server 13
Configuring SSH login 14
About SSH login 14
Configuring the device as an SSH server 14
Using the device to log in to an SSH server 16
Display and maintenance commands for CLI login 16

Accessing the device through SNMP 18

Configuring RESTful access 19

About RESTful access 19
FIPS compliance 19
Configuring RESTful access over HTTP 19
Configuring RESTful access over HTTPS 20

Controlling user access to the device 21

About login user access control 21
FIPS compliance 21
Controlling Telnet and SSH logins 21
Controlling Telnet logins 21
Controlling SSH logins 22
Example: Controlling Telnet login 22
Controlling SNMP access 23
About SNMP access control 23
Example: Controlling SNMP access 23
Configuring command authorization 23
About command authorization 23
Restrictions and guidelines 24
Procedure 24
Example: Configuring command authorization 25
Configuring command accounting 26
About command accounting 26
Restrictions and guidelines 27
Procedure 27
Example: Configuring command accounting 28

Login overview

The device supports the following types of login methods:

- **CLI login**—At the CLI, you can enter text commands to configure and manage the device.

  To log in to the CLI, you can use one of the following methods:

  - Connect to the console port.
  - Use Telnet.
  - Use SSH.
- **SNMP access**—You can run SNMP on an NMS to access the device MIB, and perform Get and Set operations to configure and manage the device.
- **RESTful access**—You can use RESTful API operations to configure and manage the device.

The first time you access the device, you can only log in to the CLI through the console port unless the device is automatically configured at startup. After login, you can change console login parameters or configure other access methods.

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Telnet and HTTP-based RESTful access are not supported in FIPS mode.

In login management related descriptions, it is assumed that the device does not enter the automatic configuration process at startup.

Using the console port for the first device access

About this task

Console login is the fundamental login method.

Prerequisites

To log in through the console port, prepare a terminal, for example, a PC. Make sure the terminal has a terminal emulation program, such as HyperTerminal or PuTTY. For information about how to use terminal emulation programs, see the programs' user guides.

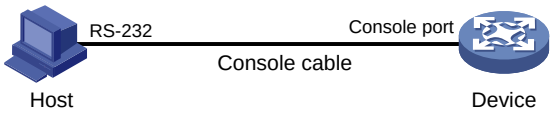
Procedure

1. Connect the DB-9 female connector of the console cable shipped with the device to the serial port of the PC.
2. Identify the console port of the device carefully and connect the RJ-45 connector of the console cable to the console port.

IMPORTANT:

- If the serial port on the PC do not support hot swapping, power off the PC before you connect the console cable to or disconnect the cable from the serial port.
- When you connect a PC to an operating device, first connect the PC end. To disconnect a PC from an operating device, first disconnect the device end.

Figure 1 Connecting a terminal to the console port



3. On the PC, launch the terminal emulation program, and create a connection that uses the serial port connected to the device. Set the port properties so the port properties match the following console port default settings:
   - **Bits per second**—9600 bps.
   - **Flow control**—None.
   - **Parity**—None.
   - **Stop bits**—1.
   - **Data bits**—8.
4. Power on the device and press **Enter** as prompted.

   The user view prompt appears. You can enter commands to configure or manage the device. To get help, enter a question mark (?).

Configuring CLI login

About CLI login

The device uses user lines (also called user interfaces) to manage CLI sessions and monitor user behavior. For a user line, you can configure access control settings, including the login authentication method and user roles.

User lines

User line types

The device supports the following types of user lines, and different user lines require different login methods:

- **Console line**—Login through the console port.
- **Virtual type terminal (VTY) line**—Telnet or SSH login.

User line numbering

A user line has an absolute number and a relative number.

An absolute number uniquely identifies a user line among all user lines. The user lines are numbered starting from 0 and incrementing by 1, in the sequence of console and VTY lines. You can use the **`display line`** command without any parameters to view supported user lines and their absolute numbers.

A relative number uniquely identifies a user line among all user lines of the same type. The number format is *user line type* + *number*. All types of user lines are numbered starting from 0 and incrementing by 1. For example, the first VTY line is VTY 0.

User line assignment

The device assigns user lines to CLI login users depending on their login methods, as described in "User lines." When a user logs in, the device checks the idle user lines for the login method, and assigns the lowest numbered user line to the user. For example, if VTY 0 and VTY 3 are idle when a user Telnets to the device, the device assigns VTY 0 to the user.

Each user line can be assigned only to one user at a time. If no user line is available, a CLI login attempt will be rejected.

Login authentication modes

You can configure login authentication to prevent illegal access to the device CLI.

In non-FIPS mode, the device supports the following login authentication modes:

- **None**—Disables authentication. This mode allows access without authentication and is insecure.
- **Password**—Requires password authentication. A user must provide the correct password at login.
- **Scheme**—Uses the AAA module to provide local or remote login authentication. A user must provide the correct username and password at login.

In FIPS mode, the device supports only the scheme authentication mode.

Different login authentication modes require different user line configurations, as shown in Table 1.

Table 1 Configuration required for different login authentication modes

Authentication mode	Configuration tasks
None	Set the authentication mode to none.
Password	1. Set the authentication mode to password. 2. Set a password.
Scheme	3. Set the authentication mode to scheme. 4. Configure login authentication methods in ISP domain view. For more information, see *Security Configuration Guide*.

User roles

A user is assigned user roles at login. The user roles control the commands available for the user. For more information about user roles, see "Configuring RBAC."

The device assigns user roles based on the login authentication mode and user type.

- In none or password authentication mode, the device assigns the user roles specified for the user line.
- In scheme authentication mode, the device uses the following rules to assign user roles:
  - For an SSH login user who uses publickey or password-publickey authentication, the device assigns the user roles specified for the local device management user with the same name.
  - For other users, the device assigns user roles according to the user role configuration of the AAA module. If the AAA server does not assign any user roles and the default user role feature is disabled, a remote AAA authentication user cannot log in.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Telnet login is not supported in FIPS mode.

Restrictions and guidelines: CLI login configuration

For commands that are available in both user line view and user line class view, the following rules apply:

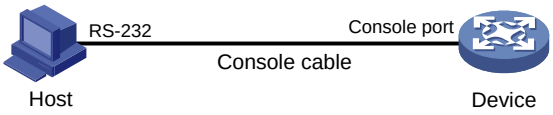
- A setting in user line view applies only to the user line. A setting in user line class view applies to all user lines of the class.
- A non-default setting in either view takes precedence over the default setting in the other view. A non-default setting in user line view takes precedence over the non-default setting in user line class view.
- A setting in user line class view takes effect only on users who log in after the setting is made. It does not affect users who are already online when the setting is made.

Configuring console login

About console, AUX, and USB login

You can connect a terminal to the console port of the device to log in and manage the device, as shown in Figure 2. For information about the login procedure, see "Using the console port for the first device access."

Figure 2 Logging in through the console port



By default, console login is enabled and does not require authentication. The default user role is network-admin for a console user. To improve device security, configure password or scheme authentication for console login immediately after you log in to the device for the first time.

Restrictions and guidelines

A console login configuration change takes effect only on users who log in after the change is made. It does not affect users who are already online when the change is made.

In FIPS mode, the device supports only scheme authentication. You cannot disable authentication or configure password authentication.

The device displays the current memory usage when you log in to the device through the console port if the following conditions exist:

- The **memory-threshold** command has been used to set free-memory thresholds.
- The amount of free memory space has decreased to or below the minor, severe, or critical alarm threshold or the early-warning threshold.

Console login configuration tasks at a glance

To configure console login, perform the following tasks:

1. Configuring console login authentication
   - Disabling authentication for console login
   - Configuring password authentication for console login
   - Configuring scheme authentication for console login
2. (Optional.) Configuring common console login settings

Configuring console login authentication

Disabling authentication for console login

1. Enter system view.
   system-view
2. Enter console line view or class view.
   - Enter console line view.
     line console *first-number* [*last-number*]
   - Enter console line class view.

line class console

3. Disable authentication.

   authentication-mode none

   By default, authentication is disabled for console login.

> **CAUTION:**
> When authentication is disabled, users can log in to the device through the line or line class without authentication. For security purpose, disable authentication with caution.

4. Assign a user role.

   user-role *role-name*

   By default, a console user is assigned the **network-admin** user role.

Configuring password authentication for console login

1. Enter system view.

   system-view

2. Enter console line view or class view.
   - Enter console line view.

     line console *first-number* [*last-number*]
   - Enter console class view.

     line class console

3. Enable password authentication.

   authentication-mode password

   By default, authentication is disabled for console login.

4. Set a password.

   set authentication password { **hash** | **simple** } *password*

   By default, no password is set.

5. Assign a user role.

   user-role *role-name*

   By default, a console user is assigned the **network-admin** user role.

Configuring scheme authentication for console login

1. Enter system view.

   system-view

2. Enter console line view or class view.
   - Enter console line view.

     line console *first-number* [*last-number*]
   - Enter console line class view.

     line class console

3. Enable scheme authentication.

   In non-FIPS mode:

   authentication-mode scheme

   By default, authentication is disabled for console login.

   In FIPS mode:

   authentication-mode scheme

By default, scheme authentication is enabled.

CAUTION:

When you enable scheme authentication, make sure an authentication user account is available. If no authentication user account is available, you cannot log in to the device through the line or line class at the next time.

4. Configure user authentication parameters in ISP domain view.

   To use local authentication, configure a local user and set the relevant attributes. To use remote authentication, configure a RADIUS, LDAP, or HWTACACS scheme. For more information, see AAA in *Security Configuration Guide*.

Configuring common console login settings

Restrictions and guidelines

Some common console login settings take effect immediately and can interrupt the current session. Use a login method different from console login to log in to the device before you change console login settings.

After you change console login settings, adjust the settings on the configuration terminal accordingly for a successful login.

Procedure

1. Enter system view.

   system-view

2. Enter console line view or class view.
   - Enter console line view.

     line console *first-number* [*last-number*]
   - Enter console line class view.

     line class console

3. Configure transmission parameters.
   - Set the transmission rate.

     speed *speed-value*

     By default, the transmission rate is 9600 bps.

     This command is not available in user line class view.
   - Specify the parity mode.

     parity { **even** | **mark** | **none** | **odd** | **space** }

     By default, a user line does not use parity.

     This command is not available in user line class view.
   - Configure flow control.

     flow-control { **hardware** | **none** | **software** }

     By default, the device does not perform flow control.

     This command is not available in user line class view.
   - Specify the number of data bits for a character.

     databits { **5** | **6** | **7** | **8** }

     The default is 8.

     This command is not available in user line class view.

Parameter	Description
7	Uses standard ASCII characters.
8	Uses extended ASCII characters.

- Specify the number of stop bits for a character.

  stopbits { **1** | **1.5** | **2** }

  The default is 1.

  Stop bits indicate the end of a character. The more the stop bits, the slower the transmission.

  This command is not available in user line class view.

4. Configure terminal attributes.
   - Enable the terminal service.

     shell

     Be default, the terminal service is enabled on all user lines.

     The **undo shell** command is not available in console line view.
   - Specify the terminal display type.

     terminal type { **ansi** | **vt100** }

     By default, the terminal display type is ANSI.

     The device supports ANSI and VT100 terminal display types. As a best practice, specify VT100 type on both the device and the configuration terminal. You can also specify the ANSI type for both sides, but a display problem might occur if a command line has more than 80 characters.
   - Set the maximum number of lines of command output to send to the terminal at a time.

     screen-length *screen-length*

     By default, the device sends a maximum of 24 lines to the terminal at a time.

     To disable pausing between screens of output, set the value to 0.
   - Set the size for the command history buffer.

     history-command max-size *value*

     By default, the buffer size is 10. The buffer for a user line can save a maximum of 10 history commands.
   - Set the CLI connection idle-timeout timer.

     idle-timeout *minutes* [*seconds*]

     By default, the CLI connection idle-timeout timer is 10 minutes.

     If no interaction occurs between the device and the user within the idle-timeout interval, the system automatically terminates the user connection on the user line.

     If you set the timeout timer to 0, the connection will not be aged out.
5. Specify the command to be automatically executed for login users on the lines.

   auto-execute command *command*

   By default, no command is specified for auto execution.

CAUTION:

Use this command with caution. If this command is used on a user line, users that log in to the device through this user line might fail to configure the system.

The device will automatically execute the specified command when a user logs in through the user line, and close the user connection after the command is executed.

This command is not available in console line view or console line class view.

6. Configure shortcut keys.
   - Specify the terminal session activation key.
     activation-key *character*
     By default, pressing **Enter** starts the terminal session.
   - Specify the escape key.
     escape-key { *character* | **default** }
     By default, pressing **Ctrl+C** terminates a command.
   - Set the user line locking key.
     lock-key *key-string*
     By default, no user line locking key is set.

Configuring Telnet login

About Telnet login

The device can act as a Telnet server to allow Telnet login, or as a Telnet client to Telnet to other devices.

Restrictions and guidelines

Telnet login is not supported in FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

A Telnet login configuration change takes effect only on users who log in after the change is made. It does not affect users who are already online when the change is made.

Telnet might be insecure because it transmits all data in plaintext form.

The device displays the current memory usage when you Telnet to the device if the following conditions exist:

- The **memory-threshold** command has been used to set free-memory thresholds.
- The amount of free memory space has decreased to or below the minor, severe, or critical alarm threshold or the early-warning threshold.

Configuring the device as a Telnet server

Telnet server configuration tasks at a glance

To configure the device as a Telnet server, perform the following tasks:

1. Enabling the Telnet server
2. Configuring Telnet login authentication
   - Disabling authentication for Telnet login
   - Configuring password authentication for Telnet login
   - Configuring scheme authentication for Telnet login
3. (Optional.) Configuring common Telnet server settings
4. (Optional.) Configuring the alarm threshold and alarm clearance threshold for Telnet login failures
5. (Optional.) Configuring common VTY line settings

Enabling the Telnet server

1. Enter system view.
   system-view
2. Enable the Telnet server.
   telnet server enable
   By default, the Telnet server is disabled.

Disabling authentication for Telnet login

1. Enter system view.
   system-view
2. Enter VTY line view or class view.
   - Enter VTY line view.
     line vty *first-number* [*last-number*]
   - Enter VTY line class view.
     line class vty
3. Disable authentication.
   authentication-mode none
   By default, password authentication is enabled for Telnet login.

CAUTION:
When authentication is disabled, users can log in to the device through the line or line class without authentication. For security purpose, disable authentication with caution.

   In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.
4. (Optional.) Assign a user role.
   user-role *role-name*
   By default, a VTY line user is assigned the **network-operator** user role.

Configuring password authentication for Telnet login

1. Enter system view.
   system-view
2. Enter VTY line view or class view.
   - Enter VTY line view.
     line vty *first-number* [*last-number*]
   - Enter VTY line class view.
     line class vty
3. Enable password authentication.
   authentication-mode password
   By default, password authentication is enabled for Telnet login.

CAUTION:
When you enable password authentication, you must also configure an authentication password for the line or line class. If no authentication password is configured, you cannot log in to the device through the line or line class at the next time.

In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

4. Set a password.

   set authentication password { **hash** | **simple** } *password*

   By default, no password is set.

5. (Optional.) Assign a user role.

   user-role *role-name*

   By default, a VTY line user is assigned the **network-operator** user role.

Configuring scheme authentication for Telnet login

1. Enter system view.

   system-view

2. Enter VTY line view or class view.

   - Enter VTY line view.

     line vty *first-number* [*last-number*]

   - Enter VTY line class view.

     line class vty

3. Enable scheme authentication.

   authentication-mode scheme

   By default, password authentication is enabled for Telnet login.

CAUTION:

When you enable scheme authentication, make sure an authentication user account is available. If no authentication user account is available, you cannot log in to the device through the line or line class at the next time.

In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

4. Configure user authentication parameters in ISP domain view.

   To use local authentication, configure a local user and set the relevant attributes.

   To use remote authentication, configure a RADIUS, LDAP, or HWTACACS scheme. For more information, see AAA in *Security Configuration Guide*.

Configuring common Telnet server settings

1. Enter system view.

   system-view

2. Set the DSCP value for outgoing Telnet packets.

   IPv4:

   telnet server dscp *dscp-value*

   IPv6:

   telnet server ipv6 dscp *dscp-value*

   By default, the DSCP value is 48.

   The DSCP value is carried in the ToS field of an IPv4 packet and in the Traffic class field of an IPv6 packet to indicate the packet transmission priority.

3. Specify the Telnet service port number.

   IPv4:

telnet server port *port-number*

IPv6:

telnet server ipv6 port *port-number*

By default, the Telnet service port number is 23.

4. Set the maximum number of concurrent Telnet users.

aaa session-limit telnet *max-sessions*

By default, the maximum number of concurrent Telnet users is 32.

Changing this setting does not affect users who are currently online. If the new limit is less than the number of online Telnet users, no additional users can Telnet in until the number drops below the new limit.

For more information about this command, see *Security Command Reference*.

Configuring the alarm threshold and alarm clearance threshold for Telnet login failures

1. Enter system view.

system-view

2. Set the alarm threshold and alarm clearance threshold for Telnet login failures in a statistics period.

telnet server login-failed threshold-alarm upper-limit *report-times* **lower-limit** *resume-times* **period** *period-time*

By default, the statistics period is five minutes and the alarm threshold and alarm clearance threshold for Telnet login failures in the statistics period is 30 and 20, respectively.

After you execute this command, if the number of Telnet login failures reaches the alarm threshold in a statistics period, an alarm message is generated. If the number of Telnet login failures drops below the alarm clearance threshold in a statistics period, an alarm clearance message is generated. This helps you obtain real-time information about Telnet logins.

Configuring common VTY line settings

1. Enter system view.

system-view

2. Enter VTY line view or class view.
   - Enter VTY line view.

     line vty *first-number* [*last-number*]
   - Enter VTY line class view.

     line class vty
3. Configure VTY terminal attributes.
   - Enable the terminal service.

     shell

     By default, the terminal service is enabled on all user lines.
   - Specify the terminal display type.

     terminal type { **ansi** | **vt100** }

     By default, the terminal display type is ANSI.
   - Set the maximum number of lines of command output to send to the terminal at a time.

     screen-length *screen-length*

     By default, the device sends a maximum of 24 lines to the terminal at a time.

     To disable pausing between screens of output, set the value to 0.
   - Set the size for the command history buffer.

     history-command max-size *value*

By default, the buffer size is 10. The buffer for a user line can save a maximum of 10 history commands.

- Set the CLI connection idle-timeout timer.

  idle-timeout *minutes* [*seconds*]

  By default, the CLI connection idle-timeout timer is 10 minutes.

  If no interaction occurs between the device and the user within the idle-timeout interval, the system automatically terminates the user connection on the user line.

  If you set the timeout timer to 0, the connection will not be aged out.

4. Specify the supported protocols.

   protocol inbound { **all** | **ssh** | **telnet** }

   By default, Telnet and SSH are supported.

   A protocol change takes effect only on users who log in after the setting is made. It does not affect users who are already online when the setting is made.

   In VTY line view, this command is associated with the **authentication-mode** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

5. Specify the command to be automatically executed for login users on the user lines.

   auto-execute command *command*

   By default, no command is specified for auto execution.

IMPORTANT:

Before you execute this command and save the configuration, make sure you can access the CLI to modify the configuration through other VTY lines or console lines.

For a VTY line, you can specify a command that is to be automatically executed when a user logs in. After executing the specified command, the system automatically disconnects the Telnet session.

6. Configure shortcut keys.

   - Specify the shortcut key for terminating a task.

     escape-key { *character* | **default** }

     The default setting is **Ctrl+C**.

   - Set the user line locking key.

     lock-key *key-string*

     By default, no user line locking key is set.

Using the device to log in to a Telnet server

About this task

You can use the device as a Telnet client to log in to a Telnet server.

Figure 3 Telnetting from the device to a Telnet server



Prerequisites

Assign an IP address to the device and obtain the IP address of the Telnet server. If the device resides on a different subnet than the Telnet server, make sure the device and the Telnet server can reach each other.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Specify the source IPv4 address or source interface for outgoing Telnet packets.

   telnet client source { **interface** *interface-type interface-number* | **ip** *ip-address* }

   By default, no source IPv4 address or source interface is specified. The device uses the primary IPv4 address of the output interface as the source address for outgoing Telnet packets.

3. Return to user view.

   quit

4. Use the device to log in to a Telnet server.

   IPv4:

   telnet *remote-host* [*service-port*] [**vpn-instance** *vpn-instance-name*] [**source** { **interface** *interface-type interface-number* | **ip** *ip-address* } | **dscp** *dscp-value*] *

   IPv6:

   telnet ipv6 *remote-host* [**-i** *interface-type interface-number*] [*port-number*] [**vpn-instance** *vpn-instance-name*] [**source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* } | **dscp** *dscp-value*] *

Configuring SSH login

About SSH login

SSH offers a secure remote login method. By providing encryption and strong authentication, it protects devices against attacks such as IP spoofing and plaintext password interception. For more information, see *Security Configuration Guide*.

The device can act as an SSH server to allow Telnet login, or as an SSH client to log in to an SSH server.

Configuring the device as an SSH server

About this task

This section provides the SSH server configuration procedure used when the SSH client authentication method is password. For more information about SSH and publickey authentication configuration, see *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Create local key pairs.

   In non-FIPS mode:

public-key local create { **dsa** | **ecdsa** [**secp192r1** | **secp256r1** | **secp384r1** | **secp521r1**] | **rsa** } [**name** *key-name*]

In FIPS mode:

public-key local create { **dsa** | **ecdsa** [**secp256r1** | **secp384r1** | **secp521r1**] | **rsa** } [**name** *key-name*]

3. Enable the SSH server.

   ssh server enable

   By default, the SSH server is disabled.

4. (Optional.) Create an SSH user and specify the authentication mode.

   ssh user *username* **service-type stelnet authentication-type password**

5. Enter VTY line view or class view.
   - Enter VTY line view.

     line vty *first-number* [*last-number*]
   - Enter VTY line class view.

     line class vty

6. Enable scheme authentication.

   In non-FIPS mode:

   authentication-mode scheme

   By default, password authentication is enabled for VTY lines.

   In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

   In FIPS mode:

   authentication-mode scheme

   By default, scheme authentication is enabled for VTY lines.

   In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

CAUTION:

When you enable scheme authentication, make sure an authentication user account is available. If no authentication user account is available, you cannot log in to the device through the line or line class at the next time.

7. (Optional.) Specify the protocols for the user lines to support.

   In non-FIPS mode:

   protocol inbound { **all** | **ssh** | **telnet** }

   By default, Telnet and SSH are supported.

   A protocol change takes effect only on users who log in after the setting is made. It does not affect users who are already online when the setting is made.

   In VTY line view, this command is associated with the **authentication-mode** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

   In FIPS mode:

   protocol inbound ssh

   By default, SSH is supported.

A protocol change takes effect only on users who log in after the setting is made. It does not affect users who are already online when the setting is made.

In VTY line view, this command is associated with the **authentication-mode** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

8. (Optional.) Set the maximum number of concurrent SSH users.

   aaa session-limit ssh *max-sessions*

   By default, the maximum number of concurrent SSH users is 32.

   Changing this setting does not affect users who are currently online. If the new limit is less than the number of online SSH users, no additional SSH users can log in until the number drops below the new limit.

   For more information about this command, see *Security Command Reference*.

9. (Optional.) Configure common settings for VTY lines:

   a. Return to system view.

      quit

   b. Configure common settings for VTY lines.

      See "Configuring common VTY line settings."

Using the device to log in to an SSH server

About this task

You can use the device as an SSH client to log in to an SSH server.

Figure 4 Logging in to an SSH server from the device



Prerequisites

Assign an IP address to the device and obtain the IP address of the SSH server. If the device resides on a different subnet than the SSH server, make sure the device and the SSH server can reach each other.

Procedure

To use the device to log in to an SSH server, execute one of the following commands in user view:

IPv4:

ssh2 *server*

IPv6:

ssh2 ipv6 *server*

To work with the SSH server, you might need to specify a set of parameters. For more information, see *Security Configuration Guide*.

Display and maintenance commands for CLI login

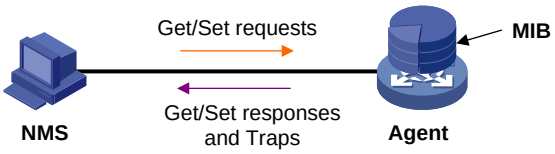
Execute **display** commands in any view.

Task	Command	Remarks
Display user line information.	**display line** [*num1* | { **console** | **vty** } *num2*] [**summary**]	N/A
Display the packet source setting for the Telnet client.	**display telnet client**	N/A
Display online CLI users.	**display users** [**all**]	N/A
Release a user line.	**free line** { *num1* | { **console** | **vty** } *num2* }	Multiple users can log in to the device to simultaneously configure the device. When necessary, you can execute this command to release some connections. You cannot use this command to release the connection you are using. This command is available in user view.
Lock the current user line and set the password for unlocking the line.	**lock**	By default, the system does not lock any user lines. This command is not supported in FIPS mode. This command is available in user view.
Lock the current user line and enable unlocking authentication.	**lock reauthentication**	By default, the system does not lock any user lines or initiate reauthentication. To unlock the locked user line, you must press **Enter** and provide the login password to pass reauthentication. This command is available in any view.
Send messages to user lines.	**send** { **all** | *num1* | { **console** | **vty** } *num2* }	This command is available in user view.

Accessing the device through SNMP

You can run SNMP on an NMS to access the device MIB and perform Get and Set operations to configure and manage the device.

Figure 5 SNMP access diagram



For more information about SNMP, see *Network Management and Monitoring Configuration Guide*.

Configuring RESTful access

About RESTful access

The device provides the Representational State Transfer application programming interface (RESTful API). Based on this API, you can use programming languages such as Python, Ruby, or Java to write programs to perform the following tasks:

- Send RESTful requests to the device to pass authentication.
- Use RESTful API operations to configure and manage the device. RESTful API operations include Get, Put, Post, and Delete.

The device supports using HTTP or HTTPS to transfer RESTful packets.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

RESTful access over HTTP is not supported in FIPS mode.

Configuring RESTful access over HTTP

1. Enter system view.
   system-view
2. (Optional.) Specify the service port number for RESTful access over HTTP.
   restful http port *port-number*
   By default, the service port number for RESTful access over HTTP is 80.
3. Enable RESTful access over HTTP.
   restful http enable
   By default, RESTful access over HTTP is disabled.
4. Create a local user and enter local user view.
   local-user *user-name* [**class manage**]
5. Configure a password for the local user.
   password [{ **hash** | **simple** } *password*]
6. (Optional.) Assign a user role to the local user.
   authorization-attribute user-role *user-role*
   The default user role is network-operator for a RESTful access user.
7. Specify the HTTP service for the local user.
   service-type http
   By default, no service type is specified for a local user.
8. (Optional.) Use an ACL to control TCP connections from RESTful clients.
   IPv4:
   http acl { *advanced-acl-number* | *basic-acl-number* }
   IPv6:

http ipv6 acl { *advanced-acl-number* | *basic-acl-number* }

By default, all RESTful clients can establish TCP connections to the device.

Configuring RESTful access over HTTPS

1. Enter system view.

   system-view

2. (Optional.) Apply an SSL server policy to the RESTful access over HTTPS service.

   restful https ssl-server-policy *policy-name*

   By default, no SSL server policy is applied to the RESTful access over HTTPS service.

   The RESTful access over HTTPS service will use the SSL server policy to enhance service security. For more information about SSL server policies, see SSL configuration in *Security Configuration Guide*.

3. (Optional.) Specify the service port number for RESTful access over HTTPS.

   restful https port *port-number*

   By default, the service port number for RESTful access over HTTPS is 443.

4. Enable RESTful access over HTTPS.

   restful https enable

   By default, RESTful access over HTTPS is disabled.

5. Create a local user and enter local user view.

   local-user *user-name* [**class manage**]

6. Configure a password for the local user.

   In non-FIPS mode:

   password [{ **hash** | **simple** } *password*]

   In FIPS mode:

   password

7. (Optional.) Assign a user role to the local user.

   authorization-attribute user-role *user-role*

   The default user role is network-operator for a RESTful access user.

8. Specify the HTTPS service for the local user.

   service-type https

   By default, no service type is specified for a local user.

9. (Optional.) Use an ACL to control TCP connections from RESTful clients.

   IPv4:

   https acl { *advanced-acl-number* | *basic-acl-number* }

   IPv6:

   https ipv6 acl { *advanced-acl-number* | *basic-acl-number* }

   By default, all RESTful clients can establish TCP connections to the device.

Controlling user access to the device

About login user access control

Use ACLs to prevent unauthorized access, and configure command authorization and accounting to monitor and control user behavior.

To control user access, specify an ACL that has rules so only users permitted by the ACL can access the device.

- If no ACL is applied, all users can access the device.
- If the ACL for Telnet, SSH, or SNMP access control does not exist or does not have rules, no Telnet, SSH, or SNMP users can access the device.
- If a VPN instance is specified in an ACL rule, the rule applies only to the packets of the VPN instance. If no VPN instance is specified in an ACL rule, the rule applies only to the packets on the public network.

For more information about ACLs, see *ACL and QoS Configuration Guide.*

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Telnet is not supported in FIPS mode.

Controlling Telnet and SSH logins

Controlling Telnet logins

1. Enter system view.

   system-view

2. Apply an ACL to control Telnet logins.

   IPv4:

   telnet server acl { *advanced-acl-number* | *basic-acl-number* | **mac** *mac-acl-number* }

   IPv6:

   telnet server ipv6 acl { **ipv6** { *advanced-acl-number* | *basic-acl-number* } | **mac** *mac-acl-number* }

   By default, no ACL is used to control Telnet logins.

3. (Optional.) Enable logging for Telnet login attempts that are denied by the Telnet login control ACL.

   telnet server acl-deny-log enable

   By default, logging is disabled for Telnet login attempts that are denied by the Telnet login control ACL.

Controlling SSH logins

1. Enter system view.

   system-view

2. Apply an ACL to control SSH logins.

   IPv4:

   ssh server acl { *advanced-acl-number* | *basic-acl-number* | **mac** *mac-acl-number* }

   IPv6:

   ssh server ipv6 acl { **ipv6** { *advanced-acl-number* | *basic-acl-number* } | **mac** *mac-acl-number* }

   By default, no ACL is used to control SSH logins.

3. (Optional.) Enable logging for SSH login attempts that are denied by the SSH login control ACL.

   ssh server acl-deny-log enable

   By default, logging is disabled for SSH login attempts that are denied by the SSH login control ACL.

For more information about **ssh** commands, see *Security Command Reference*.

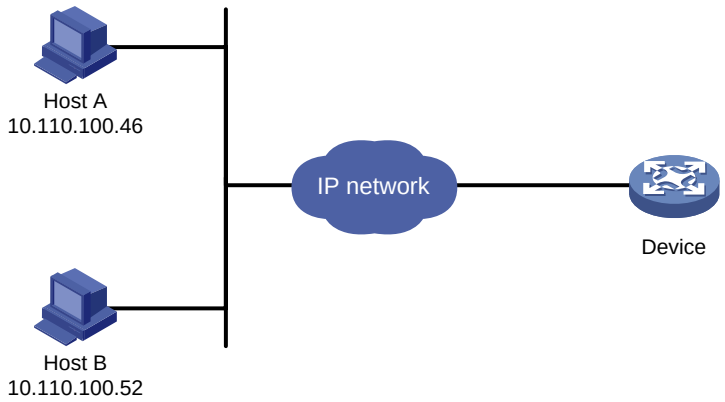
Example: Controlling Telnet login

Network configuration

As shown in Figure 6, the device is a Telnet server.

Configure the device to permit only Telnet packets sourced from Host A and Host B.

Figure 6 Network diagram



Procedure

Configure an ACL to permit packets sourced from Host A and Host B.

```
<Sysname> system-view
[Sysname] acl basic 2000 match-order config
[Sysname-acl-ipv4-basic-2000] rule 1 permit source 10.110.100.52 0
[Sysname-acl-ipv4-basic-2000] rule 2 permit source 10.110.100.46 0
[Sysname-acl-ipv4-basic-2000] quit
```

Apply the ACL to filter Telnet logins.

```
[Sysname] telnet server acl 2000
```

Controlling SNMP access

About SNMP access control

For information about SNMP access control, see SNMP in *Network Management and Monitoring Configuration Guide.*

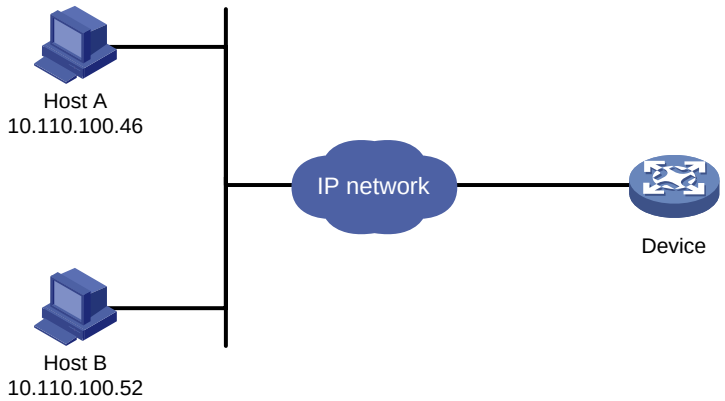
Example: Controlling SNMP access

Network configuration

As shown in Figure 7, the device is running SNMP.

Configure the device to allow Host A and Host B to access the device through SNMP.

Figure 7 Network diagram



Procedure

Create an ACL to permit packets sourced from Host A and Host B.

```
<Sysname> system-view
[Sysname] acl basic 2000 match-order config
[Sysname-acl-ipv4-basic-2000] rule 1 permit source 10.110.100.52 0
[Sysname-acl-ipv4-basic-2000] rule 2 permit source 10.110.100.46 0
[Sysname-acl-ipv4-basic-2000] quit
```

Associate the ACL with the SNMP community and the SNMP group.

```
[Sysname] snmp-agent community read aaa acl 2000
[Sysname] snmp-agent group v2c groupa acl 2000
[Sysname] snmp-agent usm-user v2c usera groupa acl 2000
```

Configuring command authorization

About command authorization

By default, commands available for a user depend only on the user's user roles. When the authentication mode is scheme, you can configure the command authorization feature to further control access to commands.

After you enable command authorization, a user can use only commands that are permitted by both the AAA scheme and user roles.

Restrictions and guidelines

When command authorization is enabled, commands available for a user vary by the user's login authentication mode.

- If authentication is disabled or password authentication is enabled, command authorization does not take effect, and the user cannot use any commands.
- If scheme authentication is enabled, commands available for a user vary by the user's access authentication method.
  - If local authentication is used, the device uses the user roles assigned to the user to perform command authorization.
  - If remote authentication is used, the remote authorization server performs command authorization to determine whether a command entered by a login user is permitted. If remote authorization fails, the device uses the user roles of a local user with the same name as the login user to determine whether the command can be used. If the authorization also fails, the login user cannot use the command.

Command authorization configuration changes in user line class view do not take effect on the current session. The changes take effect only on subsequent login sessions. Command authorization configuration changes in user line view take effect immediately on all users that access the user line.

If the remote server performs command authorization, you must configure a command authorization method in ISP domain view. For more information, see AAA configuration in *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter user line view or user line class view.
   - Enter user line view.

     line { *first-number1* [*last-number1*] | { **console** | **vty** } *first-number2* [*last-number2*] }
   - Enter user line class view.

     line class { **console** | **vty** }

   A setting in user line view applies only to the user line. A setting in user line class view applies to all user lines of the class. A non-default setting in either view takes precedence over the default setting in the other view. A non-default setting in user line view takes precedence over the non-default setting in user line class view.

   A setting in user line class view takes effect only on users who log in after the setting is made. It does not affect users who are already online when the setting is made.

3. Enable scheme authentication.

   In non-FIPS mode:

   authentication-mode scheme

   By default:
   - Authentication is disabled for console login.
   - Password authentication is enabled for VTY login.

In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

In FIPS mode:

authentication-mode scheme

By default, scheme authentication is enabled.

In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

CAUTION:

When you enable scheme authentication, make sure an authentication user account is available. If no authentication user account is available, you cannot log in to the device through the line or line class at the next time.

4. Enable command authorization.

   command authorization

   By default, command authorization is disabled, and the commands available for a user depend only on the user role.

   If the **command authorization** command is executed in user line class view, command authorization is enabled on all user lines in the class. You cannot execute the **undo command authorization** command in the view of a user line in the class.

Example: Configuring command authorization

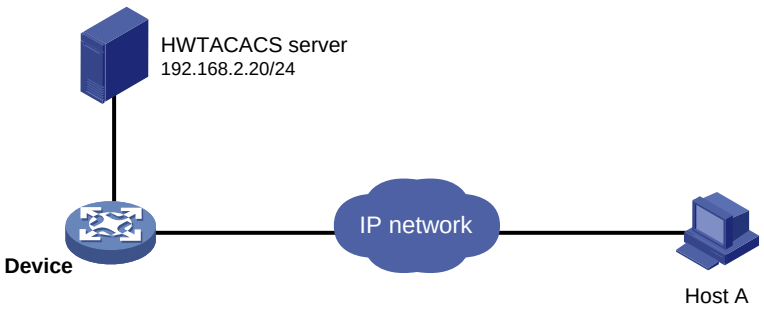
Network configuration

As shown in Figure 8, Host A needs to log in to the device to manage the device.

Configure the device to perform the following operations:

- Allow Host A to Telnet in after authentication.
- Use the HWTACACS server to control the commands that the user can execute.
- If the HWTACACS server is not available, use local authorization.

Figure 8 Network diagram



Procedure

Assign IP addresses to relevant interfaces. Make sure the device and the HWTACACS server can reach each other. Make sure the device and Host A can reach each other. (Details not shown.)

Enable the Telnet server.

```
<Device> system-view
[Device] telnet server enable
```

Enable scheme authentication for user lines VTY 0 through VTY 63.

```
[Device] line vty 0 63
[Device-line-vty0-63] authentication-mode scheme
```

Enable command authorization for the user lines.

```
[Device-line-vty0-63] command authorization
[Device-line-vty0-63] quit
```

Create HWTACACS scheme **tac**.

```
[Device] hwtacacs scheme tac
```

Configure the scheme to use the HWTACACS server at 192.168.2.20:49 for authentication and authorization.

```
[Device-hwtacacs-tac] primary authentication 192.168.2.20 49
[Device-hwtacacs-tac] primary authorization 192.168.2.20 49
```

Set the shared keys to **expert**.

```
[Device-hwtacacs-tac] key authentication simple expert
[Device-hwtacacs-tac] key authorization simple expert
```

Remove domain names from usernames sent to the HWTACACS server.

```
[Device-hwtacacs-tac] user-name-format without-domain
[Device-hwtacacs-tac] quit
```

Configure the system-defined domain (**system**).

```
[Device] domain system
```

Use HWTACACS scheme **tac** for login user authentication and command authorization. Use local authentication and local authorization as the backup method.

```
[Device-isp-system] authentication login hwtacacs-scheme tac local
[Device-isp-system] authorization command hwtacacs-scheme tac local
[Device-isp-system] quit
```

Create local user **monitor**. Set the simple password to **hello12345**, the service type to Telnet, and the default user role to level-1.

```
[Device] local-user monitor
[Device-luser-manage-monitor] password simple hello12345
[Device-luser-manage-monitor] service-type telnet
[Device-luser-manage-monitor] authorization-attribute user-role level-1
```

Configuring command accounting

About command accounting

Command accounting uses the HWTACACS server to record all executed commands to monitor user behavior on the device.

If command accounting is enabled but command authorization is not, every executed command is recorded. If both command accounting and command authorization are enabled, only authorized commands that are executed are recorded.

Restrictions and guidelines

The command accounting method can be the same as or different from the command authorization method and user login authorization method.

For the command accounting feature to take effect, you must configure a command accounting method in ISP domain view. For more information, see *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter user line view or user line class view.

   - Enter user line view.

     line { *first-number1* [*last-number1*] | { **console** | **vty** } *first-number2* [*last-number2*] }

   - Enter user line class view.

     line class { **console** | **vty** }

   A setting in user line view applies only to the user line. A setting in user line class view applies to all user lines of the class. A non-default setting in either view takes precedence over the default setting in the other view. A non-default setting in user line view takes precedence over the non-default setting in user line class view.

   A setting in user line class view takes effect only on users who log in after the setting is made. It does not affect users who are already online when the setting is made.

3. Enable scheme authentication.

   In non-FIPS mode:

   authentication-mode scheme

   By default:

   - Authentication is disabled for console login.
   - Password authentication is enabled for VTY login.

   In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

   In FIPS mode:

   authentication-mode scheme

   By default, scheme authentication is enabled.

   In VTY line view, this command is associated with the **protocol inbound** command. If one command has a non-default setting in VTY line view, the other command uses its setting in VTY line view, regardless of its setting in VTY line class view.

   CAUTION:

   When you enable scheme authentication, make sure an authentication user account is available. If no authentication user account is available, you cannot log in to the device through the line or line class at the next time.

4. Enable command accounting.

   command accounting

By default, command accounting is disabled. The accounting server does not record the commands executed by users.

If the **command accounting** command is executed in user line class view, command accounting is enabled on all user lines in the class. You cannot execute the **undo command accounting** command in the view of a user line in the class.

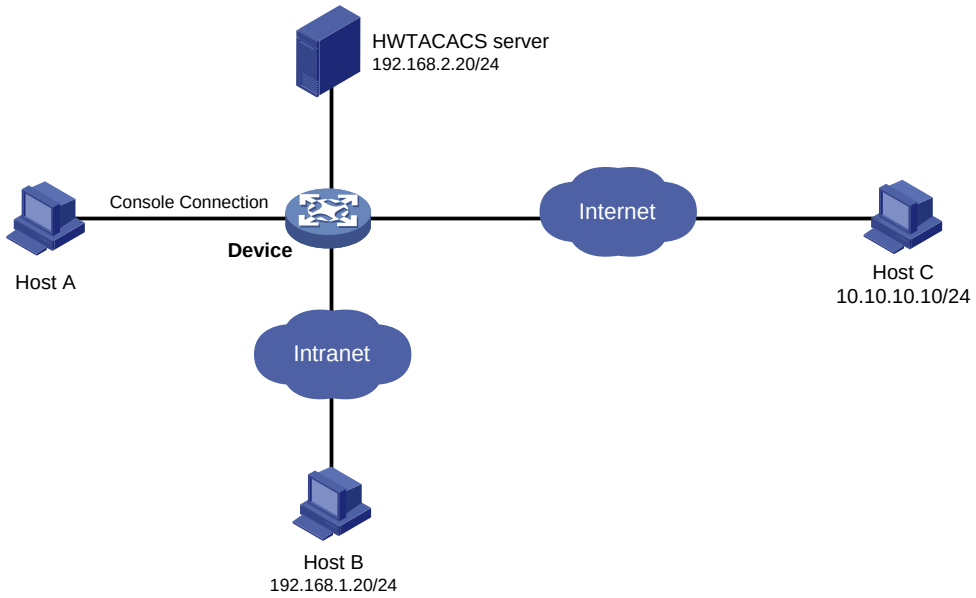
Example: Configuring command accounting

Network configuration

As shown in Figure 9, users need to log in to the device to manage the device.

Configure the device to send commands executed by users to the HWTACACS server to monitor and control user operations on the device.

Figure 9 Network diagram



Procedure

Enable the Telnet server.

```
<Device> system-view
[Device] telnet server enable
```

Enable command accounting for user line console 0.

```
[Device] line console 0
[Device-line-console0] command accounting
[Device-line-console0] quit
```

Enable command accounting for user lines VTY 0 through VTY 63.

```
[Device] line vty 0 63
[Device-line-vty0-63] command accounting
[Device-line-vty0-63] quit
```

Create HWTACACS scheme **tac**.

```
[Device] hwtacacs scheme tac
```

Configure the scheme to use the HWTACACS server at 192.168.2.20:49 for accounting.

```
[Device-hwtacacs-tac] primary accounting 192.168.2.20 49
```

Set the shared key to **expert**.

```
[Device-hwtacacs-tac] key accounting simple expert
```

Remove domain names from usernames sent to the HWTACACS server.

```
[Device-hwtacacs-tac] user-name-format without-domain
[Device-hwtacacs-tac] quit
```

Configure the system-defined domain (**system**) to use the HWTACACS scheme for command accounting.

```
[Device] domain system
[Device-isp-system] accounting command hwtacacs-scheme tac
[Device-isp-system] quit
```

Contents

Configuring FTP 2
About FTP 2
FTP file transfer modes 2
FTP operation modes 2
FIPS compliance 2
Using the device as an FTP server 2
FTP server configuration tasks at a glance 2
Restrictions and guidelines for FTP server configuration 3
Enabling the FTP server 3
Configuring client authentication and authorization 3
Configuring FTP server access control 3
Setting connection management parameters 4
Specifying an SSL server policy for SFTP connections 4
Setting the DSCP value for outgoing FTP packets 5
Configuring the alarm threshold and alarm clearance threshold for FTP login failures 5
Manually releasing FTP connections 5
Display and maintenance commands for the FTP server 5
Example: Using the device as an FTP server 6
Using the device as an FTP client 7
FTP client configuration tasks at a glance 7
Establishing an FTP connection 7
Displaying command help information 8
Displaying directories and files on the FTP server 9
Managing directories on the FTP server 9
Managing directories on the FTP client 9
Working with files on the FTP server 10
Changing to another user account 10
Maintaining and troubleshooting the FTP connection 11
Terminating the FTP connection 11
Display and maintenance commands for the FTP client 12
Example: Using the device as an FTP client 12
Configuring TFTP 13
About TFTP 13
FIPS compliance 13
Restrictions and guidelines: TFTP configuration 13
Configuring the device as a TFTP server 13
Configuring and using the IPv4 TFTP client 14
Configuring and using the IPv6 TFTP client 14

Configuring FTP

About FTP

File Transfer Protocol (FTP) is an application layer protocol for transferring files from one host to another over an IP network. It uses TCP port 20 to transfer data and TCP port 21 to transfer control commands.

FTP is based on the client/server model. The device can act as the FTP server or FTP client.

FTP file transfer modes

FTP supports the following transfer modes:

- **Binary mode**—Used to non-text files, such as **.app**, **.bin**, and **.btm** files.
- **ASCII mode**—Used to transfer text files, such as **.txt**, **.bat**, and **.cfg** files.

When the device acts as the FTP client, you can set the transfer mode (**binary** by default). When the device acts as the FTP server, the transfer mode is determined by the FTP client.

FTP operation modes

FTP can operate in either of the following modes:

- **Active mode (PORT)**—The FTP server initiates the TCP connection. This mode is not suitable when the FTP client is behind a firewall, for example, when the FTP client resides in a private network.
- **Passive mode (PASV)**—The FTP client initiates the TCP connection. This mode is not suitable when the server does not allow the client to use a random unprivileged port greater than 1024.

FTP operation mode varies depending on the FTP client program.

FIPS compliance

FTP is not supported in FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Using the device as an FTP server

To use the device as an FTP server, you must enable the FTP server and configure authentication and authorization on the device. Other commands are optional.

FTP server configuration tasks at a glance

To use the device as an FTP server, perform the following tasks:

1. Enabling the FTP server
2. Configuring client authentication and authorization
3. (Optional.) Configuring FTP server access control
4. (Optional.) Setting connection management parameters
5. (Optional.) Specifying an SSL server policy for SFTP connections

6. (Optional.) Setting the DSCP value for outgoing FTP packets
7. (Optional.) Configuring the alarm threshold and alarm clearance threshold for FTP login failures
8. (Optional.) Manually releasing FTP connections

Restrictions and guidelines for FTP server configuration

FTP might be insecure because it transmits data in plaintext form.

Enabling the FTP server

1. Enter system view.
   ```
   system-view
   ```
2. Enable the FTP server.
   ```
   ftp server enable
   ```
   By default, the FTP server is disabled.

Configuring client authentication and authorization

Perform this task on the FTP server to authenticate FTP clients and set the authorized directories that authenticated clients can access.

The following authentication modes are available:

- **Local authentication**—The device looks up the client's username and password in the local user account database. If a match is found, authentication succeeds.
- **Remote authentication**—The device sends the client's username and password to a remote authentication server for authentication. The user account is configured on the remote authentication server rather than the device.

The following authorization modes are available:

- **Local authorization**—The device assigns authorized directories to FTP clients based on the locally configured authorization attributes.
- **Remote authorization**—A remote authorization server assigns authorized directories on the device to FTP clients.

For more information about configuring authentication and authorization, see AAA in *Security Configuration Guide*.

Configuring FTP server access control

About this task

To control FTP access, specify an ACL that exists and has rules so that only FTP clients permitted by the ACL can access the FTP server.

Restrictions and guidelines

When no ACL is applied, all FTP clients can access the FTP server. To control FTP access, specify an ACL that exists and has rules so only FTP clients permitted by the ACL can access the FTP server. If you specify an ACL that does not exist or does not have rules, no FTP clients can access the FTP server.

If a VPN instance is specified in an ACL rule, the rule applies only to the packets of the VPN instance. If no VPN instance is specified in an ACL rule, the rule applies only to the packets on the public network.

The ACL takes effect only for FTP connections to be established. It does not impact existing FTP connections.

For more information about ACL, see *ACL and QoS Configuration Guide.*

If you configure FTP server access control multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.

   system-view

2. Use an ACL to control access to the FTP server.

   ftp server acl { *advanced-acl-number* | *basic-acl-number* | **ipv6** { *advanced-acl-number* | *basic-acl-number* } }

   By default, no ACL is used for access control.

3. Enable logging for FTP login attempts that are denied by the FTP login control ACL.

   ftp server acl-deny-log enable

   By default, logging is disabled for FTP login attempts that are denied by the FTP login control ACL.

Setting connection management parameters

1. Enter system view.

   system-view

2. Set the FTP connection idle-timeout timer.

   ftp timeout *minutes*

   By default, the FTP connection idle-timeout timer is 30 minutes.

   If no data transfer occurs on an FTP connection before the idle-timeout timer expires, the FTP server closes the FTP connection.

3. Set the maximum number of concurrent FTP users.

   aaa session-limit ftp *max-sessions*

   By default, the maximum number of concurrent FTP users is 32.

   Changing this setting does not affect users who are currently online. If the new limit is less than the number of online FTP users, no additional FTP users can log in until the number drops below the new limit. For more information about this command, see *Security Command Reference*.

Specifying an SSL server policy for SFTP connections

About this task

After you associate an SSL server policy with the device, a client that supports SFTP will establish a secure connection to the device to ensure data security.

Procedure

1. Enter system view.

   system-view

2. Associate an SSL server policy with the FTP server to ensure data security.

   ftp server ssl-server-policy *policy-name*

   By default, no SSL server policy is associated with the FTP server.

Setting the DSCP value for outgoing FTP packets

1. Enter system view.
   system-view
2. Set the DSCP value for outgoing FTP packets.
   IPv4:
   ftp server dscp *dscp-value*
   IPv6:
   ftp server ipv6 dscp *dscp-value*
   By default, the DSCP value is 0.

Configuring the alarm threshold and alarm clearance threshold for FTP login failures

1. Enter system view.
   system-view
2. Set the alarm threshold and alarm clearance threshold for FTP login failures in a statistics period.
   ftp server login-failed threshold-alarm upper-limit *report-times* **lower-limit** *resume-times* **period** *period-time*
   By default, the statistics period is five minutes and the alarm threshold and alarm clearance threshold for FTP login failures in the statistics period is 30 and 20, respectively.
   After you execute this command, if the number of FTP login failures reaches the alarm threshold in a statistics period, an alarm message is generated. If the number of FTP login failures drops below the alarm clearance threshold in a statistics period, an alarm clearance message is generated. This helps you obtain real-time information about FTP logins.

Manually releasing FTP connections

To manually release FTP connections, execute the following commands in user view:

- Release the FTP connection established by using a specific user account:
  free ftp user *username*
- Release the FTP connection to a specific IP address:
  free ftp user-ip [**ipv6**] *ip-address* [**port** *port*]

Display and maintenance commands for the FTP server

Execute **display** commands in any view.

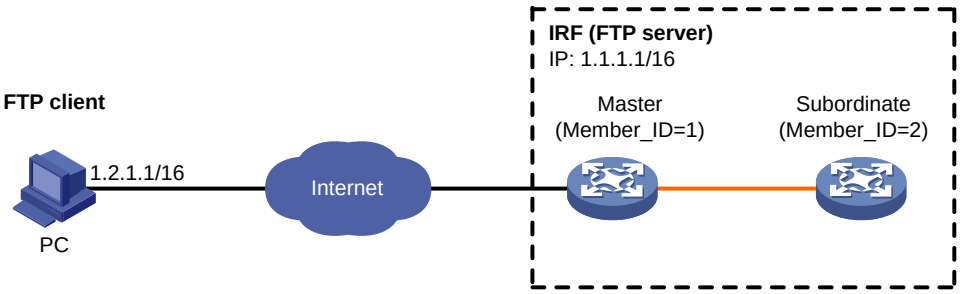
Task	Command
Display FTP server configuration and status information.	**display ftp-server**
Display detailed information about online FTP users.	**display ftp-user**

Example: Using the device as an FTP server

Network configuration

- Configure the IRF fabric as an FTP server.
- Create a local user account named **abc** on the FTP server. Set the password to **hello12345**.
- Use the user account to log in to the FTP server from the FTP client.
- Upload the **temp.bin** file from the FTP client to the FTP server.

Figure 1 Network diagram



Note: The orange line represents an IRF connection.

Procedure

1. Configure IP addresses as shown in Figure 1. Make sure the IRF fabric and the PC can reach each other. (Details not shown.)
2. Configure the FTP server:

   # Examine the storage space on the member devices. If the free space is insufficient, use the **`delete/unreserved`** *`file`* command to delete unused files. (Details not shown.)

   # Create a local user with username **abc** and password **hello12345**.

   ```
   <Sysname> system-view
   [Sysname] local-user abc class manage
   [Sysname-luser-manage-abc] password simple hello12345
   ```

   # Assign the **network-admin** user role to the user. Set the working directory to the root directory of the flash memory on the master. (To set the working directory to the root directory of the flash memory on the subordinate member, you must include the slot number in the directory path.)

   ```
   [Sysname-luser-manage-abc] authorization-attribute user-role network-admin
   work-directory flash:/
   ```

   # Assign the service type FTP to the user.

   ```
   [Sysname-luser-manage-abc] service-type ftp
   [Sysname-luser-manage-abc] quit
   ```

   # Enable the FTP server.

   ```
   [Sysname] ftp server enable
   [Sysname] quit
   ```

3. Perform FTP operations from the FTP client:

   # Log in to the FTP server at 1.1.1.1 using username **abc** and password **hello12345**.

   ```
   c:\> ftp 1.1.1.1
   Connected to 1.1.1.1.
   220 FTP service ready.
   User(1.1.1.1:(none)):abc
   ```

```
331 Password required for abc.
Password:
230 User logged in.
```

Use the ASCII mode to download configuration file **config.cfg** from the FTP server to the PC for backup.

```
ftp> ascii
200 TYPE is now ASCII
ftp> get config.cfg back-config.cfg
```

Use the binary mode to upload the **temp.bin** file from the PC to the root directory of the flash memory on the master.

```
ftp> binary
200 TYPE is now 8-bit binary
ftp> put temp.bin
```

Exit FTP.

```
ftp> bye
```

Using the device as an FTP client

FTP client configuration tasks at a glance

To use the device as an FTP server, perform the following tasks:

1. Establishing an FTP connection
2. (Optional.) Displaying command help information
3. (Optional.) Displaying directories and files on the FTP server
4. (Optional.) Managing directories on the FTP server
5. (Optional.) Working with files on the FTP server
6. (Optional.) Changing to another user account
7. (Optional.) Maintaining and troubleshooting the FTP connection
8. (Optional.) Terminating the FTP connection

Establishing an FTP connection

FTP connection establishment task list

To establish an FTP connection, perform the following tasks:

1. (Optional.) Specifying a source IP address for outgoing FTP packets
2. Establishing an FTP connection
3. Setting the FTP file transfer mode and operation mode

Restrictions and guidelines

The source IP address specified in the **ftp** command takes precedence over the one set by the **ftp client source** command.

The source IP address specified in the **ftp ipv6** command takes precedence over the one set by the **ftp client ipv6 source** command.

Specifying a source IP address for outgoing FTP packets

1. Enter system view.

   system-view

2. Specify a source IP address for outgoing FTP packets.

   IPv4:

   ftp client source { **interface** *interface-type interface-number* | **ip** *source-ip-address* }

   By default, no source IP address is specified. The device uses the primary IP address of the output interface as the source IP address.

   IPv6:

   ftp client ipv6 source { **interface** *interface-type interface-number* | **ipv6** *source-ipv6-address* }

   By default, no source IPv6 address is specified. The source address is automatically selected as defined in RFC 3484.

Establishing an FTP connection

- Log in to the FTP server from user view.

  IPv4:

  ftp [*ftp-server* [*service-port*] [**vpn-instance** *vpn-instance-name*] [**dscp** *dscp-value* | **source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* } | **-d**] *]

  IPv6:

  ftp ipv6 [*ftp-server* [*service-port*] [**vpn-instance** vpn-instance-name] [**dscp** *dscp-value* | **source** { **interface** *interface-type interface-number* | **ipv6** *source-ipv6-address* } | **-d**] * [**-i** *interface-type interface-number*]]

- Log in to the FTP server from FTP client view.

  a. Enter FTP client view.

     ftp [**ipv6**]

  b. Log in to the FTP server.

     open *server-address* [*service-port*]

Setting the FTP file transfer mode and operation mode

1. Enter FTP client view from user view.

   ftp

2. Set the file transfer mode.

   - Set the file transfer mode to ASCII.

     ascii

   - Set the file transfer mode to binary.

     binary

   The default file transfer mode is binary.

3. Change the FTP operation mode.

   passive

   The default FTP operation mode is passive.

Displaying command help information

1. Enter FTP client view from user view.

   ftp

2. Display command help information.

- **help** [*command-name*]
- **?** [*command-name*]

Displaying directories and files on the FTP server

1. Enter FTP client view from user view.
   ftp
2. Display directories and files on the FTP server.
   - **dir** [*remotefile* [*localfile*]]
   - **ls** [*remotefile* [*localfile*]]

Managing directories on the FTP server

Prerequisites

Use the **dir** or **ls** command to display the directories and files on the FTP server.

Procedure

1. Enter FTP client view from user view.
   ftp
2. Manage directories on the FTP server.
   - Display the working directory that is being accessed on the FTP server.
     pwd
   - Change the working directory on the FTP server.
     cd { *directory* | **..** | **/** }
   - Return to the upper level directory on the FTP server.
     cdup
   - Create a directory on the FTP server.
     mkdir *directory*
   - Delete a directory from the remote FTP server.
     rmdir *directory*

CAUTION:
Delete a directory from the FTP server with caution. When you delete a directory from the FTP server, make sure the directory is no longer in use.

Managing directories on the FTP client

1. Enter FTP client view from user view.
   ftp
2. Display or change the local working directory of the FTP client.
   lcd [*directory* | **/**]
   To upload a file, use this command to change to the directory where the file resides. Downloaded files are saved in the working directory.

Working with files on the FTP server

Prerequisites

Use the **dir** or **ls** command to display the directories and files on the FTP server.

Procedure

1. Enter FTP client view from user view.
   ftp
2. Work with files on the FTP server.
   - Delete a file from the FTP server permanently.
     delete *remotefile*

     CAUTION:
     Delete a file from the FTP server permanently with caution. When you delete a file from the FTP server permanently, make sure the file is no longer in use.

     This command requires that you have the delete right.
   - Rename a file.
     rename [*oldfilename* [*newfilename*]]
   - Upload a file to the FTP server.
     put *localfile* [*remotefile*]
   - Download a file from the FTP server.
     get *remotefile* [*localfile*]
   - Add the content of a file on the FTP client to a file on the FTP server.
     append *localfile* [*remotefile*]
   - Specify the retransmit marker.
     restart *marker*

     Use this command together with the **put**, **get**, or **append** command.
   - Update a local file.
     newer *remotefile*
   - Get the missing part of a file.
     reget *remotefile* [*localfile*]

Changing to another user account

About this task

After you log in to the FTP server, you can initiate an FTP authentication to change to a new account. By changing to a new account, you can get a different privilege without re-establishing the FTP connection.

Restrictions and guidelines

For successful account change, you must enter the new username and password correctly. A wrong username or password can cause the FTP connection to be disconnected.

Procedure

1. Enter FTP client view from user view.
   ftp

2. Initiate an FTP authentication on the current FTP connection.
   user *username* [*password*]

Maintaining and troubleshooting the FTP connection

About this task

After you use the device to establish an FTP connection to the FTP server, use the commands in this section to locate and troubleshoot problems about the FTP connection.

Procedure

1. Enter FTP client view from user view.
   ftp
2. Maintain and troubleshoot the FTP connection.
   - Display FTP commands supported by the FTP server.
     rhelp
   - Display help information about an FTP command that is supported by the FTP server.
     rhelp *protocol-command*
   - Display FTP server status.
     rstatus
   - Display detailed information about a directory or file on the FTP server.
     rstatus *remotefile*
   - Display FTP connection status.
     status
   - Display the system information of the FTP server.
     system
   - Enable or disable FTP operation information display.
     verbose

     By default, this function is enabled.
   - Enable FTP client debugging.
     debug

     By default, FTP client debugging is disabled.
   - Clear the reply information in the buffer.
     reset

Terminating the FTP connection

1. Enter FTP client view from user view.
   ftp
2. Terminate the connection.
   - Terminate the connection to the FTP server without exiting FTP client view.
     disconnect
     close
   - Terminate the connection to the FTP server and return to user view.
     bye
     quit

Display and maintenance commands for the FTP client

Execute the **display** command in any view.

Task	Command
Display source IP address information on the FTP client.	**display ftp client source**

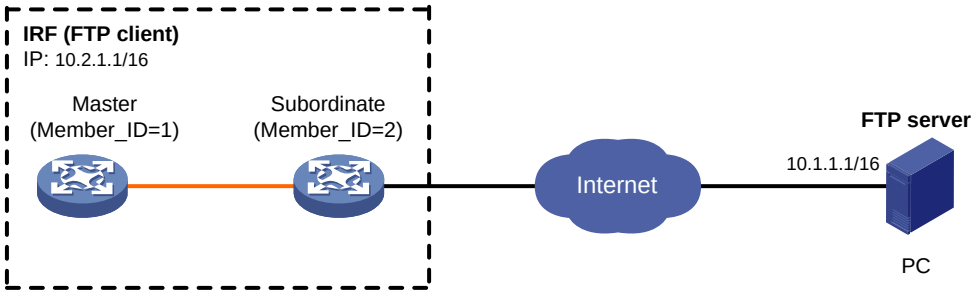
Example: Using the device as an FTP client

Network configuration

As shown in Figure 2, the PC is acting as an FTP server. A user account with username **abc** and password **hello12345** has been created on the PC.

- Use the IRF fabric as an FTP client to log in to the FTP server.
- Download the **temp.bin** file from the FTP server to the FTP client.

Figure 2 Network diagram



Note: The orange line represents an IRF connection.

Procedure

Configure IP addresses as shown in Figure 2. Make sure the IRF fabric and PC can reach each other. (Details not shown.)

Examine the storage space on the member devices. If the free space is insufficient, use the **delete/unreserved** *file* command to delete unused files. (Details not shown.)

Log in to the FTP server at 10.1.1.1 using username **abc** and password **hello12345**.

```
<Sysname> ftp 10.1.1.1
Press CTRL+C to abort.
Connected to 10.1.1.1 (10.1.1.1).
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User (10.1.1.1:(none)): abc
331 Give me your password, please
Password:
230 Logged in successfully
Remote system type is MSDOS.
ftp>
```

Set the file transfer mode to binary.

```
ftp> binary
200 TYPE is now 8-bit binary
```

Download the **temp.bin** file from the PC to the root directory of the flash memory on the master device.

```
ftp> get temp.bin
local: temp.bin remote: temp.bin
150 Connecting to port 47457
226 File successfully transferred
23951480 bytes received in 95.399 seconds (251.0 kbyte/s)
```

Download the **temp.bin** file from the PC to the root directory of the flash memory on the subordinate member (with member ID of 2).

```
ftp> get temp.bin slot2#flash:/temp.bin
```

Use the ASCII mode to upload configuration file **config.cfg** from the IRF fabric to the PC for backup.

```
ftp> ascii
200 TYPE is now ASCII
ftp> put config.cfg back-config.cfg
local: config.cfg remote: back-config.cfg
150 Connecting to port 47461
226 File successfully transferred
3494 bytes sent in 5.646 seconds (618.00 kbyte/s)
ftp> bye
221-Goodbye. You uploaded 2 and downloaded 2 kbytes.
221 Logout.
<Sysname>
```

Configuring TFTP

About TFTP

Trivial File Transfer Protocol (TFTP) is a simplified version of FTP for file transfer over secure reliable networks. TFTP uses UDP port 69 for data transmission. In contrast to TCP-based FTP, TFTP does not require authentication or complex message exchanges, and is easier to deploy. TFTP is suited for reliable network environments.

FIPS compliance

TFTP is not supported in FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Restrictions and guidelines: TFTP configuration

You can upload a file from a TFTP client to the TFTP server or download a file from the TFTP server to a TFTP client. As a best practice, specify a file name that has not been used in the destination directory. If the destination file name has been used in the destination directory, the device deletes the existing file before saving the new file.

Configuring the device as a TFTP server

1. Enter system view.

system-view

2. Enable the TFTP server.

   tftp server enable

   By default, the TFTP server is disabled.

   The device can act as the TFTP server on both IPv4 and IPv6 networks.

3. Set the TFTP server working directory.

   tftp server work-directory *directory*

   By default, the TFTP server working directory is the root directory of the default file system.

   Only a directory on the master device can be specified as the TFTP server working directory.

Configuring and using the IPv4 TFTP client

1. Enter system view.

   system-view

2. Specify the source IP address for TFTP packets sent by the TFTP client.

   tftp client source { **interface** *interface-type interface-number* | **ip** *source-ip-address* }

   By default, no source IP address is specified. The device uses the primary IP address of the output interface as the source IP address.

3. Return to user view.

   quit

4. Download or upload a file in an IPv4 network.

   tftp *tftp-server* { **get** | **put** | **sget** } *source-filename* [*destination-filename*] [**vpn-instance** *vpn-instance-name*] [**dscp** *dscp-value* | **source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }] *

   The source IP address specified in this command takes precedence over the source IP address set by using the **tftp client source** command.

Configuring and using the IPv6 TFTP client

1. Enter system view.

   system-view

2. Specify the source IPv6 address for TFTP packets sent by the TFTP client.

   tftp client ipv6 source { **interface** *interface-type interface-number* | **ipv6** *source-ipv6-address* }

   By default, no source IPv6 address is specified. The source address is automatically selected as defined in RFC 3484.

3. Return to user view.

   quit

4. Download or upload a file in an IPv6 network.

   tftp ipv6 *tftp-server* [**-i** *interface-type interface-number*] { **get** | **put** | **sget** } *source-filename* [*destination-filename*] [**vpn-instance** *vpn-instance-name*] [**dscp** *dscp-value* | **source** { **interface** *interface-type interface-number* | **ipv6** *source-ipv6-address* }] *

The source IP address specified in this command takes precedence over the one set by the **`tftp client ipv6 source`** command.

Contents

Managing file systems 1

About file system management 1
Storage media and file systems 1
Directories 2
Files 2
Specifying a directory name or file name 3
FIPS compliance 3
Restrictions and guidelines: File system management 3
Managing storage media and file systems 4
Partitioning a storage medium 4
Mounting or unmounting a file system 4
Formatting a file system 5
Repairing a file system 5
Managing files and directories 5
Setting the operation mode for files and directories 5
Displaying file and directory information 6
Displaying the contents of a text file 6
Displaying the working directory 6
Changing the working directory 6
Creating a directory 6
Renaming a file or directory 6
Copying a file 7
Moving a file 7
Deleting and restoring files 7
Deleting a directory 8
Archiving files and directories 8
Extracting files and directories 8
Compressing a file 9
Decompressing a file 9
Calculating the file digest 9
Executing a batch file 9

Managing file systems

This chapter describes how to manage file systems.

About file system management

Storage media and file systems

The device supports both fixed (the flash memory) and hot swappable (USB disk) storage media.

- The fixed storage medium has one file system.
- The hot swappable storage media can be partitioned. Each unpartitioned storage medium has one file system. On a partitioned storage medium, each partition has one file system.

Storage medium and file system naming conventions

The file system on the flash memory has the same name as the flash memory. The name has the following parts:

- Storage medium type **flash**.
- Colon (:).

A USB disk name and the file system names share the following parts:

- Storage medium type **usb**.
- Sequence number, a lower-case English letter such as a, b, or c.
- Partition number, a digit that starts at 0 and increments by 1. If the storage medium is not partitioned, the system determines that the storage medium has one partition. (The storage medium name does not contain a partition number.)
- Colon (:).

For example, the first USB disk is named **usba:**, and the file system on the first partition of the first USB disk is named **usba0:**.

IMPORTANT:

File system names are case sensitive and must be entered in lower case.

File system location

To identify a file system on the master device, you do not need to specify the file system location. To identify a file system on a subordinate member device, you must specify the file system location in the **slot***n***#** format. The *n* argument represents the IRF member ID of the member device. For example, the location is **slot2#** for a file system that resides on member device 2.

IMPORTANT:

The file system location string is case sensitive and must be entered in lower case.

Default file system

You are working with the default file system by default after you log in. To specify a file or directory on the default file system, you do not need to specify the file system name. For example, you do not need to specify any location information if you want to save the running configuration to the root directory of the default file system.

To change the default file system, use the Boot ROM menu. For more information, see the software release notes.

Directories

Directories in a file system are structured in a tree form.

Root directory

The root directory is represented by a forwarding slash (/). For example, **flash:/** represents the root directory of the flash memory.

Working directory

The working directory is also called the current directory.

Directory naming conventions

When you specify a name for a directory, follow these conventions:

- A directory name can contain letters, digits, and special characters except for asterisks (*), vertical bars (|), forward slashes (/), backward slashes (\), question marks (?), left angle brackets (<), right angle brackets (>), quotation marks ("), and colons (:).
- A directory whose name starts with a dot character (.) is a hidden directory. To prevent the system from hiding a directory, make sure the directory name does not start with a dot character.

Commonly used directories

The device has some factory-default directories. The system automatically creates directories during operation. These directories include:

- **diagfile**—Stores diagnostic information files.
- **logfile**—Stores log files.
- **seclog**—Stores security log files.
- **versionInfo**—Stores software version information files.

Files

File naming conventions

When you specify a name for a file, follow these conventions:

- A file name can contain letters, digits, and special characters except for asterisks (*), vertical bars (|), forward slashes (/), backward slashes (\), question marks (?), left angle brackets (<), right angle brackets (>), quotation marks ("), and colons (:).
- A file whose name starts with a dot character (.) is a hidden file. To prevent the system from hiding a file, make sure the file name does not start with a dot character.

Common file types

The device has some factory-default files and might create some files automatically during operation. The types of these files include:

- **.ipe file**—Compressed software image package file.
- **.bin file**—Software image file.
- **.cfg file**—Configuration file.
- **.mdb file**—Binary configuration file.
- **.log file**—Log file.

Specifying a directory name or file name

Specifying a directory name

To specify a directory, you can use the absolute path or a relative path. For example, the working directory is **flash:/**. To specify the **test2** directory in Figure 1, you can use the following methods:

- **flash:/test/test1/test2** (absolute path)
- **flash:/test/test1/test2/** (absolute path)
- **test/test1/test2** (relative path)
- **test/test1/test2/** (relative path)

Figure 1 Sample directory hierarchy



Specifying a file name

To specify a file, use the following methods:

- Enter the absolute path of the file and the file name in the format of *filesystem*/*directory1*/*directory2*/…/*directoryn*/*filename*, where *directoryn* is the directory in which the file resides.
- Enter the relative path of the file and the file name.

For example, the working directory is **flash:/**. The **samplefile.cfg** file is in the **test2** directory shown in Figure 1. To specify the file, you can use the following methods:

- **flash:/test/test1/test2/samplefile.cfg**
- **test/test1/test2/samplefile.cfg**

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Restrictions and guidelines: File system management

To avoid file system corruption, do not perform the following tasks during file system management:

- Install or remove storage media.
- Perform a master/subordinate switchover.

If you remove a storage medium while a directory or file on the medium is being accessed, the device might not recognize the medium when you reinstall it. To reinstall this kind of storage medium, perform one of the following tasks:

- If you were accessing a directory on the storage medium, change the working directory.
- If you were accessing a file on the storage medium, close the file.
- If another administrator was accessing the storage medium, unmount all partitions on the storage medium.

Make sure a USB disk is not write protected before an operation that requires the write right on the disk.

You cannot access a storage medium that is being partitioned, or a file system that is being formatted or repaired.

Before managing file systems, directories, and files, make sure you know the possible impact.

Managing storage media and file systems

Partitioning a storage medium

About this task

A storage medium can be divided into logical devices called partitions. Operations on one partition do not affect the other partitions.

Restrictions and guidelines

IMPORTANT:

Partitioning a storage medium clears all data on the medium.

The flash memory does not support partitioning.

Before partitioning a storage medium, perform the following tasks:

- Make sure no other users are accessing the medium.
- To partition a USB disk, make sure the disk is not write protected. If the disk is write protected, the partition operation will fail. To restore access to the USB disk, you must reinstall the disk or remount the file systems on the disk.

A partition must have a minimum of 32 MB of storage space.

The actual partition size and the specified partition size might have a difference of less than 5% of the storage medium's total size.

Prerequisites

Back up the files in the storage medium.

Procedure

To partition a storage medium, execute the following command in user view:

fdisk *medium* [*partition-number*]

To partition a storage medium evenly, specify the *partition-number* argument. To customize the sizes of partitions, do not specify the *partition-number* argument. The command will require you to specify a size for each partition.

Mounting or unmounting a file system

Restrictions and guidelines

You can mount or unmount only a file system that is on a hot-swappable storage medium.

You can unmount a file system only when no other users are accessing the file system.

To prevent a USB disk and the USB interface from being damaged, make sure the following requirements are met before unmounting file systems on the USB disk:

- The system has recognized the USB disk.
- The USB disk LED is not blinking.

Mounting a file system

To mount a file system, execute the following command in user view:

mount *filesystem*

File systems on a hot-swappable storage medium are automatically mounted when the storage medium is connected to the device. If the system cannot recognize a file system, you must mount the file system before you can access it.

Unmounting a file system

To unmount a file system, execute the following command in user view:

umount *filesystem*

To remove a hot-swappable storage medium from the device, you must first unmount all file systems on the storage medium to disconnect the medium from the device. Removing a connected hot-swappable storage medium might damage files on the storage medium or even the storage medium itself.

Formatting a file system

Restrictions and guidelines

You can format a file system only when no other users are accessing the file system.

Procedure

To format a file system, execute the following command in user view:

format *filesystem* [**low-level**]

CAUTION:

Formatting a file system permanently deletes all files and directories in the file system. If low-level formatting is performed, you will be unable to restore the deleted data. You cannot restore the deleted files or directories. If a startup configuration file exists in the file system, back up the file if necessary.

Repairing a file system

Restrictions and guidelines

If part of a file system is inaccessible, use this task to examine and repair the file system.

You can repair a file system only when no other users are accessing the file system.

Procedure

To repair a file system, execute the following command in user view:

fixdisk *filesystem*

Managing files and directories

Setting the operation mode for files and directories

About this task

The device supports the following operation modes:

- **alert**—The system prompts for confirmation when your operation might cause problems such as file corruption or data loss. This mode provides an opportunity for you to cancel a disruptive operation.
- **quiet**—The system does not prompt for confirmation when you perform a file or directory operation except the recycle bin emptying operation.

Procedure

1. Enter system view.
   system-view
2. Set the operation mode for files and directories.
   file prompt { **alert** | **quiet** }
   The default mode is alert.

Displaying file and directory information

To display file and directory information, execute the following command in user view:

dir [**/all**] [*file* | *directory* | **/all-filesystems**]

If multiple users perform file operations (for example, creating or deleting files or directories) at the same time, the output from this command might be incorrect.

Displaying the contents of a text file

To display the contents of a text file, execute the following command in user view:

more *file*

Displaying the working directory

To display the working directory, execute the following command in user view:

pwd

Changing the working directory

About this task

The default working directory is the root directory of the default file system on the master device.

Procedure

To change the working directory, execute the following command in user view:

cd { *directory* | **..** }

Creating a directory

To create a directory, execute the following command in user view:

mkdir *directory*

Renaming a file or directory

To rename a file or directory, execute the following command in user view:

rename { *source-file* | *source-directory* } { *dest-file* | *dest-directory* }

Copying a file

To copy a file, execute the command in user view.

In non-FIPS mode:

copy *source-file* { *dest-file* | *dest-directory* } [**vpn-instance** *vpn-instance-name*] [**source interface** *interface-type interface-number*]

In FIPS mode:

copy *source-file* { *dest-file* | *dest-directory* }

Moving a file

To move a file, execute the following command in user view:

move *source-file* { *dest-file* | *dest-directory* }

Deleting and restoring files

About this task

You can delete a file permanently or move it to the recycle bin of the file system. A file moved to the recycle bin can be restored, but a permanently deleted file cannot.

Each file system has a recycle bin. A recycle bin is a directory named **.trash** in the root directory of the file system.

Restrictions and guidelines

Files in the recycle bin occupy storage space. To release the occupied storage space, delete files from the recycle bin.

To delete files from the recycle bin, use the **reset recycle-bin** command. If you use the **delete** command, the recycle bin might not be able to operate correctly.

To display files in a recycle bin, use one of the following methods:

- Access the root directory of the file system and execute the **dir /all .trash** command.
- Access the recycle bin directory of the file system and execute the **dir** command.

Deleting a file

To delete a file, execute one of the following commands in user view:

- Delete a file by moving it to the recycle bin.
  delete *file*
- Delete a file permanently.
  delete /unreserved *file*

CAUTION:
The **delete /unreserved** *file* command deletes a file permanently. The file cannot be restored.

- Delete files from the recycle bin.

reset recycle-bin [**/force**]

CAUTION:

The files in a recycle bin can be restored by using the **undelete** command. If you delete a file from the recycle bin, that file cannot be restored. Before you delete files from a recycle bin, make sure the files are no longer in use.

Restoring a file

To restore a file from the recycle bin, execute the following command in user view:

undelete *file*

Deleting a directory

To delete a directory, execute the following command in user view:

rmdir *directory*

CAUTION:

To delete a directory, you must first delete all files and subdirectories in the directory permanently or move them to the recycle bin. If you move them to the recycle bin, executing the **rmdir** command to delete the directory will delete them permanently. Before you use the **rmdir** command to delete a directory, you must make sure the directory and its files and subdirectories are no longer in use.

Archiving files and directories

About this task

You can archive files and directories to a single file for purposes such as file backup. The original files and directories still exist.

When you archive files and directories, you can choose to compress the archive files so the archive files use less storage space.

Procedure

To archive files and directories, execute the following command in user view:

tar create [**gz**] **archive-file** *dest-file* [**verbose**] **source** { *source-file* | *source-directory* }&<1-5>

Extracting files and directories

About this task

Use this feature to extract files and directories from archive files.

Restrictions and guidelines

To specify the **screen** keyword for the **tar extract** command, first use the **tar list** command to identify the types of the archived files. As a best practice, specify the keyword only if all archived files are text files. Displaying the content of an archived non-text file that contains terminal control characters might result in garbled characters and even cause the terminal unable to operate correctly. To use the terminal again, you must close the current connection and log in to the device again.

Procedure

To extract files and directories, execute the following commands in user view:

1. (Optional.) Display archived files and directories.
   tar list archive-file *file*
2. Extract files and directories.
   tar extract archive-file *file* [**verbose**] [**screen** | **to** *directory*]

Compressing a file

To compress a file, execute the following command in user view:

gzip *file*

Decompressing a file

To decompress a file, execute the following command in user view:

gunzip *file*

Calculating the file digest

About this task

File digests are used to verify file integrity.

Procedure

To calculate the digest of a file, execute one of the following commands in user view:

- Use the SHA-256 algorithm.
  sha256sum *file*
- Use the MD5 algorithm.
  md5sum *file*

Executing a batch file

About this task

A batch file contains a set of commands. Executing a batch file executes the commands in the file one by one.

When executing an interactive command in a batch file, the system uses the default inputs.

Restrictions and guidelines

To execute a batch file on the device, create a batch file on a PC and load the batch file to the device.

As a best practice, try every command on the device to make sure the command line can be executed correctly before adding the command to a batch file. If a command is invalid or a condition for executing the command is not met, the command fails and the system continues to execute the next command.

Procedure

1. Enter system view.
   system-view
2. Execute a batch file.

execute *filename*

Contents

Managing configuration files 1

About configuration file management 1
Configuration types 1
Configuration file types and file selection process at startup 2
Next-startup configuration file redundancy 2
Configuration file content organization and format 2
Configuration rollback 3
FIPS compliance 3
Enabling configuration encryption 3
Assigning a user exclusive write access to the configuration 4
Disabling automatic system-wide next-startup configuration file operations 4
Saving the running configuration 5
Comparing configurations for their differences 6
Archiving the running configuration 7
Running configuration archive tasks at a glance 7
Setting configuration archive parameters 7
Archiving the running configuration 9
Rolling back configuration 10
Configuring configuration commit delay 11
Specifying a next-startup configuration file 12
Backing up and restoring the main next-startup configuration file 12
About backing up and restoring the main next-startup configuration file 12
Restrictions and guidelines for configuration backup and restoration 13
Prerequisites for configuration backup and restoration 13
Backing up the main next-startup configuration file to a TFTP server 13
Restoring the main next-startup configuration file from a TFTP server 13
Deleting a next-startup configuration file 13
Display and maintenance commands for configuration files 14

Managing configuration files

About configuration file management

You can manage configuration files from the CLI or the BootWare menu. The following information explains how to manage configuration files from the CLI.

A configuration file saves a set of commands for configuring software features on the device. You can save any configuration to a configuration file so the configuration can survive a reboot. You can also back up configuration files to a host for future use.

Configuration types

Initial configuration

Initial configuration is the collection of initial default settings for the configuration commands in software.

The device starts up with the initial configuration if you access the BootWare menu and select the **Skip Current System Configuration** option. In this situation, the device might also be described as starting up with empty configuration.

No commands are available to display the initial configuration. To view the initial default settings for the configuration commands, see the Default sections in the command references.

Factory defaults

Factory defaults are custom basic settings that came with the device. Factory defaults vary by device models and might differ from the initial default settings for the commands.

The device starts up with the factory defaults if no next-startup configuration files are available.

To display the factory defaults, use the **`display default-configuration`** command.

Startup configuration

The device uses startup configuration to configure software features during startup. After the device starts up, you can specify the configuration file to be loaded at the next startup. This configuration file is called the next-startup configuration file. The configuration file that has been loaded is called the current startup configuration file.

You can display the startup configuration by using one of the following methods:

- To display the contents of the current startup configuration file, execute the **`display current-configuration`** command before changing the configuration after the device reboots.
- To display the contents of the next-startup configuration file, use the **`display saved-configuration`** command.
- Use the **`display startup`** command to display names of the current startup configuration file and next-startup configuration files. Then, you can use the **`more`** command to display the contents of the specified startup configuration file.

Running configuration

The running configuration includes unchanged startup settings and new settings. The running configuration is stored in memory and is cleared at a device reboot or power off. To use the running configuration after a power cycling or reboot, save it to a configuration file.

To display the running configuration, use the **`display current-configuration`** command.

Configuration file types and file selection process at startup

When you save the configuration, the system saves the settings to a .cfg configuration file and to an .mdb file.

- A .cfg configuration file is a human-readable text file and its contents can be displayed by using the **more** command. Configuration files you specify for saving the configuration must use the .cfg extension.
- An .mdb file is a user-inaccessible binary file that has the same name as the .cfg file. The device loads an .mdb file faster than loading a .cfg file.

At startup, the device uses the following procedure to identify the configuration file to load:

1. The device searches for a valid .cfg next-startup configuration file. For more information about the file selection rules, see "Next-startup configuration file redundancy."
2. If a valid .cfg next-startup configuration file is found, the device searches for an .mdb file that has the same name and checksum as the .cfg file.
3. If a matching .mdb file is found, the device starts up with the .mdb file. If none is found, the device starts up with the .cfg file.

If no .cfg next-startup configuration files are available, the device starts up with the factory defaults.

Unless otherwise stated, the term "configuration file" in this document refers to a .cfg configuration file.

Next-startup configuration file redundancy

You can specify one main next-startup configuration file and one backup next-startup configuration file for redundancy.

At startup, the device tries to select the .cfg startup configuration in the following order:

1. The main next-startup configuration file.
2. The backup next-startup configuration file if the main next-startup configuration file does not exist or is corrupt.

If no next-startup configuration files are available, the device starts up with the factory defaults.

Configuration file content organization and format

IMPORTANT:

To run on the device, a configuration file must meet the content and format requirements. To ensure a successful configuration load, rollback, or restoration, use a configuration file created on the device. To edit a configuration file, you must edit, save, and obtain it on the device, and make sure all edits are compliant with the requirements. To avoid a configuration file becoming invalid because of incompliant settings, do not download the configuration file locally for editing.

Configuration restoration or rollback might fail for a set of interdependent commands if the operation fails for one of them.

A configuration file must meet the following requirements:

- All commands are saved in their complete form.
- No command lines contain invalid characters. The spaces between keywords or arguments must be in English.

IMPORTANT:

Some command lines (for example, the **sysname** command) cannot contain question marks

(?) or horizontal tabs (\t). If the system loads a configuration file that contains the **sysname abc??** command line at startup, the system will ignore the command line and use the default system name. If the system uses that configuration file to roll back or restore the configuration, the system name will not be rolled back or restored.

- Commands are sorted into sections by different command views, including system view, interface views, protocol views, and user line views.
- Two adjacent sections are separated by a pound sign (#).
- The configuration file starts with a pound sign (#).
- The configuration file ends with the word **return**.

The following is a sample configuration file excerpt:

```
#
local-user root class manage
 password hash
$h$6$Twd73mLrN8O2vvD5$Cz1vgdpR4KoTiRQNE9pg33gU14Br2p1VguczLSVyJLO2huV5Syx/LfDIf8ROLtV
ErJ/C31oq2rFtmNuyZf4STw==
 service-type ssh telnet terminal
 authorization-attribute user-role network-admin
 authorization-attribute user-role network-operator
#
interface Vlan-interface1
 ip address 192.168.1.84 255.255.255.0
#
```

Configuration rollback

Configuration rollback allows you to replace the running configuration with the configuration in a configuration file without rebooting the device. You can use this feature for the following purposes:

- Reverting to a previous configuration state.
- Adapting the running configuration to different network environments.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Enabling configuration encryption

About this task

Configuration encryption enables the device to use the master key to encrypt a startup configuration file automatically when it saves the running configuration. For more information about the master key, see public key management in *Security Configuration Guide*.

Restrictions and guidelines

After a configuration file is encrypted by using this feature on the device, only devices that use the same master key as the device can decrypt the encrypted configuration file.

You cannot use the **more** command to view the contents of an encrypted configuration file.

Procedure

1. Enter system view.
   system-view
2. Enable configuration encryption.
   configuration encrypt
   By default, configuration encryption is disabled.

Assigning a user exclusive write access to the configuration

About this task

In a multiuser environment, configuration conflicts or service exceptions might occur if multiple users are configuring the device.

To make sure only one user can change the running configuration, assign that user exclusive write access to the configuration. This setting locks the configuration. Then, all other users will have only read access to the configuration until the configuration is unlocked.

Restrictions and guidelines

To allow concurrent write accesses to the configuration, you must unlock the configuration by executing the **undo configuration exclusive by-user-name** *user-name* command. The *user-name* argument must be the username used to lock the configuration. To obtain this username, use the **display configuration exclusive by-user-name** command.

Prerequisites

To perform this task, you must log in to the system through scheme authentication. Scheme authentication requires a username and password.

Procedure

1. Enter system view.
   system-view
2. Assign a user exclusive write access to the configuration.
   configuration exclusive by-user-name *user-name*
   By default, the configuration is not locked. No users have exclusive write access to the configuration.

Disabling automatic system-wide next-startup configuration file operations

About this task

By default, automatic system-wide next-startup configuration file operations are enabled. The system performs the following operations on all IRF subordinate devices in addition to the master device:

- Saves the running configuration to the next-startup configuration file on each member device when you execute the **save** [**safely**] [**backup** | **main**] [**force**] [**changed**] command.
- Deletes the next-startup configuration file on each member device when you execute the **reset saved-configuration** command.

If you disable automatic system-wide next-startup configuration file operations, the system saves the running configuration or deletes the next-startup configuration file only on the master device.

Automatic system-wide operations ensure start-up configuration file consistency between the master and subordinate devices. However, a system-wide operation takes more time than an operation performed only on the master device. In addition, the amount of time required to complete a system-wide configuration operation increases as the amount of configuration data grows.

If you are disabling automatic system-wide operations for faster configuration saving, be aware that the next-startup configuration files will be inconsistent between the master device and the subordinate devices.

Procedure

1. Enter system view.

   `system-view`
2. Disable automatic system-wide next-startup configuration file operations.

   `undo standby auto-update config`

   By default, next-startup configuration file operations are automatically synchronized across the entire system.

Saving the running configuration

About this task

The running configuration is stored only in memory. If you have not saved the running configuration before a reboot, any changes you have made to the configuration will be lost and the device will restart with the factory defaults. To keep the current running configuration after a reboot, save it to a configuration file, or to the main or backup next-startup configuration file.

Restrictions and guidelines

To prevent the loss of next-startup configuration, make sure no one reboots or power cycles the device while the device is saving the running configuration.

If there are settings that must be read from the .cfg configuration file at startup, using the binary-only saving mode will cause the loss of those settings at startup. As a best practice, use this mode only when you must save a large number of table entries such as NAT entries, IP static routes, or VPN instances.

When an IRF member device splits from the IRF fabric, its settings are retained in memory but removed from the running configuration on the IRF fabric. Saving the running configuration before the IRF fabric recovers will remove the member device's settings from the next-startup configuration file.

If you have saved the running configuration before the member device rejoins the IRF fabric, perform the following steps to restore the member device settings to the next-startup configuration file:

1. Resolve the split issue.
2. Reboot the member device to rejoin the IRF fabric.
3. After the member device rejoins the IRF fabric, execute the **`display current-configuration`** command to verify that the member device's settings have been restored from memory to the running configuration.
4. Save the running configuration to the next-startup configuration file on the IRF fabric.

IMPORTANT:

To ensure a successful configuration restoration, make sure the IRF fabric has not rebooted after the member device left.

By default, the **save** [**safely**] [**backup** | **main**] [**force**] [**changed**] command saves the configuration to all IRF member devices. To save the configuration only to the master device, disable automatic system-wide next-startup configuration file operations. For more information, see "Disabling automatic system-wide next-startup configuration file operations."

Automatic configuration saving saves the running configuration to all member devices.

Automatic configuration saving saves the running configuration to both active and standby MPUs.

Automatic configuration saving saves the running configuration to each MPU on each member device.

Procedure

To save the running configuration, perform one of the following tasks in any view:

- Save the running configuration to a configuration file without specifying the file as a next-startup configuration file.

  save *file-url* [**all** | **slot** *slot-number*]

- Save the running configuration to a configuration file in the root directory of the storage medium on each member device, and specify the file as a next-startup configuration file.

  save [**safely**] [**backup** | **main**] [**force**] [**changed**]

  As a best practice, specify the **safely** keyword for reliable configuration saving.

CAUTION:

Use the **save** command with caution. This command will overwrite the settings in the target configuration file. Carefully read the messages you receive when you use this command and make sure you fully understand the impact of this command when you execute it.

Comparing configurations for their differences

About this task

You can compare configuration files or compare a configuration file with the running configuration for their differences.

If you specify the next-startup configuration for a comparison, the system selects the next-startup configuration file to be compared with in the following order:

1. The main next-startup configuration file.
2. The backup next-startup configuration file if the main next-startup configuration file is unavailable.

If both configuration files are unavailable, the system displays a message indicating that no next-startup configuration files exist.

Procedure

To compare configurations for their differences, perform one of the following tasks in any view:

- Display the differences that a configuration file, the running configuration, or the next-startup configuration has as compared with the specified source configuration file.

  display diff configfile *file-name-s* { **configfile** *file-name-d* | **current-configuration** | **startup-configuration** }

- Display the differences that a configuration file or next-startup configuration has as compared with the running configuration.

  display diff current-configuration { **configfile** *file-name-d* | **startup-configuration** }

- Display the differences that a configuration file has as compared with the next-startup configuration.

  display diff startup-configuration configfile *file-name-d*
- Display the differences that the running configuration has as compared with the next-startup configuration.
  - Method 1:

    display diff startup-configuration current-configuration
  - Method 2:

    display current-configuration diff

Archiving the running configuration

Running configuration archive tasks at a glance

To configure running configuration archive, perform the following tasks:

1. Setting configuration archive parameters
2. Archiving the running configuration

   Perform either of the following tasks:
   - Enabling automatic configuration archiving
   - Manually archiving the running configuration

Setting configuration archive parameters

About this task

Before archiving the running configuration, you must set a file directory and file name prefix for configuration archives.

The archive directory can be located on the local device or on a remote SCP server.

If you use local archiving, configuration archives are named in the format of *prefix_serial number***.cfg**, for example, **archive_1.cfg** and **archive_2.cfg**. The serial number is automatically assigned from 1 to 1000, increasing by 1. After the serial number reaches 1000, it restarts from 1.

If you change the file directory or file name prefix on the local device, the following events occur:

- The old configuration archives change to common configuration files.
- The configuration archive counter is reset. The serial number for new configuration archives starts at 1.
- The **display archive configuration** command no longer displays the old configuration archives.

The configuration archive counter does not restart when you delete configuration archives from the archive directory. However, if the device reboots after all configuration archives have been deleted, the configuration archive counter restarts. The serial number for new configuration archives starts at 1.

If you archive the running configuration to a remote SCP server, configuration archives are named in the format of *prefix_YYYYMMDD_HHMMSS***.cfg** (for example, **archive_20170526_203430.cfg**). The *YYYYMMDD_HHMMSS* string represents the device system date and time when the configuration archive was saved.

If you change the file directory or file name prefix on the remote SCP server, the **display archive configuration** command no longer displays the old configuration archives saved before the change.

Restrictions and guidelines for setting configuration archive parameters

Local archiving (the **archive configuration location** command) and remote archiving (the **archive configuration server** command) are mutually exclusive. You cannot use the two features at the same time.

IMPORTANT:

In FIPS mode, the device does not support archiving the running configuration to a remote SCP server.

With local configuration archiving, the system deletes the oldest archive to make room for the new archive after the maximum number of configuration archives is reached.

The following information applies to remote configuration archiving:

- The maximum number of configuration archives on a remote SCP server depends on the SCP server setting and is not restricted by the **archive configuration max** command.
- If the remote SCP server already contains an archive file with the same name as the new configuration archive file being transferred, the new archive file overwrites the existing file. To avoid an undesirable overwrite, back up the configuration archives on the SCP server to a different directory.

The **undo archive configuration location** command removes the local configuration archive directory and file name prefix settings, but it does not delete the configuration archives on the device. The command also performs the following operations:

- Disables both the manual and automatic configuration archiving features.
- Restores the default settings for the **archive configuration interval** and **archive configuration max** commands.
- Clears the configuration archive information displayed by using the **display archive configuration** command.

The **undo archive configuration server** command removes the remote configuration archive directory and file name prefix settings, but it does not delete the configuration archives on the server. The command also performs the following operations:

- Disables both the manual and automatic configuration archiving features.
- Restores the default setting for the **archive configuration interval** command.
- Clears the configuration archive information displayed by using the **display archive configuration** command.

Configuring local archiving parameters

1. Enter system view.

   system-view

2. Set the directory and file name prefix for archiving the running configuration to the local device.

   archive configuration location *directory* **filename-prefix** *filename-prefix*

   By default, no path or file name prefix is set for configuration archives on the device, and the system does not regularly save configuration.

   In an IRF fabric, the configuration archive directory must already exist on the master device and cannot include a member ID.

3. (Optional.) Set the maximum number of configuration archives.

   archive configuration max *file-number*

   The default number is 5.

   Change the setting depending on the amount of storage available on the device.

Configuring remote archiving parameters

1. Enter system view.

   system-view

2. Set the directory and file name prefix for archiving the running configuration on a remote SCP server.

   archive configuration server scp { *ipv4-address* | **ipv6** *ipv6-address* } [**port** *port-number*] [**vpn-instance** *vpn-instance-name*] [**directory** *directory*] [**source-ip** *source-ip-address*] **filename-prefix** *filename-prefix*

   By default, no path or file name prefix is set for archiving the running configuration to a remote SCP server.

3. Configure the username and password for accessing the SCP server.

   a. Configure the username.

      archive configuration server user *user-name*

      By default, no username is configured for accessing the SCP server.

   b. Configure the password.

      archive configuration server password { **cipher** | **simple** } *string*

      By default, no password is configured for accessing the SCP server.

   Make sure the username and password are the same as the SCP account settings on the SCP server.

Archiving the running configuration

About this task

The following are methods for archiving the running configuration:

- **Automatic configuration archiving**—The system automatically archives the running configuration at intervals as configured.
- **Manual configuration archiving**—You can disable automatic configuration archiving and manually archive the running configuration if the configuration will not be changed very often. You can also use this method before performing complicated configuration tasks. If the configuration attempt fails, you can use the archive for configuration recovery.

Restrictions and guidelines for archiving the running configuration

If you use local archiving, the configuration archive feature saves the running configuration only on the master device. To make sure the system can archive the running configuration after a master/subordinate switchover, create the configuration archive directory on all IRF members.

If a local archiving has started based on the existing archive parameters when an archive parameter is changed, the archive will still be retained in the old directory. However, the **display archive configuration** command will not display the record about this archiving.

When you modify parameters (for example, the username or password) for remote archiving, make sure the changes are consistent between the device and the server. A manual or automatic remote archiving will fail if it has started before you change the device and the server settings to be consistent.

Enabling automatic configuration archiving

1. Enter system view.

   system-view

2. Enable automatic configuration archiving and set the archiving interval.

   archive configuration interval *interval*

By default, automatic configuration archiving is disabled.

Manually archiving the running configuration

Manually archive the running configuration in user view.

archive configuration

Rolling back configuration

About this task

When you roll back the running configuration or commit target configuration command lines, the system creates a rollback point and assigns a unique commit ID to the rollback point.

The configuration rollback feature compares the running configuration against the specified rollback point or replacement configuration file and handles configuration differences as follows:

- If a command in the running configuration is not in the rollback point or replacement file, the rollback feature executes the **undo** form of the command.
- If a command in the rollback point or replacement file is not in the running configuration, the rollback feature adds the command to the running configuration.
- If a command has different settings in the running configuration and the rollback point or replacement file, the rollback feature replaces the running command setting with the setting in the rollback point or replacement file.

Restrictions and guidelines

To ensure a successful rollback, do not perform a master/subordinate switchover while the system is rolling back the configuration.

The configuration rollback feature might fail to reconfigure some commands in the running configuration for one of the following reasons:

- A command cannot be undone because prefixing the **undo** keyword to the command does not result in a valid **undo** command. For example, if the **undo** form designed for the **A** [**B**] **C** command is **undo A C**, the configuration rollback feature cannot undo the **A B C** command. This is because the system does not recognize the **undo A B C** command.
- A command (for example, a hardware-dependent command) cannot be deleted, overwritten, or undone due to system restrictions.
- The commands in different views are dependent on each other.
- Commands or command settings that the device does not support cannot be added to the running configuration.

Make sure the replacement configuration file is created by using the configuration archive feature or the **save** command on the local device. If the configuration file is not created on the local device, make sure the command lines in the configuration file are fully compatible with the local device.

If the replacement configuration file is encrypted, make sure the device can decrypt it.

Rolling back the running configuration based on a rollback point

1. Enter system view.

   system-view

2. Roll back the running configuration based on the specified rollback point.

   configuration replace { **last** *number* | **commit-id** *commit-id* | **label** *labelname* }

3. (Optional) Set the maximum number of rollback points.

   configuration commit max-number *max-number*

By default, the maximum number of rollback points is 100.

Rolling back the running configuration based on a local configuration file

1. Enter system view.
   system-view
2. Roll the running configuration back to the configuration contained in a local configuration file.
   configuration replace file *filename*

> **CAUTION:**
> This operation will cause settings not in the replacement configuration file to be lost, which might cause service interruption. When you perform configuration rollback, make sure you fully understand its impact on your network.

Configuring configuration commit delay

About this task

This feature enables the system to automatically remove the settings you made during a configuration commit delay interval if you have not manually committed them.

You specify the configuration commit delay interval by using the configuration commit delay timer. The system creates a rollback point to record the configuration status when you start the configuration commit delay timer. The settings you made during the configuration commit delay interval takes effect immediately. However, these settings will be removed automatically if you have not manually committed them before the configuration commit delay timer expires. Then, the system returns to the configuration status when the configuration commit delay timer started.

This feature prevents a misconfiguration from causing the inability to access the device and is especially useful when you configure the device remotely.

Restrictions and guidelines

When you use this feature, follow these restrictions and guidelines:

- In a multi-user context, make sure no one else is configuring the device.
- To avoid unexpected errors, do not perform any operations during the configuration rollback.
- You can reconfigure the configuration commit delay timer before it expires to shorten or extend the commit delay interval. However, the rollback point will not change.
- The configuration commit delay feature is a one-time setting. The feature is disabled with the rollback point removed when the commit delay timer expires or after a manual commit operation is performed. To use this feature again, you must restart the timer.

Procedure

1. Enter system view.
   system-view
2. Enable the configuration commit delay feature and start the commit delay timer.
   configuration commit delay *delay-time*
3. (Optional.) Commit the settings configured after the commit delay timer started.
   configuration commit

Specifying a next-startup configuration file

Restrictions and guidelines

CAUTION:
Using the **undo startup saved-configuration** command can cause an IRF split after the IRF fabric or an IRF member reboots. When you execute this command, make sure you understand its impact on your network.

As a best practice, specify different files as the main and backup next-startup configuration files.

The **undo startup saved-configuration** command changes the attribute of the main or backup next-startup configuration file to NULL instead of deleting the file.

If you use the **save** command without any parameters, the device saves the running configuration to a configuration file and automatically specifies the file as the main next-startup configuration file. For more information, see "Saving the running configuration."

Prerequisites

In an IRF fabric, make sure the specified configuration file is valid and has been saved to the root directory of a storage medium on each member device. In addition, make sure the storage media are the same type across all IRF member devices.

Procedure

1. Specify a next-startup configuration file. Choose one of the following methods:
   - Execute the following command in user view to specify a next-startup configuration file:
     startup saved-configuration *cfgfile* [**backup** | **main**]
     By default, no next-startup configuration files are specified.
   - Execute the following command in any view to save the running configuration to a file and specify the file as a next-startup configuration file:
     save [**safely**] [**backup** | **main**] [**force**]
     For more information about this command, see "Saving the running configuration."

   If you do not specify the **backup** or **main** keyword, this command specifies the configuration file as the main next-startup configuration file.
2. (Optional.) Verify the configuration. Use one of the following commands in any view:
   - Display the names of the configuration files for this startup and the next startup.
     display startup
   - Display the contents of the configuration file for the next system startup.
     display saved-configuration

Backing up and restoring the main next-startup configuration file

About backing up and restoring the main next-startup configuration file

You can back up the main next-startup configuration file to a TFTP server or restore the main next-startup configuration file from a TFTP server.

Restrictions and guidelines for configuration backup and restoration

Configuration backup and restoration are not supported in FIPS mode.

Prerequisites for configuration backup and restoration

Before you back up or restore the main next-startup configuration file, perform the following tasks:

- Make sure the following requirements are met:
  - The server is reachable.
  - The server is enabled with TFTP service.
  - You have read and write permissions to the server.
- For configuration backup, use the **display startup** command to verify that the main next-startup configuration file has been specified in user view. If no next-startup configuration file has been specified or the specified configuration file does not exist, the backup operation will fail.

Backing up the main next-startup configuration file to a TFTP server

To back up the main next-startup configuration file to a TFTP server, execute the following command in user view:

backup startup-configuration to { *ipv4-server* | **ipv6** *ipv6-server* } [*dest-filename*] [**vpn-instance** *vpn-instance-name*]

Restoring the main next-startup configuration file from a TFTP server

1. Restore the main next-startup configuration file from a TFTP server in user view.

   restore startup-configuration from { *ipv4-server* | **ipv6** *ipv6-server* } *src-filename* [**vpn-instance** *vpn-instance-name*]
2. (Optional.) Verify that the specified configuration file has been set as the main next-startup configuration file. Use one of the following commands in any view:
   - Display the names of the configuration files for this startup and the next startup.

     display startup
   - Display the contents of the configuration file for the next system startup.

     display saved-configuration

Deleting a next-startup configuration file

About this task

You can perform this task to delete a next-startup configuration file.

If both the main and backup next-startup configuration files are deleted, the device uses the factory defaults at the next startup.

To delete a file that is set as both main and backup next-startup configuration files, you must execute both the **reset saved-configuration backup** command and the **reset saved-configuration main** command. Using only one of the commands removes the specified file attribute instead of deleting the file.

For example, if the **reset saved-configuration backup** command is executed, the backup next-startup configuration file setting is set to NULL. However, the file is still used as the main file. To delete the file, you must also execute the **reset saved-configuration main** command.

Procedure

To delete a next-startup configuration file, execute the following command in user view:

reset saved-configuration [**backup** | **main**]

CAUTION:

By default, this task permanently deletes a next-startup configuration file from all IRF member devices. To delete the configuration file only from the master device, disable automatic system-wide next-startup configuration file operations. For more information, see "Disabling automatic system-wide next-startup configuration file operations."

If you do not specify the **backup** or **main** keyword, the main next-startup configuration is deleted.

Display and maintenance commands for configuration files

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display configuration archive information.	**display archive configuration**
Display configuration differences between a rollback point and the running configuration or between two rollback points.	**display configuration commit diff** [**at** *commit-id* | **last** *number* | **since** *commit-id*]
Display rollback point information.	**display configuration commit list** [*number* | **label**] [**verbose**]
Display device configuration sessions.	**display configuration sessions**
Display the running configuration.	**display current-configuration** [[**configuration** [*module-name*] | **exclude-provision** | **interface** [*interface-type* [*interface-number*]]] [**all**] | **slot** *slot-number*] [**by-section** { **begin** | **exclude** | **include** } *regular-expression*]
Display the differences that the running configuration has as compared with the next-startup configuration.	**display current-configuration diff**
Display the factory defaults.	**display default-configuration**
Display the differences between configurations.	• **display diff configfile** *file-name-s* { **configfile** *file-name-d* |

Task	Command
	current-configuration | **startup-configuration** } • **display diff current-configuration** { **configfile** *file-name-d* | **startup-configuration** } • **display diff startup-configuration** { **configfile** *file-name-d* | **current-configuration** }
Display the contents of the configuration file for the next system startup.	**display saved-configuration** [**by-section** { **begin** | **exclude** | **include** } *regular-expression*]
Display the names of the configuration files for this startup and the next startup.	**display startup**
Display the valid configuration in the current view.	**display this** [**all**] [**by-section** { **begin** | **exclude** | **include** } *regular-expression*]
Display information about user-based configuration locking.	**display configuration exclusive by-user-name**
Delete next-startup configuration files.	**reset saved-configuration** [**backup** | **main**]
Clear rollback points.	**reset configuration commit** *number*

Contents

Multichassis configuration sync 1

About multichassis configuration sync 1
Real-time configuration sync 1
Bulk configuration sync 1
Configuration backup 2
Multichassis configuration sync tasks at a glance 2
Restrictions and guidelines for multichassis configuration sync 2
Synchronizing the configurations in real time 2
Bulk synchronizing the configurations 3
Fast configuring differential settings 3
Backing up and restoring a configuration file 4
Display and maintenance commands for multichassis configuration sync 4

Multichassis configuration sync

About multichassis configuration sync

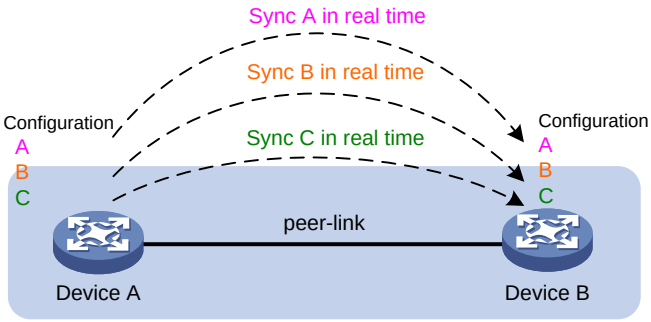
Multichassis configuration synchronization (sync) allows you to synchronize and back up configurations between multiple devices to simplify the configuration procedure, improve deployment efficiency, and optimize user experience. For example, after you configure one M-LAG member device, you can synchronize its configuration to the other M-LAG member device over the peer link for them to have identical configuration.

This feature is applicable only to the M-LAG system currently.

Real-time configuration sync

As shown in Figure 1, settings in multichassis configuration sync view of a member device are synchronized to its peer device in real time.

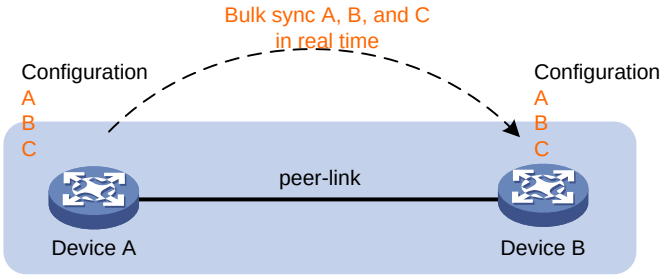
Figure 1 Real-time configuration sync



Bulk configuration sync

To fast deploy the services that are required to be consistent across the M-LAG member devices, configure the services on an M-LAG member device before M-LAG system setup. After the M-LAG system is set up and the peer link comes up, you can bulk synchronize settings on the member device to the same view for multichassis configuration sync on the M-LAG peer, as shown in Figure 2.

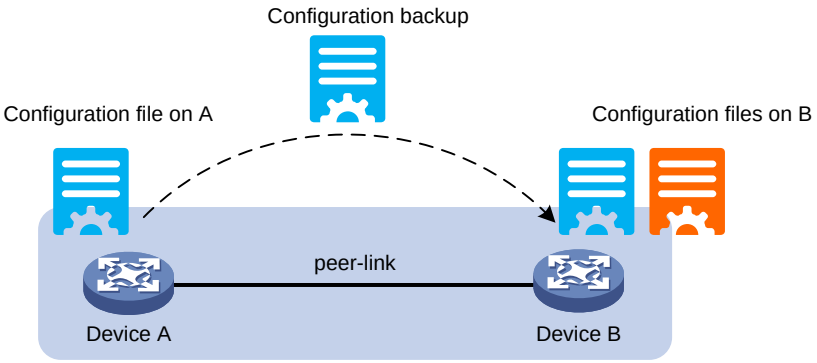
Figure 2 Bulk configuration sync



Configuration backup

If you are to update an M-LAG member device, for example, update the software version or replace the device, back up its configuration file to the M-LAG peer in advance as shown in Figure 3. After the update is completed, you can retrieve the previously saved configuration file from the peer device.

Figure 3 Configuration backup



Multichassis configuration sync tasks at a glance

To configure multichassis configuration sync, perform the following tasks:

- Synchronizing the configurations in real time
- Bulk synchronizing the configurations
- Fast configuring differential settings
- Backing up and restoring a configuration file

Restrictions and guidelines for multichassis configuration sync

Real-time or bulk configuration sync synchronizes only configurations required to be consistent across the M-LAG member devices.

Execute the **`mc-configure access-local enable`** command on both M-LAG member devices for them to access each another.

Before you use the multichassis configuration sync feature, verify that the M-LAG system is operating correctly and the peer link is up.

Synchronizing the configurations in real time

About this task

This feature enables commands executed in multichassis configuration sync view on an M-LAG member device to be synchronized to its M-LAG peer in real time.

Procedure

1. Enter system view.

   `system-view`

2. Permit the M-LAG peer to access the local device via the sync channel.

   mc-configure access-local enable

   By default, the M-LAG peer is permitted to access the local device via the sync channel.
3. Return to user view.

   quit
4. Enter multichassis configuration sync view.

   system-view mc-config

   Commands executed in any view for multichassis configuration sync will be synchronized to the M-LAG peer in real time.

Bulk synchronizing the configurations

About this task

To fast deploy the services that are required to be consistent across the M-LAG member devices, configure the services on an M-LAG member device before M-LAG system setup. After the M-LAG system is set up and the peer link comes up, you can use this feature to bulk synchronize settings on the member device to the same view for multichassis configuration sync on the M-LAG peer.

Procedure

1. Enter system view.

   system-view
2. Permit the M-LAG peer to access the local device via the sync channel.

   mc-configure access-local enable

   By default, the M-LAG peer is permitted to access the local device via the sync channel.
3. Pre-configure settings in system view and any view under system view as needed.
4. Return to user view.

   quit
5. Enter multichassis configuration sync view.

   system-view mc-config
6. Bulk synchronize the settings to the M-LAG peer.

   mc-configure batch-synchronize

   You can execute this command in any view for multichassis configuration sync.

   For example, if you execute this command in +[Sysname] view, the device will bulk synchronize the configurations in [Sysname] view and its sub-views to the peer device. If you execute this command in [Sysname1-GigabitEthernet1/0/1] view, the device will bulk synchronize the configurations in [Sysname1-GigabitEthernet1/0/1] view and its sub-views to the peer device.

Fast configuring differential settings

About this task

For settings required only on the local device, you can execute the **mc-configure switch-to-local** command in any view for multichassis configuration sync to switch to the same view of the local device and then configure the settings. To return to the original view for multichassis configuration sync after configuring the settings on the local device, execute the **mc-configure quit** command or press **Ctrl+Q**.

For the settings required only on the peer device, you can execute the **mc-configure switch-to-peer** in any view for multichassis configuration sync of the local device to switch to

the same view of the peer device and then configure the settings. To return to the original view for multichassis configuration sync after configuring the settings on the peer device, execute the **mc-configure quit** command or press **Ctrl+Q**.

Configuring the differential settings on the local device

1. Enter multichassis configuration sync view.
   system-view mc-config
2. Switch to the same view of the local device.
   [@local] will be added to the commands executed in any view of the local device.
   mc-configure switch-to-local
   You can execute this command in any view for multichassis configuration sync.
3. Configure the differential commands on the local device.
4. Return to the original view for multichassis configuration sync.
   mc-configure quit

Configuring the differential settings on the M-LAG peer

1. Enter multichassis configuration sync view.
   system-view mc-config
2. Switch to the same view of the M-LAG peer.
   [@local] will be added to the commands executed in any view of the peer device.
   mc-configure switch-to-local
   You can execute this command in any view for multichassis configuration sync.
3. Configure the differential commands on the M-LAG peer.
4. Return to the original view for multichassis configuration sync.
   mc-configure quit

Backing up and restoring a configuration file

About this task

To update an M-LAG member device, for example, update the software version or replace the device, you can the **mc-configure backup-to-peer** command to back up its configuration file to the M-LAG peer in advance. After the update is completed, you can use the **mc-configure get-from-peer** command to retrieve the previously saved configuration file from the peer device and replace the configuration file on the local device.

Procedure

Execute the following command in user view to save the local configuration file to the M-LAG peer.

mc-configure backup-to-peer *local-filename*

Execute the following command in user view to retrieve the configuration file from the M-LAG peer.

mc-configure get-from-peer *filename*

Display and maintenance commands for multichassis configuration sync

Execute **display** commands in any view.

Task	Command
Display the configuration check result for multichassis configuration sync.	**display mc-configure config-check m-lag** [{ **evpn-border** | **evpn-leaf** } [**single-homed**]]
Display the configuration comparison result for multichassis configuration sync.	**display mc-configure config-compare** [**different** | [**interface** | **verbose**] [**by-table** [**width** *width-length*]]]
Display settings not synchronized during multichassis configuration sync.	**display mc-configure unsync-info**

Contents

Upgrading software 1

About software upgrade 1
Software types 1
Software release forms 2
Upgrade methods 2
Device software upgrade 2
Application scenarios 3
Digitally signed software images 3
Restrictions and guidelines: Software upgrade 4
Upgrading device software by using the boot loader method 4
Software upgrade tasks at a glance 4
Prerequisites 4
Preloading the BootWare image to BootWare 5
Specifying startup images and completing the upgrade 5
Synchronizing startup images from the master device to subordinate devices 6
Display and maintenance commands for software image settings 6
Software upgrade examples 6
Example: Upgrading device software 6

Upgrading software

About software upgrade

Software upgrade enables you to upgrade a software version, add new features, and fix software bugs. This chapter describes software types and release forms, compares software upgrade methods, and provides the procedures for upgrading software from the CLI.

Software types

The following software types are available:

- **BootWare image**—Also called the Boot ROM image. This image contains a basic segment and an extended segment.
  - The basic segment is the minimum code that bootstraps the system.
  - The extended segment enables hardware initialization and provides system management menus. When the device cannot start up correctly, you can use the menus to load software and the startup configuration file or manage files.

  Typically, the BootWare image is integrated into the Boot image to avoid software compatibility errors.
- **Comware image**—Includes the following image subcategories:
  - **Boot image**—A .bin file that contains the Linux operating system kernel. It provides process management, memory management, file system management, and the emergency shell.
  - **System image**—A .bin file that contains the Comware kernel and standard features, including device management, interface management, configuration management, and routing.
  - **Feature image**—A .bin file that contains advanced or customized software features. You can purchase feature images as needed.
  - **Patch image**—A .bin file that is released for fixing bugs without rebooting the device. A patch image does not add or remove features.

    Patch images have the following types:

    - **Incremental patch images**—A new patch image can cover all, part, or none of the functions provided by an old patch image. A new patch image can coexist with an old patch image on the device only when the former covers none of the functions provided by the latter.
    - **Non-incremental patch images**—A new non-incremental patch image covers all functions provided by an old non-incremental patch image. Each boot, system, or feature image can have one non-incremental patch image, and these patch images can coexist on the device. The device uninstalls the old non-incremental patch image before installing a new non-incremental patch image.

    An incremental patch image and a non-incremental patch image can coexist on the device.

  Comware images that have been loaded are called current software images. Comware images specified to load at the next startup are called startup software images.

BootWare image, boot image, and system image are required for the device to operate.

You can install up to 32 .bin files on the device, including one boot image file, one system image file, and up to 30 feature and patch image files.

Software release forms

Software images are released in one of the following forms:

- Separate .bin files. You must verify compatibility between software images.
- As a whole in one .ipe package file. The images in an .ipe package file are compatible. The system decompresses the file automatically, loads the .bin images and sets them as startup software images.

NOTE:

Software image file names use the *model-comware version-im*age *type-release* format. This document uses **boot.bin** and **system.bin** as boot and system image file names.

Upgrade methods

Upgrade method	Software types	Remarks
Upgrading from the CLI by using the boot loader method	• BootWare image • Comware images (excluding incremental patches)	This method is disruptive. You must reboot the entire device to complete the upgrade.
Performing an ISSU from the CLI	Comware images	This method enables a software upgrade with a minimum amount of downtime. Use this method if possible. For more information about ISSU, see "Performing an ISSU."
Upgrading from the BootWare menu	• BootWare image • Comware images	Use this method when the device cannot start up correctly. To use this method, first connect to the console port and power cycle the device. Then, press **Ctrl+B** at prompt to access the BootWare menu. For more information about upgrading software from the BootWare menu, see the release notes for the software version. **IMPORTANT:** Use this method only when you do not have any other choice.

This chapter covers only upgrading software from the CLI by using the boot loader method.

Device software upgrade

Startup software images

To upgrade software, you must specify the upgrade files as the startup software images for the device to load at next startup. You can specify two lists of software images: one main and one backup. The device first loads the main startup software images. If the main startup software images are not available, the devices loads the backup startup software images.

Image loading process at startup

At startup, the device performs the following operations after loading and initializing BootWare:

1. Loads main images.
2. If any main image does not exist or is invalid, loads the backup images.
3. If any backup image does not exist or is invalid, checks the main or backup boot image.
4. If the main or backup boot image exist and is valid, enters the emergency shell.

If both the main and backup boot images do not exist or are invalid, the device cannot start up.

Handling the image loading failure

After the device enters the emergency shell, connect to the console port and load a system image so you can access the Comware system. For more information about using the emergency shell, see "Using the emergency shell."

If the device fails to load the images, connect to the console port, power cycle the device, and use the BootWare menu to load new startup images.

Application scenarios

Software upgrade

When the current software cannot meet the requirements of network operation or user services, you can upgrade the software to optimize device performance to meet the requirements.

Feature and patch package installation

To fix software defects without upgrading the software version, you can install patch packages compatible with the software version. To expand service features on the device, you can install the corresponding feature packages.

Digitally signed software images

The software images for the device are digitally signed for authenticity and integrity verification. This mechanism ensures that the software installed on the system is from a trusted source and has not been tampered with in the transfer, storage, or installation phase.

The system performs software digital signature verification for authenticity and integrity in the following situations:

- Before the system loads a software image during startup. If the digital signature verification fails, the system will not load the image and you will receive a digital signature verification failure message.
- When you specify a software image to upgrade the device from the BootWare menu. If the digital signature verification fails, the system will not set the image for upgrade and you will receive a digital signature verification failure message.
- Before the system loads a BootWare image to the Normal area of BootWare. If the digital signature verification fails, the system will not load the image and you will receive a digital signature verification failure message.
- When you specify a software image as a startup image through the boot loader. The system will verify the digital signature of the image before it updates the startup image list with the specified image. If the digital signature verification fails, the system will not update the startup image list and you will receive a digital signature verification failure message.
- When you specify startup images for an ISSU. The system verifies the digital signature of a software image for authenticity and integrity before it sets and loads that image as a main startup image. If the digital signature verification fails, the system will not set or load the image as a main startup image and you will receive a digital signature verification failure message. For more information about ISSU, see "Performing an ISSU."

- Before the system performs an ISSU. If the digital signature verification fails, the system will not perform the ISSU and you will receive a digital signature verification failure message. For more information about ISSU, see "Performing an ISSU."
- Before the system activates a feature or patch image. If the digital signature verification fails, the system will not activate the image and you will receive a digital signature verification failure message.

Restrictions and guidelines: Software upgrade

Do not execute the **install**, **issu**, or **boot-loader** command before the current upgrade action completes.

As a best practice, store the startup images in a fixed storage medium. If you store the startup images in a hot swappable storage medium, do not remove the hot swappable storage medium during the startup process.

Upgrading device software by using the boot loader method

Software upgrade tasks at a glance

To upgrade software, perform one of the following tasks:

1.
2. Upgrade the IRF fabric:
   a. (Optional.) Preloading the BootWare image to BootWare

      If a BootWare upgrade is required, you can perform this task to shorten the subsequent upgrade time. This task helps reduce upgrade problems caused by unexpected power failure. If you skip this task, the device upgrades the BootWare automatically when it upgrades the startup software images.

   b. Specifying startup images and completing the upgrade
3. (Optional.) Synchronizing startup images from the master device to subordinate devices

   Perform this task when the startup images on subordinate devices are not the same version as those on the master device.

Prerequisites

1. Use the **display version** command to verify the current BootWare image version and startup software version.
2. Use the release notes for the upgrade software version to evaluate the upgrade impact on your network and verify the following items:
   - Software and hardware compatibility.
   - Version and size of the upgrade software.
   - Compatibility of the upgrade software with the current BootWare image and startup software image.
3. Use the **dir** command to verify that all IRF member devices have sufficient storage space for the upgrade images. If the storage space is not sufficient on any member device, delete unused files by using the **delete /unreserved** command. For more information, see "Managing file systems."

4. Use FTP or TFTP to transfer the upgrade image file to the root directory of a file system. For more information about FTP and TFTP, see "Configuring FTP" or "Configuring TFTP." For more information about file systems, see "Managing file systems."

Preloading the BootWare image to BootWare

Perform the following steps in user view:

5.

6. Load the upgrade BootWare image to the Normal area of BootWare.

   bootrom update file *file* **slot** *slot-number-list*

   Specify the downloaded software image file for the *file* argument.

   The new BootWare image takes effect at a reboot.

NOTE:

The system will verify the digital signature of a BootWare image before it loads it to the Normal area of BootWare. If the digital signature verification fails, the system will not load the image and you will receive a digital signature verification failure message.

Specifying startup images and completing the upgrade

Perform the following steps in user view:

1. Specify main or backup startup images for all member devices.
   - Use an .ipe file:

     boot-loader file *ipe-filename* [**patch** *filename*&<1-16>] { **all** | **slot** *slot-number* } { **backup** | **main** }
   - Use .bin files:

     boot-loader file boot *filename* **system** *filename* [**feature** *filename*&<1-30>] [**patch** *filename*&<1-16>] { **all** | **slot** *slot-number* } { **backup** | **main** }

   As a best practice in a multichassis IRF fabric, specify the **all** keyword for the command. If you use the **slot** *slot-number* option to upgrade member devices one by one, version inconsistencies occur among the member devices during the upgrade.

NOTE:

The system will verify the digital signature of the specified images before it updates the startup image list with the specified images. If the digital signature verification fails, the system will not update the startup image list and you will receive a digital signature verification failure message.

2. Save the running configuration.

   save

   This step ensures that any configuration you have made can survive a reboot.
3. Reboot the IRF fabric.

   reboot
4. (Optional.) Verify the software image settings.

   display boot-loader [**slot** *slot-number*]

Verify that the current software images are the same as the startup software images.

Synchronizing startup images from the master device to subordinate devices

About this task

Perform this task when the startup images on subordinate devices are not the same version as those on the master device.

This task synchronizes startup images that are running on the master device to subordinate devices. If any of the startup images does not exist or is invalid, the synchronization fails.

The startup images synchronized to subordinate devices are set as main startup images, regardless of whether the source startup images are main or backup.

Restrictions and guidelines

If an ISSU or patch installation has been performed on the master device, use the **install commit** command to update the set of main startup images on the master device before software synchronization. This command ensures startup image consistency between the master and subordinate devices.

Procedure

Perform the following steps in user view:

1. Synchronize startup images from the master to subordinate devices.
   boot-loader update { **all** | **slot** *slot-number* }
2. Reboot the subordinate devices.
   reboot slot *slot-number* [**force**]

Display and maintenance commands for software image settings

Execute **display** commands in any view.

Task	Command
Display current software images and startup software images.	**display boot-loader** [**slot** *slot-number*]

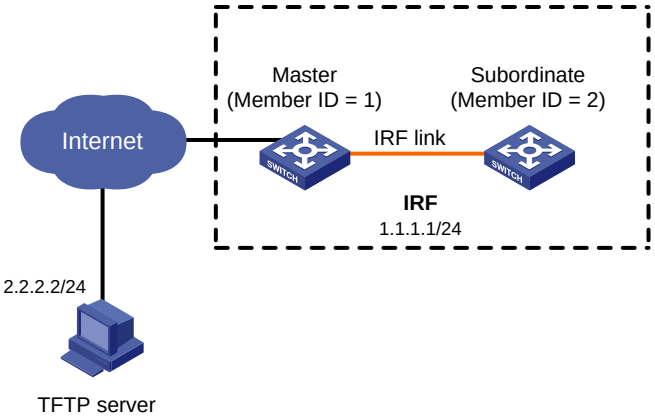
Software upgrade examples

Example: Upgrading device software

Network configuration

As shown in Figure 1, use the file **startup-a2105.ipe** to upgrade software images for the IRF fabric.

Figure 1 Network diagram



Procedure

Configure IP addresses and routes. Make sure the device and the TFTP server can reach each other.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] ip address 1.1.1.1 24
[Sysname-GigabitEthernet1/0/1]quit
[Sysname] ip route-static 2.2.2.0 24 1.1.1.2
[Sysname] quit
```

Configure TFTP settings on both the device and the TFTP server. (Details not shown.)

Display information about the current software images.

```
<Sysname> display version
```

Back up the current software images.

```
<Sysname> copy boot.bin boot_backup.bin
<Sysname> copy system.bin system_backup.bin
```

Specify **boot_backup.bin** and **system_backup.bin** as the backup startup image files for all IRF member devices.

```
<Sysname> boot-loader file boot flash:/boot_backup.bin system flash:/system_backup.bin
slot 1 backup
<Sysname> boot-loader file boot flash:/boot_backup.bin system flash:/system_backup.bin
slot 2 backup
```

Use TFTP to download the **startup-a2105.ipe** image file from the TFTP server to the root directory of the flash memory on the master device.

```
<Sysname> tftp 2.2.2.2 get startup-a2105.ipe
```

Specify **startup-a2105.ipe** as the main startup image file for all IRF member devices.

```
<Sysname> boot-loader file flash:/startup-a2105.ipe slot 1 main
<Sysname> boot-loader file flash:/startup-a2105.ipe slot 2 main
```

Verify the startup image settings.

```
<Sysname> display boot-loader
```

Reboot the device to complete the upgrade.

```
<Sysname> reboot
```

Verify that the device is running the correct software.

```
<Sysname> display version
```

Contents

Performing an ISSU 1
About ISSU 1
ISSU advantages 1
ISSU methods 1
ISSU commands 2
Operating mechanism 2
Application scenarios 4
Restrictions and guidelines: ISSU 5
Prerequisites for ISSU 5
Logging in to the device through the console port 6
Identifying availability of ISSU and licensing requirements 6
Verifying the device operating status 6
Preparing the upgrade images 6
Identifying the ISSU methods 7
Verifying the device configuration 7
Verifying feature status 7
Determining the upgrade procedure 8
Adjusting and saving the running configuration 8
Performing an ISSU step by step by using issu commands 8
Performing a compatible upgrade on a multichassis IRF fabric 8
Performing an incompatible upgrade on a multichassis IRF fabric 10
Performing an ISSU by using install commands 11
ISSU tasks at a glance 11
Decompressing an .ipe file 11
Installing and upgrading software images 11
Deactivating software images 13
Rolling back the running software images 13
Aborting a software activate or deactivate operation 14
Committing software changes 14
Verifying software images 14
Deleting inactive software images 15
Terminating the ongoing ISSU process forcibly 15
Display and maintenance commands for ISSU 15
Examples of using issu commands for ISSU 17
Example: Upgrading the system software to a compatible version 17
Example: Upgrading the system software to an incompatible version 21
Example: Rolling back the system software 25
Examples of using install commands for ISSU 29
Example: Upgrading the system software 29
Example: Rolling back a feature 32

Performing an ISSU

About ISSU

The In-Service Software Upgrade (ISSU) feature upgrades the Comware software with a minimum amount of downtime.

Unless otherwise stated, the term "upgrade" refers to both software upgrade and downgrade in ISSU.

ISSU advantages

The Comware software includes the boot image, the system image, and feature images. ISSU supports upgrading the images individually.

When you use ISSU to upgrade a feature, only the related feature images are affected. Other features can continue to provide services.

ISSU supports installing patch images to fix system bugs without a system reboot.

ISSU methods

ISSU methods are determined at software release depending on the compatibility between software versions.

ISSU supports the following upgrade types:

- **Compatible upgrade**—The new software version is compatible with the running software version. This upgrade type supports the ISSU methods in Table 1.
- **Incompatible upgrade**—The new software version is incompatible with the running software version. The two versions cannot run concurrently.

  This upgrade type supports only one upgrade method (also called incompatible upgrade). This method requires a cold reboot to upgrade both control and data planes. Incompatible upgrade disrupts service if hardware redundancy is not available.

NOTE:

For the compatibility matrix, see the ISSU upgrade type matrix in the release notes.

For information about identifying the ISSU method, see "Identifying the ISSU methods."

Table 1 ISSU methods for compatible upgrade

ISSU method	Description
Incremental upgrade: • Service Upgrade • File Upgrade	Upgrades only user mode processes that have differences between the new and old software versions. Backup processes and a main/backup process switchover are required for service continuity. • **Service upgrade**—Upgrades service features. The upgrade does not affect the operation of the features that are not being upgraded. • **File upgrade**—Upgrades hidden system program files. The system can provide services during the upgrade.
Reboot	△ **CAUTION:** The Reboot method disrupts service if hardware redundancy is not available. As a best practice, schedule the downtime carefully to minimize the upgrade impact on

ISSU method	Description
	the services. Reboots member devices to complete the software upgrade. While one member device is rebooting, the other member devices can provide services.

ISSU commands

ISSU includes the **install** and **issu** command sets. After you identify the recommended ISSU method, use Table 2 to choose the command set you want to use.

Table 2 Command set comparison

Item	issu commands	install commands
Upgrade types	• Compatible. • Incompatible.	Compatible.
Impact on the system	Large.	Small.
Technical skill requirements	Low. As a best practice, use this command set.	High. Administrators must have extensive system knowledge and understand the impact of each upgrade task on the network.

Operating mechanism

Performing an incremental upgrade on a single-chassis IRF fabric

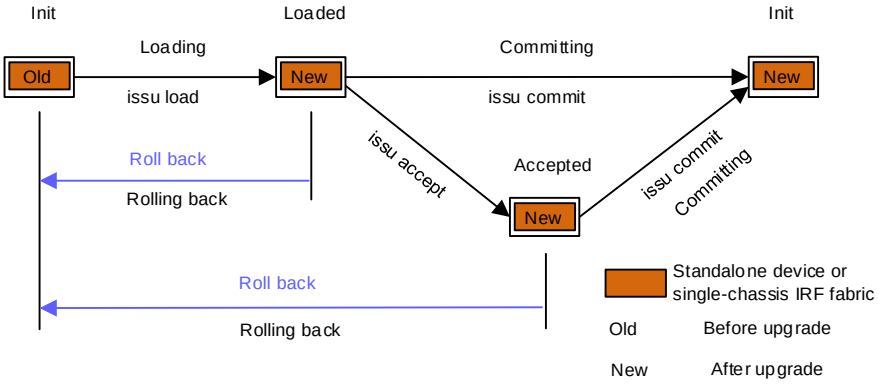
As shown in 0, the upgrade operates as follows:

1. The device is in Init state before the upgrade and enters Loading state when the **issu load** command is executed. When the upgrade completes, the device enters Loaded state.
2. The device enters Accepted state when the **issu accept** command is executed successfully to accept the upgrade to a compatible version and delete the automatic rollback timer.
3. The device enters Committing state when the **issu commit** command is executed to complete the ISSU upgrade to a compatible version, and it enters Init state when the upgrade completes.

IMPORTANT:

If a rollback is performed during the upgrade process, the system will roll back to the version before the upgrade.

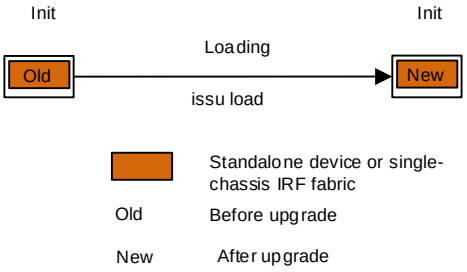
Figure 1 Performing an incremental upgrade on a standalone device or a single-chassis IRF fabric



Performing an incremental upgrade on a single-chassis IRF fabric

As shown in Figure 2, the device is in Init state before the upgrade and enters Loading state when the **issu load** command is executed. When the upgrade completes, the device enters Loaded state.

Figure 2 Performing a reboot, ISSU reboot, or incompatible upgrade on a standalone device or single-chassis IRF fabric



Performing a compatible upgrade on a multichassis IRF fabric

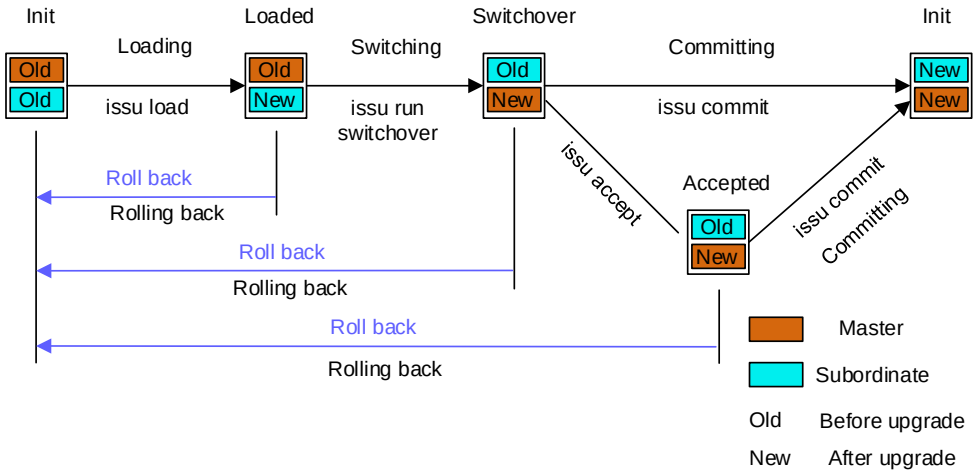
As shown in 0, the upgrade operates as follows:

1. The device is in Init state before the upgrade and enters Loading state when the **issu load** command is executed to upgrade a subordinate device in the IRF fabric. When the upgrade completes, the device enters Loaded state.
2. The device enters Switching state when the **issu run switchover** command is executed to perform an ISSU switchover. When the switchover completes, the device enters Switchover state.
3. The device enters Accepted state when the **issu accept** command is executed successfully to accept the upgrade to a compatible version and delete the automatic rollback timer.
4. The device enters Committing state when the **issu commit** command is executed to upgrade all other member devices in the IRF fabric. When the upgrade completes, the device enters Init state.

IMPORTANT:

If a rollback is performed during the upgrade process, the system will roll back to the version before the upgrade.

Figure 3 Performing a compatible upgrade on a multichassis IRF fabric



Performing an incompatible upgrade on a multichassis IRF fabric

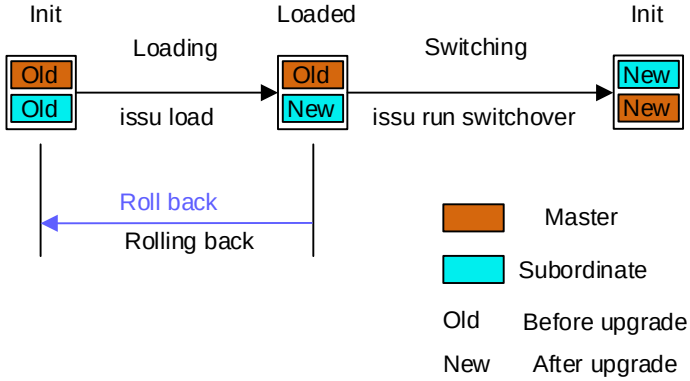
As shown in 0, the upgrade operates as follows:

1. The device is in Init state before the upgrade and enters Loading state when the **`issu load`** command is executed to upgrade a subordinate device in the IRF fabric. When the upgrade completes, the device enters Loaded state.
2. The device enters Switching state when the **`issu run switchover`** command is executed to perform an ISSU switchover. All the other members will be upgraded. When the switchover completes, the device enters Init state.

IMPORTANT:

If a rollback is performed during the upgrade process, the system will roll back to the version before the upgrade.

Figure 4 Performing an incompatible upgrade on a multichassis IRF fabric



Application scenarios

Traditional software upgrade methods require a device reboot, which will affect the services running on the system and communication service quality. To address this issue, you can use ISSU to upgrade the software to ensure uninterrupted services or a short service interruption period during the upgrade process.

Restrictions and guidelines: ISSU

IMPORTANT:

- For a successful ISSU, you must remove all commands that the new version does not support and save the running configuration. To identify the feature changes between the current version and the new version, read the release notes for the device.
- To ensure correct system operation, you must remove the commands configured for features to be uninstalled and save the running configuration before uninstalling the features.

During an ISSU, follow these restrictions and guidelines:

- Do not execute the **`install`**, **`issu`**, or **`boot-loader`** command before the current upgrade action completes.
- Do not perform any of the following tasks:
  - Reboot member devices.
  - Perform tasks that are irrelevant to the ISSU, such as modifying the configuration and displaying information.
  - Modify, delete, or rename image files.
- You cannot use both **`install`** and **`issu`** commands for the same ISSU.
- For better service continuity, strictly follow the recommended ISSU procedure. Make sure a step is completed before you proceed to the next step.
- Before executing the **`issu load`** command for the first time, use the **`display system stable state`** command to verify that the system is stable. If the **System State** field displays **Stable**, the system is stable.
- Before executing the following commands, use the **`display system stable state`** command to verify that the system is stable:
  - **issu commands**—**`issu load`**, **`issu run switchover`**, and **`issu commit`**.
  - **install commands**—**`install activate`** and **`install deactivate`**.

  If the **System State** field displays **Stable**, the system is stable.
- You may use **`issu`** commands to upgrade all or some of the software images. If you are upgrading only some of the images, make sure the new images are compatible with the images that are not to be upgraded. The upgrade will fail if a conflict exists.
- You can install up to 32 .bin files on the device, including one boot image file, one system image file, and up to 30 feature and patch image files.
- A patch image file can be incremental or non-incremental. You can install up to 30 incremental patch image files by using a single command. Because each boot, system, or feature image can have one non-incremental patch image file, you can install up to 16 non-incremental patch image files by using a single command. For more information about incremental and non-incremental patch image files, see "Upgrading software."

After an ISSU, you must log in to the device again before you can use the changed or added commands.

After an ISSU, use the **`display boot-loader`** command to verify that the new software images have been successfully loaded. The startup software images must always be in the flash memory and cannot be deleted. If you delete them, the device cannot restart after being rebooted.

Prerequisites for ISSU

To perform a successful ISSU, make sure all the preparation requirements are met.

Logging in to the device through the console port

Log in to the device through the console port. If you use Telnet or SSH, you might be disconnected from the device before the ISSU is completed.

In a multiuser environment, make sure no other administrators access the device while you are performing the ISSU.

Identifying availability of ISSU and licensing requirements

Read the software release notes to identify the following items:

- Support of the device for ISSU between the current software version and the new software version.
- Licensing requirements for the upgrade software images. If the upgrade software images require licenses, make sure the device has the required licenses. For more information about license installation, see "Managing licenses."

Verifying the device operating status

Verify the following items:

Use the **`display device`** command to verify that all components are operating correctly.

- Use the **`display irf`** command to verify that all member devices are operating correctly.
- Use the **`display cpu-usage`** command to verify that no CPU usage alarm exists on the device. As a best practice, verify that the CPU usage is lower than the CPU usage recovery threshold on each component.
- Use the **`display memory`** and **`display memory-threshold`** commands to verify that no memory alarm exists on the device. As a best practice, verify that the free memory is greater than the normal state threshold on each component.
- Use the **`display fan`** command to verify that all fans are in **Normal** state.
- Use the **`display power`** command to verify that all power supplies are in **Normal** state.
- Use the **`display environment`** command to verify that the temperature of each component is lower than the high-temperature warning threshold.
- Use the **`display counters rate inbound interface`** command to verify that all interfaces are forwarding packets correctly.
- Use the **`display interface | include CRC`** command to verify that all interfaces are forwarding packets correctly.

Preparing the upgrade images

1. Use the **`dir`** command to verify that every file system has sufficient free storage space for the upgrade images. If the storage space is not sufficient, delete unused files by using the **`delete /unreserved`** *`file-url`* command. If the files to be deleted will be used, back up the files before deleting them. You will be unable to restore a deleted file if the **`/unreserved`** keyword is used. For more information, see "Managing file systems."
2. Use FTP or TFTP to transfer upgrade image files (in .bin or .ipe) to the root directory of a file system on the master device.

Identifying the ISSU methods

1. Execute the **`display version comp-matrix file`** command to identify the recommended ISSU methods.
   - For a compatible upgrade, check the **Upgrade Way** field to identify the recommended ISSU methods.
   - For an incompatible upgrade, check the end of command output for the **Incompatible upgrade** string.

   For more information about ISSU methods, see Table 1.

Verifying the device configuration

For a successful ISSU reboot or incompatible upgrade and the IRF fabric integrity during the ISSU, a set of features must have the same configuration as expected after the upgrade. The following are the features and the commands for you to change and verify feature settings:

- System operating mode.

  `system-working-mode`

  `display system-working-mode`

Maximum number of ECMP routes.

`max-ecmp-num`

`display max-ecmp-num`

For more information about system operating mode, see device management in *Fundamentals Configuration Guide*.

For more information about ECMP routes, see basic IP routing configuration in *Layer 3—IP Routing Configuration Guide*.

Verifying feature status

For service continuity during an ISSU, configure the following feature settings:

Feature	Setting requirements
GR and NSR	Enable GR or NSR for protocols including OSPF, ISIS, BGP, and FSPF. Use the **`display`** commands to verify that the system is in stable state after GR or NSR is enabled for a protocol. For example, use the **`display ospf graceful-restart`** or **`display ospf non-stop-routing status`** command to verify that the system is in stable state after GR or NSR is enabled for OSPF.
BFD	Disable BFD for protocols including OSPF, IS-IS, RIP, BGP, VRRP, and NQA.
Ethernet link aggregation	Use the long LACP timeout interval (the **`lacp period short`** command is not configured) on all member ports in dynamic aggregation groups. Make sure all member ports are Selected ports.
IRF	Before performing an incompatible upgrade for the IRF fabric, verify that IRF MAD is disabled. To use IRF MAD, enable IRF MAD after finishing the upgrade. Configure IRF bridge MAC persistence as follows: • **Compatible upgrade**—Execute the **`irf mac-address persistent timer`** or **`irf mac-address persistent`**

Feature	Setting requirements
	always command. • **Incompatible upgrade**—Execute the **irf mac-address persistent always** command if the bridge MAC address is the MAC address of the device for which you want to execute the **issu load** command.

Determining the upgrade procedure

1. Use Table 2 to choose an upgrade command set, depending on the ISSU method.
2. Identify the hardware redundancy condition.

   ISSU can maintain service continuity only when the following conditions are met:

   - The IRF fabric has multiple members and uses the ring topology.

IMPORTANT:

If hardware redundancy is not available, service discontinuity is not avoidable during a reboot upgrade or an incompatible upgrade. Make sure you understand the impact of the upgrade on the network.

3. Choose the correct procedure from the procedures described in "Performing an ISSU step by step by using issu commands" or "Performing an ISSU by using install commands."

Adjusting and saving the running configuration

1. Remove all commands that the new software version does not support from the running configuration. To identify all feature changes between the current version and the new version, read the release notes for the device.
2. To uninstall a feature image, remove the commands configured for the feature.
3. Use the **save** command to save the running configuration.

Performing an ISSU step by step by using issu commands

Performing a compatible upgrade on a multichassis IRF fabric

Restrictions and guidelines

Upgrade a subordinate member device first. Then, upgrade the remaining member devices, including the original master.

Procedure

1. (Optional.) Configure automatic rollback:
   a. Enter system view.

      system-view

   b. Set the automatic rollback timer.

issu rollback-timer *minutes*

By default, the automatic rollback timer is set to 45 minutes.

The automatic rollback timer starts when you execute the **issu run switchover** command.

c. Return to user view.

quit

2. Verify that the system is stable.

display system stable state

The system is stable if the **System State** field displays **Stable**. For a successful ISSU, you must make sure the system is stable before you proceed to the next step.

3. Load the upgrade images as startup images on a subordinate member.

- Use .bin files:

  issu load file { **boot** *filename* | **system** *filename* | **feature** *filename*&<1-30> | **patch** *filename*&<1-30> } * **slot** *slot-number*&<1-9> [**reboot**]

- Use an .ipe file:

  issu load file ipe *ipe-filename* [**patch** *filename*&<1-30>] **slot** *slot-number*&<1-9> [**reboot**]

NOTE:

The software images for the device are digitally signed. The system verifies the digital signatures of the specified software images for authenticity and integrity before it sets or loads them as main startup images. If the digital signature verification fails, the system will not set or load the specified images as main startup images and you will receive a digital signature verification failure message.

4. Verify that the system is stable.

display system stable state

The system is stable if the **System State** field displays **Stable**. For a successful ISSU, you must make sure the system is stable before you proceed to the next step. If the system reported an error in the previous step, verify that the system is stable and then use the **issu rollback** command to roll back the upgrade.

5. Perform an ISSU switchover.

issu run switchover

This command also starts the automatic rollback timer. When the timer expires, the system automatically rolls back to the original software images.

6. (Optional.) Accept the upgrade and delete the automatic rollback timer.

issu accept

Execute this command before the automatic rollback timer expires.

7. Verify that the system is stable.

display system stable state

The system is stable if the **System State** field displays **Stable**. For a successful ISSU, you must make sure the system is stable before you proceed to the next step. If the system reported an error in the previous two steps, verify that the system is stable and then use the **issu rollback** command to roll back the upgrade.

8. Upgrade the remaining members to complete the ISSU.

issu commit slot *slot-number*

Execute this command before the automatic rollback timer expires.

CAUTION:

Repeat the previous step and this step to upgrade the remaining member devices, including the original master device. After you execute this command for one member device, wait for the member device to start up and join the IRF fabric again before you upgrade another member device. This practice ensures a successful ISSU.

9. Verify that the ISSU is finished.

   display issu state

   If the **ISSU state** field displays **Init**, the ISSU is finished.

Performing an incompatible upgrade on a multichassis IRF fabric

Restrictions and guidelines

Upgrade one or more subordinate member devices first. Then, upgrade the remaining member devices, including the original master.

Procedure

To perform an incompatible upgrade on a multichassis IRF fabric, execute the following commands in user view:

1. Verify that the system is stable.

   display system stable state

   The system is stable if the **System State** field displays **Stable**. For a successful ISSU, you must make sure the system is stable before you proceed to the next step.

2. Load the upgrade images as startup images on subordinate members.

   - Use .bin files:

     issu load file { **boot** *filename* | **system** *filename* | **feature** *filename*&<1-30> | **patch** *filename*&<1-30> } * **slot** *slot-number*&<1-9> [**reboot**]

   - Use an .ipe file:

     issu load file ipe *ipe-filename* [**patch** *filename*&<1-30>] **slot** *slot-number*&<1-9> [**reboot**]

   As a best practice on a ring-topology IRF fabric, specify half of the subordinate members for this command to reduce service interruption. Make sure the specified subordinate members are physically connected.

   NOTE:

   The software images for the device are digitally signed. The system verifies the digital signatures of the specified software images for authenticity and integrity before it sets or loads them as main startup images. If the digital signature verification fails, the system will not set or load the specified images as main startup images and you will receive a digital signature verification failure message.

3. Verify that the system is stable.

   display system stable state

The system is stable if the **System State** field displays **Stable**. For a successful ISSU, you must make sure the system is stable before you proceed to the next step. If the system reported an error in the previous step, verify that the system is stable and then use the **`issu rollback`** command to roll back the upgrade.

4. Perform an ISSU switchover to complete the ISSU process.

   `issu run switchover`

   This command upgrades the remaining members.

5. Verify that the ISSU is finished.

   `display issu state`

   If the **ISSU state** field displays **Init**, the ISSU is finished.

Performing an ISSU by using install commands

ISSU tasks at a glance

1. (Optional.) Decompressing an .ipe file
2. Installing and upgrading software images
3. (Optional.) Deactivating software images
4. (Optional.) Rolling back the running software images
5. (Optional.) Aborting a software activate or deactivate operation
6. (Optional.) Verifying software images
7. Committing software changes
8. (Optional.) Deleting inactive software images

Decompressing an .ipe file

1. (Optional.) Identify images that are included in the .ipe file.

   `display install ipe-info`

2. Decompress the .ipe file.

   `install add` *`ipe-filename filesystem`*

Installing and upgrading software images

About this task

Use this task to install new feature and patch images or upgrade the boot, system, or feature images.

Software image installation and upgrade methods

Use one of the following methods to install or upgrade software images:

- **Slot by slot**—Activate all the images on one slot, and then move to the next slot.
- **Image by image**—Activate one image on all slots before moving to another image. Before using this method, read the release notes to make sure the upgrade images are compatible with the current images. If the upgrade images are not compatible with the current images, the upgrade might fail.

Restrictions and guidelines

To install an image, you must begin with the master device. To upgrade an image, you must begin with a subordinate device.

For incremental upgrades and patch installations, the software images take effect only during the current device operation. For the software image to take effect across device reboots, you must execute the **install commit** command. For reboot upgrades, you do not need to execute this command, because the system has already modified the next startup software list when you execute the **install activate** command.

Installing or upgrading boot, system, and feature images

Execute the following commands in user view:

1. Verify that the system is stable.

   display system stable state

   The system is stable if the **System State** field displays **Stable**. For a successful ISSU, you must make sure the system is stable before you proceed to the next step.

2. (Optional.) Identify the recommended ISSU method and the possible impact of the upgrade.

   install activate { **boot** *filename* | **system** *filename* | **feature** *filename*&<1-30> } * **slot** *slot-number* **test**

3. Activate images.

   install activate { **boot** *filename* | **system** *filename* | **feature** *filename*&<1-30> } * **slot** *slot-number*

NOTE:

The software images for the device are digitally signed. The system will verify the digital signatures of the specified software images for authenticity and integrity before it activates them. If the digital signature verification fails, the system will not activate the specified images and you will receive a digital signature verification failure message.

Installing patch images

To install patch images, execute the following commands in user view:

Table 3 Verify that the system is stable.

display system stable state

The system is stable if the **System State** field displays **Stable**. For a successful installation, you must make sure the system is stable before you proceed to the next step.

1. Install patch images.

   install activate patch *filename* { **all** | **slot** *slot-number* }

   You can specify only one patch image file for this command at a time. However, you can execute this command multiple times to install multiple patch image files.

   The **install activate patch** *filename* **all** command installs the specified patch images on all hardware and the images can survive a reboot. You do not need to execute the **install commit** command for the installation.

NOTE:

The software images for the device are digitally signed. The system verifies the digital signature of a patch image for authenticity and integrity before it activates it. If the digital signature verification fails, the system will not activate the image and you will receive a digital signature verification failure message.

Deactivating software images

Restrictions and guidelines

You can deactivate only feature and patch images.

The deactivate operation only removes images from the current software image list. For the image changes to take effect after a reboot, you must perform a commit operation to remove the images from the main startup image list.

Deactivated images are still stored on the storage medium. To permanently delete the images, execute the **install remove** command. For more information, see "Deleting inactive software images."

Deactivating feature images

To deactivate feature images, execute the following commands in user view:

Table 4 Verify that the system is stable.

display system stable state

The system is stable if the **System State** field displays **Stable**. For a successful deactivate operation, you must make sure the system is stable before you proceed to the next step.

1. Deactivate feature images.

   install deactivate feature *filename*&<1-30> **slot** *slot-number*

Deactivating patch images

To deactivate patch images, execute the following commands in user view:

Table 5 Verify that the system is stable.

display system stable state

The system is stable if the **System State** field displays **Stable**. For a successful deactivate operation, you must make sure the system is stable before you proceed to the next step.

1. Deactivate patch images.

   install deactivate patch *filename* { **all** | **slot** *slot-number* }

   You can specify only one patch image file for this command at a time. However, you can execute this command multiple times to deactivate multiple patch image files.

   The **install deactivate patch** *filename* **all** command deactivates the specified patch images on all hardware and the image changes can survive a reboot. You do not need to execute the **install commit** command for the deactivation.

Rolling back the running software images

About this task

During an incremental upgrade, the system creates a rollback point for each activate or deactivate operation of a boot, system, or feature image. The system can maintain a maximum of 50 rollback points. If this limit has been reached when a rollback point is created, the system removes the earliest rollback point. You can roll back the software to any of the rollback points.

During a reboot upgrade, the system does not create rollback points. After a reboot upgrade, you can roll back the software only to the status before any activate or deactivate operations were performed.

Restrictions and guidelines

You can perform this task only before committing software changes. A commit operation deletes all rollback points.

For an incremental upgrade rollback to take effect after a reboot, you must perform a commit operation to update the main startup image list.

Procedure

To roll back the running software images, execute the following commands in user view:

1. (Optional.) Display available rollback points.
   display install rollback
2. Roll back the software.
   install rollback to { *point-id* | **original** }

Aborting a software activate or deactivate operation

About this task

While the system is activating or deactivating a software image for a service upgrade or file upgrade, you can abort the activate or deactivate operation. After an operation is aborted, the system runs with the software images that it was running with before the operation.

Procedure

To abort a software activate or deactivate operation, use one of the following methods:

- Execute the **install abort** [*job-id*] command in user view.
- Press **Ctrl+C**.

Committing software changes

About this task

When you activate or deactivate images for an incremental upgrade, or activate or deactivate patch images, the main startup image list does not update with the changes. The software changes are lost at reboot. For the changes to take effect after a reboot, you must commit the changes.

Procedure

To commit software changes, execute the following command in user view:

install commit

Verifying software images

About this task

Perform this task to verify the following items:

- **Integrity**—Verify that the boot, system, and feature images are integral.
- **Consistency**—Verify that the same active images are running across the entire system.
- **Software commit status**—Verify that the active images are committed as needed.

Procedure

To verify software images, execute the following commands in user view:

1. Verify software images.
   install verify

2. Activate or deactivate images as required.
   install { **activate** | **deactivate** }
3. Commit the software changes.
   install commit

Deleting inactive software images

About this task

After completing an ISSU, you can use this task to delete old image files permanently.

Restrictions and guidelines

This task deletes image files permanently. You cannot use the **install rollback to** command to revert the operation, or use the **install abort** command to abort the operation.

Procedure

To delete inactive software image files, execute the following command in user view:

install remove [**slot** *slot-number*] { *filename* | **inactive** }

Terminating the ongoing ISSU process forcibly

About this task

Use this feature to terminate the ISSU process if one of the following exceptions occurs and you cannot perform an upgrade or rollback:

- The ISSU status is not **Init** but the upgrade has stopped.
- The ISSU status is **Init** but the upgrade has not completed.

Procedure

1. Enter system view.
   system-view
2. Terminate the ongoing ISSU process forcibly.
   issu quit

Display and maintenance commands for ISSU

CAUTION:
The **reset install rollback oldest** command clears the specified rollback point and all rollback points earlier than the specified rollback point. You will be unable to roll back the configuration to the status when any of these rollback points was created.

Unless otherwise stated, the **display** and **reset** commands can be used during an ISSU, regardless of whether the **install** or **issu** commands are used.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command	Remarks
Display active software	**display install active**	N/A

Task	Command	Remarks
images.	[**slot** *slot-number*] [**verbose**]	
Display backup startup software images.	**display install backup** [**slot** *slot-number*] [**verbose**]	N/A
Display main startup software images.	**display install committed** [**slot** *slot-number*] [**verbose**]	N/A
Display inactive software images in the root directories of file systems.	**display install inactive** [**slot** *slot-number*] [**verbose**]	N/A
Display the software images included in an .ipe file.	**display install ipe-info** *ipe-filename*	N/A
Display ongoing ISSU activate, deactivate, and rollback operations.	**display install job**	N/A
Display ISSU log entries.	**display install log** [*log-id*] [**verbose**]	N/A
Display software image file information.	**display install package** { *filename* | **all** } [**verbose**]	N/A
Display rollback point information.	**display install rollback** [*point-id*]	The system does not record rollback points during an ISSU that uses **issu** commands.
Display the software image file that includes a specific component or file.	**display install which** { **component** *name* | **file** *filename* } [**slot** *slot-number*]	N/A
Display automatic rollback timer information.	**display issu rollback-timer**	N/A
Display ISSU status information.	**display issu state**	This command applies only to an ISSU that uses **issu** commands.
Display the ISSU methods.	**display version comp-matrix file** { **boot** *filename* | **system** *filename* | **feature** *filename*&<1-30> | **patch** *filename*&<1-30> } * **display version comp-matrix file ipe** *ipe-filename* [**patch** *filename*&<1-30>]	N/A
Clear ISSU log entries.	**reset install log-history oldest** *log-number*	N/A
Clear ISSU rollback points.	**reset install rollback oldest** *point-id*	N/A

Examples of using issu commands for ISSU

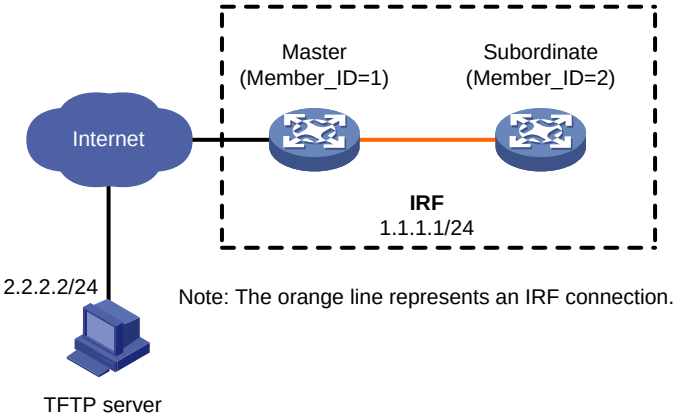
Example: Upgrading the system software to a compatible version

Network configuration

As shown in Figure 5, the IRF fabric has two members.

Upgrade the boot, system, and feature images from T0001015 to T0001016. The two versions are compatible.

Figure 5 Network diagram



Procedure

Download the upgrade image files from the TFTP server.

```
<Sysname> tftp 2.2.2.2 get boot-t0001016.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   6078464  100  6078464     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
<Sysname> tftp 2.2.2.2 get system-t0001016.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   97750016  100  97750016     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
<Sysname> tftp 2.2.2.2 get feature-t0001016.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   1008640  100  1008640     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
```

Display active software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
```

```
Active packages on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
```

Identify the recommended ISSU methods.

```
<Sysname> display version comp-matrix file boot flash:/boot-t0001016.bin system
flash:/system-t0001016.bin feature flash:/feature-t0001016.bin
Verifying digital signature of the file flash:/boot-t0001016.bin on slot
1.....................Done.
Verifying digital signature of the file flash:/system-t0001016.bin on slot
1.............. .................Done.
Verifying digital signature of the file flash:/feature-t0001016.bin on slot
1.................Done.
Identifying the upgrade methods....Done.

  Slot                          Upgrade Way
  1                             Reboot
  2                             Reboot
```

The output shows that reboot upgrades are recommended.

Save the running configuration.

```
<Sysname> save
```

Upgrade the system software on the subordinate member.

```
<Sysname> issu load file boot flash:/boot-t0001016.bin system flash:/system-t0001016.bin
feature flash:/feature-t0001016.bin slot 2
This operation will delete the rollback point information for the previous upgrade and
maybe get unsaved configuration lost. Continue? [Y/N]:y
Copying file flash:/boot-t0001016.bin to slot2#flash:/boot-t0001016.bin......Done.
Copying file flash:/system-t0001016.bin to
slot2#flash:/system-t0001016.bin.....................Done.
Copying file flash:/feature-t0001016.bin to
slot2#flash:/feature-t0001016.bin......Done.
Verifying digital signature of the file flash:/boot-t0001016.bin on slot 2...Done.
Verifying digital signature of the file flash:/system-t0001016.bin on slot
2............Done.
Verifying digital signature of the file flash:/feature-t0001016.bin on slot 2...Done.
Identifying the upgrade methods….Done.
Upgrade summary according to following table:

flash:/boot-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/system-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/feature-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016
```

```
  Slot                          Upgrade Way
  2                             Reboot
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...Done.
```

Verify that the ISSU is in a stable state.

```
<Sysname> display issu state
ISSU state: Loaded
Compatibility: Compatible
Work state: Normal
Upgrade method: Card by card
Upgraded slot:
  slot 2
Current upgrading slot: None
Previous version list:
  boot: 7.1.070, Test 0001015
  system: 7.1.070, Test 0001015
  feature: 7.1.070, Test 0001015
Previous software images:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Upgrade version list:
  boot: 7.1.070, Test 0001016
  system: 7.1.070, Test 0001016
  feature: 7.1.070, Test 0001016
Upgrade software images:
  flash:/boot-t0001016.bin
  flash:/system-t0001016.bin
  flash:/feature-t0001016.bin
```

The **Loaded** state is a stable state, which indicates that the system is waiting for a master/subordinate switchover. Slot 2 has completed the upgrade, and slot 1 has not.

Perform a master/subordinate switchover.

```
<Sysname> issu run switchover
Upgrade summary according to following table:

flash:/boot-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/system-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/feature-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016
```

```
  Slot                        Switchover Way
  1                           Master subordinate switchover
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait.........
```

Verify that the ISSU is in a stable state.

```
<Sysname> display issu state
ISSU state: Switchover
Compatibility: Compatible
Work state: Normal
Upgrade method: Card by card
Upgraded slot:
  slot 2
Current upgrading slot: None
Previous version list:
  boot: 7.1.070, Test 0001015
  system: 7.1.070, Test 0001015
  feature: 7.1.070, Test 0001015
Previous software images:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Upgrade version list:
  boot: 7.1.070, Test 0001016
  system: 7.1.070, Test 0001016
  feature: 7.1.070, Test 0001016
Upgrade software images:
  flash:/boot-t0001016.bin
  flash:/system-t0001016.bin
  flash:/feature-t0001016.bin
```

The **Switchover** state is a stable state, which indicates that the system has completed the master/subordinate switchover and is waiting for a commit operation to upgrade slot 1.

Upgrade the system software on the original master.

```
<Sysname> issu commit slot 1
Upgrade summary according to following table:

flash:/boot-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/system-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/feature-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

  Slot                        Upgrade Way
```

```
  1                               Reboot
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...Done.
```

Verifying the upgrade

Verify that the ISSU is finished.

```
<Sysname> display issu state
ISSU state: Init
Compatibility: Unknown
Work state: Normal
Upgrade method: Card by card
Upgraded slot: None
Current upgrading slot: None
Current version list:
  boot: 7.1.070, Test 0001016
  system: 7.1.070, Test 0001016
  feature: 7.1.070, Test 0001016
Current software images:
  flash:/boot-t0001016.bin
  flash:/system-t0001016.bin
  flash:/feature-t0001016.bin
```

The **ISSU state** field displays **Init**, which indicates that the ISSU is finished.

Verify that both members are running the new software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001016.bin
  flash:/system-t0001016.bin
  flash:/feature-t0001016.bin
Active packages on slot 2:
  flash:/boot-t0001016.bin
  flash:/system-t0001016.bin
  flash:/feature-t0001016.bin
```

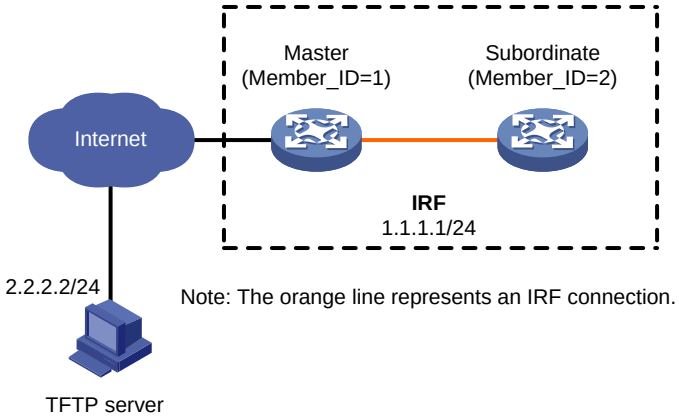
Example: Upgrading the system software to an incompatible version

Network configuration

As shown in Figure 6, the IRF fabric has two members.

Upgrade the boot, system, and feature images from T0001015 to T0001017, which is an incompatible version.

Figure 6 Network diagram



Procedure

Download the upgrade image files from the TFTP server.

```
<Sysname> tftp 2.2.2.2 get boot-t0001017.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   6078464  100  6078464     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
<Sysname> tftp 2.2.2.2 get system-t0001017.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   97750016  100  97750016     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
<Sysname> tftp 2.2.2.2 get feature-t0001017.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   1008640  100  1008640     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
```

Display active software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Active packages on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
```

Identify the recommended ISSU methods.

```
<Sysname> display version comp-matrix file boot flash:/boot-t0001017.bin system
flash:/system-t0001017.bin feature flash:/feature-t0001017.bin
Verifying digital signature of the file flash:/boot-t0001017.bin on slot
1.....................Done.
Verifying digital signature of the file flash:/system-t0001017.bin on slot
1.....................  .................Done.
```

```
Verifying digital signature of the file flash:/feature-t0001017.bin on slot
1.................Done.
Identifying the upgrade methods....Done.

Incompatible upgrade.
```

The output shows that an incompatible upgrade is recommended.

Save the running configuration.

```
<Sysname> save
```

Upgrade the system software on the subordinate member.

```
<Sysname> issu load file boot flash:/boot-t0001017.bin system flash:/system-t0001017.bin
feature flash:/feature-t0001017.bin slot 2
This operation will delete the rollback point information for the previous upgrade and
maybe get unsaved configuration lost. Continue? [Y/N]:y
Copying file flash:/boot-t0001017.bin to slot2#flash:/boot-t0001017.bin......Done.
Copying file flash:/system-t0001017.bin to
slot2#flash:/system-t0001017.bin.......................Done.
Copying file flash:/feature-t0001017.bin to
slot2#flash:/feature-t0001017.bin......Done.
Verifying digital signature of the file flash:/boot-t0001017.bin on slot 2...Done.
Verifying digital signature of the file flash:/system-t0001017.bin on slot
2............Done.
Verifying digital signature of the file flash:/feature-t0001017.bin on slot 2...Done.
Identifying the upgrade methods….Done.
Upgrade summary according to following table:

flash:/boot-t0001017.bin
  Running Version             New Version
  Test 0001015                Test 0001017

flash:/system-t0001017.bin
  Running Version             New Version
  Test 0001015                Test 0001017

flash:/feature-t0001017.bin
  Running Version             New Version
  Test 0001015                Test 0001017

  Slot                        Upgrade Way
  2                           Reboot
Upgrading software images to incompatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...Done.
```

Verify that the ISSU is in a stable state.

```
<Sysname> display issu state
ISSU state: Loaded
Compatibility: Incompatible
Work state: Normal
Upgrade method: Card by card
Upgraded slot:
```

```
  slot 2
Current upgrading slot: None
Previous version list:
  boot: 7.1.070, Test 0001015
  system: 7.1.070, Test 0001015
  feature: 7.1.070, Test 0001015
Previous software images:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Upgrade version list:
  boot: 7.1.070, Test 0001017
  system: 7.1.070, Test 0001017
  feature: 7.1.070, Test 0001017
Upgrade software images:
  flash:/boot-t0001017.bin
  flash:/system-t0001017.bin
  flash:/feature-t0001017.bin
```

The **Loaded** state is a stable state, which indicates that the system is waiting for a master/subordinate switchover. Slot 2 has completed the upgrade, and slot 1 has not.

Perform a master/subordinate switchover.

```
<Sysname> issu run switchover
Upgrade summary according to following table:

flash:/boot-t0001017.bin
  Running Version             New Version
  Test 0001015                Test 0001017

flash:/system-t0001017.bin
  Running Version             New Version
  Test 0001015                Test 0001017

flash:/feature-t0001017.bin
  Running Version             New Version
  Test 0001015                Test 0001017

  Slot                        Switchover Way
  1                           Master subordinate switchover
Upgrading software images to incompatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...Done.
```

Verifying the upgrade

Verify that the ISSU is finished.

```
<Sysname> display issu state
ISSU state: Init
Compatibility: Unknown
Work state: Normal
Upgrade method: Card by card
Upgraded slot: None
```

```
Current upgrading slot: None
Current version list:
  boot: 7.1.070, Test 0001017
  system: 7.1.070, Test 0001017
  feature: 7.1.070, Test 0001017
Current software images:
  flash:/boot-t0001017.bin
  flash:/system-t0001017.bin
  flash:/feature-t0001017.bin
```

The **ISSU state** field displays **Init**, which indicates that the ISSU is finished.

Verify that both members are running the new software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001017.bin
  flash:/system-t0001017.bin
  flash:/feature-t0001017.bin
Active packages on slot 2:
  flash:/boot-t0001017.bin
  flash:/system-t0001017.bin
  flash:/feature-t0001017.bin
```

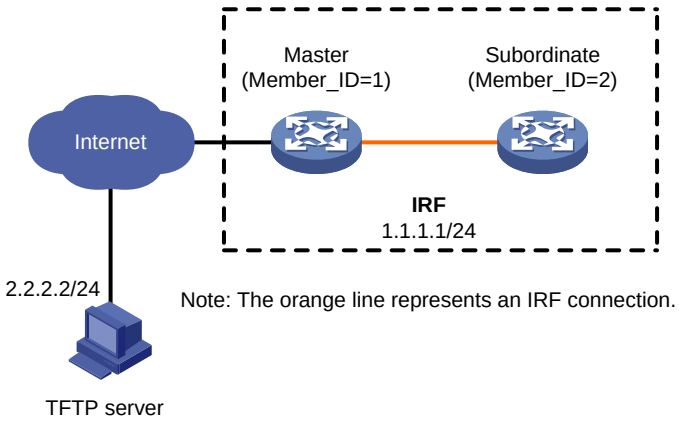
Example: Rolling back the system software

Network configuration

As shown in Figure 7, the IRF fabric has two members.

Roll back the boot, system, and feature images from T0001016 to T0001015 after upgrading them from T0001015 to T0001016. T0001016 and T0001015 are compatible.

Figure 7 Network diagram



Procedure

Download the upgrade image files from the TFTP server.

```
<Sysname> tftp 2.2.2.2 get boot-t0001016.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   6078464  100  6078464     0     0    764      0 --:--:-- --:--:-- --:--:--   810
```

```
Writing file...Done.
<Sysname> tftp 2.2.2.2 get system-t0001016.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   97750016  100  97750016     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
<Sysname> tftp 2.2.2.2 get feature-t0001016.bin
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   1008640  100  1008640     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
```

Display active software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Active packages on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
```

Identify the recommended ISSU methods.

```
<Sysname> display version comp-matrix file boot flash:/boot-t0001016.bin system
flash:/system-t0001016.bin feature flash:/feature-t0001016.bin
Verifying digital signature of the file flash:/all.ipe on slot
1................. .................Done.
Identifying the upgrade methods....Done.

  Slot                          Upgrade Way
  1                             Reboot
  2                             Reboot
```

The output shows that reboot upgrades are recommended.

Save the running configuration.

```
<Sysname> save
```

Upgrade the system software on the subordinate member.

```
<Sysname> issu load file boot flash:/boot-t0001016.bin system flash:/system-t0001016.bin
feature flash:/feature-t0001016.bin slot 2
This operation will delete the rollback point information for the previous upgrade and
maybe get unsaved configuration lost. Continue? [Y/N]:y
Copying file flash:/boot-t0001016.bin to slot2#flash:/boot-t0001016.bin......Done.
Copying file flash:/system-t0001016.bin to
slot2#flash:/system-t0001016.bin.......................Done.
Copying file flash:/feature-t0001016.bin to
slot2#flash:/feature-t0001016.bin......Done.
Verifying digital signature of the file flash:/boot-t0001016.bin on slot 2...Done.
Verifying digital signature of the file flash:/system-t0001016.bin on slot
2............Done.
Verifying digital signature of the file flash:/feature-t0001016.bin on slot 2...Done.
```

```
Identifying the upgrade methods….Done.
Upgrade summary according to following table:

flash:/boot-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/system-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/feature-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

  Slot                        Upgrade Way
  2                           Reboot
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...Done.
```

Verify that the ISSU is in a stable state.

```
<Sysname> display issu state
ISSU state: Loaded
Compatibility: Compatible
Work state: Normal
Upgrade method: Card by card
Upgraded slot:
  slot 2
Current upgrading slot: None
Previous version list:
  boot: 7.1.070, Test 0001015
  system: 7.1.070, Test 0001015
  feature: 7.1.070, Test 0001015
Previous software images:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Upgrade version list:
  boot: 7.1.070, Test 0001016
  system: 7.1.070, Test 0001016
  feature: 7.1.070, Test 0001016
Upgrade software images:
  flash:/boot-t0001016.bin
  flash:/system-t0001016.bin
  flash:/feature-t0001016.bin
```

The **Loaded** state is a stable state, which indicates that the system is waiting for a master/subordinate switchover. Slot 2 has completed the upgrade, and slot 1 has not.

Perform a master/subordinate switchover.

```
<Sysname> issu run switchover
```

```
Upgrade summary according to following table:

flash:/boot-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/system-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

flash:/feature-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

  Slot                        Switchover Way
  1                           Master subordinate switchover
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait.........
```

Verify that the ISSU is in a stable state.

```
<Sysname> display issu state
ISSU state: Switchover
Compatibility: Compatible
Work state: Normal
Upgrade method: Card by card
Upgraded slot:
  slot 2
Current upgrading slot: None
Previous version list:
  boot: 7.1.070, Test 0001015
  system: 7.1.070, Test 0001015
  feature: 7.1.070, Test 0001015
Previous software images:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Upgrade version list:
  boot: 7.1.070, Test 0001016
  system: 7.1.070, Test 0001016
  feature: 7.1.070, Test 0001016
Upgrade software images:
  flash:/boot-t0001016.bin
  flash:/system-t0001016.bin
  flash:/feature-t0001016.bin
```

The **Switchover** state is a stable state, which indicates that the system has completed the master/subordinate switchover and is waiting for a commit operation to upgrade slot 1.

Roll back the software images to T0001015.

```
<Sysname> issu rollback
```

```
This command will quit the ISSU process and roll back to the previous version. Continue?
[Y/N]:Y
```

Verify that the rollback is finished.

```
<Sysname> display issu state
ISSU state: Init
Compatibility: Unknown
Work state: Normal
Upgrade method: Card by card
Upgraded slot: None
Current upgrading slot: None
Current version list:
  boot: 7.1.070, Test 0001015
  system: 7.1.070, Test 0001015
  feature: 7.1.070, Test 0001015
Current software images:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
```

The **ISSU state** field displays **Init**, which indicates that the rollback is finished.

Verify that both members are running the old software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
Active packages on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature-t0001015.bin
```

Examples of using install commands for ISSU

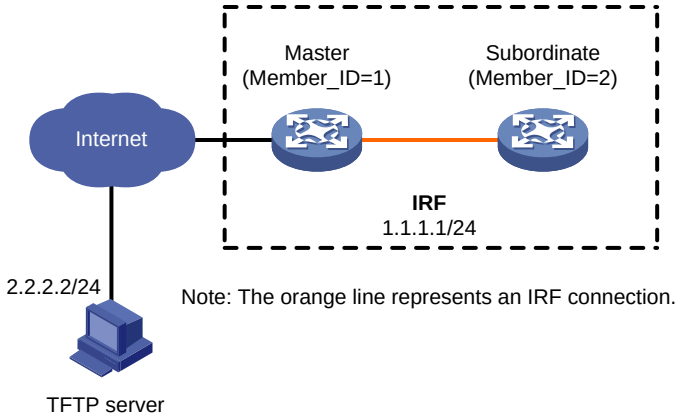
Example: Upgrading the system software

Network configuration

As shown in Figure 8, the IRF fabric has two members.

Upgrade the Feature1 feature image from T0001015 to T0001016. The two versions are compatible.

Figure 8 Network diagram



Procedure

Download the .ipe file that contains the T0001016 Feature1 feature image from the TFTP server.

```
<Sysname> tftp 2.2.2.2 get feature1-t0001016.ipe
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   256  100  256     0     0    764      0 --:--:-- --:--:-- --:--:--   810
Writing file...Done.
```

Decompress the .ipe file.

```
<Sysname> install add flash:/feature1-t0001016.ipe flash:
Verifying digital signature of the file flash:/feature1-t0001016.ipe on slot 1...Done.
Decompressing file feature1-t0001016.bin to
flash:/feature1-t0001016.bin.....................Done.
```

Display active software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001015.bin
Active packages on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001015.bin
```

Identify the recommended ISSU methods and possible impact of the upgrade.

```
<Sysname> install activate feature flash:/feature1-t0001016.bin slot 2 test
Copying file flash:/feature1-t0001016.bin to
slot2#flash:/feature1-t0001016.bin......Done.
Verifying digital signature of the file flash:/feature1-t0001016.bin on slot 2...Done.
Upgrade summary according to following table:

flash:/feature1-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

  Slot                        Upgrade Way
```

```
  2                             Service Upgrade

Influenced service according to following table on slot 2:
  flash:/feature1-t0001016.bin
         Feature1
<Sysname> install activate feature flash:/feature1-t0001016.bin slot 1 test
Verifying digital signature of the file flash:/feature1-t0001016.bin on slot 1...Done.
Upgrade summary according to following table:

flash:/feature1-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

  Slot                        Upgrade Way
  1                           Service Upgrade

Influenced service according to following table on slot 1:
  flash:/feature1-t0001016.bin
         Feature1
```

The output shows that both members need a service upgrade and the Feature1 module will be rebooted during the upgrade.

Activate the new Feature1 image to upgrade the Feature1 feature.

```
<Sysname> install activate feature flash:/feature1-t0001016.bin slot 2
Verifying digital signature of the file flash:/feature1-t0001016.bin on slot 1...Done.
flash:/feature1-t0001016.bin already exists on slot 2.
Overwrite it?[Y/N]:y
Copying file flash:/feature1-t0001016.bin to
slot2#flash:/feature1-t0001016.bin......Done.
Verifying digital signature of the file flash:/feature1-t0001016.bin on slot 2...Done.
Upgrade summary according to following table:

flash:/feature1-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

  Slot                        Upgrade Way
  2                           Service Upgrade
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...Done.
<Sysname> install activate feature flash:/feature1-t0001016.bin slot 1
Verifying digital signature of the file flash:/feature1-t0001016.bin on slot 1...Done.
Upgrade summary according to following table:

flash:/feature1-t0001016.bin
  Running Version             New Version
  Test 0001015                Test 0001016

  Slot                        Upgrade Way
```

```
  1                           Service Upgrade
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...Done.
```

Display active software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001016.bin
Active packages on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001016.bin
```

Commit the software changes.

```
<Sysname> install commit
This operation will take several minutes, please wait.........................Done.
The current software images have been saved as the startup software images.
Current software images on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001016.bin
Current software images on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001016.bin
```

Example: Rolling back a feature

Network configuration

As shown in Figure 8, the IRF fabric has two members. The Feature1 feature has been upgraded from T0001015 to T0001016. However, the software change has not been committed.

Roll back the Feature1 feature from T0001016 to T0001015.

Procedure

Display active software images.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001016.bin
Active packages on slot2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001016.bin
```

Display available rollback points.

```
<Sysname> display install rollback
  Install rollback information 1 on slot 1:
    Updating from flash:/feature1-t0001015.bin
```

```
            to flash:/feature1-t0001016.bin.
  Install rollback information 2 on slot 2:
    Updating from flash:/feature1-t0001015.bin
            to flash:/feature1-t0001016.bin.
```

Roll back the Feature1 feature to T0001015.

```
<Sysname> install rollback to original
This operation might take several minutes, please wait...Done.
```

Verify that the IRF members are running the old Feature1 image.

```
<Sysname> display install active
Active packages on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001015.bin
Active packages on slot 2:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001015.bin
```

Commit the software changes.

```
<Sysname> install commit
This operation will take several minutes, please wait.........................Done.
The current software images have been saved as the startup software images.
Current software images on slot 0:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001015.bin
Current software images on slot 1:
  flash:/boot-t0001015.bin
  flash:/system-t0001015.bin
  flash:/feature1-t0001015.bin
```

Contents

Configuring graceful insertion and removal 1

About graceful insertion and removal 1
GIR modes 1
How GIR works 1
GIR snapshots 2
GIR tasks at a glance 3
Configuring a custom profile 3
Switching to maintenance mode 5
Switching back to normal mode 6
Manually switching back to normal mode 6
Configuring GIR to automatically switch back to normal mode 7
Managing GIR snapshots 7
Manually creating a snapshot 7
Manually deleting snapshots 8
Including additional data in snapshots 8
Configuring advanced GIR features 8
Configuring the delay for the device to enter maintenance mode 8
Specifying a delay for after_maintenance snapshot creation 9
Display and maintenance commands for GIR 9
GIR configuration examples 9
Example: Configuring GIR in an ECMP-based node redundancy environment 9
Example: Configuring GIR on an M-LAG system 11

Configuring graceful insertion and removal

About graceful insertion and removal

Graceful insertion and removal (GIR) enables you to gracefully isolate the device from the network for device maintenance or upgrade. GIR minimizes service interruption by instructing the affected protocols (for example, routing protocols) to isolate the device and switch over to the redundant path. You do not need to configure graceful switchover protocol by protocol.

GIR modes

The device can operate in either of the following GIR modes:

- **Normal mode**—The device can forward traffic.
- **Maintenance mode**—The device is gracefully placed out of service for maintenance or upgrade purposes. In this mode, the device cannot forward traffic.

How GIR works

Overview

When you place the device in maintenance mode, GIR instructs each affected protocol (for example, BGP) to remove the device from the network. Then, each protocol takes action as instructed to isolate the device and switch over to the redundant path. When you change the device back to normal mode, GIR instructs the affected protocols to add the device back to the network. These protocols are called GIR-capable protocols.

To record the configuration before and after a GIR operation, GIR takes a snapshot of protocol data when the device changes to maintenance mode and when the device changes back to normal mode. You can compare the snapshots to verify configuration data consistency for each protocol.

GIR supports the following protocols: LACP, BGP, IS-IS, OSPF, and OSPFv3. For more information about how each protocol isolates the device, see its configuration guide.

Protocol isolation methods

GIR-capable protocols typically use the isolate and shutdown methods to isolate the device from the network. The following information uses OSPF for example to describe the isolation procedures.

- Isolate switching:
  - **a.** When you place the device in maintenance mode, GIR executes the **`isolate enable`** command for OSPF to increase the link cost in the LSAs advertised by the device.
  - **b.** Each OSPF neighbor of the device reselects an optimal route based on the LSAs and stops forwarding traffic to the device. After the device is fully isolated from the network, you can upgrade the device.
  - **c.** When you revert the device back to normal mode, GIR executes the **`undo isolate enable`** command for OSPF to restore its original link cost and gracefully add the device back to the network.
- Shutdown method:
  - **d.** After a process is shut down by using the **`shutdown process`** command, it sends 1-way hello packets to its neighbors

e. On receipt of the packets, the neighbors disconnect from the OSPF process and use the backup path to forward traffic.

f. The process stops receiving and sending OSPF packets.

g. The process clears its neighbor, LSDB, and OSPF route information.

Compared with the shutdown method, the isolate method uses more resources, but it can faster restore a protocol to the state before the isolation. For more information about the **`isolate enable`** and **`shutdown process`** commands for OSPF, see OSPF commands in *Layer 3—IP Routing Command Reference*.

GIR mode switching

GIR supports the following methods for you to change the device from normal mode to maintenance mode:

- **Isolate switching**—By default, GIR performs isolate switching to change the device from normal mode to maintenance mode if you do not specify shutdown or profile-based switching. With isolate switching, GIR uses the isolation commands for each GIR-capable protocol to isolate the device.
- **Shutdown switching**—GIR uses the shutdown method for each GIR-aware protocol to isolate the device when it changes the device from normal mode to maintenance mode. With this method, GIR also shuts down all interfaces except the management port.
- **Profile-based switching**—GIR executes the custom profile when it changes the system from normal mode to maintenance mode.

NOTE:

A custom profile contains a series of user-specified commands to be executed in sequence upon a mode switching. With GIR, you can configure one switch-to-maintenance profile and one switch-to-normal profile for changing to maintenance mode and normal mode, respectively.

When GIR performs an isolate or shutdown switching, it automatically writes the executed series of commands to the switch-to-maintenance custom profile if the profile is empty.

GIR supports the following methods for you to change the device from maintenance mode back to normal mode:

- **Profile-based switching**—Configure GIR to execute a custom profile for mode switching.
- **Reverse switching**—Configure GIR to perform operations in reverse to the operations performed when it changed from normal mode to maintenance mode.

When GIR performs a reverse switching, it automatically writes the executed series of commands to the switch-to-normal custom profile if the profile is empty.

GIR snapshots

A GIR snapshot records the running data of the device in NETCONF format at a point in time for system restoration or comparison of data to identify changes over time.

GIR snapshots include automatically generated snapshots and manually created snapshots.

The device automatically generates in the following situations:

- When the device enters maintenance mode, it generates a GIR snapshot named **before_maintenance** before each protocol switches traffic over to the redundant path.
- When the device returns to normal mode, it generates a GIR snapshot named **after_maintenance** after all protocols insert the device back into the traffic path and the system enters the stable state. You can use the **`display system stable state`** command to verify that the system is stable.

The device stores only the automatically generated snapshots for the most recent GIR operation.

With GIR, you can compare snapshots to identify data changes over a period of time. For example, you can identify the configuration differences of the device before and after a GIR operation by comparing the two automatically snapshots.

GIR tasks at a glance

To configure GIR, perform the following tasks:

1. Configuring a custom profile

   To perform profile-based switching, configure custom profiles.
2. Perform a mode switching operation
   a. Switching to maintenance mode
   b. Switching back to normal mode
3. Managing GIR snapshots
   - Manually creating a snapshot
   - Manually deleting snapshot
   - Including additional data in snapshots
4. Configuring advanced GIR features
   - Configuring the delay for the device to enter maintenance mode
   - Specifying a delay for after_maintenance snapshot creation

Configuring a custom profile

About this task

By default, GIR performs isolate switching to change the device from normal mode to maintenance mode. To customize the commands to execute and the interfaces to shut down for a profile-based mode switching, perform this task.

A custom profile contains a series of user-specified commands to be executed in sequence upon a mode switching.

With GIR, you can configure one switch-to-maintenance profile and one switch-to-normal profile for changing to maintenance mode and normal mode, respectively.

For more information about mode switching with GIR, see "How GIR works."

Restrictions and guidelines

CAUTION:

When you configure a custom profile, make sure you understand the impact of each command on the device and enter the commands in their correct form. Misconfiguration might cause traffic switchover failure.

When you edit a customer profile, use the following guidelines:

- You can only append commands to the end of a custom profile. You cannot remove or modify commands in a custom profile or insert a command between commands in the profile, except that you can edit the sleep instance settings.

  As a best practice, use the following steps to configure a custom profile:

  a. Add the series of commands to be executed to a text file.

  b. Verify that the commands are correct in form and order.

c. Copy the commands to the custom profile in custom profile view.

To modify a custom profile:

d. Edit the command file.

e. Delete the custom profile.

f. Copy the commands to the custom profile in custom profile view.

- The system executes the commands in a custom profile in their configuration order. If one command depends on another, enter the prerequisite command first.
- Do not add commands in user view to custom profiles. The system can execute only commands in system view or subviews of system view.
- By default, the system executes the commands in a custom profile in system view. To execute a command in a subview of system view, you must add the command used to access that subview before you add the command to execute in that subview.
- If it takes some time for a command to take effect after it is executed, you can add the **`sleep instance`** command next to it to introduce a delay. The system will wait for the delay to expire before it executes subsequent commands. If you are not sure whether a command requires a delay, add the **`sleep instance 0`** command next to it. Then, you can change the value after you determine that a delay is required.
- If you have inadvertently entered an incorrect command, you can add its **undo** form to remove it as long as that command is reversible by simply adding the **undo** keyword before it. Then, the system will automatically delete that command and its **undo** form from the profile. You cannot use this method to remove a command if it does not have an **undo** form or if the command and its **undo** form is not symmetric.
- The system provides the following mechanisms for profile optimization:
  - **Deduplication mechanism**—If you enter the same command line multiple times, the system automatically removes the duplicate commands. You do not need to do that manually.
  - **Command grouping**—The system groups commands in the same view together and then executes them in configuration order.
- Verify the settings in the custom profiles by executing the **`display gir custom-profile`** command. If you introduce irreversible misconfiguration items in a custom profile, delete and then reconfigure that profile.
- If you need to slightly optimize the series of commands used when the system changes to maintenance or normal mode, use the following method:

  g. Make sure the switch-to-maintenance or switch-to-normal custom profile is empty.

  h. Execute an isolate or shutdown mode switching.

  The system automatically populates the custom profile with the series of commands executed on mode switching.

  i. Edit the custom profile.
- The storage space for a custom profile is limited. When the storage space is full, no commands can be entered in the custom profile. As a best practice, do not enter irrelevant commands in the custom profile.

Configuring the switch-to-maintenance profile

1. Enter system view.

   `system-view`
2. Enter the view of the switch-to-maintenance profile.

   `gir custom-profile maintenance-mode`
3. Enter isolation commands.

   For isolation commands, see the command reference for each service module.

4. End the configuration and quit the view.
   config-end

Configuring the switch-to-normal profile

1. Enter system view.
   system-view
2. Enter the view of the switch-to-normal profile.
   gir custom-profile normal-mode
3. Enter the commands to remove the isolation.
   For information about the commands, see the command reference for each service module.
4. End the configuration and quit the view.
   config-end

Switching to maintenance mode

About this task

Perform this task before you maintain or upgrade the device.

The device will return to normal mode if a service module fails in switchover. The device will display the commands that failed to be executed or generate logs, but it will not roll back the commands that have been executed.

Restrictions and guidelines

CAUTION:

To maintain configuration consistency for a feature across mode switching, do not modify its configuration after you place the device in maintenance mode. To check for configuration changes, execute the **display gir snapshot compare** command after you revert the device back to normal mode.

Prerequisites

If the device is performing a snapshot operation, you must wait for the operation to end before you can change the GIR mode. Conversely, you cannot perform a snapshot operation while the device is changing the GIR mode.

Procedure

1. Verify that the system is stable.
   display system stable state
   The system is stable if the **System State** field displays **Stable**. For a successful GIR operation, you must make sure the system is stable.
2. Enter system view.
   system-view
3. Switch to maintenance mode.
   gir system-mode maintenance [**shutdown** | **custom-profile**] [**non-interactive**]

CAUTION:

- Packet loss might occur during mode switching.
- Do not restart the device or a process while GIR is switching mode. The system cannot roll back

settings issued during mode switching. If you restart the device or a disruptive process, mode switching will be interrupted, with only some of the commands issued.

4. If a service module fails in switchover, use the following methods to identify and remove the issue:
   - Remove the issue based on the error message.
   - Compare the snapshots to identify configuration inconsistencies that might have caused the issue.

     Identify snapshot information:

     display gir snapshot

     Compare the snapshots:

     display gir snapshot compare *snapshot1-name snapshot2-name* [**verbose**]
5. Verify that no traffic traverses the device before you maintain or upgrade the device.

   display counters

   This command is available in any view. This step prevents the packet loss that might occur because a protocol is slow or fails in switchover.

Switching back to normal mode

Manually switching back to normal mode

About this task

Perform this task after you maintain or upgrade the device.

The device will return to maintenance mode if a service module fails in switchover. The device will display the commands that failed to be executed or generate logs, but it will not roll back the commands that have been executed.

Prerequisites

If the device is performing a snapshot operation, you must wait for the operation to end before you can change the GIR mode. Conversely, you cannot perform a snapshot operation while the device is changing the GIR mode.

Before you change the M-LAG system in an M-LAG environment back to normal mode, execute the **display m-lag mad verbose** command to verify that no network interfaces are in M-LAG MAD DOWN state. For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

Table 1 Verify that the system is stable.

display system stable state

The system is stable if the **System State** field displays **Stable**.

1. Enter system view.

   system-view
2. Switch back to normal mode.

   undo gir system-mode maintenance [**custom-profile**] [**non-interactive**]

CAUTION:

Do not restart the device or a process while GIR is switching mode. The system cannot roll back settings issued during mode switching. If you restart the device or a disruptive process, mode switching will be interrupted, with only some of the commands issued.

3. If a service module fails in switchover, use the following methods to identify and remove the issue:
   - Remove the issue based on the error message.
   - Compare the snapshots to identify configuration inconsistencies that might cause the issue.

     Identify the snapshot names:

     display gir snapshot

     Compare the snapshots:

     display gir snapshot compare *snapshot1-name snapshot2-name* [**verbose**]

Configuring GIR to automatically switch back to normal mode

About this task

Automatic reverse switching enables the system to automatically revert back to normal mode from maintenance mode when the specified revert-delay expires.

This feature is helpful when the window of maintenance is deterministic.

Procedure

1. Enter system view.

   system-view
2. Specify the delay for the system to automatically revert to normal mode.

   gir system-mode maintenance revert-delay *value*

   By default, the system does not automatically revert to normal mode.

Managing GIR snapshots

Manually creating a snapshot

About this task

A GIR snapshot records the running data of the device in NETCONF format at a point in time for system restoration or comparison of data to identify changes over time.

Restrictions and guidelines

If the device is performing a mode switching or snapshot operation, you must wait for the operation to end before you can create a snapshot.

Procedure

1. Enter system view.

   system-view
2. Create a snapshot.

   gir snapshot create *snapshot-name description*

   The snapshot name can contain only digits, letters, and underscores (_).

Manually deleting snapshots

About this task

Perform this task to delete earliest snapshots to free up memory space.

Restrictions and guidelines

If the device is performing a mode switching or snapshot operation, you must wait for the operation to end before you can delete a snapshot.

Procedure

1. Enter system view.
   system-view
2. Delete snapshots.
   gir snapshot delete { **all** | *snapshot-name* }

Including additional data in snapshots

About this task

By default, the system does not include routing information in created snapshots, because the routing table is large in size. Perform this task to include IPv4 or IPv6 routes in created snapshots.

Procedure

1. Enter system view.
   system-view
2. Include IPv4 or IPv6 routes in snapshots.
   gir snapshot add { **ipv4routes** | **ipv6routes** }*
   By default, the system does not include routing information in created snapshots.

Configuring advanced GIR features

Configuring the delay for the device to enter maintenance mode

About this task

Execution of some commands might take some time. To ensure that all GIR applicable service modules complete traffic path switchover before you maintain the device, GIR introduces a maintenance delay for the device to enter maintenance mode.

Increase the maintenance delay if its default value is not sufficient.

Procedure

1. Enter system view.
   system-view
2. Set the maintenance delay.
   gir system-mode maintenance maintenance-delay *value*
   By default, the maintenance delay is 120 seconds.

Specifying a delay for after_maintenance snapshot creation

About this task

If the default snapshot delay value is not sufficient for a service module to restore its services, increase the delay so that GIR can take an **after_maintenance** snapshot of complete and accurate data.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Specify a delay for **after_maintenance** snapshot creation.
   ```
   gir system-mode maintenance snapshot-delay value
   ```
   By default, the delay for creation of **after_maintenance** snapshots is 120 seconds.

Display and maintenance commands for GIR

Execute **display** commands in any view.

Task	Command
Display all GIR snapshots on the device.	**display gir snapshot**
Compare GIR snapshots to view their differences.	**display gir snapshot compare** *snapshot1-name snapshot2-name* [**verbose**]
Display the current GIR mode.	**display gir system-mode**
Display the GIR custom profiles.	**display gir custom-profile** [**maintenance-mode** | **normal-mode**]

GIR configuration examples

Example: Configuring GIR in an ECMP-based node redundancy environment

Network configuration

As shown in Figure 1, Device A and Device B have ECMP routes for traffic between Internet and Device C. Before upgrading Device A, use GIR to place it in maintenance mode to prevent packet losses during upgrade. Then, change Device A back to normal mode after its software is upgraded.

Figure 1 Network diagram

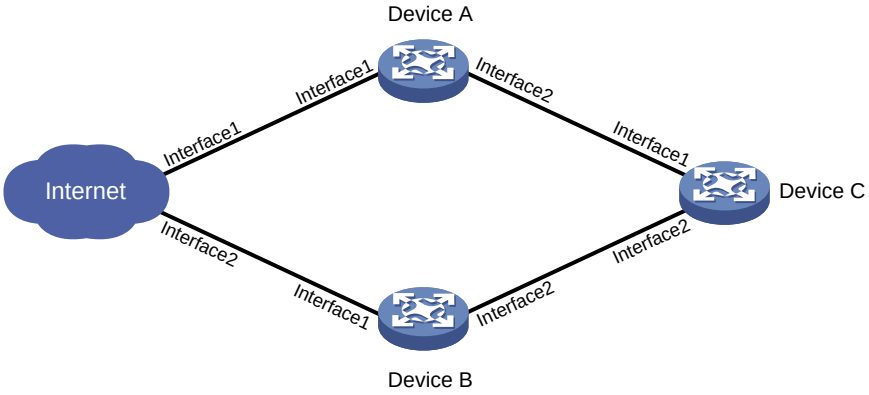


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure IP addresses and a unicast routing protocol (for example, OSPF). Make sure ECMP routes are available for traffic between the Internet and Device C to traverse Device A or Device B to reach the destination. (Details not shown.)
2. Isolate Device A from the network before you upgrade it.

   # Change Device A from the normal mode to the maintenance mode.

```
<Device A> display system stable state summary
System state      : Stable
Redundancy state  : No redundancy
NSR     state     : No standby
<Device A> system-view
[Device A] gir system-mode maintenance
Collecting commands... Please wait.
Configuration to be applied:
   ospf 1
     isolate enable
Do you want to continue? [Y/N]: y
Generated a snapshot: before_maintenance.
Applying: ospf 1
Applying: isolate enable
Waiting 120 seconds to release the CLI.
Changed to maintenance mode successfully.
```

   # Verify that the mode switching is successful on Device A.

```
[Device A] display gir system-mode
System Mode: Maintenance
Revert-delay time: 0min
Maintenance-delay time: 120s
Snapshot-delay time: 120s
```

   # Verify that traffic switching has completed on Device A.

```
[Device A] display counters rate inbound interface
Usage: Bandwidth utilization in percentage
Interface            Usage (%)   Total (pps)   Broadcast (pps)   Multicast (pps)
GE1/0/1                      0             0                --                --
GE1/0/2                      0             0                --                --

 Overflow: More than 14 digits.
       --: Not supported.
```

3. Upgrade Device A after saving the configuration. (Details not shown.)
4. Restore forwarding services on Device A:

   # Change Device A from the maintenance mode back to the normal mode after it is upgraded and rebooted.

```
<Device A> system-view
[Device A] undo gir system-mode maintenance
Collecting commands... Please wait.
Configuration to be applied:
   ospf 1
     undo isolate enable
Do you want to continue? [Y/N]: y
Applying: ospf 1
Applying: undo isolate enable
Waiting 120 seconds to generate a snapshot.
Generated a snapshot: after_maintenance.
Changed to normal mode successfully.
```

   # Verify that the mode switching is successful on Device A.

```
[Device A] display gir system-mode
System Mode: Normal
Revert-delay time: 0min
Maintenance-delay time: 120s
Snapshot-delay time: 120s
```

   # Verify that traffic switching has completed on Device A.

```
[Device A] display counters rate inbound interface
Usage: Bandwidth utilization in percentage
Interface            Usage (%)   Total (pps)   Broadcast (pps)   Multicast (pps)
GE1/0/1                    100           100                --                --
GE1/0/2                    100           100                --                --

 Overflow: More than 14 digits.
       --: Not supported.
```

Example: Configuring GIR on an M-LAG system

Network configuration

As shown in Figure 2, Device A and Device B form an M-LAG system.

Use the following method to upgrade the M-LAG system to prevent packet loss during the upgrade:

1. Upgrade Device B.
2. Perform a forward switching operation by using the isolate method on Device A.

3. Upgrade Device A.

Figure 2 Network diagram

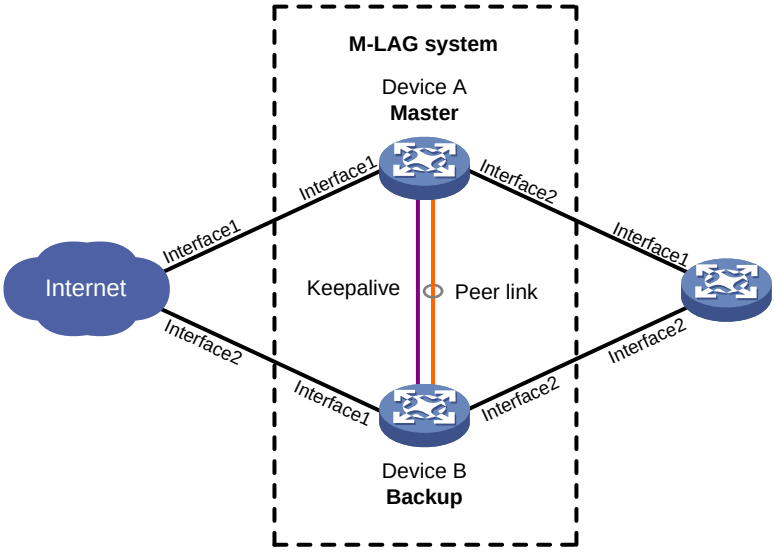


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Upgrade Device B. (Details not shown.)
2. Isolate Device A from the network before upgrade it.

   # Change Device A from the normal mode to the maintenance mode.

```
<Device A> display system stable state summary
System state      : Stable
Redundancy state  : No redundancy
NSR     state     : No standby
<Device A> system-view
[Device A] gir system-mode maintenance
Collecting commands... Please wait.
Configuration to be applied:
   ospf 1
     isolate enable
   sleep instance 1 interval 30
   link-aggregation lacp isolate
Do you want to continue? [Y/N]: y
Generated a snapshot: before_maintenance.
Applying: ospf 1
Applying: isolate enable
Applying: sleep instance 1 interval 30
Applying: link-aggregation lacp isolate
Waiting 120 seconds to release the CLI.
```

```
Changed to maintenance mode successfully.
```

Verify that the mode switching is successful on Device A.

```
[Device A] display gir system-mode
System Mode: Maintenance
Revert-delay time: 0min
Maintenance-delay time: 120s
Snapshot-delay time: 120s
```

Verify that the traffic switching has completed on Device A.

```
[Device A] display counters rate inbound interface
Usage: Bandwidth utilization in percentage
Interface            Usage (%)   Total (pps)   Broadcast (pps)   Multicast (pps)
GE1/0/1                      0             0                --                --
GE1/0/2                      0             0                --                --
GE1/0/3                      0             0                --                --

 Overflow: More than 14 digits.
       --: Not supported
```

3. Upgrade Device A after saving the configuration. (Details not shown.)
4. Restore forwarding services on Device A.

Change Device A from the maintenance mode back to the normal mode after it is upgraded and rebooted.

```
<Device A> system-view
[Device A] undo gir system-mode maintenance
Collecting commands... Please wait.
Configuration to be applied:
   undo link-aggregation lacp isolate
sleep instance 1 interval 30
   ospf 1
     undo isolate enable
Do you want to continue? [Y/N]: y
Applying: undo link-aggregation lacp isolate
Applying: sleep instance 1 interval 30
Applying: ospf 1
Applying: undo isolate enable
Waiting 120 seconds to generate a snapshot.
Generated a snapshot: after_maintenance.
Changed to normal mode successfully.
```

Verify that the mode switching is successful on Device A.

```
[Device A] display gir system-mode
System Mode: Normal
Revert-delay time: 0min
Maintenance-delay time: 120s
Snapshot-delay time: 120s
```

Verify that the traffic switching has completed on Device A.

```
[Device A] display counters rate inbound interface
Usage: Bandwidth utilization in percentage
Interface            Usage (%)   Total (pps)   Broadcast (pps)   Multicast (pps)
GE1/0/1                    100           100                --                --
```

GE1/0/2	100	100	--	--
GE1/0/3	100	100	--	--

Overflow: More than 14 digits.
--: Not supported.

Contents

Using automatic configuration 1

About automatic configuration 1
Using server-based automatic configuration 1
About server-based automatic configuration 1
Typical server-based automatic configuration network 1
Host name file 2
Selecting the interface used for automatic configuration 2
Operating mechanism 3
Server-based automatic configuration tasks at a glance 3
Preparing configuration files 4
Preparing script files 4
Configuring the file server 5
Configuring the DHCP server 5
Configuring the DNS server 8
Configuring the gateway 9
Starting and completing automatic configuration 9
Stopping automatic configuration 9
Using USB-based automatic configuration 10
About USB-based automatic configuration 10
Configuring and using index file-based automatic configuration 10
Configuring and using automatic configuration without an index file 16
Server-based automatic configuration examples 18
Example: Using a TFTP server for automatic configuration 18
Example: Using an HTTP server and Tcl scripts for automatic configuration 24
Example: Using an HTTP server and Python scripts for automatic configuration 26
Example: Using an SFTP server for automatic configuration 28
Example: Using an FTP server for automatic configuration 30
Example: Setting up an IRF fabric 32

Using automatic configuration

About automatic configuration

When the device starts up without a valid next-startup configuration file, the device searches the root directory of its default file system for the autocfg.py, autocfg.tcl, and autocfg.cfg files. Only one of files might exist in the root directory. If any one of the files exists, the device loads the file. If none of the files exists, the device uses the automatic configuration feature to obtain a set of configuration settings.

With the automatic configuration feature, the device can automatically obtain a set of configuration settings at startup. This feature simplifies network configuration and maintenance.

Automatic configuration can be implemented by using the implementation methods in Table 1.

Table 1 Automatic configuration implementation methods

Implementation method	Configuration file location	Application scenarios
Server-based automatic configuration	File server	A number of geographically distributed devices need to be configured.
USB-based automatic configuration	USB disk	• On a small network, the devices reside near to each other, and no host can be used as a file server. • On a large network, only a few devices require automatic configuration or configuration update.

If both server-based automatic configuration and USB-based automatic configuration are available, the device prefers USB-based automatic configuration.

Using server-based automatic configuration

About server-based automatic configuration

With server-based automatic configuration, a device without a configuration file can run the DHCP client to obtain a configuration file from a file server at startup.

You can deploy server-based automatic configuration on both IPv4 and IPv6 networks by using the same method. This chapter describes the tasks for deploying server-based automatic configuration on an IPv4 network.

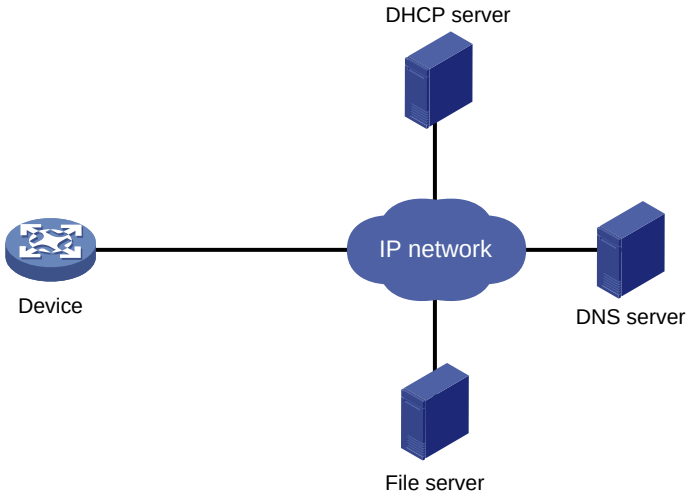
Typical server-based automatic configuration network

As shown in Figure 1, a typical server-based automatic configuration network consists of the following servers:

- **DHCP server**—Allocates an IP address to the device that performs server-based automatic configuration and notifies the device of the method to obtain the configuration file or script file for automatic configuration.
- **File server**—TFTP, HTTP, SFTP, or FTP server. The file server stores the configuration file or script file for automatic configuration.

- (Optional.) **DNS server**—Obtains the configuration file name or the IP address of the file server for the device.

Figure 1 Server-based automatic configuration network diagram



Host name file

A host name file stores mappings between host IP addresses and host names. To configure a host name file:

1. Create a host name file. The file name must be **network.cfg**.
2. Add each mapping entry in the **ip host** *host-name ip-address* format on a separate line. For example:

```
ip host host1 101.101.101.101
ip host host2 101.101.101.102
ip host client1 101.101.101.103
ip host client2 101.101.101.104
```

IMPORTANT:

The host name for a device must be the same as the name of the configuration file configured for the device. For example, the name of the configuration file for the device at 101.101.101.101 is **host1.cfg**.

Selecting the interface used for automatic configuration

The device uses the following steps to select the interface for automatic configuration:

1. Identifies the status of the management Ethernet interface at Layer 2. If the management Ethernet interface is up, the device uses that interface.
2. Identifies the status of Layer 2 Ethernet interfaces. If one or more Layer 2 Ethernet interfaces are in up state, the device uses the VLAN interface of the default VLAN.
3. Sorts all Layer 3 Ethernet interfaces in up state first in lexicographical order of interface types and then in ascending order of interface numbers. Uses the interface with the smallest interface number among the interfaces of the first interface type.

4. If no Layer 3 Ethernet interfaces are in up state, the device waits 30 seconds and goes to step 1 to try again.

For fast automatic device configuration, connect only the management Ethernet interface on each device to the network.

Operating mechanism

The process for server-based automatic configuration for a device is as follows:

1. The device selects the interface used to obtain the configuration file or script file for automatic configuration and sends an IP request message to the DHCP server. For more information about the interface selection order, see "Selecting the interface used for automatic configuration."
2. After the DHCP server receives the IP request message, it sends a reply message to the device. If the device does not receive any reply message within the timeout period of the request message, it selects the next interface according to the interface selection order and sends the IP request message again.
3. The device parses the reply message.
   - Obtains the IP address included in the reply message and uses it as the IP address for the selected interface to connect to the DHCP server. If the interface already has an IP address assigned, the device replaces the original IP address with the one included in the reply message. Once the automatic configuration process is complete, the device restores the original IP address to the interface.
   - Obtains the protocol information (HTTP, TFTP, SFTP, or FTP), configuration file name information, and file server address included in the reply message.
4. If no configuration file name is obtained or the obtained configuration file name does not meet the requirements, the device uses file **network.cfg** on the file server. If file **network.cfg** does not exist, the device uses file **device.cfg** on the file server. If file **device.cfg** does not exist, the device fails to obtain the configuration file or script file. In this case, the process returns to step 1.
5. The device resolves the domain name of the file server. If the file server address included in the reply message is a domain name rather than an IP address, the device uses the DNS server to resolve the domain name to an IP address.
6. The device downloads the configuration file or script file used for automatic configuration from the file server based on the obtained protocol, file server address, and configuration file name.
7. The device deploys the configuration in the configuration file or executes the script file.
8. The device automatically deletes the configuration file or script file.

Server-based automatic configuration tasks at a glance

To configure server-based automatic configuration, perform the following tasks:

1. Prepare the files for automatic configuration:
   - Preparing configuration files
   - Preparing script files
2. Configuring the file server
3. Configuring the DHCP server
4. (Optional.) Configuring the DNS server
5. (Optional.) Configuring the gateway
6. Starting and completing automatic configuration
7. (Optional.) Stopping automatic configuration

Preparing configuration files

Configuration file types

The device supports the configuration file types listed in Table 2.

Table 2 Configuration file types

Configuration file type	Application objects	File name requirements	Supported file server types
Dedicated configuration file	Devices that require different settings	*File name*.**cfg** For simple file name identification, use configuration file names that do not contain spaces.	• HTTP server • TFTP server • SFTP server • FTP server
Common configuration file	Devices that share all or some settings	*File name*.**cfg** For simple file name identification, use configuration file names that do not contain spaces.	• HTTP server • TFTP server • SFTP server • FTP server
Default configuration file	All devices. The file contains only common configurations that devices use to start up.	**device.cfg**	TFTP server

Identifying requirements for and preparing configuration files

1. Identify the requirements of the devices for configuration files.
2. For devices that require different configurations, prepare a configuration file for each of them and save the files to the file server.
3. For devices that share all or some configurations, save the common configurations to a .cfg file on the file server.
4. If a TFTP file server is used, you can save the common configurations that devices use to start up to the **device.cfg** file on the server. The file is assigned to a device only when the device does not have any other configuration file to use.

Preparing a host name file on the file server

If the DHCP server does not assign configuration file names, you can configure a host name file on the file server. The host name file contains the host name-IP address mappings of the devices to be automatically configured.

Preparing script files

About this task

Script files can be used for automatic software upgrade and automatic configuration.

The devices support Python scripts (.py files) and Tcl scripts (.tcl files). For more information about Python and Tcl scripts, see "Using Python" and "Using Tcl."

The devices support dedicated script files and common dedicated script files. They do not support using a default script file. For information about dedicated script files and common dedicated script files, see Table 2.

When script files are used, you cannot use a host name file to provide the host name-IP address mappings for devices.

Restrictions and guidelines

To use a Tcl script, make sure all commands in the script are supported and correctly configured. Any error in a command causes the automatic configuration process to quit.

When you use a Python script to automatically configure a device, make sure the script file does not contain command line errors (such as spelling mistakes, incorrect views, and unsupported commands for the device). If the script file contains command line errors, the current round of automatic configuration will fail and the device will proceed to the next round of automatic configuration.

When you use a Python script to automatically configure a device, the device is in the startup phase and the device is unstable. Therefore, do not use the **Stable** value in the **`display system stable state`** command output as the judgment condition in the Python script.

Procedure

- For devices that share all or some configurations, create a script file that contains the common configurations.
- For the other devices, create a separate script file for each of them.

Configuring the file server

For devices to obtain configuration information from an HTTP, TFTP, SFTP, or FTP server, start HTTP, TFTP, SFTP, or FTP service on the file server, respectively.

Configuring the DHCP server

About this task

The DHCP server assigns the following items to devices that need to be automatically configured:

- IP addresses.
- Paths of the configuration or script files.

When you configure the DHCP server, follow these guidelines:

- For devices for which you have prepared different configuration files, perform the following tasks for each of the devices on the DHCP server:
  - Create a DHCP address pool.
  - Configure a static address binding.
  - Specify a configuration file or script file.

  Because an address pool can use only one configuration file, you can specify only one static address binding for an address pool.
- For devices for which you have prepared the same configuration file, use either of the following methods:
  - Method 1:
    - Create a DHCP address pool for the devices.
    - Configure a static address binding for each of the devices in the address pool.
    - Specify the configuration file for the devices.
  - Method 2:
    - Create a DHCP address pool for the devices.
    - Specify the subnet for dynamic allocation.
    - Specify the TFTP server.
    - Specify the configuration file for the devices.

- If all devices on a subnet share the same configuration file or script file, perform the following tasks on the DHCP server:
  - Configure dynamic address allocation.
  - Specify the configuration file or script file for the devices.

  The configuration file can contain only the common settings for the devices. You can provide a method for the device administrators to change the configurations after their devices start up.

Restrictions and guidelines

When you specify a startup configuration file on a remote server for a DHCP client, follow these restrictions and guidelines:

- If a TFTP server is used, you must specify the file name. If the file name contains special characters, you must transcode the special characters according to the RFC specification.
- If an HTTP server is used, you must specify the URL of the file in the format of **http://***HTTP-server-IP-address***:***port-number***/***path***/***file-name*. If the path or file name in the URL contains special characters, you must transcode the special characters according to the RFC specification. If you do not specify a port number, the default port number applies.
- If an FTP server is used, you must specify the URL of the file in the format of **ftp://***username***:***password@FTP-server-IP-address***:***port-number***/***path***/***file-name*. If the username, password, path, or file name in the URL contains special characters, you must transcode the special characters according to the RFC specification. If you do not specify a port number, the default port number applies. If you do not specify a path, the root directory of the server applies.
- If an SFTP server is used, you must specify the URL of the file in the format of **sftp://***username***:***password@SFTP-server-IP-address***:***port-number***/***path***/***file-name*. If the username, password, path, or file name in the URL contains special characters, you must transcode the special characters according to the RFC specification. If you do not specify a port number, the default port number applies. If you do not specify a path, the root directory of the server applies.

Configuring the DHCP server when an HTTP file server is used

1. Enter system view.

   system-view
2. Enable DHCP.

   dhcp enable

   By default, DHCP is disabled.
3. Create a DHCP address pool and enter its view.

   dhcp server ip-pool *pool-name*
4. Configure the address pool.

   Choose the options to configure as needed:
   - Specify the primary subnet for the address pool.

     network *network-address* [*mask-length* | **mask** *mask*]

     By default, no primary subnet is specified.
   - Configure a static binding.

     static-bind ip-address *ip-address* [*mask-length* | **mask** *mask*] { **client-identifier** *client-identifier* | **hardware-address** *hardware-address* [**ethernet** | **token-ring**] }

     By default, no static binding is configured.

     You can configure multiple static bindings. However, one IP address can be bound to only one client. To change the binding for a DHCP client, you must remove the binding and reconfigure a binding.

5. Specify the URL of the configuration or script file in HTTP format.

   bootfile-name *url*

   By default, no configuration or script file URL is specified in HTTP format.

Configuring the DHCP server when a TFTP file server is used

1. Enter system view.

   system-view

2. Enable DHCP.

   dhcp enable

   By default, DHCP is disabled.

3. Create a DHCP address pool and enter its view.

   dhcp server ip-pool *pool-name*

4. Configure the address pool.

   Choose the options to configure as needed:

   - Specify the primary subnet for the address pool.

     network *network-address* [*mask-length* | **mask** *mask*]

     By default, no primary subnet is specified.

   - Configure a static binding.

     static-bind ip-address *ip-address* [*mask-length* | **mask** *mask*] { **client-identifier** *client-identifier* | **hardware-address** *hardware-address* [**ethernet** | **token-ring**] }

     By default, no static binding is configured.

     You can configure multiple static bindings. However, one IP address can be bound to only one client. To change the binding for a DHCP client, you must remove the binding and reconfigure a binding.

5. Specify a TFTP server.

   Choose one option as needed:

   - Specify the IP address of the TFTP server.

     tftp-server ip-address *ip-address*

     By default, no TFTP server IP address is specified.

   - Specify the name of the TFTP server.

     tftp-server domain-name *domain-name*

     By default, no TFTP server name is specified.

     If you specify a TFTP server by its name, a DNS server is required on the network.

6. Specify the name of the configuration or script file.

   bootfile-name *bootfile-name*

   By default, no configuration or script file name is specified.

Configuring the DHCP server when an SFTP file server is used

1. Enter system view.

   system-view

2. Enable DHCP.

   dhcp enable

   By default, DHCP is disabled.

3. Create a DHCP address pool and enter its view.

   dhcp server ip-pool *pool-name*

4. Configure the address pool.

   Choose the options to configure as needed:

   - Specify the primary subnet for the address pool.

     network *network-address* [*mask-length* | **mask** *mask*]

     By default, no primary subnet is specified.

   - Configure a static binding.

     static-bind ip-address *ip-address* [*mask-length* | **mask** *mask*] { **client-identifier** *client-identifier* | **hardware-address** *hardware-address* [**ethernet** | **token-ring**] }

     By default, no static binding is configured.

     You can configure multiple static bindings. However, one IP address can be bound to only one client. To change the binding for a DHCP client, you must remove the binding and reconfigure a binding.

5. Specify the URL of the configuration or script file in SFTP format.

   bootfile-name *url*

   By default, no configuration or script file URL in SFTP format is specified.

Configuring the DHCP server when an FTP file server is used

1. Enter system view.

   system-view

2. Enable DHCP.

   dhcp enable

   By default, DHCP is disabled.

3. Create a DHCP address pool and enter its view.

   dhcp server ip-pool *pool-name*

4. Configure the address pool.

   Choose the options to configure as needed:

   - Specify the primary subnet for the address pool.

     network *network-address* [*mask-length* | **mask** *mask*]

     By default, no primary subnet is specified.

   - Configure a static binding.

     static-bind ip-address *ip-address* [*mask-length* | **mask** *mask*] { **client-identifier** *client-identifier* | **hardware-address** *hardware-address* [**ethernet** | **token-ring**] }

     By default, no static binding is configured.

     You can configure multiple static bindings. However, one IP address can be bound to only one client. To change the binding for a DHCP client, you must remove the binding and reconfigure a binding.

5. Specify the URL of the configuration or script file in FTP format.

   bootfile-name *url*

   By default, no configuration or script file URL in FTP format is specified.

Configuring the DNS server

A DNS server is required in the following situations:

- **Obtaining the names of configuration files for automatic configuration when the TFTP server does not have a host name file**—In this situation, the devices that perform

server-based automatic configuration use the DNS server to resolve their IP addresses to their host names. Then, each device can use its host name to obtain the configuration file that has the same name as its host name from the TFTP server.

- **Obtaining the IP address of a file server**—If the devices obtain the domain name of the file server from DHCP reply messages, they can use the domain name to obtain the IP address of the file server through the DNS server.

For more information about DNS servers, see DNS configuration in *Layer 3—IP Services Configuration Guide*.

Configuring the gateway

If the devices to be automatically configured and the servers for automatic configuration reside in different network segments, you must perform the following tasks:

- Deploy a gateway and make sure the devices can communicate with the servers.
- Configure the DHCP relay agent feature on the gateway.
- Configure the UDP helper feature on the gateway.

  This task is required if devices send requests to a TFTP server by using broadcast packets. A device uses broadcast packets to send requests to a TFTP server in the following situations:

  - The DHCP reply does not contain the IP address or domain name of the TFTP server.
  - The IP address or domain name of the TFTP server is invalid.

  The UDP helper transforms a broadcast packet into a unicast packet and forwards the unicast packet to the file server. For more information about UDP helper, see *Layer 3—IP Services Configuration Guide*.

Starting and completing automatic configuration

1. Power on the devices to be automatically configured.
2. The devices enter the automatic configuration process. After a device obtains a configuration file and executes the file successfully, the automatic configuration process ends.
3. Save the running configuration.

   save

   As a best practice, save the running configuration on each device after the configuration file is executed successfully on the device.

   For more information about the **save** command, see configuration file management commands in *Fundamentals Command Reference*.

Stopping automatic configuration

If a device cannot obtain a configuration file for automatic configuration, the current automatic configuration attempt fails, and the device continues attempting automatic configuration. You can wait until the device automatically ends the automatic configuration process after it has made the maximum number of attempts. Alternatively, you can manually stop the automatic configuration process by using shortcut keys **Ctrl+C** or **Ctrl+D** based on the prompt information. After the automatic configuration process fails and ends, the device starts up with initial settings.

Using USB-based automatic configuration

About USB-based automatic configuration

When a device enters the automatic configuration process after it starts up with initial settings, it first attempts to load configuration from a USB disk. USB-based automatic configuration supports the following methods:

- Index file-based automatic configuration.
- Automatic configuration without an index file.

A device prefers index file-based automatic configuration to automatic configuration without an index file.

Configuring and using index file-based automatic configuration

Operating mechanism

Save an index file and configuration file on a USB disk. When a device starts, it automatically obtains and executes the configuration file on the USB disk based on the index file for automatic configuration.

The process for index file-based automatic configuration for a device is as follows:

1. The device detects that a USB disk is present.
2. Check whether the index file named **smart_config.ini** exists on the USB disk.
   - If the index file named **smart_config.ini** exists, the device enters step 3.
   - If the index file named **smart_config.ini** does not exist, the device performs automatic configuration without an index file. For more information, see "Configuring and using automatic configuration without an index file."
3. Check whether the format of the **smart_config.ini** index file is valid.
   - If the index file is valid, the device enters step 4.
   - If the index file is invalid, USB-based automatic configuration fails and the automatic configuration process ends. An error report is generated on the USB disk.
4. The device compares the data change time flag specified in the **smart_config.ini** index file with the time flag of the last USB-based automatic configuration recorded on the device.
   - If they are different, the device enters step 5.
   - If they are the same, USB-based automatic configuration fails and the automatic configuration process ends. An error report is generated on the USB disk.
5. The device obtains the configuration file from the USB disk according to the description information in the **smart_config.ini** index file and saves the configuration file to the default storage medium.
   - If the device obtains the configuration file successfully, it enters step 6.
   - If the device fails to obtain the configuration file, USB-based automatic configuration fails and the automatic configuration process ends. An error report is generated on the USB disk.
6. The device determines whether a reboot activation is required based on the obtained configuration file or the specified activation method.
   - If a reboot is not required, the device activates the files specified in the index file online and proceeds to step 7.
   - If a reboot is required, the device sets the files specified in the index file as the next-startup files. After a successful automatic reboot, the device proceeds to step 7.

7. After automatic configuration succeeds, the device uses the configuration in the configuration file obtained from the USB disk to overwrite the running configuration and sets the configuration as the next-startup configuration. The automatic configuration process ends.
8. Remove the USB disk from the device.

Prerequisites

Create an index file and save it to the root directory of the USB disk. Save the required files to the directory specified in the index file on the USB disk.

Creating an index file

The name of the index file must be **smart_config.ini**.

To use index file-based automatic configuration, you can edit the index file for USB-based automatic configuration on your PC as follows:

1. Create an empty .txt file.
2. Edit the file contents according to the format of the index file for USB-based automatic configuration.
3. Save the .txt file as **smart_config.ini**.
4. Copy the index file named **smart_config.ini** to the root directory of the USB disk.

The index file format is as follows:

```
BEGIN LSW
[GLOBAL CONFIG]
TIMESN=
AUTOOVERWRITEFILE=
AUTODELFILE=
ACTIVEMODE=
[DEVICE DESCRIPTION]
OPTION=
SN=
MAC=
AUTODELFILE=
ACTIVEMODE=
DEVICETYPE=
DIRECTORY=
SYSTEM-IPE=
HMAC=
SYSTEM-BOOT-BIN=
HMAC=
SYSTEM-SYSTEM-BIN=
HMAC=
SYSTEM-FEATURE-BIN=
HMAC=
SYSTEM-FEATURE-BIN=
HMAC=
SYSTEM-FEATURE-BIN=
HMAC=
SYSTEM-PATCH-BIN=
HMAC=
SYSTEM-CONFIG=
HMAC=
```

```
END LSW
```

Table 3 shows the description for parameters in the index file.

Table 3 Description for parameters in the index file

Parameter	Description
BEGIN LSW	Required. Start flag. It cannot be modified.
[GLOBAL CONFIG]	Required. Global configuration start flag. It cannot be modified.
TIMESN	Required. Data change time flag, in string format. It can contain 1 to 16 characters, and cannot contain spaces. Recommended format: year month day.hour minute second. For example, you can set TIMESN=20110628.080910 to represent 08:09:10 28 June 2011. Each TIMESN value corresponds to a device that will be upgraded. During the USB-based automatic configuration process, a device records its own TIMESN value before it reboots. If a reboot is not required for the upgrade, the device records its own TIMESN value when the upgrade is complete. This TIMESN value cannot be used in the next upgrade. When some devices fail to be upgraded, you must change their TIMESN values before performing another USB-based automatic configuration.
AUTODELFILE	Optional. Specifies whether to automatically delete the old software package after upgrade. Options include: • **YES**—Delete. • **NO**—Not delete By default, the value for the AUTODELFILE parameter is NO. If this parameter does not exist or the value for this parameter is empty or invalid, the default setting applies. AUTODELFILE parameters include two types, global and device-specific. • The parameters in the [GLOBAL CONFIG] area are global, and the parameters in the [DEVICE DESCRIPTION] area are device-specific. • If this parameter is set to YES or NO for a specific device, the device-specific configuration takes effect. If this parameter is not set or is empty for a specific device, the global configuration for this parameter takes effect on the device.
AUTOOVERWRITEFILE	Optional. Specifies whether to overwrite a configuration file on a device when a configuration file with the same name is copied from a USB disk to the device. • **YES**—Overwrite. • **NO**—Not overwrite. By default, the value for the AUTOOVERWRITEFILE parameter is YES. If this parameter does not exist or the value for this parameter is empty or invalid, the default setting applies.
ACTIVEMODE	Optional. Specifies the file activation method after a file is copied. Options include: • **DEFAULT**—Activates a file by its own default activation method. The default activation method is rebooting devices for system software. The default activation method is online activation without rebooting devices for configuration files and patch files. • **REBOOT**—Activates a file by rebooting the device. By default, the value for the ACTIVEMODE parameter is DEFAULT. If this parameter does not exist or the value for this

Parameter	Description
	parameter is empty or invalid, the default setting applies. ACTIVEMODE parameters include two types, global and device-specific. • The parameters in the [GLOBAL CONFIG] area are global, and the parameters in the [DEVICE DESCRIPTION] area are device-specific. • If this parameter is set to DEFAULT or REBOOT for a specific device, the device-specific configuration takes effect. If this parameter is not set or is empty for a specific device, the global configuration for this parameter takes effect on the device.
[DEVICE DESCRIPTION]	Required. Start flag for a device-specific configuration file. Each parameter in a device-specific configuration file cannot be repeated. If a device-specific configuration file contains repeated parameters, no device can match it. Each device can match only one device-specific configuration file. After a device matches a device-specific configuration file, it stops matching the subsequent device-specific configuration files.
OPTION	Optional. Validity flag for a device-specific configuration file, which indicates whether the device configuration file is valid. Options include: • **OK**—Valid. • **NOK**—Invalid. All information in the device-specific configuration file will not be used for device matching. By default, the value for the OPTION parameter is OK. If this parameter does not exist or the value for this parameter is empty or invalid, the default setting applies.
SN	Optional. Device serial number. If the value for the SN parameter is DEFAULT, the device SN will not be matched. Otherwise, the device SN will be matched. By default, the value for the SN parameter is DEFAULT. If this parameter does not exist or the value for this parameter is empty, the default setting applies.
MAC	Optional. Device MAC address, in the format of XXXX-XXXX-XXXX, where X is a hexadecimal digit. If the value for this parameter is DEFAULT, the device MAC address will not be matched. Otherwise, the device MAC address will be matched. By default, the value for the MAC parameter is DEFAULT. If this parameter does not exist or the value for this parameter is empty, the default setting applies.
DEVICETYPE	Optional. Matches the device type, for example, S12700. If the value for the DEVICETYPE parameter is DEFAULT, the device type will not be matched. Otherwise, the device type will be matched. By default, the value for the DEVICETYPE parameter is DEFAULT. If this parameter does not exist or the value for this parameter is empty, the default setting applies.
DIRECTORY	Optional. Directory for saving the software package and configuration file on the USB disk. • When the value for this parameter is empty or this parameter does not exist, the software package and configuration file will be saved in the root directory of the USB disk. • When DIRECTORY is set to /abc, the software package and configuration file will be saved in the abc directory in the USB

Parameter	Description
	disk. By default, the value for the DIRECTORY parameter is empty. The file directory format in the index file must be the same as that of the file system in the device. • The specified directory can contain a maximum of four levels. A directory must start with a slash (/), and different levels are separated by using slashes (/). However, a directory cannot end with a slash (/). For example, /abc/test is a valid directory, and /abc/test/ is an invalid directory. • The string in each level of a directory is 1 to 15 characters. • A directory cannot contain spaces or the following characters: ~ * / \ : ' " < > | ? [ ] %. A directory is case insensitive.
SYSTEM-IPE	Optional. IPE file name, which ends with .ipe.
HMAC=	Optional. The HMAC verification code for the .ipe system software package. If the system software package is used for initial deployment and the index file for USB-based automatic configuration has a verification password, you must specify this field.
SYSTEM-BOOT-BIN	Optional. Boot image file name, which ends with .bin.
HMAC=	Optional. The HMAC verification code for the Boot image file. If the Boot image file is used for initial deployment and the index file for USB-based automatic configuration has a verification password, you must specify this field.
SYSTEM-SYSTEM-BIN	Optional. System image file name, which ends with .bin.
HMAC=	Optional. The HMAC verification code for the system image file. If the system image file is used for initial deployment and the index file for USB-based automatic configuration has a verification password, you must specify this field.
SYSTEM-FEATURE-BIN	Optional. Feature image file name, which ends with .bin.
HMAC=	Optional. The HMAC verification code for the feature image file. If the feature image file is used for initial deployment and the index file for USB-based automatic configuration has a verification password, you must specify this field.
SYSTEM-PATCH-BIN	Optional. Patch file name, which ends with .bin.
HMAC=	Optional. The HMAC verification code for the patch file. If the patch file is used for initial deployment and the index file for USB-based automatic configuration has a verification password, you must specify this field.
SYSTEM-CONFIG	Optional. Configuration file name, which ends with .cfg.
HMAC=	Optional. The HMAC verification code for the configuration file. If the configuration file is used for initial deployment and the index file for USB-based automatic configuration has a verification password, you must specify this field.
END LSW	Required. End flag of the file.

CAUTION:

When editing an index file for USB-based automatic configuration, make sure the contents in the index file are correct. The devices to be upgraded are matched in the order of the DEVICE parameters in the index file from top to bottom. A device is matched in the following order: MAC > SN > DEVICETYPE. Once a match is found, the files in the matching DEVICE information are

loaded. If an error occurs during this process, the whole USB-based automatic configuration process is terminated, and an error report is output. The configuration will not be rolled back for devices that have been upgraded successfully.

When you use an index file for automatic configuration, the device will confirm the files to be loaded according to the index file as follows:

5. The device identifies whether an .ipe file is specified in the index file. If yes, the device confirms loading the .ipe file, skips step 6, and proceeds with 7. If not, the device skips step 5 and proceeds with step 6.
6. The device identifies whether the index file has boot, system, and feature files specified. If yes, the specified files are loaded. The boot and system files are required, and the feature file is optional. After the files to be loaded are confirmed, the device proceeds with step 7. If no boot, system, or feature file is specified in the index file, the device skips step 6 and proceeds with step 7.
7. The device identifies whether a patch file is specified in the index file. If yes, the device confirms loading the patch file, skips steps 7, and proceeds with step 8. If not, the device skips step 7 and proceeds with step 8.
8. The device identifies whether a. cfg configuration file is specified in the index file. If yes, the device confirms loading the .cfg configuration file. If not, the device skips the step of loading files.
9. The device performs either of the following operations depending on whether an index file verification password has been specified for USB-based automatic configuration:
   - If an index file verification password has been specified for USB-based automatic configuration, the device verifies the index file first by comparing the password with the HMAC verification code in the index file. If the verification succeeds, the device copies the configuration files from the USB drive to the local device to perform automatic configuration. If the verification fails, the device does not perform USB-based automatic configuration.
   - If an index file verification password has not been specified, the device starts loading the files to be loaded. If files fail to be loaded, the USB-based automatic configuration process stops.

NOTE:

- When editing an index file, you must press **Enter** after entering a line in the specified format and then editing new contents. After editing the index file, save the index file.
- If the keyword of a parameter is not matched or found, the parameter is considered as empty.
- When the value for a parameter in the GLOBAL CONFIG area is different from that in the DEVICE DESCRIPTION area for a matching device, the configuration in the DEVICE DESCRIPTION area takes effect on the device.

Procedure

1. Enable USB-based automatic configuration on the device:
   a. Enter system view.
      ```
      system-view
      ```
   b. Enable USB-based automatic configuration.
      ```
      autodeploy udisk enable
      ```
      By default, USB-based automatic configuration is enabled.
      For the enabling configuration to take effect if USB-based automatic configuration has been disabled, you must save the configuration and reboot the device.
   c. (Optional) Specify an index file verification password for USB-based automatic configuration.

autodeploy udisk file-verification password { **simple** | **cipher** } *string*

By default, no index file verification password is configured for USB-based automatic configuration.

d. Save the running configuration.

save

2. Create an index file.
3. Save the index file to the root directory of the USB disk, and save the files specified in the index file to the specified directory. By default, the directory is the root directory of the USB disk.
4. Insert the USB disk into interface USB1, namely, usba0:, on the active device. After USB-based automatic configuration ends, the standby device will automatically synchronize its configuration with the active device at startup.
5. Observe the LEDs of the device.

   If the configuration file fails to be deployed, you can view the log file named *fully qualified configuration file name***.log** in the root directory of the USB disk to locate and resolve the issue.
6. Use the **display current-configuration** command to verify that the configuration file has been loaded correctly. For more information about this command, see configuration file management in *Fundamentals Command Reference*.
7. Remove the USB disk.

Configuring and using automatic configuration without an index file

Operating mechanism

Save a configuration file on a USB disk. When a device starts, it automatically obtains and executes the configuration file from the USB disk for automatic configuration.

The process for USB-based automatic configuration without an index file is as follows:

1. The device detects that the USB disk is present.
2. Check whether the index file named **smart_config.ini** exists on the USB disk.
   - If the index file named **smart_config.ini** exists, the device performs index file-based automatic configuration. For more information, see "Configuring and using index file-based automatic configuration."
   - If the index file named **smart_config.ini** does not exist, proceed to step 3.
3. Manually reboot the device.
4. When the device starts up, it detects whether a .cfg configuration file exists on the USB disk.
   - If a .cfg file exists, the device proceeds to step 5.
   - If no .cfg file exists, the USB-based automatic configuration process ends.
5. After the device obtains a configuration file from the USB disk, it compares the content of the configuration file with the content of the current main next-startup configuration file on the device.
   - If the contents are the same, the device directly loads the main next-startup configuration file instead of the configuration file on the USB disk.
   - If the contents are different, the device copies the configuration file on the USB disk to the local storage medium and proceeds to step 6.

     If the device has a local file that uses the same name as the configuration file copied from the USB disk, it handles the local file as follows:

     – If the local file is the main next-startup configuration file, the device saves the file as *original-file-name*_**bak.cfg**.

- If the local file is not the main next-startup configuration file, the device directly uses the configuration file copied from the USB disk to overwrite the local file.

6. The device executes the configuration file copied from the USB disk.
   - If the execution succeeds, the device automatically runs the configuration in the configuration file as the running configuration and sets the configuration file as the main next-startup configuration file.
   - If the device fails to execute any command in the configuration file, it ignores all settings in the configuration file and loads settings in the original next-startup configuration file. If no next-startup configuration file is available, the device starts up with initial settings.

Prerequisites

1. Prepare a USB disk that has only one partition.
2. Display the serial number of the device.

   `display device manuinfo`

   For more information about this command, see device management commands in *Fundamentals Command Reference*.
3. Create a configuration file named *device serial number***.cfg** or **autodeploy.cfg**, and save the file to the root directory of the file system on the USB disk.

   If a configuration file named *device serial number***.cfg** coexists with configuration file **autodeploy.cfg**, the configuration file named *device serial number***.cfg** is used.

Procedure

1. Enable USB-based automatic configuration on the device:
   a. Enter system view.

      `system-view`

   b. Enable USB-based automatic configuration.

      `autodeploy udisk enable`

      By default, USB-based automatic configuration is enabled.

      For the enabling configuration to take effect if USB-based automatic configuration has been disabled, you must save the configuration and reboot the device.

   c. Save the running configuration.

      `save`
2. Insert the USB disk with the configuration file saved into interface USB1, namely, usba0:, on the active device. After USB-based automatic configuration ends, the standby device will automatically synchronize its configuration with the active device at startup.
3. Reboot the device and observe the LEDs of the device.

   If the configuration file fails to be deployed, you can view the log file named *fully qualified configuration file name***.log** in the root directory of the USB disk to locate and resolve the issue.
4. Use the **`display current-configuration`** command to verify that the configuration file has been loaded correctly.

   For more information about this command, see configuration file management commands in *Fundamentals Command Reference*.
5. Remove the USB disk.

Server-based automatic configuration examples

Example: Using a TFTP server for automatic configuration

Network configuration

As shown in Figure 2, two departments of a company are connected to the network through gateways (Switch B and Switch C). Access devices Switch D, Switch E, Switch F, and Switch G do not have a configuration file.

Configure the servers and gateways so the access devices can obtain a configuration file to complete the following configuration tasks:

- Enable administrators of access devices to Telnet to and manage their respective access devices.
- Require administrators to enter their respective usernames and passwords at login.

Figure 2 Network diagram

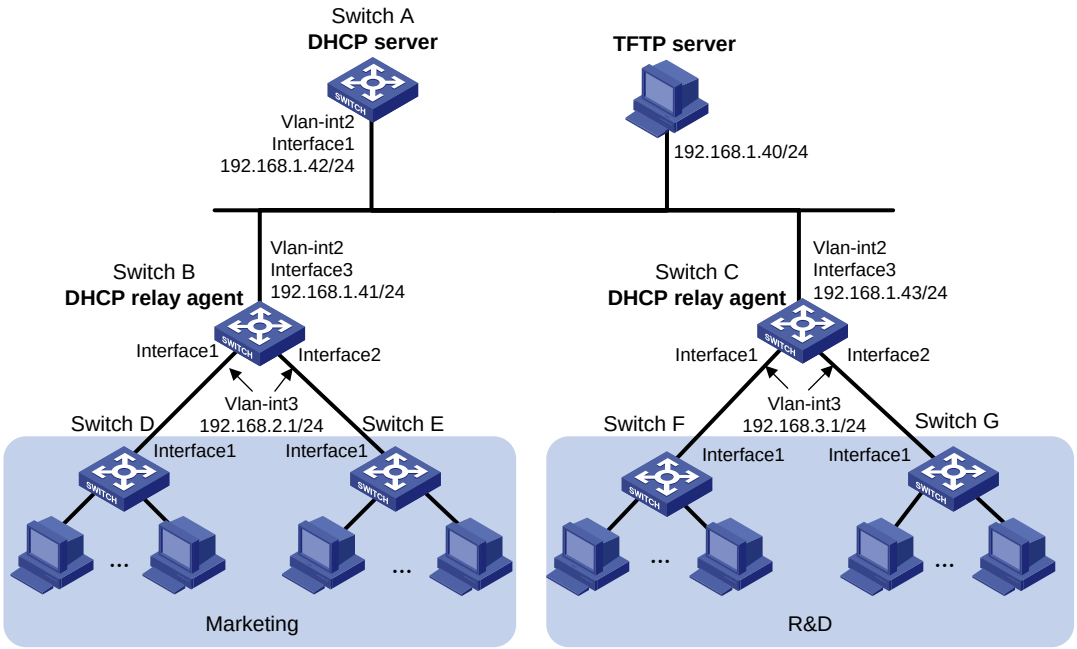


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Switch A:

   # Create a VLAN interface and assign an IP address to the interface.

   ```
   <Sysname> system-view
   [Sysname] sysname SwitchA
   [SwitchA] vlan 2
   ```

```
[SwitchA-vlan2] port gigabitethernet 1/0/1
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.1.42 24
[SwitchA-Vlan-interface2] quit
```

Enable DHCP.

```
[SwitchA] dhcp enable
```

Enable the DHCP server on VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] dhcp select server
[SwitchA-Vlan-interface2] quit
```

Configure address pool **market** to assign IP addresses on the 192.168.2.0/24 subnet to clients in the Marketing department. Specify the TFTP server, gateway, and configuration file name for the clients.

```
[SwitchA] dhcp server ip-pool market
[SwitchA-dhcp-pool-market] network 192.168.2.0 24
[SwitchA-dhcp-pool-market] tftp-server ip-address 192.168.1.40
[SwitchA-dhcp-pool-market] gateway-list 192.168.2.1
[SwitchA-dhcp-pool-market] bootfile-name market.cfg
[SwitchA-dhcp-pool-market] quit
```

Configure address pool **rd** to assign IP addresses on the 192.168.3.0/24 subnet to clients in the R&D department. Specify the TFTP server, gateway, and configuration file name for the clients.

```
[SwitchA] dhcp server ip-pool rd
[SwitchA-dhcp-pool-rd] network 192.168.3.0 24
[SwitchA-dhcp-pool-rd] tftp-server ip-address 192.168.1.40
[SwitchA-dhcp-pool-rd] gateway-list 192.168.3.1
[SwitchA-dhcp-pool-rd] bootfile-name rd.cfg
[SwitchA-dhcp-pool-rd] quit
```

Configure static routes to the DHCP relay agents.

```
[SwitchA] ip route-static 192.168.2.0 24 192.168.1.41
[SwitchA] ip route-static 192.168.3.0 24 192.168.1.43
[SwitchA] quit
```

2. Configure the gateway Switch B:

Create VLAN interfaces and assign IP addresses to the interfaces.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/3
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 192.168.1.41 24
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/1
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
```

```
[SwitchB-Vlan-interface3] ip address 192.168.2.1 24
[SwitchB-Vlan-interface3] quit
```

Enable DHCP.

```
[SwitchB] dhcp enable
```

Enable the DHCP relay agent on VLAN-interface 3.

```
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] dhcp select relay
```

Specify the DHCP server address.

```
[SwitchB-Vlan-interface3] dhcp relay server-address 192.168.1.42
```

3. Configure the gateway Switch C:

Create VLAN interfaces and assign IP addresses to the interfaces.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] vlan 2
[SwitchC-vlan2] port gigabitethernet 1/0/3
[SwitchC-vlan2] quit
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ip address 192.168.1.43 24
[SwitchC-Vlan-interface2] quit
[SwitchC] vlan 3
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] port gigabitethernet 1/0/2
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ip address 192.168.3.1 24
[SwitchC-Vlan-interface3] quit
```

Enable DHCP.

```
[SwitchC] dhcp enable
```

Enable the DHCP relay agent on VLAN-interface 3.

```
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] dhcp select relay
```

Specify the DHCP server address.

```
[SwitchC-Vlan-interface3] dhcp relay server-address 192.168.1.42
```

4. Configure the TFTP server:

On the TFTP server, create a configuration file named **market.cfg**.

```
#
 sysname Market
#
 telnet server enable
#
vlan 3
#
local-user market
 password simple build22345
 service-type telnet
 quit
#
```

```
interface Vlan-interface3
 ip address dhcp-alloc
 quit
#
interface gigabitethernet 1/0/1
 port access vlan 3
 quit
#
user-interface vty 0 63
 authentication-mode scheme
 user-role network-admin
#
return
```

On the TFTP server, create a configuration file named **rd.cfg**.

```
#
 sysname RD
#
 telnet server enable
#
vlan 3
#
local-user rd
 password simple create22345
 service-type telnet
 quit
#
interface Vlan-interface3
 ip address dhcp-alloc
 quit
#
interface gigabitethernet 1/0/1
 port access vlan 3
 quit
#
user-interface vty 0 63
 authentication-mode scheme
 user-role network-admin
#
return
```

Start TFTP service software, and specify the folder where the two configuration files reside as the working directory. (Details not shown.)

Verify that the TFTP server and DHCP relay agents can reach each other. (Details not shown.)

Verifying the configuration

1. Power on Switch D, Switch E, Switch F, and Switch G.
2. After the access devices start up, display assigned IP addresses on Switch A.

```
<SwitchA> display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration      Type
```

```
                Hardware address
192.168.2.2     3030-3066-2e65-3233-  May 6 05:21:25 2013   Auto(C)
                642e-3561-6633-2d56-
                6c61-6e2d-696e-7465-
                7266-6163-6533
192.168.2.3     3030-3066-2e65-3230-  May 6 05:22:50 2013   Auto(C)
                302e-3232-3033-2d56-
                6c61-6e2d-696e-7465-
                7266-6163-6533
192.168.3.2     3030-6530-2e66-6330-  May 6 05:23:15 2013   Auto(C)
                302e-3335-3131-2d56-
                6c61-6e2d-696e-7465-
                7266-6163-6531
192.168.3.3     3030-6530-2e66-6330-  May 6 05:24:10 2013   Auto(C)
                302e-3335-3135-2d56-
                6c61-6e2d-696e-7465-
                7266-6163-6532
```

3. Telnet to 192.168.2.2 from Switch A.

```
<SwitchA> telnet 192.168.2.2
```

4. Enter username **market** and password **build22345** as prompted. (Details not shown.) You are logged in to Switch D or Switch E.

Configuration files

- SwitchA

```
#
 sysname SwitchA
#
 dhcp enable
#
vlan 2
#
dhcp server ip-pool market
 gateway-list 192.168.2.1
 network 192.168.2.0 mask 255.255.255.0
 bootfile-name market.cfg
 tftp-server ip-address 192.168.1.40
#
dhcp server ip-pool rd
 gateway-list 192.168.3.1
 network 192.168.3.0 mask 255.255.255.0
 bootfile-name rd.cfg
 tftp-server ip-address 192.168.1.40
#
interface Vlan-interface2
 ip address 192.168.1.42 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
```

```
 port access vlan 2
#
 ip route-static 192.168.2.0 24 192.168.1.41
 ip route-static 192.168.3.0 24 192.168.1.43
#
return
```

- SwitchB

```
#
 sysname SwitchB
#
 dhcp enable
#
vlan 2 to 3
#
interface Vlan-interface2
 ip address 192.168.1.41 255.255.255.0
#
interface Vlan-interface3
 ip address 192.168.2.1 255.255.255.0
 dhcp select relay
 dhcp relay server-address 192.168.1.42
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 2
#
return
```

- SwitchC

```
#
 sysname SwitchC
#
 dhcp enable
#
vlan 2 to 3
#
interface Vlan-interface2
 ip address 192.168.1.43 255.255.255.0
#
interface Vlan-interface3
 ip address 192.168.3.1 255.255.255.0
```

```
 dhcp select relay
 dhcp relay server-address 192.168.1.42
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 2
#
return
```

Example: Using an HTTP server and Tcl scripts for automatic configuration

Network configuration

As shown in Figure 3, Switch A does not have a configuration file.

Configure the servers so Switch A can obtain a Tcl script to complete the following configuration tasks:

- Enable the administrator to Telnet to Switch A to manage Switch A.
- Require the administrator to enter the correct username and password at login.

Figure 3 Network diagram

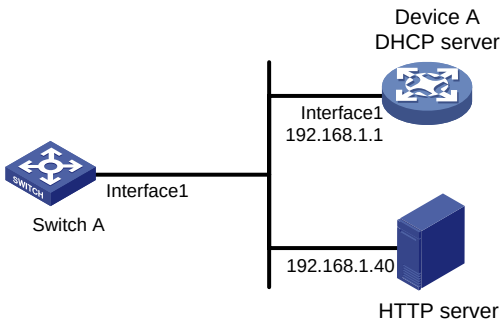


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure the DHCP server:

   # Enable DHCP.

   ```
   <Sysname> system-view
   ```

```
[Sysname] sysname DeviceA
 [DeviceA] dhcp enable
```

Configure address pool **1** to assign IP addresses on the 192.168.1.0/24 subnet to clients.

```
[DeviceA] dhcp server ip-pool 1
[DeviceA-dhcp-pool-1] network 192.168.1.0 24
```

Specify the URL of the script file for the clients.

```
[DeviceA-dhcp-pool-1] bootfile-name http://192.168.1.40/device.tcl
[DeviceA-dhcp-pool-1] quit
```

Enable the DHCP server on GigabitEthernet1/0/1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] ip address 192.168.1.1 24
[DeviceA-GigabitEthernet1/0/1] dhcp select server
[DeviceA-GigabitEthernet1/0/1] quit
```

2. Configure the HTTP server:

Create a configuration file named **device.tcl** on the HTTP server.

```
system-view
telnet server enable
local-user user
password simple hello22345
service-type telnet
quit
user-interface vty 0 63
authentication-mode scheme
user-role network-admin
quit

interface gigabitethernet 1/0/1
port link-mode route
ip address dhcp-alloc
return
```

Start HTTP service software and enable HTTP service. (Details not shown.)

Verifying the configuration

1. Power on Switch A.

2. After Switch A starts up, display assigned IP addresses on Device A.

```
<DeviceA> display dhcp server ip-in-use
IP address       Client identifier/    Lease expiration      Type
                 Hardware address
192.168.1.2      0030-3030-632e-3239-  Dec 12 17:41:15 2013  Auto(C)
                 3035-2e36-3736-622d-
                 4574-6830-2f30-2f32
```

3. Telnet to 192.168.1.2 from Device A.

```
<DeviceA> telnet 192.168.1.2
```

4. Enter username **user** and password **hello22345** as prompted. (Details not shown.)

You are logged in to Switch A.

Configuration files

```
#
```

```
 sysname DeviceA
#
 dhcp enable
#
dhcp server ip-pool 1
 network 192.168.1.0 mask 255.255.255.0
 bootfile-name http://192.168.1.40/device.tcl
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.255.0
#
return
```

Example: Using an HTTP server and Python scripts for automatic configuration

Network configuration

As shown in Figure 4, Switch A does not have a configuration file.

Configure the servers so Switch A can obtain a Python script to complete the following configuration tasks:

- Enable the administrator to Telnet to Switch A to manage Switch A.
- Require the administrator to enter the correct username and password at login.

Figure 4 Network diagram

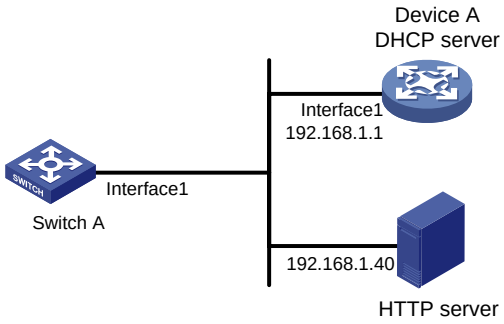


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure the DHCP server:

   # Enable DHCP.

   ```
   <Sysname> system-view
   [Sysname] sysname DeviceA
   [DeviceA] dhcp enable
   ```

   # Configure address pool **1** to assign IP addresses on the 192.168.1.0/24 subnet to clients.

```
[DeviceA] dhcp server ip-pool 1
[DeviceA-dhcp-pool-1] network 192.168.1.0 24
```

Specify the URL of the script file for the clients.

```
[DeviceA-dhcp-pool-1] bootfile-name http://192.168.1.40/device.py
[DeviceA-dhcp-pool-1] quit
```

Enable the DHCP server on GigabitEthernet1/0/1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] ip address 192.168.1.1 24
[DeviceA-GigabitEthernet1/0/1] dhcp select server
[DeviceA-GigabitEthernet1/0/1] quit
```

2. Configure the HTTP server:

Create a configuration file named **device.py** on the HTTP server.

```
#!usr/bin/python

import platformtools
platformtools.CLI('system-view ;telnet server enable ;local-user user ;password
simple hello22345 ;service-type telnet ;quit ;user-interface vty 0
63 ;authentication-mode scheme ;user-role network-admin ;quit ;interface
gigabitethernet 1/0/1 ;port link-mode route ;ip address dhcp-alloc ;return')
```

Start HTTP service software and enable HTTP service. (Details not shown.)

Verifying the configuration

1. Power on Switch A.
2. After Switch A starts up, display assigned IP addresses on Device A.

```
<DeviceA> display dhcp server ip-in-use
IP address       Client identifier/    Lease expiration      Type
                 Hardware address
192.168.1.2      0030-3030-632e-3239-  Dec 12 17:41:15 2013  Auto(C)
                 3035-2e36-3736-622d-
                 4574-6830-2f30-2f32
```

3. Telnet to 192.168.1.2 from Device A.

```
<DeviceA> telnet 192.168.1.2
```

4. Enter username **user** and password **hello22345** as prompted. (Details not shown.)

   You are logged in to Switch A.

Configuration files

```
#
 sysname DeviceA
#
 dhcp enable
#
dhcp server ip-pool 1
 network 192.168.1.0 mask 255.255.255.0
 bootfile-name http://192.168.1.40/device.py
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.255.0
```

```
#
return
```

Example: Using an SFTP server for automatic configuration

Network configuration

As shown in Figure 5, Switch A automatically obtains and executes a configuration file from the SFTP server after it starts up in order to meet the following requirements:

- The network administrator can log in to and control the switch through Telnet.
- Authentication is required for switch login to enhance security.

Figure 5 Network diagram

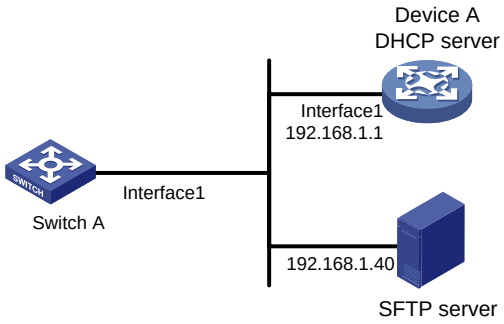


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure the DHCP server:

   # Enable DHCP, create DHCP address pool **1**, and specify primary subnet 192.168.1.0/24 for dynamic allocation in the address pool.

   ```
   <Sysname> system-view
   [Sysname] sysname DeviceA
   [DeviceA] dhcp enable
   [DeviceA] dhcp server ip-pool 1
   [DeviceA-dhcp-pool-1] network 192.168.1.0 24
   ```

   # Specify the URL of the configuration file. The username and password must be the same as those on the SFTP server.

   ```
   [DeviceA-dhcp-pool-1] bootfile-name sftp://user%dd:pass@192.168.1.40:22/device.cfg
   [DeviceA-dhcp-pool-1] quit
   ```

   # Enable the DHCP server on GigabitEthernet1/0/1.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ip address 192.168.1.1 24
   [DeviceA-GigabitEthernet1/0/1] dhcp select server
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

2. Configure the SFTP server to ensure that Switch A can successfully download configuration file **device.cfg** from the SFTP server:

   # Create configuration file **device.cfg** on the SFTP server. The file contents are as follows:

```
#
telnet server enable
#
local-user user
password simple hello12345
service-type telnet
quit
#
user-interface vty 0 63
authentication-mode scheme
user-role network-admin
quit
#
interface gigabitethernet 1/0/1
port link-mode route
ip address dhcp-alloc
return
```

Start the SFTP management software and enable the SFTP service. (Details not shown.)

Verifying the configuration

Start up Switch A without a startup configuration file. After the switch starts up successfully, display binding information about assigned IP addresses on Device A.

```
<DeviceA> display dhcp server ip-in-use
IP address       Client identifier/    Lease expiration      Type
                 Hardware address
192.168.1.2      0030-3030-632e-3239-  Dec 12 17:41:15 2013  Auto(C)
                 3035-2e36-3736-622d-
                 4574-6830-2f30-2f32
```

On Device A, Telnet to Switch A.

```
<SwitchA> telnet 192.168.1.2
```

Enter username **user** and password **hello12345**. Verify that you can Telnet to Switch A from Device A.

Configuration files

```
#
 sysname DeviceA
#
 dhcp enable
#
dhcp server ip-pool 1
 network 192.168.1.0 mask 255.255.255.0
 bootfile-name sftp://user%dd:pass@192.168.1.40:22/device.cfg
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.255.0
#
return
```

Example: Using an FTP server for automatic configuration

Network configuration

As shown in Figure 6, Switch A automatically obtains and executes a configuration file from the FTP server after it starts up in order to meet the following requirements:

- The network administrator can log in to and control the switch through Telnet.
- Authentication is required for switch login to enhance security.

Figure 6 Network diagram

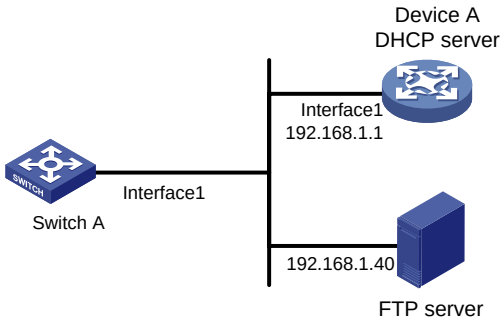


Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure the DHCP server:

   # Enable DHCP, create DHCP address pool **1**, and specify primary subnet 192.168.1.0/24 for dynamic allocation in the address pool.

   ```
   <Sysname> system-view
   [Sysname] sysname DeviceA
    [DeviceA] dhcp enable
   [DeviceA] dhcp server ip-pool 1
   [DeviceA-dhcp-pool-1] network 192.168.1.0 24
   ```

   # Specify the URL of the configuration file. The username and password must be the same as those on the FTP server.

   ```
   [DeviceA-dhcp-pool-1] bootfile-name ftp://user:pass@192.168.1.40:22/device.cfg
   [DeviceA-dhcp-pool-1] quit
   ```

   # Enable the DHCP server on GigabitEthernet1/0/1.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ip address 192.168.1.1 24
   [DeviceA-GigabitEthernet1/0/1] dhcp select server
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

2. Configure the FTP server to ensure that Switch A can successfully download configuration file **device.cfg** from the FTP server:

   # Create configuration file **device.cfg** on the FTP server. The file contents are as follows:

   ```
   #
   telnet server enable
   #
   ```

```
local-user user
password simple hello12345
service-type telnet
quit
#
user-interface vty 0 63
authentication-mode scheme
user-role network-admin
quit
#
interface gigabitethernet 1/0/1
port link-mode route
ip address dhcp-alloc
return
```

Start the FTP management software and enable the FTP service. (Details not shown.)

Verifying the configuration

Start up Switch A without a startup configuration file. After the switch starts up successfully, display binding information about assigned IP addresses on Device A.

```
<DeviceA> display dhcp server ip-in-use
IP address       Client identifier/    Lease expiration      Type
                 Hardware address
192.168.1.2      0030-3030-632e-3239-  Dec 12 17:41:15 2013  Auto(C)
                 3035-2e36-3736-622d-
                 4574-6830-2f30-2f32
```

On Device A, Telnet to Switch A.

```
<SwitchA> telnet 192.168.1.2
```

Enter username **user** and password **hello12345**. Verify that you can Telnet to Switch A from Device A.

Configuration files

```
#
 sysname DeviceA
#
 dhcp enable
#
dhcp server ip-pool 1
 network 192.168.1.0 mask 255.255.255.0
 bootfile-name ftp://user:pass@192.168.1.40:22/device.cfg
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.255.0
#
return
```

Example: Setting up an IRF fabric

Network configuration

As shown in Figure 7, Switch A and Switch B do not have a configuration file.

Configure the servers so the switches can obtain a Python script to complete their respective configurations and form an IRF fabric.

Figure 7 Network diagram

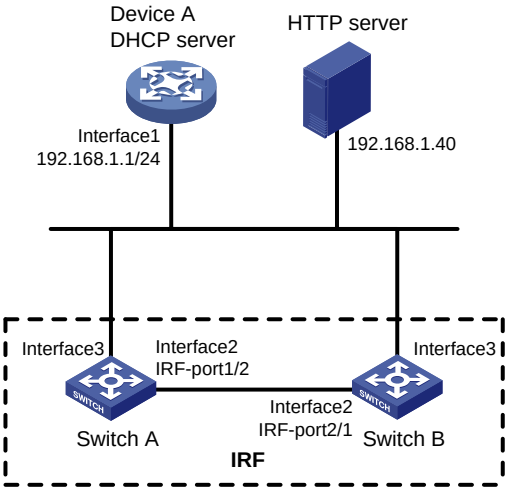


Table 9 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	Management Ethernet interface

Procedure

1. Configure Device A as the DHCP server:

   # Enable DHCP.

   ```
   <Sysname> system-view
   [Sysname] sysname DeviceA
   [DeviceA] dhcp enable
   ```

   # Configure IP pool **1** to assign IP addresses on the 192.168.1.0/24 subnet to clients.

   ```
   [DeviceA] dhcp server ip-pool 1
   [DeviceA-dhcp-pool-1] network 192.168.1.0 24
   ```

   # Specify the URL of the script file for the clients.

   ```
   [DeviceA-dhcp-pool-1] bootfile-name http://192.168.1.40/device.py
   [DeviceA-dhcp-pool-1] quit
   ```

   # Enable the DHCP server on GigabitEthernet 1/0/1.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ip address 192.168.1.1 24
   [DeviceA-GigabitEthernet1/0/1] dhcp select server
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

2. Configure the following files on the HTTP server:

File	Content	Remarks
.cfg configuration file	Commands required for IRF setup.	You can create a configuration file by copying and modifying the configuration file of an existing IRF fabric.
sn.txt	Serial numbers of the member switches.	Each SN uniquely identifies a switch. These SNs will be used for assigning a unique IRF member ID to each member switch.
(Optional.) .ipe or .bin software image file	Software images.	If the member switches are running different software versions, you must prepare the software image file used for software upgrade.
.py Python script file	Python commands that complete the following tasks: **1.** (Optional.) Verify that the flash memory has sufficient space for the files to be downloaded. **2.** Download the configuration file and **sn.txt**. **3.** (Optional.) Download the software image file and specify it as the main startup image file. **4.** Resolve **sn.txt** and assign a unique IRF member ID to each SN. **5.** Specify the configuration file as the main next-startup configuration file. **6.** Reboot the member switches.	For more information about Python script configuration, see "Using Python."

The specific content of the Python script is as follows (there are simple comments in the Python script, delete the comments when you use the script):

```
#!usr/bin/python

import comware
from time import strftime,gmtime,sleep
import signal
import os
import string
import commands
import hashlib

#python_config_file_mode
#'python_static'
#python_serial_number
#U can use 2 modes to obtain the config file
#- 'python_static'
#- 'python_serial_number'
```

```
python_config_file_mode = "python_serial_number"

# Configure the minimum remaining space required in the device flash for this software upgrade, with the unit in KB.
#Required space to copy config kickstart and system image in KB
required_space = 500000

# Configure the relevant parameters of the HTTP server: username, password, HTTP server IP address, VPN instance name, login timeout.
#transfer information
username = ""
password = ""
hostname = "192.168.1.40"
protocol = "http"
vrf = ""
config_timeout = 120
irf_timeout = 120
image_timeout = 2100

#HTTP server download path
#Server File Path
server_path = ""

# When you download it to the local device, store it in the root directory of the flash.
#Loacl File path
local_path = "flash:/"

# Name of the configuration file after being downloaded to the device.
#Local config name
config_local_name = "startup.cfg"

# Configuration file name on the HTTP server
#Server config name
config_server_name = "startup.cfg"

# Bootup software package downloaded to the local disk
#Local boot name
boot_local_name = "xxxx.ipe"

# Bootup software package on the HTTP server
#Server boot name
boot_server_name = "xxxx.ipe"

# Name of the SN file downloaded to the local disk
#Local irf name
irf_local_name = "sn.txt"

# SN file name on the HTTP server
#Server irf name
```

```
irf_server_name = "sn.txt"

# Upgrade log file name recorded on the device
python_log_name = ""

# Log upgrade method recorded on the device
#Write Log File
def write2Log(info):
  global python_log_name, local_path
  if python_log_name == "":
    try:
      python_log_name = "%s%s_python_%s_script.log" %(local_path,
strftime("%Y%m%d%H%M%S", gmtime()), os.getpid())
    except Exception as inst:
      print inst
  fd = open(python_log_name, "a")
  fd.write(info)
  fd.flush()
  fd.close()

# Obtain the device chassis number and slot number, and then calculate the local storage
path of the file by using the chassis number, slot number, and the local_path variable.
# get path according to the Chassis and Slot
def getPath(chassisID, slotID):
  global local_path
  path = ""
  obj = comware.get_self_slot()
  if (obj[0] == chassisID) and (obj[1] == slotID):
    return local_path
  if chassisID != -1:
    path = "chassis%d#" % chassisID
  if slotID != -1:
    path = "%sslot%d#%s" %(path, slotID, local_path)
  return path

# Delete the file downloaded to the local disk
#Remove File
def removeFile(filename):
  try:
    os.remove(filename)
  except os.error:
    pass

# Calculate the number of files to be deleted from the device.
#Cleanup one device temp files
def cleanDeviceFiles(str, oDevNode):
  global config_local_name, boot_local_name, irf_local_name
  sFilePath = getPath(oDevNode[0], oDevNode[1])
  if str == "error":
```

```
    removeFile("%s%s" %(sFilePath, config_local_name))
  removeFile("%s%s" %(sFilePath, boot_local_name))
  removeFile("%s%s" %(sFilePath, irf_local_name))
  removeFile("%s%s.md5" %(sFilePath, config_local_name))
  removeFile("%s%s.md5" %(sFilePath, boot_local_name))
  removeFile("%s%s.md5" %(sFilePath, irf_local_name))
  write2Log("\ndelete %s all files\n" %sFilePath)
  print "\ndelete %s all files" %sFilePath

#Cleanup files
def cleanupFiles(str):
  aSlotRange = []
  if ("get_standby_slot" in dir(comware)):
    aSlotRange = aSlotRange + comware.get_standby_slot()

  aSlotRange.append(comware.get_self_slot())
  i = 0
  while i < len(aSlotRange):
    if(aSlotRange[i] != None):
      cleanDeviceFiles(str, aSlotRange[i])
    i = i + 1

# Calculate the remaining flash space.
#Verify if free space is available to download config, boot-loader image
def verifyfreespace(path):
  global required_space
  try:
    s = os.statvfs(path)
    freespace = (s.f_bavail * s.f_frsize) /1024
    write2Log("\nthe %s free space is %s" %(path, freespace))
    print "\n####the %s free space is %s####" %(path, freespace)
    if required_space > freespace:
      write2Log("\nthe %s space is not enough" % path)
      print "\n####the %s space is not enough####" % path
      return False
  except Exception as inst:
    write2Log("\nverify %s free space exception: %s" % (path, inst))
    print "\n####verify %s free space exception: %s####" % (path, inst)
    return False
  return True

# Verify whether the remaining space on the device is sufficient.
#verify device freespace
def verifyDeviceFree(obj):
  path = getPath(obj[0], obj[1])
  if True != verifyfreespace(path):
    return False
  return True
```

```
#check all mpu free space
def verifyAllFreeSpace():
  aSlotRange = []
  if ("get_standby_slot" in dir(comware)):
    aSlotRange = aSlotRange + comware.get_standby_slot()

  aSlotRange.append(comware.get_self_slot())
  bAllEnough = True
  i = 0
  while i < len(aSlotRange):
    if(aSlotRange[i] != None) and (True != verifyDeviceFree(aSlotRange[i])):
      bAllEnough = False
    i = i + 1
  return bAllEnough

def doExit(str):
  if str == "success":
    write2Log("\nThe script is running success!")
    print "\n#### The script is running success! ####"
    cleanupFiles("success")
    comd = "reboot force"
    comware.CLI(comd, False)
    exit(0)
  if str == "error":
    write2Log("\nThe script is running failed!")
    print "\n#### The script is running failed! ####"
    cleanupFiles("error")
    exit(1)
  else:
    exit(0)

# Obtain the chassis number and slot number during software upgrade.
#get Chassis and Slot
def getChassisSlot(style):
  if style == "master":
    obj = comware.get_self_slot()
  if len(obj) <= 0:
    write2Log("\nget %s chassis and slot failed" % style)
    print "\n####get %s chassis and slot failed####" % style
    return None
  return obj

# Obtain the error information returned during software package startup.
#signal terminal handler function
def sig_handler_no_exit(signum, function):
  write2Log("\nSIGTERM Handler while configuring boot-loader variables")
  print "\n####SIGTERM Handler while configuring boot-loader variables####"
```

```
# Error information returned during software upgrade.
#signal terminal handler
def sigterm_handler(signum, function):
  write2Log("\nSIGTERM Handler")
  print "\n####SIGTERM Handler####"
  cleanupFiles("error")
  doExit("error")

# Download file from the HTTP server.
#transfer file
def doCopyFile(src = "", des = "", login_timeout = 10):
  global username, password, hostname, protocol, vrf
  print "INFO: Starting Copy of %s" % src
  try:
    removeFile(des)
    obj = comware.Transfer(protocol, hostname, src, des, vrf, login_timeout, username,
password)
    if obj.get_error() != None:
      write2Log("\ncopy %s failed: %s" % (src, obj.get_error()))
      print "\n####copy %s failed: %s####" % (src, obj.get_error())
      return False
  except Exception as inst:
    write2Log("\ncopy %s exception: %s" % (src, inst))
    print "\n####copy %s exception: %s####" % (src, inst)
    return False
  write2Log("\ncopy file %s to %s success" % (src, des))
  print "INFO: Completed Copy of %s" % src
  return True

#Get MD5SUM from md5ConfigFile
def getMD5SumGiven(keyword, filename):
  try:
    file = open(filename, "r")
    line = file.readline()
    while "" != line:
      if not string.find(line, keyword, 0, len(keyword)):
        line = line.split("=")
        line = line[1]
        line = line.strip()
        file.close()
        return line
      line = file.readline()
    file.close()
  except Exception as inst:
    write2Log("\nget %s md5 exception: %s" % (filename, inst))
    print "\n####get %s md5 exception: %s####" % (filename, inst)
  return ""
```

```
#verify MD5SUM of the file
def verifyMD5sumofFile(md5sumgiven, filename):
  if md5sumgiven == "":
    write2Log("\nverify %s md5 error: the %s md5 file is error" %(filename, filename))
    print "\n####verify %s md5 error: the %s md5 file is error####" %(filename,
filename)
    return False
  try:
    m = hashlib.md5()
    f = open(filename, 'rb')
    buffer = 8192
    while 1:
      chunk = f.read(buffer)
      if not chunk:
        break
      m.update(chunk)
    f.close()
    md5calculated = m.hexdigest()
  except Exception as inst:
    write2Log("\nverify %s md5 exception: %s" % (filename, inst))
    print "\n####verify %s md5 exception: %s####" % (filename, inst)
    return False
  if md5sumgiven == md5calculated:
    return True
  write2Log("\nverify %s md5 error: md5sumgiven is %s filemd5 is %s" %(filename,
md5sumgiven, md5calculated))
  print "\n####verify %s md5 error: md5sumgiven is %s filemd5 is %s####" %(filename,
md5sumgiven, md5calculated)
  return False

#Check MD5 file
def checkFile(src, dest):
  src = "%s.md5" % src
  destmd5 = "%s.md5" % dest
  bFlag = doCopyFile(src, destmd5, 120)
  if (True == bFlag) and (True == verifyMD5sumofFile(getMD5SumGiven("md5sum",
destmd5), dest)):
    write2Log("\ncheckFile success: %s" % destmd5)
    print "\n####checkFile success: %s####" % destmd5
    return True
  elif (True != bFlag):
    write2Log("\n%s is not exist! Don't verify the MD5 file!" % destmd5)
    print "INFO: %s is not exist! Don't verify the MD5 file!" % destmd5
    return True
  return False

#Get config file according to the mode
def getCfgFileName():
  global config_server_name
```

```
  if (python_config_file_mode == "python_serial_number") and
(os.environ.has_key('DEV_SERIAL')):
    config_server_name = "%s.cfg" % os.environ['DEV_SERIAL']
    return config_server_name
  else:
    return config_server_name

#copy file to all standby slot
def syncFileToStandby(sSrcFile, sFileName):
  try:
    aSlotRange = []
    if ("get_standby_slot" in dir(comware)):
      aSlotRange = aSlotRange + comware.get_standby_slot()

    i = 0
    while i < len(aSlotRange):
      if(aSlotRange[i] != None):
        sDestFile = "%s%s" %(getPath(aSlotRange[i][0], aSlotRange[i][1]),
sFileName)
        removeFile(sDestFile)
        open(sDestFile,"wb").write(open(sSrcFile,"rb").read())
        write2Log("\nsync file to standby %s" % (sDestFile))
        print "\n####sync file to standby %s####" % (sDestFile)
      i = i + 1
  except Exception as inst:
    write2Log("\nsync file to standby %s exception: %s" % (sSrcFile, inst))
    print "\n####sync file to standby %s exception: %s####" % (sSrcFile, inst)

# Based on the global variables defined at the beginning of the script, download the
startup software package, configuration file, and SN file.
#Procedure to copy config file using global information
def copyAndCheckFile(src, dest, timeout):
  global server_path, local_path
  srcTmp = "%s%s" % (server_path, src)
  sDestFile = "%s%s" % (local_path, dest)
  if (True == doCopyFile(srcTmp, sDestFile, timeout)) and (True == checkFile(srcTmp,
sDestFile)):
    syncFileToStandby(sDestFile, dest)
    return True
  else:
    srcTmp = "%sdefault_%s" %(server_path, src)
    if (True == doCopyFile(srcTmp, sDestFile, timeout)) and (True == checkFile(srcTmp,
sDestFile)):
      syncFileToStandby(dest)
      return True
  return False

# split the Chassis and Slot
def splitChassisSlot(chassisID, slotID):
```

```
  chassis_slot = ""
  if chassisID != -1:
    chassis_slot = " chassis %d"  % chassisID
  if slotID != -1:
    chassis_slot = "%s slot %d" %(chassis_slot, slotID)
  return chassis_slot

# Download startup software package from the HTTP server.
def copyBootImage():
  global image_timeout, local_path, boot_server_name, boot_local_name
  src = "%s" % boot_server_name
  return copyAndCheckFile(src, boot_local_name, image_timeout)

# Download configuration file from the HTTP server.
def copyCfgFile():
  global config_timeout, local_path, config_local_name
  src = "%s" % getCfgFileName()
  return copyAndCheckFile(src, config_local_name, config_timeout)

# Download the sn.txt file from the HTTP server.
def copyIrfStack():
  global irf_timeout, local_path, irf_local_name, irf_server_name
  src = "%s" % irf_server_name
  return copyAndCheckFile(src, irf_local_name, config_timeout)

# Execute the boot-loader command to specify the next-bootup main configuration file.
# Procedure to Install Boot Image
def installBoot(chassis_slot, sFile, style):
  result = None
  write2Log("\ninstall%s%s begin" %(chassis_slot, style))
  print "INFO: Install%s%s Start, Please Wait..." %(chassis_slot, style)
  comd = "boot-loader file %s%s%s" % (sFile, chassis_slot, style)
  try:
    result = comware.CLI(comd, False)
    if result == None:
      write2Log("\nboot-loader file %s%s%s failed" % (sFile, chassis_slot, style))
      print "\n####boot-loader file %s%s%s failed####" % (sFile, chassis_slot, style)
      return False
  except Exception as inst:
    write2Log("\nboot-loader %s exception: %s" % (sFile, inst))
    print "\n####boot-loader %s exception: %s####" % (sFile, inst)
    return False
  return True

#Procedure to install boot image
def installBootImage():
  global boot_local_name
  aSlotRange = [comware.get_self_slot()]
```

```
  if ("get_standby_slot" in dir(comware)):
    aSlotRange = aSlotRange + comware.get_standby_slot()
  bInstallOk = True
  i = 0
  while i < len(aSlotRange):
    sFile = "%s%s" %(getPath(aSlotRange[0][0], aSlotRange[0][1]), boot_local_name)
    if False == installBoot(splitChassisSlot(aSlotRange[i][0], aSlotRange[i][1]),
sFile, " main"):
      bInstallOk = False
    i = i + 1
  return bInstallOk

# Execute the startup saved-configuration command to specify the next-bootup
configuration file.
def startupCfg():
  global local_path, config_local_name
  result = None
  dest = "%s%s" %(local_path, config_local_name)
  write2Log("\nstartup saved-configuration %s begin" %dest)
  print "INFO: Startup Saved-configuration Start"
  comd = "startup saved-configuration %s main" % dest
  try:
    result = comware.CLI(comd, False)
    if result == None:
      write2Log("\nstartup saved-configuration %s failed" % dest)
      print "\n####startup saved-configuration %s failed####" % dest
      return False
  except Exception as inst:
    write2Log("\nstartup %s exception: %s" % (dest, inst))
    print "\n####startup %s exception: %s####" % (dest, inst)
    return False
  write2Log("\nstartup saved-configuration %s success" % dest)
  print "INFO: Completed Startup Saved-configuration"
  return True

def getIrfCfg(line, num):
  line = line.split()
  number = None
  if 3 == len(line):
    number = line[num]
  else :
    number = None
  return number

def getMemberID():
  aMemId = comware.get_self_slot()
  memId = None
  if aMemId[0] == -1 :
    memId = aMemId[1]
```

```
  else :
    memId = aMemId[0]
  return memId

def getNewMemberID():
  global irf_local_name, local_path, env
  filename = "%s%s" %(local_path, irf_local_name)
  serNum = os.environ['DEV_SERIAL']
  print "\n####Test Chassis SN or Slot SN %s" % serNum
  reNum = None
  try:
    file = open(filename, "r")
    line = file.readline()
    while "" != line:
      if (serNum == getIrfCfg(line, 0)):
        file.close()
        reNum = getIrfCfg(line, 2)
        return reNum
      line = file.readline()
    file.close()
  except Exception as inst:
    write2Log("\nget renumberID exception: %s" % inst)
    print "\n####get renumberID exception: %s####" % inst
  write2Log("\nget %s renumberID failed" % filename)
  print "\n#### get %s renumberID failed ####" % filename
  return reNum

# Check whether the device is in an IRF fabric.
def isIrfDevice():
  try:
    result = comware.CLI("display irf", False)
    if result == None:
      return False
  except Exception as inst:
    return False
  return True

# Parse the sn.txt file and use renumber to edit the IRF member ID of the device.
def getIrfComd():
  comd = None
  newMemberID = getNewMemberID()
  aMemId = comware.get_self_slot()
  if None == newMemberID:
    return None
  if False == isIrfDevice():
    comd = "system-view ; irf member %s ; chassis convert mode irf" % newMemberID
  else:
    comd = "system-view ; irf member %s renumber %s" % (getMemberID(), newMemberID)
```

```
  return comd

def stackIrfCfg():
  global env
  if (not os.environ.has_key('DEV_SERIAL')):
    write2Log("\nenviron variable 'DEV_SERIAL' is not found!")
    print "\n####environ variable 'DEV_SERIAL' is not found!####"
    return False
  comd = getIrfComd()
  if None == comd:
    return False
  result = None
  write2Log("\nstartup stack irf begin")
  print "INFO: Startup stack irf Start"
  try:
    result = comware.CLI(comd, False)
    if result == None:
      write2Log("\nstartup stack irf failed: %s" % comd)
      print "\n####startup stack irf failed: %s####" %comd
      return False
  except Exception as inst:
    write2Log("\nstartup stack irf exception: %s command: %s" % (inst, comd))
    print "\n####startup stack irf exception: %s command: %s####" % (inst, comd)
    return False
  write2Log("\nstartup stack irf success")
  print "INFO: Completed Startup Stack Irf"
  return True

#check if all standby slots are ready
def ifAllStandbyReady():
  if (("get_slot_range" in dir(comware)) == False):
    return True

  aSlotRange = comware.get_slot_range()
  bAllReady = True
  for i in range(aSlotRange["MinSlot"], aSlotRange["MaxSlot"]):
    oSlotInfo =  comware.get_slot_info(i)
    if (oSlotInfo != None) and (oSlotInfo["Role"] == "Standby") and
(oSlotInfo["Status"] == "Fail"):
      bAllReady = False
      write2Log("\nSlot %s is not ready!" %i)
      print "\n####Slot %s is not ready!####" %i
  return bAllReady

#if have any standby slot was not ready sleep for waiting
def waitStandbyReady():
  while ifAllStandbyReady() == False:
    sleep(10)
```

```
#python main

#when download file user can stop script
waitStandbyReady()
signal.signal(signal.SIGTERM, sigterm_handler)

if (True == verifyAllFreeSpace()) and (True == copyBootImage()) and (True ==
copyCfgFile()) and (True == copyIrfStack()):
  #after download file user can not stop script
  signal.signal(signal.SIGTERM, sig_handler_no_exit)
  if (True == installBootImage()) and (True == startupCfg()) and (True ==
stackIrfCfg()):
    doExit("success")
doExit("error")
```

3. Power on Switch A and Switch B.

   Switch A and Switch B will obtain the Python script file from the DHCP server and execute the script. After completing the IRF configuration, Switch A and Switch B reboot.

4. After Switch A and Switch B start up again, use a cable to connect Switch A and Switch B through their IRF physical ports.

   Switch A and Switch B will elect a master member. The subordinate member will reboot to join the IRF fabric.

Verifying the configuration

On Switch A, display IRF member devices. You can also use the **`display irf`** command on Switch B to display IRF member devices.

```
<Switch A> display irf
MemberID  Slot  Role    Priority  CPU-Mac         Description
   1      1     Standby 1         00e0-fc0f-8c02  ---
 *+2      1     Master  30        00e0-fc0f-8c14  ---
-------------------------------------------------
 * indicates the device is the master.
 + indicates the device through which the user logs in.

 The Bridge MAC of the IRF is: 000c-1000-1111
 Auto upgrade                : yes
 Mac persistent              : always
 Domain ID                   : 0
 Auto merge                  : yes
```

The output shows that the switches have formed an IRF fabric.

Configuration files

```
#
 sysname DeviceA
#
 dhcp enable
#
dhcp server ip-pool 1
 network 192.168.1.0 mask 255.255.255.0
 bootfile-name http://192.168.1.40/device.py
```

```
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.255.0
#
return
```

Contents

Managing the device 1

Device management tasks at a glance 1
Configuring the device name 2
Configuring the system ID 2
Configuring the system time 2
About the system time 2
Restrictions and guidelines for configuring the system time 2
System time configuration tasks at a glance 3
Setting the system time at the CLI 3
Obtaining the UTC time through a time protocol 3
Setting the time zone 4
Setting the daylight saving time 4
Enabling displaying the copyright statement 4
Configuring banners 4
Disabling password recovery capability 6
Disabling USB interfaces 6
Setting the device operating mode 7
Setting the system operating mode 7
Setting the port state detection timer 8
Monitoring CPU usage 8
Setting memory alarm thresholds 10
Monitoring DMA memory 13
Configuring resource monitoring 13
Setting the temperature alarm thresholds 14
Disabling the notification of hardware resource alarms 15
Configuring consistency error logging for software and hardware forwarding entries 15
Configuring parity error and ECC error logging for entries on forwarding chips 16
Configuring uncorrectable parity error and ECC error logging for entries on forwarding chips 16
Configuring device poweroff alarming 17
Ignoring absent fans 18
Verifying and diagnosing transceiver modules 18
Verifying transceiver modules 18
Diagnosing transceiver modules 19
Enabling alarm for third-party transceiver modules 19
Configuring transceiver monitoring 20
Configuring transceiver health check 20
Enabling a lane on a transceiver module 21
Setting the FEC mode for a transceiver module 21
Monitoring optical power of transceiver modules 22
Configuring RX and TX power thresholds for a transceiver module 22
Configuring RX and TX power change thresholds for a transceiver module 23
Scheduling a task 23
About task scheduling 23
Restrictions and guidelines 23
Procedure 24
Example: Scheduling a task 25
Locating devices 31
About device locating 31
Starting LED flashing 31
Stopping LED flashing 31
Rebooting the device 31
About device reboot 31
Restrictions and guidelines for device reboot 32
Rebooting the device immediately at the CLI 32
Scheduling a device reboot 32
Restoring the factory-default configuration 33
Display and maintenance commands for device management configuration 33

Managing the device

This chapter describes how to configure basic device parameters and manage the device.

Device management tasks at a glance

All device management tasks are optional. You can perform any of the tasks in any order.

- Configuring basic parameters
  - Configuring the device name
  - Configuring the system ID
  - Configuring the system time
  - Enabling displaying the copyright statement
  - Configuring banners
- Configuring security parameters
  - Disabling password recovery capability
- Adjusting device capacities
  - Setting the device operating mode
  - Setting the system operating mode
  - Setting the port state detection timer
- Monitoring the device
  - Monitoring CPU usage
  - Setting memory alarm thresholds
  - Monitoring DMA memory
  - Configuring resource monitoring
  - Setting the temperature alarm thresholds
  - Configuring consistency error logging for software and hardware forwarding entries
- Managing resources
  - Ignoring absent fans
  - Verifying and diagnosing transceiver modules
  - Configuring transceiver monitoring
  - Configuring transceiver health check
  - Enabling a lane on a transceiver module
  - Monitoring optical power of transceiver modules
    - Configuring RX and TX power thresholds for a transceiver module
    - Configuring RX and TX power change thresholds for a transceiver module
- Maintaining the device
  - Scheduling a task
  - Rebooting the device
  - Restoring the factory-default configuration

Configuring the device name

About this task

A device name (also called hostname) identifies a device in a network and is used in CLI view prompts. For example, if the device name is **Sysname**, the user view prompt is <Sysname>.

Procedure

1. Enter system view.
   system-view
2. Configure the device name.
   sysname *sysname*
   By default, the device name is **H3C**.

Configuring the system ID

About this task

You can use the system ID to indicate the position or functionality of the device or any other information.

Procedure

1. Enter system view.
   system-view
2. Configure the system ID.
   sysid *system-id*
   By default, the device does not have a system ID.

Configuring the system time

About the system time

Correct system time is essential to network management and communication. Configure the system time correctly before you run the device on the network.

The device can use one of the following methods to obtain the system time:

- Uses the locally set system time, and then uses the clock signals generated by its built-in crystal oscillator to maintain the system time.
- Periodically obtains the UTC time from an NTP source, and uses the UTC time, time zone, and daylight saving time to calculate the system time. For more information about NTP, see *Network Management and Monitoring Configuration Guide*.

The system time calculated by using the UTC time from a time source is more precise.

Restrictions and guidelines for configuring the system time

After you execute the **clock protocol none** command, the **clock datetime** command determines the system time, whether or not the time zone or daylight saving time has been configured.

If you configure or change the time zone or daylight saving time after the device obtains the system time, the device recalculates the system time. To view the system time, execute the **display clock** command.

System time configuration tasks at a glance

To configure the system time, perform the following tasks:

1. Configuring the system time

   Choose one of the following tasks:

   - Setting the system time at the CLI
   - Obtaining the UTC time through a time protocol

2. (Optional.) Setting the time zone

   Make sure each network device uses the time zone of the place where the device resides.

3. (Optional.) Setting the daylight saving time

   Make sure each network device uses the daylight saving time parameters of the place where the device resides.

Setting the system time at the CLI

1. Enter system view.

   system-view

2. Configure the device to use the local system time.

   clock protocol none

   By default, the device uses the NTP time source.

   If you execute the **clock protocol** command multiple times, the most recent configuration takes effect.

3. Return to user view.

   quit

4. Set the local system time.

   clock datetime *time date*

   By default, the system time is UTC time 00:00:00 01/01/2011.

CAUTION:

Change of the system time affects execution of system time-related features (such as scheduled tasks) and cooperative operation with other devices (such as log reporting and statistics collection). Before using this command, make sure you understand its impact on the live network.

Obtaining the UTC time through a time protocol

1. Enter system view.

   system-view

2. Specify the system time source.

   clock protocol ntp

   By default, the device uses the NTP time source.

   If you execute this command multiple times, the most recent configuration takes effect.

3. Configure time protocol parameters.

   For more information about NTP configuration, see *Network Management and Monitoring Configuration Guide*.

Setting the time zone

1. Enter system view.

   system-view
2. Set the time zone.

   clock timezone *zone-name* { **add** | **minus** } *zone-offset*

   By default, the system uses the UTC time zone.

Setting the daylight saving time

1. Enter system view.

   system-view
2. Set the daylight saving time.

   clock summer-time *name start-time start-date end-time end-date add-time*

   By default, the daylight saving time is not set.

Enabling displaying the copyright statement

About this task

This feature enables the device to display the copyright statement in the following situations:

- When a Telnet or SSH user logs in.
- When a console user quits user view. This is because the device automatically tries to restart the user session.

If you disable displaying the copyright statement, the device does not display the copyright statement in any situations.

Procedure

1. Enter system view.

   system-view
2. Enable displaying the copyright statement.

   copyright-info enable

   By default, displaying the copyright statement is enabled.

Configuring banners

About this task

Banners are messages that the system displays when a user logs in.

The system supports the following banners:

- **Legal banner**—Appears after the copyright statement.
- **Message of the Day (MOTD) banner**—Appears after the legal banner and before the login banner.

- **Login banner**—Appears only when password or scheme authentication is configured.
- **Shell banner**—Appears when a user accesses user view.
- **Shell banner**—Appears for a login user when the user accesses user view.
- **Shell banner**—Appears when the user accesses user view.
- **Shell banner**—Appears when the user accesses user view.

The system displays the banners in the following order: legal banner, MOTD banner, login banner, and shell banner.

Banner input methods

You can configure a banner by using one of the following methods:

- Input the entire command line in a single line.

  The banner cannot contain carriage returns. The entire command line, including the command keywords, the banner, and the delimiters, can have a maximum of 511 characters. The delimiters for the banner can be any printable character but must be the same. You cannot press **Enter** before you input the end delimiter.

  For example, you can configure the shell banner "Have a nice day." as follows:

  ```
  <System> system-view
  [System] header shell %Have a nice day.%
  ```

- Input the command line in multiple lines.

  The banner can contain carriage returns. A carriage return is counted as two characters.

  To input a banner configuration command line in multiple lines, use one of the following methods:

  - Press **Enter** after the final command keyword, type the banner, and end the final line with the delimiter character %. The banner plus the delimiter can have a maximum of 1999 characters.

    For example, you can configure the banner "Have a nice day." as follows:

    ```
    <System> system-view
    [System] header shell
    Please input banner content, and quit with the character '%'.
    Have a nice day.%
    ```

  - After you type the final command keyword, type any printable character as the start delimiter for the banner and press **Enter**. Then, type the banner and end the final line with the same delimiter. The banner plus the end delimiter can have a maximum of 1999 characters.

    For example, you can configure the banner "Have a nice day." as follows:

    ```
    <System> system-view
    [System] header shell A
    Please input banner content, and quit with the character 'A'.
    Have a nice day.A
    ```

  - After you type the final command keyword, type the start delimiter and part of the banner. Make sure the final character of the final string is different from the start delimiter. Then, press **Enter**, type the rest of the banner, and end the final line with the same delimiter. The banner plus the start and end delimiters can have a maximum of 2002 characters.

    For example, you can configure the banner "Have a nice day." as follows:

    ```
    <System> system-view
    [System] header shell AHave a nice day.
    Please input banner content, and quit with the character 'A'.
    A
    ```

Procedure

1. Enter system view.
   system-view
2. Configure the legal banner.
   header legal *text*
3. Configure the MOTD banner.
   header motd *text*
4. Configure the login banner.
   header login *text*
5. Configure the shell banner.
   header shell *text*

Disabling password recovery capability

About this task

Password recovery capability controls console user access to the device configuration and SDRAM from BootWare menus. For more information about BootWare menus, see the release notes.

If password recovery capability is enabled, a console user can access the device configuration without authentication to configure a new password.

If password recovery capability is disabled, console users must restore the factory-default configuration before they can configure new passwords. Restoring the factory-default configuration deletes the next-startup configuration files.

To enhance system security, disable password recovery capability.

Restrictions and guidelines

To access the device configuration without authentication, you must connect to the master device and access the BootWare menu during the device startup.

Procedure

1. Enter system view.
   system-view
2. Disable password recovery capability.
   undo password-recovery enable
   By default, password recovery capability is enabled.

Disabling USB interfaces

About this task

You can use USB interfaces to upload or download files. By default, all USB interfaces are enabled. You can disable USB interfaces as needed.

Prerequisites

Before using this command, use the **umount** command to unmount all USB partitions. For more information about this command, see *Fundamentals Command Reference*.

Procedure

1. Enter system view.

system-view

2. Disable USB interfaces.

   usb disable

   By default, all USB interfaces are enabled.

Setting the device operating mode

About this task

Supported features and feature specifications vary by device operating mode.

The device can operate in one of the following modes:

- **advance**—Advanced mode.
- **expert**—Expert mode.
- **standard**—Standard mode.

Restrictions and guidelines

Some early manufactured devices do not support the expert mode. Support for the expert mode depends on the device model.

A change to the device operating mode takes effect after a device reboot.

Procedure

1. Enter system view.

   system-view

2. Set the device operating mode.

   system-working-mode { **advance** | **expert** | **standard** }

   By default, the device operates in standard mode.

   By default, the device operates in expert mode.

Setting the system operating mode

About this task

The capacities of the MAC address table, ARP table, and routing tables vary by device operating mode. Table 1 lists the device operating modes and their feature highlights and application scenarios.

Table 1 Device operating modes and the feature highlights and application scenarios

Operating mode	Feature highlights	Recommended application scenario
normal	Uses only the default MAC address table and routing table and does not extend these tables.	A normal network environment is required.
bridging	Increases the MAC address table capacity to provide higher Layer 2 packet forwarding performance.	A large MAC address table is required.
routing	Increases the IPv4 routing table capacity to provide higher Layer 3 routing performance.	A large IPv4 or IPv6 routing table is required.
balance	Increases the capacities of the MAC address table, ARP table, and routing tables	Requirements for the MAC address table, ARP tables, and routing tables

Operating mode	Feature highlights	Recommended application scenario
	to provide more balanced Layer 2 and Layer 3 packet forwarding performance.	are balanced.
multicast	Increases the multicast table size to provide higher Layer 3 packet forwarding performance.	A large multicast table is required.

The enhanced operating modes include **enhance**, **enhance-bridging**, **enhance-routing**, and **enhance-balance**. In an enhanced operating mode, the capacity of the MAC address table, ARP table, IPV4 FIB table, ND table, IPv6 FIB table, or ACL table doubles at the expense of forwarding performance. An enhanced operating mode is suitable for networks that require large table capacities and are less sensitive to forwarding performance.

Restrictions and guidelines

To have the operating mode change take effect, you must reboot the device.

Procedure

1. Enter system view.

   system-view
2. Set the system operating mode.

   switch-mode slot *slot-number* { **balance** | **bridging** | **enhance** | **enhance-balance** | **enhance-bridging** | **enhance-routing** | **normal** [**ipv6-64**] | **routing** [**ipv6-64**] | **multicast** }

   By default, the device operates in normal mode.

Setting the port state detection timer

About this task

On a network enabled with loopback detection or the spanning tree feature, the device starts a port status detection timer when a port is shut down by a protocol. If the port remains in down state when the timer expires, the device sets the port state to the port's real physical state.

Procedure

1. Enter system view.

   system-view
2. Set the port state detection timer.

   shutdown-interval *time*

   The default setting is 30 seconds.

Monitoring CPU usage

About this task

To monitor CPU usage, the device performs the following operations:

- Samples CPU usage at 1-minute intervals, and compares the samples with CPU usage thresholds to identify the CPU usage status and send alarms or notifications accordingly.
- Samples CPU core usage at 1-minute intervals, and compares the samples with CPU core usage thresholds to identify the CPU core usage status. If a sample is greater than the threshold, the device sends a log message.

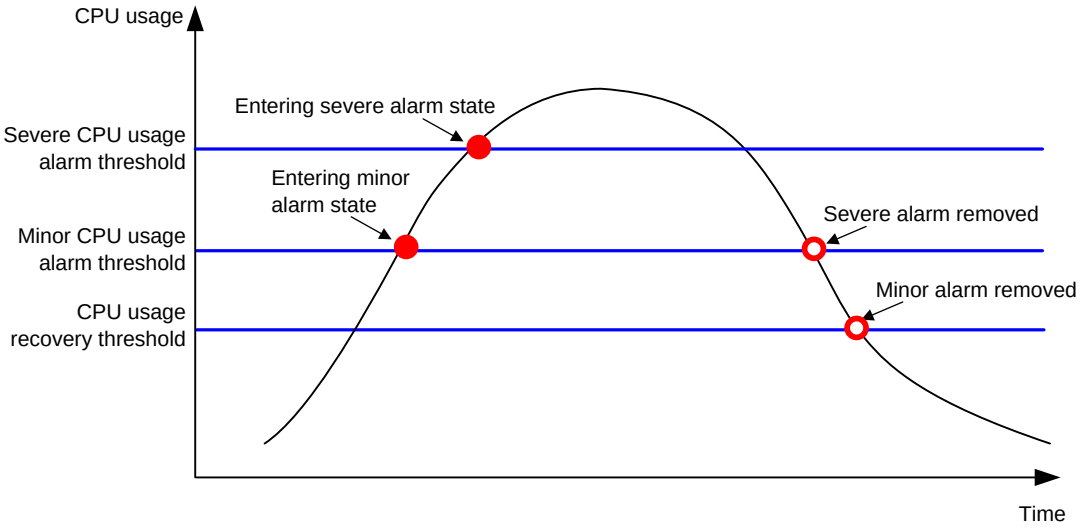
- Samples and saves CPU usage at a configurable interval if CPU usage tracking is enabled. You can use the **display cpu-usage history** command to display the historical CPU usage statistics in a coordinate system.

The device supports the following CPU usage thresholds:

- **Minor threshold**—If the CPU usage increases to or above the minor threshold but is less than the severe threshold, the CPU usage enters minor alarm state. The device sends minor alarms periodically until the CPU usage increases above the severe threshold or the minor alarm is removed.
- **Severe threshold**—If the CPU usage increases above the severe threshold, the CPU usage enters severe alarm state. The device sends severe alarms periodically until the severe alarm is removed.
- **Recovery threshold**—If the CPU usage decreases below the recovery threshold, the CPU usage enters recovered state. The device sends a recovery notification.

CPU usage alarms and notifications can be sent to NETCONF, SNMP, and the information center to be encapsulated as NETCONF events, SNMP traps and informs, and log messages. For more information, see NETCONF, SNMP, and information center in *Network Management and Monitoring Configuration Guide*.

Figure 1 CPU alarms and alarm-removed notifications



Procedure

1. Enter system view.

   system-view

2. Set the CPU usage alarm thresholds.

   monitor cpu-usage threshold *severe-threshold* { **minor-threshold** *minor-threshold* **recovery-threshold** *recovery-threshold* [**slot** *slot-number* [**cpu** *cpu-number*]] }

   The default settings are as follows:

   - **Severe CPU usage alarm threshold**—99%.
   - **Minor CPU usage alarm threshold**—80%.
   - **CPU usage recovery threshold**—60%.

CAUTION:

If you set the severe CPU usage alarm threshold to a too low value, the device will reach the threshold easily. Normal service processing will be affected.

3. Set CPU usage alarm resending intervals.

   monitor resend cpu-usage { **minor-interval** *minor-interval* | **severe-interval** *severe-interval* } * [**slot** *slot-number* [**cpu** *cpu-number*]]

   By default, the minor alarm resending interval is 300 seconds and the severe alarm resending interval is 60 seconds.

4. Set the sampling interval for CPU usage tracking.

   monitor cpu-usage interval *interval* [**slot** *slot-number* [**cpu** *cpu-number*]]

   By default, the sampling interval for CPU usage tracking is 1 minute.

5. Enable CPU usage tracking.

   monitor cpu-usage enable [**slot** *slot-number* [**cpu** *cpu-number*]]

   By default, CPU usage tracking is enabled.

6. (Optional.) Enable periodic CPU usage logging.

   monitor cpu-usage logging interval *interval-time*

   By default, periodic CPU usage logging is disabled.

Setting memory alarm thresholds

About this task

To ensure correct operation and improve memory efficiency, the system monitors the amount of free memory space in real time. If the amount of free memory space reaches the minor, severe, or critical alarm threshold, the system issues an alarm to affected service modules and processes.

The system can also issue an early warning to warn you of an approaching insufficient-memory condition.

The system monitors only the amount of free memory space. You can use the **display memory** command to display memory usage information.

The system supports these free memory thresholds:

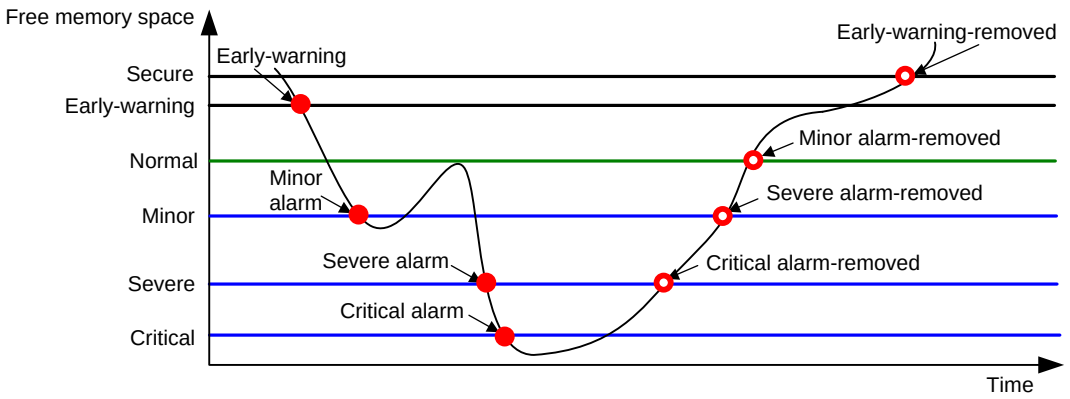
- Sufficient-memory threshold.
- Early-warning threshold.
- Normal state threshold.
- Minor alarm threshold.
- Severe alarm threshold.
- Critical alarm threshold.

Table 2 Memory alarm notifications and memory alarm-removed notifications

Notification	Triggering condition	Remarks
Early-warning notification	The amount of free memory space decreases below the early-warning threshold.	After generating and sending an early-warning notification, the system does not generate and send any additional early-warning notifications until the early warning is removed.
Minor alarm notification	The amount of free memory	After generating and sending a minor alarm

Notification	Triggering condition	Remarks
	space decreases below the minor alarm threshold.	notification, the system does not generate and send any additional minor alarm notifications until the minor alarm is removed.
Severe alarm notification	The amount of free memory space decreases below the severe alarm threshold.	After generating and sending a severe alarm notification, the system does not generate and send any additional severe alarm notifications until the severe alarm is removed.
Critical alarm notification	The amount of free memory space decreases below the critical alarm threshold.	After generating and sending a critical alarm notification, the system does not generate and send any additional critical alarm notifications until the critical alarm is removed.
Critical alarm-removed notification	The amount of free memory space increases above the severe alarm threshold.	N/A
Severe alarm-removed notification	The amount of free memory space increases above the minor alarm threshold.	N/A
Minor alarm-removed notification	The amount of free memory space increases above the normal state threshold.	N/A
Early-warning-removed notification	The amount of free memory space increases above the sufficient-memory threshold.	N/A

Figure 2 Memory alarm notifications and alarm-removed notifications



Restrictions and guidelines

If a memory alarm occurs, delete unused configuration items or disable some features to increase the free memory space. Because the memory space is insufficient, some configuration items might not be able to be deleted.

The system reboots the device if one of the following conditions is met:

- After a critical alarm occurs, the remaining free-memory value has been smaller than the critical alarm threshold for 30 seconds.
- The interval between two consecutive critical alarms is shorter than 30 seconds.
- The critical alarm has occurred three times within three minutes.

- After a critical alarm occurs, the system will periodically sample free memory space and predict if the free memory space will be exhausted within 30 seconds. If the prediction result indicates that the free memory space will be exhausted within 30 seconds, the system will reboot the device.

Once the free memory space reaches the early warning, minor, severe, or critical alarm threshold, the device will display the current memory usage information when you log in to the device through console or Telnet login, or execute every command.

After you configure this feature, if the amount of the free memory space reaches an alarm threshold, the system does not support creating a segment or entering its view. However, an existing microsegment will not be affected by this feature.

After you configure this feature, if the amount of the free memory space reaches the minor alarm threshold, the device cannot generate an asymmetric key pair through the **public-key local create** command. For more information, see PKI configuration in *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Set the memory usage threshold.

   memory-threshold [**slot** *slot-number* [**cpu** *cpu-number*]] **usage** *memory-threshold*

   By default, the memory usage threshold is 100%.

3. Set the free-memory thresholds.

   memory-threshold [**slot** *slot-number* [**cpu** *cpu-number*]] [**ratio**] **minor** *minor-value* **severe** *severe-value* **critical** *critical-value* **normal** *normal-value* [**early-warning** *early-warning-value* **secure** *secure-value*]

   The default settings are as follows:

   - **Minor alarm threshold**—386 MB.
   - **Severe alarm threshold**—270 MB.
   - **Critical alarm threshold**—154 MB.
   - **Normal state threshold**—502 MB.

   The default settings are as follows:

   - **Minor alarm threshold**—194 MB.
   - **Severe alarm threshold**—136 MB.
   - **Critical alarm threshold**—77 MB.
   - **Normal state threshold**—252 MB.
   - **Early-warning threshold**—311 MB.
   - **Sufficient-memory threshold**—350 MB.

4. Set memory depletion alarm resending intervals.

   monitor resend memory-threshold { **critical-interval** *critical-interval* | **early-warning-interval** *early-warning-interval* | **minor-interval** *minor-interval* | **severe-interval** *severe-interval* } * [**slot** *slot-number* [**cpu** *cpu-number*]]

   The following are the default settings:

   - **Early warning resending interval**—1 hour.
   - **Minor alarm resending interval**—12 hours.
   - **Severe alarm resending interval**—3 hours.
   - **Critical alarm resending interval**—1 hour.

5. (Optional.) Enable periodic memory usage logging.

   monitor memory-usage logging interval *interval-time*

   By default, periodic memory usage logging is disabled.

Monitoring DMA memory

About this task

To ensure correct operation of services that use Direct Memory Access (DMA) memory, the system monitors the amount of free DMA memory space regularly. If the amount of free DMA memory space decreases to or below the alarm threshold, the system generates a notification that the DMA memory space is insufficient. If the amount of free DMA memory space increases above the normal state threshold, the system generates a notification that the DMA memory space is sufficient.

DMA memory alarm notifications are sent to NETCONF, SNMP, and the information center to be encapsulated as NETCONF events, SNMP traps and informs, and log messages. For more information about NETCONF, SNMP, and information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Set the DMA memory thresholds.

   memory-threshold dma [**slot** *slot-number* [**cpu** *cpu-number*]] [**ratio**] **critical** *critical-value* **normal** *normal-value*

   The DMA memory alarm threshold is 2048 KB and the normal state threshold is 4096 KB.

Configuring resource monitoring

About this task

The resource monitoring feature enables the device to monitor the available amounts of types of resources, for example, the space for ARP entries. The device samples the available amounts periodically and compares the samples with resource depletion thresholds to identify the resource depletion status.

The device supports a minor resource depletion threshold and a severe resource depletion threshold for each supported resource type.

- If the available amount is equal to or less than the minor resource depletion threshold but greater than the severe resource depletion threshold, the resource type is in minor alarm state.
- If the available amount is equal to or less than the severe resource depletion threshold, the resource type is in severe alarm state.
- If the available amount increases above the minor resource depletion threshold, the resource type is in recovered state.

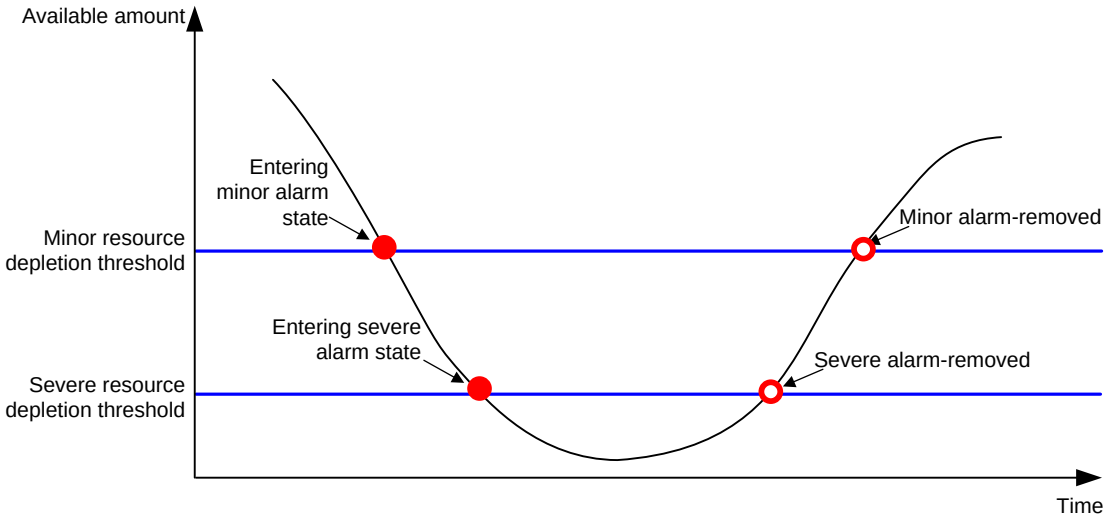
When a resource type enters severe alarm state, the device issues a severe alarm. If the resource type stays in severe alarm state, the device resends severe alarms periodically.

When a resource type enters minor alarm state, the device issues a minor alarm. If the resource type stays in minor alarm state or changes from severe alarm state to minor alarm state, the device identifies whether resending of minor resource depletion alarms is enabled. If the feature is disabled, the device does not issue additional minor alarms. If the feature is enabled, the device resends minor alarms periodically.

Resource depletion alarms can be sent to NETCONF, SNMP, and the information center to be encapsulated as NETCONF events, SNMP traps and informs, and log messages. For more

information, see NETCONF, SNMP, and information center in *Network Management and Monitoring Configuration Guide*.

Figure 3 Resource depletion alarms and alarm-removed notifications



Procedure

1. Enter system view.

   system-view

2. Set resource depletion thresholds.

   resource-monitor resource *resource-name* **slot** *slot-number* **cpu** *cpu-number* **by-percent minor-threshold** *minor-threshold* **severe-threshold** *severe-threshold*

   The default settings vary by resource type. Use the **display resource-monitor** command to display the resource depletion thresholds.

3. Specify destinations for resource depletion alarms.

   resource-monitor output { **netconf-event** | **snmp-notification** | **syslog** } *

   By default, resource depletion alarms are sent to NETCONF, SNMP, and the information center.

4. Enable resending of minor resource depletion alarms.

   resource-monitor minor resend enable

   By default, resending of minor resource depletion alarms is enabled.

Setting the temperature alarm thresholds

About this task

The device monitors its temperature based on the following thresholds:

- Low-temperature threshold.
- High-temperature warning threshold.
- High-temperature alarming threshold.

When the device temperature drops below the low-temperature threshold or reaches the high-temperature warning or alarming threshold, the device performs the following operations:

- Sends log messages and traps.

- Sets LEDs on the device panel.

Procedure

1. Enter system view.

   system-view

2. Configure the temperature alarm thresholds.

   temperature-limit slot *slot-number* **hotspot** *sensor-number lowlimit warninglimit* [*alarmlimit*]

   The defaults vary by temperature sensor model. To view the defaults, execute the **undo temperature-limit** and **display environment** commands in turn.

   The high-temperature alarming threshold must be higher than the high-temperature warning threshold, and the high-temperature warning threshold must be higher than the low-temperature threshold.

Disabling the notification of hardware resource alarms

About this task

Perform this task to disable the notification of hardware resource alarms.

Procedure

1. Enter system view.

   system-view

2. Disable the notification of hardware resource alarms.

   hardware-alarm { **syslog** | **trap** } **resource-type** { **all** | **board** | **bus** | **chip-channel** | **chip-port** | **device** | **disk** | **fan** | **interface** | **power** | **subcard** | **transceiver** | **voltage** } **disable**

   By default, the notification of all hardware resource alarms is enabled.

Configuring consistency error logging for software and hardware forwarding entries

About this task

To forward packets, the device generates hardware forwarding entries in forwarding chips and software forwarding entries in memory at the same time. The device automatically detects whether the hardware forwarding entries and the software forwarding entries are consistent.

The consistency error logging for software and hardware forwarding entries enables the device to collect statistics for consistency errors between software and hardware forwarding entries periodically. If the number of consistency errors reaches or exceeds the logging threshold, the device generates a log message.

Procedure

1. Enter system view.

   system-view

2. Set the logging threshold for consistency errors between software and hardware forwarding entries.

   parity-error consistency-check threshold *value*

By default, the logging threshold is 10.

3. Set the statistics period for consistency errors between software and hardware forwarding entries.

   parity-error consistency-check period *value*

   By default, the statistics period is 3600 seconds.

4. Enable consistency error logging for software and hardware forwarding entries.

   parity-error consistency-check log enable

   By default, consistency error logging is enabled for software and hardware forwarding entries.

Configuring parity error and ECC error logging for entries on forwarding chips

About this task

The device automatically detects parity errors and ECC errors in entries on forwarding chips. The parity error and ECC error logging feature collects parity errors and ECC error statistics periodically, and generates a log message if the number of parity errors and ECC errors reaches or exceeds the logging threshold.

Procedure

1. Enter system view.

   system-view

2. Set the parity error and ECC error statistics period for entries on forwarding chips.

   parity-error monitor period *value*

   By default, the parity error and ECC error statistics period for entries on forwarding chips is 60 seconds.

3. Set the parity error and ECC error logging threshold for entries on forwarding chips.

   parity-error monitor threshold *value*

   By default, the parity error and ECC error logging threshold for entries on forwarding chips is 5000.

4. Enable parity error and ECC error logging for entries on forwarding chips.

   parity-error monitor log enable

   By default, parity error and ECC error logging is enabled for entries on forwarding chips.

Configuring uncorrectable parity error and ECC error logging for entries on forwarding chips

About this task

The device automatically detects parity errors and ECC errors in entries on forwarding chips and tries to correct the error when an error is detected. The uncorrectable parity error and ECC error logging feature enables the device to collect statistics for uncorrectable parity errors and ECC errors periodically. If the number of uncorrectable parity errors and ECC errors reaches or exceeds the logging threshold, the device generates a log message.

Uncorrectable parity errors and ECC errors in entries on forwarding chips might affect system services. To remove the errors from the entries on forwarding chips, you can enable automatic system reboot upon generation of an uncorrectable parity error and ECC error log.

Procedure

1. Enter system view.

   system-view

2. Set the statistics period for uncorrectable parity error and ECC errors for entries on forwarding chips.

   parity-error unrecoverable period *value*

   By default, the statistics period for uncorrectable parity error and ECC errors for entries on forwarding chips is 60 seconds.

3. Set the logging threshold for uncorrectable parity errors and ECC errors for entries on forwarding chips.

   parity-error unrecoverable threshold *value*

   By default, the logging threshold for uncorrectable parity errors and ECC errors for entries on forwarding chips is 1.

4. Enable uncorrectable parity error and ECC error logging for entries on forwarding chips.

   parity-error unrecoverable log enable

   By default, uncorrectable parity error and ECC error logging is enabled for entries on forwarding chips

5. (Optional.) Enable automatic system reboot upon generation of an uncorrectable parity error and ECC error log.

   parity-error unrecoverable reboot

   By default, the device reboots automatically upon generation of an uncorrectable parity error and ECC error log.

Configuring device poweroff alarming

About this task

This feature enables the device to detect a device poweroff event and issue a poweroff alarm by sending SNMP notifications or log messages.

Restrictions and guidelines

This feature is available only for the S5590-EI switch series.

Procedure

1. Enter system view.

   system-view

2. Specify the source interface for sending the poweroff alarm.

   dying-gasp source *interface-type interface-number*

   By default, no source interface is specified. On an IPv4 network, the device uses the primary IPv4 address of the output interface for the route to the destination host as the source address. On an IPv6 network, the device selects a source IPv6 address as defined in RFC 3484.

3. Configure SNMP notification destination host settings.

   dying-gasp host { *ip-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] **snmp-trap version**{ **v1** | **v2c** | **v3** [**authentication** | **privacy**] } **securityname** *security-string*

   By default, no SNMP notification destination host settings are configured.

   You can configure the device to send poweroff alarm SNMP notifications to multiple destination hosts.

4. Configure log message destination host settings.

dying-gasp host { *ip-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] **syslog**

By default, no log message destination host settings are configured.

You can configure the device to send the poweroff alarm log messages to multiple destination hosts.

Ignoring absent fans

About this task

The device periodically polls the status of fans. Upon detecting an absent fan, the device sets the status of the fan to **Absent** and sends a log message and a trap. If the device has external fans and is well ventilated, you can use this command to configure the device to ignore absent fans. Then, the device does not send log messages or traps upon detecting absent fans.

Restrictions and guidelines

CAUTION:

This feature suppresses fan absence notifications and might cause device reboot and even device damage if the device is not well ventilated.

Before and after you configure this feature, make sure the device is well ventilated.

Procedure

1. Enter system view.

   system-view

2. Set the fan operating mode.

   fan ignore-absent

   By default, the devices does not ignore absent fans.

Verifying and diagnosing transceiver modules

Verifying transceiver modules

About this task

You can use one of the following methods to verify the genuineness of a transceiver module:

- Display the key parameters of a transceiver module, including its transceiver type, connector type, central wavelength of the transmit laser, transfer distance, and vendor name.
- Display its electronic label. The electronic label is a profile of the transceiver module and contains the permanent configuration, including the serial number, manufacturing date, and vendor name. The data was written to the transceiver module or the device's storage component during debugging or testing of the transceiver module or device.

The device regularly checks transceiver modules for their vendor names. If a transceiver module does not have a vendor name or the vendor name is not H3C, the device repeatedly outputs traps and log messages. For information about logging rules, see *Network Management and Monitoring Configuration Guide*.

Procedure

To verify transceiver modules, execute the following commands in any view:

- Display the key parameters of transceiver modules.

display transceiver interface [*interface-type interface-number*]

- Display the electrical label information of transceiver modules.

  display transceiver manuinfo interface [*interface-type interface-number*]

Diagnosing transceiver modules

About this task

The device provides the following diagnosis functions for transceiver modules:

The device provides the alarm and digital diagnosis functions for transceiver modules. When a transceiver module fails or is not operating correctly, you can perform the following tasks:

- Display transceiver alarms. The device check the alarms that exist on transceiver modules to identify the fault source.
- Display the current values of the digital diagnosis parameters on transceiver modules, that is, digital diagnosis functions for transceiver modules. The device can examine the key parameters monitored by the digital diagnosis function, including the temperature, voltage, laser bias current, TX power, and RX power.
- Display diagnosis information in an eye diagram. This function simulates the quality of data transmission from professional testing instruments. It obtains the sampling values of received data on chips and outputs the diagnostic results of transceiver modules in an eye diagram. This facilitates users in analyzing and evaluating the data transmission quality of transceiver modules.

Procedure

To diagnose transceiver modules, execute the following commands in any view:

- Display transceiver alarms.

  display transceiver alarm interface [*interface-type interface-number*]

- Display the current values of the digital diagnosis parameters on transceiver modules.

  display transceiver diagnosis interface [*interface-type interface-number*]

- Display diagnostic information for transceiver modules in an eye diagram.

  display transceiver eyediagram-diagnosis interface [*interface-type interface-number*]

Enabling alarm for third-party transceiver modules

About this task

Upon installation of a transceiver module, the system reads vendor information from it to determine whether it is a third-party transceiver module. If it is a third-party module, the device sends a notification and a log message to notify the user of replacing it with an H3C transceiver module. You can use **display transceiver manuinfo** to obtain vendor information for a transceiver module.

Procedure

1. Enter system view.

   system-view

2. Enable alarm for third-party transceiver modules.

   transceiver third-party alarm enable

By default, alarm is enabled for third-party transceiver modules

Configuring transceiver monitoring

About this task

After transceiver monitoring is enabled, the device samples the parameters of transceiver modules periodically, including the input power and output power of transceiver modules. If a sampled value reaches the alarm threshold, the device generates a log to notify users.

Restrictions and guidelines

Procedure

1. Enter system view.
   `system-view`
2. Set the transceiver monitoring interval.
   `transceiver monitor interval` *`interval`*
   By default, the transceiver monitoring interval is 600 seconds.
3. Enable transceiver monitoring.
   `transceiver monitor enable`
   By default, transceiver monitoring is disabled.

Configuring transceiver health check

About this task

The transceiver health is calculated based on the weights of five diagnosis parameters of transceiver modules, including temperature, voltage, bias current, transmit power, and receive power.

Transceiver health check enables the device to automatically check transceiver module status, and calculate transceiver health and diagnosis parameter health based on the configured weights of diagnosis parameters.

After transceiver health check is enabled and the weights for transceiver health calculation are configured, you can use the **`display transceiver health`** command to view transceiver health information. A greater health value indicates a healthier status.

Restrictions and guidelines

The sum of the weights of temperature, voltage, bias current, transmit power, and receive power for transceiver health calculation must be 100.

Procedure

1. Enter system view.
   `system-view`
2. Enable transceiver health check.
   `transceiver health check enable`
   By default, transceiver health check is disabled.
3. Configure the weights for transceiver health calculation.
   `transceiver health weight temp` *`temp-weight`* **`voltage`** *`voltage-weight`* **`bias`** *`bias-weight`* **`rxpower`** *`rxpower-weight`* **`txpower`** *`txpower-weight`*
   The default settings are as follows:
   - Temperature weight: 15.

- Voltage weight: 5.
- Bias current weight: 30.
- Receive power weight: 10.
- Transmit power weight: 40.

Enabling a lane on a transceiver module

Restrictions and guidelines

Perform this task only to verify that a transceiver module functions correctly.

Disabling a lane on a transceiver module will stop signal transmission on the lane.

The lane enablement setting is saved in a register on the transceiver module. It is not saved to the configuration file.

This feature is supported only on an SFP transceiver module.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Enable a lane on the transceiver module.
   transceiver lane [*lane-number*] **enable**
   By default, the lanes on the transceiver module are enabled.

Setting the FEC mode for a transceiver module

About this task

The forward error correction (FEC) feature corrects packet errors to improve transmission quality. It attaches correction information to a packet at the sending end, and corrects error codes generated during transmission at the receiving end based on the correction information. You can set the FEC mode as needed.

Restrictions and guidelines

Make sure you set the same FEC mode for both interfaces of a link.

Only the LMQ8511B-PC+-H3C transceiver module in the QSFP-100G-BIDI-MM850 series supports this feature.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Set the FEC mode for the transceiver module.
   transceiver fec mode { **kp4** | **berlinetta** }
   By default, the FEC mode for a transceiver module varies by device model.

Monitoring optical power of transceiver modules

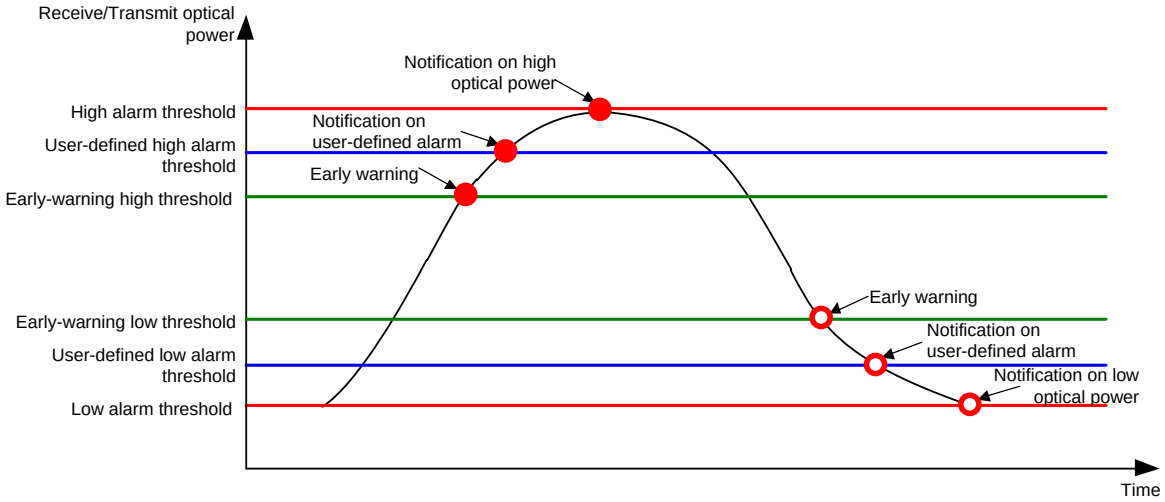
Configuring RX and TX power thresholds for a transceiver module

About this task

To detect optical power anomalies and take actions promptly, you can set alarm thresholds between the default early-warning thresholds and alarm thresholds for RX and TX power of transceiver modules, as shown in Figure 4. If the RX/TX power of a transceiver module exceeds or drops below a user-defined threshold, the system generates an alarm message.

You can execute the **display transceiver diagnosis** command to obtain the current RX/TX power, early-warning thresholds, user-defined thresholds, and alarm thresholds.

Figure 4 Configuring RX and TX power thresholds for a transceiver module



Restrictions and guidelines

This feature is available only for digital-diagnosis-capable the SFP, SFP+, SFP28, QSFP+, QSFP28 transceiver modules. It is not available for the 400G ZR, 400G ZR+, or fiber-to-copper transceiver modules.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Set the high RX power threshold for a transceiver module.
   transceiver rx-power high-threshold *power-value*
   By default, no high RX power threshold is set for a transceiver module.
4. (Optional.) Set the low RX power threshold for a transceiver module.
   transceiver rx-power low-threshold *power-value*
   By default, no low RX power threshold is set for a transceiver module.
5. (Optional.) Set the high TX power threshold for a transceiver module.

transceiver tx-power high-threshold *power-value*

By default, no high TX power threshold is set for a transceiver module.

6. (Optional.) Set the low TX power threshold for a transceiver module.

   transceiver tx-power low-threshold *power-value*

   By default, no low TX power threshold is set for a transceiver module.

Configuring RX and TX power change thresholds for a transceiver module

About this task

The device sets the initial RX optical power as the baseline for a transceiver module after the transceiver module starts up and operates correctly for the first time. After you execute the **transceiver rx-power change-threshold** or **transceiver tx-power change-threshold** command, the system checks the RX/TX optical power periodically. If the RX/TX optical power changes exceed the user-defined thresholds compared to the baseline, the system identifies it as significant changes and generates alarms.

Restrictions and guidelines

This feature is available only for digital-diagnosis-capable SFP, SFP+, SFP28, QSFP+, and QSFP28 transceiver modules. It is not available for the 400G ZR, 400G ZR+, or fiber-to-copper transceiver modules.

Procedure

1. Enter system view.

   system-view

2. Enter Ethernet interface view.

   interface *interface-type interface-number*

3. (Optional.) Configure the RX power change threshold for a transceiver module.

   transceiver rx-power change-threshold *power-value*

   By default, the RX power change threshold for a transceiver module is 5.00 dBm.

4. (Optional.) Configure the TX power change threshold for a transceiver module.

   transceiver tx-power change-threshold *power-value*

   By default, the TX power change threshold for a transceiver module is 4.00 dBm.

Scheduling a task

About task scheduling

You can schedule the device to automatically execute a command or a set of commands without administrative interference.

You can configure a periodic schedule or a non-periodic schedule. A non-periodic schedule is not saved to the configuration file and is lost when the device reboots. A periodic schedule is saved to the startup configuration file and is automatically executed periodically.

Restrictions and guidelines

- The default system time is always restored at reboot. To make sure a task schedule can be executed as expected, reconfigure the system time or configure NTP after you reboot the

device. For more information about NTP, see *Network Management and Monitoring Configuration Guide*.

- To assign a command (command A) to a job, you must first assign the job the command or commands for entering the view of command A.
- Make sure all commands in a schedule are compliant to the command syntax. The system does not check the syntax when you assign a command to a job.
- A schedule cannot contain any one of these commands: **telnet**, **ftp**, **ssh2**, and **monitor process**.
- A schedule does not support user interaction. If a command requires a yes or no answer, the system always assumes that a **Y** or **Yes** is entered. If a command requires a character string input, the system assumes that either the default character string (if any) or a null string is entered.
- A schedule is executed in the background, and no output (except for logs, traps, and debug information) is displayed for the schedule.

Procedure

1. Enter system view.
   system-view
2. Create a job.
   scheduler job *job-name*
3. Assign a command to the job.
   command *id command*
   By default, no command is assigned to a job.
   You can assign multiple commands to a job. A command with a smaller ID is executed first.
4. Exit to system view.
   quit
5. Create a schedule.
   scheduler schedule *schedule-name*
6. Assign a job to the schedule.
   job *job-name*
   By default, no job is assigned to a schedule.
   You can assign multiple jobs to a schedule. The jobs will be executed concurrently.
7. Assign user roles to the schedule.
   user-role *role-name*
   By default, a schedule has the user role of the schedule creator.
   You can assign a maximum of 64 user roles to a schedule. A command in a schedule can be executed if it is permitted by one or more user roles of the schedule.
8. Specify the execution time for the schedule.
   Choose one option as needed:
   - Execute the schedule at specific points of time.
     time at *time date*
     time once at *time* [**month-date** *month-day* | **week-day** *week-day*&<1-7>]
   - Execute the schedule after a period of time.
     time once delay *time*
   - Execute the schedule at the specified time on every specified day in a month or week.

time repeating at *time* [**month-date** [*month-day* | **last**] | **week-day** *week-day*&<1-7>]

- Execute the schedule periodically from the specified time on.

  time repeating [**at** *time* [*date*]] **interval** *interval*

By default, no execution time is specified for a schedule.

The **time** commands overwrite each other. The most recently executed command takes effect.

9. (Optional.) Set the schedule log file size limit.

   scheduler logfile size *value*

   By default, the schedule log file size limit is 16 KB.

   The schedule log file stores log messages for execution results of commands in jobs. After the limit is reached, the system deletes the oldest log messages to store the new log messages. If the remaining space of the log file is not enough for a single log message, the system truncates the message and does not store the extra part.

Example: Scheduling a task

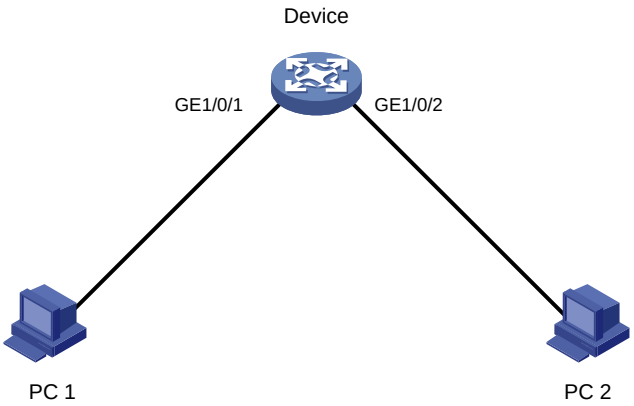
Network configuration

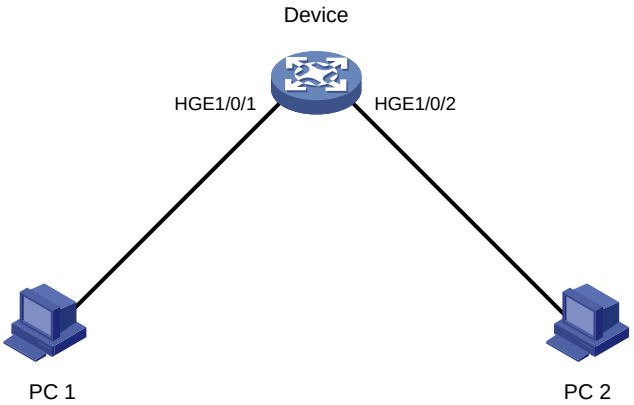
As shown in Figure 5, two interfaces of the device are connected to users.

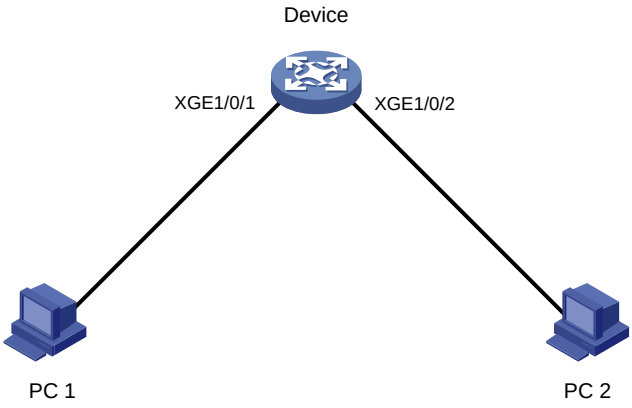
To save energy, configure the device to perform the following operations:

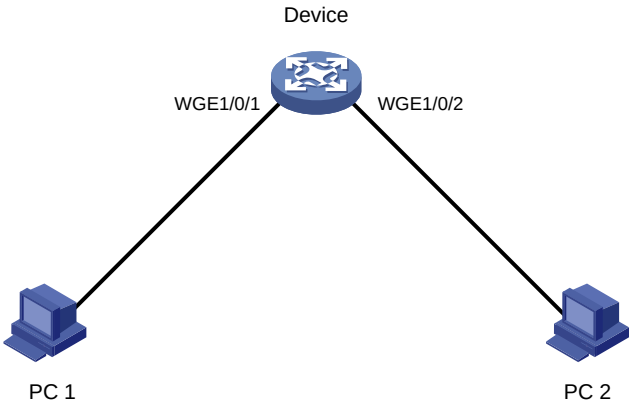
- Enable the interfaces at 8:00 a.m. every Monday through Friday.
- Disable the interfaces at 18:00 every Monday through Friday.

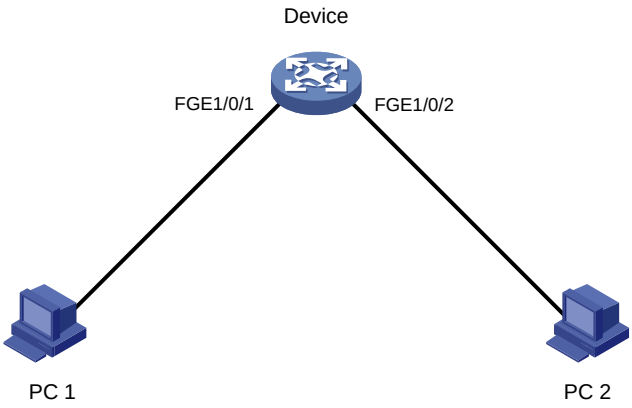
Figure 5 Network diagram

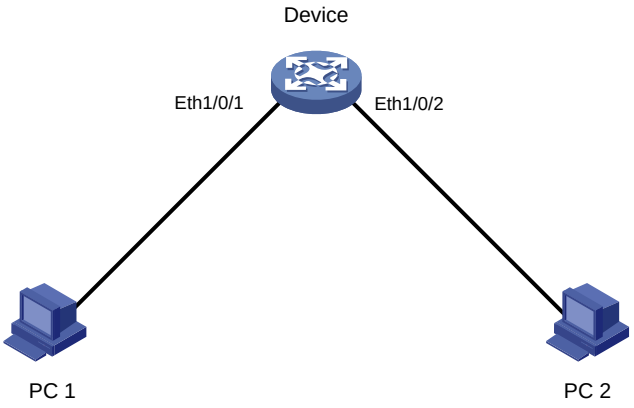


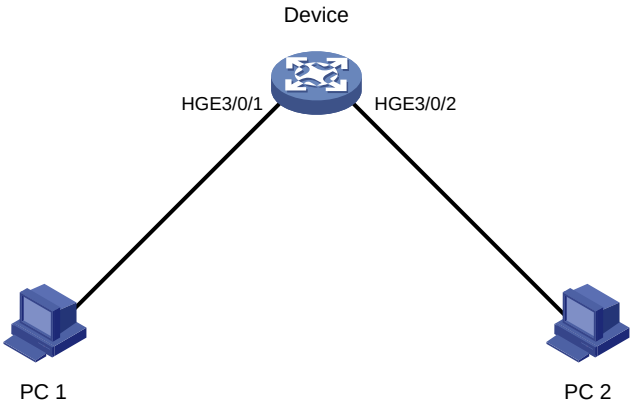


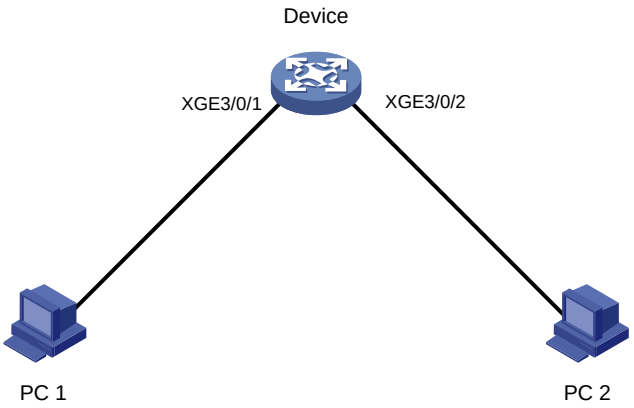

Device
XGE1/0/1
XGE1/0/2
PC 1
PC 2

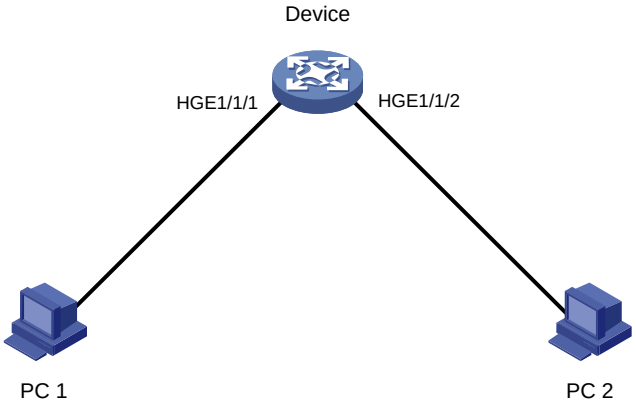

Device
WGE1/0/1
WGE1/0/2
PC 1
PC 2

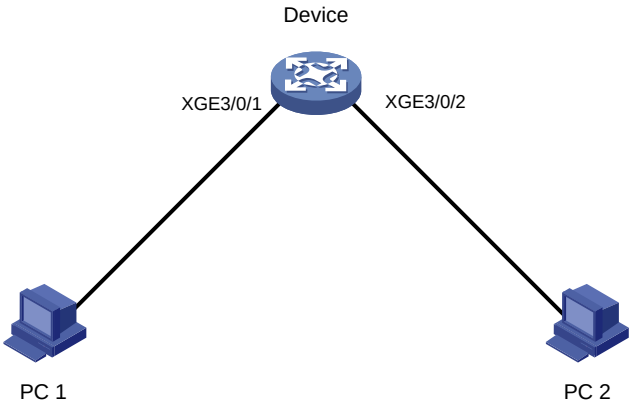

Device
FGE1/0/1
FGE1/0/2
PC 1
PC 2


Device
Eth1/0/1
Eth1/0/2
PC 1
PC 2


Device
HGE3/0/1
HGE3/0/2
PC 1
PC 2


Device
XGE3/0/1
XGE3/0/2
PC 1
PC 2





Procedure

Enter system view.

```
<Sysname> system-view
```

Configure a job for disabling interface GigabitEthernet 1/0/1.

```
[Sysname] scheduler job shutdown-GigabitEthernet1/0/1
[Sysname-job-shutdown-GigabitEthernet1/0/1] command 1 system-view
[Sysname-job-shutdown-GigabitEthernet1/0/1] command 2 interface gigabitethernet 1/0/1
[Sysname-job-shutdown-GigabitEthernet1/0/1] command 3 shutdown
[Sysname-job-shutdown-GigabitEthernet1/0/1] quit
```

Configure a job for enabling interface GigabitEthernet 1/0/1.

```
[Sysname] scheduler job start-GigabitEthernet1/0/1
[Sysname-job-start-GigabitEthernet1/0/1] command 1 system-view
[Sysname-job-start-GigabitEthernet1/0/1] command 2 interface gigabitethernet 1/0/1
[Sysname-job-start-GigabitEthernet1/0/1] command 3 undo shutdown
[Sysname-job-start-GigabitEthernet1/0/1] quit
```

Configure a job for disabling interface GigabitEthernet 1/0/2.

```
[Sysname] scheduler job shutdown-GigabitEthernet1/0/2
[Sysname-job-shutdown-GigabitEthernet1/0/2] command 1 system-view
[Sysname-job-shutdown-GigabitEthernet1/0/2] command 2 interface gigabitethernet 1/0/2
[Sysname-job-shutdown-GigabitEthernet1/0/2] command 3 shutdown
[Sysname-job-shutdown-GigabitEthernet1/0/2] quit
```

Configure a job for enabling interface GigabitEthernet 1/0/2.

```
[Sysname] scheduler job start-GigabitEthernet1/0/2
[Sysname-job-start-GigabitEthernet1/0/2] command 1 system-view
[Sysname-job-start-GigabitEthernet1/0/2] command 2 interface gigabitethernet 1/0/2
[Sysname-job-start-GigabitEthernet1/0/2] command 3 undo shutdown
[Sysname-job-start-GigabitEthernet1/0/2] quit
```

Configure a periodic schedule for enabling the interfaces at 8:00 a.m. every Monday through Friday.

```
[Sysname] scheduler schedule START-pc1/pc2
[Sysname-schedule-START-pc1/pc2] job start-GigabitEthernet1/0/1
[Sysname-schedule-START-pc1/pc2] job start-GigabitEthernet1/0/2
[Sysname-schedule-START-pc1/pc2] time repeating at 8:00 week-day mon tue wed thu fri
[Sysname-schedule-START-pc1/pc2] quit
```

Configure a periodic schedule for disabling the interfaces at 18:00 every Monday through Friday.

```
[Sysname] scheduler schedule STOP-pc1/pc2
[Sysname-schedule-STOP-pc1/pc2] job shutdown-GigabitEthernet1/0/1
[Sysname-schedule-STOP-pc1/pc2] job shutdown-GigabitEthernet1/0/2
[Sysname-schedule-STOP-pc1/pc2] time repeating at 18:00 week-day mon tue wed thu fri
[Sysname-schedule-STOP-pc1/pc2] quit
```

Verifying the configuration

Display the configuration information of all jobs.

```
[Sysname] display scheduler job
Job name: shutdown-GigabitEthernet1/0/1
 system-view
 interface gigabitethernet 1/0/1
 shutdown

Job name: shutdown-GigabitEthernet1/0/2
 system-view
 interface gigabitethernet 1/0/2
 shutdown

Job name: start-GigabitEthernet1/0/1
 system-view
 interface gigabitethernet 1/0/1
 undo shutdown

Job name: start-GigabitEthernet1/0/2
 system-view
 interface gigabitethernet 1/0/2
 undo shutdown
```

Display the schedule information.

```
[Sysname] display scheduler schedule
Schedule name        : START-pc1/pc2
Schedule type        : Run on every Mon Tue Wed Thu Fri at 08:00:00
Start time           : Wed Sep 28 08:00:00 2011
Last execution time  : Wed Sep 28 08:00:00 2011
Last completion time : Wed Sep 28 08:00:03 2011
```

```
Execution counts     : 1
------------------------------------------------------------------------
Job name                                          Last execution status
start-GigabitEthernet1/0/1                                Successful
start-GigabitEthernet1/0/2                                Successful

Schedule name        : STOP-pc1/pc2
Schedule type        : Run on every Mon Tue Wed Thu Fri at 18:00:00
Start time           : Wed Sep 28 18:00:00 2011
Last execution time  : Wed Sep 28 18:00:00 2011
Last completion time : Wed Sep 28 18:00:01 2011
Execution counts     : 1
------------------------------------------------------------------------
Job name                                          Last execution status
shutdown-GigabitEthernet1/0/1                             Successful
shutdown-GigabitEthernet1/0/2                             Successful
```

Display schedule log information.

```
[Sysname] display scheduler logfile
Job name        : start-GigabitEthernet1/0/1
Schedule name   : START-pc1/pc2
Execution time  : Wed Sep 28 08:00:00 2011
Completion time : Wed Sep 28 08:00:02 2011
--------------------------------- Job output ---------------------------------
<Sysname>system-view
System View: return to User View with Ctrl+Z.
[Sysname]interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1]undo shutdown

Job name        : start-GigabitEthernet1/0/2
Schedule name   : START-pc1/pc2
Execution time  : Wed Sep 28 08:00:00 2011
Completion time : Wed Sep 28 08:00:02 2011
--------------------------------- Job output ---------------------------------
<Sysname>system-view
System View: return to User View with Ctrl+Z.
[Sysname]interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2]undo shutdown

Job name        : shutdown-GigabitEthernet1/0/1
Schedule name   : STOP-pc1/pc2
Execution time  : Wed Sep 28 18:00:00 2011
Completion time : Wed Sep 28 18:00:01 2011
--------------------------------- Job output ---------------------------------
<Sysname>system-view
System View: return to User View with Ctrl+Z.
[Sysname]interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1]shutdown
```

```
Job name        : shutdown-GigabitEthernet1/0/2
Schedule name   : STOP-pc1/pc2
Execution time  : Wed Sep 28 18:00:00 2011
Completion time : Wed Sep 28 18:00:01 2011
--------------------------------- Job output ---------------------------------
<Sysname>system-view
System View: return to User View with Ctrl+Z.
[Sysname]interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2]shutdown
```

Locating devices

About device locating

The device provides LEDs for device locating. The **locator blink** *blink-time* command flashes the LEDs quickly for the specified period of time unless you execute the **locator blink stop** command or the specified flash duration expires.

Starting LED flashing

To start LED flashing, execute one of the following commands in user view:

locator [**slot** *slot-number*] **blink** *blink-time*

Stopping LED flashing

To stop LED flashing, execute one of the following commands in user view:

locator [**slot** *slot-number*] **blink stop**

Rebooting the device

About device reboot

The following device reboot methods are available:

- Schedule a reboot at the CLI, so the device automatically reboots at the specified time or after the specified period of time.
- Immediately reboot the device at the CLI.

  During the reboot process, the device performs the following operations:

  a. Resets all of its chips.

  b. Uses the BootWare to verify the startup software package, decompress the package, and load the images.

  c. Initializes the system.
- Power off and then power on the device. This method might cause data loss, and is the least-preferred method.

Using the CLI, you can reboot the device from a remote host.

Restrictions and guidelines for device reboot

For data security, the device does not reboot while it is performing file operations.

Rebooting the device immediately at the CLI

Prerequisites

Perform the following steps in any view:

1. Verify that the next-startup configuration file is correctly specified.

   display startup

   For more information about the **display startup** command, see *Fundamentals Command Reference*.
2. Verify that the startup image files are correctly specified.

   display boot-loader

   If one main startup image file is damaged or does not exist, you must specify another main startup image file before rebooting the device.

   For more information about the **display boot-loader** command, see *Fundamentals Command Reference*.
3. Save the running configuration to the next-startup configuration file.

   save

   To avoid configuration loss, save the running configuration before a reboot.

   For more information about the **save** command, see *Fundamentals Command Reference*.

Procedure

To reboot the device immediately at the CLI, execute one of the following commands in user view:

reboot [**slot** *slot-number*] [**force**]

CAUTION:

- A device reboot might result in a service outage. Before using this command, make sure you understand its impact on the live network.
- Use the **force** keyword to reboot the device only when the system is faulty or fails to start up normally. A forced device reboots might cause file system damage. Before using the **force** keyword to reboot the device, make sure you understand its impact.

Scheduling a device reboot

Restrictions and guidelines

The automatic reboot configuration takes effect on all member devices. It will be canceled if a master/subordinate switchover occurs.

The device supports only one device reboot schedule. If you execute the **scheduler reboot** command multiple times, the most recent configuration takes effect.

Procedure

To schedule a reboot, execute one of the following commands in user view:

- **scheduler reboot at** *time* [*date*]

- **scheduler reboot delay** *time*

By default, no device reboot time is specified.

CAUTION:
This task enables the device to reboot at a scheduled time, resulting in service interruption. Before using this command, make sure you understand its impact on the live network.

Restoring the factory-default configuration

About this task

If you want to use the device in a different scenario or you cannot troubleshoot the device by using other methods, use this task to restore the factory-default configuration.

This task does not delete **.bin** files.

Restrictions and guidelines

This feature is disruptive.

Procedure

1. Execute the following command in user view to restore the factory-default configuration for the device:
   restore factory-default
2. Reboot the device.
   reboot

CAUTION:
This command restores the device to the factory default settings. Before using this command, make sure you understand its impact on the live network.

When the command prompts you to choose whether to save the running configuration, enter **N**. If you choose to save the running configuration, the device loads the saved configuration at startup.

Display and maintenance commands for device management configuration

Execute **display** commands in any view. Execute the **reset scheduler logfile** and **display disk-information** commands in user view. Execute and **reset version-update-record** in system view.

Task	Command
Display information about active alarms.	**display alarm active** [**verbose**]
Display information about historical alarms.	**display alarm history** [**verbose**]
Display the system time, date, time zone, and daylight saving time.	**display clock**
Display the copyright statement.	**display copyright**

Task	Command
Display CPU usage statistics. The **core** keyword is mutually exclusive with the **control-plane** and **data-plane** keywords.	**display cpu-usage [control-plane | data-plane]** [**summary**] [**slot** *slot-number* [**cpu** *cpu-number* [**core** { *core-number* | **all** }]]]
Display CPU usage monitoring settings.	**display cpu-usage configuration** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display the historical CPU usage statistics in a coordinate system.	**display cpu-usage history** [**job** *job-id*] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display hardware information.	**display device** [**flash** | **usb**] [**slot** *slot-number* [**subslot** *subslot-number*] | **verbose**]
Display electronic label information for the device.	**display device manuinfo** [**slot** *slot-number*]
Display electronic label information for a fan tray.	**display device manuinfo slot** *slot-number* **fan** *fan-id*
Display electronic label information for a power module.	**display device manuinfo slot** *slot-number* **power** *power-id*
Display information about member device uptime starting from the latest startup.	**display device uptime**
Display or save diagnostic information for features and hardware modules.	**display diagnostic-information** [**hardware** | **ifmgr** | **infrastructure** | **l2** | **l3** | **service**] [**key-info**] [*filename*]
Display poweroff alarm destination host settings.	**display dying-gasp host**
Display device temperature information.	**display environment** [**slot** *slot-number*]
Display the operating states of fan trays.	**display fan** [**slot** *slot-number* [*fan-id*]] [**verbose**]
Display memory usage statistics.	**display memory** [**summary**] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display memory allocation failure information in the kernel.	**display kernel memory alloc-failure** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display memory fragment statistics in the kernel.	**display kernel memory fragment free** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display DMA memory usage statistics.	**display memory dma** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display memory alarm thresholds and statistics.	**display memory-threshold** [**slot** *slot-number* [**cpu** *cpu-number*]]

Task	Command
Display DMA memory alarm information.	**display memory-threshold dma** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display power module information.	**display power** [**slot** *slot-number* [*power-id* | **verbose**]]
Display resource monitoring information.	**display resource-monitor** [**resource** *resource-name*] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display job configuration information.	**display scheduler job** [*job-name*]
Display job execution log information.	**display scheduler logfile**
Display the automatic reboot schedule.	**display scheduler reboot**
Display schedule information.	**display scheduler schedule** [*schedule-name*]
Display system heath status information.	**display system health** [**slot** *slot-number*]
Display historical system heath status change information.	**display system health history** [**slot** *slot-number*]
Display system stability and status information.	**display system stable state**
Display system working mode information.	**display system-working-mode**
Display health information of transceiver modules.	**display transceiver health interface** [*interface-type interface-number*]
Display power information for transceiver modules.	**display transceiver power interface** [*interface-type interface-number*]
Display transceiver module status information. Only SFP transceiver modules support this command.	**display transceiver status interface** [*interface-type interface-number*]
Display system version information.	**display version**
Display startup software image upgrade records.	**display version-update-record**
Clear job execution log information.	**reset scheduler logfile**
Clear startup software image upgrade records.	**reset version-update-record**

Contents

Managing target configuration 1
About target configuration 1
Enabling target configuration management 1
Committing the target configuration command lines 2
Display and maintenance commands for target configuration 3

Managing target configuration

About target configuration

Target configuration enables you to configure and verify intended settings before merging them into the running configuration.

You can configure target configuration command lines in the following views:

- Private-mode system view and its lower-level views.
- Exclusive-mode system view and its lower-level views.

Commands available for real-time configuration are all available for target configuration.

Support for command lines in private-mode system view and in exclusive-mode system view depends on the device model. In private-mode system view and in exclusive-mode system, you do not need to commit hardware-related command lines and action command lines, because they take effect immediately.

Private-mode views and exclusive-mode views have an extra tilde sign (~) in their prompts as compared to their counterpart real-time-mode views. After you execute a command that does not take effect in real time, the prompt will have a leading asterisk sign (*) to indicate that the command has not taken effect in the current mode. After the **commit** command is executed to commit the configuration, the prompt will have a leading tilde sign (~).

For more information about views, see "Using the CLI."

Enabling target configuration management

About this task

To save memory, you can disable target configuration management by executing the **undo commit enable** command. This can reduce unnecessary resource consumption and improve system efficiency.

Restrictions and guidelines

After you enable target configuration management, you cannot enter private-mode or exclusive-mode system view by executing the **system-view** command.

If a user already enters private-mode system view, executing the **undo commit enable** command will force that user to exit private-mode system view.

Procedure

1. Enter real-time-mode system view.

   system-view
2. Enable target configuration management.

   commit enable

   By default, target configuration management is enabled.

Committing the target configuration command lines

About this task

A commit operation commits all target configuration command lines. After a successful commit operation, the system creates a configuration rollback point.

Restrictions and guidelines

As a best practice, do not execute the **commit** command frequently. The system creates a configuration rollback point after each successful commit operation. Executing the command frequently increases device load and might cause commit failures.

To save the running configuration while committing the target configuration command lines, you can use one of the following methods:

- Specify the **save-running** *filename* option when you use the **commit** command to commit the target configuration command lines. The system saves the running configuration to the specified file while committing the target configuration command lines.
- Use the **configuration commit auto-save path** command to enable automatic saving for the target configuration command lines. This command enables the device to automatically save the running configuration to a configuration file in the specified directory each time the **commit** command is executed. The directory is specified by using the *directory* argument. The prefix for the configuration file is specified by using the **filename-prefix** *filename-prefix* option.

The automatic saving feature takes effect if you do not specify the **save-running** *filename* option when executing the **commit** command. If you specify this option when executing the **commit** command, the automatic saving feature cannot take effect. The target configuration command lines are saved to the file specified by using this option.

Procedure

1. (Optional.) Specify the maximum number of rollback points:
   a. Enter real-time-mode system view.
      system-view
   b. Set the maximum number of rollback points.
      configuration commit max-number *max-number*
      By default, the maximum number of rollback points is 100.
      To create a new rollback point when the number of rollback points reaches the maximum, the system deletes the oldest rollback point.
   c. Return to user view.
      Quit
2. (Optional.) Enable automatic saving for the target configuration command lines and configure the related parameters:
   a. Enter real-time-mode system view.
      system-view
   b. Enable automatic saving for the target configuration command lines and configure the related parameters.
      configuration commit auto-save path *directory* **filename-prefix** *filename-prefix* [**user** *user-name* [**password** { **cipher** | **simple** } *string*] | **vpn-instance** *vpn-instance-name*] *
      By default, automatic saving is disabled for the target configuration command lines.

c. Return to user view.

abort

3. Enter private-mode system view or exclusive-mode system view.

system-view { **exclusive** | **private** }

4. Configure features as required.

For more information about configuring features, see the configuration guides.

5. (Optional.) Display target configuration command lines.

display target-configuration

Incorrect target configuration command lines will affect the operation of the network. Use this command to verify that all target configuration command lines are correct.

6. (Optional.) Refresh uncommitted target configuration.

refresh target-configuration

7. Commit the target configuration command lines.

commit [**best-effort**] [**force**] [**label** *labelname*] [**save-running** *filename*] [**confirmed** [*seconds* | **minutes** *minutes*]] [**show-error**] [**clear-error**] [**description** *text*]

8. (Optional.) Roll back the running configuration.

configuration replace { **file** *filename* | **last** *number* | **commit-id** *commit-id* | **label** *labelname* }

CAUTION:

This operation will cause settings not at the rollback point to be lost, which might cause service interruption. When you perform configuration rollback, make sure you fully understand its impact on your network.

To reclaim target configuration command lines that have been committed, use this command to roll back the running configuration as required. For more information about this command, see configuration file management commands in *Fundamentals Command Reference*.

Display and maintenance commands for target configuration

Execute the following commands in private-mode system view, exclusive-mode system view, or their lower-level views:

- **display target-configuration**
- **reset target-configuration**

Task	Command
Display the target configuration command lines.	**display target-configuration** [**diff** | **failure** [**brief**] | **merge** | **success**]
Clear the target configuration command lines.	**reset target-configuration** [**command** [*text*]]

Contents

Using Tcl 1

About Tcl 1
Restrictions and guidelines: Tcl 1
Using Tcl commands to configure the device 1
Restrictions and guidelines 1
Procedure 1
Executing Comware commands in Tcl configuration view 2
About executing Comware commands in Tcl configuration view 2
Restrictions and guidelines 2
Procedure 2

Using Tcl

About Tcl

Comware 7 provides a built-in tool command language (Tcl) interpreter. From user view, you can use the **tclsh** command to enter Tcl configuration view to execute the following commands:

- Tcl 8.5 commands.
- Comware commands.

The Tcl configuration view is equivalent to the user view. You can use Comware commands in Tcl configuration view in the same way they are used in user view.

Restrictions and guidelines: Tcl

To return from a subview under Tcl configuration view to the upper-level view, use the **quit** command.

To return from a subview under Tcl configuration view to the Tcl configuration view, press **Ctrl+Z**.

Using Tcl commands to configure the device

Restrictions and guidelines

When you use Tcl to configure the device, follow these restrictions and guidelines:

- You can apply Tcl environment variables to Comware commands.
- No online help information is provided for Tcl commands.
- You cannot press **Tab** to complete an abbreviated Tcl command.
- Make sure the Tcl commands can be executed correctly.
- As a best practice, log in through Telnet or SSH. You cannot stop Tcl commands by using a shortcut key or a CLI command. If a problem occurs when the Tcl commands are being executed, you can terminate the process by closing the connection if you logged in through Telnet or SSH. If you logged in from the console port, you must perform one of the following tasks:
  - Restart the device.
  - Log in to the device by using a different method, and use the **free line** command to release the console line. For more information about the command, see *Fundamentals Command Reference*.
- You can press **Ctrl+D** to abort Tcl command **read stdin**.

Procedure

1. Enter Tcl configuration view from user view.
   tclsh
2. Execute a Tcl command.
   Tcl command
3. Return from Tcl configuration view to user view.

- **tclquit**
- **quit**

Executing Comware commands in Tcl configuration view

About executing Comware commands in Tcl configuration view

To execute a Comware command in Tcl configuration view, use one of the following methods:

- Enter the Comware command directly. If a Tcl command uses the same command string as the Comware command, the Tcl command is executed.
- Prefix the Comware command with the **cli** keyword. If a Tcl command uses the same command string as the Comware command, the Comware command is executed.

Restrictions and guidelines

Follow these restrictions and guidelines when you execute Comware commands in Tcl configuration view:

- To specify a string enclosed in quotation marks (") or braces ({ and }), you must use the escape character (\) before the quotation marks or braces. For example, to specify **"a"** as the description for an interface, you must enter **description \"a\"**. If you enter **description "a"**, the description is **a**.
- For Comware commands, you can enter **?** to obtain online help or press **Tab** to complete an abbreviated command. For more information, see "Using the CLI."
- The **cli** command is a Tcl command, so you cannot enter **?** to obtain online help or press **Tab** to complete an abbreviated command.
- Successfully executed Comware commands are saved to command history buffers. You can use the upper arrow or lower arrow key to obtain executed commands.
- To execute multiple Comware commands in one operation, use one of the following methods:
  - Enter multiple Comware commands separated by semi-colons to execute the commands in the order they are entered. For example, **vlan 2;description Tech**.
  - Specify multiple Comware commands for the **cli** command, quote them, and separate them by a space and a semicolon. For example, **cli "vlan 2 ;description Tech"**.
  - Specify one Comware command for each **cli** command and separate them by a space and a semicolon. For example, **cli vlan 2 ;cli description Tech**.

Procedure

1. Enter Tcl configuration view

   tclsh
2. Execute Comware commands.
   - Execute Comware commands directly.

     Command
   - Execute Comware commands by using the **cli** command.

     cli *command*

3. Return from Tcl configuration view to user view.
    - **`tclquit`**
    - **`quit`**

Contents

Using Python 1

About Python 1
Python tasks at a glance 1
Python2 tasks at a glance 1
Executing a Python2 script 1
Entering the Python2 shell 1
Importing and using the extended Python API 1
Importing the entire extended API and using the API 2
Importing an extended API function and using the function 2
Exiting the Python shell 2
Python usage examples 2
Example: Using a Python2 script for device configuration 2

Comware 7 extended Python API 4

channel 4
CLI 5
get_error 5
get_output 6
get_self_slot 6
get_slot_info 6
get_slot_range 7
get_standby_slot 7
Transfer 8
send 9
SYSLOG 9

Using Python

About Python

Python is a simple and powerful programming language, with high-level data structures. It provides an easy and efficient method for object-oriented programming. Its concise syntax, support for dynamic input, and interpretive nature make it an ideal scripting language for rapid application development across many platforms and areas.

Python tasks at a glance

Python2 tasks at a glance

Comware 7 provides a built-in Python2 interpreter. You can use Python2 to perform the following tasks:

- Executing a Python2 script

  You can execute Python2 scripts to implement automatic device configuration.
- Enter Python2 shell to configure the device by using the following items:
  - Python 2.7 commands.
  - Python 2.7 standard API.
  - Extended API. For more information about the extended API, see "Comware 7 extended Python API."

  Perform the following tasks accordingly:

  a. Entering the Python2 shell

  b. Importing and using the extended Python API

  c. Exiting the Python shell

Executing a Python2 script

To execute a Python2 script, use the following command in user view:

python *filename*

Entering the Python2 shell

To enter the Python2 shell from user view, execute the following command:

python

Importing and using the extended Python API

To use the extended Python API, you must first import the API to Python.

Importing the entire extended API and using the API

Procedure

1. Import the entire extended API in the Python shell.
   `import platformtools`
2. Execute an extended API function.
   `platformtools.`*`api`*

Example

In the Python2 shell, use extended API function **`Transfer`** to download the **test.cfg** file from TFTP server 192.168.1.26.

```
<Sysname> python
>>> import platformtools
>>> platformtools.Transfer('tftp', '192.168.1.26', 'test.cfg', 'flash:/test.cfg',
user='', password='')
<platformtools.Transfer object at 0xb7eab0e0>
```

Importing an extended API function and using the function

Procedure

1. Import an extended API function in the Python shell.
   `from platformtools import` *`api-name`*
2. Execute an extended API function.
   `api-function`

Example

In the Python2 shell, use extended API function **`Transfer`** to download the **test.cfg** file from TFTP server 192.168.1.26.

```
<Sysname> python
>>> from platformtools import Transfer
>>> Transfer('tftp', '192.168.1.26', 'test.cfg', 'flash:/test.cfg', user='',
password='')
<platformtools.Transfer object at 0xb7e5e0e0>
```

Exiting the Python shell

To exit the Python shell, execute the following command in the Python shell.

`exit()`

Python usage examples

Example: Using a Python2 script for device configuration

Network configuration

Use a Python2 script to perform the following tasks:

- Download configuration files **main.cfg** and **backup.cfg** to the device.

- Configure the files as the main and backup configuration files for the next startup.

Figure 1 Network diagram



Procedure

Use a text editor on the PC to configure Python2 script **test.py** as follows:

```
#!usr/bin/python
import platformtools

platformtools.Transfer('tftp', '192.168.1.26', 'main.cfg', 'flash:/main.cfg')
platformtools.Transfer('tftp', '192.168.1.26', 'backup.cfg', 'flash:/backup.cfg')
platformtools.CLI('startup saved-configuration flash:/main.cfg main ;startup
saved-configuration flash:/backup.cfg backup')
```

Use TFTP to download the script to the device.

```
<Sysname> tftp 192.168.1.26 get test.py
```

Execute the script.

```
<Sysname> python flash:/test.py
<Sysname>startup saved-configuration flash:/main.cfg main
Please wait...... Done.
<Sysname>startup saved-configuration flash:/backup.cfg backup
Please wait...... Done.
```

Verifying the configuration

Display startup configuration files.

```
<Sysname> display startup
 Current startup saved-configuration file: flash:/startup.cfg
 Next main startup saved-configuration file: flash:/main.cfg
 Next backup startup saved-configuration file: flash:/backup.cfg
```

Comware 7 extended Python API

The Comware 7 extended Python API is compatible with the Python syntax.

This document uses Python2 shell as an example.

channel

Use **channel** to execute Comware 7 CLI commands and create channel objects.

Syntax

channel(*command*='')

Parameters

command: Specifies the commands to be executed. To enter multiple commands, use a space and a semicolon (;) as the delimiter. To enter a command in a view other than user view, you must first enter the commands used to enter the view. For example, you must enter **'system-view ;display interface brief | include UP'** to execute the **display interface brief | include UP** command.

Usage guidelines

This API function supports only Comware 7 commands. It does not support Linux, Python, or Tcl commands.

The API function is used for automated testing.

The list objects returned by the channel function might contain hexadecimal hidden characters such as 0x08. To hide the hexadecimal hidden characters, you can use the CLI function instead, or use the **for** loop to output the execution result of the channel function, for example:

```
<Sysname> python
>>> import platformtools
>>> c = platformtools.channel('display vlan')
>>> for value in c:
...  print value
...
display vlan
 Total VLANs: 13
 The VLANs include:
 1(default), 2-3, 10, 20, 30, 40, 100-101, 200-202, 300
<Sysname>
```

Returns

Output from the executed commands, in the format of a list object.

Examples

Display the VLAN inforamtion.

```
<Sysname> python
>>> import platformtools
>>> platformtools.channel('display vlan')
```

Sample output

```
['display vlan', ' Total VLANs: 13', ' The VLANs include:', ' 1(default), 2-3, 10, 2
```

```
0, 30, 40, 100-101, 200-202, 300, \x08\x08 ', '<Sysname>']
```

CLI

Use **CLI** to execute Comware 7 CLI commands and create CLI objects.

Syntax

CLI(*command*='', *do_print*=True)

Parameters

command: Specifies the commands to be executed. To enter multiple commands, use a space and a semicolon (;) as the delimiter. To enter a command in a view other than user view, you must first enter the commands used to enter the view. For example, you must enter **'system-view ;local-user test class manage'** to execute the **local-user test class manage** command.

do_print: Specifies whether to output the execution result:

- **True**—Outputs the execution result. This value is the default.
- **False**—Does not output the execution result.

Usage guidelines

This API function supports only Comware 7 commands. It does not support Linux, Python, or Tcl commands.

Returns

CLI objects

Examples

Add a local user named **test**.

```
<Sysname> python
>>> import platformtools
>>> platformtools.CLI('system-view ;local-user test class manage')
```

Sample output

```
<Sysname> system-view
System View: return to User View with Ctrl+Z.
[Sysname] local-user test class manage
New local user added.
<platformtools.CLI object at 0xb7f680a0>
```

get_error

Use **get_error** to get the error information from the download operation.

Syntax

Transfer.get_error()

Returns

Error information (if there is no error information, the return is null)

Examples

Download file **test.cfg** from TFTP server 1.1.1.1 and get the error information from the operation.

```
<Sysname> python
```

```
>>> import platformtools
>>> c = platformtools.Transfer('tftp', '1.1.1.1', 'test.cfg', 'flash:/test.cfg', user='', password='')
>>> c.get_error()
```

Sample output

```
"Couldn't connect to server"
```

get_output

Use **`get_output`** to get the output from executed commands.

Syntax

`CLI.get_output()`

Returns

Output from executed commands

Examples

Add a local user and get the output from the command.

```
<Sysname> python
>>> import platformtools
>>> c = platformtools.CLI('system-view ;local-user test class manage', False)
>>> c.get_output()
```

Sample output

```
['<Sysname>system-view', 'System View: return to User View with Ctrl+Z.',
'[Sysname]local-user test class manage', 'New local user added.']
```

get_self_slot

Use **`get_self_slot`** to get the member ID of the master device.

Syntax

`get_self_slot()`

Returns

A list object in the format of `[-1,`*slot-number*`]`. The *slot-number* indicates the member ID of the master device.

Examples

Get the member ID of the master device.

```
<Sysname> python
>>> import platformtools
>>> platformtools.get_self_slot()
```

Sample output

```
[-1,1]
```

get_slot_info

Use **`get_slot_info`** to get information about a member device.

Syntax

`get_slot_info()`

Returns

A dictionary object in the format of {'Slot': *slot-number*, 'Status': '*status*', 'Chassis': *chassis-number*, 'Role': '*role*', 'Cpu': *CPU-number* }. The *slot-number* argument indicates the member ID of the device. The *status* argument indicates the status of the member device. The *chassis-number* and *CPU-number* arguments are fixed at 0. The *role* argument indicates the role of the member device.

Examples

Get information about the device, a card, or a member device.

```
<Sysname> python
>>> import platformtools
>>> platformtools.get_slot_info(1)
```

Sample output

```
{'Slot': 1, 'Status': 'Normal', 'Chassis': 0, 'Role': 'Master', 'Cpu': 0}
```

get_slot_range

Use **`get_slot_range`** to get the supported IRF member ID range.

Syntax

`get_slot_range()`

Returns

A dictionary object in the format of {'MaxSlot': *max-slot-number*, 'MinSlot': *min-slot-number* }. The *max-slot-number* argument indicates the maximum member ID. The *min-slot-number* argument indicates the minimum member ID.

Examples

Get the supported IRF member ID range.

```
<Sysname> python
>>> import platformtools
>>> platformtools.get_slot_range()
```

Sample output

```
{'MaxSlot': 4, 'MinSlot': 1}
```

get_standby_slot

Use **`get_standby_slot`** to get the member IDs of the subordinate devices.

Syntax

`get_standby_slot()`

Returns

A list object in one of the following formats:

- [ ]—The IRF fabric does not have a subordinate device.
- [[-1,*slot-number*]]—The IRF fabric has only one subordinate device.
- [[-1,*slot-number1*], [-1,*slot-number2*],...]—The IRF fabric has multiple subordinate devices.

The *slot-number* arguments indicate the member IDs of the subordinate devices.

Examples

Get the member IDs of the subordinate devices.

```
<Sysname> python
>>> import platformtools
>>> platformtools.get_standby_slot()
```

Sample output

```
[]
```

Transfer

Use **Transfer** to download a file from a server.

Syntax

Transfer(*protocol*='', *host*='', *source*='', *dest*='', *vrf*='',*login_timeout*=10, *user*='', *password*='')

Parameters

protocol: Specifies the protocol used to download a file:

- **ftp**—Uses FTP.
- **tftp**—Uses TFTP.
- **http**—Uses HTTP.

host: Specifies the IP address of the remote server.

source: Specifies the name of the file to be downloaded from the remote server.

dest: Specifies a name for the downloaded file.

vrf: Specifies the VPN instance to which the remote server belongs. This argument represents the VPN instance name, a case-sensitive string of 1 to 31 characters. If the server belongs to the public network, do not specify this argument.

login_timeout: Specifies the timeout for the operation, in seconds. The default is 10.

user: Specifies the username for logging in to the server.

password: Specifies the login password.

Returns

Transfer object

Usage guidelines

If the IP address of the server is an IPv6 address, enclose the IP address with brackets ([]). For example:

```
Transfer("ftp","[2200::22]","0729.cfg","flash:/0729.cfg",user="abc",password="123456"
)
```

Examples

Download file **test.cfg** from TFTP server 192.168.1.26.

```
<Sysname> python
>>> import platformtools
>>> platformtools.Transfer('tftp', '192.168.1.26', 'test.cfg', 'flash:/test.cfg',
user='', password='')
```

Sample output

```
<platformtools.Transfer object at 0xb7f700e0>
```

send

Use **send** to generate and send a log message.

Syntax

SYSLOG.send(*digest*='', *info*='', *priority*='')

Parameters

digest: Specifies a mnemonic for the log message. The mnemonic is a case-sensitive string of 1 to 32 characters.

info: Specifies the contents for the log message, a case-sensitive string of 1 to 1024 characters.

priority: Specifies a priority for the log message, in the range of 0 to 4294967295.

Usage guidelines

You must use the **SYSLOG** function to create a syslog object before you can generate and send a log message for the object.

To enable the current terminal to display the generated log messages, first execute the **terminal monitor** command in user view.

Returns

None.

Examples

Enable the API to generate and send a log message for SNMP. Set the mnemonic to **Test**, the contents to **Try to send one message.**, and the priority to 1000.

```
<Sysname> terminal monitor
<Sysname> python
>>> import platformtools
>>> a = platformtools.SYSLOG('snmp')
>>> a.send('Test','Try to send one message.',1000)
```

Sample output

```
>>> %Jan  1 06:24:17:908 2019 Sysname SNMP/0/Test: Try to send one message.
```

SYSLOG

Use **SYSLOG** to create a syslog object.

Syntax

SYSLOG(*module*='')

Parameters

module: Specifies a name for the syslog object. The name is a case-insensitive string of 1 to 8 characters.

Usage guidelines

You must create a syslog object before you can send a log message for the object.

Returns

None.

Examples

Create a syslog object named **SNMP** and assign the object to variable **a**.

```
<Sysname> python
>>> import platformtools
>>> a = platformtools.SYSLOG('snmp')
>>> a
```

Sample output

```
<platformtools.SYSLOG object at 0xb7e180e0>
```

Contents

Managing licenses 1
About licenses 1
License types 1
Basic concepts 2
Restrictions and guidelines: License management 3
Management operation restrictions 3
File operation restrictions 3
Configuring local licensing 3
About local licensing 3
Registering and installing a license 4
License registration and installation workflow 4
Identifying the license storage 5
Compressing the license storage 5
Obtaining required information for license registration 6
Installing an activation file 6
Transferring a license 7
About this task 7
License transferring workflow 7
Obtaining an Uninstall file 8
Uninstalling a license 9
About license uninstallation 9
License uninstallation workflow 9
Recovering an activation file 9
Display and maintenance commands for license management 10
Configuring remote licensing 0
About remote licensing 0
Registering and installing a license 0
License registration and installation workflow 0
Configuring a license client 1
Transferring a license 2
About this task 2
License transferring workflow 2
Uninstalling a license 3
About license uninstallation 3
License uninstallation workflow 3
Display and maintenance commands for license client 4

Managing licenses

About licenses

To use license-based features, you need to purchase licenses from H3C and install the licenses.

To obtain information about license-based features, their licensing status, and license availability, execute the **`display license feature`** command on the device. Then, you can purchase and install licenses as needed.

License types

H3C offers preinstalled, trial (or temporary), and formal licenses. Support for preinstalled and trial licenses varies by device model. For more information about the license options available for the device, see the license matrixes document for the device.

Preinstalled licenses

Preinstalled licenses are built into a system and are available for use on initial startup of the system. Users do not need to activate them.

A preinstalled license is typically time limited and cannot be uninstalled or transferred.

When a preinstalled license expires, the license-based feature becomes unavailable. To continue to use the feature, you must purchase and install a formal license for it.

Trial licenses

Trial licenses are provided for you to verify the functionality of premium features so you can make an educated purchase decision.

A trial license is time limited and cannot be uninstalled or transferred.

When a trial license expires, the license-based feature becomes unavailable. To continue to use the feature, you must purchase and install a formal license for it.

To obtain a trial license for a feature:

1. Contact your H3C sales representative or technical support.
2. Register the trial license key with H3C License Management Platform to obtain an activation file.
3. Install the activation file to activate the feature.

Formal licenses

To gain most from a license-based feature, purchase a formal license for it.

Formal licenses are available with different validity periods and their support for uninstallation and transfer of formal licenses depends on the device model. For more information, see the license matrixes document for the device.

To install a formal license for a feature:

1. Purchase a software license certificate through an official channel.
2. Access H3C License Management Platform, and then enter the license key in the certificate and the required device information to obtain an activation file.
3. Install the activation file in the target system.

Basic concepts

The following information describes the basic concepts that you might encounter when you register, install, and manage licenses.

H3C License Management Platform

H3C License Management Platform provides product licensing services for H3C customers. You can access this system to obtain an activation file, transfer licenses, or remove the device and license binding for an uninstalled license.

H3C License Management Platform is accessible at http://www.h3c.com/en/License/.

Software license certificate

A software license certificate allows users to use a license-based feature. It contains license key, license capacity, and other information.

License key

A license key uniquely identifies a license.

- To obtain a formal license key, purchase a software license certificate. The authorization serial number in the software license certificate is the license key.
- To obtain a trial license key, contact your H3C sales representative or H3C technical support. Support for trial licenses depends on the device model. For more information, see the license matrixes document for the device.

Device serial number

A device serial number (SN or S/N) is a barcode that uniquely identifies a device. It comes with the device and must be provided when you request a license in H3C License Management Platform.

Device ID (DID) and DID file

A DID is a string of characters that uniquely identifies a hardware device. A DID file stores the DID and other information. The device comes with a DID or DID file. You must provide the DID or DID file when you request a license for the device on H3C License Management Platform.

Activation file

An activation file binds a license to a system.

To use a license-based feature on a system, you must perform the following tasks:

1. Use the license key and the system's SN and DID information to obtain an activation file from H3C License Management Platform.
2. Install the activation file on the system.

Uninstall key and Uninstall file

When you uninstall a license, an Uninstall file that contains an Uninstall key is created. The Uninstall key is required for transferring the license.

License storage

License storage is a persistent storage of fixed size for storing licensing information. This information includes the licensing state, validity period, Uninstall key or Uninstall file, and other related information.

Data in the license storage persists through reboot. This ensures licensing accuracy and continuity.

Restrictions and guidelines: License management

Management operation restrictions

- Purchase licenses from H3C official channels.
- For licenses that have been installed on the device, execute the **`display license`** command to view the license validity period. To use a license-based feature continuously, install a new license for the feature before the old license expires.
- Licenses are typically device locked. To ensure a successful licensing, use the following licensing guidelines:
  a. When you purchase a license certificate, verify the following items:
    - Make sure the license is compatible with the target device.
    - Make sure its licensed functionality and capacity meet your requirements.
  b. When you obtain an activation file, make sure the provided license key and hardware information are correct.
  c. Install the activation file on the correct target device.
- Make sure no one else is performing license management tasks while you are managing licenses on the device.

File operation restrictions

When you manage DID files, activation files, or Uninstall files, follow these restrictions and guidelines:

- To avoid licensing error, do not modify the name of a DID file, activation file, or Uninstall file, or edit the file content.
- Before you install an activation file, download the activation file to the storage media of the device such as flash memory. When installing an activation file, the device automatically copies the activation file to the license folder in the root directory of the storage media. The **license** folder stores important files for licensing. For licensed features to function correctly, do not delete or modify the license folder or the files in the license folder.

Configuring local licensing

About local licensing

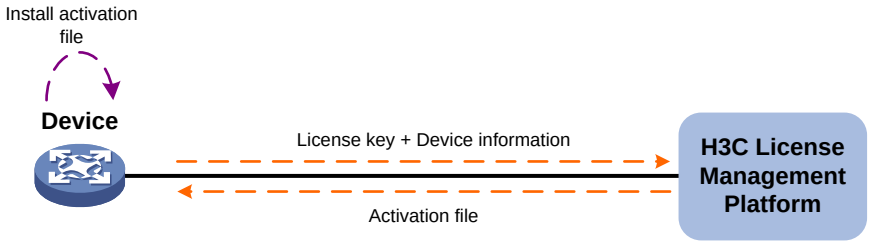
Local licensing requires license activation device by device. It is applicable to small-sized networks.

To install a license on a device:

1. Obtain the license key and the device SN and DID information of the device.
2. Access H3C License Management Platform to apply for an activation file based on the license key and the device's SN and DID information.
3. Install the activation file on the device to activate the license.

The activation file for a license is device locked. You cannot install the activation file for one device to activate the license on another device.

Figure 1 Local licensing procedure



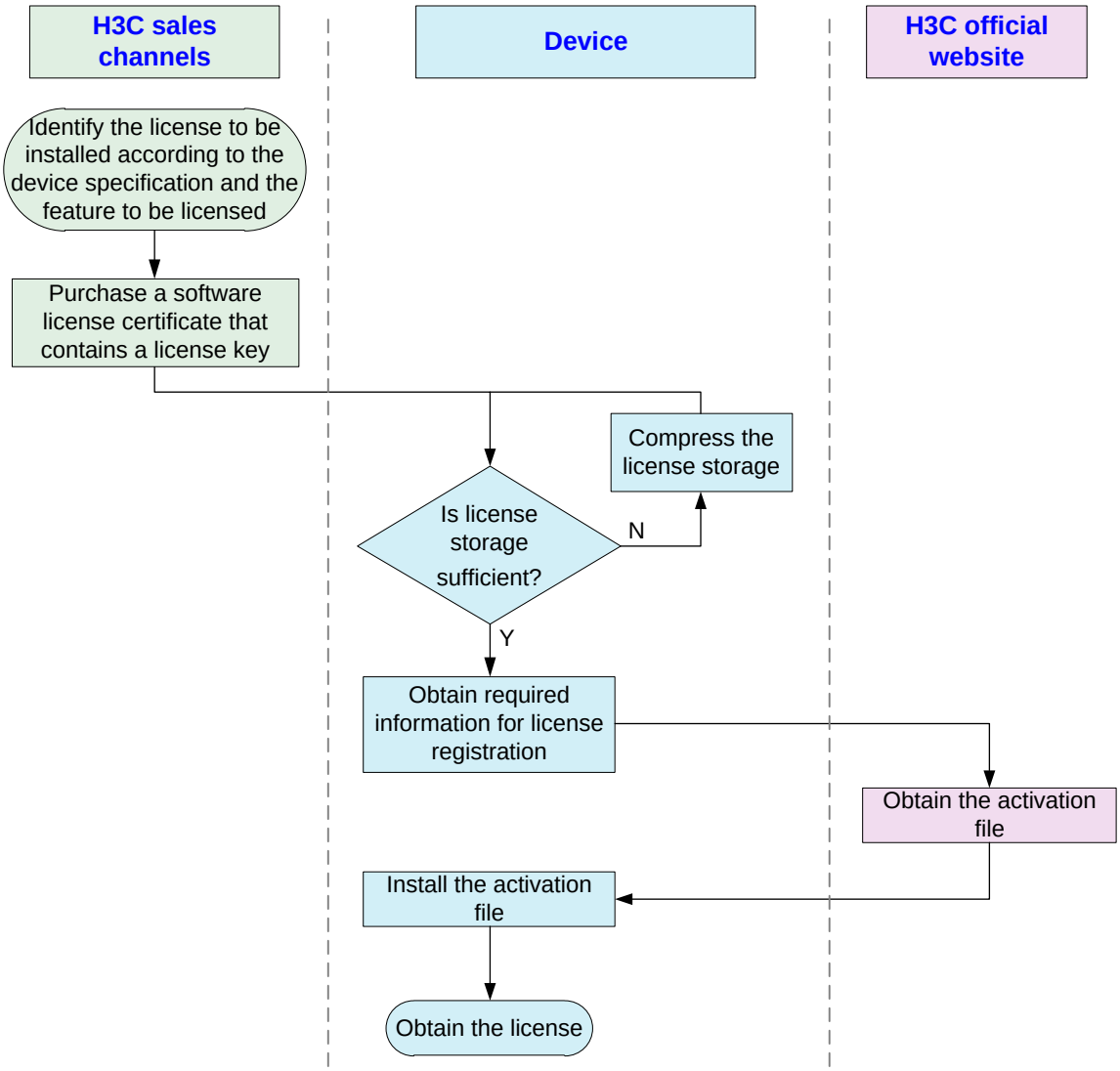
Registering and installing a license

License registration and installation workflow

The procedures for installing formal licenses and trial licenses are the same, as shown in Figure 2.

This chapter only describes the operations performed on the device. For more information, see the licensing guide for the device.

Figure 2 License registration and installation workflow



Identifying the license storage

To identify the free space of the license storage, execute the following command in any view:

`display license feature`

From the command output, view the **Total** and **Usage** fields to examine whether the remaining license storage is sufficient for installing new licenses. If the remaining license storage is not sufficient, compress the license storage.

Compressing the license storage

About this task

The license storage stores licensing information and has a fixed size.

You can compress the license storage to delete expired and uninstalled license information before requesting activation files to ensure sufficient storage space for installing new licenses.

If no license has been installed on the device, you do not need to compress the license storage.

Prerequisites

Back up the Uninstall keys or Uninstall files for the uninstalled licenses for subsequent license transfer or license uninstallation on H3C License Management Platform.

If uninstalled licenses or expired licenses exist on the device, the compression operation will make the DID or DID file change. You will be unable to install the activation file obtained by using the old DID or DID file on the device. As a best practice, install all activation files registered with the old DID or DID file before performing a compression.

If you have not installed an activation file registered with the old DID, take the following actions:

- If the license storage is sufficient, install the activation file on the device. For more information, see the license user guide.
- If the license storage is insufficient and the activation file cannot be installed after the compression, contact H3C Support.

Procedure

1. Enter system view.

   system-view

2. Compress the license storage.

   license compress slot *slot-number*

Obtaining required information for license registration

To obtain the SN and DID information, execute the following command in any view:

display license device-id slot *slot-number*

Installing an activation file

About this task

CAUTION:

Back up an activation file before you install it. If the activation file is inadvertently deleted or becomes unavailable for some other reason, you can use the backup activation file to restore the license.

To obtain a license, install an activation file for the license on the device.

Prerequisites

Use FTP or TFTP to upload the activation file to be installed to the device. If FTP is used to transfer the activation file, set it in binary mode.

Installing an activation file

1. Enter system view.

   system-view

2. Install an activation file.

   license activation-file install *file-name* **slot** *slot-number*

   You can install a single .ak file or multiple .ak files through one operation. To install multiple .ak files, save all activation files in the same directory and specify the directory as the value of the *file-name* argument.

Transferring a license

About this task

Perform this task to transfer a license that has not expired from one device to another in the same product series.

License transferring workflow

To transfer licenses from one device to another, use the workflow in Figure 3.

This chapter describes only the operations performed on the source device. For the operations performed on the target device, see "Registering and installing a license." For more information about license transferring, see the licensing guide for the device.

Figure 3 License transferring workflow

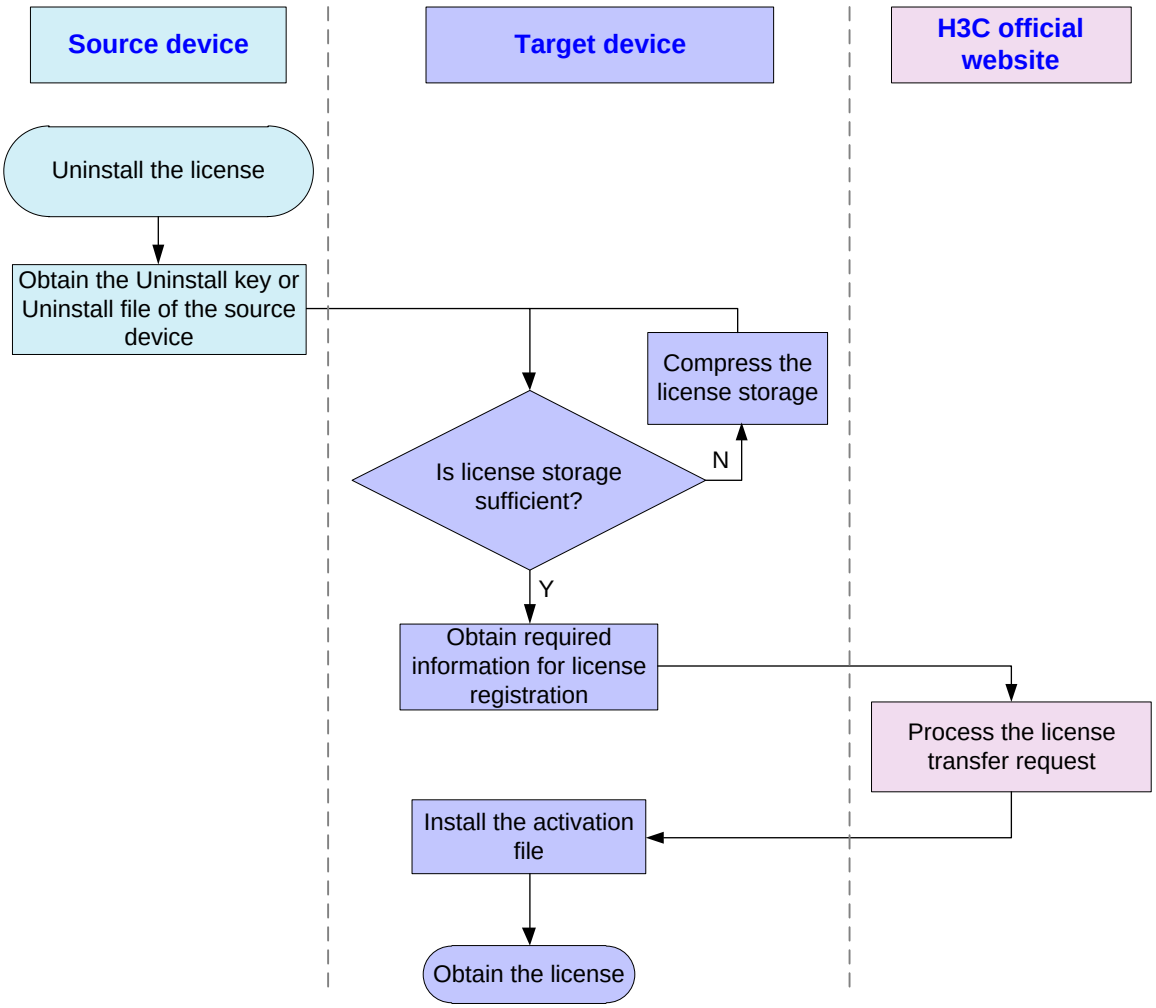
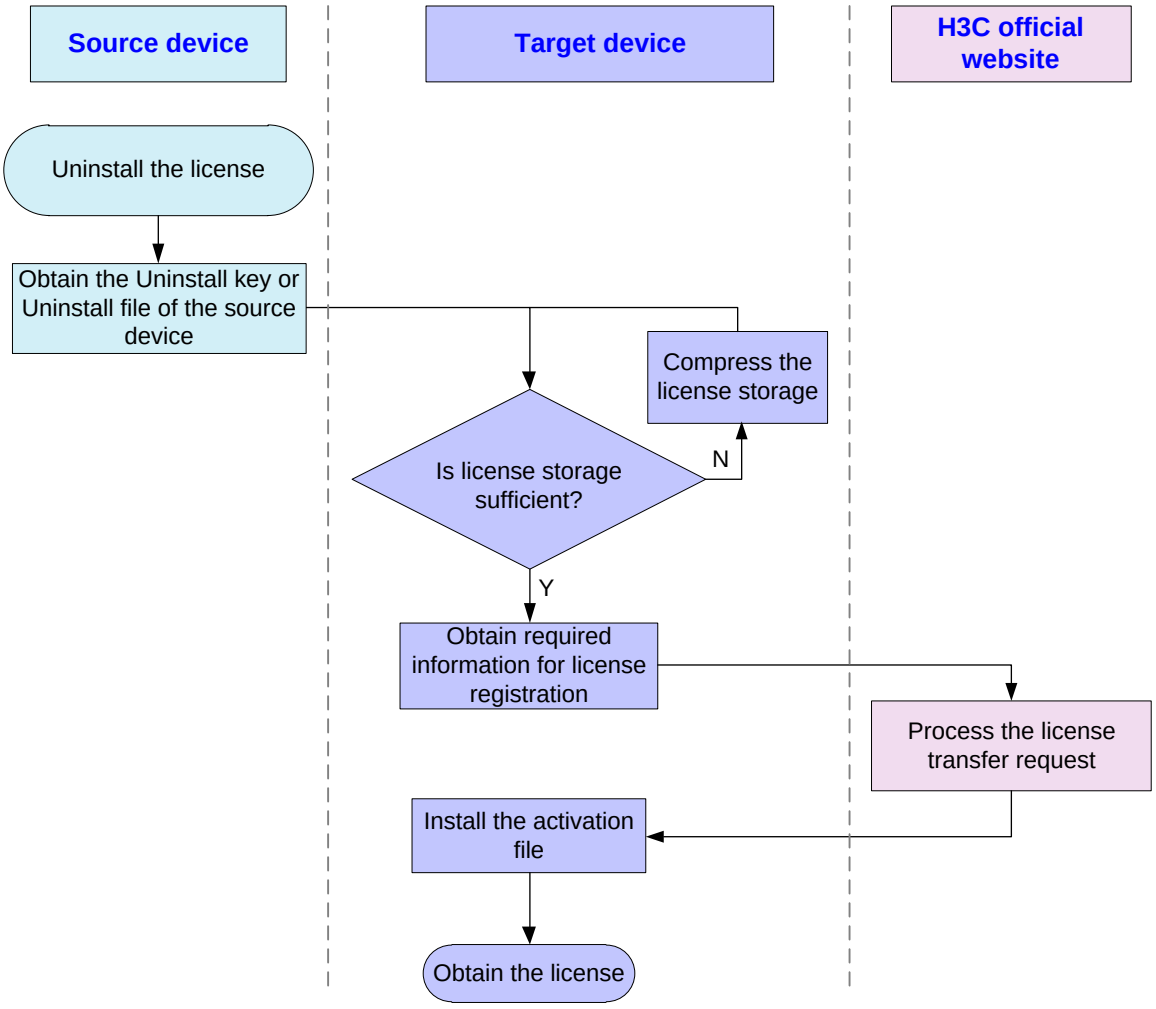


Figure 4 License transferring workflow



Obtaining an Uninstall file

1. Execute the **display license** command to view the activation file to be uninstalled.
2. Enter system view.

   system-view
3. Uninstall an activation file.

   license activation-file uninstall *license-file* **slot** *slot-number*

   You can uninstall a single .ak file or multiple .ak files through one operation. To uninstall multiple .ak files, save all activation files in the same directory and specify the directory as the value of the *license-file* argument.

Uninstalling a license

About license uninstallation

A license key is bound to a device once you activate the license key on H3C License Management Platform. You cannot bind this license key to another device. To unbind a license key for a license that has not expired from a device, obtain the Uninstall key or Uninstall file for the license on the device. Then, request for license uninstallation on H3C License Management Platform.

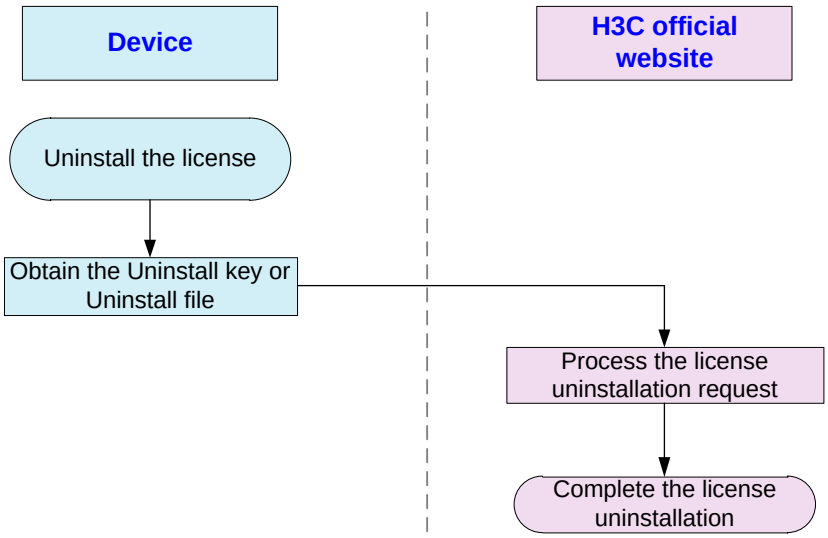
After license uninstallation is requested successfully, you can use this license key to request a new activation file on H3C License Management Platform.

The new activation file can be installed on another device for license transfer or on the local device for license restoration.

License uninstallation workflow

For information about obtaining the Uninstall key or Uninstall file for a license, see "Obtaining an Uninstall file." For more information about license uninstallation, see the licensing guide for the device.

Figure 5 License uninstallation workflow



Recovering an activation file

If you mistakenly delete an activation file, use the following procedure to recover the activation file:

1. Use the **copy** command to copy the backup activation file to the **license** folder in the root directory of the default storage media.
2. Use the **display license** command to verify that the state of the recovered activation file is **In use**.
3. Restart the device if the license state is **In use** but the licensed feature cannot function correctly.

Display and maintenance commands for license management

Execute **display** commands in any view.

Task	Command
Display detailed license information.	**display license** [**activation-file**] [**slot** *slot-number*]
Display the SN and DID information.	**display license device-id slot** *slot-number*
Display brief feature license information.	**display license feature**
Display preinstalled licenses.	**display license preinstalled**

Configuring remote licensing

About remote licensing

Remote licensing is based on the client-server architecture and applicable to large-sized networks. Remote licensing simplifies license installation and management. With remote licensing, you install licenses, uninstall licenses, or transfer licenses for all license clients from the license server.

The following is the procedure to set up remote licensing:

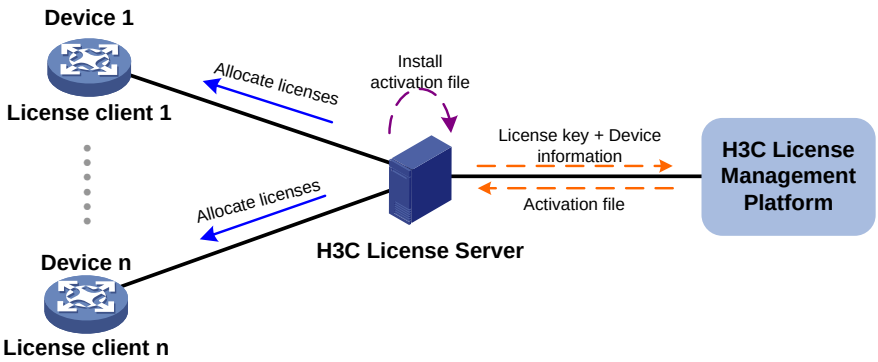
1. Install the H3C license server on a host.

   The H3C license server provides centralized management on licensing and license distribution.
2. Install the licenses required by license clients on the license server.

   If license clients request for licenses, the license server will assign the licenses to the license clients.
3. Enable a license client on each managed device and configure the parameters for communicating with the license server. The parameters include the IP address and port number of the license server and the username and password for accessing the license server.

The managed devices can then connect to the license server to obtain licenses as license clients.

Figure 6 Remote licensing



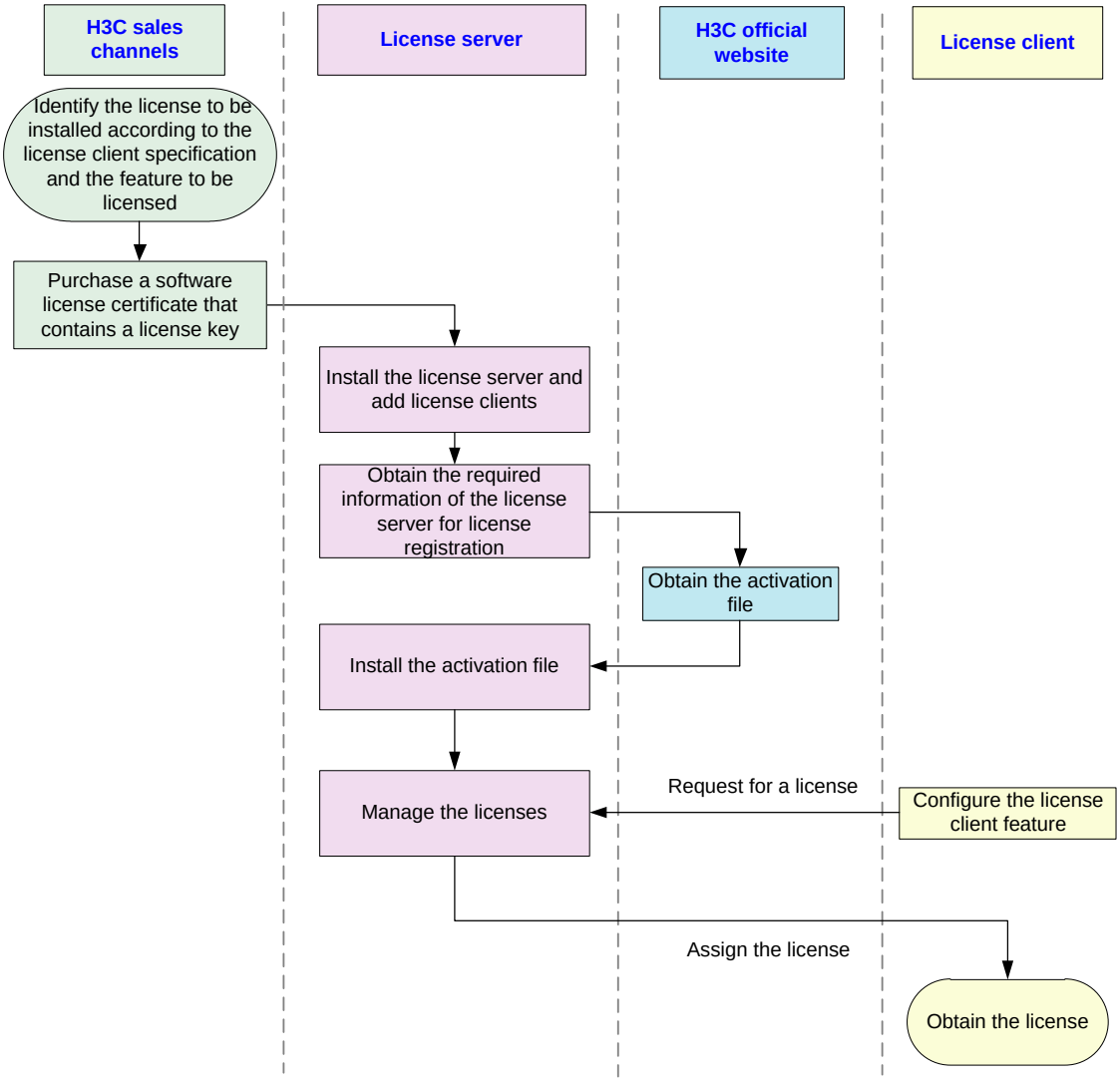
Registering and installing a license

License registration and installation workflow

The procedures for registering and installing formal licenses and trial licenses are the same, as shown in Figure 7.

This chapter only describes the license client configuration. For more information about the complete procedures, see the licensing guide for the device.

Figure 7 License registration and installation workflow



Configuring a license client

About this task

In remote licensing, licenses are installed on the license server and are assigned to a requesting license client.

Restrictions and guidelines

You can uninstall a license on a license client whether the license has expired or not. After a license is uninstalled and released from the license client, the license server withdraws the license. The withdrawn license can be assigned to other requesting license clients if the license is within the validity period.

Specifying the license server information

1. Enter system view.

   `system-view`

2. Specify the IP address and port number of the license server.

license server { **ipv4** *ipv4-address* | **ipv6** *ipv6-address* } **port** *port-number* [**vpn-instance** *vpn-instance-name*]

By default, no license server information is specified on the license client.

You can specify a maximum of four license servers on the license client. Each license server is uniquely identified by the combination of an IPv4 address and a port number.

When the upper limit is reached, you cannot specify new license servers unless you delete existing license server settings.

As a best practice, make sure only one license server is reachable. If multiple license servers are reachable, the license client selects the first license server with which the client establishes a connection.

To modify license server settings after the license client is enabled, disable the license client first.

3. Specify the username and password for the license client to access the license server.

   license client username *username* **password** { **cipher** | **simple** } *password*

   By default, no username or password is specified on the license client.

   Make sure the specified username and password on the license client are the same as those configured on the license server.

Configuring the license cache

The license cache defines the maximum number of licenses that can be cached for a license type. Without the license cache configured, the license client requests a license from the license server to use the feature and releases the license to the license server when the client does not need the feature. With the license cache configured, the license client can cache licenses, avoiding frequent license requesting from the license server.

To configure the license cache:

1. Enter system view.

   system-view

2. Configure the license cache size on the license client.

   license client config cache license-name *license-name* **size** *cache-size*

   By default, no license cache size is configured. The client does not cache licenses.

Enabling the license client

1. Enter system view.

   system-view

2. Enable the license client.

   license client enable

   By default, the license client is disabled, and the device cannot get licenses from the license server.

Transferring a license

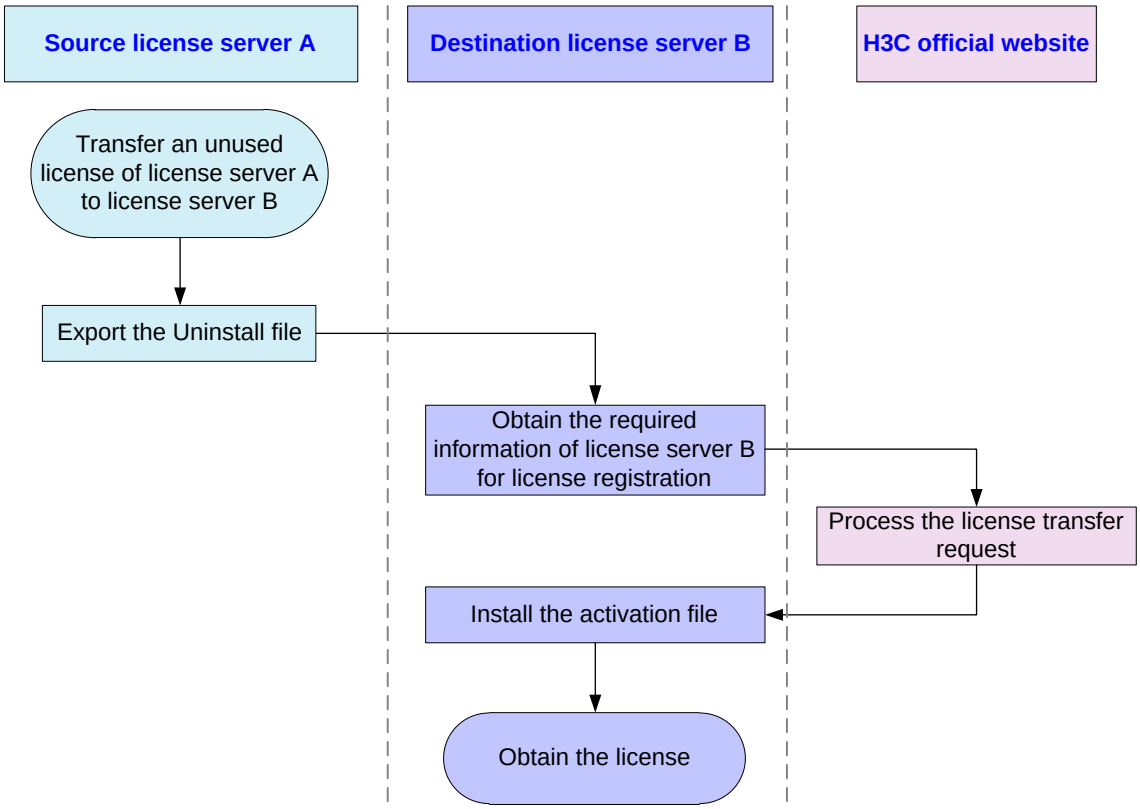
About this task

Perform this task to transfer a license that has not expired from one license server to another.

License transferring workflow

For more information about the complete license transferring procedures, see the licensing guide for the device.

Figure 8 License transferring workflow



Uninstalling a license

About license uninstallation

A license is bound to a license server once you obtain an activation file based on the license key and hardware information of the license server on H3C License Management Platform. You can bind a license only to one license server.

To transfer or restore the license before it expires, you must first uninstall its activation file from the license server, and then request license uninstallation on H3C License Management Platform.

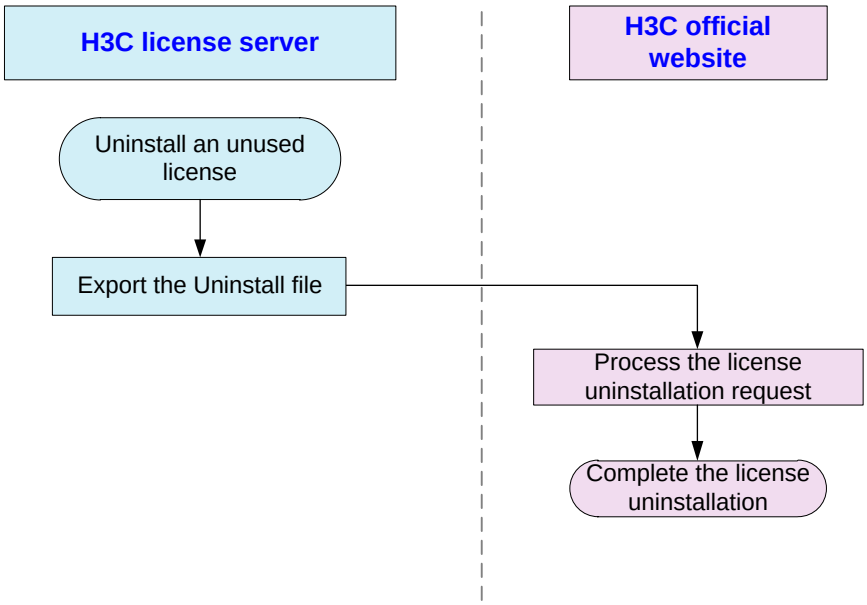
After license uninstallation is approved, the license key is available for use again. You can use it to obtain a new activation file on H3C License Management Platform to transfer the license to another license server or restore the license on the original license server.

For the support of the product for license uninstallation, see its license matrixes document.

License uninstallation workflow

For more information about the complete license uninstallation procedures, see the licensing guide for the device.

Figure 9 License uninstallation workflow



Display and maintenance commands for license client

Execute **display** commands in any view.

Task	Command
Display the license client configuration and the obtained license information on the license client.	**display license client**
Display license statistics on the license client.	**display license client statistics**

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Fundamentals Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Virtual Technologies Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the fundamentals and configuration procedures of the IRF technologies.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring an IRF fabric ····· 1

About IRF ····· 1
IRF network model ····· 1
IRF benefits ····· 1
Basic concepts ····· 2
IRF network topology ····· 4
Master election ····· 4
Interface naming conventions ····· 5
File system naming conventions ····· 5
Configuration synchronization ····· 6
Multi-active handling procedure ····· 7
MAD mechanisms ····· 9
Restrictions and guidelines: IRF configuration ····· 15
Hardware compatibility with IRF ····· 15
Software requirements for IRF ····· 15
IRF fabric size ····· 15
Candidate IRF physical interfaces ····· 16
Transceiver modules and cables selection for IRF ····· 17
IRF port connection ····· 18
IRF physical interface configuration restrictions and guidelines ····· 18
Feature compatibility and configuration restrictions with IRF ····· 19
Configuration rollback restrictions ····· 19
IRF tasks at a glance ····· 20
Planning the IRF fabric setup ····· 20
Setting up an IRF fabric ····· 20
IRF setup tasks at a glance ····· 20
Assigning a member ID to each IRF member device ····· 21
Specifying a priority for each member device ····· 22
Binding physical interfaces to IRF ports ····· 22
Bulk-configuring basic IRF settings for a member device ····· 23
Connecting IRF physical interfaces ····· 24
Accessing the IRF fabric ····· 24
Configuring MAD ····· 25
Restrictions and guidelines for MAD configuration ····· 25
Configuring LACP MAD ····· 25
Configuring BFD MAD ····· 26
Configuring ARP MAD ····· 31
Configuring ND MAD ····· 34
Excluding interfaces from the shutdown action upon detection of multi-active collision ····· 36
Recovering an IRF fabric ····· 37
Optimizing IRF settings for an IRF fabric ····· 37
Enabling IRF auto-merge ····· 37
Configuring a member device description ····· 38
Configuring IRF bridge MAC address settings ····· 38
Enabling software auto-update for software image synchronization ····· 39
Setting the IRF link down report delay ····· 40
Display and maintenance commands for IRF ····· 40
IRF configuration examples ····· 41
Example: Configuring an IRF fabric with LACP MAD enabled on an aggregate interface ····· 41
Example: Configuring an IRF fabric with BFD MAD enabled on a VLAN interface ····· 45
Example: Configuring an IRF fabric with ARP MAD enabled on a VLAN interface ····· 49
Example: Configuring an IRF fabric with ND MAD enabled on a VLAN interface ····· 54
Example: Configuring an IRF fabric with ARP MAD and management IP configured on the same management Ethernet port ····· 58
Example: Removing member devices from an IRF fabric without MAD in place ····· 62
Example: Removing member devices from an IRF fabric with MAD in place ····· 64

Configuring an IRF fabric

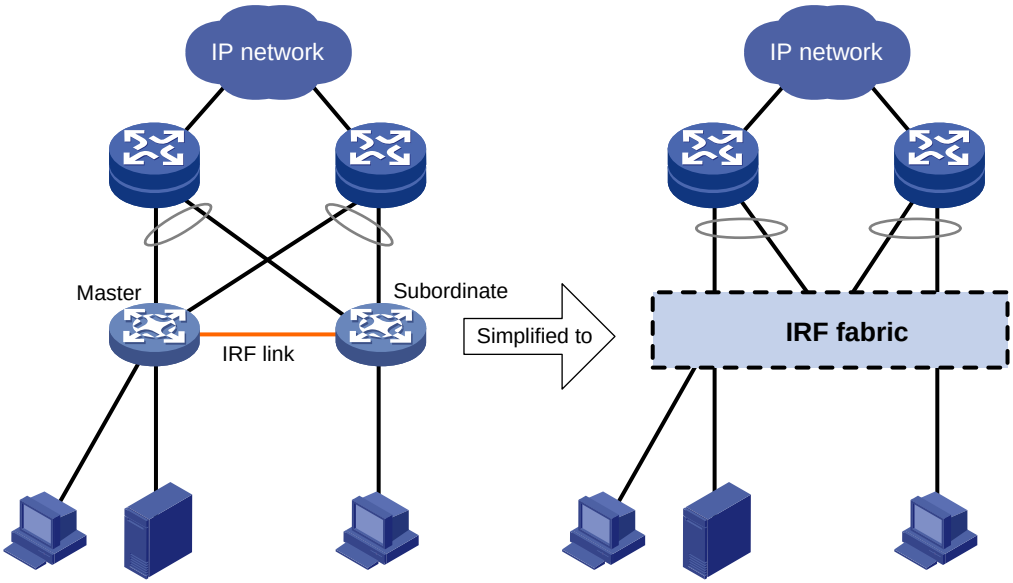
About IRF

The Intelligent Resilient Framework (IRF) technology virtualizes multiple physical devices at the same layer into one virtual fabric to provide data center class availability and scalability. IRF virtualization technology offers processing power, interaction, unified management, and uninterrupted maintenance of multiple devices.

IRF network model

Figure 1 shows an IRF fabric that has two devices, which appear as a single node to the upper-layer and lower-layer devices.

Figure 1 IRF application scenario



IRF benefits

IRF provides the following benefits:

- **Simplified topology and easy management**—An IRF fabric appears as one node and is accessible at a single IP address on the network. You can use this IP address to log in at any member device to manage all the members of the IRF fabric. In addition, you do not need to run the spanning tree feature among the IRF members.
- **1:N redundancy**—In an IRF fabric, one member acts as the master to manage and control the entire IRF fabric. All the other members process services while backing up the master. When the master fails, all the other member devices elect a new master from among them to take over without interrupting services.
- **IRF link aggregation**—You can assign several physical links between neighboring members to their IRF ports to create a load-balanced aggregate IRF connection with redundancy.
- **Multichassis link aggregation**—You can use the Ethernet link aggregation feature to aggregate the physical links between the IRF fabric and its upstream or downstream devices across the IRF members.

- **Network scalability and resiliency**—Processing capacity of an IRF fabric equals the total processing capacities of all the members. You can increase ports, network bandwidth, and processing capacity of an IRF fabric simply by adding member devices without changing the network topology.

Basic concepts

IRF member roles

IRF uses two member roles: master and standby (called subordinate throughout the documentation).

When devices form an IRF fabric, they elect a master to manage and control the IRF fabric, and all the other devices back up the master. When the master device fails, the other devices automatically elect a new master. For more information about master election, see "Master election."

IRF member ID

An IRF fabric uses member IDs to uniquely identify and manage its members. This member ID information is included as the first part of interface numbers and file paths to uniquely identify interfaces and files in an IRF fabric. Two devices cannot form an IRF fabric if they use the same member ID. A device cannot join an IRF fabric if its member ID has been used in the fabric.

Member priority

Member priority determines the possibility of a member device to be elected the master. A member with higher priority is more likely to be elected the master.

IRF port

An IRF port is a logical interface that connects IRF member devices. Every IRF-capable device has two IRF ports.

The IRF ports are named IRF-port *n*/1 and IRF-port *n*/2, where *n* is the member ID of the device. The two IRF ports are referred to as IRF-port 1 and IRF-port 2.

To use an IRF port, you must bind a minimum of one physical interface to it. The physical interfaces assigned to an IRF port automatically form an aggregate IRF link. An IRF port goes down when all its IRF physical interfaces are down.

IRF physical interface

IRF physical interfaces connect IRF member devices and must be bound to an IRF port. They forward traffic between member devices, including IRF protocol packets and data packets that must travel across IRF member devices.

IRF split

IRF split occurs when an IRF fabric breaks up into multiple IRF fabrics because of IRF link failures, as shown in Figure 2. The split IRF fabrics operate with the same IP address. IRF split causes routing and forwarding problems on the network. To quickly detect a multi-active collision, configure a minimum of one MAD mechanism (see "Configuring MAD").

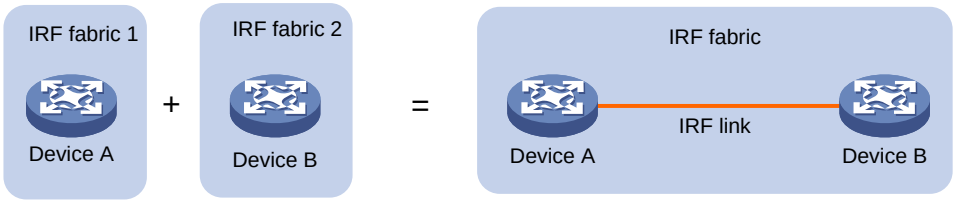
Figure 2 IRF split



IRF merge

IRF merge occurs when two split IRF fabrics reunite or when two independent IRF fabrics are united, as shown in Figure 3.

Figure 3 IRF merge



MAD

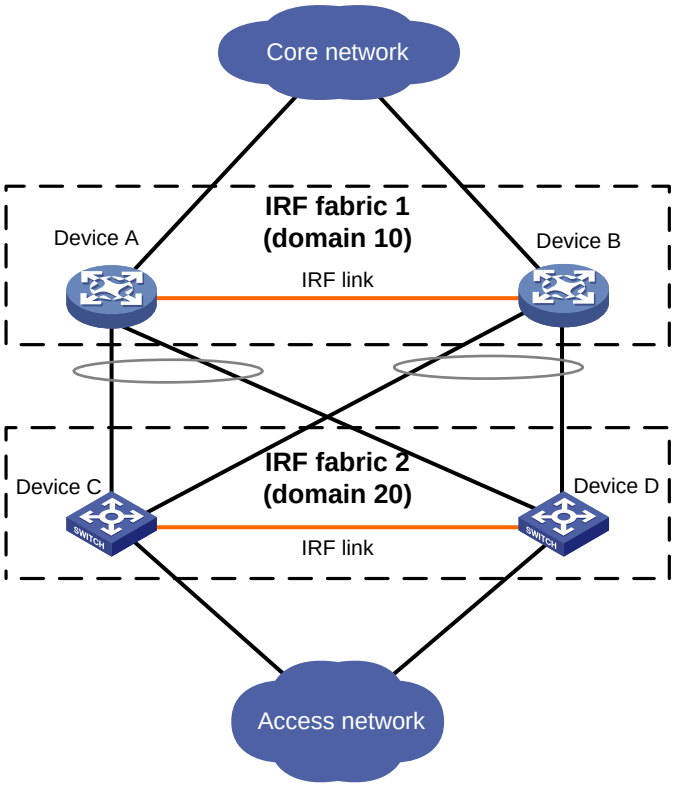
An IRF link failure causes an IRF fabric to split in two IRF fabrics operating with the same Layer 3 settings, including the same IP address. To avoid IP address collision and network problems, IRF uses multi-active detection (MAD) mechanisms to detect the presence of multiple identical IRF fabrics, handle collisions, and recover from faults.

IRF domain ID

One IRF fabric forms one IRF domain. IRF uses IRF domain IDs to uniquely identify IRF fabrics and prevent IRF fabrics from interfering with one another.

As shown in Figure 4, IRF fabric 1 contains Device A and Device B, and IRF fabric 2 contains Device C and Device D. Both fabrics use the LACP aggregate links between them for MAD. When a member device receives an extended LACPDU for MAD, it checks the domain ID to determine whether the packet is from the local IRF fabric. Then, the member device can handle the packet correctly.

Figure 4 A network that contains two IRF domains



IRF network topology

IMPORTANT:

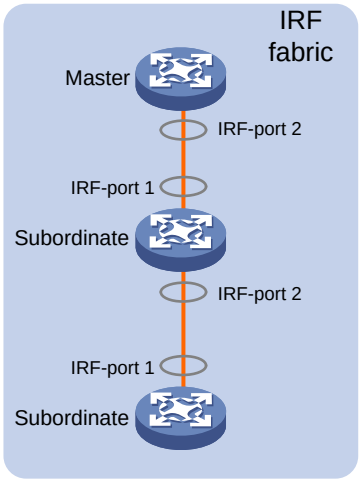
When you set up an IRF fabric, make sure the links in any of the following lists have the same bandwidth as the links in the same list:

- The bandwidth of each IRF physical link attached to the same IRF port.
- The bandwidth of each IRF aggregate link attached to the IRF ports on the same IRF member device.
- The bandwidth of each IRF aggregate link attached to the IRF ports connecting two member devices.

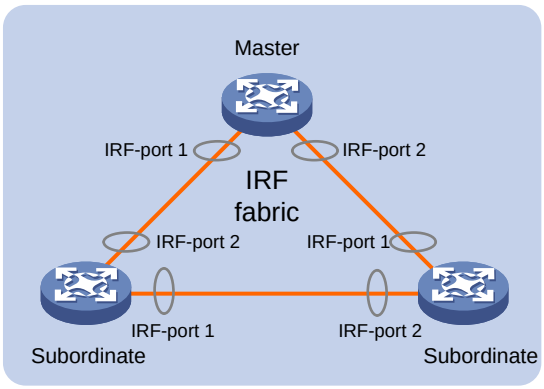
If the bandwidths do not meet the requirements, packet loss might occur for the traffic forwarded across member devices.

An IRF fabric can use a daisy-chain or ring topology. As shown in Figure 5, a ring topology is more reliable. In ring topology, the failure of one IRF link does not cause the IRF fabric to split as in daisy-chain topology. Rather, the IRF fabric changes to a daisy-chain topology without interrupting network services.

Figure 5 Daisy-chain topology vs. ring topology



Daisy-chain topology



Ring topology

IMPORTANT:

This switch series does not support building a two-member IRF fabric in ring topology.

Master election

Master election occurs each time the IRF fabric topology changes in the following situations:

- The IRF fabric is established.
- The master device fails or is removed.
- The IRF fabric splits.
- Independent IRF fabrics merge.

NOTE:

Master election does not occur when split IRF fabrics merge. For information about the master device of the merged IRF fabric, see "Failure recovery."

Master election selects a master in descending order:

1. Current master, even if a new member has higher priority.

   When an IRF fabric is being formed, all members consider themselves as the master. This rule is skipped.
2. Member with higher priority.
3. Member with the longest system uptime.

   Two members are considered to start up at the same time if the difference between their startup times is equal to or less than 10 minutes. For these members, the next tiebreaker applies.
4. Member with the lowest CPU MAC address.

For the setup of a new IRF fabric, the subordinate devices must reboot to complete the setup after the master election.

For an IRF merge, devices must reboot if they are in the IRF fabric that fails the master election.

Interface naming conventions

A physical interface is numbered in the *chassis-number/slot-number/interface-index* format.

- ***chassis-number***—Member ID of the device. The default value for this argument is 1. Any change to the member ID takes effect after a reboot.
- ***slot-number***—Slot number of the interface, which is fixed at 0.
- ***interface-index***—Interface index on the device. Interface index depends on the number of physical interfaces available on the device. To identify the index of a physical interface, examine its index mark on the chassis.

For example, GigabitEthernet 2/0/1 represents the first fixed physical interface on member device 2. Set its link type to trunk, as follows:

```
<Sysname> system-view
[Sysname] interface gigabitethernet 2/0/1
[Sysname-GigabitEthernet2/0/1] port link-type trunk
```

File system naming conventions

On a single-chassis fabric, you can use its storage device name to access its file system.

On a multichassis IRF fabric, you can use the storage device name to access the file system of the master. To access the file system of any other member device, use the name in the **slot***member-ID#storage-device-name* format.

For more information about storage device naming conventions, see *Fundamentals Configuration Guide*.

For example:

- To create and access the **test** folder under the root directory of the flash memory on the master switch:

  ```
  <Master> mkdir test
  Creating directory flash:/test... Done.
  <Master> cd test
  <Master> dir
  ```

```
Directory of flash:/test
The directory is empty.
2009784 KB total (1239680 KB free)
```

- To create and access the **test** folder under the root directory of the flash memory on member device 3:

```
<Master> mkdir slot3#flash:/test
Creating directory slot3#flash:/test... Done.
<Master> cd slot3#flash:/test
<Master> dir
Directory of slot3#flash:/test
The directory is empty.
2009784 KB total (1239652 KB free)
```

Configuration synchronization

IRF uses a strict running-configuration synchronization mechanism to make sure all member devices operate as if they were one device. When the master device fails, the remaining devices can still operate correctly to provide services.

The configuration synchronization mechanism has two phases: bulk sync at startup and real-time sync during operation.

1. **Bulk sync**—Occurs when a device starts up in the following situations:
   - The device reboots to join the IRF fabric as a new subordinate device.

     In this situation, the device takes the following actions:
     - Pulls and loads the running configuration from the master device.
     - Loads the IRF member description, priority, and IRF port bindings locally configured for it by using the **`irf member description, irf member priority, and irf-port commands, respectively.`**
     - Pushes its IRF member description, priority, and IRF port bindings to the master device.

     The configuration files on the device are retained, but they do not take effect.
   - The entire IRF fabric reboots.

     In this situation, all member devices in the IRF reboot simultaneously. The subordinate devices pull the startup configuration file from the master device and then extract the configuration from the file to run in memory.
2. **Real-time sync**—Occurs while the IRF fabric is operating.

   In this situation, the system takes the following actions:
   - Saves all configuration changes you make to the running configuration on the master device.
   - Pushes the configuration changes to the subordinate devices in real time.

Bulk sync and real-time sync ensure that all member devices in an IRF fabric run the same configuration. When the master device fails, the remaining member devices can still operate as configured.

To prevent configuration loss upon reboot of the IRF fabric, save configuration changes to all member devices by using the **`save all`** command while the IRF fabric is operating.

Multi-active handling procedure

The multi-active handling procedure includes detection, collision handling, and failure recovery.

Detection

IRF provides MAD mechanisms by extending LACP, BFD, ARP, and IPv6 ND to detect multi-active collisions. As a best practice, configure a minimum of one MAD mechanism on an IRF fabric. For more information about the MAD mechanisms and their application scenarios, see "MAD mechanisms."

For information about LACP, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*. For information about BFD, see *High Availability Configuration Guide*. For information about ARP, see *Layer 3—IP Services Configuration Guide*. For information about ND, see IPv6 basics in *Layer 3—IP Services Configuration Guide*.

Collision handling

When detecting a multi-active collision, MAD disables all IRF fabrics except one from forwarding data traffic by placing them in Recovery state. The IRF fabrics placed in Recovery state are called inactive IRF fabrics. The IRF fabric that continues to forward traffic is called the active IRF fabric.

BFD MAD and LACP MAD use the following process to handle a multi-active collision:

1. Compare the health states of split fabrics. To view their health states, use the **`display system health`** command. For more information about this command, see device management in *Fundamentals Command Reference*.
2. Set all fabrics to the Recovery state except the healthiest one.
3. Compare the number of members in each fabric if all IRF fabrics are in the same health state.
4. Set all fabrics to the Recovery state except the one that has the most members.
5. Compare the member IDs of their masters if all IRF fabrics have the same number of members.
6. Set all fabrics to the Recovery state except the one that has the lowest numbered master.
7. Shut down all common network interfaces in the Recovery-state fabrics except for the following interfaces:
   - Interfaces automatically excluded from being shut down by the system.
   - Interfaces specified by using the **`mad exclude interface`** command.

ARP MAD and ND MAD use the following process to handle a multi-active collision:

8. Compare the health states of split fabrics. To view their health states, use the **`display system health`** command. For more information about this command, see device management in *Fundamentals Command Reference*.
9. Set all fabrics to the Recovery state except the healthiest one.
10. Compare the member IDs of the masters in the IRF fabrics if all IRF fabrics have the same health state.
11. Set all fabrics to the Recovery state except the one that has the lowest numbered master.
12. Shut down all common network interfaces in the Recovery-state fabrics except for the following interfaces:
    - Interfaces automatically excluded from being shut down by the system.
    - Interfaces specified by using the **`mad exclude interface`** command.

Failure recovery

To merge two split IRF fabrics, first repair the failed IRF link and remove the IRF link failure.

When the failed IRF link between two split IRF fabrics is recovered, all member devices in the inactive IRF fabric automatically reboot to join the active IRF fabric as subordinate members. The

network interfaces that have been shut down by MAD automatically restore their original state, as shown in Figure 6.

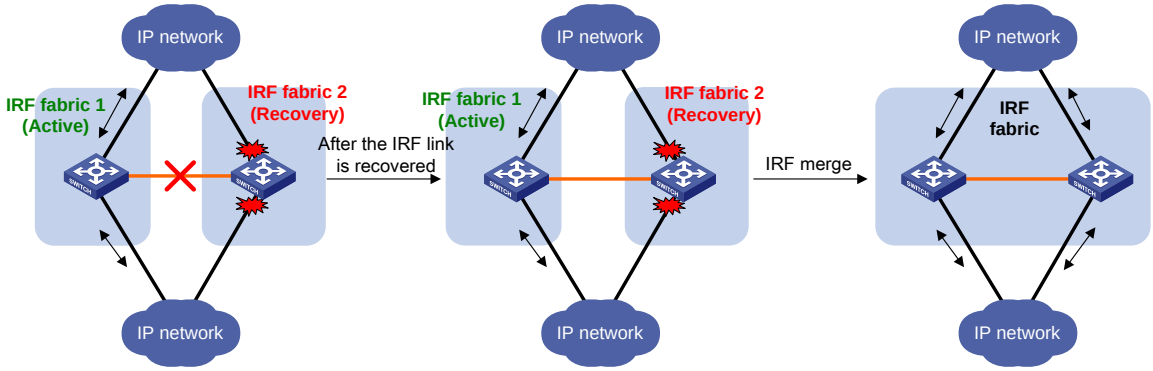
CAUTION:

If you inadvertently reboot the active IRF fabric after the failed IRF link recovers, its member devices will join the inactive IRF fabric with their network interfaces being shut down by MAD. To restore the original states of the network interfaces in the merged IRF fabric, use the **mad restore** command.

NOTE:

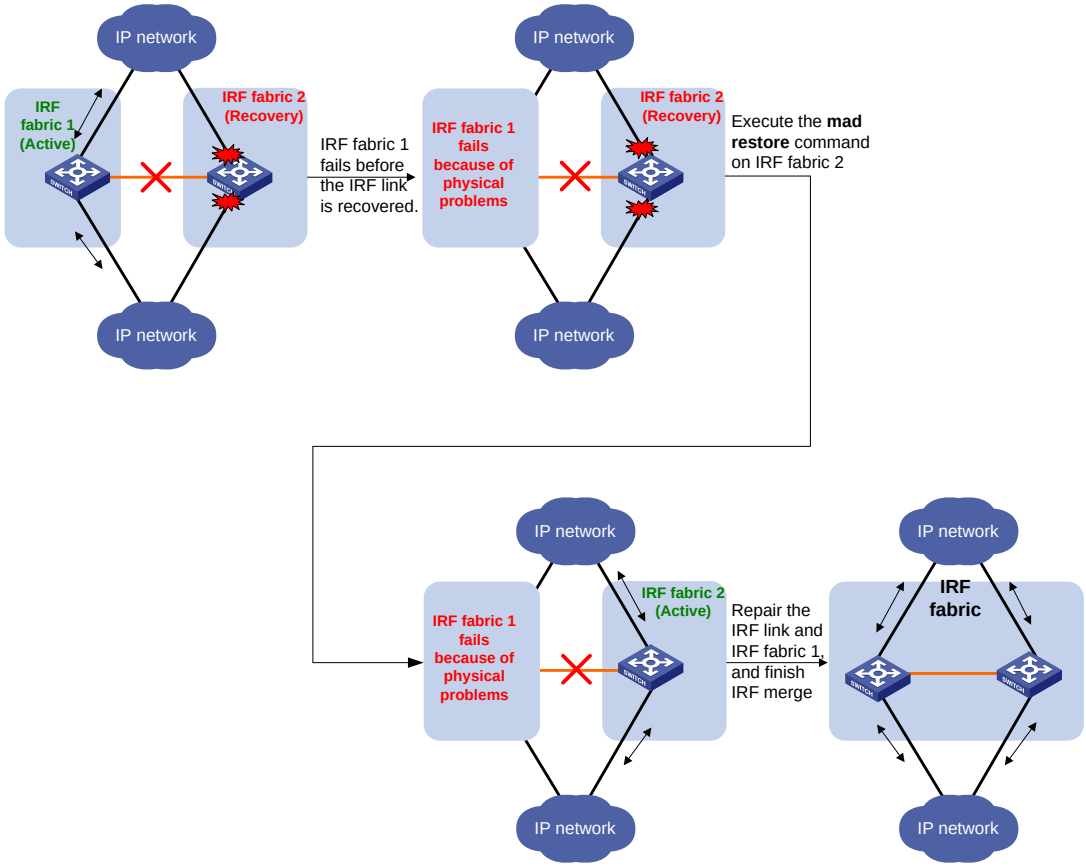
If the IRF auto-merge feature is enabled, the inactive IRF member devices will automatically reboot after the failed IRF link recovers and a manual reboot is typically not required.

Figure 6 Recovering the IRF fabric



If the active IRF fabric fails before the IRF link is recovered (see Figure 7), use the **mad restore** command on the inactive IRF fabric to recover the inactive IRF fabric. This command brings up all network interfaces that were shut down by MAD. After the IRF link is repaired, merge the two parts into a unified IRF fabric.

Figure 7 Active IRF fabric fails before the IRF link is recovered



MAD mechanisms

IRF provides MAD mechanisms by extending LACP, BFD, ARP, and IPv6 ND.

Table 1 compares the MAD mechanisms and their application scenarios.

Table 1 Comparison of MAD mechanisms

MAD mechanism	Advantages	Disadvantages	Application scenarios
LACP MAD	• Detection speed is fast. • Runs on existing aggregate links without requiring MAD-dedicated physical links or Layer 3 interfaces.	Requires an intermediate device that supports extended LACP for MAD.	Link aggregation is used between the IRF fabric and its upstream or downstream device.
BFD MAD	• Detection speed is fast. • Intermediate device, if used, can come from any vendor.	Requires MAD dedicated physical links and Layer 3 interfaces, which cannot be used for transmitting user traffic.	• No special requirements for network scenarios. • If no intermediate device is used, this mechanism is only suitable for IRF fabrics that have only two members that are geographically

MAD mechanism	Advantages	Disadvantages	Application scenarios
			close to one another.
ARP MAD	• No intermediate device is required. • Intermediate device, if used, can come from any vendor. • Does not require MAD dedicated ports.	• Detection speed is slower than BFD MAD and LACP MAD. • The spanning tree feature must be enabled if common Ethernet ports are used for ARP MAD links.	Spanning tree-enabled non-link aggregation IPv4 network scenarios if common Ethernet ports are used.
ND MAD	• No intermediate device is required. • Intermediate device, if used, can come from any vendor. • Does not require MAD dedicated ports.	• Detection speed is slower than BFD MAD and LACP MAD. • The spanning tree feature must be enabled if common Ethernet ports are used for ND MAD links.	Spanning tree-enabled non-link aggregation IPv6 network scenarios if common Ethernet ports are used.

LACP MAD

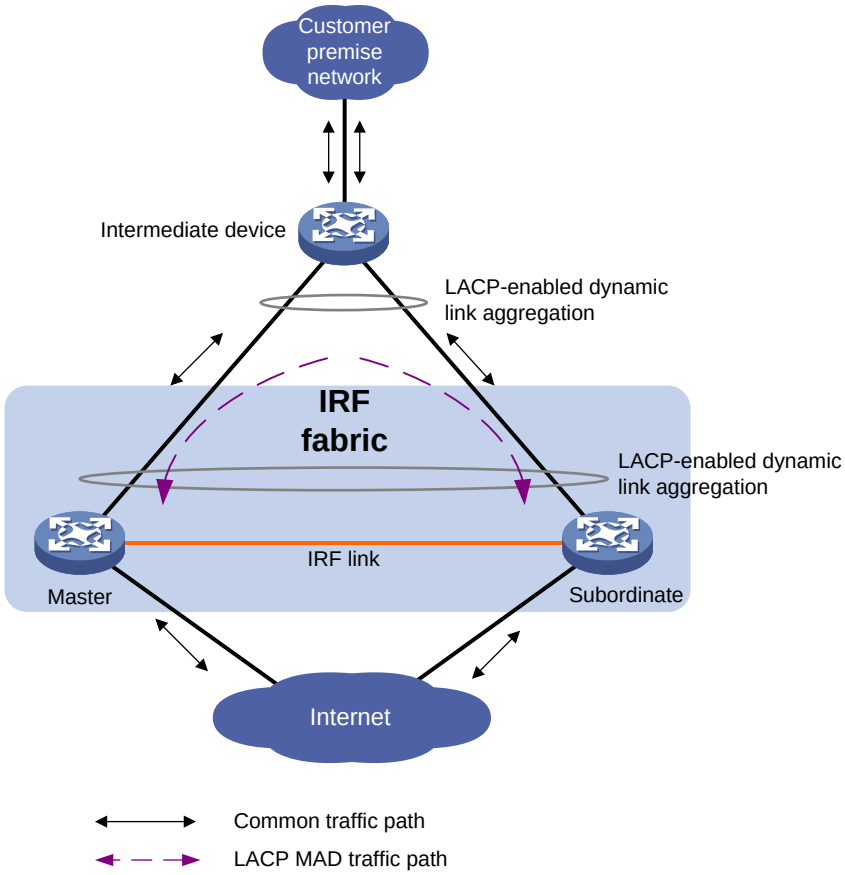
As shown in Figure 8, LACP MAD has the following requirements:

- Every IRF member must have a link with an intermediate device.
- All the links form a dynamic link aggregation group.
- The intermediate device must be a device that supports extended LACP for MAD.

The IRF member devices send extended LACPDUs that convey a domain ID and an active ID (the member ID of the master). The intermediate device transparently forwards the extended LACPDUs received from one member device to all the other member devices.

- If the domain IDs and active IDs sent by all the member devices are the same, the IRF fabric is integrated.
- If the extended LACPDUs convey the same domain ID but different active IDs, a split has occurred. LACP MAD handles this situation as described in "Collision handling."

Figure 8 LACP MAD scenario



BFD MAD

BFD MAD detects multi-active collisions by using BFD.

You can use common or management Ethernet ports for BFD MAD.

If management Ethernet ports are used, BFD MAD has the following requirements:

- Use an intermediate device and make sure each IRF member device has a BFD MAD link to the intermediate device.
- Assign each member device a MAD IP address on the master's management Ethernet port.

If common Ethernet ports are used, BFD MAD has the following requirements:

- If an intermediate device is used, each member device must have a BFD MAD link to the intermediate device. If no intermediate device is used, all member devices must have a BFD MAD link to each other.
- Assign all ports on the BFD MAD links to the same VLAN or Layer 3 aggregate interface. Assign each member device a MAD IP address on the VLAN interface or Layer 3 aggregate interface.

When you use BFD MAD, follow these restrictions and guidelines:

- As a best practice, use an intermediate device to connect IRF member devices if the IRF fabric has more than two member devices. A full mesh of IRF members might cause broadcast loops.
- As a best practice, preferentially use management Ethernet ports for BFD MAD.
- The BFD MAD links and BFD MAD VLAN (or Layer 3 aggregate interface) must be dedicated. Do not use the BFD MAD links or BFD MAD VLAN (or Layer 3 aggregate interface) for any other purposes.

- When you use a Layer 3 aggregate interface for BFD MAD, make sure its member ports do not exceed the maximum number of Selected ports allowed for an aggregation group. If the number of member ports exceeds the maximum number of Selected ports, some member ports cannot become Selected. BFD MAD will be unable to work correctly and its state will change to Faulty. For more information about setting the maximum number of Selected ports for an aggregation group, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*.

NOTE:

- The MAD addresses identify the member devices and must belong to the same subnet.
- Of all management Ethernet ports on an IRF fabric, only the master's management Ethernet port is accessible.

Figure 9 shows a typical BFD MAD scenario that uses an intermediate device. On the intermediate device, assign the ports on the BFD MAD links to the same VLAN.

Figure 10 shows a typical BFD MAD scenario that does not use an intermediate device.

With BFD MAD, the master attempts to establish BFD sessions with other member devices by using its MAD IP address as the source IP address.

- If the IRF fabric is integrated, only the MAD IP address of the master takes effect. The master cannot establish a BFD session with any other member. If you execute the **`display bfd session`** command, the state of the BFD sessions is **Down**.
- When the IRF fabric splits, the IP addresses of the masters in the split IRF fabrics take effect. The masters can establish a BFD session. If you execute the **`display bfd session`** command, the state of the BFD session between the two devices is **Up**.

Figure 9 BFD MAD scenario with an intermediate device

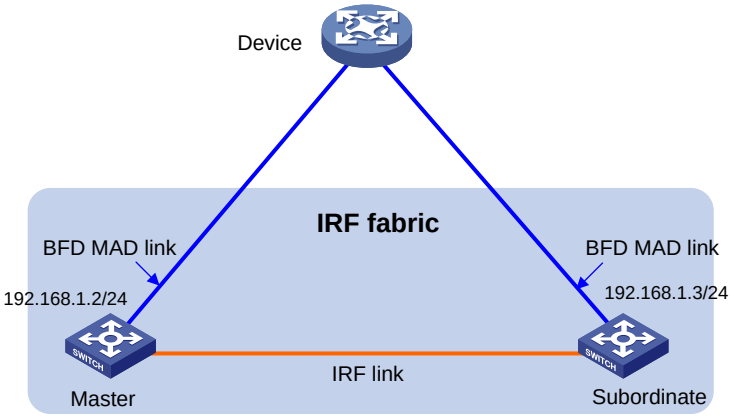
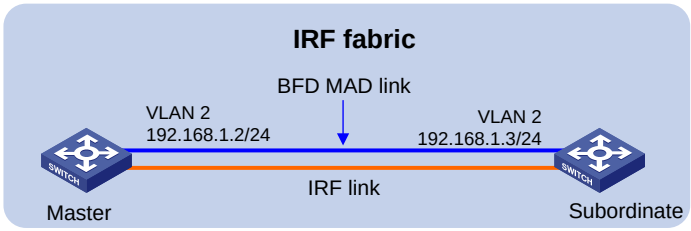


Figure 10 BFD MAD scenario without an intermediate device



ARP MAD

ARP MAD detects multi-active collisions by using extended ARP packets that convey the IRF domain ID and the active ID.

You can use common or management Ethernet ports for ARP MAD.

If management Ethernet ports are used, ARP MAD must work with an intermediate device. Make sure the following requirements are met:

- Connect a management Ethernet port on each member device to the intermediate device.
- On the intermediate device, you must assign the ports used for ARP MAD to the same VLAN.

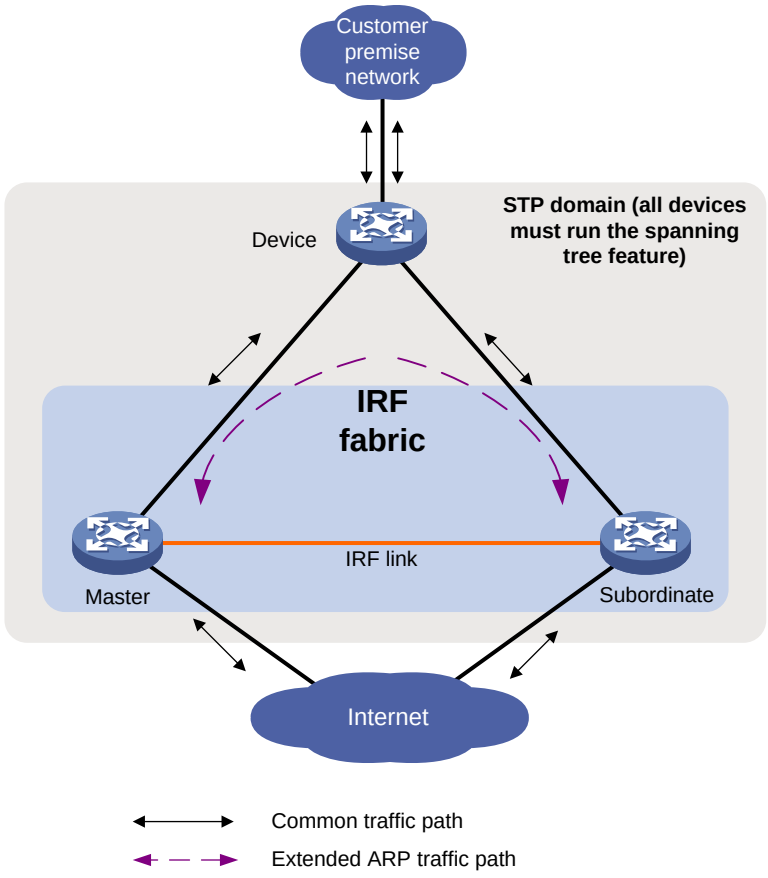
If common Ethernet ports are used, ARP MAD can work with or without an intermediate device. Make sure the following requirements are met:

- If an intermediate device is used, connect each IRF member device to the intermediate device, as shown in Figure 11. Run the spanning tree feature between the IRF fabric and the intermediate device. In this situation, data links can be used.
- If no intermediate device is used, connect each IRF member device to all other member devices. In this situation, IRF links cannot be used for ARP MAD.

Each IRF member compares the domain ID and the active ID (the member ID of the master) in incoming extended ARP packets with its domain ID and active ID.

- If the domain IDs are different, the extended ARP packet is from a different IRF fabric. The device does not continue to process the packet with the MAD mechanism.
- If the domain IDs are the same, the device compares the active IDs.
  - If the active IDs are different, the IRF fabric has split.
  - If the active IDs are the same, the IRF fabric is integrated.

Figure 11 ARP MAD scenario



ND MAD

ND MAD detects multi-active collisions by using NS packets to transmit the IRF domain ID and the active ID.

You can use common or management Ethernet ports for ND MAD.

If management Ethernet ports are used, ND MAD must work with an intermediate device. Make sure the following requirements are met:

- Connect a management Ethernet port on each member device to the intermediate device.
- On the intermediate device, you must assign the ports used for ND MAD to the same VLAN.

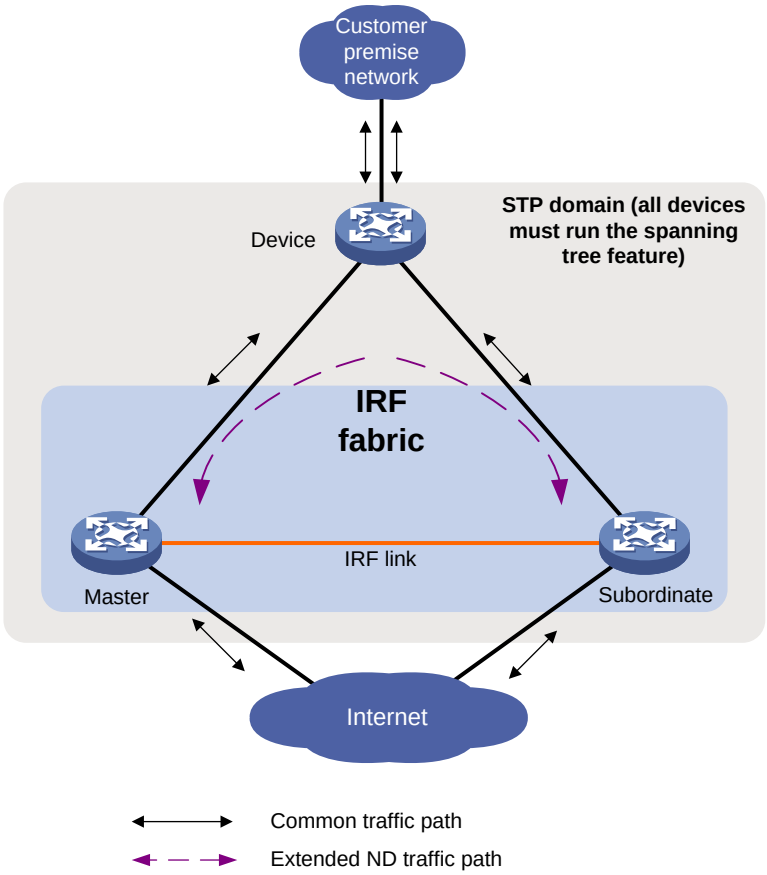
If common Ethernet ports are used, ND MAD can work with or without an intermediate device. Make sure the following requirements are met:

- If an intermediate device is used, connect each IRF member device to the intermediate device, as shown in Figure 12. Run the spanning tree feature between the IRF fabric and the intermediate device. In this situation, data links can be used.
- If no intermediate device is used, connect each IRF member device to all other member devices. In this situation, IRF links cannot be used for ND MAD.

Each IRF member device compares the domain ID and the active ID (the member ID of the master) in incoming NS packets with its domain ID and active ID.

- If the domain IDs are different, the NS packet is from a different IRF fabric. The device does not continue to process the packet with the MAD mechanism.
- If the domain IDs are the same, the device compares the active IDs.
    - If the active IDs are different, the IRF fabric has split.
    - If the active IDs are the same, the IRF fabric is integrated.

Figure 12 ND MAD scenario



Restrictions and guidelines: IRF configuration

Hardware compatibility with IRF

A switch in this series can form an IRF fabric only with switches in the same series.

Software requirements for IRF

All IRF member devices must run the same software image version. Make sure the software auto-update feature is enabled on all member devices.

IRF fabric size

An IRF fabric can contain a maximum of nine member devices.

As a best practice, use a maximum of four devices to form an IRF fabric. If an IRF fabric contains more than four member devices, do not use any network management software to manage those devices.

Candidate IRF physical interfaces

When you bind IRF physical interfaces to an IRF port, follow the restrictions and guidelines in Table 2.

Table 2 IRF port binding restrictions and guidelines

Device models	Candidate IRF physical interfaces	Binding restrictions
S5590-HI series S5500V3-HI series S5590-28T8XC-EI S5590-48T4XC-EI S5590-28S8XC-EI S5590-48S4XC-EI S5590-28P8XC-EI S5590-48P6XC-EI	• SFP+ ports on the front panel. • Ports on the rear expansion interface modules (if installed): ○ SFP+ ports except those on the LSWM2SP2PB or LSWM2SP4PB module. ○ SFP28 ports. ○ QSFP+ ports. ○ QSFP28 ports.	As a best practice, all IRF physical interfaces bound to an IRF port are operating at the same rate. IRF physical interfaces bound to an IRF port can come from the front panel, expansion interface modules, or both. Port speed requirements for IRF ports: • To use SFP+ ports as IRF physical interfaces, the ports must operate at 10 Gbps. • To use SFP28 ports as IRF physical interfaces, the ports must operate at 25 Gbps. • To use QSFP+ ports as IRF physical interfaces, the ports must operate at 40 Gbps. • To use QSFP28 ports as IRF physical interfaces, the ports must operate at 40 Gbps or 100 Gbps.
S5590-24UXM4YC-EI S5590-48UXM4YC-EI S5590-24X4YC-EI	• SFP28 ports on the front panel. • Ports on the rear expansion interface modules (if installed): ○ SFP+ ports except those on the LSWM2SP2PB or LSWM2SP4PB module. ○ SFP28 ports. ○ QSFP+ ports. ○ QSFP28 ports.	As a best practice, all IRF physical interfaces bound to an IRF port are operating at the same rate. IRF physical interfaces bound to an IRF port can come from the front panel, expansion interface modules, or both. Port speed requirements for IRF ports: • To use SFP+ ports as IRF physical interfaces, the ports must operate at 10 Gbps. • To use SFP28 ports as IRF physical interfaces, the ports must operate at 25 Gbps. • To use QSFP+ ports as IRF physical interfaces, the ports must operate at 40 Gbps. • To use QSFP28 ports as IRF physical interfaces, the ports must operate at 40 Gbps or 100 Gbps.

Device models	Candidate IRF physical interfaces	Binding restrictions
S5590-48TS4X2QC-EI	• QSFP+ ports on the front panel. • Ports on the rear expansion interface modules (if installed): ◦ SFP+ ports except those on the LSWM2SP2PB or LSWM2SP4PB module. ◦ SFP28 ports. ◦ QSFP+ ports. ◦ QSFP28 ports.	As a best practice, all IRF physical interfaces bound to an IRF port are operating at the same rate. IRF physical interfaces bound to an IRF port can come from the front panel, expansion interface modules, or both. Port speed requirements for IRF ports: • To use SFP+ ports as IRF physical interfaces, the ports must operate at 10 Gbps. • To use SFP28 ports as IRF physical interfaces, the ports must operate at 25 Gbps. • To use QSFP+ ports as IRF physical interfaces, the ports must operate at 40 Gbps. • To use QSFP28 ports as IRF physical interfaces, the ports must operate at 40 Gbps or 100 Gbps.
S5590-24UXMC-EI	• 100M, 1G, 2.5G, 5G, or 10G BASE-T ports on the front panel. • Ports on the rear expansion interface modules (if installed): ◦ SFP+ ports except those on the LSWM2SP2PB or LSWM2SP4PB module. ◦ SFP28 ports. ◦ QSFP+ ports. ◦ QSFP28 ports.	As a best practice, all IRF physical interfaces bound to an IRF port are operating at the same rate. IRF physical interfaces bound to an IRF port can come from the front panel, expansion interface modules, or both. Port speed requirements for IRF ports: • To use 100M, 1G, 2.5G, 5G, or 10G BASE-T ports as IRF physical interfaces, the ports must operate at 10 Gbps. • To use SFP+ ports as IRF physical interfaces, the ports must operate at 10 Gbps. • To use SFP28 ports as IRF physical interfaces, the ports must operate at 25 Gbps. • To use QSFP+ ports as IRF physical interfaces, the ports must operate at 40 Gbps. • To use QSFP28 ports as IRF physical interfaces, the ports must operate at 40 Gbps or 100 Gbps.

Transceiver modules and cables selection for IRF

When you select transceiver modules and cables, follow these restrictions and guidelines:

- To connect SFP+ ports in a long distance, use SFP+ transceiver modules and fibers. To connect SFP+ ports in a short distance, use SFP+ cables.

- To connect SFP28 ports in a long distance, use SFP28 transceiver modules and fibers. To connect SFP28 ports in a short distance, use SFP28 cables.
- To connect QSFP+ ports in a long distance, use QSFP+ transceiver modules and fibers. To connect QSFP+ ports in a short distance, use QSFP+ cables.
- For QSFP28 ports:

  Use the following media for a long-distance connection:

  - QSFP28 transceiver modules and fibers.
  - QSFP+ transceiver modules and fibers.

  Use QSFP28 or QSFP+ cables for a short-distance connection.
- Use category 5e twisted-pair cables to connect 100M/1G/2.5G/5G/10GBASE-T Ethernet ports.

For more information about types of transceiver modules and cables available on the switch for IRF setup, see the device installation guide. For more information about the transceiver models and cables available on the switch, see hardware information and specifications documentation for the switch.

For more information about the transceiver modules and DAC cables, see the device installation guide and *H3C Transceiver Modules User Guide*.

NOTE:

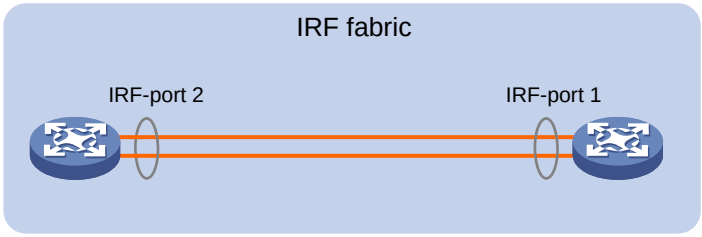
The transceiver modules and DAC cables available for the device are subject to change over time. For the most up-to-date list of transceiver modules and DAC cables, contact your H3C sales representative.

IRF port connection

When you connect two neighboring IRF members, follow these restrictions and guidelines:

- You must connect the physical interfaces of IRF-port 1 on one member to the physical interfaces of IRF-port 2 on the other.
- For high availability, bind multiple physical interfaces to an IRF port. You can bind a maximum of eight physical interfaces to an IRF port.

Figure 13 Connecting IRF physical interfaces



IRF physical interface configuration restrictions and guidelines

Command configuration restrictions

On a physical interface bound to an IRF port, you can execute only the following commands:

- Basic interface commands, including **shutdown** and **description**. For more information about these commands, see Ethernet interface commands in *Layer 2—LAN Switching Command Reference*.

- The **flow-interval** command, which sets the statistics polling interval on an interface. For more information about this command, see Ethernet interface commands in *Layer 2—LAN Switching Command Reference*.
- The **itu-channel** command for setting the ITU channel number for a transceiver module. For more information about this command, see device management commands in *Fundamentals Command Reference*.
- MAC address table configuration commands, including the **mac-address static source-check enable** command. For more information about these commands, see *Layer 2—LAN Switching Command Reference*.
- LLDP commands, including:
  - **lldp admin-status**.
  - **lldp check-change-interval**.
  - **lldp enable**.
  - **lldp encapsulation snap**.
  - **lldp notification remote-change enable**.
  - **lldp tlv-enable**.

  For more information about these commands, see LLDP in *Layer 2—LAN Switching Command Reference*.

Suppressing SNMP notifications of packet drops on IRF physical interfaces

Before an IRF member device forwards a packet, it examines its forwarding path in the IRF fabric for a loop. If a loop exists, the device discards the packet on the source interface of the looped path. This loop elimination mechanism will drop a large number of broadcast packets on the IRF physical interfaces.

To suppress SNMP notifications of packet drops that do not require attention, do not monitor packet forwarding on the IRF physical interfaces.

Feature compatibility and configuration restrictions with IRF

Routing settings

To form an IRF fabric, all member devices must use the same settings for the following routing features:

- Maximum number of ECMP routes (set by using the **max-ecmp-num** command).

For more information about the routing features, see basic IP routing in *Layer 3—IP Routing Configuration Guide*.

Port speed mode

To use S5590-HI switches to set up an IRF fabric, all candidate member devices must use the same port speed mode for expansion interface modules. By default, the port speed mode is 0 for expansion interface modules. To configure the port speed mode for expansion interface modules, use the **port-speed-mode** command. For the command configuration to take effect on a device, you must save the configuration and reboot the device. For more information about this command, see Ethernet interface commands in *Layer 2—LAN Switching Command Reference*.

Configuration rollback restrictions

The configuration rollback feature cannot roll back the following IRF settings:

- Member device description (set by using the **irf member description** command).
- Member device priority (set by using the **irf member priority** command).

- IRF physical interface and IRF port bindings (set by using the **`port group interface`** command).

For more information about the configuration rollback feature, see configuration file management in *Fundamentals Configuration Guide*.

IRF tasks at a glance

To configure IRF, perform the following tasks:

1. Configuring MAD

   Configure a minimum of one MAD mechanism on an IRF fabric. For the MAD compatibility, see "MAD mechanism compatibility."

   - Configuring LACP MAD
   - Configuring BFD MAD
   - Configuring ARP MAD
   - Configuring ND MAD
   - Excluding interfaces from the shutdown action upon detection of multi-active collision

     This feature excludes an interface from the shutdown action for management or other special purposes when an IRF fabric transits to the Recovery state.
   - Recovering an IRF fabric
2. (Optional.) Optimizing IRF settings for an IRF fabric
   - Enabling IRF auto-merge

     When two IRF fabrics merge, this feature enables the IRF fabric that failed the master election to automatically reboot all its member devices to complete the merge.
   - Configuring a member device description
   - Configuring IRF bridge MAC address settings
   - Enabling software auto-update for software image synchronization

     This feature automatically synchronizes the current software images of the master to devices that are attempting to join the IRF fabric.
   - Setting the IRF link down report delay

Planning the IRF fabric setup

Consider the following items when you plan an IRF fabric:

- Hardware compatibility and restrictions.
- IRF fabric size.
- Master device.
- Member ID and priority assignment scheme.
- Fabric topology and cabling scheme.
- IRF physical interfaces.

Setting up an IRF fabric

IRF setup tasks at a glance

To set up an IRF fabric, perform the following tasks:

1. Configure member IDs, priorities, and IRF physical interfaces separately.
   a. Assigning a member ID to each IRF member device
   b. (Optional.) Specifying a priority for each member device
   c. Binding physical interfaces to IRF ports

   Skip these tasks if you configure member IDs, priorities, domain ID, and IRF physical interfaces in bulk.
2. Bulk-configuring basic IRF settings for a member device

   Skip this task if you configure member IDs, priorities, domain ID, and IRF physical interfaces separately.
3. Connecting IRF physical interfaces
4. Accessing the IRF fabric

Assigning a member ID to each IRF member device

Restrictions and guidelines

CAUTION:

An IRF member ID change can invalidate member ID-related settings and cause data loss. Make sure you fully understand its impact on your live network.

To create an IRF fabric, you must assign a unique IRF member ID to each member device.

The new member ID of a device takes effect at a reboot. After the device reboots, the settings on all member ID-related physical resources (including common physical network interfaces) are removed, regardless of whether you have saved the configuration.

Procedure

Table 3 Enter system view.

system-view

2. Assign a member ID to a member device.

   irf member *member-id* **renumber** *new-member-id*

   The default IRF member ID is 1.

CAUTION:

An IRF member ID change can invalidate member ID-related settings and cause data loss. Make sure you fully understand its impact on your live network.

3. (Optional.) Save the configuration.

   save

   If you have bound physical interfaces to IRF ports or assigned member priority, you must perform this step for these settings to take effect after the reboot.
4. Return to user view.

   quit
5. Reboot the device.

   reboot [**slot** *slot-number*] [**force**]

Specifying a priority for each member device

About this task

IRF member priority represents the possibility for a device to be elected the master in an IRF fabric. A larger priority value indicates a higher priority.

A change to member priority affects the election result at the next master election, but it does not cause an immediate master re-election.

Procedure

1. Enter system view.
   system-view
2. Specify a priority for the device.
   irf member *member-id* **priority** *priority*
   The default IRF member priority is 1.

Binding physical interfaces to IRF ports

Restrictions and guidelines

Select qualified physical interfaces as IRF physical interfaces as described in "Candidate IRF physical interfaces."

Make sure the IRF physical interfaces of an IRF port use the same binding mode.

After binding physical interfaces to IRF ports for the first time, you must use the **irf-port-configuration active** command to activate the settings on the IRF ports.

The system activates the IRF port settings automatically only in the following situations:

- The configuration file that the device starts with contains IRF port bindings.
- You are adding physical interfaces to an IRF port (in UP state) after an IRF fabric is formed.

Procedure

1. Enter system view.
   system-view
2. Enter interface view or interface range view.
   - Enter interface view.
     interface *interface-type interface-number*
   - Enter interface range view. Choose one of the following commands:
     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>
     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]

   To shut down a range of IRF physical interfaces, enter interface range view.

   To shut down one IRF physical interface, enter its interface view.
3. Shut down the physical interfaces.
   shutdown
   By default, interfaces on the device are disabled (in **ADM** or **Administratively Down** state).
   You must always shut down a physical interface before binding it to an IRF port or removing the binding. If the system prevents you from shutting down an interface, follow the system instructions to disable its peer interface.

4. Return to system view.

   quit

5. Enter IRF port view.

   irf-port *member-id/irf-port-number*

6. Bind each physical interface to the IRF port.

   port group interface *interface-type interface-number* [**mode enhanced**]

   By default, no physical interfaces are bound to an IRF port.

   Repeat this step to assign multiple physical interfaces to the IRF port.

7. Return to system view.

   quit

8. Enter interface view or interface range view.
   - Enter interface view.

     interface *interface-type interface-number*

   - Enter interface range view. Choose one of the following commands:

     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>

     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]

9. Bring up the physical interfaces.

   undo shutdown

10. Return to system view.

    quit

11. Save the configuration.

    save

    Activating IRF port configurations causes IRF merge and reboot. To avoid data loss, save the running configuration to the startup configuration file before you perform the operation.

12. Activate the IRF port settings.

    irf-port-configuration active

Bulk-configuring basic IRF settings for a member device

About this task

Use the easy IRF feature to bulk-configure basic IRF settings for a member device, including the member ID, domain ID, priority, and IRF port bindings.

The easy IRF feature provides the following configuration methods:

- **Interactive method**—Enter the **easy-irf** command without parameters. The system will guide you to set the parameters step by step.
- **Non-interactive method**—Enter the **easy-irf** command with parameters.

As a best practice, use the interactive method if you are new to IRF.

Restrictions and guidelines

CAUTION:

- Use caution when you change the member ID of an IRF member device. An IRF member ID uniquely identifies a device in an IRF fabric. An IRF member ID change can invalidate member ID-related settings, including interface and file path settings, and cause data loss. Make sure you fully understand its impact on the live network.
- The member device reboots immediately after you specify a new member ID for it. Make sure you are aware of the impact of this operation on the network.

If you execute the **easy-irf** command multiple times, the following settings take effect:

- The most recent settings for the member ID, domain ID, and priority.
- IRF port bindings added through repeated executions of the command. To remove an IRF physical interface from an IRF port, you must use the **undo port group interface** command in IRF port view.

If you specify IRF physical interfaces by using the interactive method, you must also follow these restrictions and guidelines:

- Do not enter spaces between the interface type and interface number.
- Use a comma (,) to separate two physical interfaces. No spaces are allowed between interfaces.

Procedure

1. Enter system view.

   system-view

2. Bulk-configure basic IRF settings for the device.

   easy-irf [**member** *member-id* [**renumber** *new-member-id*] **domain** *domain-id* [**priority** *priority*] [**irf-port1** *interface-list1*] [**irf-port2** *interface-list2*]]

   Make sure the new member ID is unique in the IRF fabric to which the device will be added.

Connecting IRF physical interfaces

Follow the restrictions in "IRF port connection" to connect IRF physical interfaces as well as based on the topology and cabling scheme. The devices perform master election. The member devices that fail the master election automatically reboot to form an IRF fabric with the master device.

Accessing the IRF fabric

The IRF fabric appears as one device after it is formed. You configure and manage all IRF members at the CLI of the master. All settings you have made are automatically propagated to the IRF members.

The following methods are available for accessing an IRF fabric:

- **Local login**—Log in through the console port of any member device.
- **Remote login**—Log in at a Layer 3 interface on any member device by using methods including Telnet and SNMP.

When you log in to an IRF fabric, you are placed at the CLI of the master, regardless of at which member device you are logged in.

For more information, see login configuration in *Fundamentals Configuration Guide*.

Configuring MAD

Restrictions and guidelines for MAD configuration

MAD mechanism compatibility

As a best practice, configure a minimum of one MAD mechanism on an IRF fabric for prompt IRF split detection. Because MAD mechanisms use different collision handling processes, follow these restrictions and guidelines when you configure multiple MAD mechanisms on an IRF fabric:

- Do not configure LACP MAD together with ARP MAD or ND MAD.
- Do not configure BFD MAD together with ARP MAD or ND MAD.

Assigning IRF domain IDs

An IRF fabric has only one IRF domain ID. You can change the IRF domain ID by using the following commands: **irf domain**, **mad enable**, **mad arp enable**, or **mad nd enable**. The IRF domain IDs configured by using these commands overwrite each other.

If LACP MAD, ARP MAD, or ND MAD runs between two IRF fabrics, assign each fabric a unique IRF domain ID. (For BFD MAD, this task is optional.)

Actions on interfaces shut down by MAD

To prevent a multi-active collision from causing network issues, avoid using the **undo shutdown** command to bring up the interfaces shut down by a MAD mechanism on a Recovery-state IRF fabric.

Configuring LACP MAD

1. Enter system view.
   system-view
2. Assign a domain ID to the IRF fabric.
   irf domain *domain-id*
   The default IRF domain ID is 0.

CAUTION:
Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the remaining member devices in the IRF fabric.

3. Create an aggregate interface and enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*

   Perform this step also on the intermediate device.
4. Configure the aggregation group to operate in dynamic aggregation mode.
   link-aggregation mode dynamic
   By default, an aggregation group operates in static aggregation mode.
   LACP MAD takes effect only on dynamic aggregate interfaces.

Perform this step also on the intermediate device.

5. Enable LACP MAD.

   mad enable

   By default, LACP MAD is disabled.

6. Return to system view.

   quit

7. Enter Ethernet interface view or interface range view.
   - Enter Ethernet interface view.

     interface *interface-type interface-number*
   - Enter interface range view. Choose one of the following commands:

     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>

     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]

   To assign a range of ports to the aggregation group, enter interface range view.

   To assign one port to the aggregation group, enter Ethernet interface view.

8. Assign the Ethernet port or the range of Ethernet ports to the specified aggregation group.

   port link-aggregation group *group-id*

   Multichassis link aggregation is allowed.

   Perform this step also on the intermediate device.

Configuring BFD MAD

Restrictions and guidelines for configuring BFD MAD

As a best practice, use the following procedure to set up BFD MAD:

1. Choose a BFD MAD link scheme as described in "BFD MAD."
2. Configure BFD MAD.
3. Connect the BFD MAD links.

When you configure BFD MAD on a VLAN interface, follow these restrictions and guidelines:

Category	Restrictions and guidelines
BFD MAD VLAN	• Do not enable BFD MAD on VLAN-interface 1. • If you are using an intermediate device, perform the following tasks: ○ On the IRF fabric and the intermediate device, create a VLAN for BFD MAD. ○ On the IRF fabric and the intermediate device, assign the ports of BFD MAD links to the BFD MAD VLAN. ○ On the IRF fabric, create a VLAN interface for the BFD MAD VLAN. • Make sure the IRF fabrics on the network use different BFD MAD VLANs. • Make sure the BFD MAD VLAN contains only ports on the BFD MAD links. Exclude a port from the BFD MAD VLAN if that port is not on a BFD MAD link. If you have assigned that port to all VLANs by using the **port trunk permit vlan all** command, use the **undo port trunk permit** command to exclude that port from the BFD MAD VLAN. • After an IRF fabric split, the system handles the BFD MAD VLAN and its interface in the IRF fabric that lost the master election as follows: ○ Retains the BFD MAD VLAN interface.

Category	Restrictions and guidelines
	○ Shuts down all Layer 2 Ethernet interfaces in the BFD MAD VLAN except one on the master device for BFD MAD detection. The interfaces that shut down are placed in MAD down state.
BFD MAD VLAN and feature compatibility	Do not use the BFD MAD VLAN and its member ports for any purpose other than BFD MAD. • Use only the **mad bfd enable** and **mad ip address** commands on the BFD MAD-enabled VLAN interface. If you configure other features, both BFD MAD and other features on the interface might run incorrectly. • Disable the spanning tree feature on any Layer 2 Ethernet ports in the BFD MAD VLAN. The MAD feature is mutually exclusive with the spanning tree feature. • You can use multiple Layer 2 Ethernet ports to establish BFD MAD links directly between the member devices or across a relay device, for redundancy. To avoid loops in this situation, take the following actions: ○ Create one Layer 2 aggregate interface for each member device. ○ Assign all the ports used on each member device for BFD MAD to one of those aggregate interfaces. ○ Make sure each aggregate interface contains only ports on the same member device. ○ Assign the aggregate interfaces to the BFD MAD VLAN.
MAD IP address	• To avoid network issues, only use the **mad ip address** command to configure IP addresses on the BFD MAD-enabled VLAN interface. Do not configure an IP address by using the **ip address** command or configure a VRRP virtual address on the BFD MAD-enabled VLAN interface. • Make sure all the MAD IP addresses are on the same subnet. • To make sure BFD MAD can operate correctly, do not use the BFD MAD IP address to create BFD sessions for other services.

When you configure BFD MAD on a Layer 3 aggregate interface, follow these restrictions and guidelines:

Category	Restrictions and guidelines
BFD MAD-enabled Layer 3 aggregate interface	• Make sure the Layer 3 aggregate interface operates in static aggregation mode. • Make sure the member ports in the aggregation group do not exceed the maximum number of Selected ports allowed for an aggregation group. If the number of member ports exceeds the maximum number of Selected ports, some member ports cannot become Selected. BFD MAD will be unable to work correctly and its state will change to Faulty.
BFD MAD VLAN	If a switch is used as the intermediate device, use a VLAN as the BFD MAD VLAN on the intermediate device. • If the intermediate device acts as a BFD MAD intermediate device for multiple IRF fabrics, assign different BFD MAD VLANs to the IRF fabrics. • Assign the ports on the BFD MAD links for the same IRF fabric to the same VLAN. Do not assign the ports to an aggregate interface. If the ports are hybrid ports, make sure these ports are untagged members of their PVIDs. • Do not use the BFD MAD VLAN on the intermediate device for any purposes other than BFD MAD. • Make sure the BFD MAD VLAN on the intermediate device contains only ports on the BFD MAD links. Exclude a port from the BFD MAD VLAN if that port is not on a BFD MAD link. If you have assigned that port to all VLANs by using the **port trunk permit vlan all** command,

Category	Restrictions and guidelines
	use the **undo port trunk permit** command to exclude that port from the BFD MAD VLAN.
BFD MAD-enabled Layer 3 aggregate interface and feature compatibility	Use only the **mad bfd enable** and **mad ip address** commands on the BFD MAD-enabled interface. If you configure other features, both BFD MAD and other features on the interface might run incorrectly.
MAD IP address	• To avoid network issues, only use the **mad ip address** command to configure IP addresses on the BFD MAD-enabled interface. Do not configure an IP address by using the **ip address** command or configure a VRRP virtual address on the BFD MAD-enabled interface. • Make sure all the MAD IP addresses are on the same subnet. • To make sure BFD MAD can operate correctly, do not use the BFD MAD IP address to create BFD sessions for other services.

When you configure BFD MAD on a management Ethernet port, follow these restrictions and guidelines:

Category	Restrictions and guidelines
Management Ethernet ports for BFD MAD	Connect a management Ethernet port on each IRF member device to the common Ethernet ports on the intermediate device.
BFD MAD VLAN	• On the intermediate device (if any), create a VLAN for BFD MAD, and assign the ports used for BFD MAD to the VLAN. On the IRF fabric, you do not need to assign the management Ethernet ports to the VLAN. • Make sure the IRF fabrics on the network use different BFD MAD VLANs. • Make sure the BFD MAD VLAN on the intermediate device contains only ports on the BFD MAD links.
MAD IP address	• Use the **mad ip address** command instead of the **ip address** command to configure MAD IP addresses on the BFD MAD-enabled management Ethernet ports. • Make sure all the MAD IP addresses are on the same subnet. • To make sure BFD MAD can operate correctly, do not use the BFD MAD IP address to create BFD sessions for other services.

Configuring BFD MAD on a VLAN interface

1. Enter system view.
   system-view
2. (Optional.) Assign a domain ID to the IRF fabric.
   irf domain *domain-id*
   By default, the domain ID of an IRF fabric is 0.

CAUTION:
Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the remaining member devices in the IRF fabric.

3. Create a VLAN dedicated to BFD MAD.
   vlan *vlan-id*
   By default, only VLAN 1 exists.
   Do not enable BFD MAD on VLAN-interface 1.
   Perform this step also on the intermediate device (if any).

4. Return to system view.
   quit
5. Enter Ethernet interface view or interface range view.
   - Enter Ethernet interface view.
     interface *interface-type interface-number*
   - Enter interface range view. Choose one of the following commands:
     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>
     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]

   To assign a range of ports to the BFD MAD VLAN, enter interface range view.
   To assign one port to the BFD MAD VLAN, enter Ethernet interface view.
6. Assign the port or the range of ports to the BFD MAD VLAN.
   - Assign the ports to the VLAN as access ports.
     port access vlan *vlan-id*
   - Assign the ports to the VLAN as trunk ports.
     port trunk permit vlan *vlan-id*
   - Assign the ports to the VLAN as hybrid ports.
     port hybrid vlan *vlan-id* { **tagged** | **untagged** }

   The link type of BFD MAD ports can be access, trunk, or hybrid.
   The default link type of a port is access.
   Perform this step also on the intermediate device (if any).
7. Return to system view.
   quit
8. Enter VLAN interface view.
   interface vlan-interface *vlan-interface-id*
9. Enable BFD MAD.
   mad bfd enable
   By default, BFD MAD is disabled.
10. Assign a MAD IP address to a member device on the VLAN interface.
    mad ip address *ip-address* { *mask* | *mask-length* } **member** *member-id*
    By default, no MAD IP addresses are configured on any VLAN interfaces.
    Repeat this step to assign a MAD IP address to each member device on the VLAN interface.

Configuring BFD MAD on a Layer 3 aggregate interface

1. Enter system view.
   system-view
2. (Optional.) Assign a domain ID to the IRF fabric.
   irf domain *domain-id*
   By default, the domain ID of an IRF fabric is 0.

CAUTION:

Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the remaining member devices in the IRF fabric.

3. Create a Layer 3 aggregate interface for BFD MAD.

   interface route-aggregation *interface-number*

4. Return to system view.

   quit

5. Enter interface view or interface range view.

   - Enter Ethernet interface view.

     interface *interface-type interface-number*

   - Enter interface range view. Choose one of the following commands:

     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>

     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]

   To assign a range of ports to the aggregation group for the aggregate interface, enter interface range view.

   To assign one port to the aggregation group for the aggregate interface, enter Ethernet interface view.

6. Assign the port or the range of ports to the aggregation group for the aggregate interface.

   port link-aggregation group *number*

7. Return to system view.

   quit

8. Enter Layer 3 aggregate interface view.

   interface route-aggregation *interface-number*

9. Enable BFD MAD.

   mad bfd enable

   By default, BFD MAD is disabled.

10. Assign a MAD IP address to a member device on the Layer 3 aggregate interface.

    mad ip address *ip-address* { *mask* | *mask-length* } **member** *member-id*

    By default, no MAD IP addresses are configured on aggregate interfaces.

    Repeat this step to assign a MAD IP address to each member device on the aggregate interface.

Configuring BFD MAD on a management Ethernet port

1. Enter system view.

   system-view

2. (Optional.) Assign a domain ID to the IRF fabric.

   irf domain *domain-id*

   By default, the domain ID of an IRF fabric is 0.

CAUTION:

Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the

remaining member devices in the IRF fabric.

3. Enter management Ethernet interface view.
   interface m-gigabitethernet *interface-number*
   Of all management Ethernet ports on an IRF fabric, only the master's management Ethernet port is accessible.
4. Enable BFD MAD.
   mad bfd enable
   By default, BFD MAD is disabled.
5. Assign a MAD IP address to each member device.
   mad ip address *ip-address* { *mask* | *mask-length* } **member** *member-id*
   By default, no MAD IP addresses are configured.

Configuring ARP MAD

Restrictions and guidelines for configuring ARP MAD

As a best practice, use the following procedure to set up ARP MAD:

1. Choose an ARP MAD link scheme as described in "ARP MAD."
2. Configure ARP MAD.
3. Connect the ARP MAD links.

When you configure ARP MAD on a VLAN interface, follow these restrictions and guidelines:

Category	Restrictions and guidelines
ARP MAD VLAN	• Do not enable ARP MAD on VLAN-interface 1. • If you are using an intermediate device, perform the following tasks: ○ On the IRF fabric and the intermediate device, create a VLAN for ARP MAD. ○ On the IRF fabric and the intermediate device, assign the ports of ARP MAD links to the ARP MAD VLAN. ○ On the IRF fabric, create a VLAN interface for the ARP MAD VLAN. • Do not use the ARP MAD VLAN for any other purposes.
ARP MAD and feature configuration	If an intermediate device is used, make sure the following requirements are met: • Run the spanning tree feature between the IRF fabric and the intermediate device to ensure that there is only one ARP MAD link in forwarding state. For more information about the spanning tree feature and its configuration, see *Layer 2—LAN Switching Configuration Guide*. • Enable the IRF fabric to change its bridge MAC address as soon as the address owner leaves. • If the intermediate device is also an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection.

When you configure ARP MAD on a management Ethernet port, follow these restrictions and guidelines:

Category	Restrictions and guidelines
Management Ethernet ports for ARP MAD	Connect a management Ethernet port on each member device to the common Ethernet ports on the intermediate device.

Category	Restrictions and guidelines
ARP MAD VLAN	On the intermediate device, create a VLAN for ARP MAD, and assign the ports used for ARP MAD to the VLAN. On the IRF fabric, you do not need to assign the management Ethernet ports to the VLAN.
ARP MAD and feature configuration	• Enable the IRF fabric to change its bridge MAC address as soon as the address owner leaves. • If the intermediate device is also an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection.

Configuring ARP MAD on a VLAN interface

1. Enter system view.

   system-view
2. Assign a domain ID to the IRF fabric.

   irf domain *domain-id*

   The default IRF domain ID is 0.

   CAUTION:

   Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the remaining member devices in the IRF fabric.
3. Configure the IRF bridge MAC address to change as soon as the address owner leaves.

   undo irf mac-address persistent

   By default, the IRF bridge MAC address does not change after the address owner leaves.

   CAUTION:

   IRF bridge MAC address change will cause transient traffic disruption.
4. Create a VLAN dedicated to ARP MAD.

   vlan *vlan-id*

   By default, only VLAN 1 exists.

   Do not configure ARP MAD on VLAN-interface 1.

   Perform this task also on the intermediate device (if any).
5. Return to system view.

   quit
6. Enter Ethernet interface view or interface range view.
   - Enter Ethernet interface view.

     interface *interface-type interface-number*
   - Enter interface range view. Choose one of the following commands:

     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>

     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]

   To assign a range of ports to the ARP MAD VLAN, enter interface range view.

   To assign one port to the ARP MAD VLAN, enter Ethernet interface view.
7. Assign the port or the range of ports to the ARP MAD VLAN.
   - Assign the ports to the VLAN as access ports.

port access vlan *vlan-id*

- Assign the ports to the VLAN as trunk ports.
  port trunk permit vlan *vlan-id*
- Assign the ports to the VLAN as hybrid ports.
  port hybrid vlan *vlan-id* { **tagged** | **untagged** }

The link type of ARP MAD ports can be access, trunk, or hybrid.

The default link type of a port is access.

Perform this task also on the intermediate device (if any).

8. Return to system view.
   quit
9. Enter VLAN interface view.
   interface vlan-interface *vlan-interface-id*
10. Assign the interface an IP address.
    ip address *ip-address* { *mask* | *mask-length* }

    By default, no IP addresses are assigned to any VLAN interfaces.
11. Enable ARP MAD.
    mad arp enable

    By default, ARP MAD is disabled.

Configuring ARP MAD on a management Ethernet port

1. Enter system view.
   system-view
2. Assign a domain ID to the IRF fabric.
   irf domain *domain-id*

   The default IRF domain ID is 0.

CAUTION:

Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the remaining member devices in the IRF fabric.

3. Configure the IRF bridge MAC address to change as soon as the address owner leaves.
   undo irf mac-address persistent

   By default, the IRF bridge MAC address does not change after the address owner leaves.

CAUTION:

IRF bridge MAC address change will cause transient traffic disruption.

4. Enter management Ethernet interface view.
   interface m-gigabitethernet *interface-number*

   Of all management Ethernet ports on an IRF fabric, only the master's management Ethernet port is accessible.
5. Assign an IP address to the management Ethernet port.
   ip address ip-address { *mask* | *mask-length* }

   By default, no IP addresses are configured.
6. Enable ARP MAD.
   mad arp enable

By default, ARP MAD is disabled.

Configuring ND MAD

Restrictions and guidelines for configuring ND MAD

As a best practice, use the following procedure to set up ND MAD:

1. Choose an ND MAD link scheme as described in "ND MAD."
2. Configure ND MAD.
3. Connect the ND MAD links.

When you configure ND MAD on a VLAN interface, follow these restrictions and guidelines:

Category	Restrictions and guidelines
ND MAD VLAN	• Do not enable ND MAD on VLAN-interface 1. • If you are using an intermediate device, perform the following tasks: ○ On the IRF fabric and the intermediate device, create a VLAN for ND MAD. ○ On the IRF fabric and the intermediate device, assign the ports of ND MAD links to the ND MAD VLAN. ○ On the IRF fabric, create a VLAN interface for the ND MAD VLAN. • Do not use the ND MAD VLAN for any other purposes.
ND MAD and feature configuration	If an intermediate device is used, make sure the following requirements are met: • Run the spanning tree feature between the IRF fabric and the intermediate device to ensure that there is only one ND MAD link in forwarding state. For more information about the spanning tree feature and its configuration, see *Layer 2—LAN Switching Configuration Guide*. • Enable the IRF fabric to change its bridge MAC address as soon as the address owner leaves. • If the intermediate device is also an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection.

When you configure ND MAD on a management Ethernet port, follow these restrictions and guidelines:

Category	Restrictions and guidelines
Management Ethernet ports for ND MAD	Connect a management Ethernet port on each member device to the common Ethernet ports on the intermediate device.
ND MAD VLAN	On the intermediate device, create a VLAN for ND MAD, and assign the ports used for ND MAD to the VLAN. On the IRF fabric, you do not need to assign the management Ethernet ports to the VLAN.
ND MAD and feature configuration	• Enable the IRF fabric to change its bridge MAC address as soon as the address owner leaves. • If the intermediate device is also an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection.

Configuring ND MAD on a VLAN interface

1. Enter system view.
   `system-view`
2. Assign a domain ID to the IRF fabric.
   `irf domain` *`domain-id`*

The default IRF domain ID is 0.

CAUTION:

Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the remaining member devices in the IRF fabric.

3. Configure the IRF bridge MAC address to change as soon as the address owner leaves.

   undo irf mac-address persistent

   By default, the IRF bridge MAC address does not change after the address owner leaves.

CAUTION:

IRF bridge MAC address change will cause transient traffic disruption.

4. Create a VLAN dedicated to ND MAD.

   vlan *vlan-id*

   By default, only VLAN 1 exists.

   Do not configure ND MAD on VLAN-interface 1.

   Perform this task also on the intermediate device (if any).

5. Return to system view.

   quit

6. Enter Ethernet interface view or interface range view.

   - Enter Ethernet interface view.

     interface *interface-type interface-number*

   - Enter interface range view. Choose one of the following commands:

     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>

     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]

   To assign a range of ports to the ND MAD VLAN, enter interface range view.

   To assign one port to the ND MAD VLAN, enter Ethernet interface view.

7. Assign the port or the range of ports to the ND MAD VLAN.

   - Assign the ports to the VLAN as access ports.

     port access vlan *vlan-id*

   - Assign the ports to the VLAN as trunk ports.

     port trunk permit vlan *vlan-id*

   - Assign the ports to the VLAN as hybrid ports.

     port hybrid vlan *vlan-id* { **tagged** | **untagged** }

   The link type of ND MAD ports can be access, trunk, or hybrid.

   The default link type of a port is access.

   Perform this task also on the intermediate device (if any).

8. Return to system view.

   quit

9. Enter VLAN interface view.

   interface vlan-interface *interface-number*

10. Assign the interface an IPv6 address.

ipv6 address { *ipv6-address/prefix-length* | *ipv6-address prefix-length* }

By default, no IPv6 addresses are assigned to a VLAN interface.

11. Enable ND MAD.

mad nd enable

By default, ND MAD is disabled.

Configuring ND MAD on a management Ethernet port

1. Enter system view.

system-view

2. Assign a domain ID to the IRF fabric.

irf domain *domain-id*

The default IRF domain ID is 0.

CAUTION:

Changing the IRF domain ID of an IRF member device will remove that member device from the IRF fabric. This member device will be unable to exchange IRF protocol packets with the remaining member devices in the IRF fabric.

3. Configure the IRF bridge MAC address to change as soon as the address owner leaves.

undo irf mac-address persistent

By default, the IRF bridge MAC address does not change after the address owner leaves.

CAUTION:

IRF bridge MAC address change will cause transient traffic disruption.

4. Enter management Ethernet interface view.

interface m-gigabitethernet *interface-number*

Of all management Ethernet ports on an IRF fabric, only the master's management Ethernet port is accessible.

5. Assign an IPv6 address to the management Ethernet port.

ipv6 address { *ipv6-address/pre-length* | *ipv6 address pre-length* }

By default, no IPv6 addresses are assigned to a management Ethernet port.

6. Enable ND MAD.

mad nd enable

By default, ND MAD is disabled.

Excluding interfaces from the shutdown action upon detection of multi-active collision

About this task

When an IRF fabric transits to the Recovery state, the system automatically excludes the following network interfaces from being shut down:

- IRF physical interfaces.
- Interfaces used for BFD MAD.
- Member interfaces of an aggregate interface if the aggregate interface is excluded from being shut down.

You can exclude an interface from the shutdown action for management or other special purposes. For example:

- Exclude a port from the shutdown action so you can Telnet to the port for managing the device.
- Exclude a VLAN interface and its Layer 2 ports from the shutdown action so you can log in through the VLAN interface.

Restrictions and guidelines

When you exclude a logical interface from the shutdown action, you must also exclude its associated physical interfaces. If you fail to do so, the logical interface will be unable to forward packets when the device is placed in Recovery state and the associated physical interfaces are shut down. For example, if a VLAN interface is used for remote login, you must exclude the VLAN interface and its associated physical interfaces from the shutdown action.

Procedure

1. Enter system view.

   system-view

2. Configure an interface to not shut down when the IRF fabric transits to the Recovery state.

   mad exclude interface *interface-type interface-number*

   By default, all network interfaces on a Recovery-state IRF fabric are shut down, except for the network interfaces automatically excluded by the system.

Recovering an IRF fabric

About this task

If the active IRF fabric fails before the IRF link is recovered, perform this task on the inactive IRF fabric to recover the inactive IRF fabric for traffic forwarding. The manual recovery operation brings up all interfaces that were shut down by MAD on the inactive IRF fabric.

Procedure

1. Enter system view.

   system-view

2. Recover the inactive IRF fabric.

   mad restore

Optimizing IRF settings for an IRF fabric

Enabling IRF auto-merge

About this task

When two IRF fabrics merge, you must reboot the member devices in the IRF fabric that fails in the master election. The auto-merge feature enables the IRF fabric to automatically reboot all its member devices to complete the merge.

If this feature is disabled, you must manually reboot the devices that failed the master election to complete the merge.

Procedure

1. Enter system view.

   system-view

2. Enable IRF auto-merge.

irf auto-merge enable

By default, this feature is enabled.

Configuring a member device description

1. Enter system view.

   system-view
2. Configure a description for a member device.

   irf member *member-id* **description** *text*

   By default, no member device description is configured.

Configuring IRF bridge MAC address settings

About this task

The bridge MAC address of a system must be unique on a switched LAN. IRF bridge MAC address identifies an IRF fabric by Layer 2 protocols (for example, LACP) on a switched LAN.

By default, an IRF fabric uses the bridge MAC address of the master as the IRF bridge MAC address. After the master leaves, the IRF bridge MAC address persists for a period of time or permanently depending on the IRF bridge MAC persistence setting. When the IRF bridge MAC persistence timer expires, the IRF fabric uses the bridge MAC address of the current master as the IRF bridge MAC address.

In special occasions that require a fixed special IRF bridge MAC address, you can specify that MAC address as the IRF bridge MAC address. For example, when you replace an IRF fabric as a whole, you can configure the new IRF fabric with the IRF bridge MAC address of the existing IRF fabric before the replacement to minimize service interruption.

The IRF bridge MAC persistence setting does not take effect on the manually specified IRF bridge MAC address.

If IRF fabric merge occurs, IRF determines the IRF bridge MAC address of the merged IRF fabric as follows:

1. When IRF fabrics merge, IRF ignores the IRF bridge MAC addresses and checks the bridge MAC address of each member device in the IRF fabrics. IRF merge fails if any two member devices have the same bridge MAC address.
2. After IRF fabrics merge, the merged IRF fabric uses the bridge MAC address of the merging IRF fabric that won the master election as the IRF bridge MAC address.

Restrictions and guidelines for IRF bridge MAC address configuration

CAUTION:

Bridge MAC address change will cause transient traffic disruption.

When you configure IRF bridge MAC persistence, follow these restrictions and guidelines:

- If ARP MAD or ND MAD is used with the spanning tree feature, you must disable IRF bridge MAC persistence by using the **undo irf mac-address persistent** command. In addition, do not specify a MAC address as the IRF bridge MAC address.
- If the IRF fabric has multichassis aggregate links, do not use the **undo irf mac-address persistent** command. Use of this command might cause traffic disruption.

Configuring IRF bridge MAC persistence

1. Enter system view.

   system-view

2. Configure IRF bridge MAC persistence.
   - Retain the bridge MAC address permanently even if the address owner has left the fabric.
     irf mac-address persistent always
   - Retain the bridge MAC address for 6 minutes after the address owner leaves the fabric.
     irf mac-address persistent timer
   - Change the bridge MAC address as soon as the address owner leaves the fabric.
     undo irf mac-address persistent

   By default, the IRF bridge MAC address does not change after the address owner leaves.

   The **irf mac-address persistent timer** command avoids unnecessary bridge MAC address changes caused by device reboot, transient link failure, or purposeful link disconnection.

Specifying a MAC address as the IRF bridge MAC address

1. Enter system view.
   system-view
2. Specify a MAC address as the IRF bridge MAC address.
   irf mac-address *mac-address*

   By default, an IRF fabric uses the bridge MAC address of the master as the IRF bridge MAC address.

CAUTION:
IRF bridge MAC address change will cause transient traffic disruption.

If an IRF fabric splits after you configure the IRF bridge MAC address, both the split IRF fabrics use the configured bridge MAC address as the IRF bridge MAC address.

Enabling software auto-update for software image synchronization

About this task

The software auto-update feature automatically synchronizes the current software images of the master to devices that are attempting to join the IRF fabric.

To join an IRF fabric, a device must use the same software images as the master in the fabric.

When you add a device to the IRF fabric, software auto-update compares the startup software images of the device with the current software images of the IRF master. If the two sets of images are different, the device automatically performs the following operations:

1. Downloads the current software images of the master.
2. Sets the downloaded images as its main startup software images.
3. Reboots with the new software images to rejoin the IRF fabric.

You must manually update the new device with the software images running on the IRF fabric if software auto-update is disabled.

Restrictions and guidelines

To ensure a successful software auto-update in a multi-user environment, prevent anyone from rebooting member devices during the auto-update process. To inform administrators of the auto-update status, configure the information center to output the status messages to configuration terminals (see *Network Management and Monitoring Configuration Guide*).

Make sure the device you are adding to the IRF fabric has sufficient storage space for the new software images.

If sufficient storage space is not available, the device automatically deletes the current software images. If the reclaimed space is still insufficient, the device cannot complete the auto-update. You must reboot the device, and then access the Boot menu to delete files.

Procedure

1. Enter system view.
   `system-view`
2. Enable software auto-update.
   `irf auto-update enable`
   By default, software auto-update is enabled.

Setting the IRF link down report delay

About this task

To prevent frequent IRF splits and merges during link flapping, configure the IRF ports to delay reporting link down events.

An IRF port does not report a link down event to the IRF fabric immediately after its link changes from up to down. If the IRF link state is still down when the delay is reached, the port reports the change to the IRF fabric.

IRF ports do not delay link up events. They report the link up event immediately after the IRF link comes up.

Restrictions and guidelines

Make sure the IRF link down report delay is shorter than the heartbeat or hello timeout settings of upper-layer protocols (for example, OSPF). If the report delay is longer than the timeout setting of a protocol, unnecessary recalculations might occur.

Set the delay to 0 seconds in the following situations:

- The IRF fabric requires a fast master/subordinate or IRF link switchover.
- The RRPP, BFD, or GR feature is used.
- You want to shut down an IRF physical interface or reboot an IRF member device. (After you complete the operation, reconfigure the delay depending on the network condition.)

Procedure

1. Enter system view.
   `system-view`
2. Set the IRF link down report delay.
   `irf link-delay` *`interval`*
   The default IRF link down report delay is 4 seconds (4000 milliseconds).

Display and maintenance commands for IRF

Execute **`display`** commands in any view.

Task	Command
Display information about all IRF members.	**`display irf`**

Task	Command
Display the IRF fabric topology.	**display irf topology**
Display IRF link information.	**display irf link**
Display IRF configuration.	**display irf configuration**
Display MAD configuration.	**display mad** [**verbose**]

IRF configuration examples

The IRF configuration examples show how to set up IRF fabrics that use different MAD mechanisms.

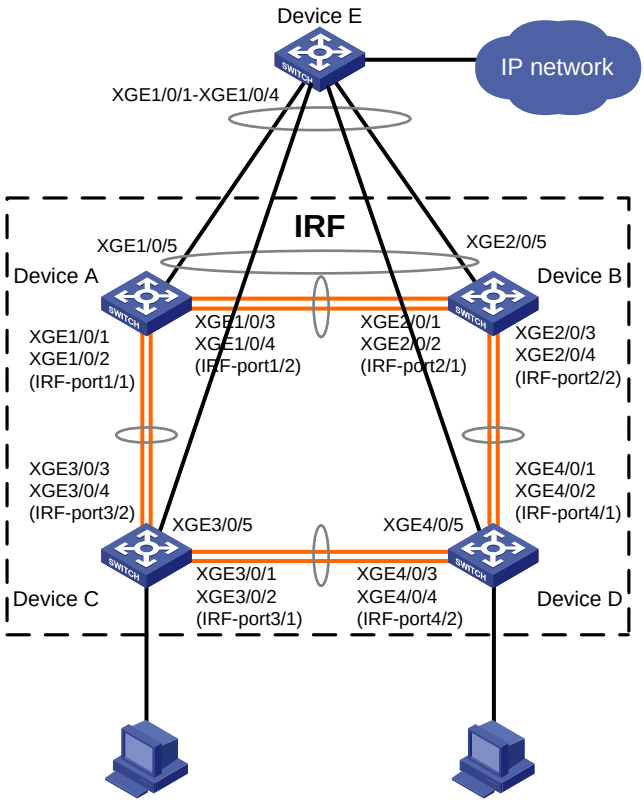
Example: Configuring an IRF fabric with LACP MAD enabled on an aggregate interface

Network configuration

As shown in Figure 14, set up a four-chassis IRF fabric at the access layer of the enterprise network.

Configure LACP MAD on the multichassis aggregation to Device E, which supports extended LACP.

Figure 14 Network diagram



Procedure

1. Configure settings for IRF setup on Device A:

Shut down the physical interfaces used for IRF links. In this example, the physical interfaces are shut down in batch. For more information, see *Layer 2—LAN Switching Configuration Guide*.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 1/0/1 and Ten-GigabitEthernet 1/0/2 to IRF-port 1/1.

```
[Sysname] irf-port 1/1
[Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/1
[Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/2
[Sysname-irf-port1/1] quit
```

Bind Ten-GigabitEthernet 1/0/3 and Ten-GigabitEthernet 1/0/4 to IRF-port 1/2.

```
[Sysname] irf-port 1/2
[Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/3
[Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/4
[Sysname-irf-port1/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

2. Configure settings for IRF setup on Device B:

Change the member ID of Device B to 2 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 2
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device B to Device A as shown in Figure 14, and log in to Device B. (Details not shown.)

Shut down the physical interfaces for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 2/0/1 and Ten-GigabitEthernet 2/0/2 to IRF-port 2/1.

```
[Sysname] irf-port 2/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/2
[Sysname-irf-port2/1] quit
```

Bind Ten-GigabitEthernet 2/0/3 and Ten-GigabitEthernet 2/0/4 to IRF-port 2/2.

```
[Sysname] irf-port 2/2
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/3
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/4
[Sysname-irf-port2/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

The two devices perform master election, and the one that has lost the election reboots to form an IRF fabric with the master.

3. Configure settings for IRF setup on Device C:

Change the member ID of Device C to 3 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 3
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device C to Device A as shown in Figure 14, and log in to Device C. (Details not shown.)

Shut down the physical interfaces for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 3/0/1 and Ten-GigabitEthernet 3/0/2 to IRF-port 3/1.

```
[Sysname] irf-port 3/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/2
[Sysname-irf-port3/1] quit
```

Bind Ten-GigabitEthernet 3/0/3 and Ten-GigabitEthernet 3/0/4 to IRF-port 3/2.

```
[Sysname] irf-port 3/2
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/3
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/4
[Sysname-irf-port3/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device C reboots to join the IRF fabric.

4. Configure settings for IRF setup on Device D:

Change the member ID of Device D to 4 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 4
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
```

```
<Sysname> reboot
```

Connect Device D to Device B and Device C as shown in Figure 14, and log in to Device D. (Details not shown.)

Shut down the physical interfaces.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 4/0/1 and Ten-GigabitEthernet 4/0/2 to IRF-port 4/1.

```
[Sysname] irf-port 4/1
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/1
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/2
[Sysname-irf-port4/1] quit
```

Bind Ten-GigabitEthernet 4/0/3 and Ten-GigabitEthernet 4/0/4 to IRF-port 4/2.

```
[Sysname] irf-port 4/2
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/3
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/4
[Sysname-irf-port4/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device D reboots to join the IRF fabric. A four-chassis IRF fabric is formed.

5. Configure LACP MAD on the IRF fabric:

Set the domain ID of the IRF fabric to 1.

```
<Sysname> system-view
[Sysname] irf domain 1
```

Create a dynamic aggregate interface and enable LACP MAD.

```
[Sysname] interface bridge-aggregation 2
[Sysname-Bridge-Aggregation2] link-aggregation mode dynamic
[Sysname-Bridge-Aggregation2] mad enable
You need to assign a domain ID (range: 0-4294967295)
[Current domain ID is: 1]:
The assigned domain ID is: 1
[Sysname-Bridge-Aggregation2] quit
```

Assign Ten-GigabitEthernet 1/0/5, Ten-GigabitEthernet 2/0/5, Ten-GigabitEthernet 3/0/5, and Ten-GigabitEthernet 4/0/5 to the aggregate interface.

```
[Sysname] interface range ten-gigabitethernet 1/0/5 ten-gigabitethernet 2/0/5
ten-gigabitethernet 3/0/5 ten-gigabitethernet 4/0/5
[Sysname-if-range] port link-aggregation group 2
[Sysname-if-range] quit
```

6. Configure Device E as the intermediate device for LACP MAD:

CAUTION:

If the intermediate device is also an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection. False detection causes IRF split.

Create a dynamic aggregate interface.

```
<Sysname> system-view
[Sysname] interface bridge-aggregation 2
[Sysname-Bridge-Aggregation2] link-aggregation mode dynamic
[Sysname-Bridge-Aggregation2] quit
```

Assign Ten-GigabitEthernet 1/0/1, Ten-GigabitEthernet 1/0/2, Ten-GigabitEthernet 1/0/3, and Ten-GigabitEthernet 1/0/4 to the aggregate interface.

```
[Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
[Sysname-if-range] port link-aggregation group 2
[Sysname-if-range] quit
```

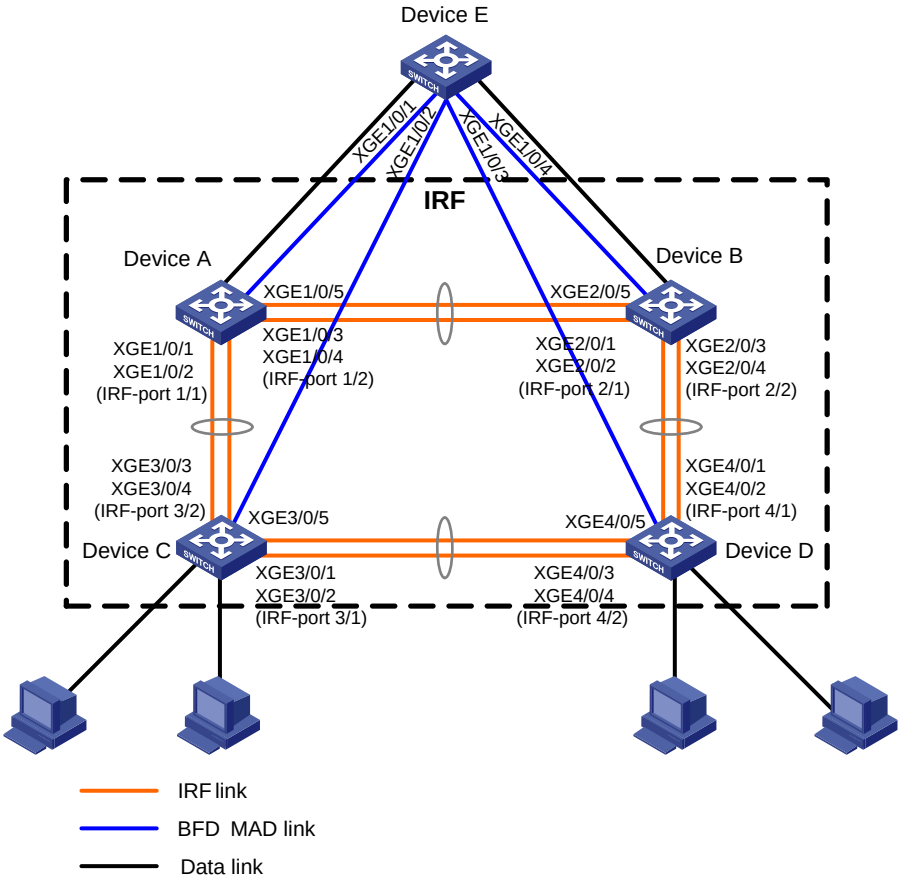
Example: Configuring an IRF fabric with BFD MAD enabled on a VLAN interface

Network configuration

As shown in Figure 15, set up a four-chassis IRF fabric at the distribution layer of the enterprise network.

- Configure BFD MAD on the IRF fabric and set up BFD MAD links between each member device and the intermediate device.
- Disable the spanning tree feature on the ports used for BFD MAD, because the two features conflict with each other.

Figure 15 Network diagram



Procedure

1. Configure settings for IRF setup on Device A:

   # Shut down the physical interfaces used for IRF links.

   ```
   <Sysname> system-view
   [Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
   [Sysname-if-range] shutdown
   [Sysname-if-range] quit
   ```

   # Bind Ten-GigabitEthernet 1/0/1 and Ten-GigabitEthernet 1/0/2 to IRF-port 1/1.

   ```
   [Sysname] irf-port 1/1
   [Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/1
   [Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/2
   [Sysname-irf-port1/1] quit
   ```

   # Bind Ten-GigabitEthernet 1/0/3 and Ten-GigabitEthernet 1/0/4 to IRF-port 1/2.

   ```
   [Sysname] irf-port 1/2
   [Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/3
   [Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/4
   [Sysname-irf-port1/2] quit
   ```

   # Bring up the physical interfaces and save the configuration.

   ```
   [Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
   [Sysname-if-range] undo shutdown
   [Sysname-if-range] quit
   ```

```
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

2. Configure settings for IRF setup on Device B:

Change the member ID of Device B to 2 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 2
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device B to Device A as shown in Figure 15, and log in to Device B. (Details not shown.)

Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 2/0/1 and Ten-GigabitEthernet 2/0/2 to IRF-port 2/1.

```
[Sysname] irf-port 2/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/2
[Sysname-irf-port2/1] quit
```

Bind Ten-GigabitEthernet 2/0/3 and Ten-GigabitEthernet 2/0/4 to IRF-port 2/2.

```
[Sysname] irf-port 2/2
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/3
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/4
[Sysname-irf-port2/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

The two devices perform master election, and the one that has lost the election reboots to form an IRF fabric with the master.

3. Configure settings for IRF setup on Device C:

Change the member ID of Device C to 3 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 3
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device C to Device A as shown in Figure 15, and log in to Device C. (Details not shown.)

Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
```

```
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 3/0/1 and Ten-GigabitEthernet 3/0/2 to IRF-port 3/1.

```
[Sysname] irf-port 3/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/2
[Sysname-irf-port3/1] quit
```

Bind Ten-GigabitEthernet 3/0/3 and Ten-GigabitEthernet 3/0/4 to IRF-port 3/2.

```
[Sysname] irf-port 3/2
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/3
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/4
[Sysname-irf-port3/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device C reboots to join the IRF fabric.

4. Configure settings for IRF setup on Device D:

Change the member ID of Device D to 4 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 4
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device D to Device B and Device C as shown in Figure 15, and log in to Device D. (Details not shown.)

Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 4/0/1 and Ten-GigabitEthernet 4/0/2 to IRF-port 4/1.

```
[Sysname] irf-port 4/1
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/1
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/2
[Sysname-irf-port4/1] quit
```

Bind Ten-GigabitEthernet 4/0/3 and Ten-GigabitEthernet 4/0/4 to IRF-port 4/2.

```
[Sysname] irf-port 4/2
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/3
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/4
[Sysname-irf-port4/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
```

```
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device D reboots to join the IRF fabric. A four-chassis IRF fabric is formed.

5. Configure BFD MAD on the IRF fabric:

   # Create VLAN 3, and add Ten-GigabitEthernet 1/0/5, Ten-GigabitEthernet 2/0/5, Ten-GigabitEthernet 3/0/5, and Ten-GigabitEthernet 4/0/5 to VLAN 3.

```
[Sysname] vlan 3
[Sysname-vlan3] port ten-gigabitethernet 1/0/5 ten-gigabitethernet 2/0/5
ten-gigabitethernet 3/0/5 ten-gigabitethernet 4/0/5
[Sysname-vlan3] quit
```

   # Create VLAN-interface 3, and configure a MAD IP address for each member device on the VLAN interface.

```
[Sysname] interface vlan-interface 3
[Sysname-Vlan-interface3] mad bfd enable
[Sysname-Vlan-interface3] mad ip address 192.168.2.1 24 member 1
[Sysname-Vlan-interface3] mad ip address 192.168.2.2 24 member 2
[Sysname-Vlan-interface3] mad ip address 192.168.2.3 24 member 3
[Sysname-Vlan-interface3] mad ip address 192.168.2.4 24 member 4
[Sysname-Vlan-interface3] quit
```

   # Disable the spanning tree feature on Ten-GigabitEthernet 1/0/5, Ten-GigabitEthernet 2/0/5, Ten-GigabitEthernet 3/0/5, and Ten-GigabitEthernet 4/0/5.

```
[Sysname] interface range ten-gigabitethernet 1/0/5 ten-gigabitethernet 2/0/5
ten-gigabitethernet 3/0/5 ten-gigabitethernet 4/0/5
[Sysname-if-range] undo stp enable
[Sysname-if-range] quit
```

6. Configure Device E as the intermediate device for BFD MAD:

   # Create VLAN 3, and assign Ten-GigabitEthernet 1/0/1, Ten-GigabitEthernet 1/0/2, Ten-GigabitEthernet 1/0/3, and Ten-GigabitEthernet 1/0/4 to VLAN 3 for forwarding BFD MAD packets.

```
<DeviceE> system-view
[DeviceE] vlan 3
[DeviceE-vlan3] port ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
[DeviceE-vlan3] quit
```

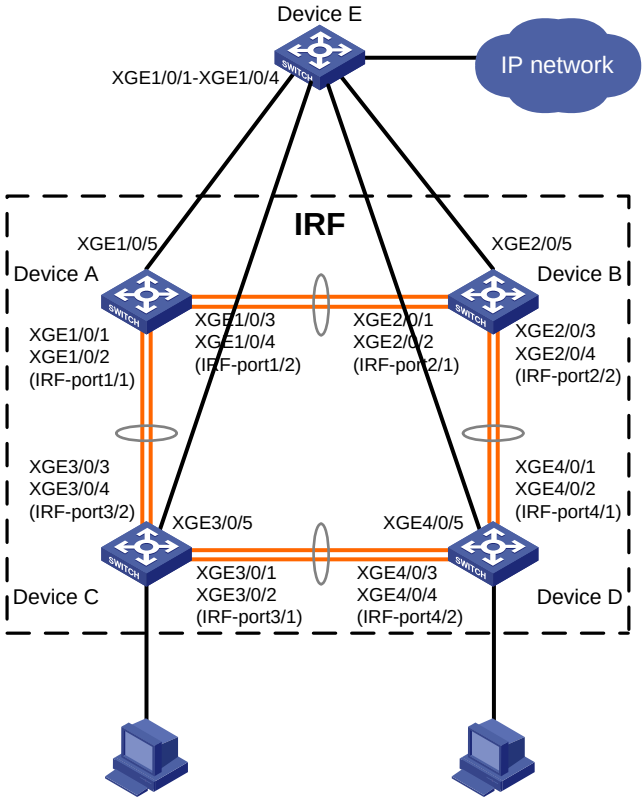
Example: Configuring an IRF fabric with ARP MAD enabled on a VLAN interface

Network configuration

As shown in Figure 16, set up a four-chassis IRF fabric in the enterprise network.

- Configure ARP MAD on the IRF fabric and use the links connected to Device E for transmitting ARP MAD packets.
- To prevent loops, run the spanning tree feature between Device E and the IRF fabric.

Figure 16 Network diagram



Procedure

1. Configure settings for IRF setup on Device A:

   # Shut down the physical interfaces used for IRF links.

   ```
   <Sysname> system-view
   [Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
   [Sysname-if-range] shutdown
   [Sysname-if-range] quit
   ```

   # Bind Ten-GigabitEthernet 1/0/1 and Ten-GigabitEthernet 1/0/2 to IRF-port 1/1.

   ```
   [Sysname] irf-port 1/1
   [Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/1
   [Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/2
   [Sysname-irf-port1/1] quit
   ```

   # Bind Ten-GigabitEthernet 1/0/3 and Ten-GigabitEthernet 1/0/4 to IRF-port 1/2.

   ```
   [Sysname] irf-port 1/2
   [Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/3
   [Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/4
   [Sysname-irf-port1/2] quit
   ```

   # Bring up the physical interfaces and save the configuration.

   ```
   [Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
   [Sysname-if-range] undo shutdown
   [Sysname-if-range] quit
   [Sysname] save
   ```

   # Activate the IRF port configuration.

   ```
   [Sysname] irf-port-configuration active
   ```

2. Configure settings for IRF setup on Device B:

   # Change the member ID of Device B to 2 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 2
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

   # Connect Device B to Device A as shown in Figure 16, and log in to Device B. (Details not shown.)

   # Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

   # Bind Ten-GigabitEthernet 2/0/1 and Ten-GigabitEthernet 2/0/2 to IRF-port 2/1.

```
[Sysname] irf-port 2/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/2
[Sysname-irf-port2/1] quit
```

   # Bind Ten-GigabitEthernet 2/0/3 and Ten-GigabitEthernet 2/0/4 to IRF-port 2/2.

```
[Sysname] irf-port 2/2
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/3
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/4
[Sysname-irf-port2/2] quit
```

   # Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

   # Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

   The two devices perform master election, and the one that has lost the election reboots to form an IRF fabric with the master.

3. Configure settings for IRF setup on Device C:

   # Change the member ID of Device C to 3 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 3
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

   # Connect Device C to Device A as shown in Figure 16, and log in to Device C. (Details not shown.)

   # Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

   # Bind Ten-GigabitEthernet 3/0/1 and Ten-GigabitEthernet 3/0/2 to IRF-port 3/1.

```
[Sysname] irf-port 3/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/2
[Sysname-irf-port3/1] quit
```

Bind Ten-GigabitEthernet 3/0/3 and Ten-GigabitEthernet 3/0/4 to IRF-port 3/2.

```
[Sysname] irf-port 3/2
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/3
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/4
[Sysname-irf-port3/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device C reboots to join the IRF fabric.

4. Configure settings for IRF setup on Device D:

Change the member ID of Device D to 4 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 4
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device D to Device B and Device C as shown in Figure 16, and log in to Device D. (Details not shown.)

Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 4/0/1 and Ten-GigabitEthernet 4/0/2 to IRF-port 4/1.

```
[Sysname] irf-port 4/1
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/1
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/2
[Sysname-irf-port4/1] quit
```

Bind Ten-GigabitEthernet 4/0/3 and Ten-GigabitEthernet 4/0/4 to IRF-port 4/2.

```
[Sysname] irf-port 4/2
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/3
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/4
[Sysname-irf-port4/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device D reboots to join the IRF fabric. A four-chassis IRF fabric is formed.

5. Configure ARP MAD on the IRF fabric:

Enable the spanning tree feature globally. Map the ARP MAD VLAN to MSTI 1 in the MST region.

```
<Sysname> system-view
[Sysname] stp global enable
[Sysname] stp region-configuration
[Sysname-mst-region] region-name arpmad
[Sysname-mst-region] instance 1 vlan 3
[Sysname-mst-region] active region-configuration
[Sysname-mst-region] quit
```

Configure the IRF fabric to change its bridge MAC address as soon as the address owner leaves.

```
[Sysname] undo irf mac-address persistent
```

Set the domain ID of the IRF fabric to 1.

```
[Sysname] irf domain 1
```

Create VLAN 3, and assign Ten-GigabitEthernet 1/0/5, Ten-GigabitEthernet 2/0/5, Ten-GigabitEthernet 3/0/5, and Ten-GigabitEthernet 4/0/5 to VLAN 3.

```
[Sysname] vlan 3
[Sysname-vlan3] port ten-gigabitethernet 1/0/5 ten-gigabitethernet 2/0/5
ten-gigabitethernet 3/0/5 ten-gigabitethernet 4/0/5
[Sysname-vlan3] quit
```

Create VLAN-interface 3, assign it an IP address, and enable ARP MAD on the interface.

```
[Sysname] interface vlan-interface 3
[Sysname-Vlan-interface3] ip address 192.168.2.1 24
[Sysname-Vlan-interface3] mad arp enable
You need to assign a domain ID (range: 0-4294967295)
[Current domain ID is: 1]:
The assigned domain ID is: 1
```

6. Configure Device E as the intermediate device for ARP MAD:

CAUTION:

If the intermediate device is also in an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection. False detection causes IRF split.

Enable the spanning tree feature globally. Map the ARP MAD VLAN to MSTI 1 in the MST region.

```
<DeviceE> system-view
[DeviceE] stp global enable
[DeviceE] stp region-configuration
[DeviceE-mst-region] region-name arpmad
[DeviceE-mst-region] instance 1 vlan 3
[DeviceE-mst-region] active region-configuration
[DeviceE-mst-region] quit
```

Create VLAN 3, and assign Ten-GigabitEthernet 1/0/1, Ten-GigabitEthernet 1/0/2, Ten-GigabitEthernet 1/0/3, and Ten-GigabitEthernet 1/0/4 to VLAN 3 for forwarding ARP MAD packets.

```
[DeviceE] vlan 3
```

```
[DeviceE-vlan3] port ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
[DeviceE-vlan3] quit
```

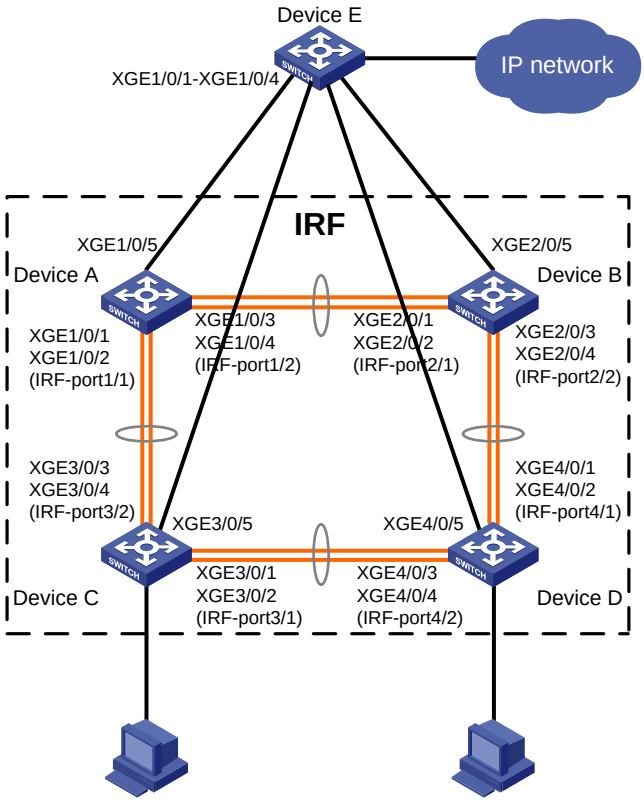
Example: Configuring an IRF fabric with ND MAD enabled on a VLAN interface

Network configuration

As shown in Figure 17, set up a four-chassis IRF fabric in the IPv6 enterprise network.

- Configure ND MAD on the IRF fabric and use the links connected to Device E for transmitting ND MAD packets.
- To prevent loops, run the spanning tree feature between Device E and the IRF fabric.

Figure 17 Network diagram



Procedure

1. Configure settings for IRF setup on Device A:

   # Shut down the physical interfaces used for IRF links.

   ```
   <Sysname> system-view
   [Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
   [Sysname-if-range] shutdown
   [Sysname-if-range] quit
   ```

   # Bind Ten-GigabitEthernet 1/0/1 and Ten-GigabitEthernet 1/0/2 to IRF-port 1/1.

   ```
   [Sysname] irf-port 1/1
   [Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/1
   [Sysname-irf-port1/1] port group interface ten-gigabitethernet 1/0/2
   ```

```
[Sysname-irf-port1/1] quit
```

Bind Ten-GigabitEthernet 1/0/3 and Ten-GigabitEthernet 1/0/4 to IRF-port 1/2.

```
[Sysname] irf-port 1/2
[Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/3
[Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/4
[Sysname-irf-port1/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

2. Configure settings for IRF setup on Device B:

Change the member ID of Device B to 2 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 2
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device B to Device A as shown in Figure 17, and log in to Device B. (Details not shown.)

Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 2/0/1 and Ten-GigabitEthernet 2/0/2 to IRF-port 2/1.

```
[Sysname] irf-port 2/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/1
[Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/2
[Sysname-irf-port2/1] quit
```

Bind Ten-GigabitEthernet 2/0/3 and Ten-GigabitEthernet 2/0/4 to IRF-port 2/2.

```
[Sysname] irf-port 2/2
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/3
[Sysname-irf-port2/2] port group interface ten-gigabitethernet 2/0/4
[Sysname-irf-port2/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

The two devices perform master election, and the one that has lost the election reboots to form an IRF fabric with the master.

3. Configure settings for IRF setup on Device C:

Change the member ID of Device C to 3 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 3
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device C to Device A as shown in Figure 17, and log in to Device C. (Details not shown.)

Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 3/0/1 and Ten-GigabitEthernet 3/0/2 to IRF-port 3/1.

```
[Sysname] irf-port 3/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/1
[Sysname-irf-port3/1] port group interface ten-gigabitethernet 3/0/2
[Sysname-irf-port3/1] quit
```

Bind Ten-GigabitEthernet 3/0/3 and Ten-GigabitEthernet 3/0/4 to IRF-port 3/2.

```
[Sysname] irf-port 3/2
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/3
[Sysname-irf-port3/2] port group interface ten-gigabitethernet 3/0/4
[Sysname-irf-port3/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 3/0/1 to ten-gigabitethernet 3/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device C reboots to join the IRF fabric.

4. Configure settings for IRF setup on Device D:

Change the member ID of Device D to 4 and reboot the device to have the change take effect.

```
<Sysname> system-view
[Sysname] irf member 1 renumber 4
Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
[Sysname] quit
<Sysname> reboot
```

Connect Device D to Device B and Device C as shown in Figure 17, and log in to Device D. (Details not shown.)

Shut down the physical interfaces used for IRF links.

```
<Sysname> system-view
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
[Sysname-if-range] shutdown
[Sysname-if-range] quit
```

Bind Ten-GigabitEthernet 4/0/1 and Ten-GigabitEthernet 4/0/2 to IRF-port 4/1.

```
[Sysname] irf-port 4/1
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/1
```

```
[Sysname-irf-port4/1] port group interface ten-gigabitethernet 4/0/2
[Sysname-irf-port4/1] quit
```

Bind Ten-GigabitEthernet 4/0/3 and Ten-GigabitEthernet 4/0/4 to IRF-port 4/2.

```
[Sysname] irf-port 4/2
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/3
[Sysname-irf-port4/2] port group interface ten-gigabitethernet 4/0/4
[Sysname-irf-port4/2] quit
```

Bring up the physical interfaces and save the configuration.

```
[Sysname] interface range ten-gigabitethernet 4/0/1 to ten-gigabitethernet 4/0/4
[Sysname-if-range] undo shutdown
[Sysname-if-range] quit
[Sysname] save
```

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device D reboots to join the IRF fabric. A four-chassis IRF fabric is formed.

5. Configure ND MAD on the IRF fabric:

Enable the spanning tree feature globally. Map the ND MAD VLAN to MSTI 1 in the MST region.

```
<Sysname> system-view
[Sysname] stp global enable
[Sysname] stp region-configuration
[Sysname-mst-region] region-name ndmad
[Sysname-mst-region] instance 1 vlan 3
[Sysname-mst-region] active region-configuration
[Sysname-mst-region] quit
```

Configure the IRF fabric to change its bridge MAC address as soon as the address owner leaves.

```
[Sysname] undo irf mac-address persistent
```

Set the domain ID of the IRF fabric to 1.

```
[Sysname] irf domain 1
```

Create VLAN 3, and add Ten-GigabitEthernet 1/0/5, Ten-GigabitEthernet 2/0/5, Ten-GigabitEthernet 3/0/5, and Ten-GigabitEthernet 4/0/5 to VLAN 3.

```
[Sysname] vlan 3
[Sysname-vlan3] port ten-gigabitethernet 1/0/5 ten-gigabitethernet 2/0/5
ten-gigabitethernet 3/0/5 ten-gigabitethernet 4/0/5
[Sysname-vlan3] quit
```

Create VLAN-interface 3, assign it an IPv6 address, and enable ND MAD on the interface.

```
[Sysname] interface vlan-interface 3
[Sysname-Vlan-interface3] ipv6 address 2001::1 64
[Sysname-Vlan-interface3] mad nd enable
You need to assign a domain ID (range: 0-4294967295)
[Current domain ID is: 1]:
The assigned domain ID is: 1
```

6. Configure Device E as the intermediate device for ND MAD:

CAUTION:

If the intermediate device is also in an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection. False detection causes IRF split.

Enable the spanning tree feature globally. Map the ND MAD VLAN to MSTI 1 in the MST region.

```
<DeviceE> system-view
[DeviceE] stp global enable
[DeviceE] stp region-configuration
[DeviceE-mst-region] region-name ndmad
[DeviceE-mst-region] instance 1 vlan 3
[DeviceE-mst-region] active region-configuration
[DeviceE-mst-region] quit
```

Create VLAN 3, and add Ten-GigabitEthernet 1/0/1, Ten-GigabitEthernet 1/0/2, Ten-GigabitEthernet 1/0/3, and Ten-GigabitEthernet 1/0/4 to VLAN 3 for forwarding ND MAD packets.

```
[DeviceE] vlan 3
[DeviceE-vlan3] port ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/4
[DeviceE-vlan3] quit
```

Example: Configuring an IRF fabric with ARP MAD and management IP configured on the same management Ethernet port

IMPORTANT:

This example is also applicable to a two-chassis IRF fabric running ND MAD, except that you must configure ND MAD rather than ARP MAD.

Network configuration

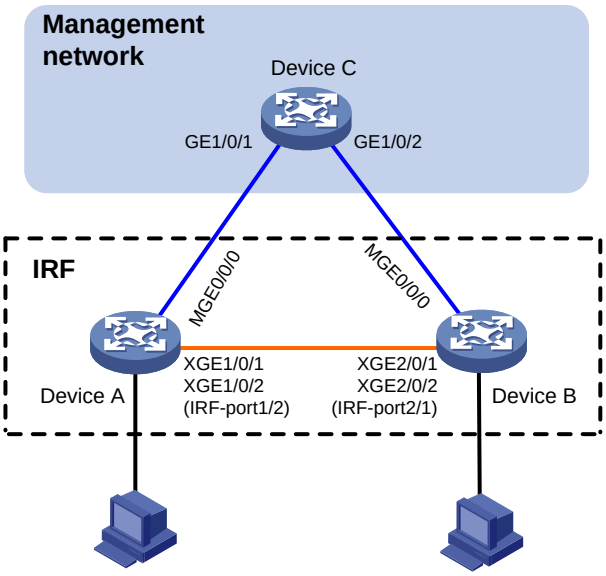
As shown in Figure 18, use Device A and Device B to set up a two-chassis IRF fabric.

Use management Ethernet port M-GigabitEthernet 0/0/0 to perform ARP MAD for IRF split detection and to provide device access for management purposes.

Configure management IP addresses for the IRF fabric and each member device on the management port to meet the following requirements:

- The administrator can use the same management IP address to access the IRF fabric before and after an active/subordinate switchover.
- The administrator can access each member device for failure recovery through their respective management IP addresses after the IRF fabric splits.

Figure 18 Network diagram



Restrictions and guidelines

When you configure management IP addresses, follow these restrictions and guidelines:

- To make sure you can access the IRF fabric at the same IP address after a master/subordinate switchover, use the **ip address** *ip-address* { *mask-length* | *mask* } command to configure a management IP address on the management port, in this example, on M-GigabitEthernet 0/0/0. You can use this address to access the IRF fabric as long as the fabric has not split. This address might be inaccesible for an IP conflict after the IRF fabric splits, because it exists on all member devices. In the context of IRF, this address is called a global management IP address.
- To access each member device for failure recovery after the IRF fabric splits, use the **ip address** *ip-address* { *mask-length* | *mask* } **irf-member** *member-id* command to configure a management IP address for each member device on the management port, in this example, on M-GigabitEthernet 0/0/0. In the context of IRF, this address is called a member-specific management IP address.
- Make sure the management IP addresses assigned to all IRF member devices on the same management Ethernet port belong to the same subnet. Make sure the management IP addresses assigned to an IRF member device on different management Ethernet ports belong to different subnets.
- You must manually make sure the management IP addresses configured on the subordinate devices will not cause IP address conflicts on the network. When the IRF fabric is running correctly, only the IP addresses of the management Ethernet ports on the master device take effect. The management IP addresses of the management Ethernet ports on the subordinate devices do not take effect. If a management IP address on a subordinate device conflicts with an IP address in the network, the system cannot detect the IP conflict, causing network issues after an IRF split.

NOTE:

When the IRF fabric is integrated, you can access the IRF fabric by using either the global management IP address or the management IP address specific to the master device. As a best practice to prevent the management IP address from changing upon a master/subordinate switchover, use the global management IP address as long as the IRF fabric is integrated.

When you configure ARP MAD on a management Ethernet port, follow the restrictions and guidelines in "Configuring ARP MAD."

Setting up the IRF fabric

1. Configure settings for IRF setup on Device A:

   # Shut down the physical interfaces used for IRF connection. This example uses Ten-GigabitEthernet 1/0/1 and Ten-GigabitEthernet 1/0/2 for IRF connection.

   ```
   <Sysname> system-view
   [Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/2
   [Sysname-if-range] shutdown
   [Sysname-if-range] quit
   ```

   # Bind Ten-GigabitEthernet 1/0/1 and Ten-GigabitEthernet 1/0/2 to IRF-port 1/2.

   ```
   [Sysname] irf-port 1/2
   [Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/1
   [Sysname-irf-port1/2] port group interface ten-gigabitethernet 1/0/2
   [Sysname-irf-port1/2] quit
   ```

   # Bring up the physical interfaces.

   ```
   [Sysname] interface range ten-gigabitethernet 1/0/1 to ten-gigabitethernet 1/0/2
   [Sysname-if-range] undo shutdown
   [Sysname-if-range] quit
   ```

   # Save the running configuration to the next-startup configuration file.

   ```
   [Sysname] save
   ```

   # Activate the IRF port configuration.

   ```
   [Sysname] irf-port-configuration active
   ```

2. Configure settings for IRF setup on Device B:

   # Assign member ID 2 to Device B, and reboot the device to have the change take effect.

   ```
   <Sysname> system-view
   [Sysname] irf member 1 renumber 2
   Renumbering the member ID may result in configuration change or loss. Continue? [Y/N]:y
   [Sysname] quit
   <Sysname> reboot
   ```

   # Shut down the physical interfaces used for IRF connection. This example uses Ten-GigabitEthernet 2/0/1 and Ten-GigabitEthernet 2/0/2 for IRF connection.

   ```
   <Sysname> system-view
   [Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/2
   [Sysname-if-range] shutdown
   [Sysname-if-range] quit
   ```

   # Bind Ten-GigabitEthernet 2/0/1 and Ten-GigabitEthernet 2/0/2 to IRF-port 2/1.

   ```
   [Sysname] irf-port 2/1
   [Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/1
   [Sysname-irf-port2/1] port group interface ten-gigabitethernet 2/0/2
   [Sysname-irf-port2/1] quit
   ```

   # Bring up the physical interfaces.

   ```
   [Sysname] interface range ten-gigabitethernet 2/0/1 to ten-gigabitethernet 2/0/2
   [Sysname-if-range] undo shutdown
   [Sysname-if-range] quit
   ```

   # Save the running configuration to the next-startup configuration file.

   ```
   [Sysname] save
   ```

Connect Device B to Device A, as shown in Figure 18.

Activate the IRF port configuration.

```
[Sysname] irf-port-configuration active
```

Device A and Device B perform master election. The device that has failed master election automatically reboots to form an IRF fabric with the other device. A two-chassis IRF fabric is formed.

Configuring management IP addresses

Assign IP address 192.168.1.1/24 to the IRF fabric on management Ethernet port M-GigabitEthernet 0/0/0.

```
<Sysname> system-view
[Sysname] interface m-gigabitethernet 0/0/0
[Sysname-M-GigabitEthernet0/0/0] ip address 192.168.1.1 24
```

On management Ethernet port M-GigabitEthernet 0/0/0, assign IP addresses 192.168.1.101/24 and 192.168.1.102/24 to IRF member device 1 (Device A) and IRF member device 2 (Device B), respectively. After the IRF fabric splits, the administrator can use IP addresses 192.168.1.101/24 and 192.168.1.102/24 to log in to Device A and Device B, respectively.

```
[Sysname-M-GigabitEthernet0/0/0] ip address 192.168.1.101 24 irf-member 1
[Sysname-M-GigabitEthernet0/0/0] ip address 192.168.1.102 24 irf-member 2
[Sysname-M-GigabitEthernet0/0/0] quit
```

Exclude management Ethernet port M-GigabitEthernet 0/0/0 from being shut down by MAD upon detection of multi-active collisions.

```
[Sysname] mad exclude interface m-gigabitethernet 0/0/0
```

Configuring ARP MAD

1. Configure the IRF fabric:

Configure the IRF fabric to change its bridge MAC address as soon as the address owner leaves.

```
[Sysname] undo irf mac-address persistent
```

Set the domain ID of the IRF fabric to 1.

```
[Sysname] irf domain 1
```

Enable ARP MAD on management Ethernet port M-GigabitEthernet 0/0/0.

```
[Sysname] interface m-gigabitethernet 0/0/0
[Sysname-M-GigabitEthernet0/0/0] mad arp enable
You need to assign a domain ID (range: 0-4294967295)
[Current domain ID is: 1]:
The assigned domain ID is: 1
[Sysname-M-GigabitEthernet0/0/0] quit
```

Save the running configuration to the next-startup configuration file.

```
[Sysname] save
```

2. Configure Device C (a switch that supports ARP) as the intermediate device for ARP MAD:

CAUTION:

If the intermediate device is also in an IRF fabric, assign the two IRF fabrics different domain IDs for correct split detection. False detection causes IRF split.

Create VLAN 100.

```
<DeviceC> system-view
[DeviceC] vlan 100
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to VLAN 100 for forwarding ARP MAD packets.

```
[DeviceC-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/2
[DeviceC-vlan100] quit
```

Verifying the configuration

Disconnect the IRF links between Device A and Device B. Verify that you can use 192.168.1.101 and 192.168.1.102 to log in to Device A and Device B, respectively. (Details not shown.)

Execute the **`display mad verbose`** command on Device A and Device B to verify that MAD is correctly functioning, as follows:

- The **Multi-active recovery state** field displays **Yes** on one device and displays **No** on the other device.
- On each device, the IRF physical interfaces and management Ethernet port M-GigabitEthernet 0/0/0 are excluded from being shut down by MAD.

Execute the **`display interface brief`** command on Device A and Device B to verify that the MAD shutdown action is performed correctly, as follows:

- On the device in Recovery state, all network interfaces have been shut down by MAD except the IRF physical interfaces and management Ethernet port M-GigabitEthernet 0/0/0.
- On the device not in Recovery state, no network interfaces are shut down by MAD.

Example: Removing member devices from an IRF fabric without MAD in place

Network configuration

As shown in Figure 19 or Figure 20, an IRF fabric is in daisy-chain or ring topology. The IRF fabric contains four member devices, in which member 1 is the master device. MAD is not configured on the IRF fabric.

Remove member devices 3 and 4 from the IRF fabric.

Figure 19 Network diagram (daisy-chain topology)

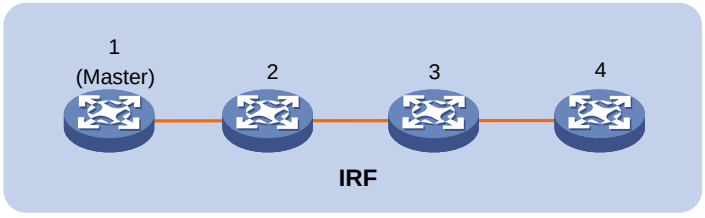
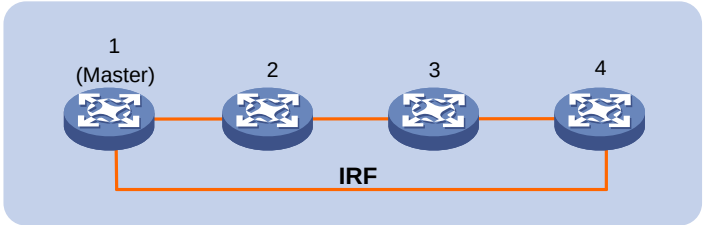


Figure 20 Network diagram (ring topology)



Procedure

1. Enable the IRF fabric to change its bridge MAC address as soon as the address owner leaves the fabric.

```
<System> system-view
[System] undo irf mac-address persistent
```

2. Shut down the network interfaces that connect member devices 3 and 4 to the upstream and downstream devices. (Details not shown.)
3. Use one of the following methods to split the IRF fabric into two IRF fabrics (IRF 1 and IRF 2):
   - Disconnect the physical IRF links between IRF 1 and IRF 2, as shown in Figure 21 or Figure 22.

     Figure 21 Disconnecting physical IRF links (daisy-chain topology)

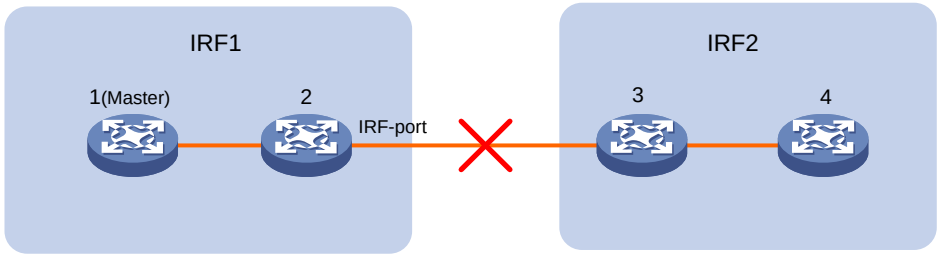
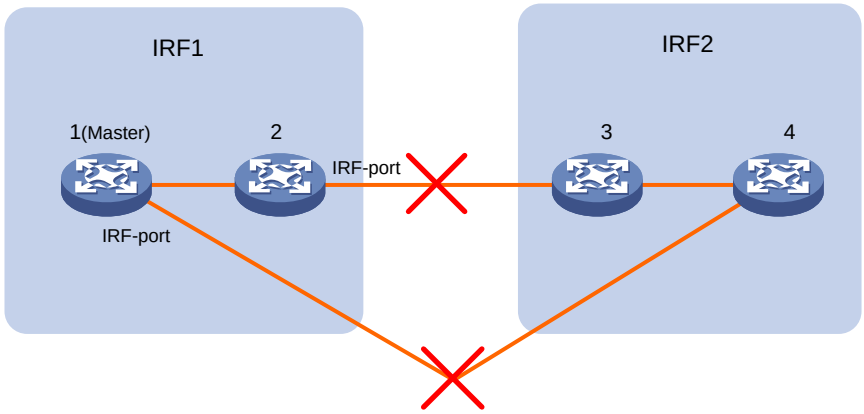
     

     Figure 22 Disconnecting physical IRF links (ring topology)

   - Shut down the IRF physical interfaces that connect IRF 1 and IRF 2, as shown in Figure 21 or Figure 22. (Details not shown.)

   The IRF fabric splits into IRF 1 and IRF 2. Because MAD is not configured, both IRF 1 and IRF 2 are active. The member devices in IRF 2 automatically elect a master on split.
4. Handle IRF 2:
   - To retain IRF 2 in the network, log in to IRF 2, modify the IP addresses that conflict with those on IRF 1 and change its domain ID. (Details not shown.)
   - To remove IRF 2 from the network, disconnect all physical network connections from the IRF fabric and remove IRF configuration from all member devices in the IRF fabric. (Details not shown.)
5. Bring up the network interfaces that have been shut down on the removed member devices. (Details not shown.)
6. Remove IRF port bindings from unused IRF physical interfaces on IRF 1. (Details not shown.)

Example: Removing member devices from an IRF fabric with MAD in place

Network configuration

As shown in Figure 23 or Figure 24, an IRF fabric is in daisy-chain or ring topology. The IRF fabric contains four member devices, in which member 1 is the master device. MAD is configured on the IRF fabric.

Remove member devices 3 and 4 from the IRF fabric.

Figure 23 Network diagram (daisy-chain topology)

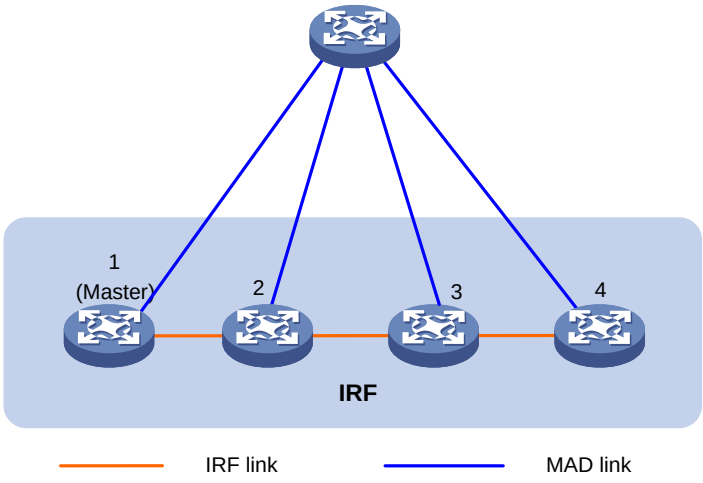
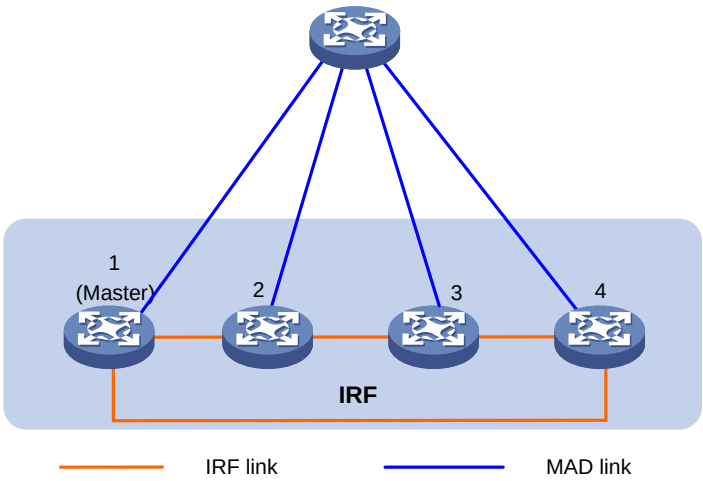


Figure 24 Network diagram (ring topology)



Procedure

1. Enable the IRF fabric to change its bridge MAC address as soon as the address owner leaves the fabric.

   ```
   <System> system-view
   [System] undo irf mac-address persistent
   ```

2. Shut down the network interfaces that connect member devices 3 and 4 to the upstream and downstream devices. (Details not shown.)
3. Use one of the following methods to split the IRF fabric into two IRF fabrics (IRF 1 and IRF 2):

- Disconnect the physical IRF links between IRF 1 and IRF 2, as shown in Figure 25 or Figure 26.

Figure 25 Disconnecting physical IRF links (daisy-chain topology)

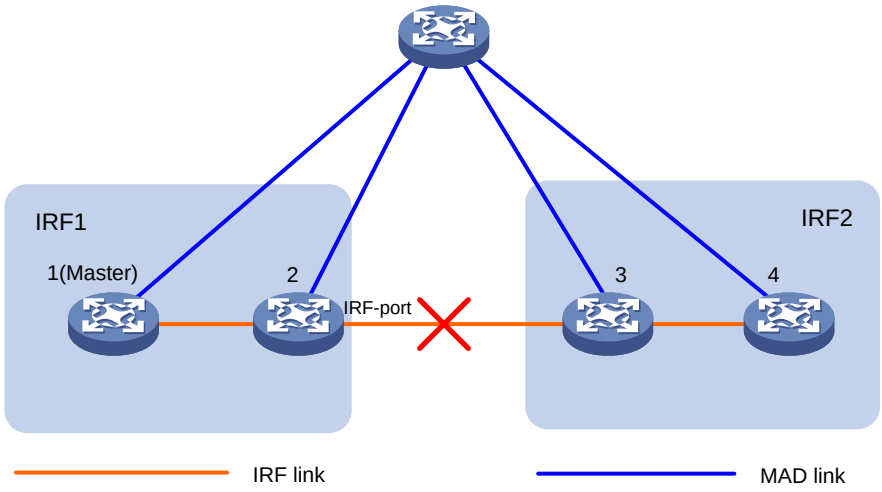
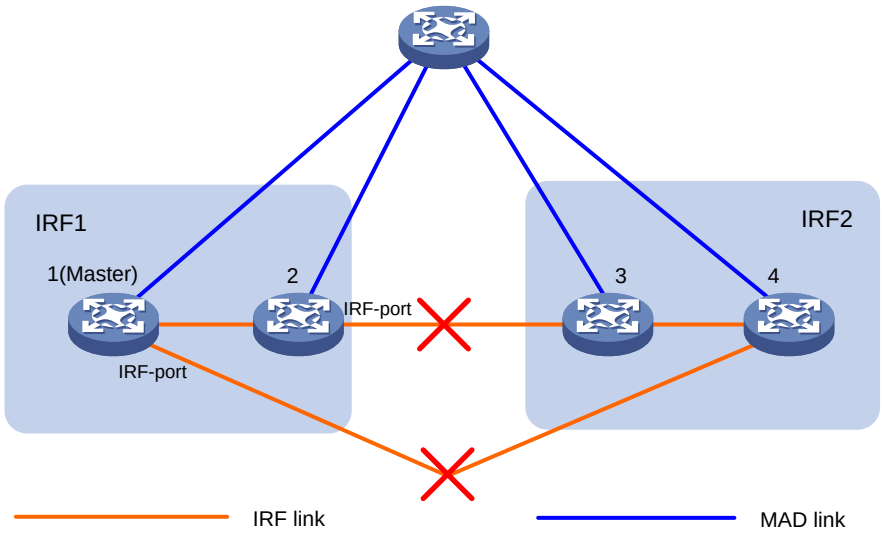


Figure 26 Disconnecting physical IRF links (ring topology)



- Shut down the IRF physical interfaces that connect IRF 1 and IRF 2, as shown in Figure 25 or Figure 26. (Details not shown.)

The IRF fabric splits into IRF 1 and IRF 2. Because MAD is configured, IRF 1 remains active and IRF 2 is placed in Recovery state and becomes inactive.

4. Handle IRF 2:
   - To retain IRF 2 in the network, log in to IRF 2, modify the IP addresses that conflict with those on IRF 1, change its domain ID, and recover IRF 2.

     ```
     <System> system-view
     [System] irf domain 2
     [System] mad restore
     ```

   - To remove IRF 2 from the network, disconnect all physical network connections from the IRF fabric and remove IRF configuration from all member devices in the IRF fabric. (Details not shown.)

5. Bring up the network interfaces that have been shut down on the removed member devices. (Details not shown.)
6. Remove IRF port bindings from unused IRF physical interfaces on IRF 1. (Details not shown.)

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Layer 2—LAN Switching Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes LAN switching fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring the MAC address table 1
About the MAC address table 1
How a MAC address entry is created 1
Types of MAC address entries 1
MAC address table tasks at a glance 2
Configuring MAC address entries 2
About MAC address entry-based frame forwarding 2
Restrictions and guidelines for MAC address entry configuration 3
Prerequisites for MAC address entry configuration 3
Adding or modifying a static or dynamic MAC address entry 3
Adding or modifying a blackhole MAC address entry 4
Adding or modifying a multiport unicast MAC address entry 4
Setting the aging timer for dynamic MAC address entries 5
Disabling MAC address learning 6
About disabling MAC address learning 6
Disabling MAC address learning on an interface 6
Disabling MAC address learning on a VLAN 6
Setting the MAC learning limit 7
Setting the MAC learning limit on interfaces 7
Setting the MAC learning limit for a VLAN 8
Configuring the unknown frame forwarding rule after the MAC learning limit is reached 8
About unknown frame forwarding rule configuration 8
Configuring the device to forward unknown frames after the MAC learning limit on an interface is reached 8
Configuring the device to forward unknown frames after the MAC learning limit for a VLAN is reached 9
Configuring MAC address move notifications and suppression 9
Disabling static source check 10
About this task 10
Disabling static source check on an interface 10
Enabling SNMP notifications for the MAC address table 11
Display and maintenance commands for MAC address table 11
MAC address table configuration examples 12
Example: Configuring the MAC address table 12
Configuring MAC Information 14
About MAC Information 14
Enabling MAC Information 14
Configuring the MAC Information mode 14
Setting the MAC change notification interval 15
Setting the MAC Information queue length 15
MAC Information configuration examples 16
Example: Configuring MAC Information 16

Configuring the MAC address table

About the MAC address table

An Ethernet device uses a MAC address table to forward frames. A MAC address entry includes a destination MAC address, an outgoing interface, and a VLAN ID. When the device receives a frame, it uses the destination MAC address of the frame to look for a match in the MAC address table.

- The device forwards the frame out of the outgoing interface in the matching entry if a match is found.
- The device floods the frame in the VLAN of the frame if no match is found.

How a MAC address entry is created

The entries in the MAC address table include entries automatically learned by the device and entries manually added.

MAC address learning

The device can automatically populate its MAC address table by learning the source MAC addresses of incoming frames on each interface.

The device performs the following operations to learn the source MAC address of incoming packets:

1. Checks the source MAC address (for example, MAC-SOURCE) of the frame.
2. Looks up the source MAC address in the MAC address table.
    - The device updates the entry if an entry is found.
    - The device adds an entry for MAC-SOURCE and the incoming port if no entry is found.

When the device receives a frame destined for MAC-SOURCE after learning this source MAC address, the device performs the following operations:

3. Finds the MAC-SOURCE entry in the MAC address table.
4. Forwards the frame out of the port in the entry.

The device performs the learning process for each incoming frame with an unknown source MAC address until the table is fully populated.

Manually configuring MAC address entries

Dynamic MAC address learning does not distinguish between illegitimate and legitimate frames, which can invite security hazards. When Host A is connected to Port A, a MAC address entry will be learned for the MAC address of Host A (for example, MAC A). When an illegal user sends frames with MAC A as the source MAC address to Port B, the device performs the following operations:

1. Learns a new MAC address entry with Port B as the outgoing interface and overwrites the old entry for MAC A.
2. Forwards frames destined for MAC A out of Port B to the illegal user.

As a result, the illegal user obtains the data of Host A. To improve the security for Host A, manually configure a static entry to bind Host A to Port A. Then, the frames destined for Host A are always sent out of Port A. Other hosts using the forged MAC address of Host A cannot obtain the frames destined for Host A.

Types of MAC address entries

A MAC address table can contain the following types of entries:

- **Static entries**—A static entry is manually added to forward frames with a specific destination MAC address out of the associated interface, and it never ages out. A static entry has higher priority than a dynamically learned one.
- **Dynamic entries**—A dynamic entry can be manually configured or dynamically learned to forward frames with a specific destination MAC address out of the associated interface. A dynamic entry might age out. A manually configured dynamic entry has the same priority as a dynamically learned one.
- **Blackhole entries**—A blackhole entry is manually configured and never ages out. A blackhole entry is configured for filtering out frames with a specific source or destination MAC address. For example, to block all frames destined for or sourced from a user, you can configure the MAC address of the user as a blackhole MAC address entry. A blackhole entry has higher priority than a dynamically learned one.
- **Multiport unicast entries**—A multiport unicast entry is manually added to send frames with a specific unicast destination MAC address out of multiple ports, and it never ages out. A multiport unicast entry has higher priority than a dynamically learned one.

A static or blackhole MAC address entry can overwrite a dynamic MAC address entry. A dynamic MAC address entry cannot overwrite a static, dynamic, blackhole, or multiport unicast MAC address entry. A static entry, a blackhole entry, and a multiport unicast entry cannot overwrite one another.

A multiport unicast MAC address entry does not affect learning the corresponding dynamic MAC address entry. For the same MAC address, a multiport unicast MAC address entry and a dynamic MAC address entry can coexist, and the multiport unicast MAC address takes priority.

This document does not cover the configuration of static multicast MAC address entries. For more information about configuring static multicast MAC address entries, see IGMP snooping in *IP Multicast Configuration Guide*.

MAC address table tasks at a glance

All MAC address table configuration tasks are optional.

To configure the MAC address table, perform the following tasks:

- Configuring MAC address entries
  - Adding or modifying a static or dynamic MAC address entry
  - Adding or modifying a blackhole MAC address entry
  - Adding or modifying a multiport unicast MAC address entry
- Setting the aging timer for dynamic MAC address entries
- Configuring MAC address learning
  - Disabling MAC address learning
  - Setting the MAC learning limit
  - Configuring the unknown frame forwarding rule after the MAC learning limit is reached
- Configuring MAC address move notifications and suppression
- Disabling static source check
- Enabling SNMP notifications for the MAC address table

Configuring MAC address entries

About MAC address entry-based frame forwarding

A frame whose source MAC address matches different types of MAC address entries is processed differently.

Type	Description
Static MAC address entry	By default, if the interface in the entry is different from the frame's ingress interface, the device drops the packet. If the interface in the entry is the same as the frame's ingress interface, the device forwards the frame according to the destination MAC address. With the static source check disabled, the device forwards the frame according to the destination MAC address regardless of whether the frame's ingress interface is the same as that in the entry.
Multiport unicast MAC address entry	Learns the source MAC address of the frame, generates a dynamic MAC address entry for that MAC address, and forwards the frame. The multiport unicast MAC address entry has higher priority than the dynamic MAC address entry in traffic forwarding.
Blackhole MAC address entry	Drops the frame.
Dynamic MAC address entry	• Learns the MAC address of the frames received on a different interface from that in the entry and overwrites the original entry. • Forwards the frame received on the same interface as that in the entry and updates the aging timer for the entry.

Restrictions and guidelines for MAC address entry configuration

The manually configured static, blackhole, and multiport unicast MAC address entries cannot survive a reboot if you do not save the configuration. The manually configured dynamic MAC address entries are lost upon reboot whether or not you save the configuration.

Do not configure reserved MAC addresses as static, dynamic, blackhole, or multiport unicast MAC addresses. The reserved MAC addresses are the bridge MAC address and its subsequent higher 37 MAC addresses. For information about bridge MAC addresses, see IRF in *Virtual Technologies Configuration Guide*.

You cannot add or modify dynamic MAC address entries for interfaces in the down state.

Prerequisites for MAC address entry configuration

Before manually configuring a MAC address entry for an interface, make sure the VLAN in the entry has been created.

Adding or modifying a static or dynamic MAC address entry

Adding or modifying a static or dynamic MAC address entry globally

1. Enter system view.

   system-view

2. Add or modify a static or dynamic MAC address entry.

   mac-address { **dynamic** | **static** } *mac-address* **interface** *interface-type interface-number* **vlan** *vlan-id*

   By default, no MAC address entry is configured globally.

   Make sure you have assigned the interface to the VLAN.

Adding or modifying a static or dynamic MAC address entry on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Add or modify a static or dynamic MAC address entry.
   mac-address { **dynamic** | **static** } *mac-address* **vlan** *vlan-id*

   By default, no MAC address entry is configured on an interface.

   Make sure you have assigned the interface to the VLAN.

Adding or modifying a blackhole MAC address entry

1. Enter system view.
   system-view
2. Add or modify a blackhole MAC address entry.
   mac-address blackhole *mac-address* **vlan** *vlan-id*

   By default, no blackhole MAC address entry is configured.

Adding or modifying a multiport unicast MAC address entry

About this task

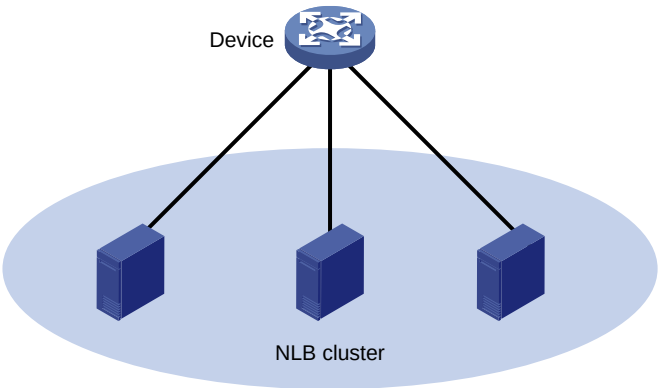
You can configure a multiport unicast MAC address entry to associate a unicast destination MAC address with multiple ports. The frame with a destination MAC address matching the entry is sent out of multiple ports.

For example, in NLB unicast mode (see Figure 1):

- All servers within a cluster uses the cluster's MAC address as their own address.
- Frames destined for the cluster are forwarded to every server in the group.

In this case, you can configure a multiport unicast MAC address entry on the device connected to the server group. Then, the device forwards the frame destined for the server group to every server through all ports connected to the servers within the cluster.

Figure 1 NLB cluster



You can configure a multiport unicast MAC address entry globally or on an interface.

Configuring a multiport unicast MAC address entry globally

1. Enter system view.
   system-view
2. Add or modify a multiport unicast MAC address entry.
   mac-address multiport *mac-address* **interface** *interface-list* **vlan** *vlan-id*
   By default, no multiport unicast MAC address entry is configured globally.
   Make sure you have assigned the interface to the VLAN.

Configuring a multiport unicast MAC address entry on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Add the interface to a multiport unicast MAC address entry.
   mac-address multiport *mac-address* **vlan** *vlan-id*
   By default, no multiport unicast MAC address entry is configured on an interface.
   Make sure you have assigned the interface to the VLAN.

Setting the aging timer for dynamic MAC address entries

About this task

For security and efficient use of table space, the MAC address table uses an aging timer for each dynamic MAC address entry. If a dynamic MAC address entry is not updated before the aging timer expires, the device deletes the entry. This aging mechanism ensures that the MAC address table can promptly update to accommodate latest network topology changes.

A stable network requires a longer aging interval, and an unstable network requires a shorter aging interval.

An aging interval that is too long might cause the MAC address table to retain outdated entries. As a result, the MAC address table resources might be exhausted, and the MAC address table might fail to update its entries to accommodate the latest network changes.

An interval that is too short might result in removal of valid entries, which would cause unnecessary floods and possibly affect the device performance.

To reduce floods on a stable network, set a long aging timer or disable the timer to prevent dynamic entries from unnecessarily aging out. Reducing floods improves the network performance. Reducing flooding also improves the security because it reduces the chances for a data frame to reach unintended destinations.

Procedure

1. Enter system view.
   system-view

2. Set the aging timer for dynamic MAC address entries.

   mac-address timer { **aging** *seconds* | **no-aging** }

   By default, the aging timer is 300 seconds for dynamic MAC address entries.

Disabling MAC address learning

About disabling MAC address learning

MAC address learning is enabled by default. To prevent the MAC address table from being saturated when the device is experiencing attacks, disable MAC address learning. For example, you can disable MAC address learning to prevent the device from being attacked by a large amount of frames with different source MAC addresses.

After MAC address learning is disabled, the device immediately deletes existing dynamic MAC address entries.

After MAC address learning is disabled, the device immediately deletes existing dynamic MAC address entries.

Disabling MAC address learning on an interface

About this task

When global MAC address learning is enabled, you can disable MAC address learning on a single interface.

Restrictions and guidelines

The **mac-address mac-learning enable** command and its **undo** form do not take effect on a Layer 2 aggregate interface acting as a peer-link interface, because MAC address learning is always disabled on that interface. For more information about peer-link interfaces, see "Configuring M-LAG."

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*

3. Disable MAC address learning on the interface.

   undo mac-address mac-learning enable

   By default, MAC address learning is enabled on an interface.

Disabling MAC address learning on a VLAN

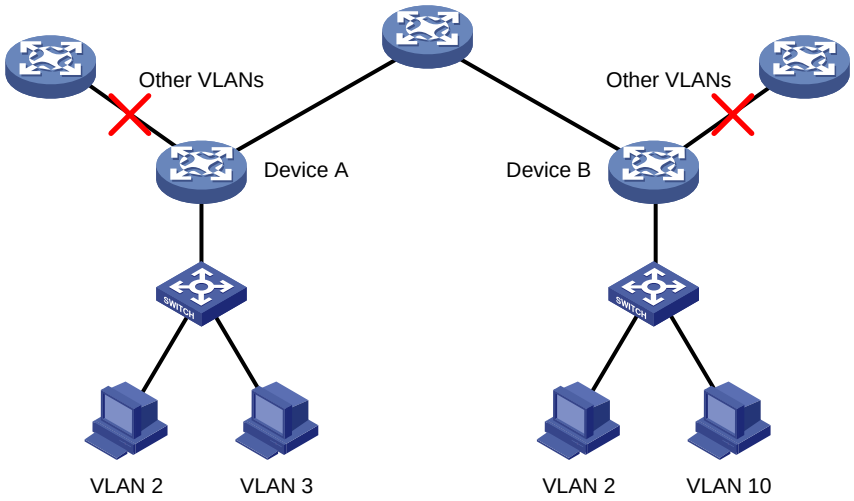
About this task

When global MAC address learning is enabled, you can disable MAC address learning on a per-VLAN basis.

As shown in Figure 2, VLAN 2, VLAN 3, and VLAN 10 are configured on the network to isolate users from different departments. To save MAC address entry resources on the core device and deny

external access to these VLANs, you can disable MAC address learning for the VLANs other than VLAN 2, VLAN 3, and VLAN 10 on Device A and Device B.

Figure 2 Network diagram



Procedure

1. Enter system view.
   `system-view`
2. Enter VLAN view.
   `vlan` *`vlan-id`*
3. Disable MAC address learning on the VLAN.
   `undo mac-address mac-learning enable`
   By default, MAC address learning on the VLAN is enabled.

Setting the MAC learning limit

Setting the MAC learning limit on interfaces

About this task

This feature limits the MAC address table size. A large MAC address table will degrade forwarding performance.

Restrictions and guidelines

The MAC learning limit does not control the number of MAC addresses learned in voice VLANs. For more information, see "Configuring voice VLANs."

Procedure

1. Enter system view.
   `system-view`
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     `interface` *`interface-type interface-number`*
   - Enter Layer 2 aggregate interface view.
     `interface bridge-aggregation` *`interface-number`*

3. Set the MAC learning limit on the interface.

   mac-address max-mac-count *count*

   The MAC learning limit on an interface is restricted by the **switch-mode** command. For more information about the **switch-mode** command, see *Fundamentals Command Reference*.

Setting the MAC learning limit for a VLAN

About this task

You can limit the number of MAC addresses that can be learned for a VLAN. A large MAC address table will degrade forwarding performance.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Set the MAC leaning limit for the VLAN.

   mac-address max-mac-count *count*

   The MAC learning limit for a VLAN is restricted by the **switch-mode** command. For more information about the **switch-mode** command, see *Fundamentals Command Reference.*

Configuring the unknown frame forwarding rule after the MAC learning limit is reached

In this document, unknown frames refer to frames whose source MAC addresses are not in the MAC address table.

About unknown frame forwarding rule configuration

You can enable or disable forwarding of unknown frames after the MAC learning limit is reached.

Configuring the device to forward unknown frames after the MAC learning limit on an interface is reached

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*

3. Configure the device to forward unknown frames received on the interface after the MAC learning limit on the interface is reached.

   mac-address max-mac-count enable-forwarding

   By default, the device can forward unknown frames received on an interface after the MAC learning limit on the interface is reached.

Configuring the device to forward unknown frames after the MAC learning limit for a VLAN is reached

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Configure the device to forward unknown frames received on interfaces in the VLAN after the MAC learning limit for the VLAN is reached.
   mac-address max-mac-count enable-forwarding
   By default, the device can forward unknown frames received on interfaces in a VLAN after the MAC learning limit for the VLAN is reached.

Configuring MAC address move notifications and suppression

About this task

The outgoing interface for a MAC address entry learned on interface A is changed to interface B when the following conditions exist:

- Interface B receives a packet with the MAC address as the source MAC address.
- Interface B belongs to the same VLAN as interface A.

In this case, the MAC address is moved from interface A to interface B, and a MAC address move occurs.

The MAC address move notifications feature enables the device to output MAC address move logs when MAC address moves are detected.

If a MAC address is continuously moved between the two interfaces, Layer 2 loops might occur. To detect and locate loops, you can view the MAC address move information. To display the MAC address move records after the device is started, use the **display mac-address mac-move** command.

If the system detects that MAC address moves occur frequently on an interface, you can configure MAC address move suppression to shut the interface down. The interface automatically goes up after a suppression interval. Or, you can manually bring up the interface.

Restrictions and guidelines

After you configure MAC address move notifications, the system sends only log messages to the information center module. If the device is also configured with the **snmp-agent trap enable mac-address** command, the system also sends SNMP notifications to the SNMP module.

Procedure

1. Enter system view.
   system-view
2. Enter MAC address move notifications and optionally specify a MAC move detection interval.
   mac-address notification mac-move [**interval** *interval*]
   By default, MAC address move notifications are disabled.
3. (Optional.) Set MAC address move suppression parameters.

mac-address notification mac-move suppression { **interval** *interval* | **threshold** *threshold* }

By default, the suppression interval is 30 seconds, and the suppression threshold is 3.

For the MAC address move suppression parameters to take effect, enable the MAC address move suppression on a port.

4. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
5. Enable MAC address move suppression.
   mac-address notification mac-move suppression
   By default, MAC address move suppression is disabled.

Disabling static source check

About this task

By default, the static source check feature is enabled on an interface. The check identifies whether a received frame meets the following conditions:

- The source MAC address of the frame matches a static MAC address entry.
- The incoming interface of the frame is different from the outgoing interface in the entry.

If the frame meets both conditions, the device drops the frame.

When this feature is disabled, the device does not perform the check for a received frame. It can forward the frame whether or not the frame meets the conditions.

Disabling static source check on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation { *interface-number* | *interface-number.subnumber* }
   - Enter IRF physical interface view.
     interface *interface-type interface-number*
3. Disable the static source check feature on the interface.
   undo mac-address static source-check enable
   By default, the static source check feature is enabled on an interface.

Enabling SNMP notifications for the MAC address table

About this task

To report critical MAC address move events to an NMS, enable SNMP notifications for the MAC address table. For MAC address move event notifications to be sent correctly, you must also configure SNMP on the device.

When SNMP notifications are disabled for the MAC address table, the device sends the generated logs to the information center. To display the logs, configure the log destination and output rule configuration in the information center.

For more information about SNMP and information center configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for the MAC address table.

   snmp-agent trap enable mac-address [**mac-move**]

   By default, SNMP notifications are enabled for the MAC address table.

   If you execute this command without specifying any keywords, the command enables all types of SNMP notifications for the MAC address table. If you specify a keyword, the command enables only the specified type of SNMP notifications for the MAC address table.

Display and maintenance commands for MAC address table

Execute **display** commands in any view.

Task	Command
Display MAC address table information.	**display mac-address** [*mac-address* [**vlan** *vlan-id*] | [[**dynamic** | **static**] [**interface** *interface-type interface-number*] | **blackhole** | **multiport**] [**vlan** *vlan-id*] [**count**]]
Display the aging timer for dynamic MAC address entries.	**display mac-address aging-time**
Display the system or interface MAC address learning state.	**display mac-address mac-learning** [**interface** *interface-type interface-number*]
Display the MAC address move records.	**display mac-address mac-move** [**slot** *slot-number*]

Task	Command
Display MAC address statistics.	**display mac-address statistics**
Clear the MAC address move records for all VLANs after the device is started.	**reset mac-address mac-move**

MAC address table configuration examples

Example: Configuring the MAC address table

Network configuration

As shown in Figure 3:

- Host A at MAC address 000f-e235-dc71 is connected to GigabitEthernet 1/0/1 of Device and belongs to VLAN 10.
- Host B at MAC address 000f-e235-abcd, which behaved suspiciously on the network, also belongs to VLAN 10.

Configure the MAC address table as follows:

- To prevent MAC address spoofing, add a static entry for Host A in the MAC address table of Device.
- To drop all frames destined for Host B, add a blackhole MAC address entry for Host B.
- Set the aging timer to 500 seconds for dynamic MAC address entries.

Figure 3 Network diagram

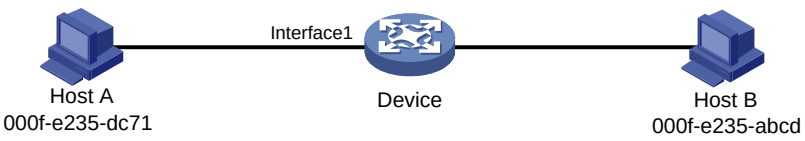


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Create VLAN-interface 10, and assign an IP address and subnet mask to the interface.

```
<Device> system-view
[Device] vlan 10
[Device-vlan10] quit
[Device] interface Vlan-interface 10
[Device-Vlan-interface10] ip address 192.168.1.201 24
[Device-Vlan-interface10] quit
```

Configure GigabitEthernet1/0/1 as an access port and add it to VLAN 10.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] port link-type access
[Device-GigabitEthernet1/0/1] port access vlan 10
[Device-GigabitEthernet1/0/1] quit
```

Add a static MAC address entry for MAC address 000f-e235-dc71 on GigabitEthernet 1/0/1 that belongs to VLAN 10.

```
[Device] mac-address static 000f-e235-dc71 interface gigabitethernet 1/0/1 vlan 10
```

Add a blackhole MAC address entry for MAC address 000f-e235-abcd that belongs to VLAN 10.

```
[Device] mac-address blackhole 000f-e235-abcd vlan 10
```

Set the aging timer to 500 seconds for dynamic MAC address entries.

```
[Device] mac-address timer aging 500
```

Verifying the configuration

Display the static MAC address entries for GigabitEthernet 1/0/1.

```
[Device] display mac-address static interface gigabitethernet 1/0/1
MAC Address      VLAN ID    State            Port/Nickname            Aging
000f-e235-dc71   10         Static           GE1/0/1                  N
```

Display the blackhole MAC address entries.

```
[Device] display mac-address blackhole
MAC Address      VLAN ID    State            Port/Nickname            Aging
000f-e235-abcd   10         Blackhole        N/A                      N
```

Display the aging time of dynamic MAC address entries.

```
[Device] display mac-address aging-time
MAC address aging time: 500s.
```

Configuration files

```
#
interface Vlan-interface10
 ip address 192.168.1.201 255.255.255.0
#
interface GigabitEthernet1/0/1
 port access vlan 10
 mac-address static 000f-e235-dc71 vlan 10
#
 mac-address blackhole 000f-e235-abcd vlan 10
#
 mac-address timer aging 500
#
```

Configuring MAC Information

About MAC Information

The MAC Information feature can generate syslog messages or SNMP notifications when MAC address entries are learned or deleted. You can use these messages to monitor user's leaving or joining the network and analyze network traffic.

The MAC Information feature buffers the MAC change syslog messages or SNMP notifications in a queue. The device overwrites the oldest MAC address change written into the queue with the most recent MAC address change when the following conditions exist:

- The MAC change notification interval does not expire.
- The queue has been exhausted.

To send a syslog message or SNMP notification immediately after it is created, set the queue length to zero.

Enabling MAC Information

Restrictions and guidelines

For MAC Information to take effect, you must enable MAC Information both globally and on interfaces.

Procedure

1. Enter system view.
   system-view
2. Enable MAC Information globally.
   mac-address information enable
   By default, MAC Information is globally disabled.
3. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
4. Enable MAC Information on the interface.
   mac-address information enable { **added** | **deleted** }
   By default, MAC Information is disabled on an interface.

Configuring the MAC Information mode

About this task

The following MAC Information modes are available for sending MAC address changes:

- **Syslog**—The device sends syslog messages to notify MAC address changes. The device sends syslog messages to the information center, which then outputs them to the monitoring terminal. For more information about information center, see *Network Management and Monitoring Configuration Guide*.
- **Trap**—The device sends SNMP notifications to notify MAC address changes. The device sends SNMP notifications to the NMS. For more information about SNMP, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Configure the MAC Information mode.
   mac-address information mode { **syslog** | **trap** }
   The default setting is **trap**.

Setting the MAC change notification interval

About this task

To prevent syslog messages or SNMP notifications from being sent too frequently, you can set the MAC change notification interval to a larger value.

Procedure

1. Enter system view.
   system-view
2. Set the MAC change notification interval.
   mac-address information interval *interval*
   The default setting is 1 second.

Setting the MAC Information queue length

About this task

If the MAC Information queue length is 0, the device sends a syslog message or SNMP notification immediately after learning or deleting a MAC address.

If the MAC Information queue length is not 0, the device stores MAC changes in the queue:

- The device overwrites the oldest MAC change written into the queue with the most recent MAC change when the following conditions exist:
  - The MAC change notification interval does not expire.
  - The queue has been exhausted.
- The device sends syslog messages or SNMP notifications only if the MAC change notification interval expires.

Procedure

1. Enter system view.
   system-view
2. Set the MAC Information queue length.
   mac-address information queue-length *value*
   The default setting is 50.

MAC Information configuration examples

Example: Configuring MAC Information

Network configuration

Enable MAC Information on Device A in Figure 4 to generate logs about MAC address changes and display them on the screen of a Telnet client PC.

Figure 4 Network diagram



Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Configure Device A

Create VLAN 10, and assign an IP address and subnet mask to the interface.

```
<DeviceA> system-view
[DeviceA] vlan 10
[DeviceA-vlan10] quit
[DeviceA] interface vlan-interface 10
[DeviceA-Vlan-interface10] ip address 192.168.1.193 24
[DeviceA-Vlan-interface10] quit
```

ConfigureGigabitEthernet1/0/1 as a trunk port, and assign the port to VLAN 10.

```
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 10
[DeviceA-GigabitEthernet1/0/1] return
```

Enable log output to the terminal

```
<DeviceA> terminal monitor
```

Enable the information center.

```
<DeviceA> system-view
[DeviceA] info-center enable
```

Enable MAC Information globally.

```
[DeviceA] mac-address information enable
```

Configure the MAC Information mode as syslog.

```
[DeviceA] mac-address information mode syslog
```

Enable MAC Information on GigabitEthernet1/0/1 to enable the port to record MAC address change information when the interface performs either of the following operations:

- Learns a new MAC address.
- Deletes an existing MAC address.

```
[DeviceA] interface gigabitethernet 1/0/1
```

```
[DeviceA-GigabitEthernet1/0/1] mac-address information enable added
[DeviceA-GigabitEthernet1/0/1] mac-address information enable deleted
[DeviceA-GigabitEthernet1/0/1] quit
```

Set the MAC Information queue length to 100.

```
[DeviceA] mac-address information queue-length 100
```

Set the MAC change notification interval to 20 seconds.

```
[DeviceA] mac-address information interval 20
```

Verifying the configuration

When the PC pings Device A, Device A dynamically learns the source MAC address of the PC on GigabitEthernet1/0/1 and adds a new entry to its MAC address table. Verify that Device A generates log messages when new MAC address entries are learned on GigabitEthernet1/0/1.

```
<DeviceA>%Jan  6 12:43:32:526 2025 Device A MAC/6/MAC_UPDATEMSG:
-Action=Add-MacAddr=7278-a742-54f8-VlanID=10-MacType=1-IfName=GigabitEthernet1/0/1;
MAC address table changed.
```

When GigabitEthernet1/0/1 is shut down on Device A, all MAC address entries on this interface are deleted. Verify that Device A generates log messages when MAC address entries are deleted on GigabitEthernet1/0/1.

```
<DeviceA>%Jan  6 12:46:17:326 2025 Device A MAC/6/MAC_UPDATEMSG:
-Action=Delete-MacAddr=7278-a742-54f8-VlanID=10-MacType=1-IfName=GigabitEthernet1/0/1;
MAC address table changed.
```

Configuration files

```
#
 info-center enable
#
 mac-address information queue-length 100
 mac-address information interval 20
 mac-address information mode syslog
 mac-address information enable
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 10
 mac-address information enable added
 mac-address information enable deleted
#
interface Vlan-interface10
 ip address 192.168.1.193 255.255.255.0
#
```

Contents

Bulk configuring interfaces 1
About interface bulk configuration 1
Restrictions and guidelines: Bulk interface configuration 1
Procedure 2
Display and maintenance commands for bulk interface configuration 2

Bulk configuring interfaces

About interface bulk configuration

You can enter interface range view to bulk configure multiple interfaces with the same feature instead of configuring them one by one. For example, you can execute the **shutdown** command in interface range view to shut down a range of interfaces.

To configure interfaces in bulk, you must configure an interface range and enter its view by using the **interface range** or **interface range name** command.

The interface range created by using the **interface range** command is not saved to the running configuration. You cannot use the interface range repeatedly. To create an interface range that can be used repeatedly, use the **interface range name** command.

Restrictions and guidelines: Bulk interface configuration

When you bulk configure interfaces in interface range view, follow these restrictions and guidelines:

- In interface range view, only the commands supported by the first interface in the specified interface list (alphabetically sorted) are available for configuration.
- Before you configure an interface as the first interface in an interface range, make sure you can enter the view of the interface by using the **interface** *interface-type* { *interface-number* | *interface-number.subnumber* } command.
- Do not assign both an aggregate interface and any of its member interfaces to an interface range. Some commands, after being executed on both an aggregate interface and its member interfaces, can break up the aggregation.
- Understand that the more interfaces you specify, the longer the command execution time.
- To guarantee bulk interface configuration performance, configure fewer than 1000 interface range names.

The device does not output prompt or alarm messages during the bulk interface configuration process. Make sure you are fully aware of the support and impacts of bulk interface configuration.

- After a command is executed in interface range view, one of the following situations might occur:
  - The system displays an error message and stays in interface range view. This means that the execution failed on one or multiple member interfaces.
    - If the execution failed on the first member interface, the command is not executed on any member interfaces.
    - If the execution failed on a non-first member interface, the command takes effect on the remaining member interfaces.
  - The system returns to system view. This means that:
    - The command is supported in both system view and interface view.
    - The execution failed on a member interface in interface range view and succeeded in system view.
    - The command is not executed on the subsequent member interfaces.

You can use the **display this** command to verify the configuration in interface view of each member interface. In addition, if the configuration in system view is not needed, use the **undo** form of the command to remove the configuration.

Procedure

1. Enter system view.

   system-view
2. Create an interface range and enter interface range view.
   - **Create an interface range without specifying a name.**

     interface range { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>
   - **Create a named interface range.**

     interface range name *name* [**interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]
   - **Add or delete interfaces for an existing interface range.**

     interface range name *name* [**add** | **delete**] **interface** { *interface-type interface-number* [**to** *interface-type interface-number*] } &<1-24>]
3. (Optional.) Display commands available for the first interface in the interface range.

   Enter a question mark (?) at the interface range prompt.
4. Use available commands to configure the interfaces.

   Available commands depend on the interface.
5. (Optional.) Verify the configuration.

   display this

Display and maintenance commands for bulk interface configuration

Execute the **display** command in any view.

Task	Command
Display information about the interface ranges created by using the **interface range name** command.	**display interface range** [**name** *name*]

Contents

M-LAG overview 1

About M-LAG 1
M-LAG network model 1
M-LAG benefits 2
M-LAG setup mechanism 3
M-LAG system setup process 3
DRCP 3
Device role calculation 4
Configuration consistency check 6
Keepalive and failover mechanism 11
MAD mechanism 12
M-LAG traffic forwarding 12
M-LAG loop avoidance mechanism 12
Unicast traffic forwarding 13
BUM traffic forwarding 15
M-LAG failure handling 16
M-LAG standalone mode 16
M-LAG failure handling mechanisms 17
Mechanisms to handle concurrent peer link and keepalive link failures 19
Extended M-LAG features 21
Dual-active gateways on an M-LAG system 21
M-LAG active/standby mode 22
M-LAG sequence number check 26
M-LAG packet authentication 27
Multichassis configuration synchronization 27
Real-time configuration sync 27
Bulk configuration sync 27
Configuration backup 28
Protocols and standards 28

Configuring M-LAG 28

Restrictions and guidelines: M-LAG configuration 28
Software version requirements 28
M-LAG configuration 29
Compatibility with other features 29
Support for DRNI commands 31
M-LAG tasks at a glance 32
Configuring multichassis configuration sync 33
Restrictions and guidelines for multichassis configuration sync 33
Synchronizing the configurations in real time 33
Bulk synchronizing the configurations 34
Fast configuring differential settings 34
Backing up and restoring a configuration file 35
Configuring M-LAG system settings 35
Configuring the M-LAG system MAC address 35
Setting the M-LAG system number 36
Setting the M-LAG system priority 36
Setting the M-LAG role priority of the device 37
Enabling M-LAG standalone mode on an M-LAG member device 37
Configuring M-LAG keepalive settings 38
Restrictions and guidelines for configuring M-LAG keepalive settings 38
Configuring M-LAG keepalive packet parameters 38
Setting the M-LAG keepalive interval and timeout timer 39
Associating the keepalive link with a track entry 39
Configuring M-LAG MAD 40
About this task 40
Restrictions and guidelines for M-LAG MAD 41

Configuring the default M-LAG MAD action on network interfaces ····· 41
Excluding an interface from the shutdown action by M-LAG MAD ····· 42
Excluding all logical interfaces from the shutdown action by M-LAG MAD ····· 42
Specifying interfaces to be shut down by M-LAG MAD when the M-LAG system splits ····· 43
Enabling M-LAG MAD DOWN state persistence ····· 43
Configuring an M-LAG interface ····· 44
Configuring an M-LAG uplink interface ····· 45
Configuring the flapping detection interval for M-LAG uplink interfaces ····· 46
Setting up a peer link ····· 47
Specifying a Layer 2 aggregate interface or VXLAN tunnel interface as the peer-link interface ····· 47
Reserving VLANs for the peer link ····· 48
Enabling the short DRCP timeout timer on the peer-link interface or an M-LAG interface ····· 48
Enabling the peer-link interface to retain MAC address entries for down single-homed devices ····· 49
Configuring the M-LAG active/standby mode ····· 49
Assigning a M-LAG virtual IP address to an interface ····· 50
Configuring M-LAG extra VLANs ····· 51
Setting the mode of configuration consistency check ····· 52
Disabling configuration consistency check ····· 53
Setting the keepalive hold timer for identifying the cause of peer link down events ····· 53
Configuring M-LAG system auto-recovery ····· 53
Setting the data restoration interval ····· 54
Enabling M-LAG sequence number check ····· 55
Enabling M-LAG packet authentication ····· 55
Displaying and maintaining M-LAG ····· 55
M-LAG configuration examples ····· 56
Example: Configuring basic M-LAG functions ····· 56
Example: Configuring Layer 3 gateways on an M-LAG system ····· 63
Example: Configure dual-active gateways on an M-LAG system ····· 77
Comparison between M-LAG and DRNI commands ····· 95
Configuration commands ····· 95

M-LAG overview

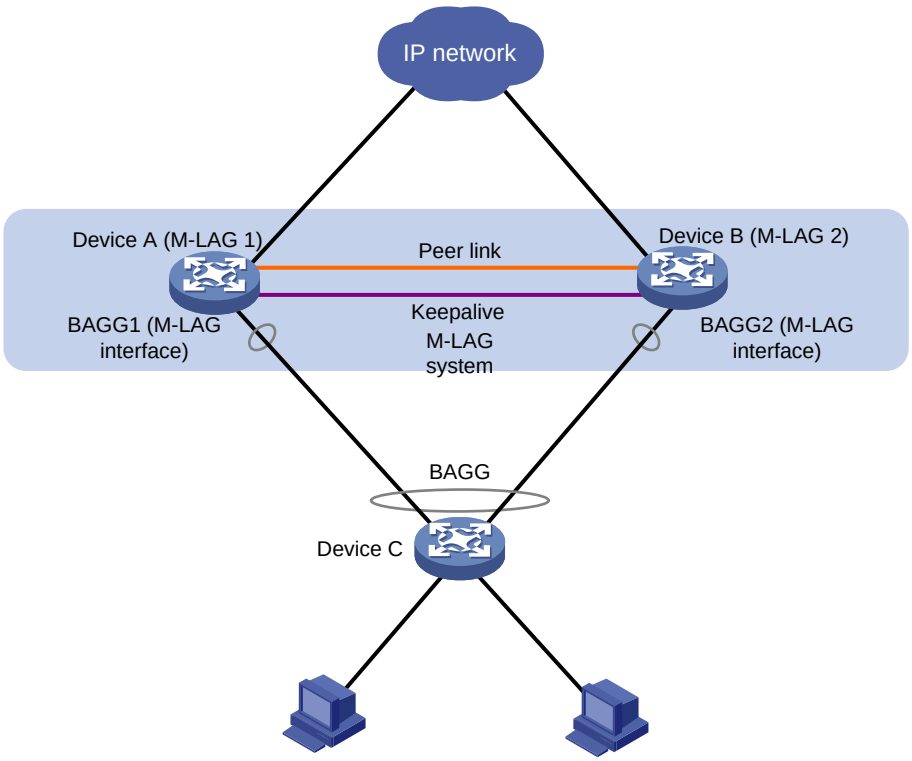
About M-LAG

Multichassis Link Aggregation (M-LAG) virtualizes two physical devices into one system through multichassis link aggregation.

M-LAG network model

As shown in Figure 1, M-LAG virtualizes two devices into an M-LAG system, which connects to the remote aggregation system through a multichassis aggregate link. To the remote aggregation system, the M-LAG system is one device.

Figure 1 M-LAG network model



The M-LAG member devices are M-LAG peers to each other. For features that require centralized traffic processing (for example, spanning tree), an M-LAG member device is assigned the primary or secondary role based on its M-LAG role priority. The secondary M-LAG member device passes the traffic of those features to the primary M-LAG member device for processing. If the M-LAG member devices in an M-LAG system have the same M-LAG role priority, the device with the lower bridge MAC address is assigned the primary role.

M-LAG defines the following interface roles for each M-LAG member device:

- **M-LAG interface**—Layer 2 aggregate interface connected to the remote aggregation system. M-LAG interfaces connected to the same remote aggregation system belong to one M-LAG group. In Figure 1, Bridge-Aggregation 1 on Device A and Bridge-Aggregation 2 on Device B belong to the same M-LAG group. M-LAG interfaces in an M-LAG group form a multichassis aggregate link.

- **Peer-link interface**—Interface connected to the M-LAG peer for internal control. Each M-LAG member device has only one peer-link interface. The peer-link interfaces of the M-LAG member devices transmit M-LAG protocol packets and data packets through the peer link established between them. An M-LAG system has only one peer link.

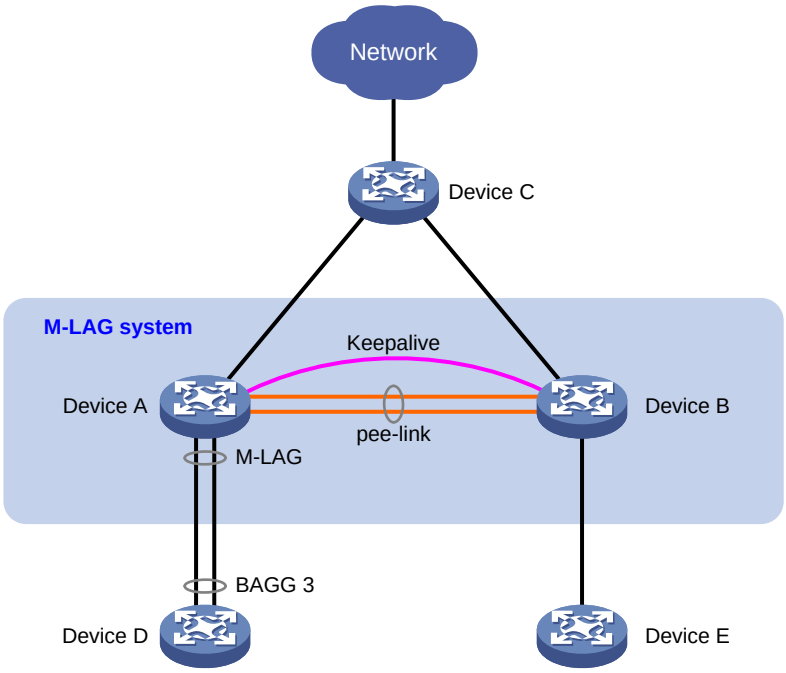
M-LAG member devices use a keepalive link to monitor each other's state. For more information about the keepalive mechanism, see "Keepalive and failover mechanism."

If a device is attached to only one of the M-LAG member devices in an M-LAG system, that device is a single-homed device. Single-homed access includes the following types:

- **M-LAG single-homed access**—Connects to one of the M-LAG member devices through an M-LAG interface.
- **Non-M-LAG single-homed access**—Connects to one of the M-LAG member devices through a non-M-LAG interface.

As shown in Figure 2, Device D connects to the M-LAG system in M-LAG single-homed mode, while Device E connects to the M-LAG system in non-M-LAG single-homed mode. Device D and Device E connect to the M-LAG system in single-homed mode. The MAC address entries and ARP/ND entries of Device D and E are backed up between the M-LAG member devices. This provides a backup path for north-south traffic, enhancing reliability.

Figure 2 Connecting to an M-LAG network in single-homed access mode



M-LAG benefits

Traditional aggregate links rely on interoperation between single devices. If a link or peer device fails, normal communication with the peer device might fail. The M-LAG technology addresses this issue by aggregating ports from different devices in the system into one logical interface, effectively preventing single point failures from affecting the overall link and ensuring reliable data flow transmission. Additionally, the M-LAG system features a loop-free Layer 2 topology, allowing for more efficient use of links. Two M-LAG member devices can be upgraded independently, reducing service interruption time.

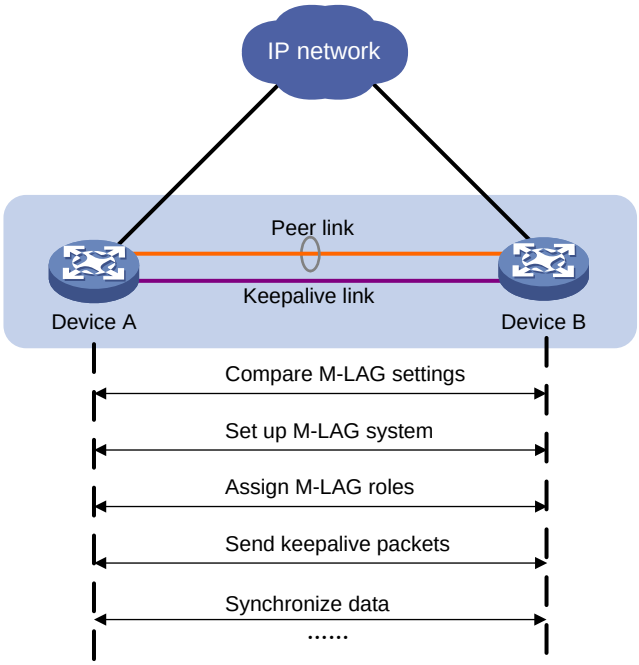
M-LAG setup mechanism

M-LAG system setup process

As shown in Figure 3, two devices perform the following operations to form an M-LAG system:

1. Send DRCPDUs over the peer link to each other and compare the DRCPDUs to determine the M-LAG system stackability and device roles:
   a. Compare the M-LAG system settings. The devices can form an M-LAG system if they have consistent M-LAG system settings.
   b. Determine the device roles as described in "Device role calculation."
   c. Perform configuration consistency check. For more information, see "Configuration consistency check."
2. Send keepalive packets over the keepalive link after primary M-LAG member election to verify that the peer system is operating correctly.
3. Synchronize configuration data by sending DRCPDUs over the peer link. The configuration data includes MAC address entries and ARP entries.

Figure 3 M-LAG system setup process



DRCP

M-LAG uses H3C proprietary Distributed Relay Control Protocol (DRCP) for multichassis link aggregation. DRCP runs on the peer link and uses distributed relay control protocol data units (DRCPDUs) to advertise the M-LAG configuration out of peer-link interfaces and M-LAG interfaces.

DRCP operating mechanism

M-LAG-enabled devices use DRCPDUs for the following purposes:

- Exchange DRCPDUs through M-LAG interfaces to determine whether they can form an M-LAG system.
- Exchange DRCPDUs through peer-link interfaces to negotiate the peer link state.

DRCP timeout timers

DRCP uses a timeout mechanism to specify the amount of time that a peer-link interface or M-LAG interface must wait to receive DRCPDUs before it determines that the peer interface is down. This timeout mechanism provides the following timer options:

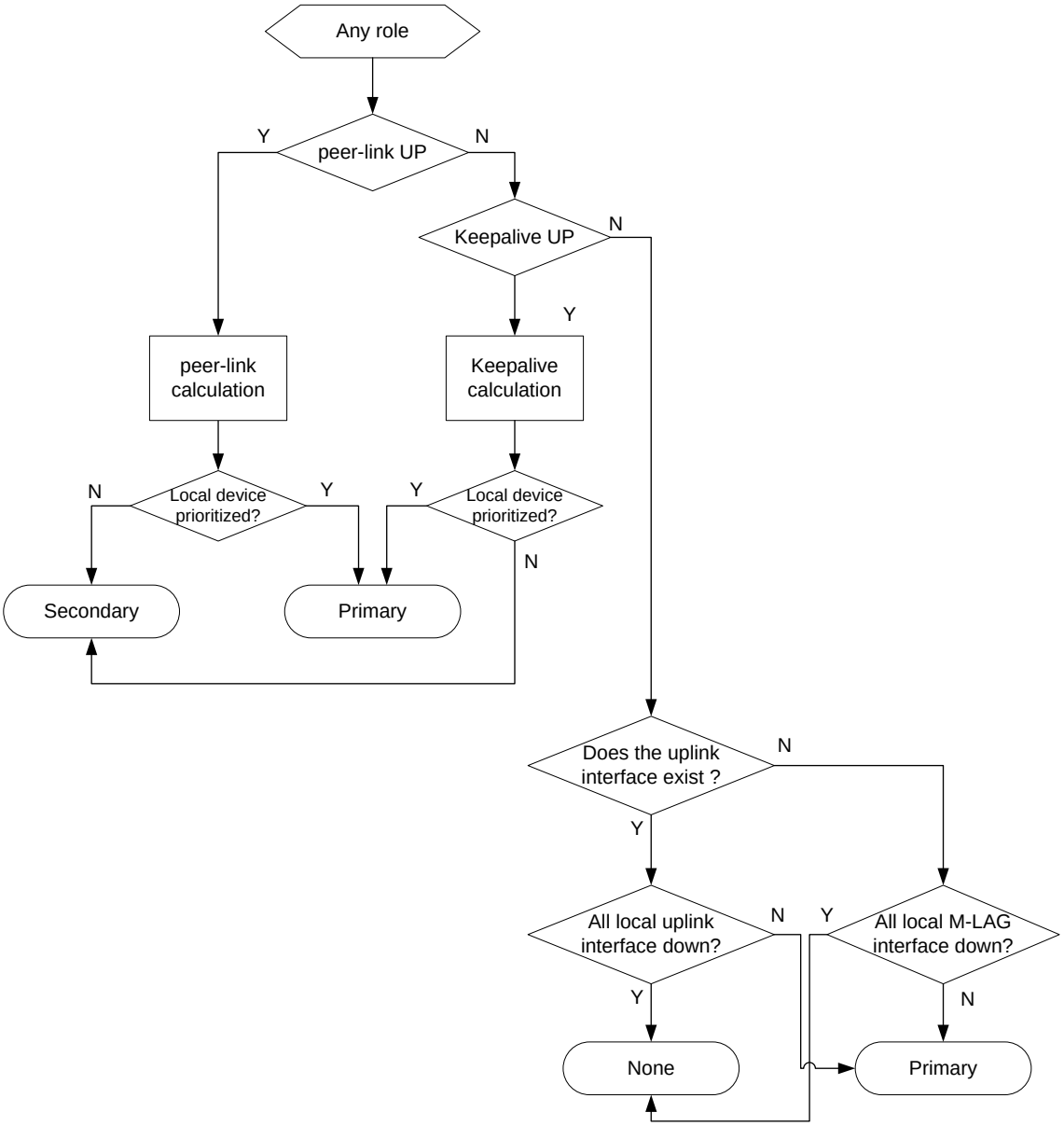
- Short DRCP timeout timer, which is fixed at 3 seconds. If this timer is used, the peer interface sends one DRCPDU every second.
- Long DRCP timeout timer, which is fixed at 90 seconds. If this timer is used, the peer interface sends one DRCPDU every 30 seconds.

Short DRCP timeout timer enables the M-LAG member devices to detect a peer interface down event more quickly than the long DRCP timeout timer. However this benefit is at the expense of bandwidth and system resources.

Device role calculation

As shown in Figure 4, the role of an M-LAG member device can be primary, secondary, or none after role calculation.

Figure 4 M-LAG role calculation process



Role calculation rules

M-LAG calculates the roles of the M-LAG member devices according to the following rules:

- The M-LAG roles are determined upon M-LAG system initialization triggered by M-LAG system setup or reboot of an M-LAG member device.
- If the peer link is up, the M-LAG member devices exchange DRCPDUs over the peer link to determine which of them to take the primary role.
- If the peer link is down while the keepalive link is up, the M-LAG member devices exchange keepalive packets over the link to determine their roles.
- If both the peer link and the keepalive link are down, an M-LAG member device takes the primary role if it has available uplink interfaces or M-LAG interfaces.

Factors in role calculation

When the peer link or keepalive link is up, the M-LAG member devices exchange the following information to determine which of them takes the primary role:

1. (Applicable only when the peer link fails but the keepalive link is operating correctly.) When both M-LAG member devices have configured uplink interfaces, the one with the uplink interface that comes up first is preferred. If both devices do not have uplink interfaces configured, skip this comparison factor.
2. Status of M-LAG interfaces. An M-LAG member device takes the primary role if it has available M-LAG interfaces. This status is skipped if role calculation is performed over the peer link.
3. Device roles before calculation. If one device already has the primary role, the primary device retains its role.
4. M-LAG MAD DOWN state. If one device has not placed any network interfaces in M-LAG MAD DOWN state, it becomes the primary device.
5. Health state. The healthier device takes the primary role.

   To view the health state of the device, execute the **`display system health`** command. The smaller the health state value, the healthier the device is. The health state value is 0 if the device is running without faults. For more information about the **`display system health`** command, see device management commands in *Fundamentals Command Reference*.
6. M-LAG role priority. The device with higher M-LAG role priority takes the primary role.
7. Bridge MAC address. The device with a lower bridge MAC address takes the primary role.

The device that has failed the election takes the secondary role.

Configuration consistency check

During M-LAG system setup, M-LAG member devices exchange the configuration and perform configuration consistency check to verify their consistency in the following configurations:

- **Type 1 configuration**—Settings that affect traffic forwarding of the M-LAG system. If an inconsistency in type 1 configuration is detected, the secondary M-LAG member device shuts down its M-LAG interfaces.
- **Type 2 configuration**—Settings that affect only service features. If an inconsistency in type 2 configuration is detected, the secondary M-LAG member device disables the affected service features, but it does not shut down its M-LAG interfaces.

To prevent interface flapping, the M-LAG system performs configuration consistency check when half the data restoration interval elapses.

NOTE:

The data restoration interval specifies the maximum amount of time for the secondary M-LAG member device to synchronize data with the primary M-LAG member device during M-LAG system setup. For more information, see "Setting the data restoration interval."

Type 1 configuration

Type 1 configuration consistency check is performed both globally and on M-LAG interfaces. Table 1 and Table 2 show settings that type 1 configuration contains.

Table 1 Global type 1 configuration

Setting	Details
Peer-link interface link type	Peer-link interface link type, including access, hybrid, and trunk.
PVID on the peer-link interface	PVID on the peer-link interface.
Spanning tree state	• Global spanning tree state. • VLAN-specific spanning tree state. M-LAG checks the VLAN-specific spanning tree state only when PVST is enabled.

Setting	Details
Spanning tree mode	Spanning tree mode, including STP, RSTP, PVST, and MSTP.
MST region settings	• MST region name. • MST region revision level. • VLAN-to-MSTI mappings.

Table 2 M-LAG interface type 1 configuration

Setting	Details
Aggregation mode	Aggregation mode, including static and dynamic.
Spanning tree state	Interface-specific spanning tree state.
Link type	Interface link type, including access, hybrid, and trunk.
PVID	Interface PVID.

Type 2 configuration

Type 2 configuration consistency check is performed both globally and on M-LAG interfaces. Table 3 and Table 4 show settings that type 2 configuration contains.

Table 3 Global type 2 configuration

Setting	Details
VLANs permitted by the peer-link interface	VLANs permitted by the peer-link interface. The M-LAG system compares tagged VLANs prior to untagged VLANs.
VLAN interfaces	Up VLAN interfaces of which the VLANs contain the peer-link interface.
VLAN interface status	Whether a VLAN interface is in administratively down state.
IPv4 address of a VLAN interface	IPv4 address assigned to a VLAN interface.
IPv6 address of a VLAN interface	IPv6 address assigned to a VLAN interface.
Virtual IPv4 address of the VRRP group on a VLAN interface	Virtual IPv4 address of the VRRP group configured on a VLAN interface on the VRRP master.
Global BPDU guard	Global status of BPDU guard.
MAC aging timer	Aging timer for dynamic MAC address entries.
Global state of port security.	Whether port security is enabled globally.
Authentication load sharing mode for users on port security-enabled M-LAG interfaces	Authentication load sharing mode for users on port security-enabled M-LAG interfaces: • **Centralized**—The primary M-LAG member device authenticates users. • **Local**—Each M-LAG member device authenticates their local users. • **Odd-MAC**—The local M-LAG member device authenticates odd-MAC users on all M-LAG interfaces of the M-LAG system. • **Even-MAC**—The local M-LAG member device authenticates even-MAC users on all M-LAG interfaces of the M-LAG system.
MAC move	Status of the MAC move feature.

Setting	Details
MAC move mode	Port security MAC move mode: • **Port**—Allows an authenticated online user to move between ports on the device. • **VLAN**—Allows an authenticated online user to move between VLANs on a trunk or hybrid port. • **All**—Allows an authenticated online user to move between ports on the device or VLANs on a trunk or hybrid port.
State of traffic statistics for 802.1X and MAC authentication users	Whether traffic statistics collection is enabled globally for 802.1X and MAC authentication users.
Global state of 802.1X authentication	Whether 802.1X authentication is enabled globally.
802.1X authentication method	802.1X authentication method: • **Chap**—Performs EAP termination and uses CHAP to communicate with the RADIUS server. • **Eap**—Relays EAP packets and supports any of the EAP authentication methods to communicate with the RADIUS server. • **Pap**—Performs EAP termination and uses PAP to communicate with the RADIUS server.
EAD rule timer of 802.1X authentication	EAD rule timer value for 802.1X authentication.
Handshake timer of 802.1X authentication	Handshake timer value for 802.1X authentication.
Offline detection timer of 802.1X authentication	Offline detection timer value for 802.1X authentication.
Quiet timer of 802.1X authentication	Quiet timer value for 802.1X authentication.
Periodic reauthentication timer of 802.1X authentication	Periodic reauthentication timer value for 802.1X authentication.
Server timeout timer of 802.1X authentication	Server timeout timer value for 802.1X authentication.
Client timeout timer of 802.1X authentication	Client timeout timer value for 802.1X authentication.
Username request timeout timer of 802.1X authentication	Username request timeout timer value for 802.1X authentication.
Aging timer for users in Auth-Fail VLANs of 802.1X authentication	Aging timer value for users in Auth-Fail VLANs of 802.1X authentication.
Aging timer for users in Auth-Fail VSIs of 802.1X authentication	Aging timer value for users in Auth-Fail VSIs of 802.1X authentication.
Aging timer for users in critical VLANs of 802.1X authentication	Aging timer value for users in critical VLANs of 802.1X authentication.
Aging timer for users in critical VSIs of 802.1X authentication	Aging timer value for users in critical VSIs of 802.1X authentication.
Aging timer for users in guest VLANs of 802.1X authentication	Aging timer value for users in guest VLANs of 802.1X authentication.
Aging timer for users in guest VSIs of 802.1X authentication	Aging timer value for users in guest VSIs of 802.1X authentication.
Global state of MAC authentication	Whether MAC authentication is enabled globally.
MAC authentication method	MAC authentication method:

Setting	Details
	• **Chap**—CHAP authentication. • **Pap**—PAP authentication.
Offline detection timer of MAC authentication	Offline detection timer value for MAC authentication.
Quiet timer of MAC authentication	Quiet timer value for MAC authentication.
Periodic reauthentication timer of MAC authentication	Periodic reauthentication timer value for MAC authentication.
Server timeout timer of MAC authentication	Server timeout timer value for MAC authentication.
Aging timer for temporary MAC authentication users	Aging timer value for temporary MAC authentication users.
Aging timer for users in critical VLANs of MAC authentication	Aging timer value for users in critical VLANs of MAC authentication.
Aging timer for users in critical VSIs of MAC authentication	Aging timer value for users in critical VSIs of MAC authentication.
Aging timer for users in guest VLANs of MAC authentication	Aging timer value for users in guest VLANs of MAC authentication.
Aging timer for users in guest VSIs of MAC authentication	Aging timer value for users in guest VSIs of MAC authentication.
Aging timer for MAC address entries created for temporary MAC authentication users	Aging timer value for MAC address entries created for temporary MAC authentication users.
VSI name	Name of a VSI that has ACs on an M-LAG interface.
VXLAN ID	VXLAN ID of a VSI.
Gateway interface	VSI interface associated with a VSI.
VSI interface number	Number of a VSI interface.
MAC address of a VSI interface	MAC address assigned to a VSI interface.
IPv4 address of a VSI interface	IPv4 address assigned to a VSI interface.
IPv6 address of a VSI interface	IPv6 address assigned to a VSI interface.
Physical state of a VSI interface	Physical link state of a VSI interface.
Protocol state of a VSI interface	Data link layer state of a VSI interface.

The device displays the following global type 2 settings only when VLAN or VLAN interface configuration inconsistency exists:

- VLAN interface status.
- IPv4 address of a VLAN interface.
- IPv6 address of a VLAN interface.
- Virtual IPv4 address of the VRRP group on a VLAN interface.

Table 4 M-LAG interface type 2 configuration

Setting	Details
VLANs permitted by an M-LAG interface	VLANs permitted by an M-LAG interface. The M-LAG system compares tagged VLANs prior to untagged VLANs.

Setting	Details
Using port speed as the prioritized criterion for reference port selection	Whether an M-LAG interface uses port speed as the prioritized criterion for reference port selection.
Ignoring port speed in setting the aggregation states of member ports	Whether an M-LAG interface ignores port speed in setting the aggregation states of member ports.
Root guard status	Status of root guard.
Port security mode	Port security mode: • Autolearn. • Mac-authentication. • Mac-and-userlogin-secure-ext. • Mac-else-userlogin-secure. • Mac-else-userlogin-secure-ext. • Secure. • Userlogin. • Userlogin-secure. • Userlogin-secure-ext. • Userlogin-secure-or-mac. • Userlogin-secure-or-mac-ext. • Userlogin-withoui.
Secure MAC address aging type	Secure MAC address aging type: • **Inactivity**—Inactivity aging feature together with the aging timer. • **Periodical**—Timer aging only.
State of 802.1X authentication on an interface	Whether 802.1X authentication is enabled on an interface.
802.1X online user handshake	Status of the 802.1X online user handshake feature.
802.1X multicast trigger	Status of the 802.1X multicast trigger feature.
802.1X unicast trigger	Status of the 802.1X unicast trigger feature.
802.1X authentication offline detection	State of 802.1X authentication offline detection on an M-LAG interface: • **Enabled**. • **Disabled**.
Periodic reauthentication timer of 802.1X authentication	Periodic reauthentication timer of 802.1X authentication.
802.1X authentication critical profile	802.1X authentication critical profile.
802.1X critical VLAN	802.1X critical VLAN which an M-LAG interface is assigned to.
802.1X packet detection	State of 802.1X packet detection.
802.1X packet detection retries	Number of 802.1X packet detection retries.
State of MAC authentication on an interface	Whether MAC authentication is enabled on an interface.
Parallel processing of MAC authentication and 802.1X authentication on a port	Status of parallel processing of MAC authentication and 802.1X authentication on a port.
MAC authentication offline detection	State of MAC authentication offline detection on an M-LAG interface: • **Enabled**.

Setting	Details
	• **Disabled**.
MAC authentication VLAN mode for users moving from one VLAN to another	MAC authentication VLAN mode for users moving from one VLAN to another on an M-LAG interface: • **Single VLAN**—Single-VLAN mode. • **Multiple VLAN**—Multi-VLAN mode.
MAC authentication delay timer	MAC authentication delay timer value.
Periodic reauthentication timer of MAC authentication	Periodic reauthentication timer value for MAC authentication.
MAC authentication critical profile	MAC authentication critical profile name.
Critical VLAN of MAC authentication	Critical VLAN of MAC authentication.
MAC authentication packet detection	State of MAC authentication packet detection.
MAC authentication packet detection retries	Number of MAC authentication packet detection retries.
Online detection interval of Web authentication	Online detection interval of Web authentication.

Keepalive and failover mechanism

H3C provides proprietary keepalive mechanism to detect the availability of the M-LAG member devices.

In normal conditions, an M-LAG system can have only one primary device, which is responsible for processing control messages and distributing data-plane traffic for load sharing. When a failure occurs in the M-LAG system, both M-LAG member devices might consider the peer device unavailable and both attempt to become the primary device, a situation known as a dual-primary conflict. A dual-primary conflict can lead to issues such as loops, failure of traffic load sharing, and flapping of forwarding entries. To prevent dual-primary conflicts, M-LAG implements a dual-primary detection mechanism that uses the keepalive mechanism to ensure that only one device operates as the primary device.

For the secondary M-LAG member device to monitor the state of the primary device, you must establish a Layer 3 keepalive link between the M-LAG member devices.

The M-LAG member devices periodically send keepalive packets over the keepalive link. If an M-LAG member device has not received keepalive packets from the peer when the keepalive timeout timer expires, it determines that the keepalive link is down. When both the keepalive link and the peer link are down, an M-LAG member device acts depending on its role.

- When the local device is the primary device and has uplink interfaces configured, it remains as the primary device if any uplink interface is in up state. If no uplink interface in up state exists, the local device's role changes to None.
- When the local device is the primary device and no uplink interface is configured on it, the local device remains as the primary device if an M-LAG interface in up state exists. If no M-LAG interface in up state exists, the local device's role changes to None.
- When the local device is a secondary device, it becomes the primary device. If the local device is configured with uplink interfaces, it will remain as the primary device as long as an uplink interface in up state exists on the local device. If no uplink interface in up state exists, the local device's role will change to None.

- When the local device is a secondary device, it becomes the primary device. If the local device does not have an uplink interface configured, it will remain as the primary device as long as an M-LAG interface in up state exists on the device. If no M-LAG interface in up state exists, the local device's role will change to None.

A device with the None role cannot send or receive keepalive packets. Its keepalive link stays in the down state.

If the keepalive link is down while the peer link is up, the M-LAG member devices prompt you to check for keepalive link issues.

If the keepalive link is up while the peer link is down, the M-LAG member devices elect a primary device based on the information in the keepalive packets.

MAD mechanism

M-LAG MAD

A multi-active collision occurs if the peer link goes down while the keepalive link is up. To avoid network issues, M-LAG MAD shuts down all network interfaces on the secondary M-LAG member device except those manually or automatically excluded.

When the peer link comes up, the secondary M-LAG member device starts a delay timer and begins to restore table entries (including MAC address entries and ARP entries) from the primary M-LAG member device. When the delay timer expires, the secondary M-LAG member device brings up all network interfaces placed in M-LAG MAD DOWN state.

M-LAG MAD DOWN state persistence

Both of the M-LAG member devices might take the primary role if both of them have M-LAG interfaces in up state after the following series of events occur:

1. The peer link goes down while the keepalive link is up. Then, M-LAG MAD shuts down all network interfaces on the secondary M-LAG member device except those excluded from the shutdown action by M-LAG MAD.
2. The keepalive link also goes down. Then, the secondary M-LAG member device brings up the network interfaces in M-LAG MAD DOWN state and sets its role to primary.

M-LAG MAD DOWN state persistence helps avoid the forwarding issues that might occur in the multi-active situation that occurs because the keepalive link goes down while the peer link is down.

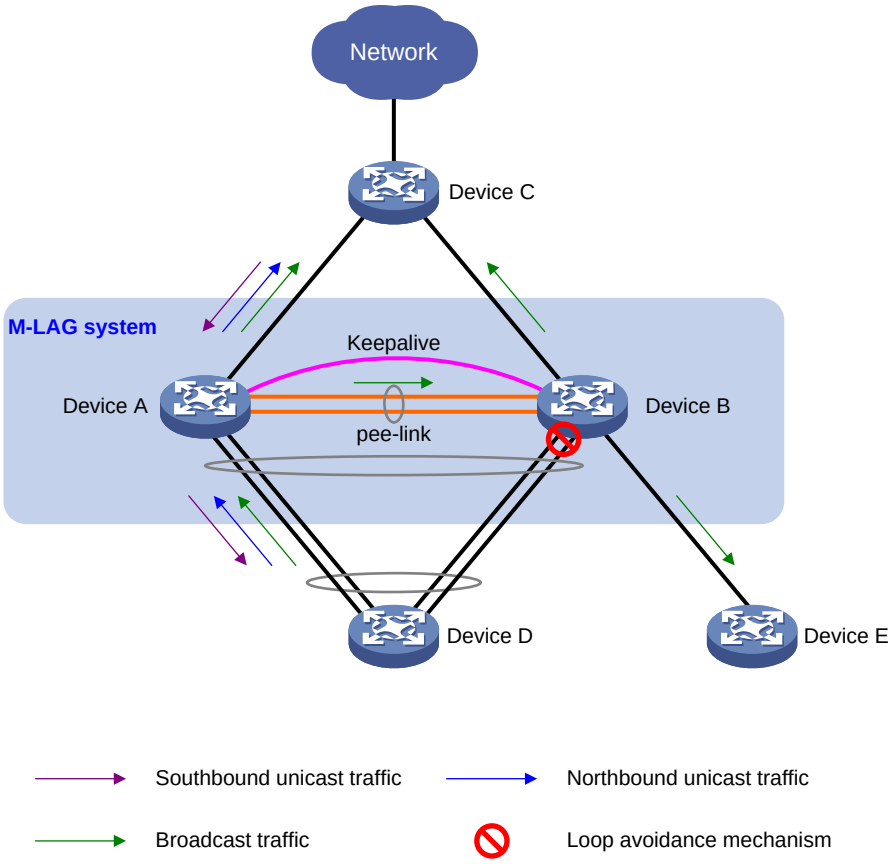
To prevent packet loss after the peer link recovers from a link fault, the secondary member device should try to synchronize entries (such as the MAC address entries and ARP entries) within the data restoration interval. Subsequently, interfaces in M-LAG MAD DOWN state on the device will restore to up.

M-LAG traffic forwarding

M-LAG loop avoidance mechanism

An M-LAG system is a loop-free network. As shown in Figure 5, when north-south unicast traffic reaches the M-LAG system, it is preferentially forwarded directly locally. Typically, the peer link does not forward data traffic. Only BUM traffic is forwarded to the peer M-LAG member device through the peer link. The M-LAG device does not forward traffic received from the peer link out through the local M-LAG interface, thus preventing loops. This is the loop avoidance mechanism of M-LAG.

Figure 5 Schematic diagram for the M-LAG loop avoidance mechanism



NOTE:

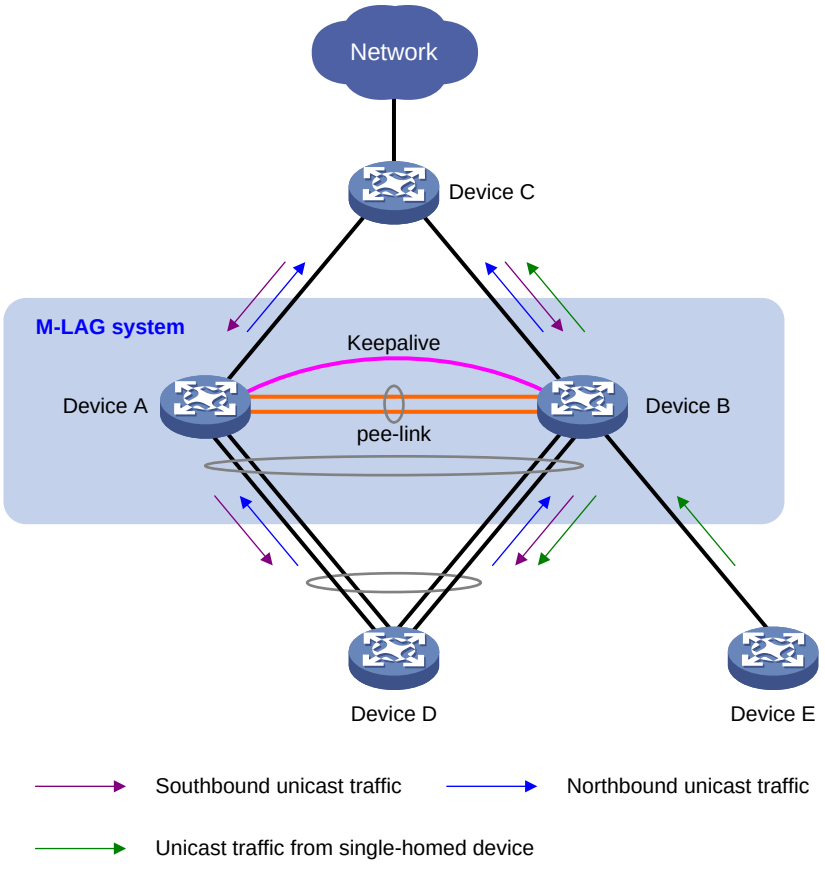
The preceding loop avoidance mechanism takes effect on dual-homed devices but does not take effect on single-homed devices. Single-homed devices will still receive traffic forwarded from the M-LAG peer link.

Unicast traffic forwarding

As shown in Figure 6, Device D and Device E connect to the M-LAG system. The known unicast traffic is forwarded as follows:

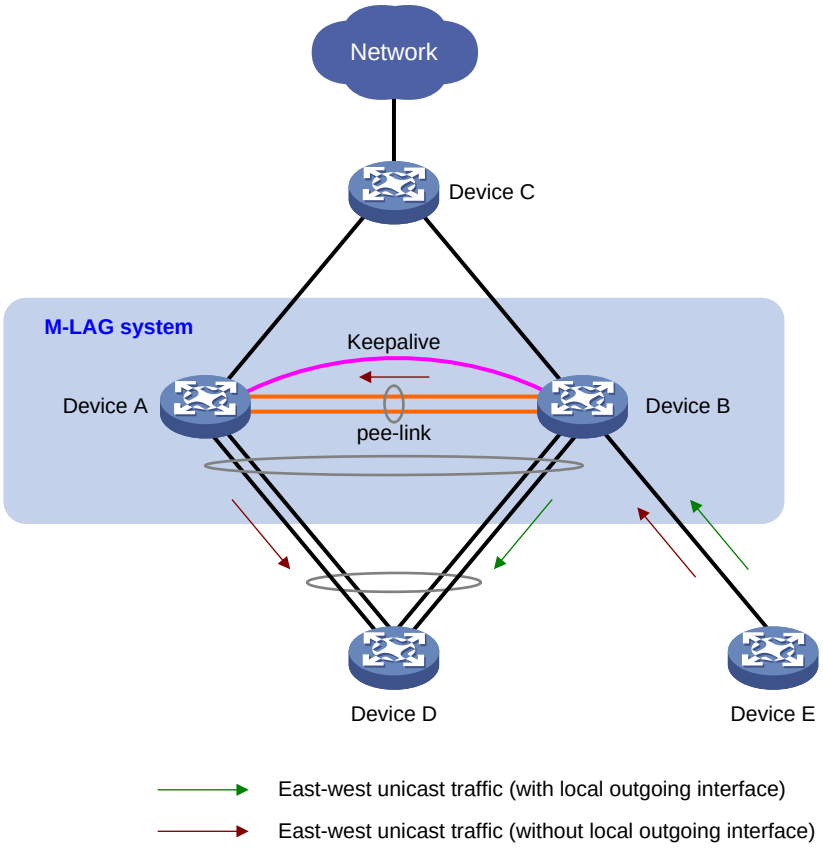
- For north-south unicast traffic, the M-LAG system preferentially forwards the traffic directly upon receiving the traffic, instead of sending it to the peer through the peer link.
- For east-west unicast traffic (traffic forwarded between access devices), the traffic is preferentially forwarded locally on the M-LAG member devices.

Figure 6 Schematic diagram for unicast traffic forwarding



As shown in Figure 7, for east-west traffic, if the traffic is received and no local outgoing interface is available, the traffic will traverse through the peer link and be forwarded by the peer M-LAG member device.

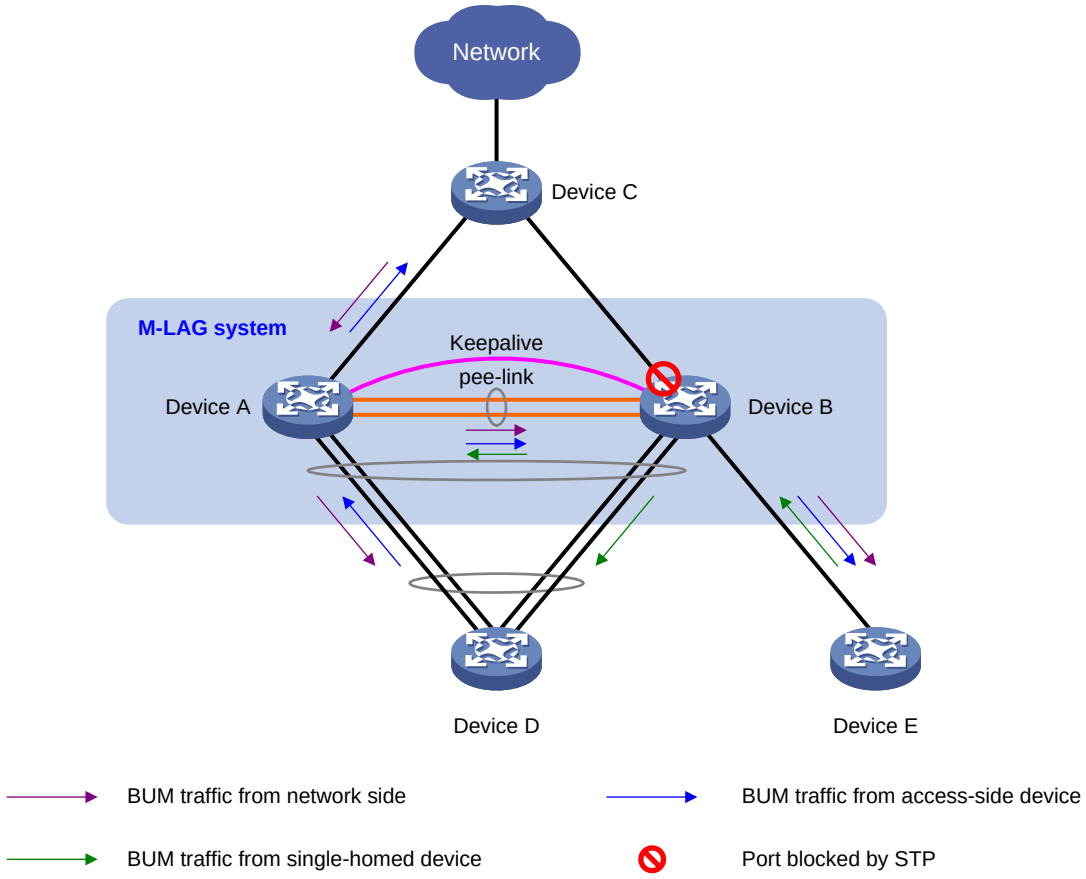
Figure 7 Schematic diagram east-west traffic forwarding



BUM traffic forwarding

As shown in Figure 8, BUM traffic will be forwarded through the peer link, but the loop avoidance mechanism prevents an access device from receiving its own BUM traffic.

Figure 8 Schematic diagram for BUM traffic forwarding



M-LAG failure handling

M-LAG standalone mode

The M-LAG member devices might both operate with the primary role to forward traffic if they have M-LAG interfaces in up state after the M-LAG system splits. M-LAG standalone mode helps avoid traffic forwarding issues in this multi-active situation by allowing only the member ports in the M-LAG interfaces on one member device to forward traffic.

The following information describes the operating mechanism of this feature.

The M-LAG member devices change to M-LAG standalone mode when they detect that both the peer link and the keepalive link are down. In addition, the secondary M-LAG member device changes its role to primary.

In M-LAG standalone mode, the LACPDUs sent out of an M-LAG interface by each M-LAG member device contain the interface-specific LACP system MAC address and LACP system priority.

The Selected state of the member ports in the M-LAG interfaces in an M-LAG group depends on their LACP system MAC address and LACP system priority. If an M-LAG interface has a lower LACP system priority value or LACP system MAC address, the member ports in that M-LAG interface become Selected to forward traffic. If those Selected ports fail, the member ports in the M-LAG interface on the other M-LAG member device become Selected to forward traffic.

NOTE:

An M-LAG member device changes to M-LAG standalone mode only when it detects that both the peer link and the keepalive link are down. It does not change to M-LAG standalone mode when the peer M-LAG member device reboots.

M-LAG failure handling mechanisms

M-LAG interface failure handling mechanism

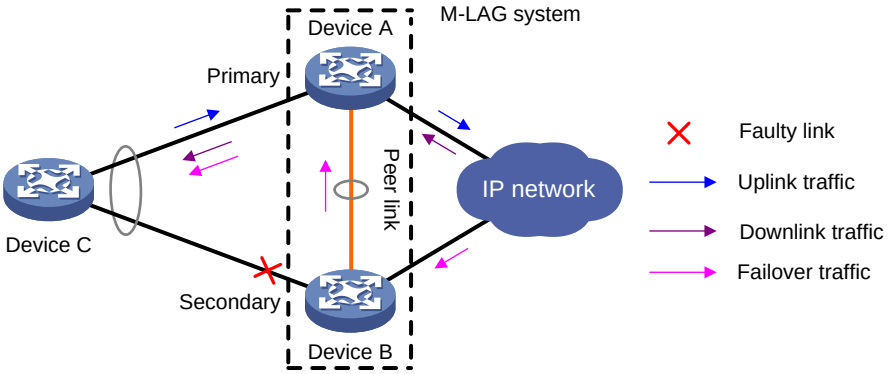
When the M-LAG interface of one M-LAG member device fails, the M-LAG system forwards traffic through the other M-LAG member device.

As shown in Figure 9, Device A and Device B form an M-LAG system, to which Device C is attached through a multichassis aggregation. If traffic to Device C arrives at Device B after the M-LAG interface connected Device B to Device C has failed, the M-LAG system forwards the traffic as follows:

1. Device B sends the traffic to Device A over the peer link.
2. Device A forwards the downlink traffic received from the peer link to Device C.

After the faulty M-LAG interface comes up, Device B forwards traffic to Device C through the M-LAG interface.

Figure 9 M-LAG interface failure handling mechanism



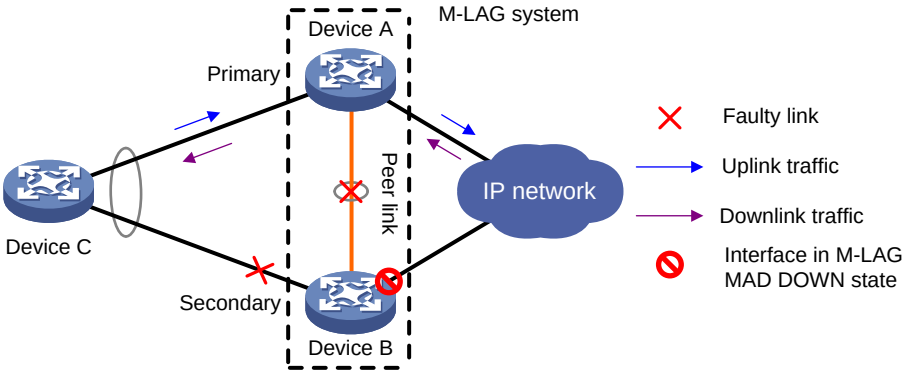
Peer link failure handling mechanism

As shown in Figure 10, multi-active collision occurs if the peer link goes down while the keepalive link is up. To avoid network issues, the secondary M-LAG device sets all network interfaces to M-LAG MAD DOWN state, except for the interfaces excluded from the shutdown action by M-LAG MAD.

In this situation, the primary M-LAG member device forwards all traffic for the M-LAG system.

When the peer-link interface comes up, the secondary M-LAG member device does not bring up the network interfaces immediately. Instead, it starts a delay timer and begins to recover data from the primary M-LAG member device. When the delay timer expires, the secondary M-LAG member device brings up all network interfaces.

Figure 10 Peer link failure handling mechanism

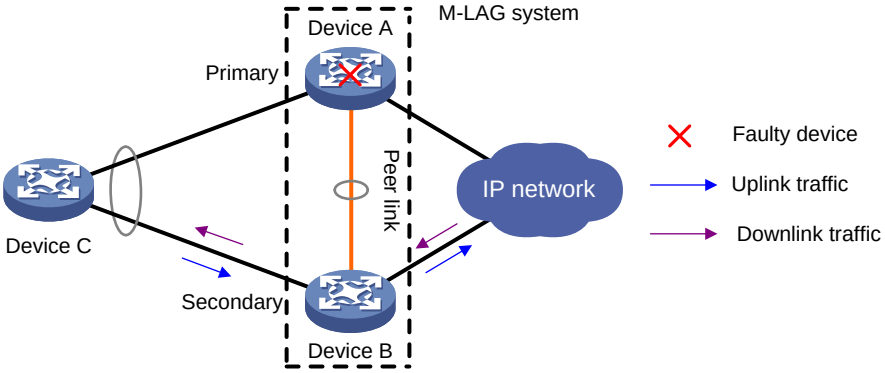


Device failure handling mechanism

As shown in Figure 11, when the primary M-LAG member device fails, the secondary M-LAG member device takes over the primary role to forward all traffic for the M-LAG system. When the faulty device recovers, it becomes the secondary M-LAG member device.

When the secondary M-LAG member device fails, the primary M-LAG member device forwards all traffic for the M-LAG system.

Figure 11 Device failure handling mechanism

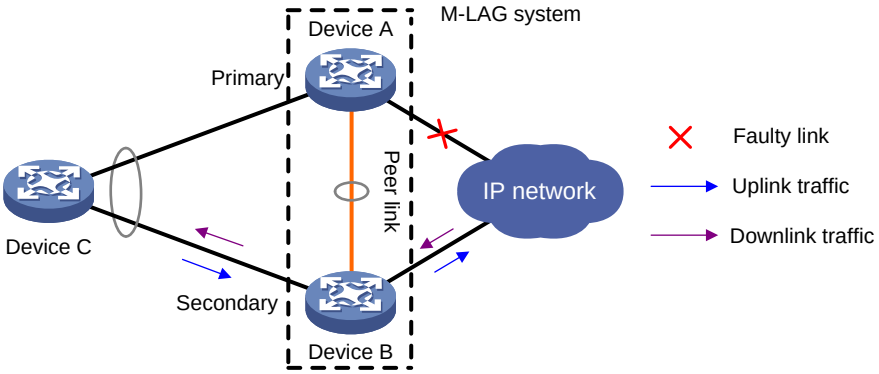


Uplink failure handling mechanism

Uplink failure does not interrupt traffic forwarding of the M-LAG system. As shown in Figure 12, when the uplink of Device A fails, Device A passes traffic destined for the IP network to Device B for forwarding.

To enable faster traffic switchover in response to an uplink failure and minimize traffic losses, configure Monitor Link to associate the M-LAG interfaces with the uplink interfaces. When the uplink interface of an M-LAG member device fails, that device shuts down its M-LAG interface for the other M-LAG member device to forward all traffic of Device C. For more information about Monitor Link, see *High Availability Configuration Guide*.

Figure 12 Uplink failure handling mechanism



Mechanisms to handle concurrent peer link and keepalive link failures

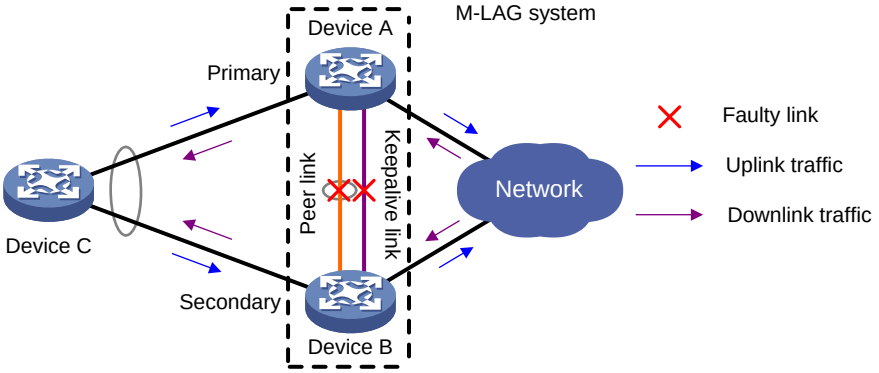
When both the peer link and the keepalive link are down, the M-LAG member devices handle this situation depending on your configuration.

Default failure handling mechanism

Figure 13 shows the default mechanism to handle peer link and keepalive link failures when the M-LAG standalone mode and M-LAG MAD DOWN state persistency features are not configured.

- If the peer link goes down while the keepalive link is up, the M-LAG member devices negotiate their roles over the keepalive link. M-LAG MAD shuts down all network interfaces on the secondary M-LAG member device except those excluded from the shutdown action by M-LAG MAD.
- If the keepalive link goes down while the peer link is down, the secondary M-LAG member device sets its role to primary and brings up the network interfaces in M-LAG MAD DOWN state to forward traffic. In this situation, both of the M-LAG member devices might operate with the primary role to forward traffic. Forwarding errors might occur because the M-LAG member devices cannot synchronize MAC address entries over the peer link.
- If the keepalive link is down before the peer link goes down, M-LAG MAD will not place network interfaces in M-LAG MAD DOWN state. Both M-LAG member devices can operate with the primary role to forward traffic.

Figure 13 Default failure handling mechanism

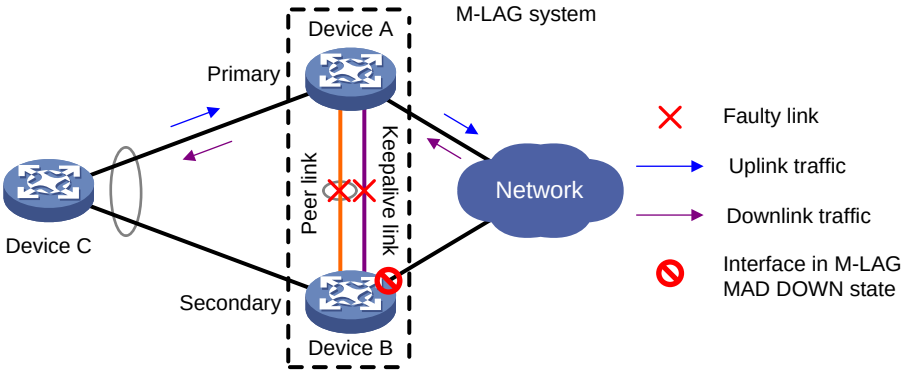


Failure handling mechanism with M-LAG MAD DOWN state persistence

Figure 14 shows the mechanism to handle peer link and keepalive link failures when the M-LAG MAD DOWN state persistence feature is configured.

- If the peer link goes down while the keepalive link is up, the M-LAG member devices negotiate their roles over the keepalive link. M-LAG MAD shuts down all network interfaces on the secondary M-LAG member device except those excluded from the shutdown action by M-LAG MAD.
- If the keepalive link goes down while the peer link is down, the secondary M-LAG member device sets its role to primary, but it does not bring up the network interfaces in M-LAG MAD DOWN state. Only the original primary member device can forward traffic.
- If the keepalive link is down before the peer link goes down, M-LAG MAD will not place network interfaces in M-LAG MAD DOWN state. Both M-LAG member devices can operate with the primary role to forward traffic.

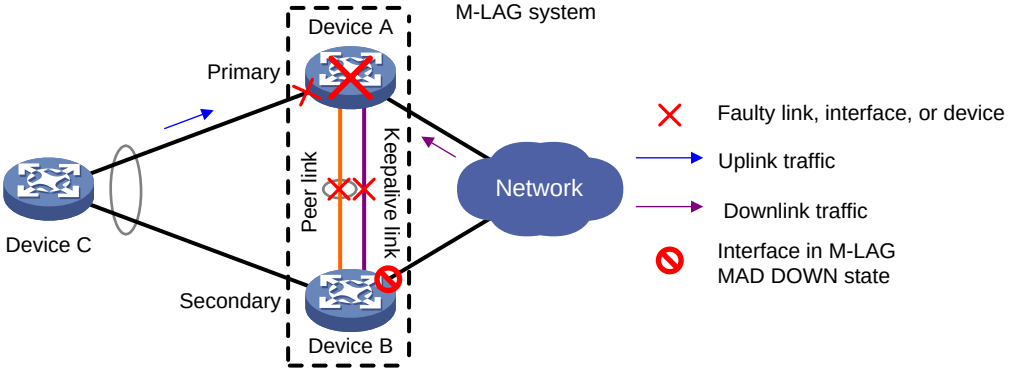
Figure 14 Failure handling mechanism with M-LAG MAD DOWN state persistence



As shown in Figure 15, you can bring up the interfaces in M-LAG MAD DOWN state on the secondary M-LAG member device for it to forward traffic if the following conditions exist:

- Both the peer link and the keepalive link are down.
- The primary M-LAG member device fails or its M-LAG interface fails.

Figure 15 Bringing up the interfaces in M-LAG MAD DOWN state



Failure handling mechanism with M-LAG standalone mode

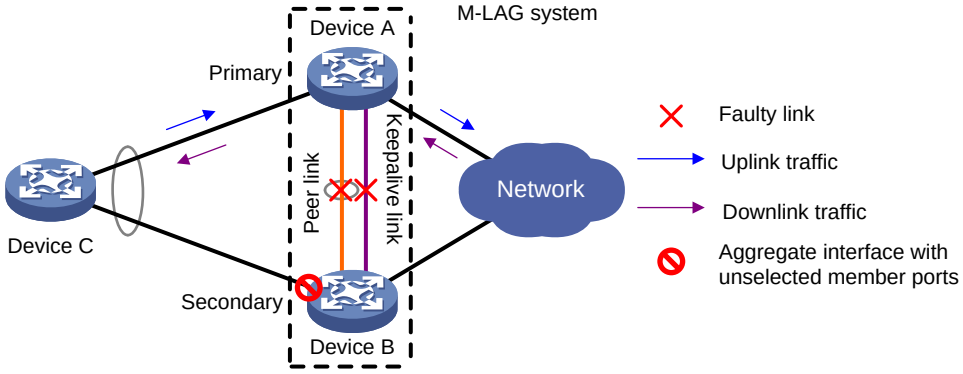
Figure 16 shows the mechanism to handle peer link and keepalive link failures when the M-LAG standalone mode feature is configured.

- If the peer link goes down while the keepalive link is up, the M-LAG member devices negotiate their roles over the keepalive link. M-LAG MAD shuts down all network interfaces on the

secondary M-LAG member device except those excluded from the shutdown action by M-LAG MAD.

- If the keepalive link goes down while the peer link is down, both M-LAG member devices change to M-LAG standalone mode. The secondary M-LAG member device sets its role to primary and brings up its network interfaces in M-LAG MAD DOWN state. In M-LAG standalone mode, only the aggregation member ports on one M-LAG member device can become Selected to forward traffic. For more information about how M-LAG standalone mode operates, see "M-LAG standalone mode."
- If the keepalive link is down before the peer link goes down, both M-LAG member devices change to M-LAG standalone mode.

Figure 16 Failure handling mechanism with M-LAG standalone mode



Extended M-LAG features

Dual-active gateways on an M-LAG system

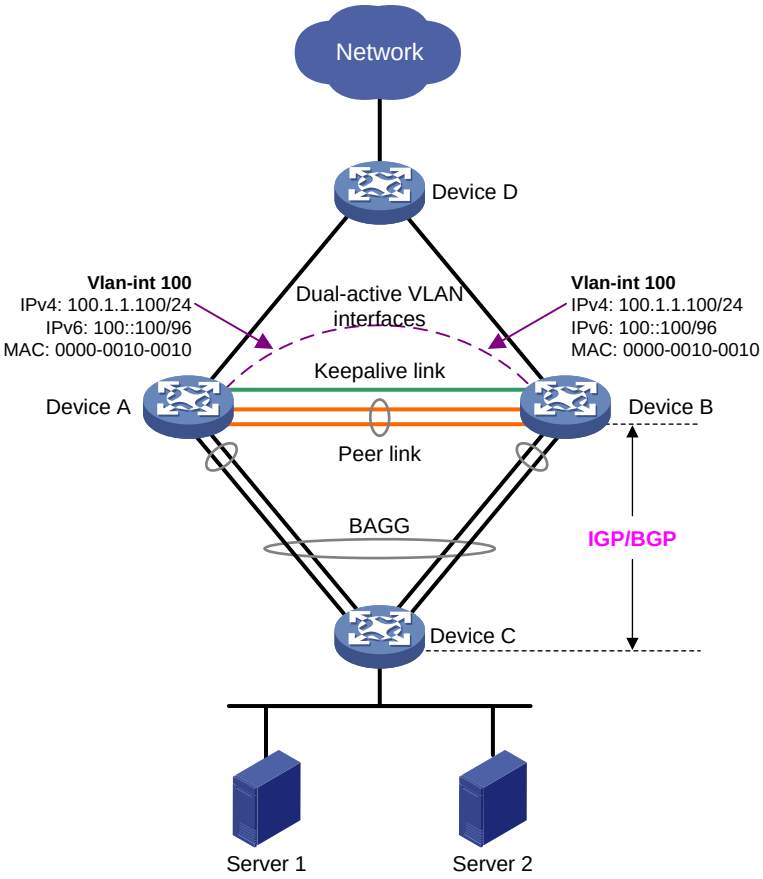
When a Layer 3 network is dualhomed to an M-LAG system, each M-LAG member device uses a logical interface as the gateway interface for the Layer 3 network. Both logical interfaces are active to perform Layer 3 forwarding, and they use the same IP address and MAC address. When both links to the M-LAG system are operating correctly, the links load share traffic to balance bandwidth usage. When one of the links fails, the M-LAG system switches the traffic on the failed link to the other link to ensure high availability.

Typically, dual-active gateways are deployed if downstream devices run a dynamic routing protocol to communicate with the M-LAG system. As shown in Figure 17, IGP or BGP runs between Device C and the M-LAG system. To configure dual-active VLAN interfaces, perform the following tasks on Device A and Device B:

1. Create VLAN interfaces for the same VLAN, VLAN-interface 100 for example, assign the same IPv4, IPv6, and MAC addresses to the VLAN interfaces, and assign the M-LAG interfaces to that VLAN.
2. Assign different M-LAG VIPs to the VLAN interfaces.

Device A and Device B use their M-LAG VIPs to set up IGP or BGP peer relationships with Device C and with each other.

Figure 17 Layer 3 access to an M-LAG system through dynamic routing protocols

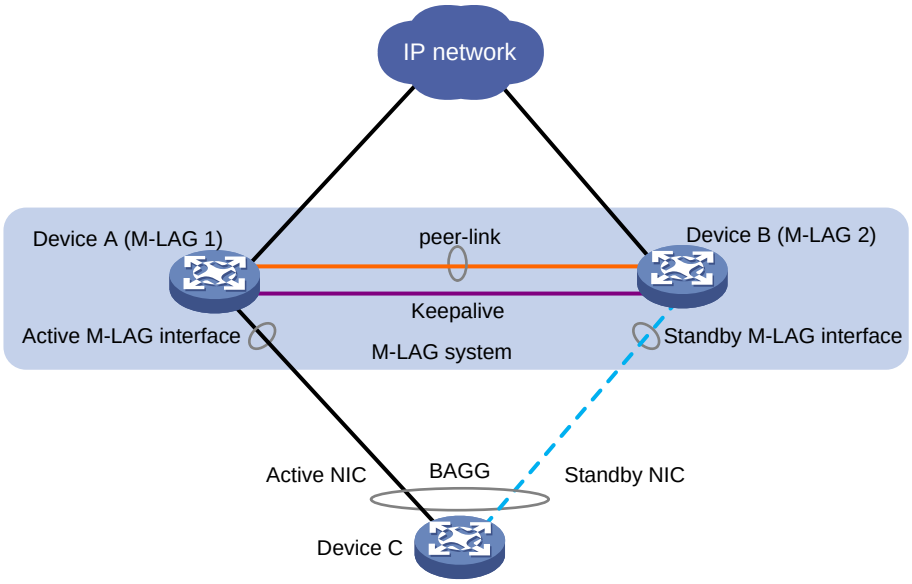


M-LAG active/standby mode

About M-LAG active/standby mode

By default, M-LAG interfaces on two M-LAG member devices forward traffic for load sharing. However, when a device with active and standby NICs connects to the M-LAG system, only the active NIC can send, receive, and process traffic. As a result, the dual-active mode of M-LAG interfaces for load sharing is not applicable. In this scenario, configure the M-LAG interfaces to operate in active/standby mode. Then, only the M-LAG interface connected to the active NIC can forward traffic.

Figure 18 M-LAG active/standby mode network model



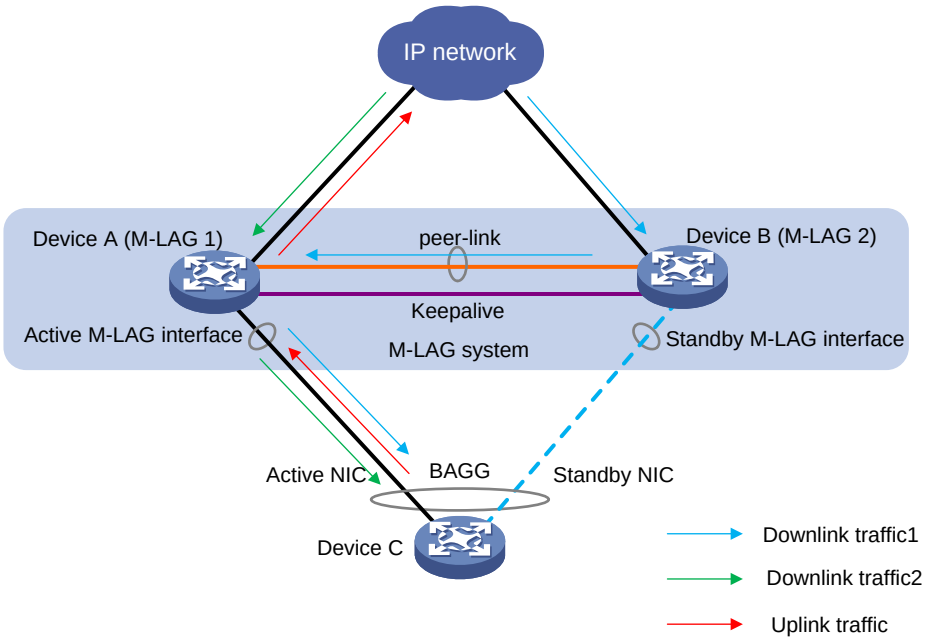
As shown in Figure 18, in active/standby mode, only the M-LAG interface connected to the active NIC sends and receives traffic in normal conditions, and this M-LAG interface is called the active M-LAG interface. The M-LAG interface connected to the standby NIC acts as a backup and does not send or receive traffic, and this M-LAG interface is called the standby M-LAG interface. When the active M-LAG interface fails, the standby M-LAG interface takes over to forward traffic.

In an M-LAG system, the active and standby M-LAG interfaces are elected based on protocol packets sent by the active NIC. The M-LAG interface that receives protocol packets from the active NIC becomes the active M-LAG interface, while the M-LAG interface on the other end becomes the standby M-LAG interface. The roles of the active/standby M-LAG interfaces do not change or restore proactively. If an active/standby NIC switchover occurs on the device connected to M-LAG interfaces, the M-LAG interface that receives the protocol packets for election changes to trigger a re-election. Then, the roles of the active/standby M-LAG interfaces change.

In the current software version, the following types of protocol packets are supported for election: ARP, DHCP, IGMP, MLD, and ND.

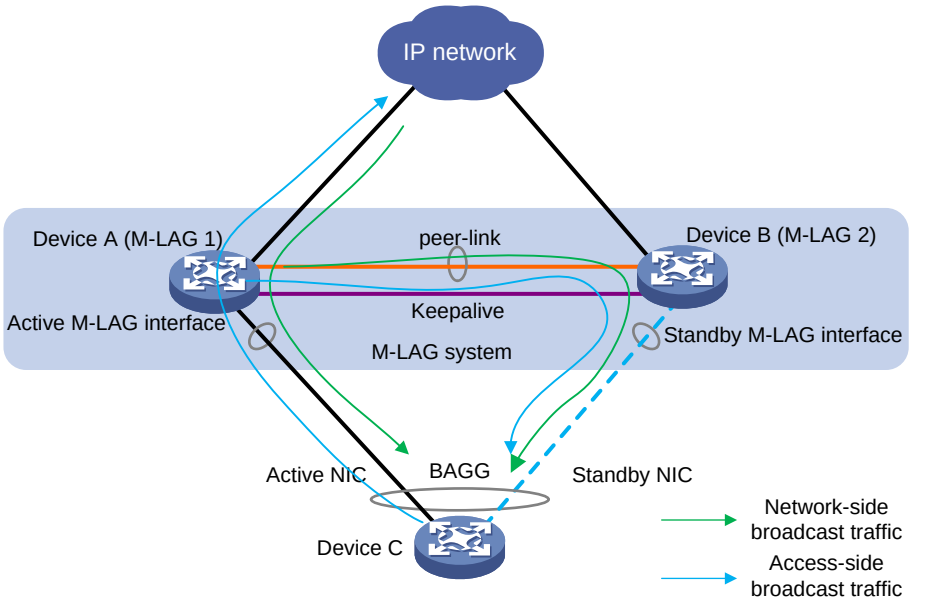
Traffic forwarding mechanism in M-LAG active/standby mode

Figure 19 M-LAG unicast and multicast traffic forwarding mechanism in normal state



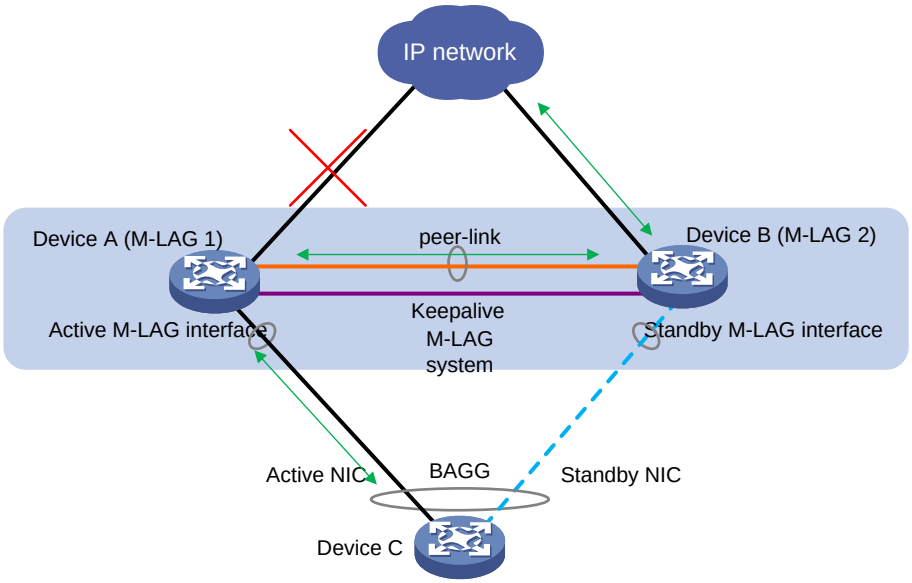
As shown in Figure 19, in normal state, the access device only sends uplink unicast and multicast traffic out the active NIC. After the active M-LAG interface receives the traffic, the M-LAG interface directly forwards the traffic out the uplink interface of the local M-LAG member device. For downlink unicast and multicast traffic, if the network-side traffic is sent to the device with the active M-LAG interface, the device forwards the traffic directly out the active M-LAG interface to the access device. If the network-side traffic is sent to the device with the standby M-LAG interface, the device must reroute the traffic through the peer link to the device with the active M-LAG interface. Then, the device with the active M-LAG interface forwards the traffic out the active M-LAG interface to the access device.

Figure 20 Broadcast traffic forwarding mechanism in normal state



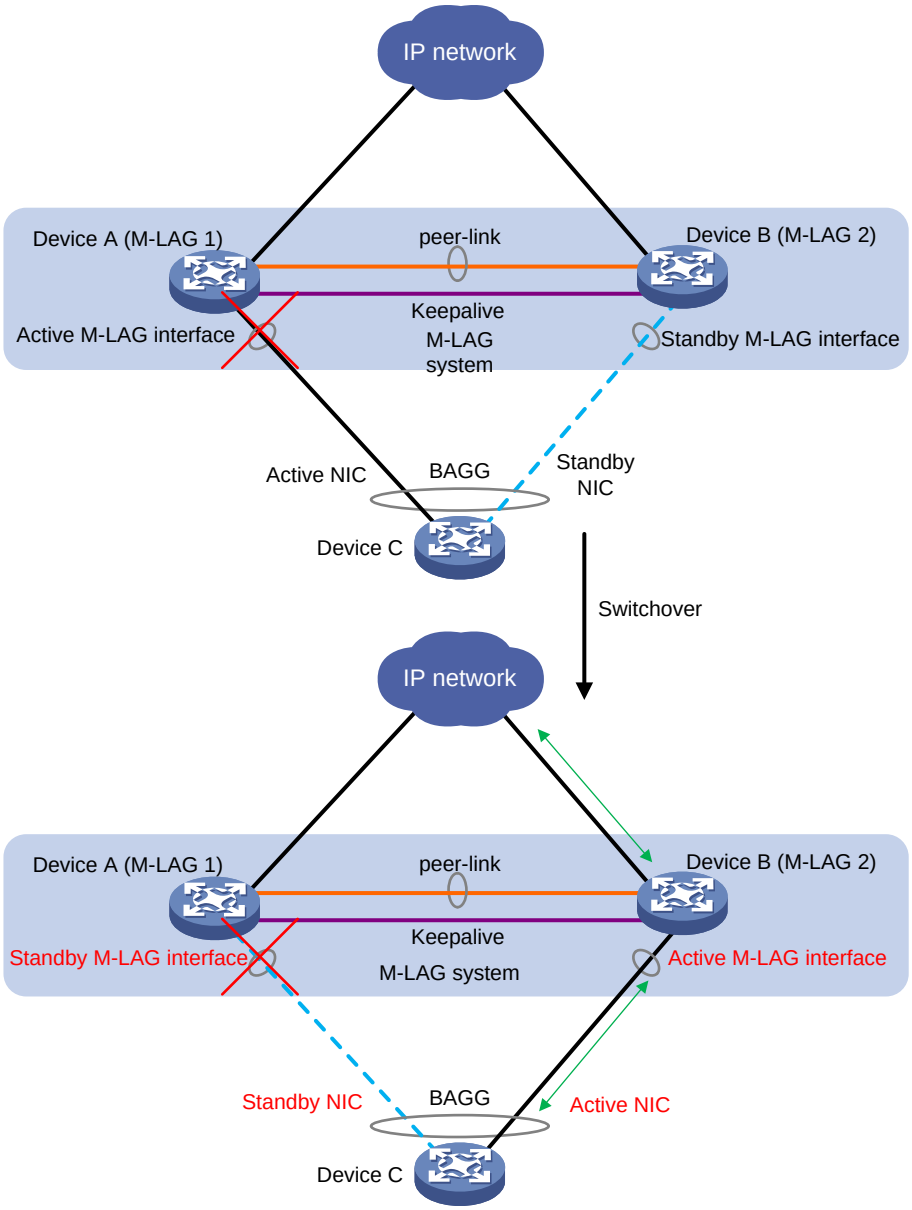
As shown in Figure 20, in normal state, whether the broadcast traffic comes from the network side or the access side, the device with the active M-LAG interface forwards the received broadcast traffic to all interfaces. The peer M-LAG device also forwards the received broadcast traffic to the access device through the M-LAG interface. Therefore, the access device will receive the broadcast traffic that it sends out. However, no loop will occur because the standby NIC does not process traffic.

Figure 21 Traffic forwarding mechanism when the uplink fails but the active M-LAG interface is normal



As shown in Figure 21, when the uplink of the device with the active M-LAG interface fails but the active M-LAG interface is normal, both uplink and downlink traffic need to bypass the device with the standby M-LAG interface for forwarding. The access device still receives traffic through the active NIC.

Figure 22 Traffic forwarding mechanism when the active M-LAG interface is unavailable



As shown in Figure 22, the active M-LAG interface becomes unavailable when a downlink fault or device fault occurs on the device with the active M-LAG interface. When the active NIC of the access device detects the link fault, the access device performs an active/standby NIC switchover. After the original standby M-LAG interface receives protocol packets sent by the new active NIC, it triggers a re-election and becomes the active M-LAG interface. Subsequently, the new active M-LAG interface sends and receives traffic to and from the access device.

M-LAG sequence number check

M-LAG sequence number check protects M-LAG member devices from replay attacks.

With this feature enabled, the M-LAG member devices insert a sequence number into each outgoing DRCPDU or keepalive packet and the sequence number increases by 1 for each sent packet. When receiving a DRCPDU or keepalive packet, the M-LAG member devices check its sequence number and drop the packet if the check result is either of the following:

- The sequence number of the packet is the same as that of a previously received packet.

- The sequence number of the packet is smaller than that of the most recently received packet.

M-LAG packet authentication

M-LAG packet authentication prevents DRCPDU and keepalive packet tampering from causing link flapping.

With this feature enabled, the M-LAG member devices compute a message digest by using an authentication key for each outgoing DRCPDU or keepalive packet and insert the message digest into the packet. When receiving a DRCPDU or keepalive packet, an M-LAG member device computes a message digest and compares it with the message digest in the packet. If the message digests match, the packet passes authentication. If the message digests do not match, the device drops the packet.

Multichassis configuration synchronization

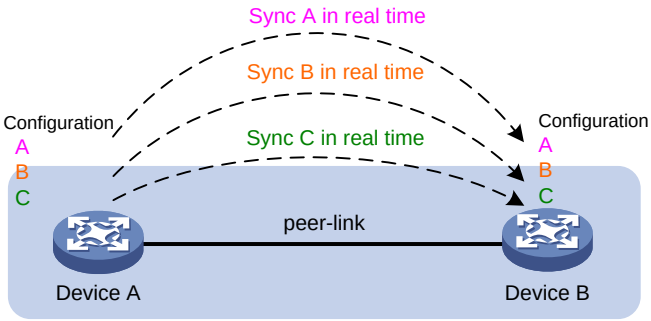
Multichassis configuration synchronization (sync) allows you to synchronize and back up configurations between multiple devices to simplify the configuration procedure, improve deployment efficiency, and optimize user experience. For example, after you configure one M-LAG member device, you can synchronize its configuration to the other M-LAG member device over the peer link for them to have identical configuration.

This feature is applicable only to an M-LAG system currently.

Real-time configuration sync

As shown in Figure 23, settings in multichassis configuration sync view of a member device are synchronized to its peer device in real time.

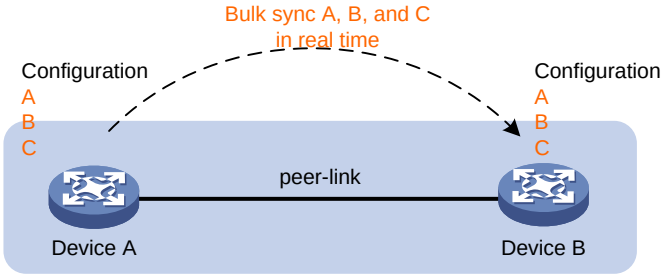
Figure 23 Real-time configuration sync



Bulk configuration sync

To fast deploy the services that are required to be consistent across the M-LAG member devices, configure the services on an M-LAG member device before M-LAG system setup. After the M-LAG system is set up and the peer link comes up, you can bulk synchronize settings on the member device to the same view for multichassis configuration sync on the M-LAG peer, as shown in Figure 24.

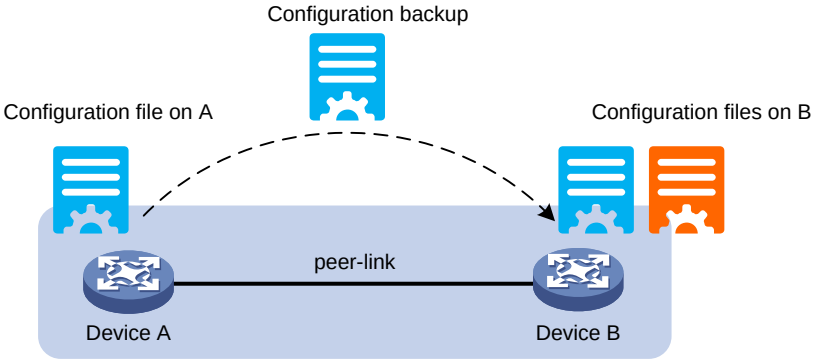
Figure 24 Bulk configuration sync



Configuration backup

If you are to update an M-LAG member device, for example, update the software version or replace the device, back up its configuration file to the M-LAG peer in advance as shown in Figure 25. After the update is completed, you can retrieve the previously saved configuration file from the peer device.

Figure 25 Configuration backup



Protocols and standards

IEEE P802.1AX-REV™/D4.4c, *Draft Standard for Local and Metropolitan Area Networks*

Configuring M-LAG

Restrictions and guidelines: M-LAG configuration

Software version requirements

The M-LAG member devices in an M-LAG system must use the same software version.

During an upgrade, the CLI of the peer M-LAG member device is locked after the local M-LAG member device is upgraded and rebooted. You cannot operate the CLI of the peer M-LAG member device. To unlock the CLI of the peer M-LAG member device, perform a primary/secondary role switchover or restart the M-LAG process of the device.

M-LAG configuration

For the M-LAG member devices to be identified as one M-LAG system, you must configure the same M-LAG system MAC address and M-LAG system priority on them. You must assign different M-LAG system numbers to the M-LAG member devices.

The M-LAG interfaces in the same M-LAG group must use different LACP system MAC addresses.

To ensure correct forwarding, delete M-LAG configuration from an M-LAG member device if it leaves its M-LAG system.

As a best practice to reduce the impact of interface flapping on upper-layer services, use the **link-delay** command to configure the same link delay settings on the peer-link interfaces.

In an M-LAG system, two peer-link interfaces must have the same configuration for the maximum jumbo frame length.

As a best practice, establish a directly connected physical link between the M-LAG member devices as the keepalive link and use this link for the sole purpose of keepalive detection. When the keepalive link uses the IP address of a VLAN interface, do not assign the peer-link interface to the VLAN corresponding to the VLAN-interface. If you do that, the keepalive mechanism might fail.

To configure dual-active gateways for an M-LAG system, first set up the M-LAG system.

Compatibility with other features

An M-LAG member device does not support the following service features. Do not configure these service features on an M-LAG system.

- MVRP.
- Service quality analysis.
- IPsec.
- DNS.

For correct traffic forwarding, make sure the M-LAG member devices are consistent in service feature settings.

GIR

Before you change an M-LAG system back to normal mode by using the **undo gir system-mode maintenance** command, execute the **display m-lag mad verbose** command to verify that no network interfaces are in M-LAG MAD DOWN state. For information about GIR, see *Fundamentals Configuration Guide*.

IRF

M-LAG cannot work correctly on an IRF fabric. Do not configure M-LAG on an IRF fabric. For more information about IRF, see *Virtual Technologies Configuration Guide*.

MAC address table

If the M-LAG system has a large number of MAC address entries, set the MAC aging timer to a higher value than 20 minutes as a best practice. To set the MAC aging timer, use the **mac-address timer** command.

The MAC address learning feature is always disabled on the peer-link Layer 2 aggregate interface.

For more information about the MAC address table, see "Configuring the MAC address table."

Ethernet link aggregation

Do not configure automatic link aggregation on an M-LAG system.

The aggregate interfaces in an S-MLAG group cannot be used as M-LAG interfaces or peer-link interfaces.

You cannot configure link aggregation management subnets on an M-LAG system.

When you configure an M-LAG interface, follow these restrictions and guidelines:

- The **link-aggregation selected-port minimum** command does not take effect on an M-LAG interface.
- If you execute the **display link-aggregation verbose** command for an M-LAG interface, the displayed system ID contains the M-LAG system MAC address and the M-LAG system priority.
- If the reference port is a member port of an M-LAG interface, the **display link-aggregation verbose** command displays the reference port on both M-LAG member devices.

For more information about Ethernet link aggregation, see "Configuring Ethernet link aggregation."

Port isolation

Do not assign peer-link interfaces to a port isolation group. For more information about port isolation, see "Configuring port isolation."

Loop detection

Member devices in an M-LAG system must have the same loop detection configuration. For information about loop detection, see "Configuring loop detection."

Spanning tree

When the spanning tree protocol is enabled for an M-LAG system, follow these restrictions and guidelines:

- Make sure the M-LAG member devices have the same spanning tree configuration. Violation of this rule might cause network flapping. The configuration includes:
  - Global spanning tree configuration.
  - Spanning tree configuration on the peer-link interface.
  - Spanning tree configuration on M-LAG interfaces.
- Peer-link interfaces of the M-LAG system do not participate in spanning tree calculation.
- The M-LAG member devices still use the M-LAG system MAC address after the M-LAG system splits, which will cause spanning tree calculation issues. To avoid the issues, enable M-LAG standalone mode on the M-LAG member devices before the M-LAG system splits.

For more information about spanning tree, see "Configuring spanning tree."

ICMP

An M-LAG system does not support sending ICMPv4 or ICMPv6 redirect messages. These features are configurable with the **ip redirects enable** and **ipv6 redirects enable** commands.

DHCP snooping

Do not use DHCP snooping and QinQ together on an M-LAG system. DHCP snooping cannot operate correctly when QinQ is configured.

AAA

On an M-LAG system, you must specify a virtual IP address as the source IP address of outgoing RADIUS packets.

When RADIUS DAS is enabled, an M-LAG system does not support shutting down or rebooting the access ports for 802.1X authentication users or reauthenticating the users through CoA messages.

For more information about AAA, see *Security Configuration Guide*.

Portal

Before you configure portal authentication to support M-LAG, you must complete the following tasks on both the primary and secondary M-LAG member devices:

- Enable portal authentication.
- Configure IPv4 portal authentication or IPv6 portal authentication.
- Set the portal authentication mode.
- Enable or disable MAC-based quick portal authentication.
- Set the free-traffic threshold.
- Enable portal fail-permit for a portal authentication server.
- Configure the URL of the portal Web server.
- Configure portal user roaming.
- Enable or disable portal authentication to support IPv4/IPv6 dual stack.

Make sure the above configuration on both the primary and secondary M-LAG member devices is the same.

For information about portal configuration, see *Security Configuration Guide*.

Port security

In an M-LAG system, the M-LAG member devices exchange configuration information with each other to detect whether configuration conflicts exist. The detection does not affect packet forwarding on the M-LAG member devices. If configuration conflicts exist, the M-LAG member devices do not allow new users to come online.

For information about port security configuration, see *Security Configuration Guide*.

Mirroring

For a mirroring group, do not assign the source port to an aggregation group other than the one that accommodates the destination port, egress port, or reflector port. If the source port is in a different aggregation group than the other ports, mirrored LACPDUs will be transmitted between the aggregation groups and cause aggregate interface flapping.

Support for DRNI commands

The M-LAG feature was named DRNI in earlier software versions. For compatibility with earlier software versions, the device supports both M-LAG and DRNI commands.

Table 5 Difference between M-LAG and DRNI commands

Feature name	Keywords	Example
M-LAG	**m-lag**, **mlag**, **peer-link**	**m-lag system-number** *system-number* As a best practice, use M-LAG commands.
DRNI	**drni**, **drmac**, **ipp**	**drni system-number** *system-number* To use configuration files created from an earlier software version, use DRNI commands. The system recognizes only the complete syntaxes of DRNI commands. It does not support displaying available keywords and arguments in response to a question mark (?) or automatically completing the last keyword or argument in response to the Tab key. If you execute a DRNI command, the system converts it into the corresponding M-LAG command and saves the M-LAG

		command in the configuration file.

M-LAG and DRNI commands do not differ in the configuration method or functionality. For more information about the keyword differences, see "Comparison between M-LAG and DRNI commands."

M-LAG tasks at a glance

To configure M-LAG, perform the following tasks:

1. Configuring multichassis configuration sync
2. Configuring M-LAG system settings
   - Configuring the M-LAG system MAC address
   - Setting the M-LAG system number
   - Setting the M-LAG system priority
3. Setting the M-LAG role priority of the device
4. (Optional.) Enabling M-LAG standalone mode on an M-LAG member device
5. Configuring M-LAG keepalive settings
   - Configuring M-LAG keepalive packet parameters
   - Setting the M-LAG keepalive interval and timeout timer
   - Associating the keepalive link with a track entry
6. Configuring M-LAG MAD
   - Configuring the default M-LAG MAD action on network interfaces
   - Excluding an interface from the shutdown action by M-LAG MAD
   - Excluding all logical interfaces from the shutdown action by M-LAG MAD
   - Specifying interfaces to be shut down by M-LAG MAD when the M-LAG system splits
   - Enabling M-LAG MAD DOWN state persistence
7. Configuring interfaces on the M-LAG system
   - Configuring an M-LAG interface
   - Configuring an M-LAG uplink interface
   - Configuring the flapping detection interval for M-LAG uplink interfaces
   - Setting up a peer link
   - (Optional.) Enabling the short DRCP timeout timer on the peer-link interface or an M-LAG interface
   - (Optional.) Configuring the M-LAG active/standby mode
8. (Optional.) Enabling the peer-link interface to retain MAC address entries for down single-homed devices
9. (Optional.) Assigning a M-LAG virtual IP address to an interface

   Configure M-LAG VIPs for devices to communicate with the M-LAG system by using dynamic routing protocols.
10. (Optional.) Configuring M-LAG extra VLANs
11. (Optional.) Configuring configuration consistency check
    - Setting the mode of configuration consistency check
    - (Optional.) Disabling configuration consistency check

Configuration consistency check might fail when you upgrade the M-LAG member devices in an M-LAG system. To prevent the M-LAG system from falsely shutting down M-LAG interfaces, temporarily disable configuration consistency check.

12. Configuring M-LAG timers
    - (Optional.) Setting the keepalive hold timer for identifying the cause of peer link down events
    - Configuring M-LAG system auto-recovery
    - (Optional.) Setting the data restoration interval
13. (Optional.) Configuring M-LAG security features
    - Enabling M-LAG sequence number check
    - Enabling M-LAG packet authentication

Configuring multichassis configuration sync

Restrictions and guidelines for multichassis configuration sync

Real-time or bulk configuration sync synchronizes only configurations required to be consistent across the M-LAG member devices.

Execute the **mc-configure access-local enable** command on both M-LAG member devices for them to access each another.

Before you use the multichassis configuration sync feature, verify that the M-LAG system is operating correctly and the peer link is up.

For more information about the multichassis configuration sync commands, see *Fundamentals Command Reference*.

Synchronizing the configurations in real time

About this task

This feature enables commands executed in multichassis configuration sync view on an M-LAG member device to be synchronized to its M-LAG peer in real time.

Procedure

1. Enter system view.
   system-view
2. Permit the M-LAG peer to access the local device via the sync channel.
   mc-configure access-local enable
   By default, the M-LAG peer is permitted to access the local device via the sync channel.
3. Return to user view.
   quit
4. Enter multichassis configuration sync view.
   system-view mc-config
   Commands executed in any view for multichassis configuration sync will be synchronized to the M-LAG peer in real time.

Bulk synchronizing the configurations

About this task

To fast deploy the services that are required to be consistent across the M-LAG member devices, configure the services on an M-LAG member device before M-LAG system setup. After the M-LAG system is set up and the peer link comes up, you can use this feature to bulk synchronize settings on the member device to the same view for multichassis configuration sync on the M-LAG peer.

Procedure

1. Enter system view.
   `system-view`
2. Permit the M-LAG peer to access the local device via the sync channel.
   `mc-configure access-local enable`
   By default, the M-LAG peer is permitted to access the local device via the sync channel.
3. Configure required settings in system view or any view under system view.
4. Return to user view.
   `quit`
5. Enter multichassis configuration sync view.
   `system-view mc-config`
6. Bulk synchronize the settings to the M-LAG peer.
   `mc-configure batch-synchronize`
   You can execute this command in any view for multichassis configuration sync.

Fast configuring differential settings

About this task

For settings required only on the local device, you can execute the **`mc-configure switch-to-local`** command in any view for multichassis configuration sync to switch to the same view of the local device and then configure the settings. To return to the original view for multichassis configuration sync after configuring the settings on the local device, execute the **`mc-configure quit`** command or press **Ctrl+Q**.

For the settings required only on the peer device, you can execute the **`mc-configure switch-to-peer`** in any view for multichassis configuration sync of the local device to switch to the same view of the peer device and then configure the settings. To return to the original view for multichassis configuration sync after configuring the settings on the peer device, execute the **`mc-configure quit`** command or press **Ctrl+Q**.

Configuring the differential settings on the local device

1. Enter multichassis configuration sync view.
   `system-view mc-config`
2. Switch to the same view of the local device.
   [@local] will be added to the commands executed in any view of the local device.
   `mc-configure switch-to-local`
   You can execute this command in any view for multichassis configuration sync.
3. Configure the differential commands on the local device.
4. Return to the original view for multichassis configuration sync.
   `mc-configure quit`

Configuring the differential settings on the M-LAG peer

1. Enter multichassis configuration sync view.
   system-view mc-config
2. Switch to the same view of the M-LAG peer.
   [@peer] will be added to the commands executed in any view of the peer device.
   mc-configure switch-to-local
   You can execute this command in any view for multichassis configuration sync.
3. Configure the differential commands on the M-LAG peer.
4. Return to the original view for multichassis configuration sync.
   mc-configure quit

Backing up and restoring a configuration file

About this task

To update an M-LAG member device, for example, update the software version or replace the device, you can the **mc-configure backup-to-peer** command to back up its configuration file to the M-LAG peer in advance. After the update is completed, you can use the **mc-configure get-from-peer** command to retrieve the previously saved configuration file from the peer device and replace the configuration file on the local device.

Procedure

Execute the following command in user view to save the local configuration file to the M-LAG peer.

mc-configure backup-to-peer *local-filename*

Execute the following command in user view to retrieve the configuration file from the M-LAG peer.

mc-configure get-from-peer *filename*

Configuring M-LAG system settings

Configuring the M-LAG system MAC address

Restrictions and guidelines

On an M-LAG system, M-LAG interfaces in the same M-LAG group must use the same LACP system MAC address. As a best practice, use the bridge MAC address of one M-LAG member device as the M-LAG system MAC address.

Changing the M-LAG system MAC address causes M-LAG system split. When you perform this task on a live network, make sure you are fully aware of its impact.

You can configure the M-LAG system MAC address on an aggregate interface only after it is configured as an M-LAG interface.

You can configure the M-LAG system MAC address globally and in aggregate interface view. The global M-LAG system MAC address takes effect on all aggregation groups. On an aggregate interface, the interface-specific M-LAG system MAC address takes precedence over the global M-LAG system MAC address.

Procedure

1. Enter system view.
   system-view
2. Configure the M-LAG system MAC address.

m-lag system-mac { *bridge-address* | *mac-address* }

By default, the M-LAG system MAC address is not configured.

3. Enter Layer 2 aggregate interface view.

   interface bridge-aggregation *interface-number*

4. Set the M-LAG system MAC address on the aggregate interface.

   port m-lag system-mac *mac-address*

   By default, the M-LAG system MAC address is not configured.

Setting the M-LAG system number

Restrictions and guidelines

Changing the M-LAG system number causes M-LAG system split. When you perform this task on a live network, make sure you are fully aware of its impact.

You must assign different M-LAG system numbers to the M-LAG member devices in an M-LAG system.

Procedure

1. Enter system view.

   system-view

2. Set the M-LAG system number.

   m-lag system-number *system-number*

   By default, the M-LAG system number is not set.

Setting the M-LAG system priority

About this task

An M-LAG system uses its M-LAG system priority as the system LACP priority to communicate with the remote aggregation system.

Restrictions and guidelines

Changing the M-LAG system priority in system view causes M-LAG system split. When you perform this task on a live network, make sure you are fully aware of its impact.

You must configure the same M-LAG system priority for the M-LAG interfaces in the same M-LAG group.

You can configure the M-LAG system priority on an aggregate interface only after it is configured as an M-LAG interface.

You can configure the M-LAG system priority globally and in aggregate interface view. The global M-LAG system priority takes effect on all aggregation groups. On an aggregate interface, the interface-specific M-LAG system priority takes precedence over the global M-LAG system priority.

Procedure

1. Enter system view.

   system-view

2. Set the M-LAG system priority.

   m-lag system-priority *system-priority*

   By default, the M-LAG system priority is 32768.

3. Enter Layer 2 aggregate interface view.

   interface bridge-aggregation *interface-number*

4. Set the M-LAG system priority on the aggregate interface.

   port m-lag system-priority *priority*

   By default, the M-LAG system priority is 32768.

Setting the M-LAG role priority of the device

About this task

M-LAG assigns the primary or secondary role to an M-LAG member device based on its M-LAG role priority. The smaller the priority value, the higher the priority. If the M-LAG member devices in an M-LAG system use the same M-LAG role priority, the device with the lower bridge MAC address is assigned the primary role.

Restrictions and guidelines

To prevent a primary/secondary role switchover from causing network flapping, avoid changing the M-LAG priority assignment after the M-LAG system is established.

Procedure

1. Enter system view.

   system-view

2. Set the M-LAG role priority of the device.

   m-lag role priority *priority-value*

   By default, the M-LAG role priority of the device is 32768.

Enabling M-LAG standalone mode on an M-LAG member device

About this task

Perform this task to avoid forwarding issues in the multi-active situation that might occur after both the peer link and the keepalive link are down.

M-LAG standalone mode helps avoid traffic forwarding issues in this multi-active situation by allowing only the member ports in the M-LAG interfaces on one member device to forward traffic. For more information about this mode, see "M-LAG standalone mode."

When you configure this feature, you can configure a delay to prevent an unnecessary mode change because of transient link down issues.

Restrictions and guidelines

An M-LAG member device changes to M-LAG standalone mode only when it detects that both the peer link and the keepalive link are down. It does not change to M-LAG standalone mode when the peer M-LAG member device reboots. As a best practice, set the delay that the device must wait before changing to M-LAG standalone mode based on the cause of simultaneous failure of the peer link and keepalive link:

- If the cause is device poweroff, set the delay to be longer than the time required to finish a device reboot.
- For any other causes, reduce the delay to speed up transition to M-LAG standalone mode.

As a best practice, enable M-LAG standalone mode on both M-LAG member devices.

If the MAC address of an M-LAG system does not correspond to any actual MAC address on any M-LAG member device, and the M-LAG interface contains multiple aggregation member ports, enabling the M-LAG standalone mode might cause the aggregation member ports of the M-LAG

interface to flap. Make sure you are fully aware of the impact of this feature when you use it on a live network.

In a single-level M-LAG network, before you enable M-LAG standalone mode on an M-LAG member device, make sure its LACP system priority is higher than that of the remote aggregation system. This restriction ensures that the reference port is on the remote aggregation system and prevents the interfaces attached to the M-LAG system from flapping.

Procedure

1. Enter system view.

   system-view

2. Enable M-LAG standalone mode.

   m-lag standalone enable [**delay** *delay-time*]

   By default, M-LAG standalone mode is disabled.

Configuring M-LAG keepalive settings

Restrictions and guidelines for configuring M-LAG keepalive settings

As a best practice, establish a direct physical link between the M-LAG member devices as the keepalive link and use that link only for keepalive detection purposes. Make sure the link has both Layer 2 and Layer 3 connectivity.

Make sure the two ends use the same keepalive settings. M-LAG member devices check the peer keepalive settings for consistency. If an inconsistency is found, the device will prompt for configuration revision.

Configuring M-LAG keepalive packet parameters

About this task

Perform this task to specify the parameters for sending M-LAG keepalive packets, such as its source and destination IP addresses.

The device accepts only keepalive packets that are sourced from the specified destination IP address. The keepalive link goes down if the device receives keepalive packets sourced from any other IP address. To improve reliability, use both the **m-lag keepalive** { **ip** | **ipv6** } and **m-lag keepalive standby** { **ip** | **ipv6** } commands to configure the primary and standby keepalive links.

In normal conditions, the primary keepalive link is used to transmit keepalive packets. When the primary keepalive link does not receive keepalive packets within a certain time, the standby link is used to transmit keepalive packets.

If the primary and standby keepalive links both fail to receive keepalive packets within a certain time, they are both considered DOWN, and the system detects a keepalive link fault.

Restrictions and guidelines

Make sure the source and destination IP addresses of M-LAG keepalive packets have Layer 3 connectivity.

If you do not specify the **source** keyword, the keepalive interface might fail to be excluded from the shutdown action by M-LAG MAD. For the keepalive interface to be excluded from the shutdown action by M-LAG MAD, you must specify the **source** keyword.

Make sure the M-LAG member devices in an M-LAG system use the same keepalive destination UDP port.

The standby keepalive link uses the same UDP port as the primary link, so you do not need to configure the UDP port.

Procedure

1. Enter system view.

   system-view

2. Configure the primary keepalive link.

   m-lag keepalive { **ip** | **ipv6** } **destination** { *ipv4-address* | *ipv6-address* } [**source** { *ipv4-address* | *ipv6-address* } | **udp-port** *udp-number* | **vpn-instance** *vpn-instance-name*] *

   By default, no primary keepalive link is configured. If you do not specify a source IP address or destination UDP port when you execute this command, the IP address of the outgoing interface and UDP port 6400 are used, respectively.

3. (Optional.) Configure the standby keepalive link.

   m-lag keepalive standby { **ip** | **ipv6** } **destination** { *ipv4-address* | *ipv6-address* } [**source** { *ipv4-address* | *ipv6-address* } | **vpn-instance** *vpn-instance-name*] *

   By default, no standby keepalive link is configured for M-LAG.

Setting the M-LAG keepalive interval and timeout timer

About this task

The device sends keepalive packets at the specified interval to its M-LAG peer. If the device has not received a keepalive packet from the M-LAG peer before the keepalive timeout timer expires, the device determines that the keepalive link is down.

Restrictions and guidelines

The local M-LAG keepalive timeout timer must be two times the M-LAG keepalive interval of the peer at minimum.

Configure the same M-LAG keepalive interval on the M-LAG member devices in the M-LAG system.

Procedure

1. Enter system view.

   system-view

2. Set the M-LAG keepalive interval and timeout timer.

   m-lag keepalive interval *interval* [**timeout** *timeout*]

   By default, the M-LAG keepalive interval is 1000 milliseconds, and the M-LAG keepalive timeout timer is 5 seconds.

Associating the keepalive link with a track entry

About this task

Associate a track entry with the keepalive link to fast detect state changes of the keepalive link. For more information about track entries, see Track configuration in *High Availability Configuration Guide*.

Restrictions and guidelines

As a best practice, do not specify the track entry associated with M-LAG MAD when executing **m-lag keepalive track** command. If you do that, the M-LAG system state might flap.

Procedure

1. Enter system view.
   system-view
2. Associate the keepalive link with a track entry.
   m-lag keepalive track *track-entry-number*
   By default, no track entry is associated with the keepalive link.

Configuring M-LAG MAD

About this task

M-LAG MAD configuration methods

When you configure M-LAG MAD, use either of the following methods:

- To shut down all network interfaces on the secondary M-LAG member device except a few special-purpose interfaces that must be retained in up state:
  - Set the default M-LAG MAD action to M-LAG MAD DOWN. For more information, see "Configuring the default M-LAG MAD action on network interfaces."
  - Exclude interfaces from being shut down by M-LAG MAD. For more information, see "Excluding an interface from the shutdown action by M-LAG MAD."

  This method is applicable to most network environments.
- To have the secondary M-LAG member device retain a large number of interfaces in up state and shut down the remaining interfaces:
  - Set the default M-LAG MAD action to NONE. For more information, see "Configuring the default M-LAG MAD action on network interfaces."
  - Specify network interfaces that must be shut down by M-LAG MAD. For more information, see "Specifying interfaces to be shut down by M-LAG MAD when the M-LAG system splits."

  Use this method if you must retain a large number of logical interfaces (for example, VLAN, aggregate, loopback, and tunnel interfaces) in up state.

List of automatically included interfaces

M-LAG MAD will always shut down the ports in the system-configured included port list if the device acts as the secondary M-LAG member device when the M-LAG system splits.

This list contains aggregation member ports of M-LAG interfaces. To identify system-configured included ports, execute the **display m-lag mad verbose** command.

List of automatically excluded interfaces

M-LAG MAD will not shut down the ports in the following list when the M-LAG system splits:

- System-configured excluded port list in M-LAG MAD:
  - Peer-link interface.
  - Aggregation member interfaces if a Layer 2 aggregate interface is used as the peer-link interface.
  - M-LAG interfaces.
  - Keepalive interfaces, interfaces used for setting up the keepalive link.
  - VLAN interfaces.

- VSI interfaces.
- Management interfaces.

  To identify these interfaces, execute the **`display m-lag mad verbose`** command.

- Network interfaces used for special purposes, including:
  - Interfaces placed in a loopback test by using the **`loopback`** command. For more information about this command, see Ethernet interface commands in *Layer 2—LAN Switching Command Reference*.
  - Interfaces in a mirroring group.
  - Interfaces forced to stay up by using the **`port-up mode`** command. For more information about this command, see Ethernet interface commands in *Layer 2—LAN Switching Command Reference.*

Restrictions and guidelines for M-LAG MAD

When the M-LAG system splits, M-LAG MAD takes the same action on an aggregate interface and its member ports. For example, if M-LAG MAD shuts down an aggregate interface, it also shuts down the member ports of the aggregate interface.

If you specify an aggregation member port in the **`m-lag mad include interface`** or **`m-lag mad exclude interface`** command, the configuration takes precedence over the M-LAG MAD action on the aggregate interface.

Configuring the default M-LAG MAD action on network interfaces

About this task

You can configure M-LAG MAD to take either of the following default actions on network interfaces if the device acts as the secondary M-LAG member device when the M-LAG system splits:

- **M-LAG MAD DOWN**—M-LAG MAD will shut down all network interfaces on the secondary M-LAG member device when the M-LAG system splits, except the interfaces excluded manually or by the system.
- **NONE**—M-LAG MAD will not shut down any network interfaces when the M-LAG system splits, except the interfaces configured manually or by the system to be shut down by M-LAG MAD.

Restrictions and guidelines

The M-LAG MAD DOWN action will not take effect on the interfaces listed in "List of automatically excluded interfaces."

The M-LAG MAD DOWN action will always take on the interfaces listed in "List of automatically included interfaces," even if the default M-LAG MAD action is NONE.

Procedure

1. Enter system view.

   `system-view`

2. Configure the default M-LAG MAD action to take on network interfaces on the secondary M-LAG member device when the M-LAG system splits.

   `m-lag mad default-action { down | none }`

   By default, M-LAG MAD shuts down network interfaces on the secondary M-LAG member device.

Excluding an interface from the shutdown action by M-LAG MAD

About this task

By default, M-LAG MAD automatically excludes the interfaces listed in "List of automatically excluded interfaces" when it shuts down network interfaces on the secondary M-LAG member device.

To specify additional interfaces that cannot be shut down, perform this task.

You typically perform this task when the default M-LAG MAD action is set to M-LAG MAD DOWN.

Restrictions and guidelines

The M-LAG MAD DOWN action is always taken on interfaces listed in "List of automatically included interfaces." You cannot disable the action by excluding those interfaces.

To view interfaces excluded from the MAD shutdown action, see the **Excluded ports (user-configured)** field in the output from the **`display m-lag mad verbose`** command.

If you exclude an interface that is already in M-LAG MAD DOWN state from the MAD shutdown action, the interface stays in that state. It will not come up automatically.

Peer-link interface, member ports in the Layer 2 aggregation group of the peer-link interface, M-LAG interfaces, management Ethernet interface, keepalive interfaces, VLAN interfaces, and VSI interfaces are excluded by the system from the MAD shutdown action.

When an M-LAG member device reboots, its keepalive interface might remain in the MAD DOWN state after recovery. This prevents the M-LAG system from being established correctly. Resolve this issue in either of the following methods:

- Use the **`m-lag mad exclude interface`** command to manually exclude the keepalive interface from the shutdown action by M-LAG MAD.
- Specify the **`source`** keyword when you use the **`m-lag keepalive`** { **`ip`** | **`ipv6`** } or **`m-lag keepalive standby`** { **`ip`** | **`ipv6`** } command to configure the M-LAG keepalive link.

As a best practice, when you exclude a logical interface from the shutdown action by M-LAG MAD, also exclude its bound physical interfaces from the shutdown action.

Procedure

1. Enter system view.

   `system-view`
2. Exclude an interface from the shutdown action by M-LAG MAD.

   `m-lag mad exclude interface` *`interface-type interface-number`*

   By default, M-LAG MAD shuts down all network interfaces when detecting a multi-active collision, except for the network interfaces set by the system to not shut down.

Excluding all logical interfaces from the shutdown action by M-LAG MAD

About this task

To retain a large number of logical interfaces (for example, aggregate, loopback, and tunnel interfaces, and aggregation member ports) in up state, you can exclude all logical interfaces from the shutdown action by M-LAG MAD.

Peer-link interface, member ports in the Layer 2 aggregation group of the peer-link interface, M-LAG interfaces, management Ethernet interface, keepalive interfaces, and VLAN interfaces are excluded by the system from the MAD shutdown action.

Peer-link interface, member ports in the Layer 2 aggregation group of the peer-link interface, M-LAG interfaces, management Ethernet interface, keepalive interfaces, VLAN interfaces, and VSI interfaces are excluded by the system from the MAD shutdown action.

Restrictions and guidelines

The **m-lag mad exclude interface** and **m-lag mad include interface** commands take precedence over the **m-lag mad exclude logical-interfaces** command.

Procedure

1. Enter system view.

   system-view

2. Exclude all logical interfaces from the shutdown action by M-LAG MAD.

   m-lag mad exclude logical-interfaces

   By default, M-LAG MAD shuts down all network interfaces when it detects a multi-active collision, except for the network interfaces set by the system to not shut down.

Specifying interfaces to be shut down by M-LAG MAD when the M-LAG system splits

About this task

By default, M-LAG MAD automatically shuts down the interfaces listed in "List of automatically included interfaces" if the device is the secondary M-LAG member device when the M-LAG system splits.

To specify additional interfaces to be shut down by M-LAG MAD, perform this task.

You typically perform this task when the default M-LAG MAD action is set to NONE.

Restrictions and guidelines

The M-LAG MAD DOWN action will not take effect on the interfaces listed in "List of automatically excluded interfaces."

Procedure

1. Enter system view.

   system-view

2. Specify interfaces to be shut down by M-LAG MAD when the M-LAG system splits.

   m-lag mad include interface *interface-type interface-number*

   By default, the user-configured included port list does not contain any ports.

Enabling M-LAG MAD DOWN state persistence

About this task

M-LAG MAD DOWN state persistence helps avoid the multi-active situation by preventing the secondary M-LAG member device from bringing up the network interfaces in M-LAG MAD DOWN state. For more information about this feature, see "M-LAG MAD DOWN state persistence" and "Failure handling mechanism with M-LAG MAD DOWN state persistence."

You can bring up the interfaces in M-LAG MAD DOWN state on the secondary M-LAG member device for it to forward traffic if the following conditions exist:

- The primary M-LAG member device fails while the peer link is down.
- The M-LAG MAD DOWN state persists on the secondary M-LAG member device.

To prevent packet loss after the peer link recovers from a link fault, the secondary member device should try to synchronize entries (such as the MAC address entries and ARP entries) within the data restoration interval. Subsequently, interfaces in M-LAG MAD DOWN state on the device will restore to up.

Procedure

1. Enter system view.

   system-view
2. Enable M-LAG MAD DOWN state persistence.

   m-lag mad persistent

   By default, the secondary M-LAG member device brings up interfaces in M-LAG MAD DOWN state when its role changes to primary.
3. (Optional.) Bring up the interfaces in M-LAG MAD DOWN state.

   m-lag mad restore

   Execute this command only when both the peer link and the keepalive link are down.

Configuring an M-LAG interface

About this task

If an M-LAG group contains only one M-LAG interface, that interface is called a single-homed M-LAG interface. By default, M-LAG does not allow access through single-homed M-LAG interfaces, which means M-LAG MAD shuts down an M-LAG interface if it is the only member in its M-LAG group.

To ensure traffic forwarding for a device single-homed to an M-LAG interface, allow the M-LAG interface to be the single member in its M-LAG group. M-LAG MAD will not shut down the single-homed M-LAG interface, and the device will not perform configuration consistency check on the interface.

Restrictions and guidelines

Specify the **allow-single-member** keyword only when the network is managed by a controller. For example, Device A and Device B form an M-LAG system, and they separately set up OpenFlow channels with the controller. If the OpenFlow channel of Device A is disconnected, the controller will notify Device B of this event. Then, Device B will permit access of single-homed peer devices on the M-LAG interfaces with the **allow-single-member** keyword configured. Those M-LAG interfaces will not be shut down by M-LAG MAD.

The device can have multiple M-LAG interfaces. However, you can assign a Layer 2 aggregate interface to only one M-LAG group.

A Layer 2 aggregate interface cannot operate as both peer-link interface and M-LAG interface.

To improve forwarding efficiency, exclude the M-LAG interface on the secondary M-LAG member device from the shutdown action by M-LAG MAD. This action enables the M-LAG interface to forward traffic immediately after a multi-active collision is removed without having to wait for the secondary M-LAG member device to complete entry restoration.

If you execute the **port m-lag group** command multiple times, the most recent configuration takes effect.

To prevent loops when you assign a single-homed aggregate interface to an M-LAG group, use the following procedure:

1. Assign the aggregate interface to the M-LAG group.
2. Assign ports to the aggregation group of the aggregate interface.

When you remove a single-homed M-LAG interface from its M-LAG group, use the following procedure:

3. Remove the member ports from the aggregation group of the M-LAG interface.
4. Remove the M-LAG interface from the M-LAG group.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 aggregate interface view.
   interface bridge-aggregation *interface-number*
3. Assign the aggregate interface to an M-LAG group.
   port m-lag group *group-id* [**allow-single-member**]
   As a best practice, specify the **allow-single-member** keyword for a dynamic aggregate interface.

Configuring an M-LAG uplink interface

About this task

Use this feature to configure the interface that connects the M-LAG device to the uplink network and determine the role election for the M-LAG system based on the interface state.

When the peer link of an M-LAG system is down but can still receive keepalive packets, the M-LAG member device whose uplink interface comes up first becomes the primary device if both M-LAG member devices have configured uplink interfaces. All user traffic between the users and the uplink network is then forwarded solely through the primary device.

If the local M-LAG member device does not receive the keepalive packets transmitted by the peer M-LAG member device within the specified time and the peer link is down, the local M-LAG member device considers the peer M-LAG device as down.

- When the local device is the primary device and has uplink interfaces configured, it remains as the primary device if any uplink interface is in up state. If no uplink interface in up state exists, the local device's role changes to None.
- When the local device is the primary device and no uplink interface is configured on it, the local device remains as the primary device if an M-LAG interface in up state exists. If no M-LAG interface in up state exists, the local device's role changes to None.
- When the local device is a secondary device, it becomes the primary device. If the local device is configured with uplink interfaces, it will remain as the primary device as long as an uplink interface in up state exists on the local device. If no uplink interface in up state exists, the local device's role will change to None.
- When the local device is a secondary device, it becomes the primary device. If the local device does not have an uplink interface configured, it will remain as the primary device as long as an M-LAG interface in up state exists on the device. If no M-LAG interface in up state exists, the local device's role will change to None.

When the device is of the None role, it cannot send or receive keepalive packets, and the keepalive link is down.

Restrictions and guidelines

You cannot configure aggregation member ports or peer-link interfaces as uplink interfaces by using this feature.

To configure an M-LAG uplink interface, configure it on both M-LAG member devices. This prevents abnormal M-LAG device role elections.

Procedure

1. Enter system view.

   system-view

2. Enter the Layer 2 Ethernet interface view, Layer 3 Ethernet interface view, Layer 2 aggregate interface view, or Layer 3 aggregate interface view.

   interface *interface-type interface-number*

3. Configure the interface as an M-LAG uplink interface.

   m-lag uplink-port enable

   By default, no M-LAG uplink interface is configured.

Configuring the flapping detection interval for M-LAG uplink interfaces

About this task

M-LAG uplink interfaces are used to elect the primary and secondary M-LAG member devices when the peer link is down but the keepalive link remains up. If an uplink interface on the local M-LAG member device comes up first, that M-LAG member device is elected as the primary device. Then, all traffic is forwarded through the primary device. However, the uplink interfaces on an M-LAG member device might frequently flap, which can cause packet loss for traffic on the primary device. To prevent packet loss caused by uplink interface flapping, configure the flapping detection interval for M-LAG uplink interfaces on both M-LAG member devices. The system elects the device with less frequent or no flappings within the detection interval as the primary device.

Within the flapping detection interval, each M-LAG member device counts the flapping frequency of its uplink interfaces. The system calculates the flapping frequency through dividing the number of flaps by the detection interval. A higher flapping frequency gives a lower priority in primary/secondary device election. The two M-LAG member devices notify each other through keepalive packets and elect the M-LAG member device with the best uplink quality as the primary device based on flapping frequency.

Restrictions and guidelines

For this feature to take effect, configure M-LAG uplink interfaces and the flapping detection interval on both M-LAG member devices.

If you configure multiple interfaces as M-LAG uplink interfaces, the system calculates the flapping frequency for those M-LAG uplink interfaces through dividing the total flaps of all M-LAG uplink interfaces by the detection interval.

Procedure

1. Enter system view.

   system-view

2. Configure the flapping detection interval for M-LAG uplink interfaces.

   m-lag uplink-port flapping-detection timer *value*

   By default, an M-LAG member device does not perform flapping detection for M-LAG uplink interfaces, and no flapping detection interval is configured.

Setting up a peer link

Specifying a Layer 2 aggregate interface or VXLAN tunnel interface as the peer-link interface

About this task

On a peer-link interface, you can specify the **auto-negotiate** keyword to enable M-LAG system parameter autonegotiation. The M-LAG member devices will negotiate the M-LAG system parameters not manually configured as follows:

- **M-LAG system MAC address**—The M-LAG member devices compare their bridge MAC addresses and use the lower value as the M-LAG system MAC address.
- **M-LAG system number**—The M-LAG system number is 1 for the M-LAG member device with the lower bridge MAC address and 2 for the other M-LAG member device.
- **M-LAG system priority**—The default M-LAG system priority is used by both M-LAG member devices.

Restrictions and guidelines

A Layer 2 aggregate interface cannot operate as both peer-link interface and M-LAG interface. Make sure the bandwidth of the peer-link interface is higher than that of an M-LAG interface.

An M-LAG member device can have only one peer-link interface.

If you specify an aggregate interface as a peer-link interface, the following rules apply:

- If the aggregate interface uses the default VLAN settings, the system changes the link type of the aggregate interface to trunk, configures the interface to permit all VLANs, and sets the PVID to 4094.
- If the link type of the aggregate interface is trunk, the system configures the interface to also permit VLAN 4094 and changes the PVID to 4094.
- In any other cases, the system does not change the VLAN settings.

The device does not change the VLAN settings of an aggregate interface when you remove its peer-link interface role.

As a best practice to reduce the impact of interface flapping on upper-layer services, execute the **link-delay** command on the peer-link interface. For more information about this command, see Ethernet link aggregation commands in *Layer 2—LAN Switching Command Reference*.

By default, MAC address learning is enabled on the peer-link interface. This feature is not configurable on the peer-link interface. For more information about the MAC address learning feature, see "Configuring the MAC address table."

To prevent data synchronization failure, you must set the same maximum jumbo frame length on the peer-link interfaces of the M-LAG member devices. For more information about jumbo frames, see "Configuring Ethernet link aggregation."

You must use the **shutdown** command to shut down an interface before you remove it from the peer-link interface's aggregation group.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*

3. Specify the interface as the peer-link interface.
   port m-lag peer-link *port-number* [**auto-negotiate**]

Reserving VLANs for the peer link

About this task

Reserved VLANs enable Layer 3 multicast traffic to fail over via the peer link between the M-LAG member devices in an M-LAG system with Layer 3 multicast forwarding configured. After you reserve VLANs for the peer link, the peer-link interfaces will be assigned to the reserved VLANs and thus act as traffic outgoing interfaces of multicast forwarding entries.

For more information about configuring Layer 3 multicast forwarding on an M-LAG system, see PIM configuration and IPv6 PIM configuration in *IP Multicast Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Reserve the VLAN for the peer link.
   m-lag peer-link reserved

   By default, a VLAN is not reserved for the peer link.

Enabling the short DRCP timeout timer on the peer-link interface or an M-LAG interface

About this task

By default, the peer-link interface or an M-LAG interface uses the 90-second long DRCP timeout timer. To detect peer interface down events more quickly, enable the 3-second short DRCP timeout timer on the interface.

Restrictions and guidelines

To avoid traffic interruption during an ISSU or M-LAG process restart, disable the short DRCP timeout timer before you perform an ISSU or M-LAG process restart. For more information about ISSU, see *Fundamentals Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Enable the short DRCP timeout timer.
   m-lag drcp period short

   By default, an interface uses the long DRCP timeout timer (90 seconds).

Enabling the peer-link interface to retain MAC address entries for down single-homed devices

About this task

When an M-LAG member device detects that the link to a single-homed device goes down, the peer-link interface takes the following actions:

- Deletes the MAC address entries for the single-homed device.
- Sends a message to the peer peer-link interface for it to delete the affected MAC address entries.

If the link to a single-homed device flaps constantly, the peer-link interface repeatedly deletes and adds MAC address entries for the device. This situation increases floods of unicast traffic destined for the single-homed device.

To reduce flood traffic, enable the peer-link interface to retain MAC address entries for single-homed devices. After the links to single-homed devices go down, the affected MAC address entries age out on expiration of the MAC aging timer instead of being deleted immediately. The timer is set by using the **mac-address timer** command. For more information about this command, see MAC address table commands in *Layer 2—LAN Switching Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enable the peer-link interface to retain MAC address entries for single-homed devices.
   m-lag peer-link mac-address hold
   By default, the peer-link interface does not retain MAC address entries for single-homed devices when the devices go down.

Configuring the M-LAG active/standby mode

Restrictions and guidelines

If you configure the M-LAG interface on one end to operate in active/standby mode, you must also configure the peer M-LAG interface numbered the same to operate in active/standby mode and use the same type of protocol packets for election.

The M-LAG active/standby mode is applicable only when the following conditions exist:

- Active/standby NICs are connected to an M-LAG interface.
- The Layer 2 aggregation group of the M-LAG interface operates in static aggregation mode.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 aggregate interface view.
   interface bridge-aggregation *interface-number*
3. Assign the Layer 2 aggregate interface to an M-LAG group in active/standby mode.
   port m-lag group *group-id* **active-standby**
   By default, a Layer 2 aggregate interface is not assigned to an M-LAG group.
4. Configure the type of protocol packets used for active/standby M-LAG interface election.

port m-lag active-standby election { **arp** | **dhcp** | **igmp** | **mld** | **nd** } *

By default, the type of protocol packets used for active/standby M-LAG interface election is not specified, and the active/standby mode does not take effect.

Assigning a M-LAG virtual IP address to an interface

About this task

M-LAG virtual IP addresses (VIPs) allow devices to communicate with the M-LAG system by using dynamic routing protocols.

To ensure correct traffic forwarding, assign M-LAG virtual IP addresses to the following interfaces on the M-LAG system:

- VLAN interfaces that act as dual-active gateways for the same VLAN.
- Loopback interfaces that offer AAA and 802.1X authentication services. For more information, see AAA configuration in *Security Configuration Guide*.

When both M-LAG member devices act as gateways for dualhomed user-side devices, the gateway interfaces (VLAN interfaces) on the M-LAG member devices use the same IP address and MAC address. In this scenario, the M-LAG member devices cannot set up neighbor relationships with the user-side devices. To resolve this issue, assign VIPs to the gateway interfaces and configure routing protocols such as BGP, OSPF, and OSPFv3 to use the VIPs for neighbor relationship setup.

When dual-active gateways exist on the M-LAG system, you must assign unique VIPs to the gateway interfaces on the M-LAG member devices and configure both VIPs to be active. When you assign a virtual MAC address to a VLAN interface, make sure the virtual MAC address is identical to the MAC address assigned to the VLAN interface by using the **mac-address** command.

Restrictions and guidelines

When you assign multiple M-LAG VIPs to an interface, follow these restrictions and guidelines:

- You can assign a maximum of two virtual IPv4 or IPv6 addresses to an interface.
- If you configure different virtual MAC addresses for a virtual IPv4 or IPv6 address, the most recent configuration takes effect.
- You cannot configure the same virtual MAC address for multiple virtual IPv4 or IPv6 addresses.
- When you assign a virtual IPv4 or IPv6 address to VLAN interfaces, you must configure the same virtual MAC address for the virtual IPv4 or IPv6 address on both M-LAG member devices.
- To assign a virtual IPv6 address to an interface, specify the **link-local** keyword only if the address is used for OSPFv3 peer relationship setup.

If you assign both virtual IPv4 and IPv6 addresses to VLAN interfaces, make sure the virtual IPv4 and IPv6 addresses that use the same virtual MAC address are in the same state on the M-LAG member devices.

Assigning M-LAG VIPs to a VLAN interface

1. Enter system view.

   system-view

2. Enter VLAN interface view.

   interface vlan-interface *interface-number*

3. Assign a virtual IPv4 address to the VLAN interface.

   port m-lag virtual-ip *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**] [**virtual-mac** *mac-address*]

By default, no virtual IPv4 addresses are assigned to interfaces.

4. Assign a virtual IPv6 address to the VLAN interface.

   port m-lag ipv6 virtual-ip *ipv6-address* { *prefix-length* [**active** | **standby**] [**virtual-mac** *mac-address*] | **link-local** }

   By default, no virtual IPv6 addresses are assigned to interfaces.

Assigning M-LAG VIPs to a loopback interface

1. Enter system view.

   system-view

2. Enter loopback interface view.

   interface loopback *interface-number*

3. Assign a virtual IPv4 address to the loopback interface.

   port m-lag virtual-ip *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**]

   By default, no virtual IPv4 addresses are assigned to interfaces.

4. Assign a virtual IPv6 address to the loopback interface.

   port m-lag ipv6 virtual-ip *ipv6-address* { *prefix-length* [**active** | **standby**] | **link-local** }

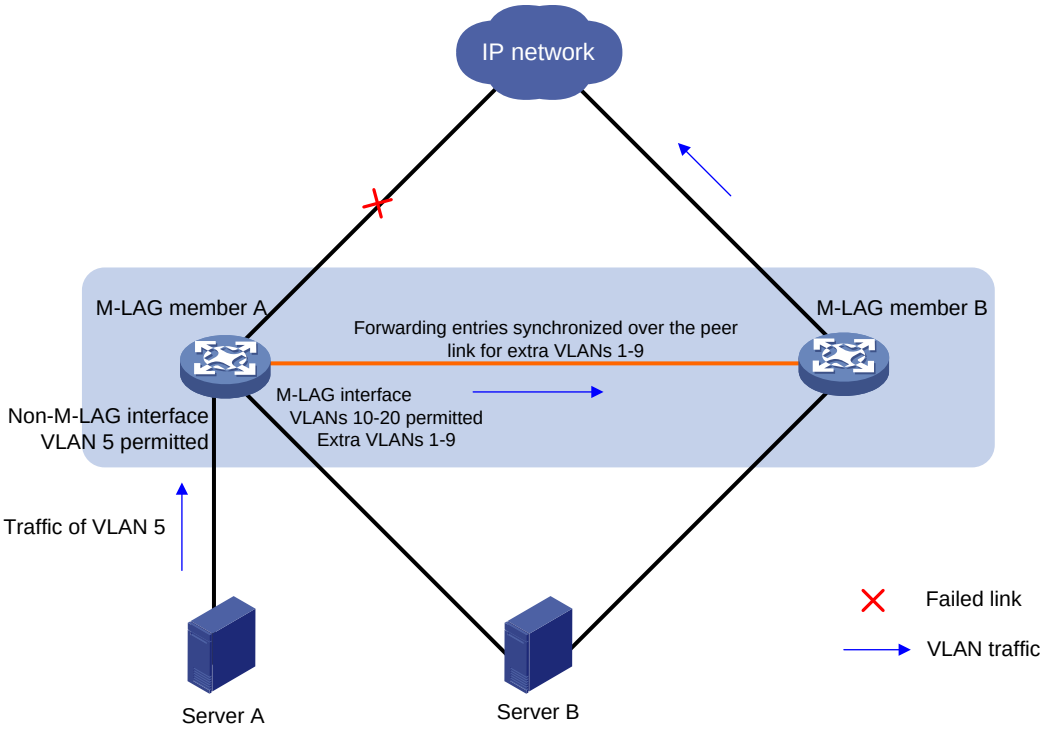
   By default, no virtual IPv6 addresses are assigned to interfaces.

Configuring M-LAG extra VLANs

About this task

For VLANs permitted only on non-M-LAG interfaces, the device processes traffic locally. For the device to synchronize forwarding entries for such VLANs, configure them as extra VLANs. The device will synchronize forwarding entries for the extra VLANs (such as ARP entries) even though they are not permitted on M-LAG interfaces, as shown in Figure 26. When the uplink of one M-LAG member device fails, traffic of the extra VLANs can fail over to the other M-LAG member device through the peer link.

Figure 26 M-LAG extra VLANs



Procedure

1. Enter system view.

   system-view
2. Configure M-LAG extra VLANs.

   m-lag extra-vlan *vlan-id-list*

   By default, no M-LAG extra VLANs are configured.

Setting the mode of configuration consistency check

About this task

The device handles configuration inconsistency depending on the mode of configuration consistency check.

- For type 1 configuration inconsistency:
  - The device generates log messages if loose mode is enabled.
  - The device shuts down M-LAG interfaces and generates log messages if strict mode is enabled.
- For type 2 configuration inconsistency, the device only generates log messages, whether strict or loose mode is enabled.

Procedure

1. Enter system view.

   system-view
2. Set the mode of configuration consistency check.

m-lag consistency-check mode { **loose** | **strict** }

By default, configuration consistency check uses strict mode.

Disabling configuration consistency check

About this task

To ensure that the M-LAG system can operate correctly, M-LAG by default performs configuration consistency check when the M-LAG system is set up.

Configuration consistency check might fail when you upgrade the M-LAG member devices in an M-LAG system. To prevent the M-LAG system from falsely shutting down M-LAG interfaces, you can temporarily disable configuration consistency check.

Restrictions and guidelines

Make sure the M-LAG member devices use the same setting for configuration consistency check.

Procedure

1. Enter system view.

   system-view

2. Disable configuration consistency check.

   m-lag consistency-check disable

   By default, configuration consistency check is enabled.

Setting the keepalive hold timer for identifying the cause of peer link down events

About this task

The keepalive hold timer starts when the peer link goes down. The keepalive hold timer specifies the amount of time that the device uses to identify the cause of a peer link down event.

- If the device receives keepalive packets from the M-LAG peer before the timer expires, the peer link is down because the peer link fails.
- If the device does not receive keepalive packets from the M-LAG peer before the timer expires, the peer link is down because the peer M-LAG member device fails.

Procedure

1. Enter system view.

   system-view

2. Set the keepalive hold timer.

   m-lag keepalive hold-time *value*

   By default, the keepalive hold timer is 3 seconds.

Configuring M-LAG system auto-recovery

About this task

If only one M-LAG member device recovers after the entire M-LAG system reboots, auto-recovery enables that member device to remove its M-LAG interfaces from the M-LAG DOWN interface list.

- If an uplink interface is configured on the device, an uplink interface in up state must exist for the device to become the primary device and forward user traffic normally. If no uplink interface in

up state exists on the device, the device will remain in the None role, and user traffic cannot be forwarded.

- If no uplink interface is configured on the device, an M-LAG interface in up state must exist for the device to become the primary device and forward user traffic normally. If no M-LAG interface in up state exists on the device, the device will remain in the None role, and user traffic cannot be forwarded.

If auto-recovery is disabled, that M-LAG member device will be stuck in the None role with all its M-LAG interfaces being M-LAG DOWN after it recovers.

Restrictions and guidelines

Both member M-LAG devices in an M-LAG system become primary devices if the following conditions exist:

- Both primary and secondary member devices restart due to a fault.
- Neither device receives DRCP or keepalive packets and both devices have M-LAG or uplink interfaces in up state (consider only uplink interfaces if they exist, and consider M-LAG interfaces otherwise) before the automatic recovery timer expires.

In this case, you must check the peer link and keepalive link and resolve any faults.

To avoid incorrect role preemption, make sure the reload delay timer is longer than the amount of time required for the device to restart.

If you execute both the **m-lag default-auto-recovery enable** and **m-lag auto-recovery reload-delay** commands, the **m-lag auto-recovery reload-delay** command takes effect.

Procedure

1. Enter system view.

   system-view

2. Set the M-LAG system reload delay timer.

   m-lag auto-recovery reload-delay *delay-value*

   By default, the device does not automatically recover.

Setting the data restoration interval

About this task

The data restoration interval specifies the maximum amount of time for the secondary M-LAG device to synchronize data with the primary M-LAG device during M-LAG system setup. Within the data restoration interval, the secondary M-LAG device sets all network interfaces to M-LAG MAD DOWN state, except for the interfaces excluded from the shutdown action by M-LAG MAD.

When the data restoration interval expires, the secondary M-LAG member device brings up all network interfaces.

Restrictions and guidelines

Increase the data restoration interval as needed for the following purposes:

- Avoid packet loss and forwarding failure that might occur when the amount of data is large or when you perform an ISSU between the M-LAG member devices.
- Avoid M-LAG interface flapping that might occur if type 1 configuration consistency check fails after the M-LAG interfaces come up upon expiration of the data restoration interval.

Procedure

1. Enter system view.

   system-view

2. Set the data restoration interval.

   m-lag restore-delay *value*

   By default, the data restoration interval is 300 seconds.

Enabling M-LAG sequence number check

Restrictions and guidelines

As a best practice to improve security, use M-LAG sequence number check together with M-LAG packet authentication.

After one M-LAG member device reboots, the other M-LAG member device might receive and accept the packets that were intercepted by an attacker before the reboot. As a best practice, change the authentication key after an M-LAG member device reboots.

Procedure

1. Enter system view.

   system-view

2. Enable M-LAG sequence number check.

   m-lag sequence enable

   By default, M-LAG sequence number check is disabled.

Enabling M-LAG packet authentication

Restrictions and guidelines

For successful authentication, configure the same authentication key for the M-LAG member devices.

Procedure

1. Enter system view.

   system-view

2. Enable M-LAG packet authentication and configure an authentication key.

   m-lag authentication key { **simple** | **cipher** } *string*

   By default, M-LAG packet authentication is disabled.

Displaying and maintaining M-LAG

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about the configuration consistency check done by M-LAG.	**display m-lag consistency** { **type1** | **type2** } { **global** | **interface** *interface-type interface-number* }
Display the status of configuration consistency check.	**display m-lag consistency-check status**
Display DRCPDU statistics.	**display m-lag drcp statistics** [**interface** *interface-type interface-number*]
Display M-LAG keepalive packet	**display m-lag keepalive**

Task	Command
statistics.	
Display detailed M-LAG MAD information.	**display m-lag mad verbose**
Display M-LAG role information.	**display m-lag role**
Display brief information about the peer-link interface and M-LAG interfaces.	**display m-lag summary**
Display the M-LAG system settings.	**display m-lag system**
Display detailed information about the peer-link interface and M-LAG interfaces.	**display m-lag verbose** [**interface bridge-aggregation** *interface-number*]
Display M-LAG troubleshooting information.	**display m-lag troubleshooting** [**m-lag-interface** | **peer-link** | **keepalive**] [**history**] [*count*]
Display M-LAG VIPs.	**display m-lag virtual-ip** [**interface** *interface-type interface-number*]
Display the M-LAG uplink interface information.	**display m-lag uplink-port information**
Clear DRCPDU statistics.	**reset m-lag drcp statistics** [**interface** *interface-list*]
Clear M-LAG troubleshooting records.	**reset m-lag troubleshooting history**

M-LAG configuration examples

Example: Configuring basic M-LAG functions

Network configuration

As shown in Figure 27, configure M-LAG on Device A and Device B to establish a multichassis aggregate link with Device C.

Figure 27 Network diagram

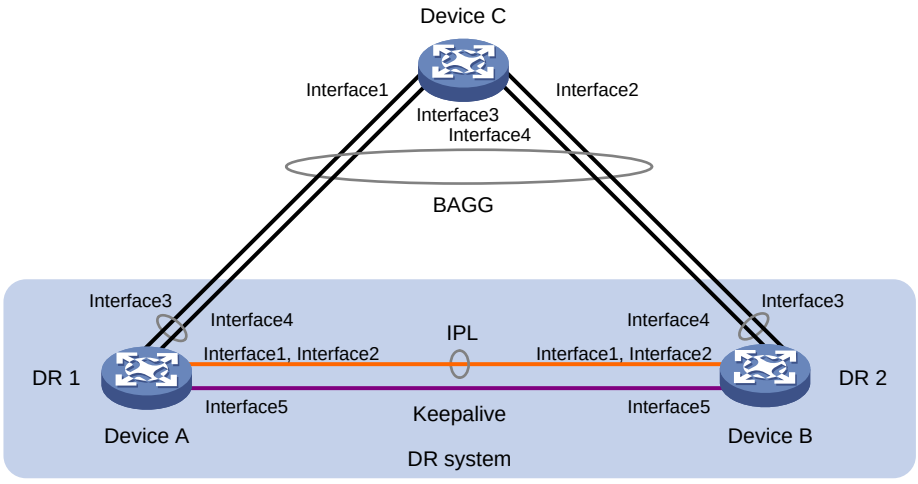


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

1. Configure Device A:

Configure M-LAG system settings.

```
<Sysname> system-view
[Sysname] sysname DeviceA
[DeviceA] m-lag system-mac 1-1-1
[DeviceA] m-lag system-number 1
[DeviceA] m-lag system-priority 123
```

Configure M-LAG keepalive packet parameters.

```
[DeviceA] m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
```

Set the link mode of GigabitEthernet 1/0/5 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[DeviceA] interface gigabitethernet 1/0/5
[DeviceA-GigabitEthernet1/0/5] port link-mode route
[DeviceA-GigabitEthernet1/0/5] ip address 1.1.1.2 24
[DeviceA-GigabitEthernet1/0/5] quit
```

Exclude the interface used for M-LAG keepalive detection (GigabitEthernet 1/0/5) from the shutdown action by M-LAG MAD.

```
[DeviceA] m-lag mad exclude interface gigabitethernet 1/0/5
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 3.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 3
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 3
[DeviceA-GigabitEthernet1/0/2] quit
```

Specify Bridge-Aggregation 3 as the peer-link interface.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] port m-lag peer-link 1
[DeviceA-Bridge-Aggregation3] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
[DeviceA] interface bridge-aggregation 4
[DeviceA-Bridge-Aggregation4] link-aggregation mode dynamic
```

```
[DeviceA-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 4.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 4
[DeviceA-GigabitEthernet1/0/3] quit
[DeviceA] interface gigabitethernet 1/0/4
[DeviceA-GigabitEthernet1/0/4] port link-aggregation group 4
[DeviceA-GigabitEthernet1/0/4] quit
```

Assign Bridge-Aggregation 4 to M-LAG group 4.

```
[DeviceA] interface bridge-aggregation 4
[DeviceA-Bridge-Aggregation4] port m-lag group 4
[DeviceA-Bridge-Aggregation4] quit
```

2. Configure Device B:

Configure M-LAG system settings.

```
<Sysname> system-view
[Sysname] sysname DeviceB
[DeviceB] m-lag system-mac 1-1-1
[DeviceB] m-lag system-number 2
[DeviceB] m-lag system-priority 123
```

Configure M-LAG keepalive packet parameters.

```
[DeviceB] m-lag keepalive ip destination 1.1.1.2 source 1.1.1.1
```

Set the link mode of GigabitEthernet 1/0/5 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[DeviceB] interface gigabitethernet 1/0/5
[DeviceB-GigabitEthernet1/0/5] port link-mode route
[DeviceB-GigabitEthernet1/0/5] ip address 1.1.1.1 24
[DeviceB-GigabitEthernet1/0/5] quit
```

Exclude the interface used for M-LAG keepalive detection (GigabitEthernet 1/0/5) from the shutdown action by M-LAG MAD.

```
[DeviceB] m-lag mad exclude interface gigabitethernet 1/0/5
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3.

```
[DeviceB] interface bridge-aggregation 3
[DeviceB-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 3.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-aggregation group 3
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-aggregation group 3
[DeviceB-GigabitEthernet1/0/2] quit
```

Specify Bridge-Aggregation 3 as the peer-link interface.

```
[DeviceB] interface bridge-aggregation 3
[DeviceB-Bridge-Aggregation3] port m-lag peer-link 1
[DeviceB-Bridge-Aggregation3] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
[DeviceB] interface bridge-aggregation 4
[DeviceB-Bridge-Aggregation4] link-aggregation mode dynamic
```

```
[DeviceB-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 4.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-aggregation group 4
[DeviceB-GigabitEthernet1/0/3] quit
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port link-aggregation group 4
[DeviceB-GigabitEthernet1/0/4] quit
```

Assign Bridge-Aggregation 4 to M-LAG group 4.

```
[DeviceB] interface bridge-aggregation 4
[DeviceB-Bridge-Aggregation4] port m-lag group 4
[DeviceB-Bridge-Aggregation4] quit
```

3. Configure Device C:

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] interface bridge-aggregation 4
[DeviceC-Bridge-Aggregation4] link-aggregation mode dynamic
[DeviceC-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to aggregation group 4.

```
[DeviceC] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[DeviceC-if-range] port link-aggregation group 4
[DeviceC-if-range] quit
```

Verifying the configuration

Verify that the keepalive link is working correctly on Device A.

```
[DeviceA] display m-lag keepalive
Neighbor keepalive link status (cause): Up
Neighbor is alive for: 104 s, 16 ms
Keepalive packet transmission status:
  Sent: Successful
  Received: Successful
Last received keepalive packet information:
  Source IP address: 1.1.1.1
  Time: 2019/09/11 09:21:51
  Action: Accept

M-LAG keepalive parameters:
Destination IP address: 1.1.1.1
Source IP address: 1.1.1.2
Keepalive UDP port : 6400
Keepalive VPN name : N/A
Keepalive interval : 1000 ms
Keepalive timeout  : 5 sec
Keepalive hold time: 3 sec
```

Verify that the peer-link interface and the M-LAG interface are working correctly on Device A.

```
[DeviceA] display m-lag summary
Flags: A -- Aggregate interface down, B -- No peer M-LAG interface configured
```

```
       C -- Configuration consistency check failed

Peer-link interface: BAGG3
Peer-link interface state (cause): UP
Keepalive link state (cause): UP

                    M-LAG interface information
M-LAG IF      M-LAG group  Local state (cause)  Peer state  Remaining down time(s)
BAGG4         4            UP                   UP          -
[DeviceA] display m-lag verbose
Flags: A -- Home_Gateway, B -- Neighbor_Gateway, C -- Other_Gateway,
       D -- PeerLink_Activity, E -- DRCP_Timeout, F -- Gateway_Sync,
       G -- Port_Sync, H -- Expired
Peer-link interface/Peer-link interface ID: BAGG3/1
State: UP
Cause: -
Local DRCP flags/Peer DRCP flags: ABDFG/ABDFG
Local Selected ports (index): GE1/0/1 (260), GE1/0/2 (261)
Peer Selected ports indexes: 260, 261

M-LAG interface/M-LAG group ID: BAGG4/4
Local M-LAG interface state: UP
Peer M-LAG interface state: UP
M-LAG group state: UP
Local M-LAG interface down cause: -
Remaining M-LAG DOWN time: -
Local M-LAG interface LACP MAC: Config=0001-0001-0001, Effective=0001-0001-0001
Peer M-LAG interface LACP MAC: Config=0001-0001-0001, Effective=0001-0001-0001
Local M-LAG interface LACP priority: Config=123, Effective=123
Peer M-LAG interface LACP priority: Config=123, Effective=123
Local DRCP flags/Peer DRCP flags: ABDFG/ABDFG
Local Selected ports (index): GE1/0/3 (258), GE1/0/4 (259)
Peer Selected ports indexes: 258, 259
```

Verify that all member ports of aggregation group 4 are in Selected state on Device C, which indicates a successful link aggregation between the M-LAG system and Device C.

```
[DeviceC] display link-aggregation verbose bridge-aggregation 4
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired
Aggregate Interface: Bridge-Aggregation4
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0x8000, 2e56-cbae-0600
```

```
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/1(R)          S        32768    1        1                      {ACDEF}
  GE1/0/2             S        32768    2        1                      {ACDEF}
  GE1/0/3             S        32768    3        1                      {ACDEF}
  GE1/0/4             S        32768    4        1                      {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/1             32768    16387    40004    0x7b  , 0001-0001-0001 {ACDEF}
  GE1/0/2             32768    16388    40004    0x7b  , 0001-0001-0001 {ACDEF}
  GE1/0/3             32768    32771    40004    0x7b  , 0001-0001-0001 {ACDEF}
  GE1/0/4             32768    32772    40004    0x7b  , 0001-0001-0001 {ACDEF}
```

Configuration files

- Device A:

```
 sysname DeviceA
#
 undo mac-address static source-check enable
#
interface Bridge-Aggregation3
 link-aggregation mode dynamic
 port m-lag peer-link 1
 undo mac-address static source-check enable
#
interface Bridge-Aggregation4
 link-aggregation mode dynamic
 port m-lag group 4
#
interface GigabitEthernet1/0/5
 ip address 1.1.1.2 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-aggregation group 3
#
interface GigabitEthernet1/0/2
 port link-aggregation group 3
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port link-aggregation group 4
#
interface GigabitEthernet1/0/4
 port link-mode bridge
 port link-aggregation group 4
#
 m-lag mad exclude interface GigabitEthernet1/0/5
 m-lag system-mac 0001-0001-0001
 m-lag system-number 1
 m-lag system-priority 123
```

```
 m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
#
```

- Device B:

```
 sysname DeviceB
#
undo mac-address static source-check enable
#
interface Bridge-Aggregation3
 link-aggregation mode dynamic
 port m-lag peer-link 1
 undo mac-address static source-check enable

#
interface Bridge-Aggregation4
 link-aggregation mode dynamic
 port m-lag group 4
#
interface GigabitEthernet1/0/5
 port link-mode route
 ip address 1.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-aggregation group 3
#
interface GigabitEthernet1/0/2
 port link-aggregation group 3
#
interface GigabitEthernet1/0/3
 port link-aggregation group 4
#
interface GigabitEthernet1/0/4
 port link-aggregation group 4
#
 m-lag mad exclude interface GigabitEthernet1/0/5
 m-lag system-mac 0001-0001-0001
 m-lag system-number 2
 m-lag system-priority 123
 m-lag keepalive ip destination 1.1.1.2 source 1.1.1.1
#
```

- Device C:

```
 sysname DeviceC
#
interface Bridge-Aggregation4
 link-aggregation mode dynamic
#
interface GigabitEthernet1/0/1
```

```
 port link-aggregation group 4
#
interface GigabitEthernet1/0/2
 port link-aggregation group 4
#
interface GigabitEthernet1/0/3
 port link-aggregation group 4
#
interface GigabitEthernet1/0/4
 port link-aggregation group 4
#
```

Example: Configuring Layer 3 gateways on an M-LAG system

Network configuration

As shown in Figure 28:

- Configure Device A and Device B as an M-LAG system to establish one multichassis aggregate link with Device C and one with Device D.
- Set up a keepalive link between GigabitEthernet 1/0/5 of Device A and GigabitEthernet 1/0/5 of Device B, and exclude the interfaces from the shutdown action by M-LAG MAD.
- Configure two VRRP groups on Device A and Device B to provide gateway services for VLAN 100 and VLAN 200. Configure Device A as the master of the VRRP groups.

Figure 28 Network diagram

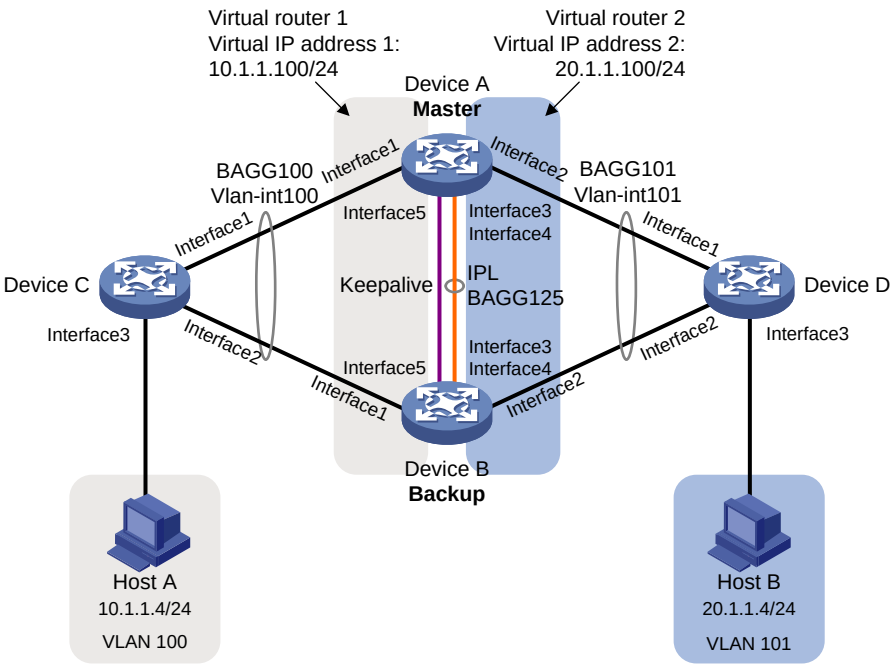


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

1. Configure Device A:

Configure M-LAG system settings.

```
<Sysname> system-view
[Sysname] sysname DeviceA
[DeviceA] m-lag system-mac 1-1-1
[DeviceA] m-lag system-number 1
[DeviceA] m-lag system-priority 123
```

Configure M-LAG keepalive parameters.

```
[DeviceA] m-lag keepalive ip destination 1.1.1.2 source 1.1.1.1
```

Set the link mode of GigabitEthernet 1/0/5 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[DeviceA] interface gigabitethernet 1/0/5
[DeviceA-GigabitEthernet1/0/5] port link-mode route
[DeviceA-GigabitEthernet1/0/5] ip address 1.1.1.1 24
[DeviceA-GigabitEthernet1/0/5] quit
```

Exclude the interface used for M-LAG keepalive detection (GigabitEthernet 1/0/5) from the shutdown action by M-LAG MAD.

```
[DeviceA] m-lag mad exclude interface gigabitethernet 1/0/5
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 125.

```
[DeviceA] interface bridge-aggregation 125
[DeviceA-Bridge-Aggregation125] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation125] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 125.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 125
[DeviceA-GigabitEthernet1/0/3] quit
[DeviceA] interface GigabitEthernet 1/0/4
[DeviceA-GigabitEthernet1/0/4] port link-aggregation group 125
[DeviceA-GigabitEthernet1/0/4] quit
```

Specify Bridge-Aggregation 125 as the peer-link interface.

```
[DeviceA] interface bridge-aggregation 125
[DeviceA-Bridge-Aggregation125] port m-lag peer-link 1
[DeviceA-Bridge-Aggregation125] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 100, and assign it to M-LAG group 1.

```
[DeviceA] interface bridge-aggregation 100
[DeviceA-Bridge-Aggregation100] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation100] port m-lag group 1
[DeviceA-Bridge-Aggregation100] quit
```

Assign GigabitEthernet 1/0/1 to aggregation group 100.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 100
[DeviceA-GigabitEthernet1/0/1] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 101, and assign it to M-LAG group 2.

```
[DeviceA] interface bridge-aggregation 101
[DeviceA-Bridge-Aggregation101] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation101] port m-lag group 2
[DeviceA-Bridge-Aggregation101] quit
```

Assign GigabitEthernet 1/0/2 to aggregation group 101.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 101
[DeviceA-GigabitEthernet1/0/2] quit
```

Create VLAN 100 and VLAN 101.

```
[DeviceA] vlan 100
[DeviceA-vlan100] quit
[DeviceA] vlan 101
[DeviceA-vlan101] quit
```

Set the link type of Bridge-Aggregation 100 to trunk, and assign it to VLAN 100.

```
[DeviceA] interface bridge-aggregation 100
[DeviceA-Bridge-Aggregation100] port link-type trunk
[DeviceA-Bridge-Aggregation100] port trunk permit vlan 100
[DeviceA-Bridge-Aggregation100] quit
```

Set the link type of Bridge-Aggregation 101 to trunk, and assign it to VLAN 101.

```
[DeviceA] interface bridge-aggregation 101
[DeviceA-Bridge-Aggregation101] port link-type trunk
[DeviceA-Bridge-Aggregation101] port trunk permit vlan 101
[DeviceA-Bridge-Aggregation101] quit
```

Create VLAN-interface 100 and VLAN-interface 101, and assign IP addresses to them.

```
[DeviceA] interface vlan-interface 100
[DeviceA-vlan-interface100] ip address 10.1.1.1 24
[DeviceA-vlan-interface100] quit
[DeviceA] interface vlan-interface 101
[DeviceA-vlan-interface101] ip address 20.1.1.1 24
[DeviceA-vlan-interface101] quit
```

Exclude VLAN-interface 100 and VLAN-interface 101 from the shutdown action by M-LAG MAD.

```
[DeviceA] m-lag mad exclude interface vlan-interface 100
[DeviceA] m-lag mad exclude interface vlan-interface 101
```

Configure OSPF.

```
[DeviceA] ospf
[DeviceA-ospf-1] import-route direct
[DeviceA-ospf-1] area 0
[DeviceA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[DeviceA-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[DeviceA-ospf-1-area-0.0.0.0] quit
[DeviceA-ospf-1] quit
```

Create VRRP group 1 on VLAN-interface 100 and set its virtual IP address to 10.1.1.100.

```
[DeviceA] interface vlan-interface 100
[DeviceA-Vlan-interface100] vrrp vrid 1 virtual-ip 10.1.1.100
```

Set the priority of Device A (primary M-LAG member device) to 200 for it to become the master in VRRP group 1.

```
[DeviceA-Vlan-interface100] vrrp vrid 1 priority 200
[DeviceA-Vlan-interface100] quit
```

Create VRRP group 2 on VLAN-interface 101 and set its virtual IP address to 20.1.1.100.

```
[DeviceA] interface vlan-interface 101
[DeviceA-Vlan-interface101] vrrp vrid 2 virtual-ip 20.1.1.100
```

Set the priority of Device A (primary M-LAG member device) to 200 for it to become the master in VRRP group 2.

```
[DeviceA-Vlan-interface101] vrrp vrid 2 priority 200
[DeviceA-Vlan-interface101] quit
```

2. Configure Device B:

Configure M-LAG system settings.

```
<Sysname> system-view
[Sysname] sysname DeviceB
[DeviceB] m-lag system-mac 1-1-1
[DeviceB] m-lag system-number 2
[DeviceB] m-lag system-priority 123
```

Configure M-LAG keepalive parameters.

```
[DeviceB] m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
```

Set the link mode of GigabitEthernet 1/0/5 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[DeviceB] interface gigabitethernet 1/0/5
[DeviceB-GigabitEthernet1/0/5] port link-mode route
[DeviceB-GigabitEthernet1/0/5] ip address 1.1.1.2 24
[DeviceB-GigabitEthernet1/0/5] quit
```

Exclude the interface used for M-LAG keepalive detection (GigabitEthernet 1/0/5) from the shutdown action by M-LAG MAD.

```
[DeviceB] m-lag mad exclude interface gigabitethernet 1/0/5
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 125.

```
[DeviceB] interface bridge-aggregation 125
[DeviceB-Bridge-Aggregation125] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation125] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 125.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-aggregation group 125
[DeviceB-GigabitEthernet1/0/3] quit
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port link-aggregation group 125
[DeviceB-GigabitEthernet1/0/4] quit
```

Specify Bridge-Aggregation 125 as the peer-link interface.

```
[DeviceB] interface bridge-aggregation 125
[DeviceB-Bridge-Aggregation125] port m-lag peer-link 1
[DeviceB-Bridge-Aggregation125] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 100, and assign it to M-LAG group 1.

```
[DeviceB] interface bridge-aggregation 100
[DeviceB-Bridge-Aggregation100] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation100] port m-lag group 1
[DeviceB-Bridge-Aggregation100] quit
```

Assign GigabitEthernet 1/0/1 to aggregation group 100.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-aggregation group 100
[DeviceB-GigabitEthernet1/0/1] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 101, and assign it to M-LAG group 2.

```
[DeviceB] interface bridge-aggregation 101
[DeviceB-Bridge-Aggregation101] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation101] port m-lag group 2
[DeviceB-Bridge-Aggregation101] quit
```

Assign GigabitEthernet 1/0/2 to aggregation group 101.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-aggregation group 101
[DeviceB-GigabitEthernet1/0/2] quit
```

Create VLAN 100 and VLAN 101.

```
[DeviceB] vlan 100
[DeviceB-vlan100] quit
[DeviceB] vlan 101
[DeviceB-vlan101] quit
```

Set the link type of Bridge-Aggregation 100 to trunk, and assign it to VLAN 100.

```
[DeviceB] interface bridge-aggregation 100
[DeviceB-Bridge-Aggregation100] port link-type trunk
[DeviceB-Bridge-Aggregation100] port trunk permit vlan 100
[DeviceB-Bridge-Aggregation100] quit
```

Set the link type of Bridge-Aggregation 101 to trunk, and assign it to VLAN 101.

```
[DeviceB] interface bridge-aggregation 101
[DeviceB-Bridge-Aggregation101] port link-type trunk
[DeviceB-Bridge-Aggregation101] port trunk permit vlan 101
[DeviceB-Bridge-Aggregation101] quit
```

Create VLAN-interface 100 and VLAN-interface 101, and assign IP addresses to them.

```
[DeviceB] interface vlan-interface 100
[DeviceB-vlan-interface100] ip address 10.1.1.2 24
[DeviceB-vlan-interface100] quit
[DeviceB] interface vlan-interface 101
[DeviceB-vlan-interface101] ip address 20.1.1.2 24
[DeviceB-vlan-interface101] quit
```

Exclude VLAN-interface 100 and VLAN-interface 101 from the shutdown action by M-LAG MAD.

```
[DeviceB] m-lag mad exclude interface vlan-interface 100
[DeviceB] m-lag mad exclude interface vlan-interface 101
```

Configure OSPF.

```
[DeviceB] ospf
[DeviceB-ospf-1] import-route direct
[DeviceB-ospf-1] area 0
```

```
[DeviceB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[DeviceB-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[DeviceB-ospf-1-area-0.0.0.0] quit
[DeviceB-ospf-1] quit
```

Create VRRP group 1 on VLAN-interface 100 and set its virtual IP address to 10.1.1.100.

```
[DeviceB] interface vlan-interface 100
[DeviceB-Vlan-interface100] vrrp vrid 1 virtual-ip 10.1.1.100
[DeviceB-Vlan-interface100] quit
```

Create VRRP group 2 on VLAN-interface 101 and set its virtual IP address to 20.1.1.100.

```
[DeviceB] interface vlan-interface 101
[DeviceB-Vlan-interface101] vrrp vrid 2 virtual-ip 20.1.1.100
[DeviceB-Vlan-interface101] quit
```

3. Configure Device C:

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 100.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] interface bridge-aggregation 100
[DeviceC-Bridge-Aggregation100] link-aggregation mode dynamic
[DeviceC-Bridge-Aggregation100] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 100.

```
[DeviceC] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/2
[DeviceC-if-range] port link-aggregation group 100
[DeviceC-if-range] quit
```

Create VLAN 100.

```
[DeviceC] vlan 100
[DeviceC-vlan100] quit
```

Set the link type of Bridge-Aggregation 100 to trunk, and assign it to VLAN 100.

```
[DeviceC] interface bridge-aggregation 100
[DeviceC-Bridge-Aggregation100] port link-type trunk
[DeviceC-Bridge-Aggregation100] port trunk permit vlan 100
[DeviceC-Bridge-Aggregation100] quit
```

Set the link type of GigabitEthernet 1/0/3 to trunk, and assign it to VLAN 100.

```
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] port link-type trunk
[DeviceC-GigabitEthernet1/0/3] port trunk permit vlan 100
[DeviceC-GigabitEthernet1/0/3] quit
```

Create VLAN-interface 100, and assign it an IP address.

```
[DeviceC] interface vlan-interface 100
[DeviceC-vlan-interface100] ip address 10.1.1.3 24
[DeviceC-vlan-interface100] quit
```

Configure OSPF.

```
[DeviceC] ospf
[DeviceC-ospf-1] import-route direct
[DeviceC-ospf-1] area 0
[DeviceC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[DeviceC-ospf-1-area-0.0.0.0] quit
[DeviceC-ospf-1] quit
```

4. Configure Device D:

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 101.

```
<Sysname> system-view
[Sysname] sysname DeviceD
[DeviceD] interface bridge-aggregation 101
[DeviceD-Bridge-Aggregation101] link-aggregation mode dynamic
[DeviceD-Bridge-Aggregation101] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 101.

```
[DeviceD] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/2
[DeviceD-if-range] port link-aggregation group 101
[DeviceD-if-range] quit
```

Create VLAN 101.

```
[DeviceD] vlan 101
[DeviceD-vlan101] quit
```

Set the link type of Bridge-Aggregation 101 to trunk, and assign it to VLAN 101.

```
[DeviceD] interface bridge-aggregation 101
[DeviceD-Bridge-Aggregation101] port link-type trunk
[DeviceD-Bridge-Aggregation101] port trunk permit vlan 101
[DeviceD-Bridge-Aggregation101] quit
```

Set the link type of GigabitEthernet 1/0/3 to trunk, and assign it to VLAN 101.

```
[DeviceD] interface gigabitethernet 1/0/3
[DeviceD-GigabitEthernet1/0/3] port link-type trunk
[DeviceD-GigabitEthernet1/0/3] port trunk permit vlan 101
[DeviceD-GigabitEthernet1/0/3] quit
```

Create VLAN-interface 101, and assign it an IP address.

```
[DeviceD] interface vlan-interface 101
[DeviceD-vlan-interface101] ip address 20.1.1.3 24
[DeviceD-vlan-interface101] quit
```

Configure OSPF.

```
[DeviceD] ospf
[DeviceD-ospf-1] import-route direct
[DeviceD-ospf-1] area 0
[DeviceD-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[DeviceD-ospf-1-area-0.0.0.0] quit
[DeviceD-ospf-1] quit
```

Verifying the configuration

Verify that Device A has formed an M-LAG system with Device B.

```
[DeviceA] display m-lag summary
Flags: A -- Aggregate interface down, B -- No peer M-LAG interface configured
       C -- Configuration consistency check failed

Peer-link interface: BAGG125
Peer-link interface state (cause): UP
Keepalive link state (cause): UP

                    M-LAG interface information
M-LAG IF      M-LAG group  Local state (cause)  Peer state  Remaining down time(s)
```

```
BAGG100       1            UP                  UP          -
BAGG101       2            UP                  UP          -
[DeviceA] display m-lag verbose
Flags: A -- Home_Gateway, B -- Neighbor_Gateway, C -- Other_Gateway,
       D -- PeerLink_Activity, E -- DRCP_Timeout, F -- Gateway_Sync,
       G -- Port_Sync, H -- Expired

Peer-link interface/Peer-link interface ID: BAGG125/1
State: UP
Cause: -
Local DRCP flags/Peer DRCP flags: ABDFG/ABDFG
Local Selected ports (index): GE1/0/3 (260), GE1/0/4 (261)
Peer Selected ports indexes: 260, 261

M-LAG interface/M-LAG group ID: BAGG100/1
Local M-LAG interface state: UP
Peer M-LAG interface state: UP
M-LAG group state: UP
Local M-LAG interface down cause: -
Remaining M-LAG DOWN time: -
Local M-LAG interface LACP MAC: Config=N/A, Effective=0001-0001-0001
Peer M-LAG interface LACP MAC: Config=N/A, Effective=0001-0001-0001
Local M-LAG interface LACP priority: Config=32768, Effective=123
Peer M-LAG interface LACP priority: Config=32768, Effective=123
Local DRCP flags/Peer DRCP flags: ABDFG/ABDFG
Local Selected ports (index): GE1/0/1 (258)
Peer Selected ports indexes: 258

M-LAG interface/M-LAG group ID: BAGG101/2
Local M-LAG interface state: UP
Peer M-LAG interface state: UP
M-LAG group state: UP
Local M-LAG interface down cause: -
Remaining M-LAG DOWN time: -
Local M-LAG interface LACP MAC: Config=N/A, Effective=0001-0001-0001
Peer M-LAG interface LACP MAC: Config=N/A, Effective=0001-0001-0001
Local M-LAG interface LACP priority: Config=32768, Effective=123
Peer M-LAG interface LACP priority: Config=32768, Effective=123
Local DRCP flags/Peer DRCP flags: ABDFG/ABDFG
Local Selected ports (index): GE1/0/2 (259)
Peer Selected ports indexes: 259
```

Verify that Device C and Device D have correctly set up aggregate links with the M-LAG system.

```
[DeviceC] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
```

```
        G -- Defaulted, H -- Expired

Aggregate Interface: Bridge-Aggregation100
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0x8000, a03b-0694-0300
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/1             S        32768    1        1                      {ACDEF}
  GE1/0/2             S        32768    2        1                      {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/1(R)          32768    16386    40001    0x7b  , 0001-0001-0001 {ACDEF}
  GE1/0/2             32768    32770    40001    0x7b  , 0001-0001-0001 {ACDEF}
[DeviceD] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregate Interface: Bridge-Aggregation101
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0x8000, a03b-0d51-0400
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/1             S        32768    1        1                      {ACDEF}
  GE1/0/2             S        32768    2        1                      {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/1(R)          32768    16387    40002    0x7b  , 0001-0001-0001 {ACDEF}
  GE1/0/2             32768    32771    40002    0x7b  , 0001-0001-0001 {ACDEF}
```

Verify that Device A is the master in VRRP group 1 and VRRP group 2.

```
[DeviceA] display vrrp
IPv4 Virtual Router Information:
 Running mode : Standard
 Total number of virtual routers : 2
 Interface          VRID  State        Running Adver   Auth            Virtual
                                         Pri     Timer   Type            IP
 ------------------------------------------------------------------------------
 Vlan100            1     Master       200     100     None            10.1.1.100
 Vlan101            2     Master       200     100     None            20.1.1.100
```

```
[DeviceB] display vrrp
IPv4 Virtual Router Information:
 Running mode : Standard
 Total number of virtual routers : 2
 Interface          VRID  State        Running Adver   Auth            Virtual
                                         Pri     Timer   Type            IP
 -----------------------------------------------------------------------------
 Vlan100            1     Backup       100     100     None            10.1.1.100
 Vlan101            2     Backup       100     100     None            20.1.1.100
```

Verify that Device C and Device D have established OSPF neighbor relationships with Device A and Device B.

```
[DeviceC] display ospf peer

         OSPF Process 1 with Router ID 10.1.1.3
               Neighbor Brief Information

 Area: 0.0.0.0
 Router ID       Address         Pri Dead-Time  State             Interface
 20.1.1.1        10.1.1.1        1   37         Full/DR           Vlan100
 20.1.1.2        10.1.1.2        1   32         Full/BDR          Vlan100
[DeviceD] display ospf peer

         OSPF Process 1 with Router ID 20.1.1.3
               Neighbor Brief Information

 Area: 0.0.0.0
 Router ID       Address         Pri Dead-Time  State             Interface
 20.1.1.1        20.1.1.1        1   38         Full/DR           Vlan101
 20.1.1.2        20.1.1.2        1   37         Full/BDR          Vlan101
```

Verify that Host A and Host B can ping each other. (Details not shown.)

Configuration files

- Device A:

```
 sysname DeviceA
#
ospf 1
 import-route direct
 area 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 20.1.1.0 0.0.0.255
#
 undo mac-address static source-check enable
#
vlan 100 to 101
#
interface Bridge-Aggregation100
 port link-type trunk
 port trunk permit vlan 1 100
```

```
 link-aggregation mode dynamic
 port m-lag group 1
#
interface Bridge-Aggregation101
 port link-type trunk
 port trunk permit vlan 1 101
 link-aggregation mode dynamic
 port m-lag group 2
#
interface Bridge-Aggregation125
 port link-type trunk
 port trunk permit vlan all
 link-aggregation mode dynamic
 port m-lag peer-link 1
#
interface Vlan-interface100
 ip address 10.1.1.1 255.255.255.0
 vrrp vrid 1 virtual-ip 10.1.1.100
 vrrp vrid 1 priority 200
#
interface Vlan-interface101
 ip address 20.1.1.1 255.255.255.0
 vrrp vrid 2 virtual-ip 20.1.1.100
 vrrp vrid 2 priority 200
#
interface GigabitEthernet1/0/5
 port link-mode route
 ip address 1.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 100
 port link-aggregation group 100
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 101
 port link-aggregation group 101
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port link-aggregation group 125
#
interface GigabitEthernet1/0/4
 port link-mode bridge
 port link-aggregation group 125
#
 m-lag mad exclude interface GigabitEthernet1/0/5
```

```
 m-lag mad exclude interface Vlan-interface100
 m-lag mad exclude interface Vlan-interface101
 m-lag system-mac 0001-0001-0001
 m-lag system-number 1
 m-lag system-priority 123
 m-lag keepalive ip destination 1.1.1.2 source 1.1.1.1
#
```

- Device B:

```
 sysname DeviceB
#
ospf 1
 import-route direct
 area 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 20.1.1.0 0.0.0.255
#
 undo mac-address static source-check enable
#
vlan 100 to 101
#
interface Bridge-Aggregation100
 port link-type trunk
 port trunk permit vlan 1 100
 link-aggregation mode dynamic
 port m-lag group 1
#
interface Bridge-Aggregation101
 port link-type trunk
 port trunk permit vlan 1 101
 link-aggregation mode dynamic
 port m-lag group 2
#
interface Bridge-Aggregation125
 port link-type trunk
 port trunk permit vlan all
 link-aggregation mode dynamic
 port m-lag peer-link 1
#
interface Vlan-interface100
 ip address 10.1.1.2 255.255.255.0
 vrrp vrid 1 virtual-ip 10.1.1.100
#
interface Vlan-interface101
 ip address 20.1.1.2 255.255.255.0
 vrrp vrid 2 virtual-ip 20.1.1.100
#
interface GigabitEthernet1/0/5
```

```
 port link-mode route
 ip address 1.1.1.2 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 100
 port link-aggregation group 100
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 101
 port link-aggregation group 101
#
interface GigabitEthernet1/0/3
 port link-aggregation group 125
#
interface GigabitEthernet1/0/4
 port link-aggregation group 125
#
 m-lag mad exclude interface GigabitEthernet1/0/5
 m-lag mad exclude interface Vlan-interface100
 m-lag mad exclude interface Vlan-interface101
 m-lag system-mac 0001-0001-0001
 m-lag system-number 2
 m-lag system-priority 123
 m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
#
```

- Device C:

```
 sysname DeviceC
#
ospf 1
 import-route direct
 area 0.0.0.0
  network 10.1.1.0 0.0.0.255
#
vlan 100
#
interface Bridge-Aggregation100
 port link-type trunk
 port trunk permit vlan 1 100
 link-aggregation mode dynamic
#
interface Vlan-interface100
 ip address 10.1.1.3 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-type trunk
```

```
 port trunk permit vlan 1 100
 port link-aggregation group 100
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 100
 port link-aggregation group 100
#
interface GigabitEthernet1/0/3
 port link-type trunk
 port trunk permit vlan 1 100
#
```

- Device D:

```
 sysname DeviceD
#
ospf 1
 import-route direct
 area 0.0.0.0
  network 20.1.1.0 0.0.0.255
#
vlan 101
#
interface Bridge-Aggregation101
 port link-type trunk
 port trunk permit vlan 1 101
 link-aggregation mode dynamic
#
interface Vlan-interface101
 ip address 20.1.1.3 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 101
 port link-aggregation group 101
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 101
 port link-aggregation group 101
#
interface GigabitEthernet1/0/3
 port link-type trunk
 port trunk permit vlan 1 101
#
```

Example: Configure dual-active gateways on an M-LAG system

Network configuration

As shown in Figure 29, set up an M-LAG system with Device A and Device B and configure dual-active gateways for the servers on Device A and Device B.

Figure 29 Network diagram

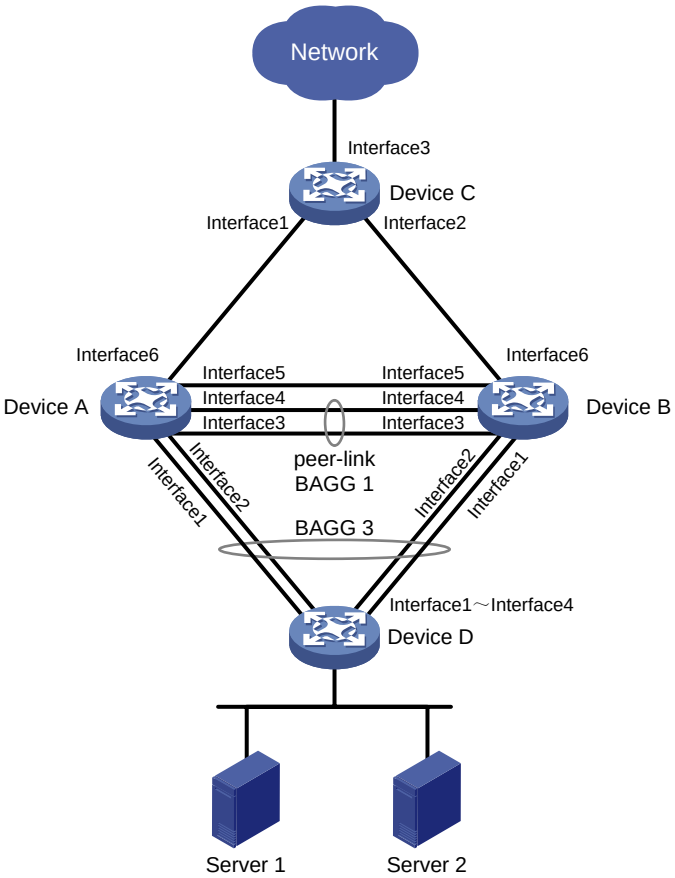


Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5
Interface6	GigabitEthernet1/0/6

Restrictions and guidelines

In this example, all devices use the factory default settings. When you apply this configuration example to a live environment, make sure the existing configurations do not conflict with the configurations in this example.

You must configure the same M-LAG system MAC address on the M-LAG member devices in the same M-LAG system. You must assign unique M-LAG system MAC addresses to different M-LAG systems.

To avoid loops, do not configure local proxy ARP or ND on the downstream VLAN interfaces of the M-LAG member devices. For more information about local proxy ARP or ND, see proxy ARP configuration and IPv6 neighbor discovery in *Layer 3—IP Services Configuration Guide*.

Procedure

1. Configure Device A to form an M-LAG system:

Configure M-LAG system settings.

```
<Sysname> system-view
[Sysname] sysname DeviceA
[DeviceA] m-lag system-mac 2-2-2
[DeviceA] m-lag system-number 1
[DeviceA] m-lag system-priority 123
```

Configure M-LAG keepalive parameters.

```
[DeviceA] m-lag keepalive ip destination 21.1.1.2 source 21.1.1.1
```

Set the link mode of GigabitEthernet 1/0/5 to Layer 3, and assign the interface IPv4 and IPv6 addresses. The IPv4 address will be used as the source IP address of keepalive packets.

```
[DeviceA] interface gigabitethernet 1/0/5
[DeviceA-GigabitEthernet1/0/5] port link-mode route
[DeviceA-GigabitEthernet1/0/5] ip address 21.1.1.1 24
[DeviceA-GigabitEthernet1/0/5] ipv6 address 21::1 64
[DeviceA-GigabitEthernet1/0/5] quit
```

Exclude the interface used for M-LAG keepalive detection (GigabitEthernet 1/0/5) from the shutdown action by M-LAG MAD.

```
[DeviceA] m-lag mad exclude interface gigabitethernet 1/0/5
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 1.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation1] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 1.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/3] quit
[DeviceA] interface gigabitethernet 1/0/4
[DeviceA-GigabitEthernet1/0/4] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/4] quit
```

Configure Bridge-Aggregation 1 as the peer-link interface.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] port m-lag peer-link 1
[DeviceA-Bridge-Aggregation1] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3, and assign it to M-LAG group 1.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation3] port m-lag group 1
[DeviceA-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 3.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 3
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 3
[DeviceA-GigabitEthernet1/0/2] quit
```

Create VLAN 102 and VLAN 103.

```
[DeviceA] vlan 102 103
```

Set the link type of Bridge-Aggregation 3 to trunk, and assign it to VLAN 102 and VLAN 103.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] port link-type trunk
[DeviceA-Bridge-Aggregation3] port trunk permit vlan 102 103
[DeviceA-Bridge-Aggregation3] quit
```

2. Configure Device B to form an M-LAG system:

Configure M-LAG system settings.

```
<Sysname> system-view
[Sysname] sysname DeviceB
[DeviceB] m-lag system-mac 2-2-2
[DeviceB] m-lag system-number 2
[DeviceB] m-lag system-priority 123
```

Configure M-LAG keepalive parameters.

```
[DeviceB] m-lag keepalive ip destination 21.1.1.1 source 21.1.1.2
```

Set the link mode of GigabitEthernet 1/0/5 to Layer 3, and assign the interface IPv4 and IPv6 addresses. The IPv4 address will be used as the source IP address of keepalive packets.

```
[DeviceB] interface gigabitethernet 1/0/5
[DeviceB-GigabitEthernet1/0/5] port link-mode route
[DeviceB-GigabitEthernet1/0/5] ip address 21.1.1.2 255.255.255.0
[DeviceB-GigabitEthernet1/0/5] ipv6 address 21::2 64
[DeviceB-GigabitEthernet1/0/5] quit
```

Exclude the interface used for M-LAG keepalive detection (GigabitEthernet 1/0/5) from the shutdown action by M-LAG MAD.

```
[DeviceB] m-lag mad exclude interface gigabitethernet 1/0/5
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 1.

```
[DeviceB] interface bridge-aggregation 1
[DeviceB-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation1] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 1.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-aggregation group 1
[DeviceB-GigabitEthernet1/0/3] quit
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port link-aggregation group 1
[DeviceB-GigabitEthernet1/0/4] quit
```

Configure Bridge-Aggregation 1 as the peer-link interface.

```
[DeviceB] interface bridge-aggregation 1
[DeviceB-Bridge-Aggregation1] port m-lag peer-link 1
[DeviceB-Bridge-Aggregation1] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3, and assign it to M-LAG group 1.

```
[DeviceB] interface bridge-aggregation 3
[DeviceB-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation3] port m-lag group 1
[DeviceB-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 3.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-aggregation group 3
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-aggregation group 3
[DeviceB-GigabitEthernet1/0/2] quit
```

Create VLAN 102 and VLAN 103.

```
[DeviceB] vlan 102 103
```

Set the link type of Bridge-Aggregation 3 to trunk, and assign it to VLAN 102 and VLAN 103.

```
[DeviceB] interface bridge-aggregation 3
[DeviceB-Bridge-Aggregation3] port link-type trunk
[DeviceB-Bridge-Aggregation3] port trunk permit vlan 102 103
[DeviceB-Bridge-Aggregation3] quit
```

3. Configure dual-active gateways on Device A after the M-LAG system is set up:

Create VLAN-interface 102 and assign it an IPv4 address and an IPv6 address for the interface to act as an IPv4 and IPv6 gateway.

```
[DeviceA] interface vlan-interface 102
[DeviceA-Vlan-interface102] ip address 10.102.0.1 255.255.255.0
[DeviceA-Vlan-interface102] ipv6 address 10:102::1 64
```

Assign M-LAG virtual IPv4 and IPv6 addresses to VLAN-interface 102 for Device A to establish OSPF and OSPFv3 neighbor relationships with Device D.

```
[DeviceA-Vlan-interface102] port m-lag virtual-ip 10.102.0.3 24 active
[DeviceA-Vlan-interface102] port m-lag ipv6 virtual-ip FE80::2 link-local
```

Assign a MAC address to VLAN-interface 102.

```
[DeviceA-Vlan-interface102] mac-address 3-3-3
[DeviceA-Vlan-interface102] quit
```

Exclude VLAN-interface 102 from the shutdown action by M-LAG MAD.

```
[DeviceA] m-lag mad exclude interface vlan-interface 102
```

Create VLAN-interface 103 and assign it an IPv4 address and an IPv6 address for the interface to act as an IPv4 and IPv6 gateway.

```
[DeviceA] interface vlan-interface 103
[DeviceA-Vlan-interface103] ip address 10.103.0.1 24
[DeviceA-Vlan-interface103] ipv6 address 10:103::1 64
```

Assign M-LAG virtual IPv4 and IPv6 addresses to VLAN-interface 103 for Device A to establish BGP and BGP4+ peer relationships with Device D.

```
[DeviceA-Vlan-interface103] port m-lag virtual-ip 10.103.0.3 24 active
[DeviceA-Vlan-interface103] port m-lag ipv6 virtual-ip 10:103::3 64 active
```

Assign a MAC address to VLAN-interface 103.

```
[DeviceA-Vlan-interface103] mac-address 4-4-4
[DeviceA-Vlan-interface103] quit
```

Exclude VLAN-interface 103 from the shutdown action by M-LAG MAD.

```
[DeviceA] m-lag mad exclude interface vlan-interface 103
```

Assign an IPv4 address and an IPv6 address to uplink interface GigabitEthernet 1/0/6.

```
[DeviceA] interface gigabitethernet 1/0/6
[DeviceA-GigabitEthernet1/0/6] port link-mode route
[DeviceA-GigabitEthernet1/0/6] ip address 32.1.1.1 24
[DeviceA-GigabitEthernet1/0/6] ipv6 address 32:1::1 64
[DeviceA-GigabitEthernet1/0/6] quit
```

Configure OSPF.

```
[DeviceA] ospf 1 router-id 11.1.1.1
[DeviceA-ospf-1] area 0
[DeviceA-ospf-1-area-0.0.0.0] quit
[DeviceA-ospf-1] quit
[DeviceA] interface vlan-interface 102
[DeviceA-Vlan-interface102] ospf 1 area 0
[DeviceA-Vlan-interface102] ospf peer sub-address enable 10.102.0.3
[DeviceA-Vlan-interface102] quit
[DeviceA] interface gigabitethernet 1/0/6
[DeviceA-GigabitEthernet1/0/6] ospf 1 area 0
[DeviceA-GigabitEthernet1/0/6] quit
```

Configure OSPFv3.

```
[DeviceA] ospfv3 1
[DeviceA-ospfv3-1] router-id 10.1.1.1
[DeviceA-ospfv3-1] area 0
[DeviceA-ospfv3-1-area-0.0.0.0] quit
[DeviceA-ospfv3-1] quit
[DeviceA] interface vlan-interface 102
[DeviceA-Vlan-interface102] ospfv3 1 area 0
[DeviceA-Vlan-interface102] quit
[DeviceA] interface gigabitethernet 1/0/6
[DeviceA-GigabitEthernet1/0/6] ospfv3 1 area 0
[DeviceA-GigabitEthernet1/0/6] quit
```

Configure BGP.

```
[DeviceA] bgp 100
[DeviceA-bgp-default] router-id 12.1.1.1
[DeviceA-bgp-default] peer 32.1.1.2 as-number 100
[DeviceA-bgp-default] peer 10.103.0.2 as-number 100
[DeviceA-bgp-default] peer 10.103.0.2 source-address 10.103.0.3
[DeviceA-bgp-default] address-family ipv4
[DeviceA-bgp-default-ipv4] peer 10.103.0.2 enable
[DeviceA-bgp-default-ipv4] network 10.103.0.0  255.255.255.0
[DeviceA-bgp-default-ipv4] peer 32.1.1.2 enable
[DeviceA-bgp-default-ipv4] network 32.1.1.0 255.255.255.0
[DeviceA-bgp-default-ipv4] quit
```

Configure BGP4+.

```
[DeviceA-bgp-default] peer 32:1::2 as-number 100
[DeviceA-bgp-default] peer 10:103::2 as-number 100
[DeviceA-bgp-default] peer 10:103::2 source-address 10:103::3
[DeviceA-bgp-default] address-family ipv6
```

```
[DeviceA-bgp-default-ipv6] peer 10:103::2 enable
[DeviceA-bgp-default-ipv6] network 10:103::2  64
[DeviceA-bgp-default-ipv6] peer 32:1::2 enable
[DeviceA-bgp-default-ipv6] network  32:1::2 64
[DeviceA-bgp-default-ipv6] quit
[DeviceA-bgp-default] quit
```

4. Configure dual-active gateways on Device B after the M-LAG system is set up:

Create VLAN-interface 102 and assign it an IPv4 address and an IPv6 address for the interface to act as an IPv4 and IPv6 gateway.

```
[DeviceB] interface vlan-interface 102
[DeviceB-Vlan-interface102] ip address 10.102.0.1 24
[DeviceB-Vlan-interface102] ipv6 address 10:102::1 64
```

Assign M-LAG virtual IPv4 and IPv6 addresses to VLAN-interface 102 for Device A to establish OSPF and OSPFv3 neighbor relationships with Device D.

```
[DeviceB-Vlan-interface102] port m-lag virtual-ip 10.102.0.4 24 active
[DeviceB-Vlan-interface102] port m-lag ipv6 virtual-ip FE80::4 link-local
```

Assign a MAC address to VLAN-interface 102.

```
[DeviceB-Vlan-interface102] mac-address 3-3-3
[DeviceB-Vlan-interface102] quit
```

Exclude VLAN-interface 102 from the shutdown action by M-LAG MAD.

```
[DeviceB] m-lag mad exclude interface vlan-interface 102
```

Create VLAN-interface 103 and assign it an IPv4 address and an IPv6 address for the interface to act as an IPv4 and IPv6 gateway.

```
[DeviceB] interface vlan-interface 103
[DeviceB-Vlan-interface103] ip address 10.103.0.1 24
[DeviceB-Vlan-interface103] ipv6 address 10:103::1 64
```

Assign M-LAG virtual IPv4 and IPv6 addresses to VLAN-interface 103 for Device A to establish BGP and BGP4+ peer relationships with Device D.

```
[DeviceB-Vlan-interface103] port m-lag virtual-ip 10.103.0.4 24 active
[DeviceB-Vlan-interface103] port m-lag ipv6 virtual-ip 10:103::4 64 active
```

Assign a MAC address to VLAN-interface 103.

```
[DeviceB-Vlan-interface103] mac-address 4-4-4
[DeviceB-Vlan-interface103] quit
```

Exclude VLAN-interface 103 from the shutdown action by M-LAG MAD.

```
[DeviceB] m-lag mad exclude interface vlan-interface 103
```

Assign an IPv4 address and an IPv6 address to uplink interface GigabitEthernet 1/0/6.

```
[DeviceB] interface gigabitethernet 1/0/6
[DeviceB-GigabitEthernet1/0/6] port link-mode route
[DeviceB-GigabitEthernet1/0/6] ip address 33.1.1.1 24
[DeviceB-GigabitEthernet1/0/6] ipv6 address 33:1::1 64
[DeviceB-GigabitEthernet1/0/6] quit
```

Configure OSPF.

```
[DeviceB] ospf 1 router-id 11.1.1.2
[DeviceB-ospf-1] area 0
[DeviceB-ospf-1-area-0.0.0.0] quit
[DeviceB-ospf-1] quit
[DeviceB] interface vlan-interface 102
[DeviceB-Vlan-interface102] ospf 1 area 0
```

```
[DeviceB-Vlan-interface102] ospf peer sub-address enable 10.102.0.4
[DeviceB-Vlan-interface102] quit
[DeviceB] interface gigabitethernet 1/0/6
[DeviceB-GigabitEthernet1/0/6] ospf 1 area 0
[DeviceB-GigabitEthernet1/0/6] quit
```

Configure OSPFv3.

```
[DeviceB] ospfv3 1
[DeviceB-ospfv3-1] router-id 10.1.1.2
[DeviceB-ospfv3-1] area 0
[DeviceB-ospfv3-1-area-0.0.0.0] quit
[DeviceB-ospfv3-1] quit
[DeviceB] interface vlan-interface 102
[DeviceB-Vlan-interface102] ospfv3 1 area 0
[DeviceB-Vlan-interface102] quit
[DeviceB] interface gigabitethernet 1/0/6
[DeviceB-GigabitEthernet1/0/6] ospfv3 1 area 0
[DeviceB-GigabitEthernet1/0/6] quit
```

Configure BGP.

```
[DeviceB] bgp 100
[DeviceB-bgp-default] router-id 12.1.1.2
[DeviceB-bgp-default] peer 33.1.1.2 as-number 100
[DeviceB-bgp-default] peer 10.103.0.2 as-number 100
[DeviceB-bgp-default] peer 10.103.0.2 source-address 10.103.0.4
[DeviceB-bgp-default] address-family ipv4
[DeviceB-bgp-default-ipv4] peer 10.103.0.2 enable
[DeviceB-bgp-default-ipv4] network 10.103.0.0  255.255.255.0
[DeviceB-bgp-default-ipv4] peer 33.1.1.2 enable
[DeviceB-bgp-default-ipv4] network 33.1.1.0 255.255.255.0
[DeviceB-bgp-default-ipv4] quit
```

Configure BGP4+.

```
[DeviceB-bgp-default] peer 33:1::2 as-number 100
[DeviceB-bgp-default] peer 10:103::2 as-number 100
[DeviceB-bgp-default] peer 10:103::2 source-address 10:103::4
[DeviceB-bgp-default] address-family ipv6
[DeviceB-bgp-default-ipv6] peer 10:103::2 enable
[DeviceB-bgp-default-ipv6] network 10:103::2  64
[DeviceB-bgp-default-ipv6] peer 33:1::2 enable
[DeviceB-bgp-default-ipv6] network  33:1::2 64
[DeviceB-bgp-default-ipv6] quit
[DeviceB-bgp-default] quit
```

5. Configure Device C:

Assign IPv4 and IPv6 addresses to both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] port link-mode route
[DeviceC-GigabitEthernet1/0/1] ip address 32.1.1.2 24
[DeviceC-GigabitEthernet1/0/1] ipv6 address 32:1::2 64
```

```
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] port link-mode route
[DeviceC-GigabitEthernet1/0/2] ip address 33.1.1.2 24
[DeviceC-GigabitEthernet1/0/2] ipv6 address 33:1::2 64
[DeviceC-GigabitEthernet1/0/2] quit
```

Configure OSPF.

```
[DeviceC] ospf 1 router-id 11.1.1.3
[DeviceC-ospf-1] area 0
[DeviceC-ospf-1-area-0.0.0.0] quit
[DeviceC-ospf-1] quit
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] ospf 1 area 0
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] ospf 1 area 0
[DeviceC-GigabitEthernet1/0/2] quit
```

Configure OSPFv3.

```
[DeviceC] ospfv3 1
[DeviceC-ospfv3-1] router-id 10.1.1.3
[DeviceC-ospfv3-1] area 0
[DeviceC-ospfv3-1-area-0.0.0.0] quit
[DeviceC-ospfv3-1] quit
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] ospfv3 1 area 0
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] ospfv3 1 area 0
[DeviceC-GigabitEthernet1/0/2] quit
```

Configure BGP.

```
[DeviceC] bgp 100
[DeviceC-bgp-default] router-id 12.1.1.3
[DeviceC-bgp-default] peer 33.1.1.1 as-number 100
[DeviceC-bgp-default] peer 32.1.1.1 as-number 100
[DeviceC-bgp-default] address-family ipv4
[DeviceC-bgp-default-ipv4] peer 33.1.1.1 enable
[DeviceC-bgp-default-ipv4] network 33.1.1.0  255.255.255.0
[DeviceC-bgp-default-ipv4] peer 32.1.1.1 enable
[DeviceC-bgp-default-ipv4] network 32.1.1.0 255.255.255.0
[DeviceC-bgp-default-ipv4] quit
```

Configure BGP4+.

```
[DeviceC-bgp-default] peer 33:1::1 as-number 100
[DeviceC-bgp-default] peer 32:1::1 as-number 100
[DeviceC-bgp-default] address-family ipv6
[DeviceC-bgp-default-ipv6] peer 32:1::1 enable
[DeviceC-bgp-default-ipv6] peer 33:1::1 enable
[DeviceC-bgp-default-ipv6] network 32:1:: 64
[DeviceC-bgp-default-ipv6] network 33:1:: 64
```

```
[DeviceC-bgp-default-ipv6] quit
[DeviceC-bgp-default] quit
```

6. Configure Device D:

Create VLAN 102 and VLAN 103.

```
<Sysname> system-view
[Sysname] sysname DeviceD
[DeviceD] vlan 102 103
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3.

```
[DeviceD] interface bridge-aggregation 3
[DeviceD-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceD-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to aggregation group 3.

```
[DeviceD] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[DeviceD-if-range] port link-aggregation group 3
[DeviceD-if-range] quit
```

Configure Bridge-Aggregation 3 as a trunk port, and assign it to VLAN 102 and VLAN 103.

```
[DeviceD] interface bridge-aggregation 3
[DeviceD-Bridge-Aggregation3] port link-type trunk
[DeviceD-Bridge-Aggregation3] port trunk permit vlan 102 103
[DeviceD-Bridge-Aggregation3] quit
```

Create VLAN-interface102 and assign IPv4 and IPv6 addresses to it.

```
[DeviceD] interface vlan-interface 102
[DeviceD-Vlan-interface102] ip address 10.102.0.2 24
[DeviceD-Vlan-interface102] ipv6 address 10:102::2 64
[DeviceD-Vlan-interface102] quit
```

Create VLAN-interface103 and assign IPv4 and IPv6 addresses to it.

```
[DeviceD] interface vlan-interface 103
[DeviceD-Vlan-interface103] ip address 10.103.0.2 24
[DeviceD-Vlan-interface103] ipv6 address 10:103::2 64
[DeviceD-Vlan-interface103] quit
```

Configure OSPF.

```
[DeviceD] ospf 1 router-id 11.1.1.4
[DeviceD-ospf-1] area 0
[DeviceD-ospf-1-area-0.0.0.0] quit
[DeviceD-ospf-1] quit
[DeviceD] interface vlan-interface 102
[DeviceD-Vlan-interface102] ospf 1 area 0
[DeviceD-Vlan-interface102] quit
[DeviceD] interface vlan-interface 103
[DeviceD-Vlan-interface103] ospf 1 area 0
[DeviceD-Vlan-interface103] quit
```

Configure OSPFv3.

```
[DeviceD] ospfv3 1
[DeviceD-ospfv3-1] router-id 10.1.1.4
[DeviceD-ospfv3-1] area 0
[DeviceD-ospfv3-1-area-0.0.0.0] quit
[DeviceD-ospfv3-1] quit
```

```
[DeviceD] interface vlan-interface 102
[DeviceD-Vlan-interface102] ospfv3 1 area 0
[DeviceD-Vlan-interface102] quit
[DeviceD] interface vlan-interface 103
[DeviceD-Vlan-interface103] ospfv3 1 area 0
[DeviceD-Vlan-interface103] quit
```

Configure BGP.

```
[DeviceD] bgp 100
[DeviceD-bgp-default] router-id 12.1.1.4
[DeviceD-bgp-default] peer 10.103.0.3 as-number 100
[DeviceD-bgp-default] peer 10.103.0.4 as-number 100
[DeviceD-bgp-default] address-family ipv4
[DeviceD-bgp-default-ipv4] peer 10.103.0.3 enable
[DeviceD-bgp-default-ipv4] peer 10.103.0.4 enable
[DeviceD-bgp-default-ipv4] network 10.103.0.3  255.255.255.0
[DeviceD-bgp-default-ipv4] quit
```

Configure BGP4+.

```
[DeviceD-bgp-default] peer 10:103::3 as-number 100
[DeviceD-bgp-default] peer 10:103::4 as-number 100
[DeviceD-bgp-default] address-family ipv6
[DeviceD-bgp-default-ipv6] peer 10:103::3  enable
[DeviceD-bgp-default-ipv6] peer 10:103::4 enable
[DeviceD-bgp-default-ipv6] network 10:103::3 64
[DeviceD-bgp-default-ipv6] quit
[DeviceD-bgp-default] quit
```

Verifying the configuration

The following tasks use Device A as an example.

Verify that Device A has formed an M-LAG system with Device B.

```
[DeviceA] display m-lag summary
Flags: A -- Aggregate interface down, B -- No peer M-LAG interface configured
       C -- Configuration consistency check failed

Peer-link interface: BAGG1
Peer-link interface state (cause): UP
Keepalive link state (cause): UP

                    M-LAG interface information
M-LAG IF      M-LAG group  Local state (cause)  Peer state  Remaining down time(s)
BAGG3         1            UP                   UP          -
[DeviceA] display m-lag verbose
Flags: A -- Home_Gateway, B -- Neighbor_Gateway, C -- Other_Gateway,
       D -- PeerLink_Activity, E -- DRCP_Timeout, F -- Gateway_Sync,
       G -- Port_Sync, H -- Expired

Peer-link interface/Peer-link interface ID: BAGG1/1
State: UP
Cause: -
```

```
Local DRCP flags/Peer DRCP flags: ABDFG/ABDFG
Local Selected ports (index): GE1/0/3 (260), GE1/0/4 (261)
Peer Selected ports indexes: 260, 261

M-LAG interface/M-LAG group ID: BAGG3/1
Local M-LAG interface state: UP
Peer M-LAG interface state: UP
M-LAG group state: UP
Local M-LAG interface down cause: -
Remaining M-LAG DOWN time: -
Local M-LAG interface LACP MAC: Config=N/A, Effective=0002-0002-0002
Peer M-LAG interface LACP MAC: Config=N/A, Effective=0002-0002-0002
Local M-LAG interface LACP priority: Config=32768, Effective=123
Peer M-LAG interface LACP priority: Config=32768, Effective=123
Local DRCP flags/Peer DRCP flags: ABDFG/ABDFG
Local Selected ports (index): GE1/0/1 (258), GE1/0/2 (259)
Peer Selected ports indexes: 258, 259
```

Verify that Device A have established OSPF neighbor relationships.

```
[DeviceA] display ospf peer

         OSPF Process 1 with Router ID 11.1.1.1
               Neighbor Brief Information

 Area: 0.0.0.0
 Router ID       Address         Pri Dead-Time  State             Interface
 11.1.1.3        32.1.1.2        1   39         Full/BDR          GE1/0/6
 11.1.1.4        10.102.0.2      1   31         Full/DROther      Vlan102
 11.1.1.2        10.102.0.4      1   37         Full/DR           Vlan102
```

Verify that Device A have established OSPFv3 neighbor relationships.

```
[DeviceA] display ospfv3 peer

               OSPFv3 Process 1 with Router ID 10.1.1.1

 Area: 0.0.0.0
--------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 10.1.1.3        1   Full/BDR          00:00:34  0      GE1/0/6
 10.1.1.2        1   Full/BDR          00:00:32  0      Vlan102
 10.1.1.4        1   Full/DROther      00:00:39  0      Vlan102
```

Verify that Device A have established BGP peer relationships.

```
[DeviceA] display bgp peer ipv4

 BGP local router ID: 12.1.1.1
 Local AS number: 100
 Total number of peers: 2                 Peers in established state: 2

  * - Dynamically created peer
  ^ - Peer created through link-local address
```

```
 Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

 10.103.0.2             100        6        6    0       1 00:01:50 Established
 32.1.1.2               100        5        5    0       2 00:00:01 Established
```

Verify that Device A have established BGP4+ peer relationships.

```
[DeviceA] display bgp peer ipv6

 BGP local router ID: 12.1.1.1
 Local AS number: 100
 Total number of peers: 2                Peers in established state: 2

  * - Dynamically created peer
  ^ - Peer created through link-local address
  Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

  10:103::2              100        9        9    0       1 00:04:26 Established
  32:1::2                100        6        6    0       2 00:02:13 Established
```

Configuration files

- Device A:

```
 sysname DeviceA
#
ospf 1 router-id 11.1.1.1
 area 0.0.0.0
#
ospfv3 1
 router-id 10.1.1.1
 area 0.0.0.0
#
 undo mac-address static source-check enable
#
vlan 102 to 103
#
interface Bridge-Aggregation1
 port link-type trunk
 port trunk permit vlan all
 link-aggregation mode dynamic
 port m-lag peer-link 1
 undo mac-address static source-check enable
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 link-aggregation mode dynamic
 port m-lag group 1
#
interface Vlan-interface102
 ip address 10.102.0.1 255.255.255.0
 ospf 1 area 0.0.0.0
```

```
 ospf peer sub-address enable 10.102.0.3
 ospfv3 1 area 0.0.0.0
 port m-lag virtual-ip 10.102.0.3 255.255.255.0 active
 port m-lag ipv6 virtual-ip FE80::2 link-local
 mac-address 0003-0003-0003
 ipv6 address 10:102::1/64
#
interface Vlan-interface103
 ip address 10.103.0.1 255.255.255.0
 port m-lag virtual-ip 10.103.0.3 255.255.255.0 active
 port m-lag ipv6 virtual-ip 10:103::3 64 active
 mac-address 0004-0004-0004
 ipv6 address 10:103::1/64
#
interface GigabitEthernet1/0/5
 port link-mode route
 ip address 21.1.1.1 255.255.255.0
 ipv6 address 21::1/64
#
interface GigabitEthernet1/0/6
 port link-mode route
 ip address 32.1.1.1 255.255.255.0
 ospf 1 area 0.0.0.0
 ospfv3 1 area 0.0.0.0
 ipv6 address 32:1::1/64
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 port link-aggregation group 3
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 port link-aggregation group 3
#
interface GigabitEthernet1/0/3
 port link-aggregation group 1
#
interface GigabitEthernet1/0/4
 port link-aggregation group 1
#
bgp 100
 router-id 12.1.1.1
 peer 10.103.0.2 as-number 100
 peer 10.103.0.2 source-address 10.103.0.3
 peer 32.1.1.2 as-number 100
 peer 10:103::2 as-number 100
```

```
 peer 10:103::2 source-address 10:103::3
 peer 32:1::2 as-number 100
 #
 address-family ipv4 unicast
  network 10.103.0.0 255.255.255.0
  network 32.1.1.0 255.255.255.0
  peer 10.103.0.2 enable
  peer 32.1.1.2 enable
 #
 address-family ipv6 unicast
  network 10:103:: 64
  network 32:1:: 64
  peer 10:103::2 enable
  peer 32:1::2 enable
#
 m-lag mad exclude interface GigabitEthernet1/0/5
 m-lag mad exclude interface Vlan-interface102
 m-lag mad exclude interface Vlan-interface103
 m-lag system-mac 0002-0002-0002
 m-lag system-number 1
 m-lag system-priority 123
 m-lag keepalive ip destination 21.1.1.2 source 21.1.1.1
#
```

- Device B:

```
sysname DeviceB
#
ospf 1 router-id 11.1.1.2
 area 0.0.0.0
#
ospfv3 1
 router-id 10.1.1.2
 area 0.0.0.0
#
 undo mac-address static source-check enable
#
vlan 102 to 103
#
interface Bridge-Aggregation1
 port link-type trunk
 port trunk permit vlan all
 link-aggregation mode dynamic
 port m-lag peer-link 1
 undo mac-address static source-check enable
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 102 to 103
```

```
 link-aggregation mode dynamic
 port m-lag group 1
#
interface Vlan-interface102
 ip address 10.102.0.1 255.255.255.0
 ospf 1 area 0.0.0.0
 ospf peer sub-address enable 10.102.0.4
 ospfv3 1 area 0.0.0.0
 port m-lag virtual-ip 10.102.0.4 255.255.255.0 active
 port m-lag ipv6 virtual-ip FE80::4 link-local
 mac-address 0003-0003-0003
 ipv6 address 10:102::1/64
#
interface Vlan-interface103
 ip address 10.103.0.1 255.255.255.0
 port m-lag virtual-ip 10.103.0.4 255.255.255.0 active
 port m-lag ipv6 virtual-ip 10:103::4 64 active
 mac-address 0004-0004-0004
 ipv6 address 10:103::1/64
#
interface GigabitEthernet1/0/5
 port link-mode route
 ip address 21.1.1.2 255.255.255.0
 ipv6 address 21::2/64
#
interface GigabitEthernet1/0/6
 port link-mode route
 ip address 33.1.1.1 255.255.255.0
 ospf 1 area 0.0.0.0
 ospfv3 1 area 0.0.0.0
 ipv6 address 33:1::1/64
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 port link-aggregation group 3
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 port link-aggregation group 3
#
interface GigabitEthernet1/0/3
 port link-aggregation group 1
#
interface GigabitEthernet1/0/4
 port link-aggregation group 1
#
```

```
bgp 100
 router-id 12.1.1.2
 peer 10.103.0.2 as-number 100
 peer 10.103.0.2 source-address 10.103.0.4
 peer 33.1.1.2 as-number 100
 peer 10:103::2 as-number 100
 peer 10:103::2 source-address 10:103::4
 peer 33:1::2 as-number 100
 #
 address-family ipv4 unicast
  network 10.103.0.0 255.255.255.0
  network 33.1.1.0 255.255.255.0
  peer 10.103.0.2 enable
  peer 33.1.1.2 enable
 #
 address-family ipv6 unicast
  network 10:103:: 64
  network 33:1:: 64
  peer 10:103::2 enable
  peer 33:1::2 enable
#
 m-lag mad exclude interface GigabitEthernet1/0/5
 m-lag mad exclude interface Vlan-interface102
 m-lag mad exclude interface Vlan-interface103
 m-lag system-mac 0002-0002-0002
 m-lag system-number 2
 m-lag system-priority 123
 m-lag keepalive ip destination 21.1.1.1 source 21.1.1.2
#
```

- Device C:

```
#
 sysname DeviceC
#
ospf 1 router-id 11.1.1.3
 area 0.0.0.0
#
ospfv3 1
 router-id 10.1.1.3
 area 0.0.0.0
#
interface GigabitEthernet1/0/1
 port link-mode route
 combo enable fiber
 ip address 32.1.1.2 255.255.255.0
 ospf 1 area 0.0.0.0
 ospfv3 1 area 0.0.0.0
 ipv6 address 32:1::2/64
```

```
#
interface GigabitEthernet1/0/2
 port link-mode route
 combo enable fiber
 ip address 33.1.1.2 255.255.255.0
 ospf 1 area 0.0.0.0
 ospfv3 1 area 0.0.0.0
 ipv6 address 33:1::2/64
#
bgp 100
 router-id 12.1.1.3
 peer 32.1.1.1 as-number 100
 peer 33.1.1.1 as-number 100
 peer 32:1::1 as-number 100
 peer 33:1::1 as-number 100
 #
 address-family ipv4 unicast
  network 32.1.1.0 255.255.255.0
  network 33.1.1.0 255.255.255.0
  peer 32.1.1.1 enable
  peer 33.1.1.1 enable
 #
 address-family ipv6 unicast
  network 32:1:: 64
  network 33:1:: 64
  peer 32:1::1 enable
  peer 33:1::1 enable
#
```

- Device D:

```
 sysname DeviceD
#
ospf 1 router-id 11.1.1.4
 area 0.0.0.0
#
ospfv3 1
 router-id 10.1.1.4
 area 0.0.0.0
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 link-aggregation mode dynamic
#
interface Vlan-interface102
 ip address 10.102.0.2 255.255.255.0
 ospf 1 area 0.0.0.0
 ospfv3 1 area 0.0.0.0
```

```
 ipv6 address 10:102::2/64
#
interface Vlan-interface103
 ip address 10.103.0.2 255.255.255.0
 ospf 1 area 0.0.0.0
 ospfv3 1 area 0.0.0.0
 ipv6 address 10:103::2/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 combo enable fiber
 port link-aggregation group 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 combo enable fiber
 port link-aggregation group 3
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 combo enable fiber
 port link-aggregation group 3
#
interface GigabitEthernet1/0/4
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 102 to 103
 combo enable fiber
 port link-aggregation group 3
#
bgp 100
 router-id 12.1.1.4
 peer 10.103.0.3 as-number 100
 peer 10.103.0.4 as-number 100
 peer 10:103::3 as-number 100
 peer 10:103::4 as-number 100
 #
 address-family ipv4 unicast
  network 10.103.0.0 255.255.255.0
  peer 10.103.0.3 enable
  peer 10.103.0.4 enable
 #
```

```
 address-family ipv6 unicast
  network 10:103:: 64
  peer 10:103::3 enable
  peer 10:103::4 enable
#
```

Comparison between M-LAG and DRNI commands

Configuration commands

Feature	View	M-LAG command	DRNI command
MAC address table	System view	**mac-address m-lag forwarding-conversational-learning** **undo mac-address m-lag forwarding-conversational-learning**	**mac-address drni forwarding-conversational-learning** **undo mac-address drni forwarding-conversational-learning**
M-LAG	Any view	**display m-lag consistency** { **type1** | **type2** } { **global** | **interface** *interface-type interface-number* }	**display drni consistency** { **type1** | **type2** } { **global** | **interface** *interface-type interface-number* }
M-LAG	Any view	**display m-lag consistency-check status**	**display drni consistency-check status**
M-LAG	Any view	**display m-lag drcp statistics** [**interface** *interface-type interface-number*]	**display drni drcp statistics** [**interface** *interface-type interface-number*]
M-LAG	Any view	**display m-lag keepalive**	**display drni keepalive**
M-LAG	Any view	**display m-lag mad verbose**	**display drni mad verbose**
M-LAG	Any view	**display m-lag role**	**display drni role**
M-LAG	Any view	**display m-lag summary**	**display drni summary**
M-LAG	Any view	**display m-lag system**	**display drni system**
M-LAG	Any view	**display m-lag troubleshooting** [**m-lag-interface** | **peer-link** | **keepalive**] [**history**] [*count*]	**display drni troubleshooting** [**dr** | **ipp** | **keepalive**] [**history**] [*count*]
M-LAG	Any view	**display m-lag verbose** [**interface** *interface-type interface-number*]	**display drni verbose** [**interface** *interface-type interface-number*]
M-LAG	Any view	**display m-lag virtual-ip** [**interface** *interface-type*	**display drni virtual-ip** [**interface**

		interface-number]	*interface-type interface-number*]
M-LAG	System view	**m-lag authentication key** { **simple** | **cipher** } *string* **undo m-lag authentication key**	**drni authentication key** { **simple** | **cipher** } *string* **undo drni authentication key**
M-LAG	System view	**m-lag auto-recovery reload-delay** *delay-value* **undo m-lag auto-recovery reload-delay**	**drni auto-recovery reload-delay** *delay-value* **undo drni auto-recovery reload-delay**
M-LAG	System view	**m-lag consistency-check disable** **undo m-lag consistency-check disable**	**drni consistency-check disable** **undo drni consistency-check disable**
M-LAG	System view	**m-lag consistency-check mode** { **loose** | **strict** } **undo m-lag consistency-check mode**	**drni consistency-check mode** { **loose** | **strict** } **undo drni consistency-check mode**
M-LAG	Layer 2 aggregate interface view Tunnel interface view	**m-lag drcp period short** **undo m-lag drcp period**	**drni drcp period short** **undo drni drcp period**
M-LAG	System view	**m-lag peer-link mac-address hold** **undo m-lag peer-link mac-address hold**	**drni ipp mac-address hold** **undo drni ipp mac-address hold**
M-LAG	System view	**m-lag keepalive** { **ip** | **ipv6** } **destination** { *ipv4-address* | *ipv6-address* } [**source** { *ipv4-address* | *ipv6-address* } | **udp-port** *udp-number* | **vpn-instance** *vpn-instance-name*] * **undo m-lag keepalive** { **ip** | **ipv6** }	**drni keepalive** { **ip** | **ipv6** } **destination** { *ipv4-address* | *ipv6-address* } [**source** { *ipv4-address* | *ipv6-address* } | **udp-port** *udp-number* | **vpn-instance** *vpn-instance-name*] * **undo drni keepalive** { **ip** | **ipv6** }
M-LAG	System view	**m-lag keepalive hold-time** *value* **undo m-lag keepalive hold-time**	**drni keepalive hold-time** *value* **undo drni keepalive hold-time**
M-LAG	System view	**m-lag keepalive interval** *interval* [**timeout** *timeout*] **undo m-lag keepalive**	**drni keepalive interval** *interval* [**timeout** *timeout*] **undo drni keepalive**

		interval	**interval**
M-LAG	System view	**m-lag mad default-action** { **down** | **none** } **undo m-lag mad default-action**	**drni mad default-action** { **down** | **none** } **undo drni mad default-action**
M-LAG	System view	**m-lag mad exclude interface** *interface-type interface-number* **undo m-lag mad exclude interface** *interface-type interface-number*	**drni mad exclude interface** *interface-type interface-number* **undo drni mad exclude interface** *interface-type interface-number*
M-LAG	System view	**m-lag mad exclude logical-interfaces** **undo m-lag mad exclude logical-interfaces**	**drni mad exclude logical-interfaces** **undo drni mad exclude logical-interfaces**
M-LAG	System view	**m-lag mad include interface** *interface-type interface-number* **undo m-lag mad include interface** *interface-type interface-number*	**drni mad include interface** *interface-type interface-number* **undo drni mad include interface** *interface-type interface-number*
M-LAG	System view	**m-lag mad persistent** **undo m-lag mad persistent**	**drni mad persistent** **undo drni mad persistent**
M-LAG	System view	**m-lag mad restore**	**drni mad restore**
M-LAG	System view	**m-lag restore-delay** *value* **undo m-lag restore-delay**	**drni restore-delay** *value* **undo drni restore-delay**
M-LAG	System view	**m-lag role priority** *priority-value* **undo m-lag role priority**	**drni role priority** *priority-value* **undo drni role priority**
M-LAG	System view	**m-lag sequence enable** **undo m-lag sequence enable**	**drni sequence enable** **undo drni sequence enable**
M-LAG	System view	**m-lag standalone enable** [**delay** *delay-time*] **undo m-lag standalone enable** [**delay**]	**drni standalone enable** [**delay** *delay-time*] **undo drni standalone enable** [**delay**]
M-LAG	System view	**m-lag system-mac** { *bridge-address* | *mac-address* } **undo m-lag system-mac**	**drni system-mac** { *bridge-address* | *mac-address* } **undo drni system-mac**
M-LAG	System view	**m-lag system-number** *system-number* **undo m-lag system-number**	**drni system-number** *system-number* **undo drni system-number**

M-LAG	System view	**m-lag system-priority** *priority* **undo m-lag system-priority**	**drni system-priority** *priority* **undo drni system-priority**
M-LAG	Layer 2 aggregate interface view	**port m-lag group** *group-id* [**allow-single-member**] **undo port m-lag group**	**port drni group** *group-id* [**allow-single-member**] **undo port drni group**
M-LAG	Layer 2 aggregate interface view Tunnel interface view	**port m-lag peer-link** *port-id* **undo port m-lag peer-link**	**port drni intra-portal-port** *port-id* **undo port drni intra-portal-port**
M-LAG	VLAN interface view	**port m-lag ipv6 virtual-ip** *ipv6-address* { *prefix-length* [**active** | **standby**] [**virtual-mac** *mac-address*] | **link-local** } **undo port m-lag ipv6 virtual-ip** [*ipv6-address*]	**port drni ipv6 virtual-ip** *ipv6-address* { *prefix-length* [**active** | **standby**] [**virtual-mac** *mac-address*] | **link-local** } **undo port drni ipv6 virtual-ip** [*ipv6-address*]
M-LAG	Loopback interface view	**port m-lag ipv6 virtual-ip** *ipv6-address* { *prefix-length* [**active** | **standby**] | **link-local** } **undo port m-lag ipv6 virtual-ip** [*ipv6-address*]	**port drni ipv6 virtual-ip** *ipv6-address* { *prefix-length* [**active** | **standby**] | **link-local** } **undo port drni ipv6 virtual-ip** [*ipv6-address*]
M-LAG	VSI interface view	**port m-lag ipv6 virtual-ip** *ipv6-address* { *prefix-length* [**active** | **standby**] | **link-local** } **undo port m-lag ipv6 virtual-ip** [*ipv6-address*]	**port drni ipv6 virtual-ip** *ipv6-address* { *prefix-length* [**active** | **standby**] | **link-local** } **undo port drni ipv6 virtual-ip** [*ipv6-address*]
M-LAG	Layer 2 aggregate interface view	**port m-lag system-mac** *mac-address* **undo port m-lag system-mac**	**port drni system-mac** *mac-address* **undo port drni system-mac**
M-LAG	Layer 2 aggregate interface view	**port m-lag system-priority** *priority* **undo port m-lag system-priority**	**port drni system-priority** *priority* **undo port drni system-priority**
M-LAG	VLAN interface view	**port m-lag virtual-ip** *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**] [**virtual-mac**	**port drni virtual-ip** *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**] [**virtual-mac**

		mac-address] **undo port m-lag virtual-ip** [*ipv4-address*]	*mac-address*] **undo port drni virtual-ip** [*ipv4-address*]
M-LAG	Loopback interface view	**port m-lag virtual-ip** *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**] **undo port m-lag virtual-ip** [*ipv4-address*]	**port drni virtual-ip** *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**] **undo port drni virtual-ip** [*ipv4-address*]
M-LAG	VSI interface view	**port m-lag virtual-ip** *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**] **undo port m-lag virtual-ip** [*ipv4-address*]	**port drni virtual-ip** *ipv4-address* { *mask-length* | *mask* } [**active** | **standby**] **undo port drni virtual-ip** [*ipv4-address*]
M-LAG	User view	**reset m-lag drcp statistics** [**interface** *interface-list*]	**reset drni drcp statistics** [**interface** *interface-list*]
M-LAG	User view	**reset m-lag troubleshooting history**	**reset drni troubleshooting history**
DHCP	Any view	**display dhcp snooping m-lag-statistics** [**old-version**]	**display dhcp snooping drni-statistics** [**old-version**]
DHCP	Any view	**display dhcp snooping m-lag-status**	**display dhcp snooping drni-status**
DHCP	User view	**reset dhcp snooping m-lag-statistics**	**reset dhcp snooping drni-statistics**
DHCPv6	Any view	**display ipv6 dhcp snooping m-lag-status**	**display ipv6 dhcp snooping drni-status**
DHCPv6	Any view	**display ipv6 dhcp snooping m-lag-statistics** [**old-version**]	**display ipv6 dhcp snooping drni-statistics** [**old-version**]
DHCPv6	User view	**reset ipv6 dhcp snooping m-lag-statistics**	**reset ipv6 dhcp snooping drni-statistics**
AAA	RADIUS scheme view	**nas-ip** [**m-lag** { **local** | **peer** }] { *ipv4-address* | **interface** *interface-type interface-number* | **ipv6** *ipv6-address* } **undo nas-ip** [**m-lag** { **local** | **peer** }] [**interface** | **ipv6**]	**nas-ip** [**drni** { **local** | **peer** }] { *ipv4-address* | **interface** *interface-type interface-number* | **ipv6** *ipv6-address* } **undo nas-ip** [**drni** { **local** | **peer** }] [**interface** | **ipv6**]

802.1X	Any view	**display dot1x connection** [**open**] [[**m-lag** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **slot** *slot-number* | **user-name** *name-string*] | **user-mac** *mac-address*]	**display dot1x connection** [**open**] [[**drni** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **slot** *slot-number* | **user-name** *name-string*] | **user-mac** *mac-address*]
MAC authentication	Any view	**display mac-authentication connection** [**open**] [[**m-lag** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **slot** *slot-number* | **user-name** *user-name*] | **user-mac** *mac-address*]	**display mac-authentication connection** [**open**] [[**drni** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **slot** *slot-number* | **user-name** *user-name*] | **user-mac** *mac-address*]
Port security	System view	**port-security m-lag load-sharing-mode** { **centralized** | **distributed** { **even-mac** | **local** | **odd-mac** } } **undo port-security m-lag load-sharing-mode**	**port-security drni load-sharing-mode** { **centralized** | **distributed** { **even-mac** | **local** | **odd-mac** } } **undo port-security drni load-sharing-mode**
Port security	Any view	**display port-security static-user connection** [[**m-lag** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **slot** *slot-number* | **user-name** *user-name*] | { **ip** | **ipv6** } *ip-address* | **mac** *mac-address*]	**display port-security static-user connection** [[**drni** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **slot** *slot-number* | **user-name** *user-name*] | { **ip** | **ipv6** } *ip-address* | **mac** *mac-address*]

Port security	Any view	**display port-security access-user** [**m-lag** [**local** | **peer**]] [**access-type** { **dot1x** | **mac-auth** | **web-auth** | **static** } | **domain** *domain-name* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **slot** *slot-number*] *	**display port-security access-user** [**drni** [**local** | **peer**]] [**access-type** { **dot1x** | **mac-auth** | **web-auth** | **static** } | **domain** *domain-name* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **slot** *slot-number*] *
Portal	System view	**portal m-lag load-sharing-mode** { **centralized** | **distributed** { **even-ip** | **odd-ip** } } **undo portal m-lag load-sharing-mode**	**portal drni load-sharing-mode** { **centralized** | **distributed** { **even-ip** | **odd-ip** } } **undo portal drni load-sharing-mode**
Portal	System view	**portal m-lag traffic backup** { **interval** *interval-value* | **threshold** *threshold-value* }* **undo portal m-lag traffic backup**	**portal drni traffic backup** { **interval** *interval-value* | **threshold** *threshold-value* }* **undo portal drni traffic backup**
Track	System view	**track** *track-entry-number* **mlag-mad-status** **undo** *track track-entry-number*	**track** *track-entry-number* **drni-mad-status** **undo track** *track-entry-number*
VXLAN	System view	**display l2vpn m-lag arp** [**local** | **remote**] [**count** | { **public-instance** | **vpn-instance** *vpn-instance-name* } [**count** | *ip-address*]]	**display l2vpn drni arp** [**local** | **remote**] [**count** | { **public-instance** | **vpn-instance** *vpn-instance-name* } [**count** | *ip-address*]]
VXLAN	System view	**display l2vpn m-lag arp suppression** [**local** | **remote**] [**count** | **vsi** *vsi-name* [**count** | *ip-address*]]	**display l2vpn drni arp suppression** [**local** | **remote**] [**count** | **vsi** *vsi-name* [**count** | *ip-address*]]
VXLAN	System view	**display l2vpn m-lag mac-address** [**local** | **remote**] [**count** | **vsi** *vsi-name* [**count** | *mac-address*]]	**display l2vpn drni mac-address** [**local** | **remote**] [**count** | **vsi** *vsi-name* [**count** | *mac-address*]]

VXLAN	System view	**display l2vpn m-lag nd** [**local** | **remote**] [**count** | { **public-instance** | **vpn-instance** *vpn-instance-name* } [*ipv6-address* | **count**]]	**display l2vpn drni nd** [**local** | **remote**] [**count** | { **public-instance** | **vpn-instance** *vpn-instance-name* } [*ipv6-address* | **count**]]
VXLAN	System view	**display l2vpn m-lag nd suppression** [**local** | **remote**] [**count** | **vsi** *vsi-name* [*ipv6-address* | **count**]]	**display l2vpn drni nd suppression** [**local** | **remote**] [**count** | **vsi** *vsi-name* [*ipv6-address* | **count**]]
EVPN	System view	**display evpn m-lag synchronized-mac** [**vsi** *vsi-name*] [**count**]	**display evpn drni synchronized-mac** [**vsi** *vsi-name*] [**count**]
EVPN	System view	**evpn m-lag group** { *virtual-vtep-ipv4* | *virtual-vtep-ipv6* } **undo evpn m-lag group**	**evpn drni group** { *virtual-vtep-ipv4* | *virtual-vtep-ipv6* } **undo evpn drni group**
EVPN	System view	**evpn m-lag local** { *local-ipv4-address* **remote** *remote-ipv4-address* | *local-ipv6-address* **remote** *remote-ipv6-address* } **undo evpn m-lag local**	**evpn drni local** { *local-ipv4-address* **remote** *remote-ipv4-address* | *local-ipv6-address* **remote** *remote-ipv6-address* } **undo evpn drni local**
EVPN	System view	**l2vpn m-lag peer-link ac-match-rule vxlan-mapping** **undo l2vpn m-lag peer-link ac-match-rule vxlan-mapping**	**l2vpn drni peer-link ac-match-rule vxlan-mapping** **undo l2vpn drni peer-link ac-match-rule vxlan-mapping**

Contents

Configuring port isolation 1
About port isolation 1
Assigning a port to an isolation group 1
Display and maintenance commands for port isolation 2
Port isolation configuration examples 2
Example: Configuring port isolation 2

Configuring port isolation

About port isolation

The port isolation feature isolates Layer 2 traffic for data privacy and security without using VLANs.

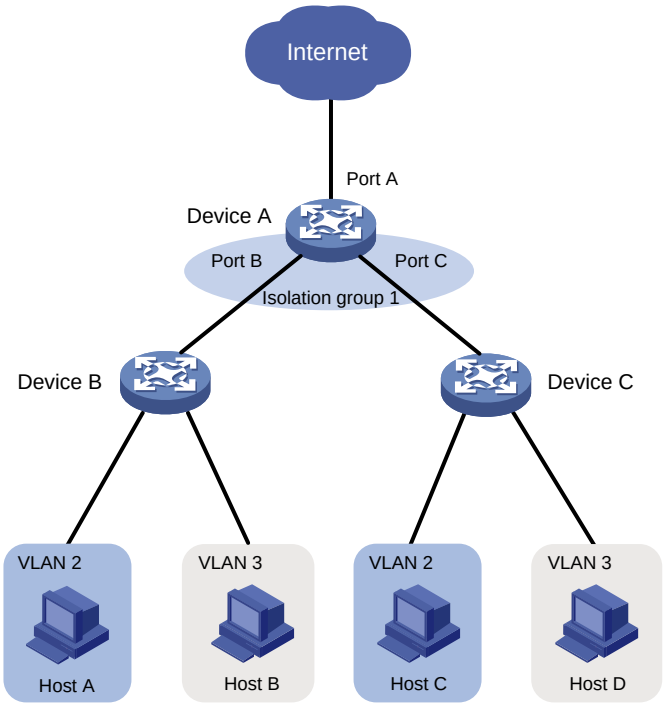
Ports in an isolation group cannot communicate with each other. However, they can communicate with ports outside the isolation group.

You can configure community VLANs in an isolation group. Ports in an isolation group can communicate with each other if they belong to a community VLAN.

As shown in Figure 1:

- To isolate Host A from Host C, and Host B from Host D, assign Port B and Port C on Device A to isolation group 1.
- To enable Layer 2 communication between Host B and Host D, specify VLAN 3 as a community VLAN in isolation group 1. Host B and Host D belong to VLAN 3.

Figure 1 Community VLANs in an isolation group



Assigning a port to an isolation group

About this task

The device supports multiple isolation groups, which can be configured manually. The number of ports assigned to an isolation group is not limited.

Restrictions and guidelines

- You can assign a port to only one isolation group.
- The configuration in Layer 2 Ethernet interface view applies only to the interface.

- The configuration in Layer 2 aggregate interface view applies to the Layer 2 aggregate interface and its aggregation member ports. If the device fails to apply the configuration to the aggregate interface, it does not assign any aggregation member port to the isolation group. If the failure occurs on an aggregation member port, the device skips the port and continues to assign other aggregation member ports to the isolation group.
- Port isolation is not supported on an IRF fabric in the current software version.

Procedure

1. Enter system view.
   system-view
2. Create an isolation group.
   port-isolate group *group-id* [**peer-link**]
3. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
4. Assign the port to the isolation group.
   port-isolate enable group *group-id*
   By default, the port is not in any isolation group.

Display and maintenance commands for port isolation

Execute **display** commands in any view.

Task	Command
Display isolation group information.	**display port-isolate group** [*group-id*]

Port isolation configuration examples

Example: Configuring port isolation

Network configuration

As shown in Figure 2:

- LAN users Host A, Host B, and Host C are connected to GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 on the device, respectively.
- The device connects to the Internet through GigabitEthernet 1/0/4.

Configure the device to provide Internet access for the hosts, and isolate them from one another at Layer 2.

Figure 2 Network diagram

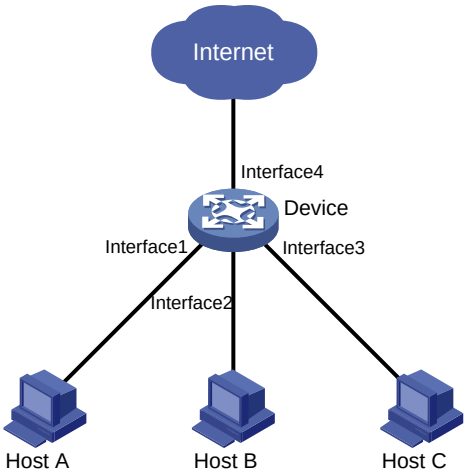


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

Create isolation group 2.

```
<Device> system-view
[Device] port-isolate group 2
```

Assign GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 to isolation group 2.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] port-isolate enable group 2
[Device-GigabitEthernet1/0/1] quit
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] port-isolate enable group 2
[Device-GigabitEthernet1/0/2] quit
[Device] interface gigabitethernet 1/0/3
[Device-GigabitEthernet1/0/3] port-isolate enable group 2
[Device-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Display information about isolation group 2.

```
[Device] display port-isolate group 2
 Port isolation group information:
 Group ID: 2
 Group members:
    GigabitEthernet1/0/1          GigabitEthernet1/0/2
    GigabitEthernet1/0/3
```

The output shows that GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 are assigned to isolation group 2. As a result, Host A, Host B, and Host C are isolated from one another at layer 2.

Contents

Configuring VLANs 1
About VLANs 1
VLAN frame encapsulation 1
VLAN types 2
Port-based VLANs 2
MAC-based VLANs 3
IP subnet-based VLANs 5
Protocol-based VLANs 6
Layer 3 communication between VLANs 6
Protocols and standards 6
Configuring a VLAN 6
Restrictions and guidelines 6
VLAN configuration tasks at a glance 6
Creating VLANs 6
Enabling packet statistics for the VLAN 7
Configuring port-based VLANs 7
Restrictions and guidelines for port-based VLANs 7
Assigning an access port to a VLAN 8
Assigning a trunk port to a VLAN 8
Assigning a hybrid port to a VLAN 9
Configuring MAC-based VLANs 10
Restrictions and guidelines for MAC-based VLANs 10
Configuring static MAC-based VLAN assignment 10
Configuring dynamic MAC-based VLAN assignment 10
Configuring server-assigned MAC-based VLAN 12
Enabling ARP flood suppression 12
Configuring IP subnet-based VLANs 13
Configuring protocol-based VLANs 14
Configuring a VLAN group 15
Configuring VLAN interfaces 15
Restrictions and guidelines 15
VLAN interfaces configuration tasks at a glance 15
Prerequisites 15
Creating a VLAN interface 16
Restoring the default settings for the VLAN interface 16
Enabling the function of setting VLAN tags for sent protocol packets on an interface 17
Display and maintenance commands for VLANs 17
VLAN configuration examples 18
Example: Configuring port-based VLANs 18
Example: Configuring MAC-based VLANs 22
Example: Configuring IP subnet-based VLANs 29
Example: Configuring protocol-based VLANs 36
Configuring super VLANs 44
About super VLANs 44
Restrictions and guidelines: Super VLAN configuration 44
Super VLAN tasks at a glance 44
Creating a sub-VLAN 44

Configuring a super VLAN ····45
Configuring a super VLAN interface····45
Display and maintenance commands for super VLANs····46
Super VLAN configuration examples ····46
Example: Configuring a super VLAN ····46
Configuring private VLAN ····51
About private VLAN····51
Restrictions and guidelines: Private VLAN configuration····51
Private VLAN tasks at a glance ····52
Creating a primary VLAN ····52
Creating secondary VLANs····52
Associating the primary VLAN with secondary VLANs ····53
Configuring the uplink port ····53
Configuring a downlink port····54
Configuring Layer 3 communication for secondary VLANs ····55
Display and maintenance commands for the private VLAN····56
Private VLAN configuration examples ····56
Example: Configuring promiscuous ports ····56
Example: Configuring trunk promiscuous ports ····62
Example: Configuring trunk promiscuous and trunk secondary ports····69
Configuring voice VLANs ····78
About voice VLANs ····78
Working mechanism····78
Methods of identifying IP phones····78
Advertising the voice VLAN information to IP phones····79
IP phone access methods····79
Voice VLAN assignment modes····80
Cooperation of voice VLAN assignment modes and IP phones ····81
Security mode and normal mode of voice VLANs····82
Restrictions and guidelines: Voice VLAN configuration ····83
Voice VLAN tasks at a glance····83
Configuring the QoS priority settings for voice traffic····83
Configuring the ACL resource occupation mode of voice VLAN ····84
Configuring voice VLAN assignment modes for a port ····85
Configuring a port to operate in automatic voice VLAN assignment mode····85
Configuring a port to operate in manual voice VLAN assignment mode····86
Enabling LLDP for automatic IP phone discovery····87
Configuring LLDP or CDP to advertise a voice VLAN ····87
Configuring LLDP to advertise a voice VLAN ····87
Configuring CDP to advertise a voice VLAN····87
Display and maintenance commands for voice VLANs ····88
Voice VLAN configuration examples····89
Example: Configuring automatic voice VLAN assignment mode ····89
Example: Configuring manual voice VLAN assignment mode····94

Configuring VLANs

About VLANs

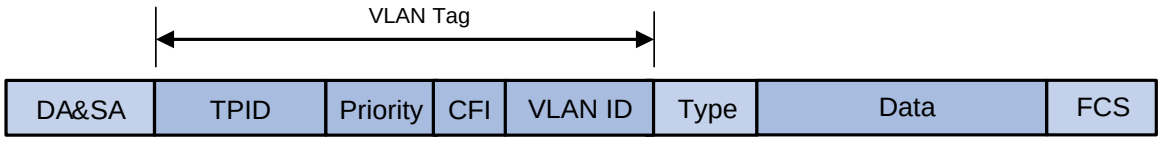
The Virtual Local Area Network (VLAN) technology divides a physical LAN into multiple logical LANs. It has the following benefits:

- **Security**—Hosts in the same VLAN can communicate with one another at Layer 2, but they are isolated from hosts in other VLANs at Layer 2.
- **Broadcast traffic isolation**—Each VLAN is a broadcast domain that limits the transmission of broadcast packets.
- **Flexibility**—A VLAN can be logically divided on a workgroup basis. Hosts in the same workgroup can be assigned to the same VLAN, regardless of their physical locations.

VLAN frame encapsulation

To identify Ethernet frames from different VLANs, IEEE 802.1Q inserts a four-byte VLAN tag between the destination and source MAC address (DA&SA) field and the Type field.

Figure 1 VLAN tag placement and format



A VLAN tag includes the following fields:

- **TPID**—16-bit tag protocol identifier that indicates whether a frame is VLAN-tagged. By default, the hexadecimal TPID value 8100 identifies a VLAN-tagged frame. A device vendor can set the TPID to a different value. For compatibility with a neighbor device, set the TPID value on the device to be the same as the neighbor device.For more information about setting the TPID value, see QinQ commands in Layer 2—*LAN Switching Command Reference*.
- **Priority**—3-bit long, identifies the 802.1p priority of the frame. For more information, see *ACL and QoS Configuration Guide*.
- **CFI**—1-bit long canonical format indicator that indicates whether the MAC addresses are encapsulated in the standard format when packets are transmitted across different media. Available values include:
  - **0 (default)**—The MAC addresses are encapsulated in the standard format.
  - **1**—The MAC addresses are encapsulated in a non-standard format.

  This field is always set to 0 for Ethernet.
- **VLAN ID**—12-bit long, identifies the VLAN to which the frame belongs. The VLAN ID range is 0 to 4095. VLAN IDs 0 and 4095 are reserved, and VLAN IDs 1 to 4094 are user configurable.

The way a network device handles an incoming frame depends on whether the frame has a VLAN tag and the value of the VLAN tag (if any).

Ethernet supports encapsulation formats Ethernet II, 802.3/802.2 LLC, 802.3/802.2 SNAP, and 802.3 raw. The Ethernet II encapsulation format is used here. For information about the VLAN tag fields in other frame encapsulation formats, see related protocols and standards.

For a frame that has multiple VLAN tags, the device handles it according to its outermost VLAN tag and transmits its inner VLAN tags as the payload.

VLAN types

The following VLAN types are available:

- Port-based VLAN.
- MAC-based VLAN.
- IP subnet-based VLAN.
- Protocol-based VLAN.

If all these types of VLANs are configured on a port, the port processes packets in the following descending order of priority by default:

- MAC-based VLAN.
- IP subnet-based VLAN.
- Protocol-based VLAN.
- Port-based VLAN.

Port-based VLANs

Port-based VLANs group VLAN members by port. A port forwards packets from a VLAN only after it is assigned to the VLAN.

Port link type

You can set the link type of a port to access, trunk, or hybrid. The port link type determines whether the port can be assigned to multiple VLANs. The link types use the following VLAN tag handling methods:

- **Access**—An access port can forward packets only from one VLAN and send these packets untagged. An access port is typically used in the following conditions:
  - Connecting to a terminal device that does not support VLAN packets.
  - In scenarios that do not distinguish VLANs.
- **Trunk**—A trunk port can forward packets from multiple VLANs. Except packets from the port VLAN ID (PVID), packets sent out of a trunk port are VLAN-tagged. Ports connecting network devices are typically configured as trunk ports.
- **Hybrid**—A hybrid port can forward packets from multiple VLANs. The tagging status of the packets forwarded by a hybrid port depends on the port configuration.

PVID

The PVID identifies the default VLAN of a port. Untagged packets received on a port are considered as the packets from the port PVID.

An access port can join only one VLAN. The VLAN to which the access port belongs is the PVID of the port. A trunk or hybrid port supports multiple VLANs and the PVID configuration.

How ports of different link types handle frames

Actions	Access	Trunk	Hybrid
In the inbound direction for an untagged frame	Tags the frame with the PVID tag.	• If the PVID is permitted on the port, tags the frame with the PVID tag. • If not, drops the frame.	
In the inbound direction for a tagged frame	• Receives the frame if its VLAN ID is the same as the PVID. • Drops the frame if	• Receives the frame if its VLAN is permitted on the port. • Drops the frame if its VLAN is not permitted on the port.	

Actions	Access	Trunk	Hybrid
	its VLAN ID is different from the PVID.		
In the outbound direction	Removes the VLAN tag and sends the frame.	• Removes the tag and sends the frame if the frame carries the PVID tag and the port belongs to the PVID. • Sends the frame without removing the tag if its VLAN is carried on the port but is different from the PVID.	Sends the frame if its VLAN is permitted on the port. The tagging status of the frame depends on the **`port hybrid vlan`** command configuration.

MAC-based VLANs

The MAC-based VLAN feature assigns hosts to a VLAN based on their MAC addresses. This feature is also called user-based VLAN because VLAN configuration remains the same regardless of a user's physical location.

Static MAC-based VLAN assignment

Use static MAC-based VLAN assignment in networks that have a small number of VLAN users. To configure static MAC-based VLAN assignment on a port, perform the following tasks:

1. Create MAC-to-VLAN entries.
2. Enable the MAC-based VLAN feature on the port.
3. Assign the port to the MAC-based VLAN.

A port configured with static MAC-based VLAN assignment processes a received frame as follows before sending the frame out:

- For an untagged frame, the port determines its VLAN ID in the following workflow:
  a. The port first performs a fuzzy match as follows:
     - Searches for the MAC-to-VLAN entries whose masks are not all Fs.
     - Performs a logical AND operation on the source MAC address and each of these masks.

       If an AND operation result matches the MAC address in a MAC-to-VLAN entry, the port tags the frame with the VLAN ID specific to this entry.

  b. If the fuzzy match fails, the port performs an exact match. It searches for MAC-to-VLAN entries whose masks are all Fs. If the source MAC address of the frame exactly matches the MAC address of a MAC-to-VLAN entry, the port tags the frame with the VLAN ID specific to this entry.

  c. If no matching VLAN ID is found, the port determines the VLAN for the packet by using the following matching order:
     - IP subnet-based VLAN.
     - Protocol-based VLAN.
     - Port-based VLAN.

     When a match is found, the port tags the packet with the matching VLAN ID.
- For a tagged frame, the port determines whether the VLAN ID of the frame is permitted on the port.

- If the VLAN ID of the frame is permitted on the port, the port forwards the frame.
- If the VLAN ID of the frame is not permitted on the port, the port drops the frame.

Dynamic MAC-based VLAN assignment

When you cannot determine the target MAC-based VLANs of a port, use dynamic MAC-based VLAN assignment on the port. To use dynamic MAC-based VLAN assignment, perform the following tasks:

1. Create MAC-to-VLAN entries.
2. Enable the MAC-based VLAN feature on the port.
3. Enable dynamic MAC-based VLAN assignment on the port.

Dynamic MAC-based VLAN assignment uses the following workflow, as shown in Figure 2:

1. When a port receives a frame, it first determines whether the frame is tagged.
   - If the frame is tagged, the port gets the source MAC address of the frame.
   - If the frame is untagged, the port selects a VLAN for the frame by using the following matching order:
     - MAC-based VLAN (fuzzy and exact MAC address match).
     - IP subnet-based VLAN.
     - Protocol-based VLAN.
     - Port-based VLAN.

     After tagging the frame with the selected VLAN, the port gets the source MAC address of the frame.
2. The port uses the source MAC address and VLAN of the frame to match the MAC-to VLAN entries.
   - If the source MAC address of the frame exactly matches the MAC address in a MAC-to-VLAN entry, the port checks whether the VLAN ID of the frame matches the VLAN in the entry.
     - If the two VLAN IDs match, the port joins the VLAN and forwards the frame.
     - If the two VLAN IDs do not match, the port drops the frame.
   - If the source MAC address of the frame does not exactly match any MAC addresses in MAC-to-VLAN entries, the port checks whether the VLAN ID of the frame is its PVID.
     - If the VLAN ID of the frame is the PVID of the port, the port determines whether it allows the PVID.

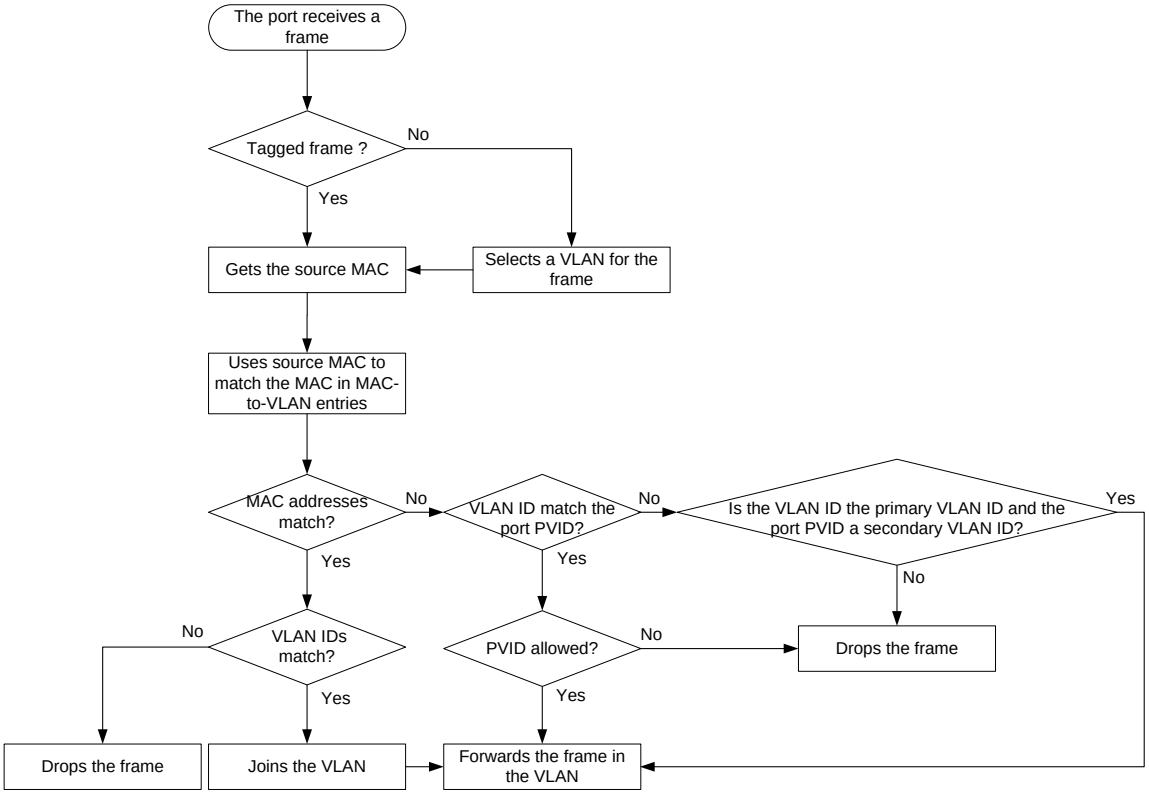
       If the PVID is allowed, the port generates a dynamic MAC address entry with the source MAC address of the packet as the destination MAC address, and forwards the frame within the PVID. If the PVID is not allowed, the port drops the frame.

IMPORTANT:

If the device has generated a dynamic MAC address entry, packets whose source MAC address is the destination MAC address in the dynamic MAC address entry can still be forwarded by the port, even if the PVID configuration of the port is changed and the VLAN ID of the packet no longer meets the above conditions. The port drops the frame only when the generated dynamic MAC address entry ages out and the frame no longer meets the above conditions.

     - If the VLAN ID of the frame is not the PVID of the port, the port determines whether the VLAN ID is the primary VLAN ID and the port PVID is a secondary VLAN ID. If yes, the port forwards the frame. Otherwise, the port drops the frame.

Figure 2 Flowchart for processing a frame in dynamic MAC-based VLAN assignment



Server-assigned MAC-based VLAN

Use this feature with access authentication, such as MAC-based 802.1X authentication, to implement secure and flexible terminal access.

To implement server-assigned MAC-based VLAN, perform the following tasks:

1. Configure the server-assigned MAC-based VLAN feature on the access device.
2. Configure username-to-VLAN entries on the access authentication server.

When a user passes authentication of the access authentication server, the server assigns the authorization VLAN information for the user to the device. The device then performs the following operations:

1. Generates a MAC-to-VLAN entry by using the source MAC address of the user packet and the authorization VLAN information. The authorization VLAN is a MAC-based VLAN.

   The generated MAC-to-VLAN entry cannot conflict with the existing static MAC-to-VLAN entries. If a confliction exists, the dynamic MAC-to-VLAN entry cannot be generated.
2. Assigns the port that connects the user to the MAC-based VLAN.

When the user goes offline, the device automatically deletes the MAC-to-VLAN entry and removes the port from the MAC-based VLAN. For more information about 802.1X and MAC authentication, see *Security Configuration Guide*.

IP subnet-based VLANs

The IP subnet-based VLAN feature assigns untagged packets to VLANs based on their source IP addresses and subnet masks.

Use this feature when untagged packets from an IP subnet or IP address must be transmitted in a VLAN.

Protocol-based VLANs

The protocol-based VLAN feature assigns inbound packets to different VLANs based on their protocol types and encapsulation formats. The protocols available for VLAN assignment include IP, IPX, and AT. The encapsulation formats include Ethernet II, 802.3 raw, 802.2 LLC, and 802.2 SNAP.

This feature associates the available network service types with VLANs and facilitates network management and maintenance.

Layer 3 communication between VLANs

Hosts of different VLANs use VLAN interfaces to communicate at Layer 3. VLAN interfaces are virtual interfaces that do not exist as physical entities on devices. For each VLAN, you can create one VLAN interface and assign an IP address to it. The VLAN interface acts as the gateway of the VLAN to forward packets at Layer 3. When the IP addresses of multiple VLAN interfaces are on different subnets, you must configure IP routing protocols to enable Layer 3 communication between these IP addresses. Then, the devices can forward packets at Layer 3 across subnets between different VLANs. For more information about IP routing protocols, see *Layer 3—IP Routing Configuration Guide*.

Protocols and standards

IEEE 802.1Q, *IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks*

Configuring a VLAN

Restrictions and guidelines

- As the system default VLAN, VLAN 1 cannot be created or deleted.
- Before you delete a dynamic VLAN or a VLAN locked by an application, you must first remove the configuration from the VLAN.

VLAN configuration tasks at a glance

To configure VLANs, perform the following tasks:

1. Creating VLANs

Creating VLANs

1. Enter system view.

   system-view
2. Create one or multiple VLANs.
   - Create a VLAN and enter its view.

     vlan *vlan-id*
   - Create multiple VLANs and enter VLAN view.

     Create VLANs.

     vlan { *vlan-id-list* | **all** }

     Enter VLAN view.

vlan *vlan-id*

By default, only the system default VLAN (VLAN 1) exists.

3. (Optional.) Set a name for the VLAN.

   name *text*

   By default, the name of a VLAN is **VLAN** *vlan-id.* The *vlan-id* argument specifies the VLAN ID in a four-digit format. If the VLAN ID has fewer than four digits, leading zeros are added. For example, the name of VLAN 100 is **VLAN 0100**.

4. (Optional.) Configure the description for the VLAN.

   description *text*

   By default, the description of a VLAN is **VLAN** *vlan-id.* The *vlan-id* argument specifies the VLAN ID in a four-digit format. If the VLAN ID has fewer than four digits, leading zeros are added. For example, the default description of VLAN 100 is **VLAN 0100**.

Enabling packet statistics for the VLAN

About this task

When you need to examine or troubleshoot the network, you can enable packet statistics for a VLAN to monitor the statistics of packets in the VLAN. The VLAN packet statistics include statistics on unicast, multicast, and broadcast packets.

Restrictions and guidelines

Disable packet statistics for a VLAN to save system resources when you do not need this feature.

The system does not automatically clear the packet statistics for VLANs. To clear the packet statistics for a VLAN, use the **reset vlan statistics** command.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Enable packet statistics for the VLAN.

   statistics enable

   By default, packet statistics is disabled for a VLAN.

Configuring port-based VLANs

Restrictions and guidelines for port-based VLANs

- When you use the **undo vlan** command to delete the PVID of a port, either of the following events occurs depending on the port link type:
  - For an access port, the PVID of the port changes to VLAN 1.
  - For a hybrid or trunk port, the PVID setting of the port does not change.

  You can use a nonexistent VLAN as the PVID for a hybrid or trunk port, but not for an access port.
- As a best practice, set the same PVID for a local port and its peer.
- To prevent a port from dropping untagged packets or PVID-tagged packets, assign the port to its PVID.

Assigning an access port to a VLAN

About this task

You can assign an access port to a VLAN in VLAN view or interface view.

Assigning one or multiple access ports to a VLAN in VLAN view

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Assign one or multiple access ports to the VLAN.
   port *interface-list*
   By default, all ports belong to VLAN 1.

Assigning an access port to a VLAN in interface view

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Set the port link type to access.
   port link-type access
   By default, all ports are access ports.
4. Assign the access port to a VLAN.
   port access vlan *vlan-id*
   By default, all access ports belong to VLAN 1.

Assigning a trunk port to a VLAN

About this task

A trunk port supports multiple VLANs. You can assign it to a VLAN in interface view.

Restrictions and guidelines

To change the link type of a port from trunk to hybrid, set the link type to access first.

To enable a trunk port to transmit packets from its PVID, you must assign the trunk port to the PVID by using the **port trunk permit vlan** command.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

interface bridge-aggregation *interface-number*

3. Set the port link type to trunk.
   port link-type trunk
   By default, all ports are access ports.
4. Assign the trunk port to the specified VLANs.
   port trunk permit vlan { *vlan-id-list* | **all** }
   By default, a trunk port permits only VLAN 1.
5. (Optional.) Set the PVID for the trunk port.
   port trunk pvid vlan *vlan-id*
   The default setting is VLAN 1.

Assigning a hybrid port to a VLAN

About this task

A hybrid port supports multiple VLANs. You can assign it to the specified VLANs in interface view. Make sure the VLANs have been created.

Restrictions and guidelines

To change the link type of a port from trunk to hybrid, set the link type to access first.

To enable a hybrid port to transmit packets from its PVID, you must assign the hybrid port to the PVID by using the **port hybrid vlan** command.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Set the port link type to hybrid.
   port link-type hybrid
   By default, all ports are access ports.
4. Assign the hybrid port to the specified VLANs.
   port hybrid vlan *vlan-id-list* { **tagged** | **untagged** }
   By default, the hybrid port is an untagged member of the VLAN to which the port belongs when its link type is **access**.
5. (Optional.) Set the PVID for the hybrid port.
   port hybrid pvid vlan *vlan-id*
   By default, the PVID of a hybrid port is the ID of the VLAN to which the port belongs when its link type is **access**.

Configuring MAC-based VLANs

Restrictions and guidelines for MAC-based VLANs

- MAC-based VLANs are available only on hybrid ports.
- Do not configure a VLAN as both a super VLAN and a MAC-based VLAN.
- The MAC-based VLAN feature is mainly configured on downlink ports of user access devices. Do not use this feature with link aggregation.

Configuring static MAC-based VLAN assignment

1. Enter system view.

   system-view
2. Create a MAC-to-VLAN entry.

   mac-vlan mac-address *mac-address* [**mask** *mac-mask*] **vlan** *vlan-id* [**dot1p** *priority*]

   By default, no MAC-to-VLAN entries exist.
3. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
4. Set the port link type to hybrid.

   port link-type hybrid

   By default, all ports are access ports.
5. Assign the hybrid port to the MAC-based VLANs.

   port hybrid vlan *vlan-id-list* { **tagged** | **untagged** }

   By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is **access**.
6. Enable the MAC-based VLAN feature.

   mac-vlan enable

   By default, this feature is disabled.
7. (Optional.) Configure the interface to assign VLANs based on the MAC address preferentially.

   vlan precedence mac-vlan

   By default, the system assigns VLANs based on the MAC address preferentially when both the MAC-based VLAN and IP subnet-based VLAN are configured on a port.

Configuring dynamic MAC-based VLAN assignment

About this task

For successful dynamic MAC-based VLAN assignment, use static VLANs when you create MAC-to-VLAN entries.

When a port joins a VLAN specified in the MAC-to-VLAN entry, one of the following events occurs depending on the port configuration:

- If the port has not been configured to allow packets from the VLAN to pass through, the port joins the VLAN as an untagged member.

- If the port has been configured to allow packets from the VLAN to pass through, the port configuration remains the same.

The 802.1p priority of the VLAN in a MAC-to-VLAN entry determines the transmission priority of the matching packets.

Restrictions and guidelines

- If you configure both static and dynamic MAC-based VLAN assignments on a port, dynamic MAC-based VLAN assignment takes effect.
- As a best practice to ensure correct operation of 802.1X and MAC authentication, do not use dynamic MAC-based VLAN assignment with 802.1X or MAC authentication.
- As a best practice, do not both configure dynamic MAC-based VLAN assignment and disable MAC address learning on a port. If the two features are configured together on a port, the port forwards only packets exactly matching the MAC-to-VLAN entries and drops inexactly matching packets.
- As a best practice, do not configure both dynamic MAC-based VLAN assignment and the MAC learning limit on a port.

  If the two features are configured together on a port and the port learns the configured maximum number of MAC address entries, the port processes packets as follows:
  - Forwards only packets matching the MAC address entries learnt by the port.
  - Drops unmatching packets.
- As a best practice, do not use dynamic MAC-based VLAN assignment with MSTP. In MSTP mode, if a port is blocked in the MSTI of its target VLAN, the port drops the received packets instead of delivering them to the CPU. As a result, the port will not be dynamically assigned to the target VLAN.
- As a best practice, do not use dynamic MAC-based VLAN assignment with PVST. In PVST mode, if the target VLAN of a port is not permitted on the port, the port is placed in blocked state. The port drops the received packets instead of delivering them to the CPU. As a result, the port will not be dynamically assigned to the target VLAN.
- As a best practice, do not configure both dynamic MAC-based VLAN assignment and automatic voice VLAN assignment mode on a port. They can have a negative impact on each other.

Procedure

1. Enter system view.

   system-view
2. Create a MAC-to-VLAN entry.

   mac-vlan mac-address *mac-address* **vlan** *vlan-id* [**dot1p** *priority*]

   By default, no MAC-to-VLAN entries exist.
3. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
4. Set the port link type to hybrid.

   port link-type hybrid

   By default, all ports are access ports.
5. Enable the MAC-based VLAN feature.

   mac-vlan enable

   By default, MAC-based VLAN is disabled.
6. Enable dynamic MAC-based VLAN assignment.

   mac-vlan trigger enable

   By default, dynamic MAC-based VLAN assignment is disabled.

The VLAN assignment for a port is triggered only when the source MAC address of its receiving packet exactly matches the MAC address in a MAC-to-VLAN entry.

7. (Optional.) Configure the system to assign VLANs based on the MAC address preferentially.

   vlan precedence mac-vlan

   By default, the system assigns VLANs based on the MAC address preferentially when both the MAC-based VLAN and IP subnet-based VLAN are configured on a port.

8. (Optional.) Disable the port from forwarding packets that fail the exact MAC address match in its PVID.

   port pvid forbidden

   By default, when a port receives packets whose source MAC addresses fail the exact match, the port forwards them in its PVID.

Configuring server-assigned MAC-based VLAN

1. Enter system view.

   system-view

2. Enter interface view.

   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*

3. Set the port link type to hybrid.

   port link-type hybrid

   By default, all ports are access ports.

4. Assign the hybrid port to the MAC-based VLANs.

   port hybrid vlan *vlan-id-list* { **tagged** | **untagged** }

   By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is **access**.

5. Enable the MAC-based VLAN feature.

   mac-vlan enable

   By default, MAC-based VLAN is disabled.

6. (Optional.) Configure the interface to assign VLANs based on the MAC address preferentially.

   vlan precedence mac-vlan

   By default, the system assigns VLANs based on the MAC address preferentially when both the MAC-based VLAN and IP subnet-based VLAN are configured on a port.

7. Configure 802.1X or MAC authentication.

   For more information, see *Security Command Reference*.

Enabling ARP flood suppression

About this task

Excessive ARP packets in the network lead to increased CPU load on the device, affecting normal processing of other user operations. To resolve this issue, you can enable ARP flood suppression on the device to reduce the flood of ARP packets and ensure normal network communication.

With this feature enabled, the device can monitor the ARP packets passing through it. When other devices initiate ARP resolution again, the device can responce to the ARP requests to suppress the network flood caused by ARP resolution.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ARP flood suppression.
   arp suppression enable
   By default, ARP flood suppression is disabled.

Configuring IP subnet-based VLANs

Restrictions and guidelines

This feature is available only on hybrid ports, and it processes only untagged packets.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Associate the VLAN with an IP subnet or IP address.
   ip-subnet-vlan [*ip-subnet-index*] **ip** *ip-address* [*mask*]
   By default, a VLAN is not associated with an IP subnet or IP address.
   A multicast subnet or a multicast address cannot be associated with a VLAN.
4. Return to system view.
   quit
5. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
6. Set the port link type to hybrid.
   port link-type hybrid
   By default, all ports are access ports.
7. Assign the hybrid port to the specified IP subnet-based VLANs.
   port hybrid vlan *vlan-id-list* { **tagged** | **untagged** }
   By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is **access**.
8. Associate the hybrid port with the specified IP subnet-based VLAN.
   port hybrid ip-subnet-vlan vlan *vlan-id*
   By default, a hybrid port is not associated with a subnet-based VLAN.

Configuring protocol-based VLANs

About this task

A protocol-based VLAN has one or multiple protocol templates. A protocol template defines a protocol type and an encapsulation format as the match criteria to match inbound packets. Each protocol template has a unique index in the protocol-based VLAN. All protocol templates in a protocol-based VLAN have the same VLAN ID.

For a port to assign inbound packets to protocol-based VLANs, perform the following tasks:

- Assign the port to the protocol-based VLANs.
- Associate the port with the protocol templates of the protocol-based VLANs.

When an untagged packet arrives at the port, the port processes the packet as follows:

- If the protocol type and encapsulation format in the packet match a protocol template, the port tags the packet with the VLAN tag specific to the protocol template.
- If no protocol templates are matched, the port tags the packet with its PVID.

Restrictions and guidelines

The voice VLAN in automatic mode processes only tagged voice traffic. Do not configure a VLAN as both a protocol-based VLAN and a voice VLAN.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Associate the VLAN with a protocol template.

   protocol-vlan [*protocol-index*] { **at** | **ipv4** | **ipv6** | **ipx** { **ethernetii** | **llc** | **raw** | **snap** } | **mode** { **ethernetii etype** *etype-id* | **llc** { **dsap** *dsap-id* [**ssap** *ssap-id*] | **ssap** *ssap-id* } | **snap etype** *etype-id* } }

   By default, a VLAN is not associated with a protocol template.

4. Exit VLAN view.

   quit

5. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*

6. Set the port link type to hybrid.

   port link-type hybrid

   By default, all ports are access ports.

7. Assign the hybrid port to the specified protocol-based VLANs.

   port hybrid vlan *vlan-id-list* { **tagged** | **untagged** }

   By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is **access**.

8. Associate the hybrid port with the specified protocol-based VLAN.

   port hybrid protocol-vlan vlan *vlan-id* { *protocol-index* [**to** *protocol-end*] | **all** }

By default, a hybrid port is not associated with a protocol-based VLAN.

Configuring a VLAN group

About this task

A VLAN group includes a set of VLANs.

On an authentication server, a VLAN group name represents a group of authorization VLANs. When an 802.1X or MAC authentication user passes authentication, the authentication server assigns a VLAN group name to the device. The device then uses the received VLAN group name to match the locally configured VLAN group names. If a match is found, the device selects a VLAN from the group and assigns the VLAN to the user. For more information about 802.1X and MAC authentication, see *Security Configuration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Create a VLAN group and enter its view.
   `vlan-group` *`group-name`*
3. Add VLANs to the VLAN group.
   `vlan-list` *`vlan-id-list`*
   By default, no VLANs exist in a VLAN group.
   You can add multiple VLAN lists to a VLAN group.

Configuring VLAN interfaces

Restrictions and guidelines

- You cannot create VLAN interfaces for sub-VLANs. For more information about sub-VLANs, see "Configuring super VLANs."
- You cannot create VLAN interfaces for secondary VLANs that have the following characteristics:
  - Associated with the same primary VLAN.
  - Enabled with Layer 3 communication in VLAN interface view of the primary VLAN interface.

  For more information about secondary VLANs, see "Configuring private VLAN."
- A VLAN interface can come up only when one or more Ethernet interfaces in the corresponding VLAN are up.

VLAN interfaces configuration tasks at a glance

To configure VLAN interfaces, perform the following tasks:

1. Creating a VLAN interface
2. Restoring the default settings for the VLAN interface

Prerequisites

Before you create a VLAN interface for a VLAN, create the VLAN first.

Creating a VLAN interface

1. Enter system view.
   system-view
2. Create a VLAN interface and enter its view.
   interface vlan-interface *interface-number*
3. Assign an IP address to the VLAN interface.
   ip address *ip-address* { *mask* | *mask-length* } [**sub**]
   By default, no IP address is assigned to a VLAN interface.
4. (Optional.) Configure the description for the VLAN interface.
   description *text*
   The default setting is the VLAN interface name. For example, **Vlan-interface1 Interface**.
5. (Optional.) Set the MTU for the VLAN interface.
   mtu *size*
   By default, the MTU of a VLAN interface is 1500 bytes.
6. (Optional.) Set a MAC address for the VLAN interface.
   mac-address *mac-address*
   By default, no MAC address is set for a VLAN interface.
7. (Optional.) Set the expected bandwidth for the interface.
   bandwidth *bandwidth-value*
   By default, the expected bandwidth (in kbps) is the interface baud rate divided by 1000.
8. Bring up the VLAN interface.
   undo shutdown
   By default, a VLAN interface is not manually shut down.

Restoring the default settings for the VLAN interface

Restrictions and guidelines

This feature might fail to restore the default settings for some commands for reasons such as command dependencies or system restrictions. Use the **display this** command in interface view to identify these commands, and then use their **undo** forms or follow the command reference to restore their default settings. If your restoration attempt still fails, follow the error message instructions to resolve the problem.

Procedure

1. Enter system view.
   system-view
2. Enter a VLAN interface view.
   interface vlan-interface *interface-number*
3. Restore the default settings for the VLAN interface.
   default

CAUTION:

This feature might interrupt ongoing network services. Make sure you are fully aware of the impact of this feature when you use it on a live network.

Enabling the function of setting VLAN tags for sent protocol packets on an interface

About this task

With this feature enabled on an interface, the interface records all layers of VLAN tags of received LACP or BFD protocol packets. When the interface sends the same protocol packets, the interface adds the recorded VLAN tags to the sent protocol packets, regardless of whether the link type of the interface allows these packets to carry these VLAN tags. For more information about LACP, see Ethernet link aggregation configuration in *Layer 2—LAN Switching Configuration Guide*. For more information about BFD, see BFD configuration in *High Availability Configuration Guide*.

Restrictions and guidelines

Executing the **port outbound-vlan-tag enable** command and then its **undo** form on an interface will cause BFD sessions on the interface to go down. Perform this operation with caution.

Executing this command on an aggregation group member port does not affect the other member ports in the same aggregation group or affect the Selected state of aggregation group member ports. To ensure configuration consistency within an aggregation group, manually execute this command on all aggregation group member ports.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
3. Enable the function of setting VLAN tags for sent protocol packets on the interface.
   port outbound-vlan-tag enable

Display and maintenance commands for VLANs

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ARP suppression entries.	**display arp suppression vlan** [*vlan-id*] [**slot** *slot-number*] [**count**] **display arp suppression vlan** *vlan-id* **ip** *ip-address* [**slot** *slot-number*]
Display VLAN interface information.	**display interface** [**vlan-interface** [*interface-number*]] [**brief** [**description** | **down**]]
Display information about IP subnet-based VLANs that are associated with the specified ports.	**display ip-subnet-vlan interface** { *interface-type interface-number1* [**to** *interface-type interface-number2*] | **all** }
Display information about IP subnet-based VLANs.	**display ip-subnet-vlan vlan** { *vlan-id1* [**to** *vlan-id2*] | **all** }
Display hybrid ports or trunk ports on the device.	**display port** { **hybrid** | **trunk** }

Task	Command
Display information about protocol-based VLANs that are associated with the specified ports.	**display protocol-vlan interface** { *interface-type interface-number1* [**to** *interface-type interface-number2*] | **all** }
Display information about protocol-based VLANs.	**display protocol-vlan vlan** { *vlan-id1* [**to** *vlan-id2*] | **all** }
Display VLAN information.	**display vlan** [*vlan-id1* [**to** *vlan-id2*] | **all** | **dynamic** | **reserved** | **static**]
Display brief VLAN information.	**display vlan brief**
Display packet statistics for a VLAN.	**display vlan** *vlan-id* **statistics** [**slot** *slot-number*]
Display VLAN group information.	**display vlan-group** [*group-name*]
Clear ARP flood suppression entries on VSIs	**reset arp suppression vlan** [*vlan-id* **ip** *ip-address* | *vlan-id*]
Clear statistics on a VLAN interface.	**reset counters interface** [**vlan-interface** [*interface-number*]]
Clear packet statistics for a VLAN.	**reset vlan** *vlan-id* **statistics** [**slot** *slot-number*]
Display MAC-to-VLAN entries.	**display mac-vlan** { **all** | **dynamic** | **mac-address** *mac-address* [**mask** *mac-mask*] | **static** | **vlan** *vlan-id* }
Display all ports that are enabled with the MAC-based VLAN feature.	**display mac-vlan interface**

VLAN configuration examples

Example: Configuring port-based VLANs

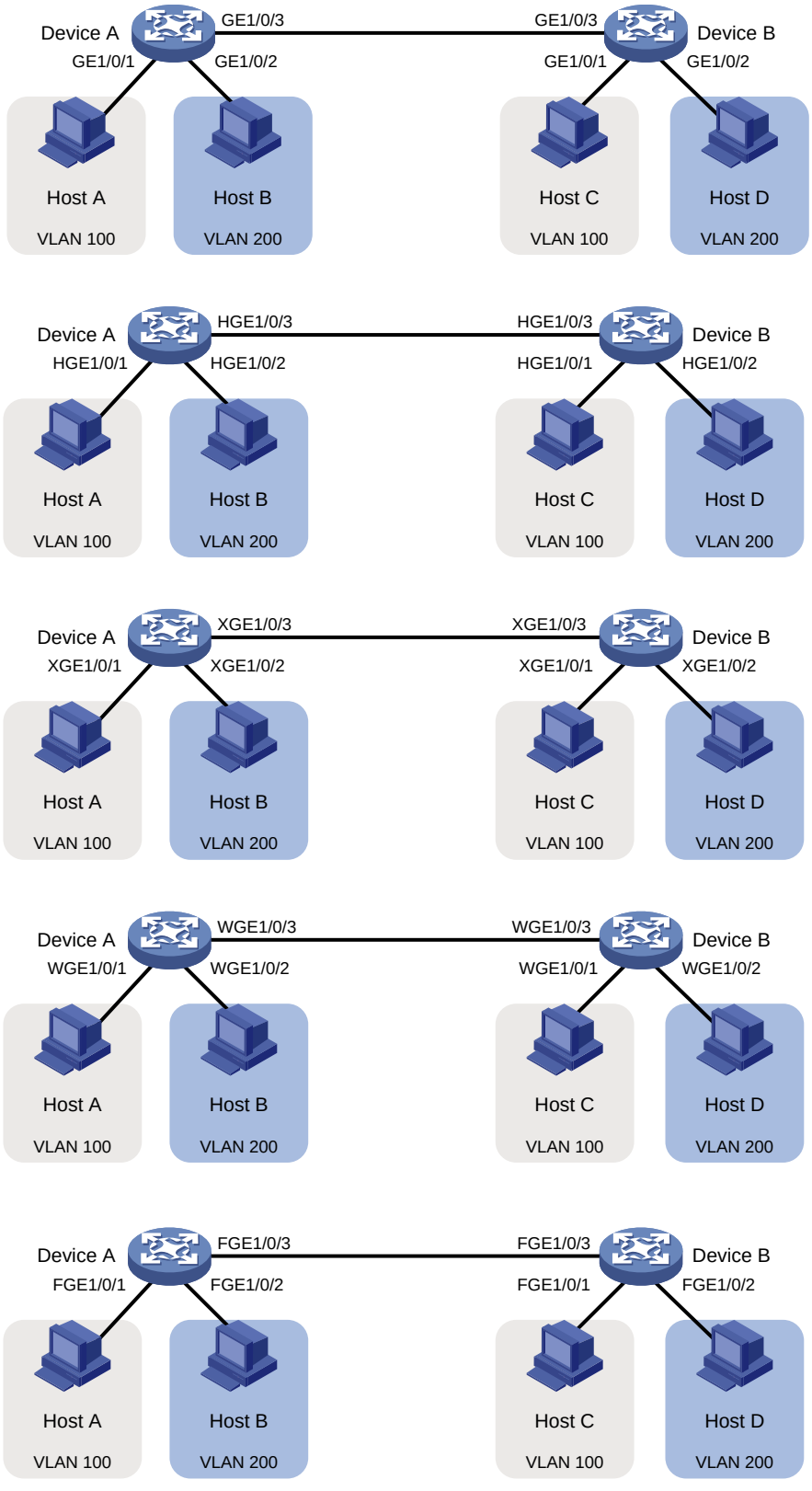
Network configuration

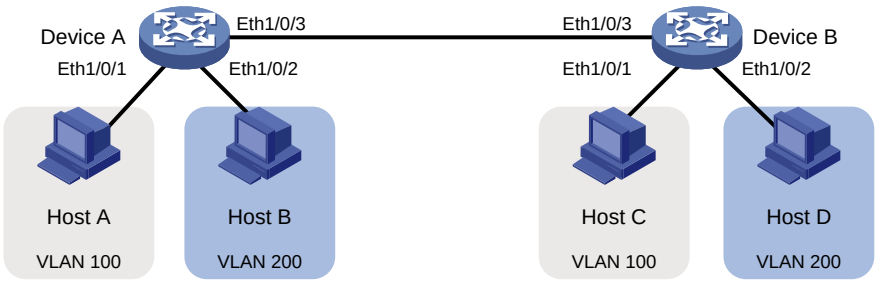
As shown in Figure 3:

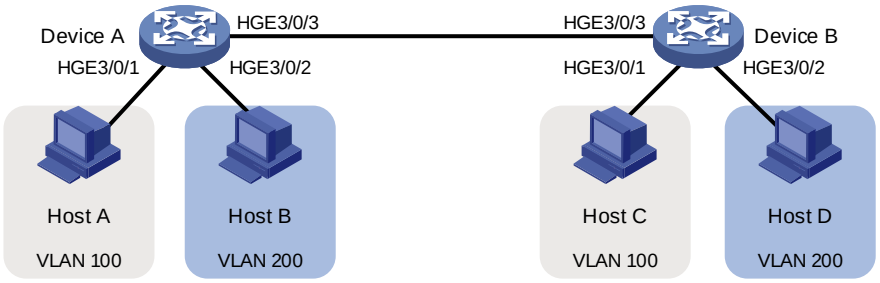
- Host A and Host C belong to Department A. VLAN 100 is assigned to Department A.
- Host B and Host D belong to Department B. VLAN 200 is assigned to Department B.

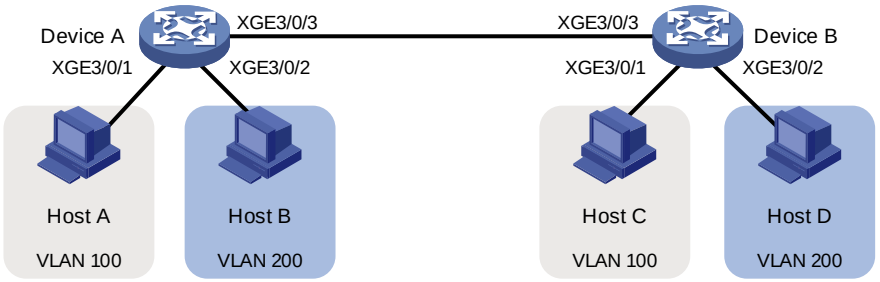
Configure port-based VLANs so that only hosts in the same department can communicate with each other.

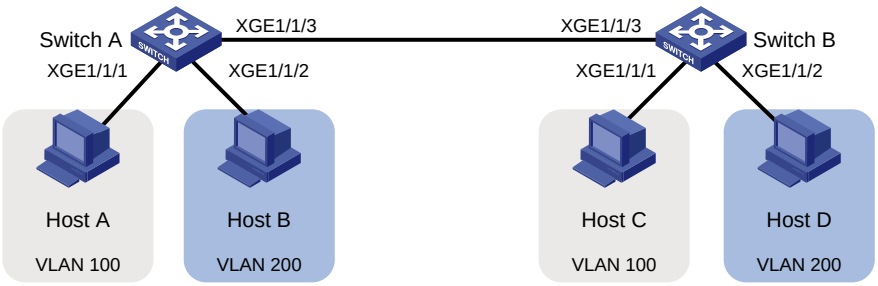
Figure 3 Network diagram

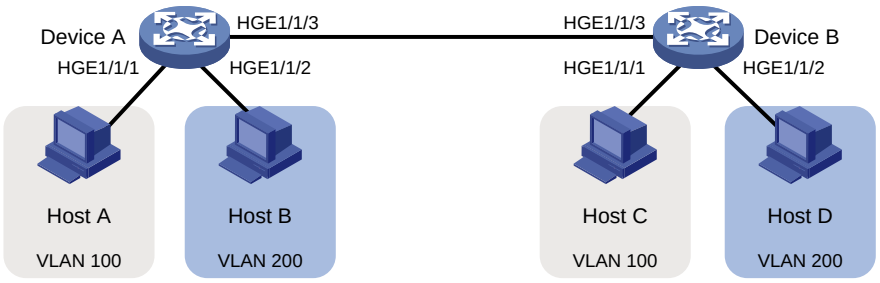


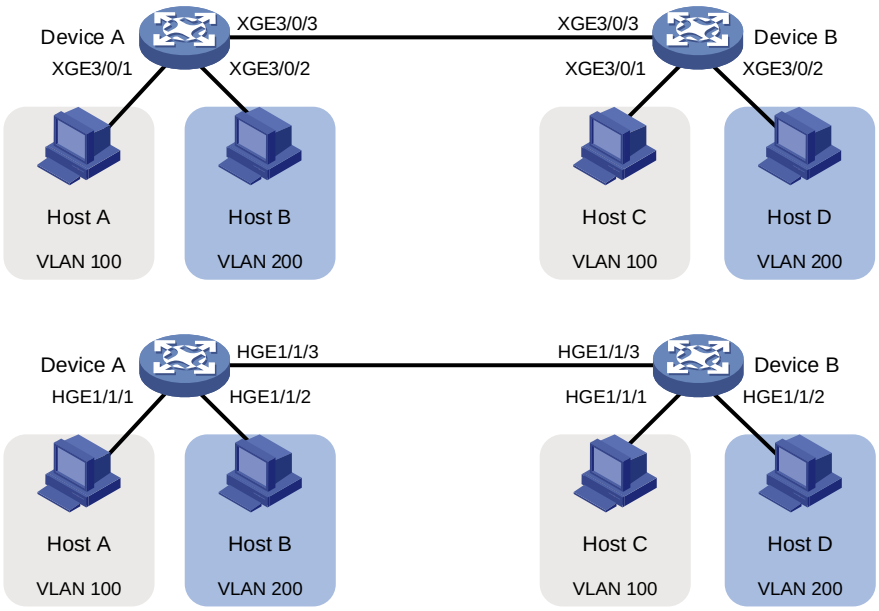

Device A
Eth1/0/3
Eth1/0/3
Device B
Eth1/0/1
Eth1/0/2
Eth1/0/1
Eth1/0/2
Host A
VLAN 100
Host B
VLAN 200
Host C
VLAN 100
Host D
VLAN 200


Device A
HGE3/0/3
HGE3/0/3
Device B
HGE3/0/1
HGE3/0/2
HGE3/0/1
HGE3/0/2
Host A
VLAN 100
Host B
VLAN 200
Host C
VLAN 100
Host D
VLAN 200


Device A
XGE3/0/3
XGE3/0/3
Device B
XGE3/0/1
XGE3/0/2
XGE3/0/1
XGE3/0/2
Host A
VLAN 100
Host B
VLAN 200
Host C
VLAN 100
Host D
VLAN 200


Switch A
XGE1/1/3
XGE1/1/3
Switch B
XGE1/1/1
XGE1/1/2
XGE1/1/1
XGE1/1/2
Host A
VLAN 100
Host B
VLAN 200
Host C
VLAN 100
Host D
VLAN 200


Device A
HGE1/1/3
HGE1/1/3
Device B
HGE1/1/1
HGE1/1/2
HGE1/1/1
HGE1/1/2
Host A
VLAN 100
Host B
VLAN 200
Host C
VLAN 100
Host D
VLAN 200



Procedure

1. Configure Device A:

   # Create VLAN 100, and assign GigabitEthernet 1/0/1 to VLAN 100.

   ```
   <DeviceA> system-view
   [DeviceA] vlan 100
   [DeviceA-vlan100] port gigabitethernet 1/0/1
   [DeviceA-vlan100] quit
   ```

   # Create VLAN 200, and assign GigabitEthernet 1/0/2 to VLAN 200.

   ```
   [DeviceA] vlan 200
   [DeviceA-vlan200] port gigabitethernet 1/0/2
   [DeviceA-vlan200] quit
   ```

   # Configure GigabitEthernet 1/0/3 as a trunk port, and assign the port to VLANs 100 and 200.

   ```
   [DeviceA] interface gigabitethernet 1/0/3
   [DeviceA-GigabitEthernet1/0/3] port link-type trunk
   [DeviceA-GigabitEthernet1/0/3] port trunk permit vlan 100 200
   Please wait... Done.
   ```

2. Configure Device B in the same way Device A is configured. (Details not shown.)
3. Configure hosts:
   a. Configure Host A and Host C to be on the same IP subnet. For example, 192.168.100.0/24.
   b. Configure Host B and Host D to be on the same IP subnet. For example, 192.168.200.0/24.

Verifying the configuration

Verify that Host A and Host C can ping each other, but they both fail to ping Host B and Host D. (Details not shown.)

Verify that Host B and Host D can ping each other, but they both fail to ping Host A and Host C. (Details not shown.)

Verify that VLANs 100 and 200 are correctly configured on Device A.

```
[DeviceA-GigabitEthernet1/0/3] display vlan 100
 VLAN ID: 100
 VLAN type: Static
 Route interface: Not configured
```

```
 Description: VLAN 0100
 Name: VLAN 0100
 Tagged ports:
    GigabitEthernet1/0/3(U)
 Untagged ports:
    GigabitEthernet1/0/1(U)
[DeviceA-GigabitEthernet1/0/3] display vlan 200
 VLAN ID: 200
 VLAN type: Static
 Route interface: Not configured
 Description: VLAN 0200
 Name: VLAN 0200
 Tagged ports:
    GigabitEthernet1/0/3(U)
 Untagged ports:
    GigabitEthernet1/0/2(U)
```

Example: Configuring MAC-based VLANs

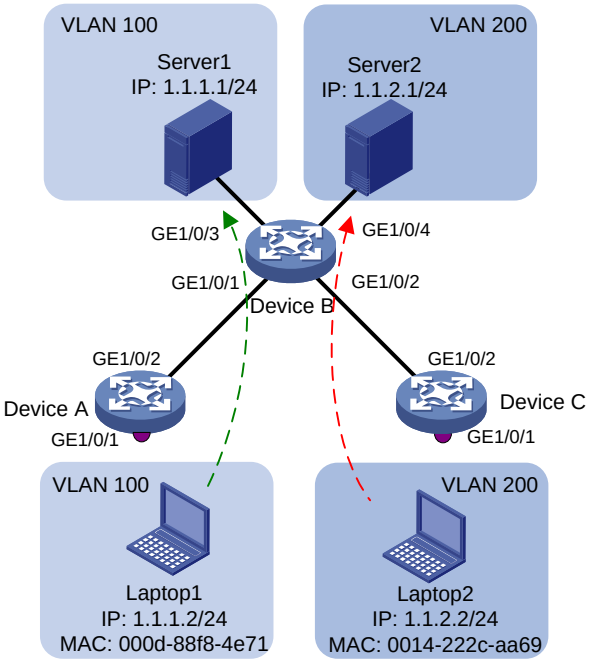
Network configuration

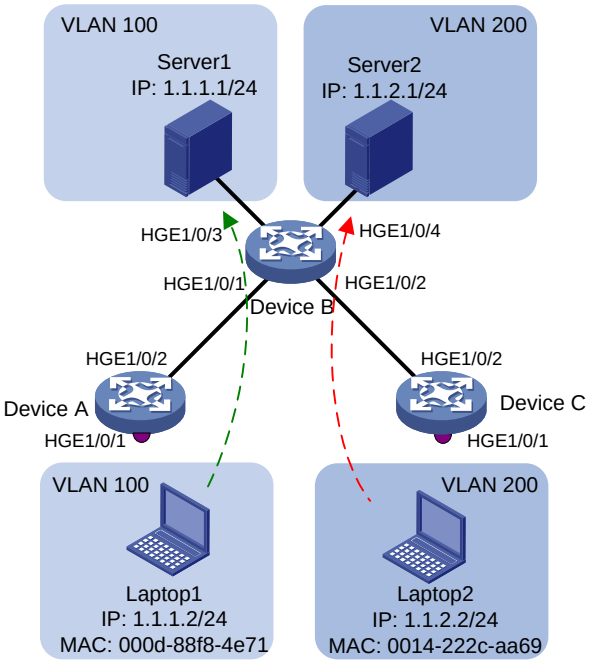
As shown in Figure 4:

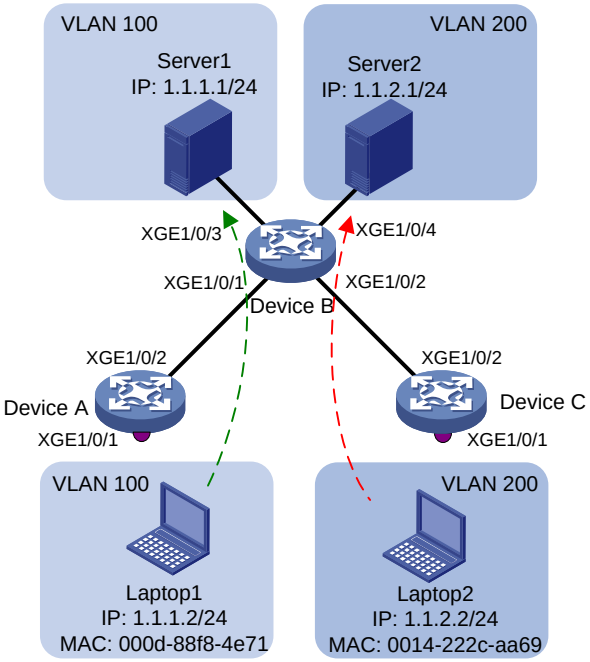
- GigabitEthernet 1/0/1 of Device A and Device C are each connected to a meeting room. Laptop 1 and Laptop 2 are used for meetings and might be used in either of the two meeting rooms.
- One department uses VLAN 100 and owns Laptop 1. The other department uses VLAN 200 and owns Laptop 2.

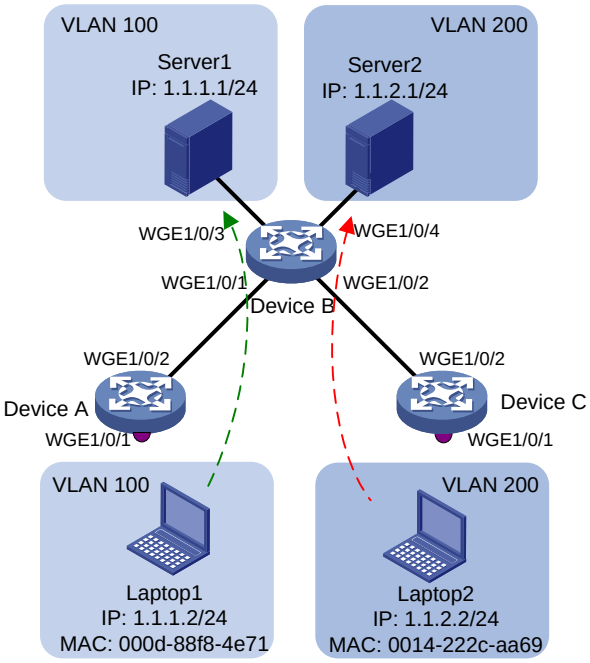
Configure MAC-based VLANs, so that Laptop 1 and Laptop 2 can access Server 1 and Server 2, respectively, no matter which meeting room they are used in.

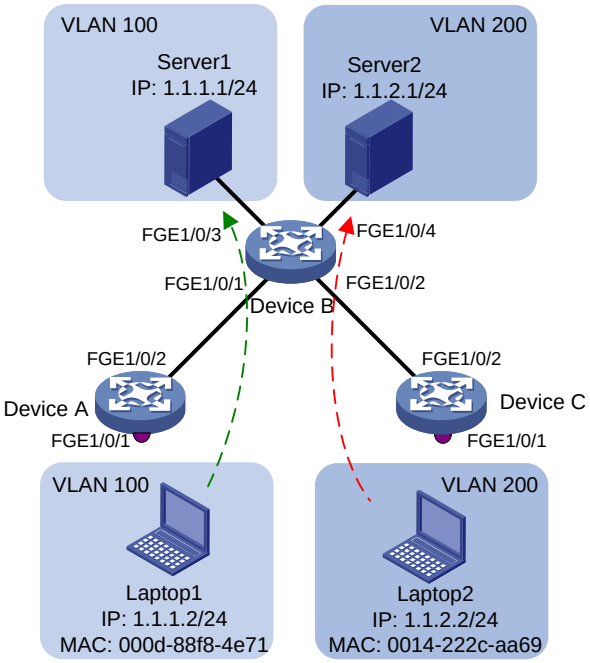
Figure 4 Network diagram

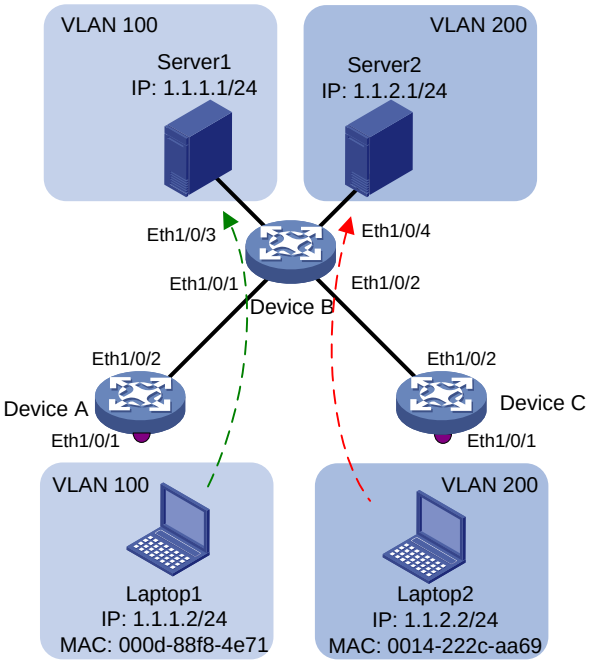


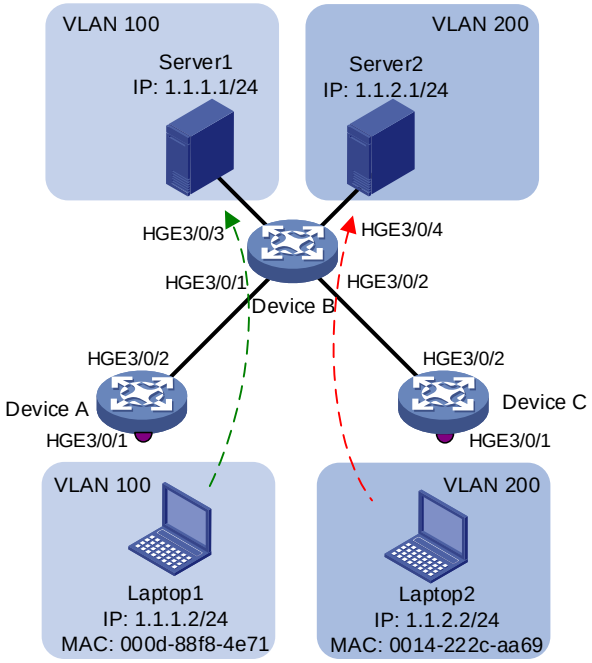

VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
HGE1/0/3
HGE1/0/4
HGE1/0/1
HGE1/0/2
Device B
HGE1/0/2
HGE1/0/2
Device A
Device C
HGE1/0/1
HGE1/0/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69

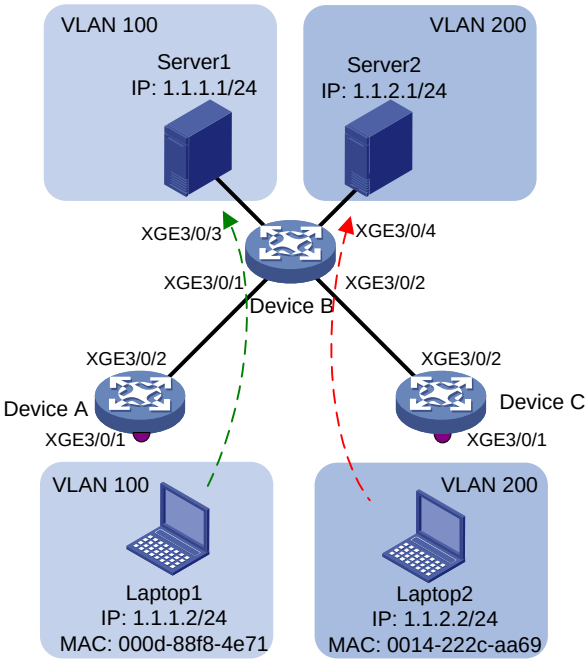

VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
XGE1/0/3
XGE1/0/4
XGE1/0/1
XGE1/0/2
Device B
XGE1/0/2
XGE1/0/2
Device A
Device C
XGE1/0/1
XGE1/0/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69

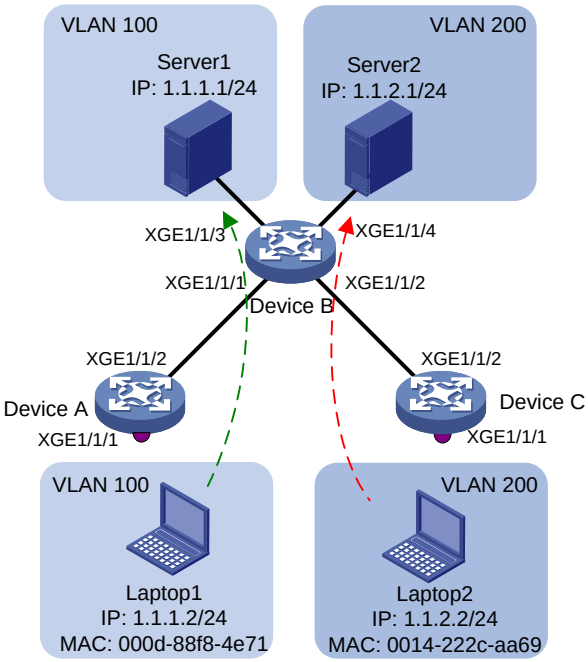

VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
WGE1/0/3
WGE1/0/4
WGE1/0/1
WGE1/0/2
Device B
WGE1/0/2
WGE1/0/2
Device A
Device C
WGE1/0/1
WGE1/0/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69

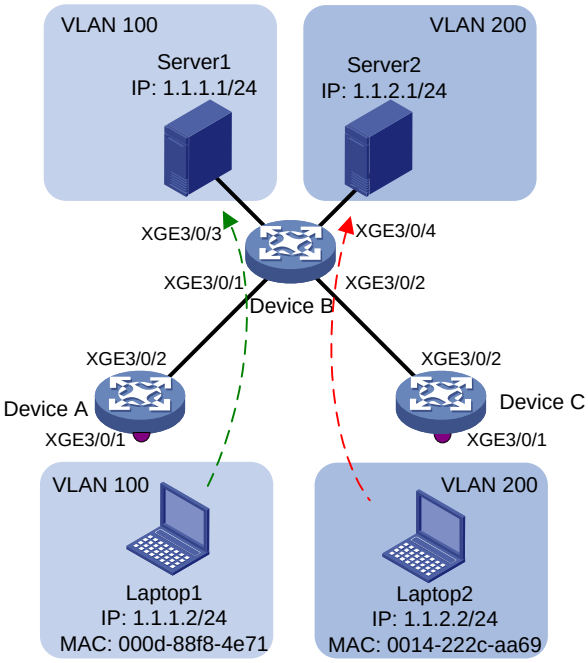

VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
FGE1/0/3
FGE1/0/4
FGE1/0/1
FGE1/0/2
Device B
FGE1/0/2
FGE1/0/2
Device A
Device C
FGE1/0/1
FGE1/0/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69

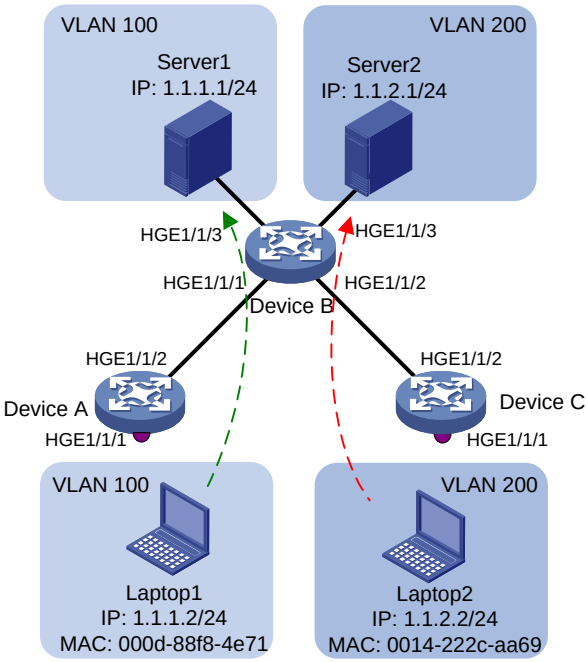

VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
Eth1/0/3
Eth1/0/4
Eth1/0/1
Eth1/0/2
Device B
Eth1/0/2
Eth1/0/2
Device A
Device C
Eth1/0/1
Eth1/0/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69


VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
HGE3/0/3
HGE3/0/4
HGE3/0/1
HGE3/0/2
Device B
HGE3/0/2
HGE3/0/2
Device A
Device C
HGE3/0/1
HGE3/0/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69


VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
XGE3/0/3
XGE3/0/4
XGE3/0/1
XGE3/0/2
Device B
XGE3/0/2
XGE3/0/2
Device A
Device C
XGE3/0/1
XGE3/0/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69


VLAN 100
Server1
IP: 1.1.1.1/24
VLAN 200
Server2
IP: 1.1.2.1/24
XGE1/1/3
XGE1/1/4
XGE1/1/1
XGE1/1/2
Device B
XGE1/1/2
XGE1/1/2
Device A
Device C
XGE1/1/1
XGE1/1/1
VLAN 100
VLAN 200
Laptop1
IP: 1.1.1.2/24
MAC: 000d-88f8-4e71
Laptop2
IP: 1.1.2.2/24
MAC: 0014-222c-aa69





Procedure

1. Configure Device A:

   # Create VLANs 100 and 200.

   ```
   <DeviceA> system-view
   [DeviceA] vlan 100
   [DeviceA-vlan100] quit
   [DeviceA] vlan 200
   [DeviceA-vlan200] quit
   ```

   # Associate the MAC addresses of Laptop 1 and Laptop 2 with VLANs 100 and 200, respectively.

```
[DeviceA] mac-vlan mac-address 000d-88f8-4e71 vlan 100
[DeviceA] mac-vlan mac-address 0014-222c-aa69 vlan 200
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign it to VLANs 100 and 200 as an untagged VLAN member.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type hybrid
[DeviceA-GigabitEthernet1/0/1] port hybrid vlan 100 200 untagged
```

Enable the MAC-based VLAN feature on GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] mac-vlan enable
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure the uplink port (GigabitEthernet 1/0/2) as a trunk port, and assign it to VLANs 100 and 200.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 100 200
[DeviceA-GigabitEthernet1/0/2] quit
```

2. Configure Device B:

Create VLAN 100, and assign GigabitEthernet 1/0/3 to VLAN 100.

```
<DeviceB> system-view
[DeviceB] vlan 100
[DeviceB-vlan100] port gigabitethernet 1/0/3
[DeviceB-vlan100] quit
```

Create VLAN 200 and assign GigabitEthernet 1/0/4 to VLAN 200.

```
[DeviceB] vlan 200
[DeviceB-vlan200] port gigabitethernet 1/0/4
[DeviceB-vlan200] quit
```

Configure GigabitEthernet 1/0/1 as a trunk port, and assign the port to VLANs 100 and 200.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 100 200
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port, and assign the port to VLANs 100 and 200.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 100 200
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device C in the same way as the Device A is configured. (Details not shown.)

Verifying the configuration

Verify that Laptop 1 can access only Server 1, and Laptop 2 can access only Server 2. (Details not shown.)

Verify the MAC-to-VLAN entries on Device A and Device C, for example, on Device A.

```
[DeviceA] display mac-vlan all
  The following MAC VLAN addresses exist:
  S:Static  D:Dynamic
  MAC address      Mask             VLAN ID   Dot1p      State
  000d-88f8-4e71   ffff-ffff-ffff   100       0          S
  0014-222c-aa69   ffff-ffff-ffff   200       0          S
```

```
Total MAC VLAN address count: 2
```

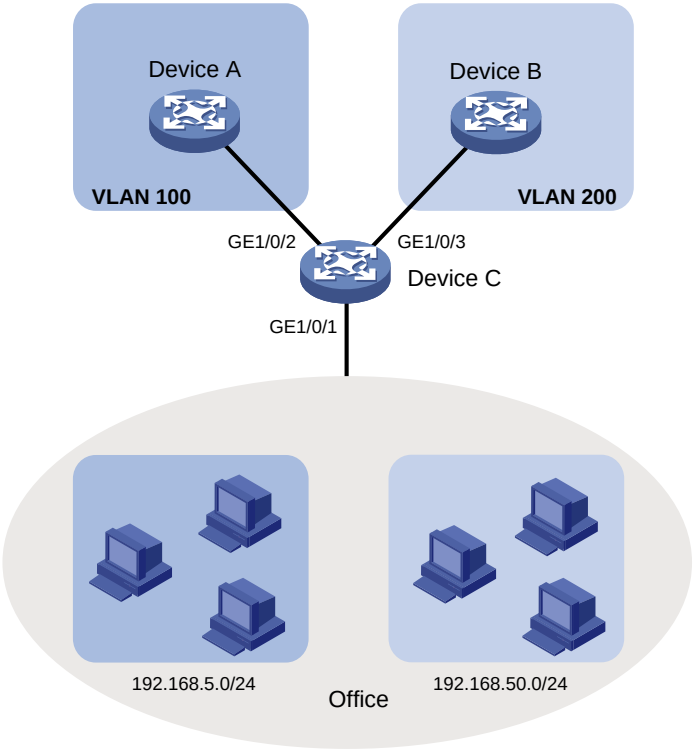
Example: Configuring IP subnet-based VLANs

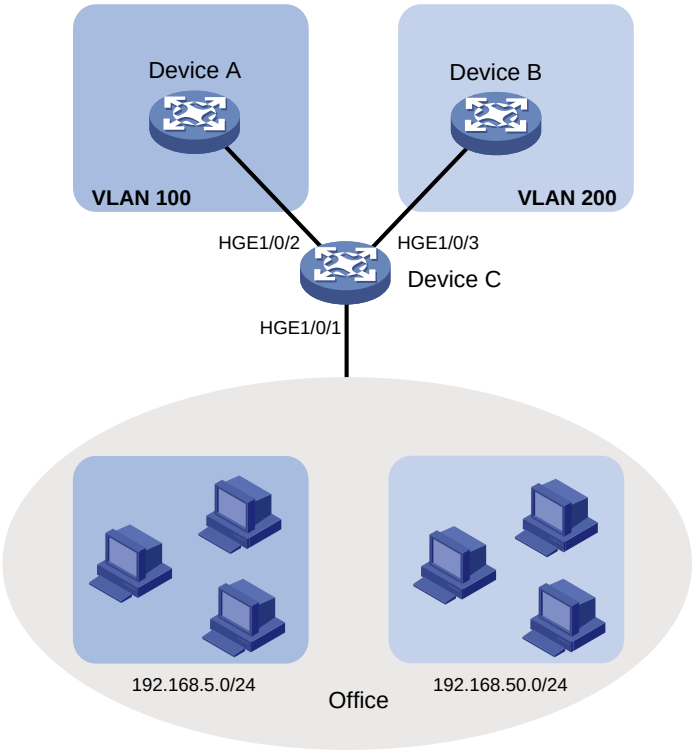
Network configuration

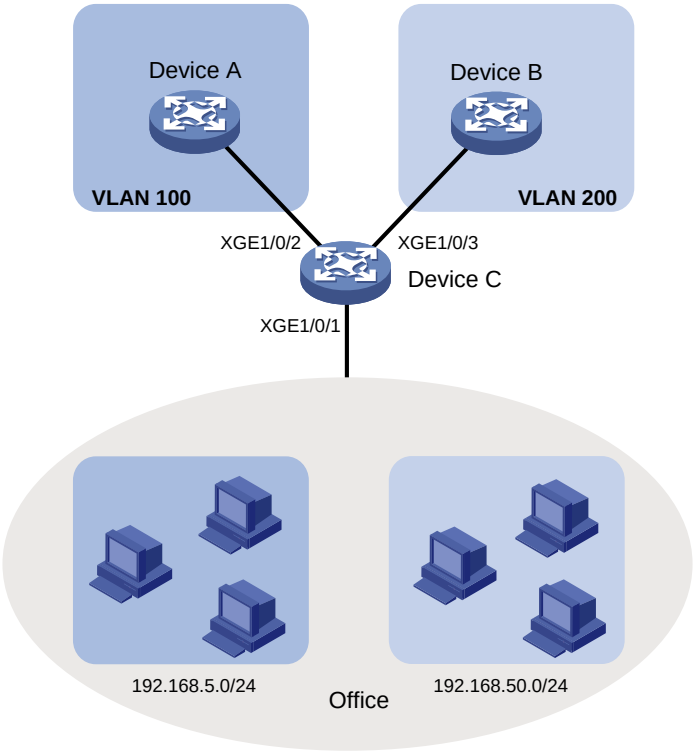
As shown in Figure 5, the hosts in the office belong to different IP subnets.

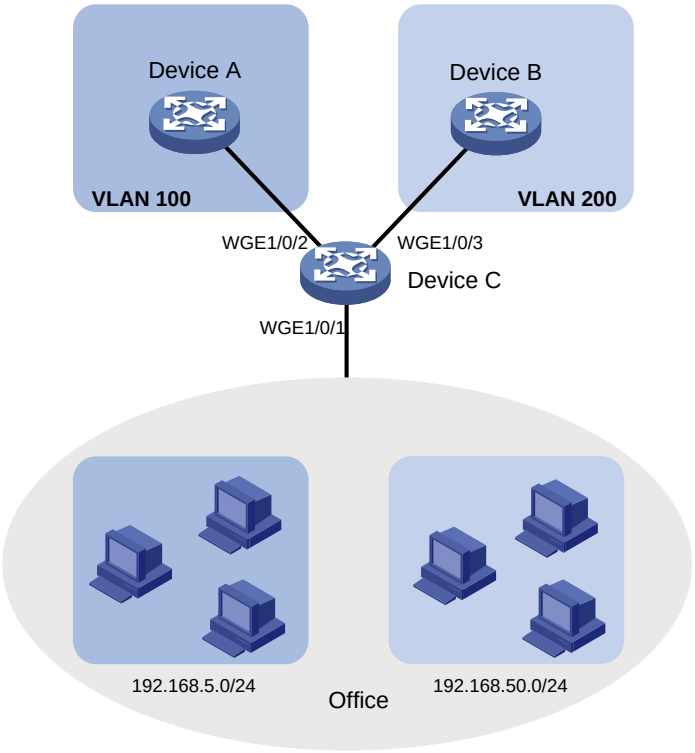
Configure Device C to transmit packets from 192.168.5.0/24 and 192.168.50.0/24 in VLANs 100 and 200, respectively.

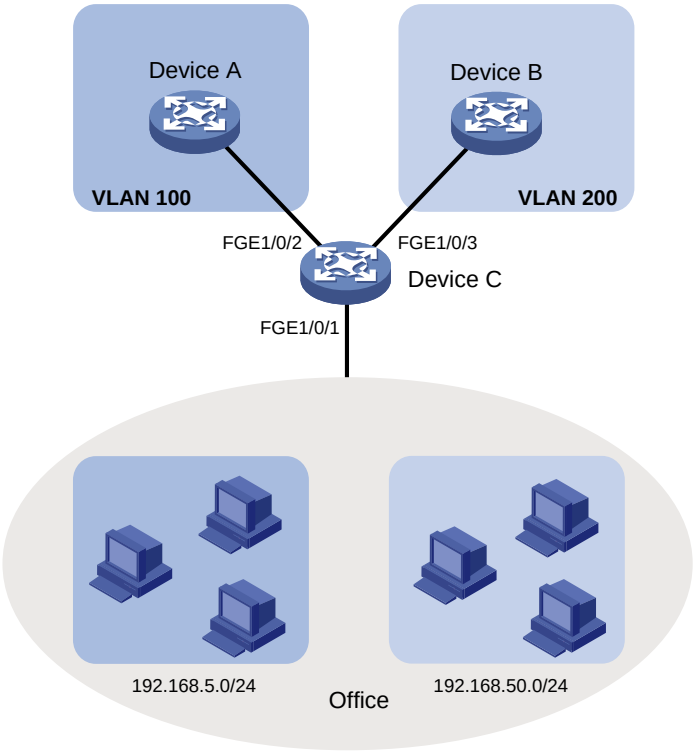
Figure 5 Network diagram

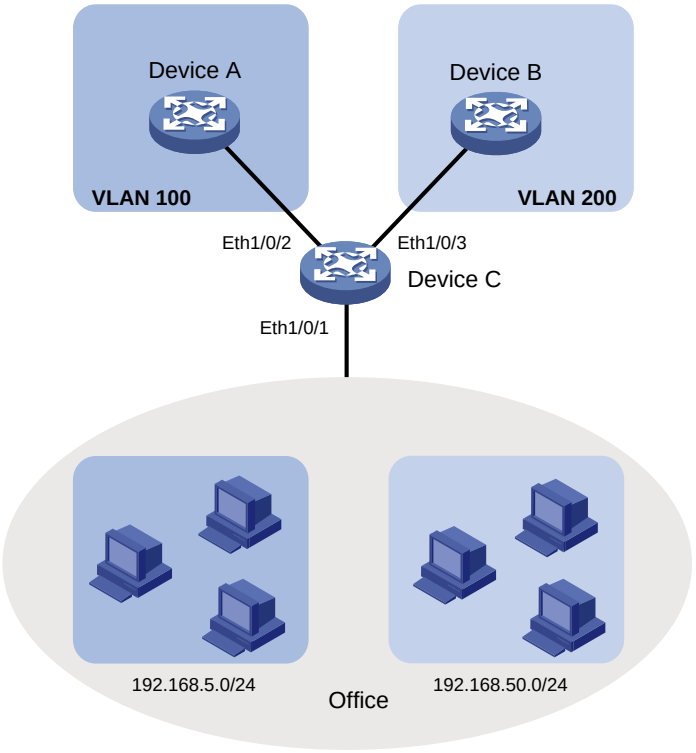


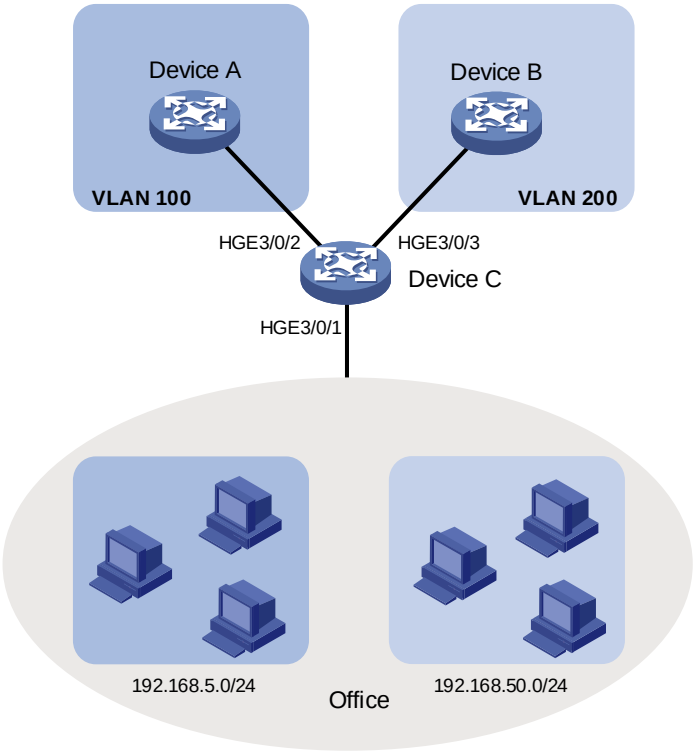

Device A
Device B
VLAN 100
VLAN 200
HGE1/0/2
HGE1/0/3
Device C
HGE1/0/1
192.168.5.0/24
Office
192.168.50.0/24

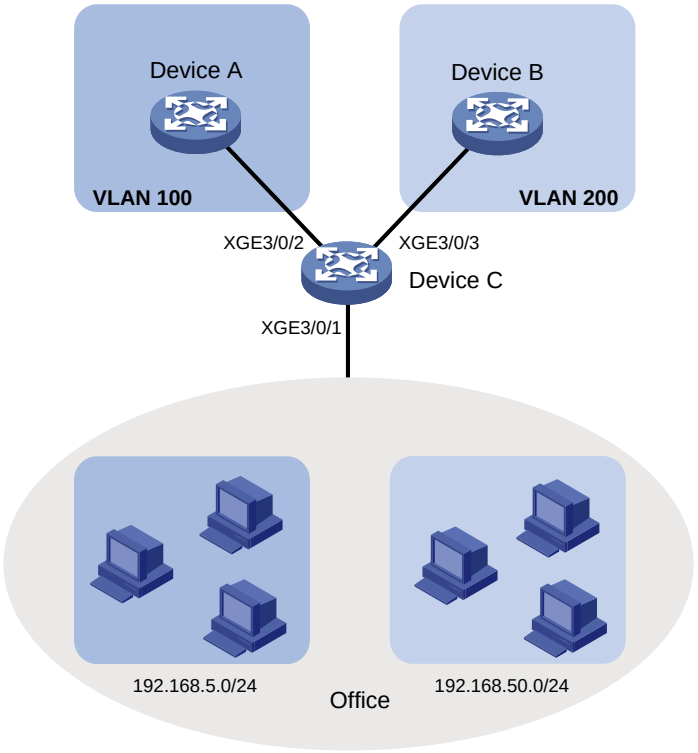

Device A
Device B
VLAN 100
VLAN 200
XGE1/0/2
XGE1/0/3
Device C
XGE1/0/1
192.168.5.0/24
Office
192.168.50.0/24

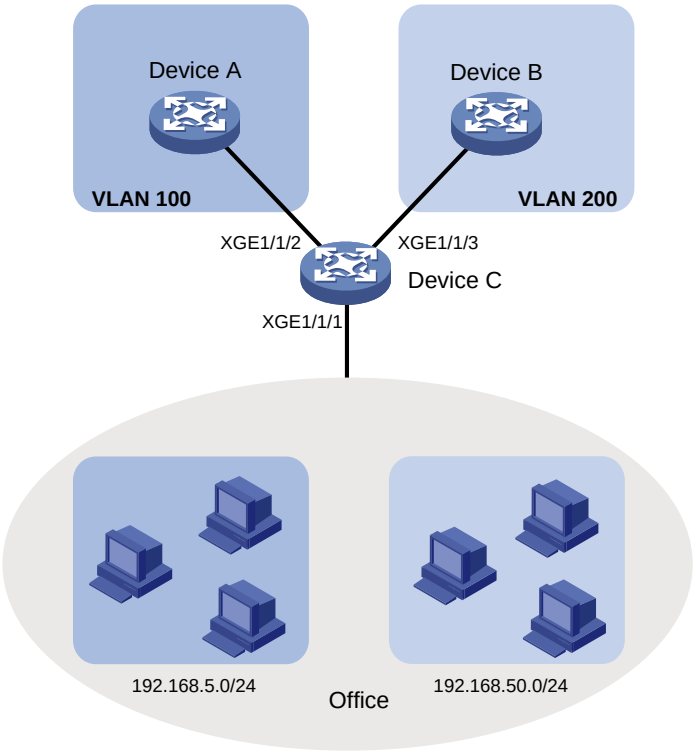

Device A
Device B
VLAN 100
VLAN 200
WGE1/0/2
WGE1/0/3
Device C
WGE1/0/1
192.168.5.0/24
Office
192.168.50.0/24

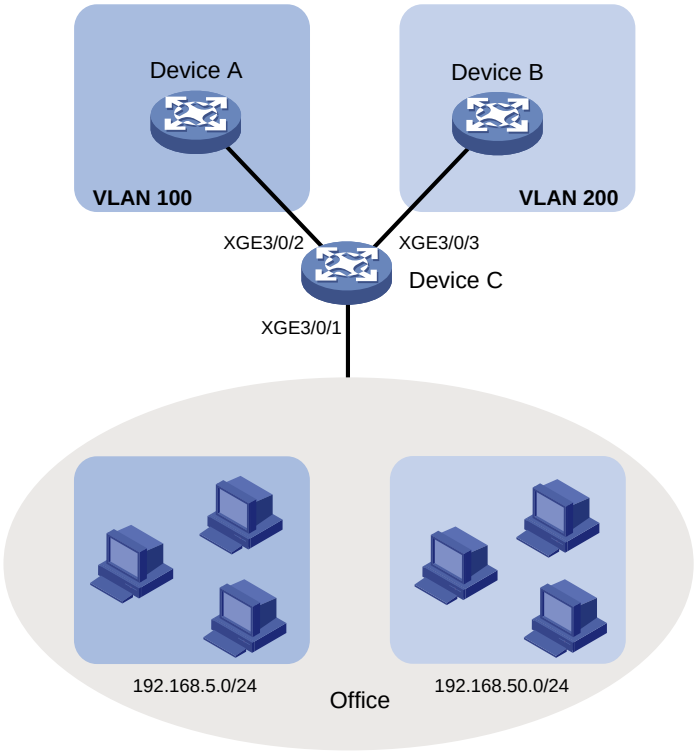

Device A
Device B
VLAN 100
VLAN 200
FGE1/0/2
FGE1/0/3
Device C
FGE1/0/1
192.168.5.0/24
Office
192.168.50.0/24

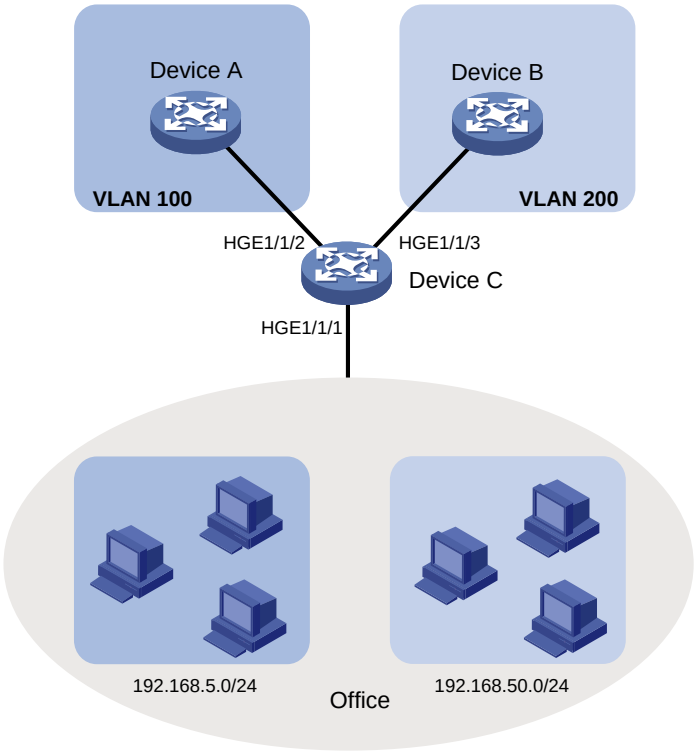

Device A
Device B
VLAN 100
VLAN 200
Eth1/0/2
Eth1/0/3
Device C
Eth1/0/1
192.168.5.0/24
Office
192.168.50.0/24


Device A
Device B
VLAN 100
VLAN 200
HGE3/0/2
HGE3/0/3
Device C
HGE3/0/1
192.168.5.0/24
Office
192.168.50.0/24


Device A
Device B
VLAN 100
VLAN 200
XGE3/0/2
XGE3/0/3
Device C
XGE3/0/1
192.168.5.0/24
Office
192.168.50.0/24


Device A
Device B
VLAN 100
VLAN 200
XGE1/1/2
XGE1/1/3
Device C
XGE1/1/1
192.168.5.0/24
Office
192.168.50.0/24





Procedure

1. Configure Device C:

   # Associate IP subnet 192.168.5.0/24 with VLAN 100.

   ```
   <DeviceC> system-view
   [DeviceC] vlan 100
   ```

```
[DeviceC-vlan100] ip-subnet-vlan ip 192.168.5.0 255.255.255.0
[DeviceC-vlan100] quit
```

Associate IP subnet 192.168.50.0/24 with VLAN 200.

```
[DeviceC] vlan 200
[DeviceC-vlan200] ip-subnet-vlan ip 192.168.50.0 255.255.255.0
[DeviceC-vlan200] quit
```

Configure GigabitEthernet 1/0/2 as a hybrid port, and assign it to VLAN 100 as a tagged VLAN member.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] port link-type hybrid
[DeviceC-GigabitEthernet1/0/2] port hybrid vlan 100 tagged
[DeviceC-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a hybrid port, and assign it to VLAN 200 as a tagged VLAN member.

```
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] port link-type hybrid
[DeviceC-GigabitEthernet1/0/3] port hybrid vlan 200 tagged
[DeviceC-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign it to VLANs 100 and 200 as an untagged VLAN member.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] port link-type hybrid
[DeviceC-GigabitEthernet1/0/1] port hybrid vlan 100 200 untagged
```

Associate GigabitEthernet 1/0/1 with the IP subnet-based VLANs 100 and 200.

```
[DeviceC-GigabitEthernet1/0/1] port hybrid ip-subnet-vlan vlan 100
[DeviceC-GigabitEthernet1/0/1] port hybrid ip-subnet-vlan vlan 200
[DeviceC-GigabitEthernet1/0/1] quit
```

2. Configure Device A and Device B to forward packets from VLANs 100 and 200, respectively. (Details not shown.)

Verifying the configuration

Verify the IP subnet-based VLAN configuration on Device C.

```
[DeviceC] display ip-subnet-vlan vlan all
 VLAN ID: 100
  Subnet index      IP address      Subnet mask
  0                 192.168.5.0     255.255.255.0

 VLAN ID: 200
  Subnet index      IP address      Subnet mask
  0                 192.168.50.0    255.255.255.0
```

Verify the IP subnet-based VLAN configuration on GigabitEthernet 1/0/1 of Device C.

```
[DeviceC] display ip-subnet-vlan interface gigabitethernet 1/0/1
 Interface: GigabitEthernet1/0/1
  VLAN ID   Subnet index    IP address       Subnet mask       Status
  100       0               192.168.5.0      255.255.255.0     Active
  200       0               192.168.50.0     255.255.255.0     Active
```

Example: Configuring protocol-based VLANs

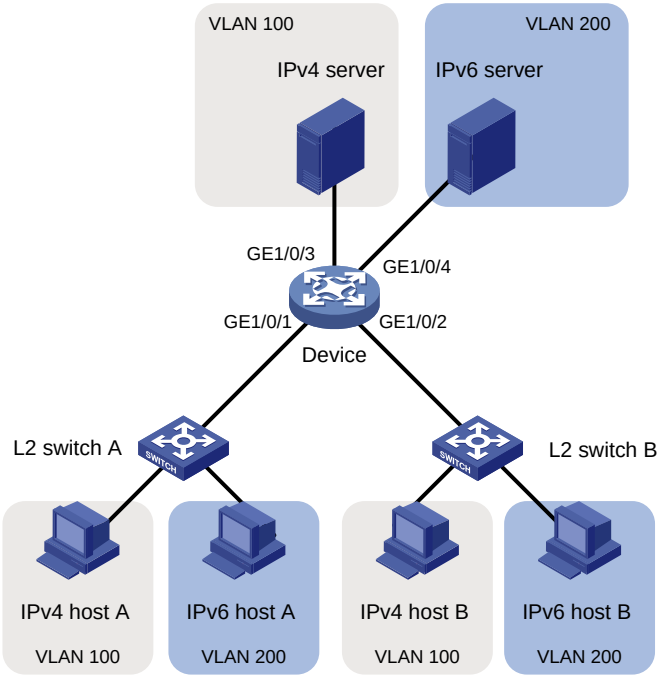
Network configuration

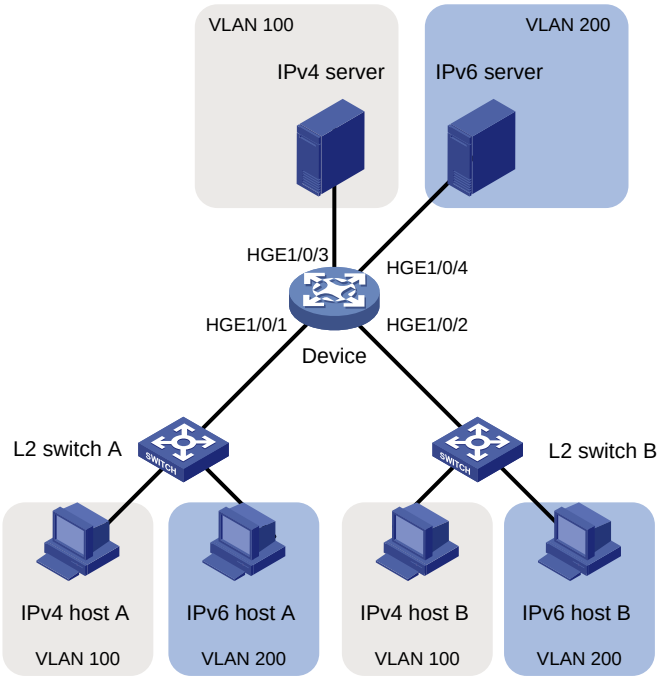
As shown in Figure 6:

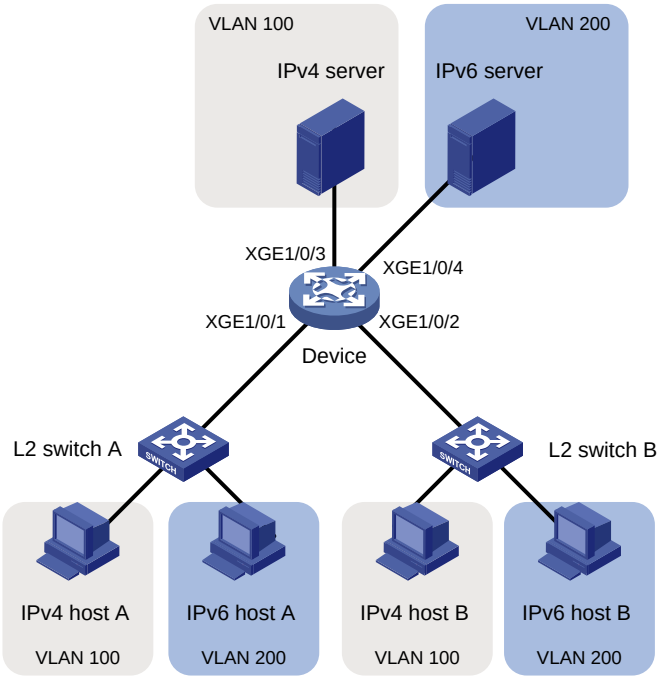
- The majority of hosts in a lab environment run the IPv4 protocol.
- The other hosts run the IPv6 protocol for teaching purposes.

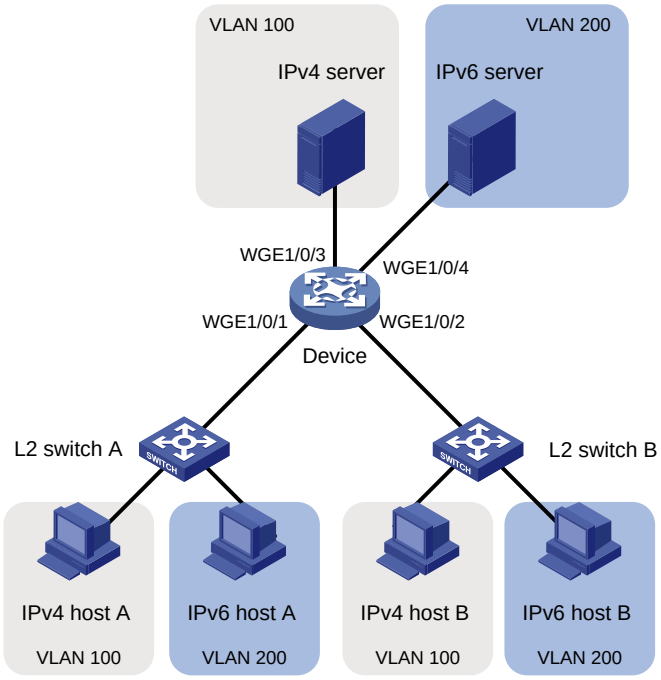
To isolate IPv4 and IPv6 traffic at Layer 2, configure protocol-based VLANs to associate the IPv4 and ARP protocols with VLAN 100, and associate the IPv6 protocol with VLAN 200.

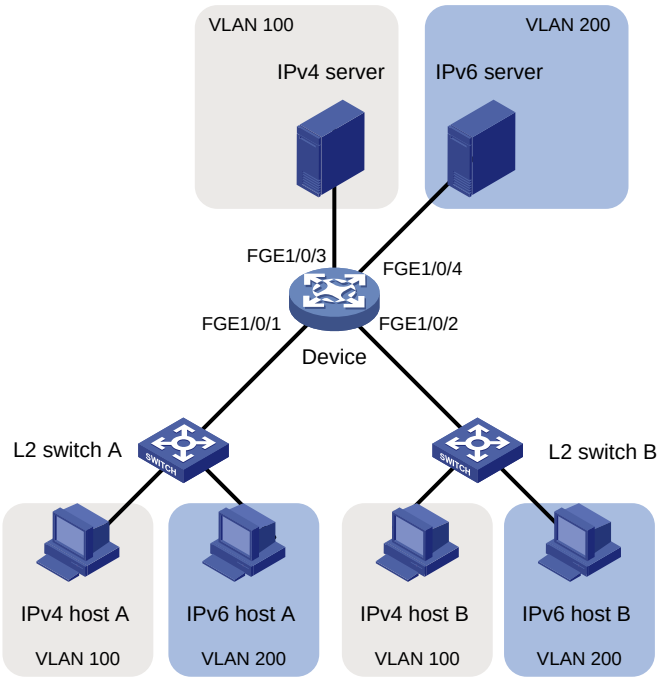
Figure 6 Network diagram

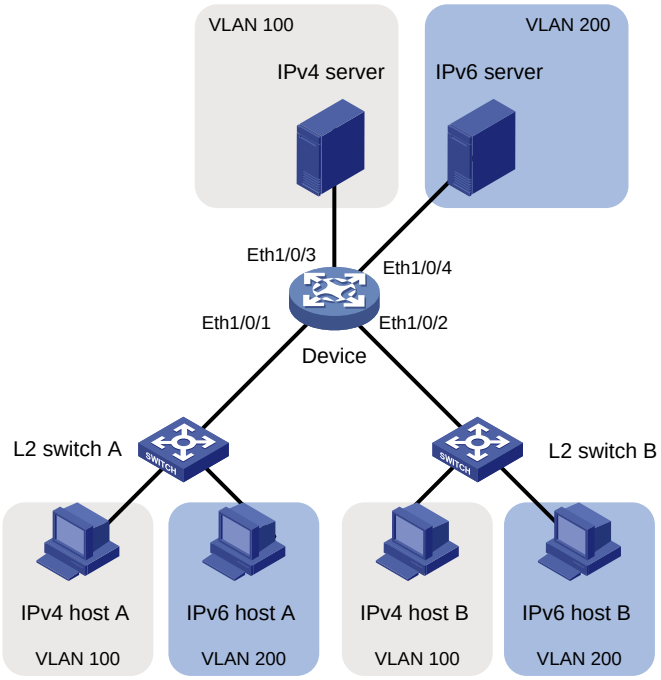


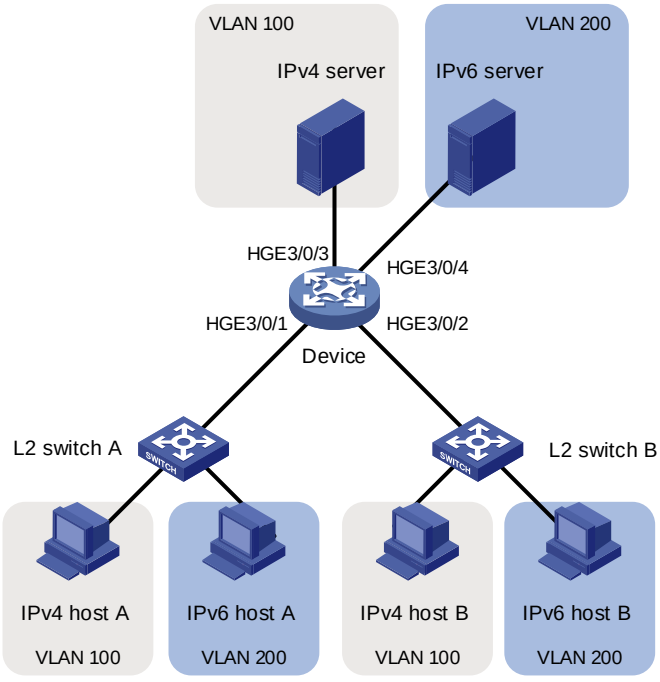

VLAN 100
VLAN 200
IPv4 server
IPv6 server
HGE1/0/3
HGE1/0/4
HGE1/0/1
HGE1/0/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200

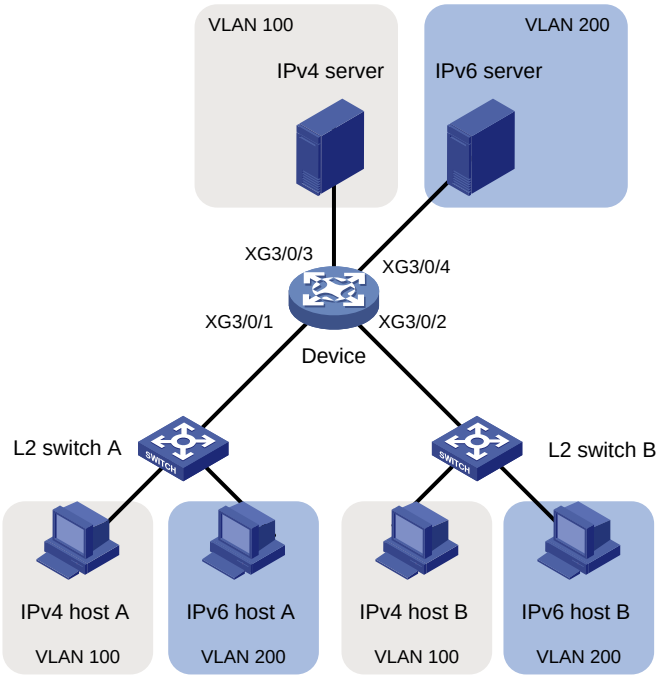

VLAN 100
VLAN 200
IPv4 server
IPv6 server
XGE1/0/3
XGE1/0/4
XGE1/0/1
XGE1/0/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200

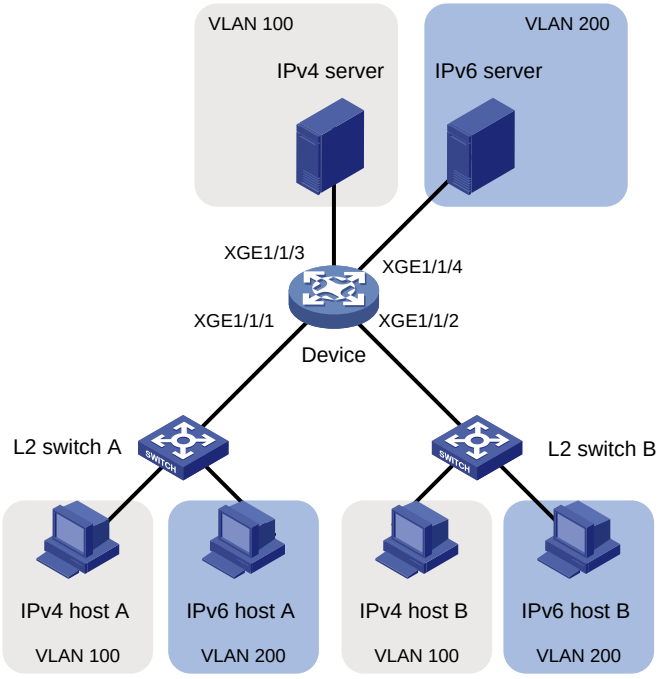

VLAN 100
VLAN 200
IPv4 server
IPv6 server
WGE1/0/3
WGE1/0/4
WGE1/0/1
WGE1/0/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200

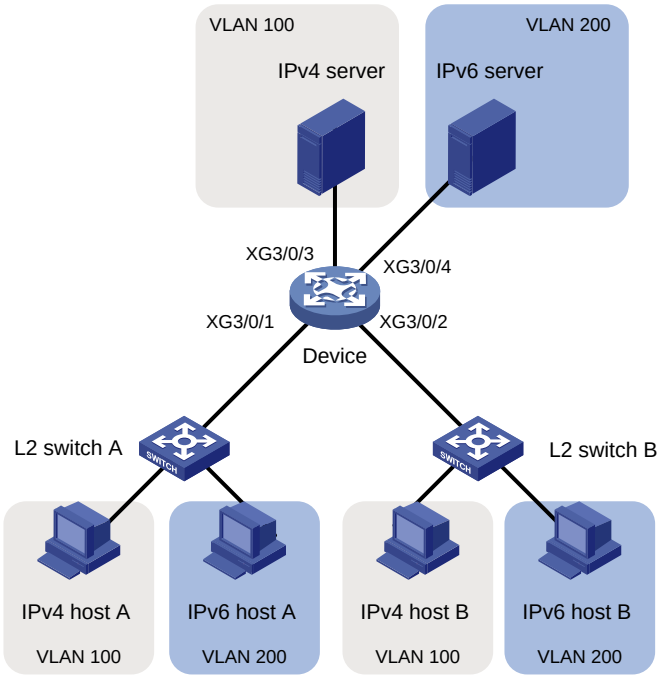

VLAN 100
VLAN 200
IPv4 server
IPv6 server
FGE1/0/3
FGE1/0/4
FGE1/0/1
FGE1/0/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200

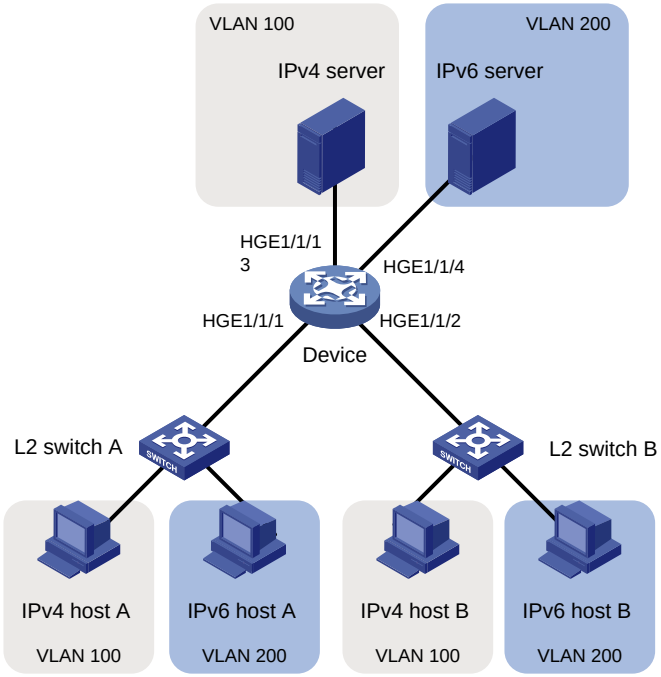

VLAN 100
VLAN 200
IPv4 server
IPv6 server
Eth1/0/3
Eth1/0/4
Eth1/0/1
Eth1/0/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200


VLAN 100
VLAN 200
IPv4 server
IPv6 server
HGE3/0/3
HGE3/0/4
HGE3/0/1
HGE3/0/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200


VLAN 100
VLAN 200
IPv4 server
IPv6 server
XG3/0/3
XG3/0/4
XG3/0/1
XG3/0/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200


VLAN 100
VLAN 200
IPv4 server
IPv6 server
XGE1/1/3
XGE1/1/4
XGE1/1/1
XGE1/1/2
Device
L2 switch A
L2 switch B
IPv4 host A
VLAN 100
IPv6 host A
VLAN 200
IPv4 host B
VLAN 100
IPv6 host B
VLAN 200





Procedure

In this example, L2 Switch A and L2 Switch B use the factory configuration.

1. Configure Device:

   # Create VLAN 100, and configure the description for VLAN 100 as **protocol VLAN for IPv4**.

   ```
   <Device> system-view
   [Device] vlan 100
   [Device-vlan100] description protocol VLAN for IPv4
   ```

   # Assign GigabitEthernet 1/0/3 to VLAN 100.

   ```
   [Device-vlan100] port gigabitethernet 1/0/3
   ```

```
[Device-vlan100] quit
```

Create VLAN 200, and configure the description for VLAN 200 as **protocol VLAN for IPv6**.

```
[Device] vlan 200
[Device-vlan200] description protocol VLAN for IPv6
```

Assign GigabitEthernet 1/0/4 to VLAN 200.

```
[Device-vlan200] port gigabitethernet 1/0/4
```

Configure VLAN 200 as a protocol-based VLAN, and create an IPv6 protocol template with the index 1 for VLAN 200.

```
[Device-vlan200] protocol-vlan 1 ipv6
[Device-vlan200] quit
```

Configure VLAN 100 as a protocol-based VLAN. Create an IPv4 protocol template with the index 1, and create an ARP protocol template with the index 2. (In Ethernet II encapsulation, the protocol type ID for ARP is 0806 in hexadecimal notation.)

```
[Device] vlan 100
[Device-vlan100] protocol-vlan 1 ipv4
[Device-vlan100] protocol-vlan 2 mode ethernetii etype 0806
[Device-vlan100] quit
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign it to VLANs 100 and 200 as an untagged VLAN member.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] port link-type hybrid
[Device-GigabitEthernet1/0/1] port hybrid vlan 100 200 untagged
```

Associate GigabitEthernet 1/0/1 with the IPv4 and ARP protocol templates of VLAN 100 and the IPv6 protocol template of VLAN 200.

```
[Device-GigabitEthernet1/0/1] port hybrid protocol-vlan vlan 100 1 to 2
[Device-GigabitEthernet1/0/1] port hybrid protocol-vlan vlan 200 1
[Device-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a hybrid port, and assign it to VLANs 100 and 200 as an untagged VLAN member.

```
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] port link-type hybrid
[Device-GigabitEthernet1/0/2] port hybrid vlan 100 200 untagged
```

Associate GigabitEthernet 1/0/2 with the IPv4 and ARP protocol templates of VLAN 100 and the IPv6 protocol template of VLAN 200.

```
[Device-GigabitEthernet1/0/2] port hybrid protocol-vlan vlan 100 1 to 2
[Device-GigabitEthernet1/0/2] port hybrid protocol-vlan vlan 200 1
[Device-GigabitEthernet1/0/2] quit
```

2. Configure hosts and servers:
   a. Configure IPv4 Host A, IPv4 Host B, and IPv4 server to be on the same network segment (192.168.100.0/24, for example). (Details not shown.)
   b. Configure IPv6 Host A, IPv6 Host B, and IPv6 server to be on the same network segment (2001::1/64, for example). (Details not shown.)

Verifying the configuration

1. Verify the following:
   - The hosts and the server in VLAN 100 can successfully ping one another. (Details not shown.)
   - The hosts and the server in VLAN 200 can successfully ping one another. (Details not shown.)

- The hosts or the server in VLAN 100 cannot ping the hosts or server in VLAN 200. (Details not shown.)

2. Verify the protocol-based VLAN configuration:

Display protocol-based VLANs on Device.

```
[Device] display protocol-vlan vlan all
 VLAN ID: 100
  Protocol index  Protocol type
  1               IPv4
  2               Ethernet II Etype 0x0806

 VLAN ID: 200
  Protocol index  Protocol type
  1               IPv6
```

Display protocol-based VLANs on the ports of Device.

```
[Device] display protocol-vlan interface all
 Interface: GigabitEthernet1/0/1
  VLAN ID  Protocol index  Protocol type             Status
  100      1               IPv4                     Active
  100      2               Ethernet II Etype 0x0806  Active
  200      1               IPv6                     Active

 Interface: GigabitEthernet 1/0/2
  VLAN ID  Protocol index  Protocol type             Status
  100      1               IPv4                     Active
  100      2               Ethernet II Etype 0x0806  Active
  200      1               IPv6                     Active
```

Configuring super VLANs

About super VLANs

Hosts in a VLAN typically use IP addresses in the same subnet. For Layer 3 interoperability with other VLANs, you can create a VLAN interface for the VLAN and assign an IP address to it. This requires a large number of IP addresses.

The super VLAN feature was introduced to save IP addresses. A super VLAN is associated with multiple sub-VLANs. These sub-VLANs use the VLAN interface of the super VLAN (also known as a super VLAN interface) as the gateway for Layer 3 communication.

You can create a VLAN interface for a super VLAN and assign an IP address to it. However, you cannot create a VLAN interface for a sub-VLAN. You can assign a physical port to a sub-VLAN, but you cannot assign a physical port to a super VLAN. Sub-VLANs are isolated at Layer 2.

To enable Layer 3 communication between sub-VLANs, perform the following tasks:

1. Create a super VLAN and the VLAN interface for the super VLAN.
2. Enable local proxy ARP or ND on the super VLAN interface as follows:
   - In an IPv4 network, enable local proxy ARP on the super VLAN interface. The super VLAN can then process ARP requests and replies sent from the sub-VLANs.
   - In an IPv6 network, enable local proxy ND on the super VLAN interface. The super VLAN can then process the NS and NA messages sent from the sub-VLANs.

Restrictions and guidelines: Super VLAN configuration

- The VLAN of a MAC address-to-VLAN entry cannot be configured as a super VLAN.
- A VLAN cannot be configured as both a super VLAN and a guest VLAN, Auth-Fail VLAN, or critical VLAN. For more information about guest VLANs, Auth-Fail VLANs, and critical VLANs, see *Security Configuration Guide*.
- A VLAN cannot be configured as both a super VLAN and a sub-VLAN.
- Layer 2 multicast configuration for super VLANs does not take effect because they do not have physical ports.

Super VLAN tasks at a glance

To configure a super VLAN, perform the following tasks:

1. Creating a sub-VLAN
2. Configuring a super VLAN
3. Configuring a super VLAN interface

Creating a sub-VLAN

1. Enter system view.

   `system-view`
2. Create a sub-VLAN.

vlan *vlan-id-list*

By default, only the system default VLAN (VLAN 1) exists.

Configuring a super VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Configure the VLAN as a super VLAN.
   supervlan
   By default, a VLAN is not a super VLAN.
4. Associate the super VLAN with the sub-VLANs.
   subvlan *vlan-id-list*
   Make sure the sub-VLANs already exist before associating them with a super VLAN.

Configuring a super VLAN interface

Restrictions and guidelines

As a best practice, do not configure VRRP for a super VLAN interface because the configuration affects network performance. For more information about VRRP, see *High Availability Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Create a VLAN interface and enter its view.
   interface vlan-interface *interface-number*
   The value for the *interface-number* argument must be the super VLAN ID.
3. Configure an IP address for the super VLAN interface.
   IPv4:
   ip address *ip-address* { *mask-length* | *mask* } [**sub**]
   IPv6:
   ipv6 address { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }
   By default, no IP address is configured for a VLAN interface.
4. Configure Layer 3 communication between sub-VLANs by enabling local proxy ARP or ND.
   IPv4:
   local-proxy-arp enable
   By default:
   - Sub-VLANs cannot communicate with each other at Layer 3.
   - Local proxy ARP is disabled.

   For more information about local proxy ARP, see *Layer 3—IP Services Configuration Guide*.
   IPv6:
   local-proxy-nd enable
   By default:

- Sub-VLANs cannot communicate with each other at Layer 3.
- Local proxy ND is disabled.

For more information about local proxy ND, see *Layer 3—IP Services Configuration Guide*.

Display and maintenance commands for super VLANs

Execute **display** commands in any view.

Task	Command
Display information about super VLANs and their associated sub-VLANs.	**display supervlan** [*supervlan-id*]

Super VLAN configuration examples

Example: Configuring a super VLAN

Network configuration

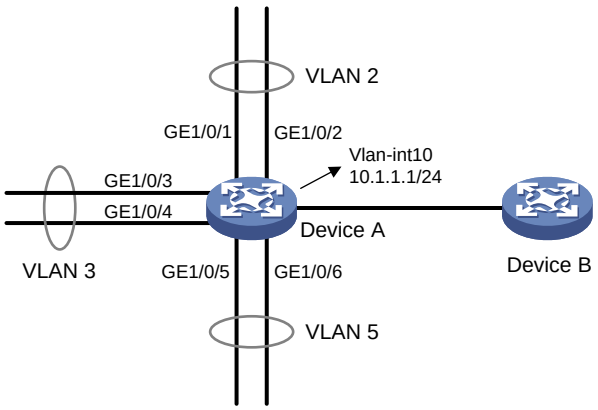
As shown in Figure 7:

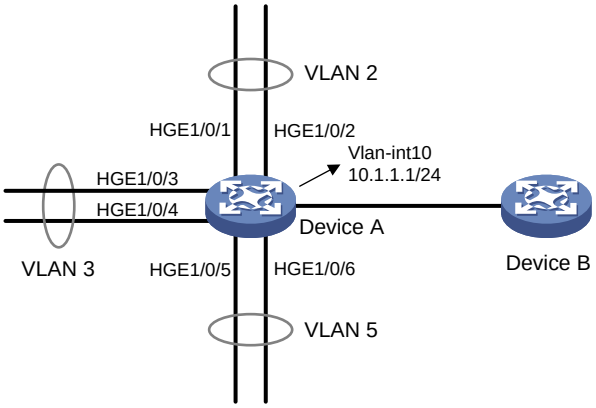
- GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are in VLAN 2.
- GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 are in VLAN 3.
- GigabitEthernet 1/0/5 and GigabitEthernet 1/0/6 are in VLAN 5.

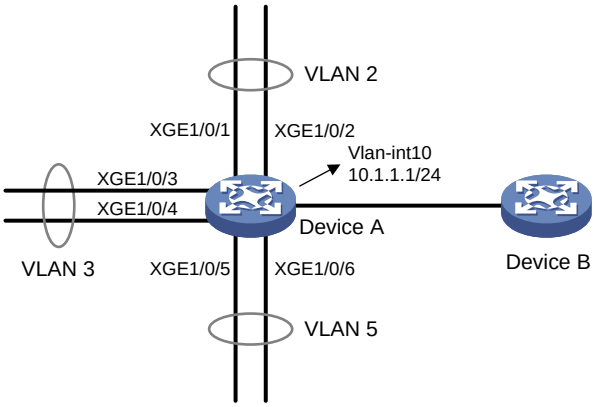
To save IP addresses and enable sub-VLANs to be isolated at Layer 2 but interoperable at Layer 3, perform the following tasks:

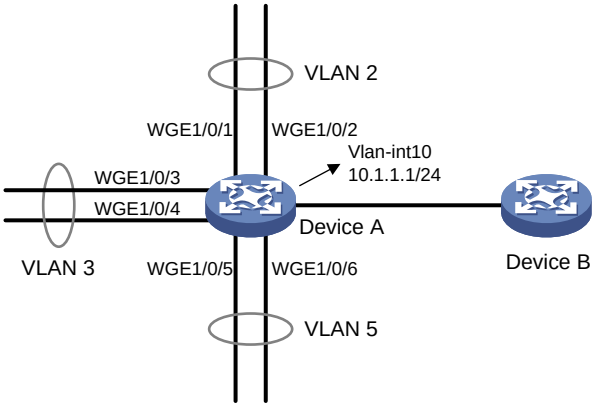
- Create a super VLAN and assign an IP address to its VLAN interface.
- Associate the super VLAN with VLANs 2, 3, and 5.

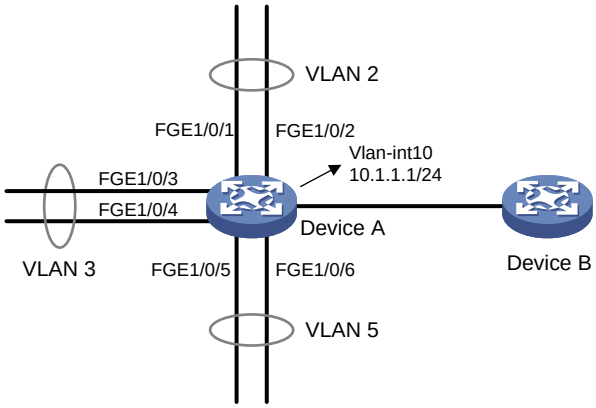
Figure 7 Network diagram

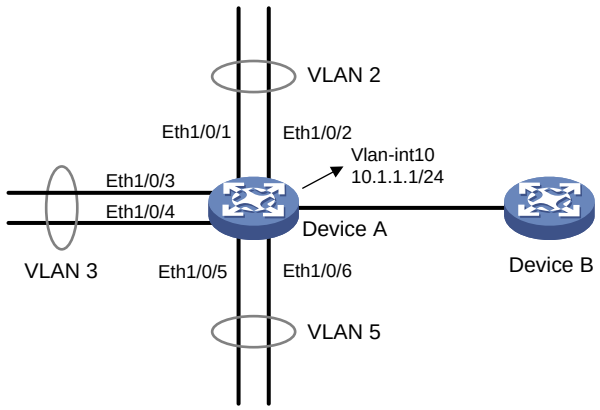


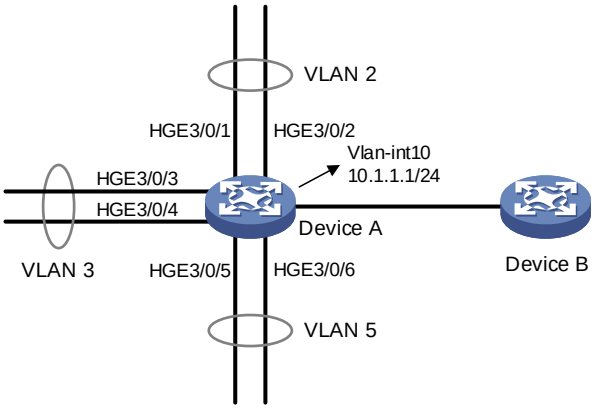

VLAN 2
HGE1/0/1
HGE1/0/2
Vlan-int10
10.1.1.1/24
HGE1/0/3
HGE1/0/4
Device A
Device B
VLAN 3
HGE1/0/5
HGE1/0/6
VLAN 5

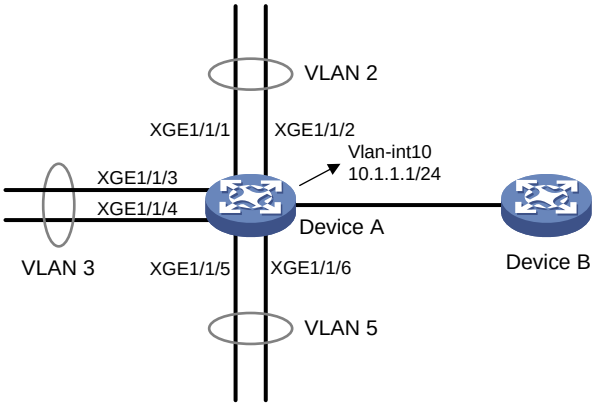

VLAN 2
XGE1/0/1
XGE1/0/2
Vlan-int10
10.1.1.1/24
XGE1/0/3
XGE1/0/4
Device A
Device B
VLAN 3
XGE1/0/5
XGE1/0/6
VLAN 5

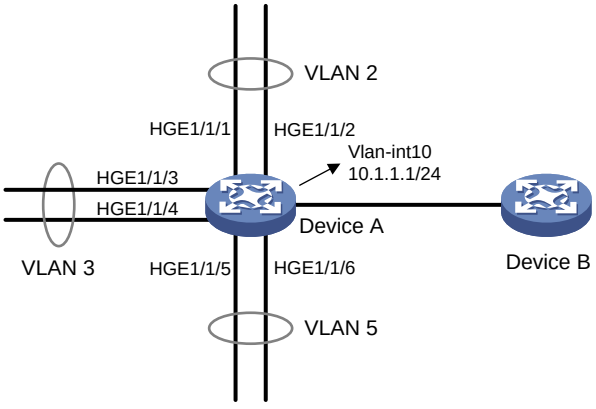

VLAN 2
WGE1/0/1
WGE1/0/2
Vlan-int10
10.1.1.1/24
WGE1/0/3
WGE1/0/4
Device A
Device B
VLAN 3
WGE1/0/5
WGE1/0/6
VLAN 5

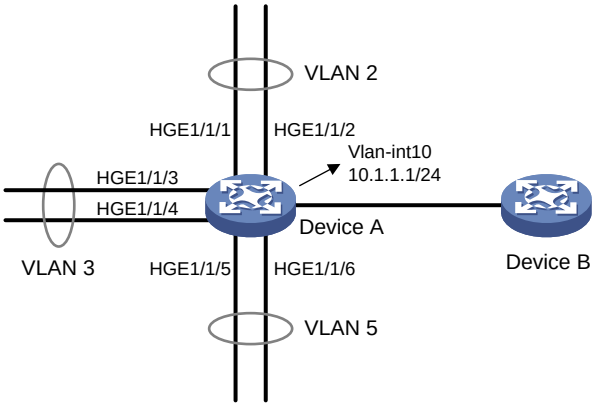

VLAN 2
FGE1/0/1
FGE1/0/2
Vlan-int10
10.1.1.1/24
FGE1/0/3
FGE1/0/4
Device A
Device B
VLAN 3
FGE1/0/5
FGE1/0/6
VLAN 5


VLAN 2
Eth1/0/1
Eth1/0/2
Vlan-int10
10.1.1.1/24
Eth1/0/3
Eth1/0/4
Device A
Device B
VLAN 3
Eth1/0/5
Eth1/0/6
VLAN 5


VLAN 2
HGE3/0/1
HGE3/0/2
Vlan-int10
10.1.1.1/24
HGE3/0/3
HGE3/0/4
Device A
Device B
VLAN 3
HGE3/0/5
HGE3/0/6
VLAN 5


VLAN 2
XGE1/1/1
XGE1/1/2
Vlan-int10
10.1.1.1/24
XGE1/1/3
XGE1/1/4
Device A
Device B
VLAN 3
XGE1/1/5
XGE1/1/6
VLAN 5


VLAN 2
HGE1/1/1
HGE1/1/2
Vlan-int10
10.1.1.1/24
HGE1/1/3
HGE1/1/4
Device A
Device B
VLAN 3
HGE1/1/5
HGE1/1/6
VLAN 5



Procedure

Create VLAN 10.

```
<DeviceA> system-view
[DeviceA] vlan 10
[DeviceA-vlan10] quit
```

Create VLAN-interface 10, and assign IP address 10.1.1.1/24 to it.

```
[DeviceA] interface vlan-interface 10
[DeviceA-Vlan-interface10] ip address 10.1.1.1 255.255.255.0
```

Enable local proxy ARP.

```
[DeviceA-Vlan-interface10] local-proxy-arp enable
[DeviceA-Vlan-interface10] quit
```

Create VLAN 2, and assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the VLAN.

```
[DeviceA] vlan 2
[DeviceA-vlan2] port gigabitethernet 1/0/1 gigabitethernet 1/0/2
[DeviceA-vlan2] quit
```

Create VLAN 3, and assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to the VLAN.

```
[DeviceA] vlan 3
[DeviceA-vlan3] port gigabitethernet 1/0/3 gigabitethernet 1/0/4
[DeviceA-vlan3] quit
```

Create VLAN 5, and assign GigabitEthernet 1/0/5 and GigabitEthernet 1/0/6 to the VLAN.

```
[DeviceA] vlan 5
[DeviceA-vlan5] port gigabitethernet 1/0/5 gigabitethernet 1/0/6
[DeviceA-vlan5] quit
```

Configure VLAN 10 as a super VLAN, and associate sub-VLANs 2, 3, and 5 with the super VLAN.

```
[DeviceA] vlan 10
[DeviceA-vlan10] supervlan
[DeviceA-vlan10] subvlan 2 3 5
[DeviceA-vlan10] quit
[DeviceA] quit
```

Verifying the configuration

Display information about super VLAN 10 and its associated sub-VLANs.

```
<DeviceA> display supervlan
 Super VLAN ID: 10
 Sub-VLAN ID: 2-3 5
```

```
VLAN ID: 10
VLAN type: Static
It is a super VLAN.
Route interface: Configured
Ipv4 address: 10.1.1.1
Ipv4 subnet mask: 255.255.255.0
Description: VLAN 0010
Name: VLAN 0010
Tagged ports:   None
Untagged ports: None
VLAN ID: 2
VLAN type: Static
It is a sub VLAN.
Route interface: Configured
Ipv4 address: 10.1.1.1
Ipv4 subnet mask: 255.255.255.0
Description: VLAN 0002
Name: VLAN 0002
Tagged ports:   None
Untagged ports:
   GigabitEthernet1/0/1
   GigabitEthernet1/0/2
VLAN ID: 3
VLAN type: Static
It is a sub VLAN.
Route interface: Configured
Ipv4 address: 10.1.1.1
Ipv4 subnet mask: 255.255.255.0
Description: VLAN 0003
Name: VLAN 0003
Tagged ports:   None
Untagged ports:
   GigabitEthernet1/0/3
   GigabitEthernet1/0/4
VLAN ID: 5
VLAN type: Static
It is a sub VLAN.
Route interface: Configured
Ipv4 address: 10.1.1.1
Ipv4 subnet mask: 255.255.255.0
Description: VLAN 0005
Name: VLAN 0005
Tagged ports:   None
Untagged ports:
   GigabitEthernet1/0/5
   GigabitEthernet1/0/6
```

Configuring private VLAN

About private VLAN

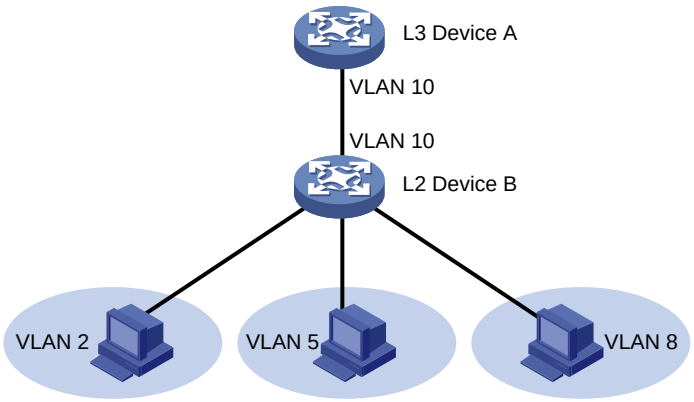
VLAN technology provides a method for isolating traffic from customers. At the access layer of a network, customer traffic must be isolated for security or accounting purposes. If VLANs are assigned on a per-user basis, a large number of VLANs will be required.

The private VLAN feature saves VLAN resources. It uses a two-tier VLAN structure as follows:

- **Primary VLAN**—Used for connecting the upstream device. A primary VLAN can be associated with multiple secondary VLANs. The upstream device identifies only the primary VLAN.
- **Secondary VLANs**—Used for connecting users. Secondary VLANs are isolated at Layer 2.

As shown in Figure 8, the private VLAN feature is enabled on L2 Device B. VLAN 10 is the primary VLAN. VLANs 2, 5, and 8 are secondary VLANs that are associated with VLAN 10. L3 Device A is only aware of VLAN 10.

Figure 8 Private VLAN example



If the private VLAN feature is configured on a Layer 3 device, use one of the following methods on the Layer 3 device to enable Layer 3 communication. Layer 3 communication might be required between secondary VLANs that are associated with the same primary VLAN, or between secondary VLANs and other networks.

- Method 1:
  - **a.** Create VLAN interfaces for the secondary VLANs.
  - **b.** Assign IP addresses to the secondary VLAN interfaces.
- Method 2:
  - **c.** Enable Layer 3 communication between the secondary VLANs that are associated with the primary VLAN.
  - **d.** Create the VLAN interface for the primary VLAN and assign an IP address to it. (Do not create secondary VLAN interfaces if you use this method.)
  - **e.** Enable local proxy ARP or ND on the primary VLAN interface.

Restrictions and guidelines: Private VLAN configuration

- Make sure the following requirements are met:

    - For a promiscuous port:
        - The primary VLAN is the PVID of the port.
        - The port is an untagged member of the primary VLAN and secondary VLANs.
    - For a host port:
        - The PVID of the port is a secondary VLAN.
        - The port is an untagged member of the primary VLAN and the secondary VLAN.
    - A trunk promiscuous or trunk secondary port must be a tagged member of the primary VLANs and the secondary VLANs.
- After you configure a primary VLAN, the system automatically synchronizes the dynamic MAC address entries of the primary VLAN with the dynamic MAC address entries of the secondary VLANs. After you configure Layer 2 isolation for ports in a secondary VLAN, a port with the following configuration permits all VLAN packets in the inbound direction and forwards them according to VLAN-related configurations in the outbound orientation.
    - The port operates in trunk secondary mode in that VLAN.
    - The port operates in host mode and has the PVID as that VLAN.
- VLAN 1 (system default VLAN) does not support the private VLAN configuration.

Private VLAN tasks at a glance

To configure a private VLAN, perform the following tasks:

1. Creating a primary VLAN
2. Creating secondary VLANs
3. Associating the primary VLAN with secondary VLANs
4. Configuring the uplink port
5. Configuring a downlink port
6. (Optional.) Configuring Layer 3 communication for secondary VLANs

Creating a primary VLAN

1. Enter system view.
   system-view
2. Create a VLAN and enter VLAN view.
   vlan *vlan-id*
3. Configure the VLAN as a primary VLAN.
   private-vlan primary
   By default, a VLAN is not a primary VLAN.

Creating secondary VLANs

1. Enter system view.
   system-view
2. Create one or multiple secondary VLANs.
   vlan { *vlan-id-list* | **all** }

Associating the primary VLAN with secondary VLANs

1. Enter system view.
   system-view
2. Create enter VLAN view of the primary VLAN.
   vlan *vlan-id*
3. Associate the primary VLAN with the secondary VLANs.
   private-vlan secondary *vlan-id-list*
   By default, a primary VLAN is not associated with any secondary VLANs.

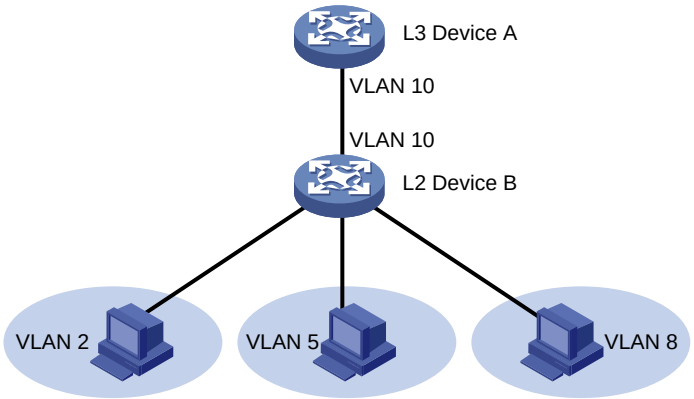
Configuring the uplink port

About this task

Configure the uplink port (for example, the port connecting L2 Device B to L3 Device A in Figure 9) as follows:

- If the port allows only one primary VLAN, configure the port as a promiscuous port of the primary VLAN. The promiscuous port can be automatically assigned to the primary VLAN and its associated secondary VLANs.
- If the port allows multiple primary VLANs, configure the port as a trunk promiscuous port of the primary VLANs. The trunk promiscuous port can be automatically assigned to the primary VLANs and their associated secondary VLANs.

Figure 9 Private VLAN example



Procedure

1. Enter system view.
   system-view
2. Enter interface view of the uplink port.
   interface *interface-type interface-number*
3. Configure the uplink port as a promiscuous or trunk promiscuous port of the specified VLANs.
   - Configure the uplink port as a promiscuous port of the specified VLAN.
     port private-vlan *vlan-id* **promiscuous**
   - Configure the uplink port as a trunk promiscuous port of the specified VLANs.

port private-vlan *vlan-id-list* **trunk promiscuous**

By default, a port is not a promiscuous or trunk promiscuous port of any VLANs.

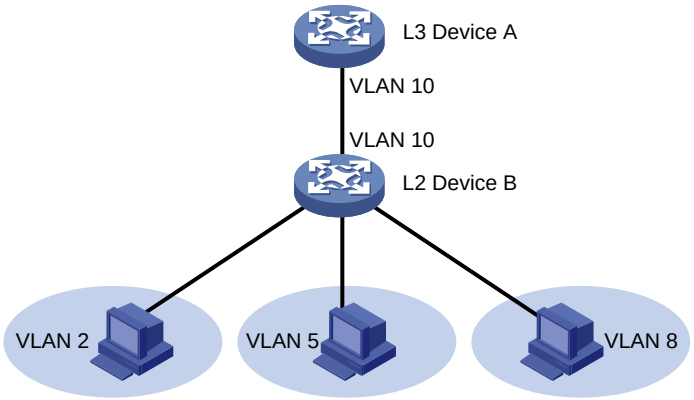
Configuring a downlink port

About this task

Configure a downlink port as follows:

- If a downlink port allows only one secondary VLAN (for example, the port connecting L2 Device B to a host in Figure 10), configure the port as a host port. The host port can be automatically assigned to the secondary VLAN and its associated primary VLAN.
- If a downlink port allows multiple secondary VLANs, configure the port as a trunk secondary port. The trunk secondary port can be automatically assigned to the secondary VLANs and their associated primary VLANs.

Figure 10 Private VLAN example



Procedure

1. Enter system view.
   system-view
2. Enter interface view of the downlink port.
   interface *interface-type interface-number*
3. Assign the downlink port to secondary VLANs.
   a. Set the link type of the port.
      port link-type { **access** | **hybrid** | **trunk** }
   b. Assign the access port to the specified VLAN.
      port access vlan *vlan-id*
   c. Assign the trunk port to the specified VLANs.
      port trunk permit vlan { *vlan-id-list* | **all** }
   d. Assign the hybrid port to the specified VLANs.
      port hybrid vlan *vlan-id-list* { **tagged** | **untagged** }

   Select substep b, c, or d depending on the port link type.
4. Configure the downlink port as a host or trunk secondary port.
   - Configure the downlink port as a host port.
     port private-vlan host
   - Configure the downlink port as a trunk secondary port of the specified VLANs.

port private-vlan *vlan-id-list* **trunk secondary**

By default, a port is not a host or trunk secondary port.

5. Return to system view.

   quit

6. Enter VLAN view of a secondary VLAN.

   vlan *vlan-id*

7. (Optional.) Enable Layer 2 communication for ports in the same secondary VLAN. Choose one command as needed:

   undo private-vlan isolated

   private-vlan community

   By default, ports in the same secondary VLAN can communicate with each other at Layer 2.

Configuring Layer 3 communication for secondary VLANs

Restrictions and guidelines

When the downlink port is operating in trunk secondary mode, it does not support Layer 3 communication between the specified secondary VLANs in the same primary VLAN.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN interface view of the primary VLAN interface.

   interface vlan-interface *interface-number*

3. Enable Layer 3 communication between secondary VLANs that are associated with the primary VLAN.

   private-vlan secondary *vlan-id-list*

   By default, secondary VLANs cannot communicate with each other at Layer 3.

4. Assign an IP address to the primary VLAN interface.

   IPv4:

   ip address *ip-address* { *mask-length* | *mask* } [**sub**]

   IPv6:

   ipv6 address { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }

   By default, no IP address is configured for a VLAN interface.

5. Enable local proxy ARP or ND.

   IPv4:

   local-proxy-arp enable

   By default, local proxy ARP is disabled.

   For more information about local proxy ARP, see *Layer 3—IP Services Configuration Guide*.

   IPv6:

   local-proxy-nd enable

   By default, local proxy ND is disabled.

   For more information about local proxy ND, see *Layer 3—IP Services Configuration Guide*.

Display and maintenance commands for the private VLAN

Execute **display** commands in any view.

Task	Command
Display information about primary VLANs and the secondary VLANs associated with each primary VLAN.	**display private-vlan** [*primary-vlan-id*]

Private VLAN configuration examples

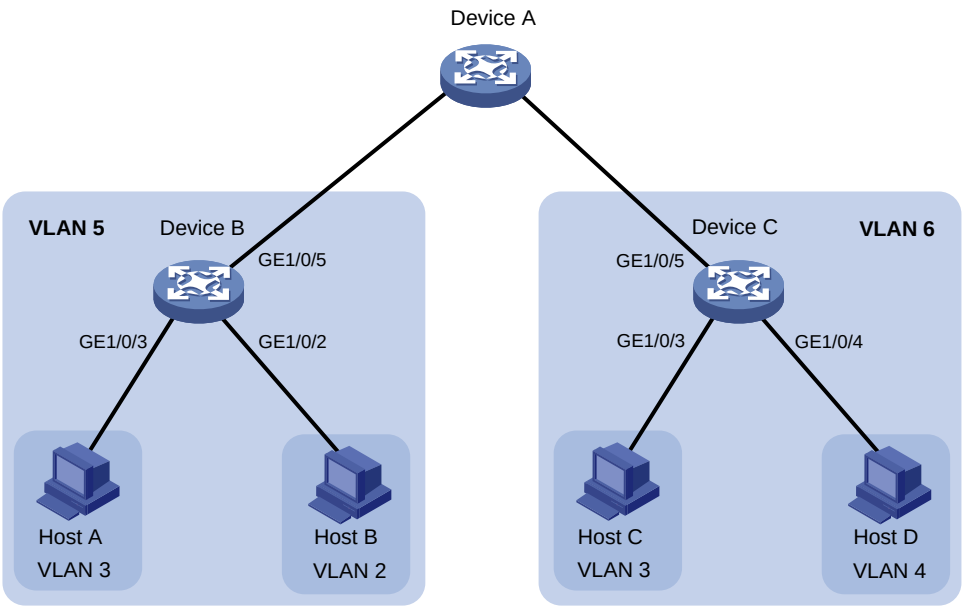
Example: Configuring promiscuous ports

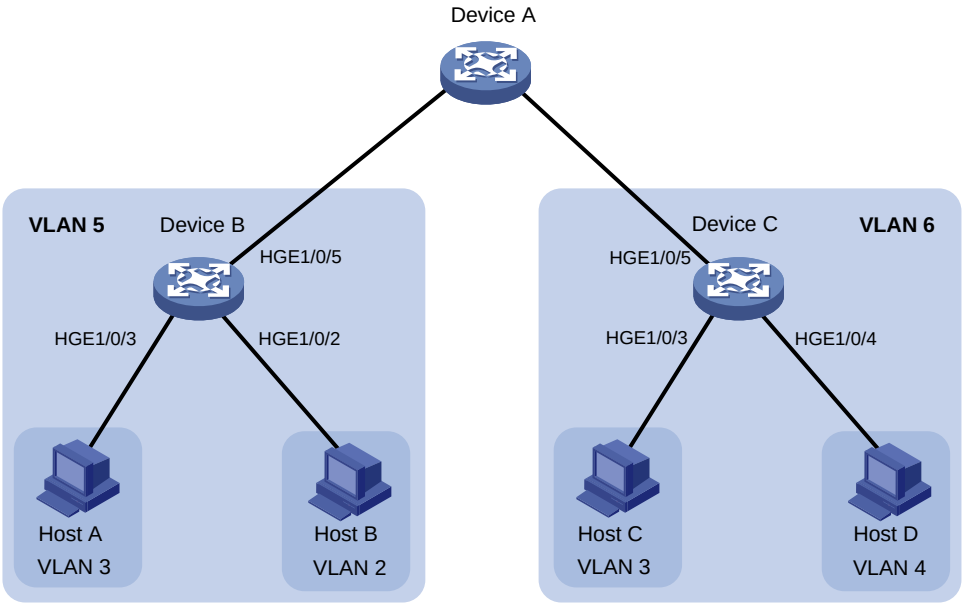
Network configuration

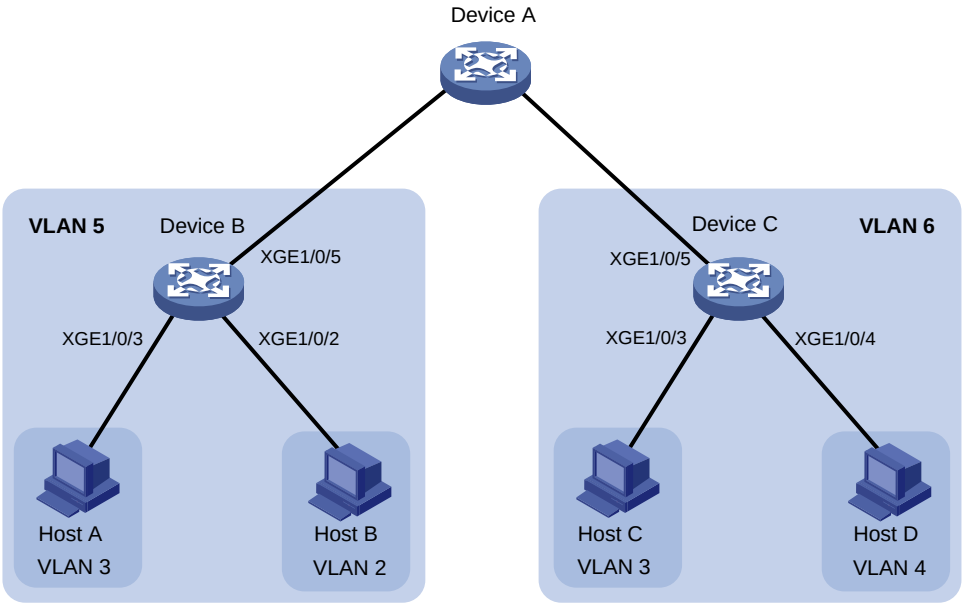
As shown in Figure 11, configure the private VLAN feature to meet the following requirements:

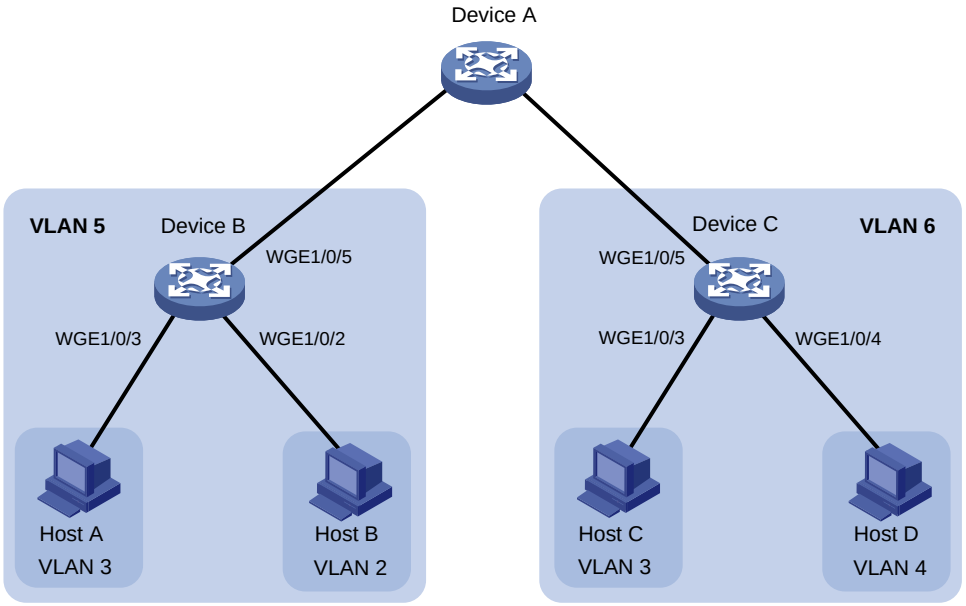
- On Device B, VLAN 5 is a primary VLAN that is associated with secondary VLANs 2 and 3. GigabitEthernet 1/0/5 is in VLAN 5. GigabitEthernet 1/0/2 is in VLAN 2. GigabitEthernet 1/0/3 is in VLAN 3.
- On Device C, VLAN 6 is a primary VLAN that is associated with secondary VLANs 3 and 4. GigabitEthernet 1/0/5 is in VLAN 6. GigabitEthernet 1/0/3 is in VLAN 3. GigabitEthernet 1/0/4 is in VLAN 4.
- Device A is aware of only VLAN 5 on Device B and VLAN 6 on Device C.

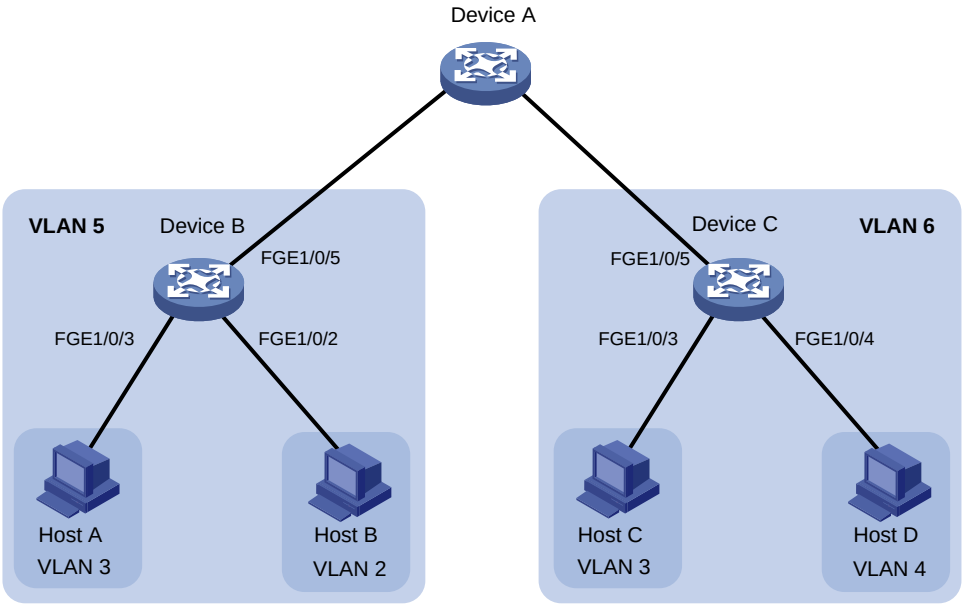
Figure 11 Network diagram

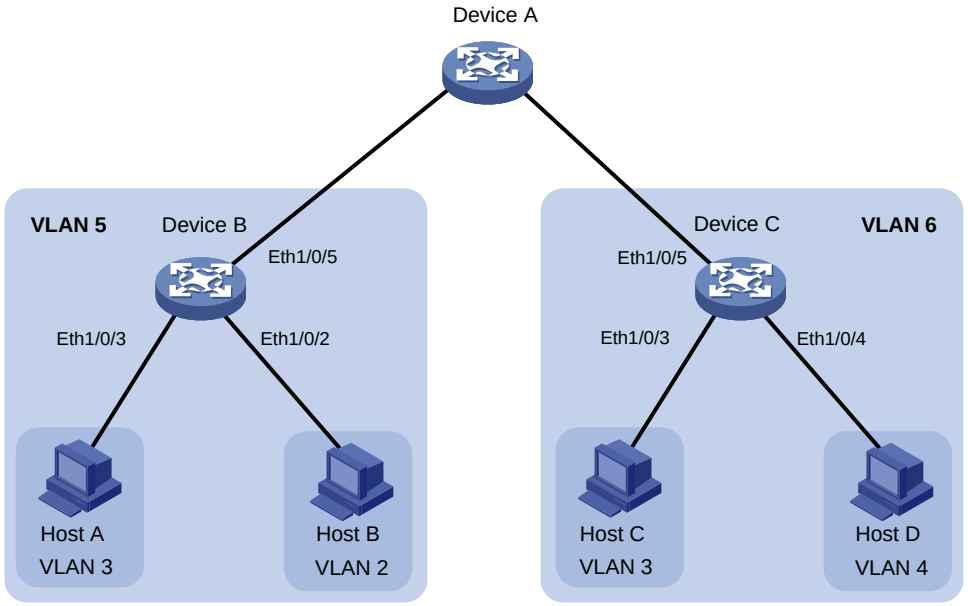


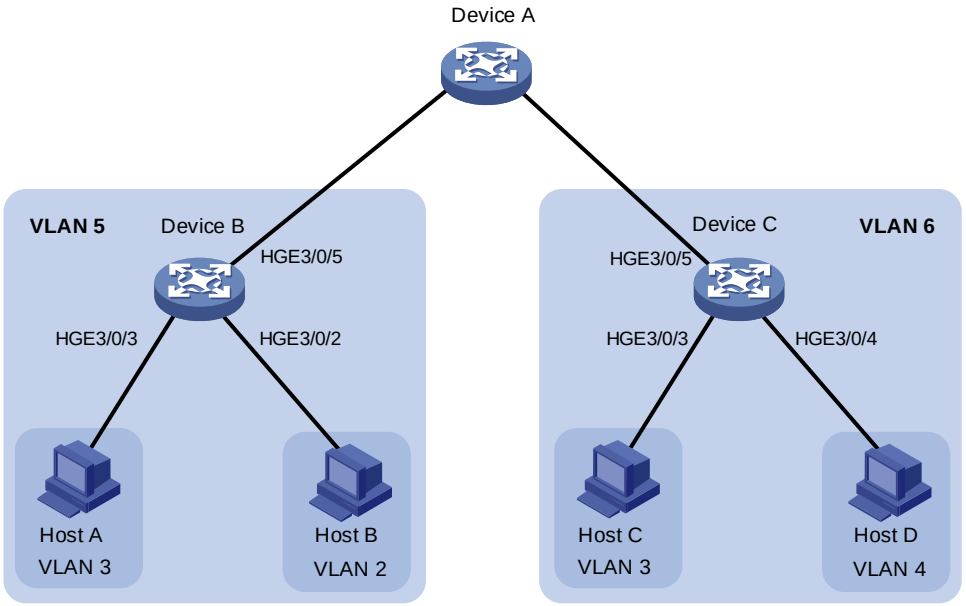

Device A
VLAN 5
Device B
HGE1/0/5
HGE1/0/3
HGE1/0/2
Device C
VLAN 6
HGE1/0/5
HGE1/0/3
HGE1/0/4
Host A
VLAN 3
Host B
VLAN 2
Host C
VLAN 3
Host D
VLAN 4

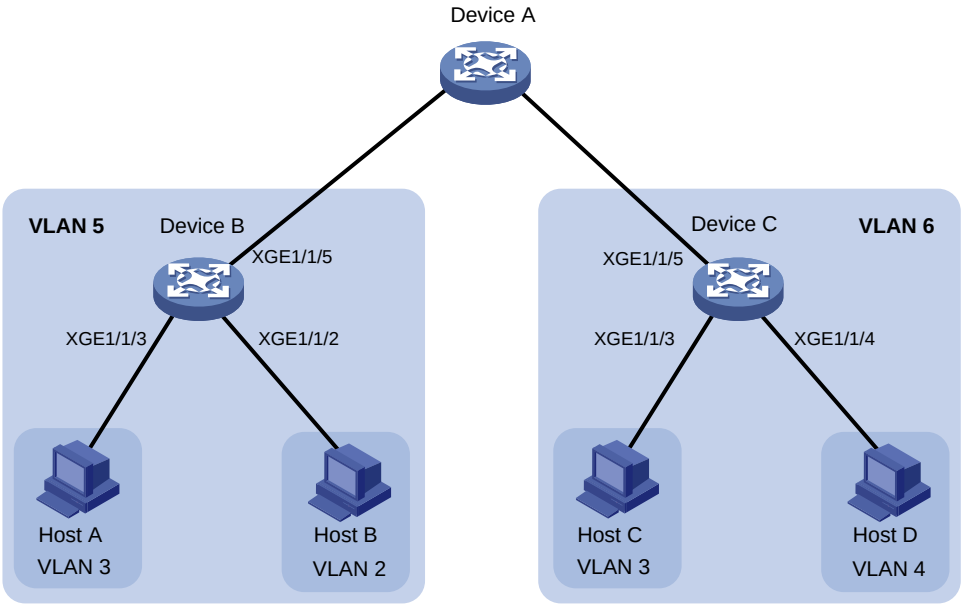

Device A
VLAN 5
Device B
XGE1/0/5
XGE1/0/3
XGE1/0/2
Device C
VLAN 6
XGE1/0/5
XGE1/0/3
XGE1/0/4
Host A
VLAN 3
Host B
VLAN 2
Host C
VLAN 3
Host D
VLAN 4

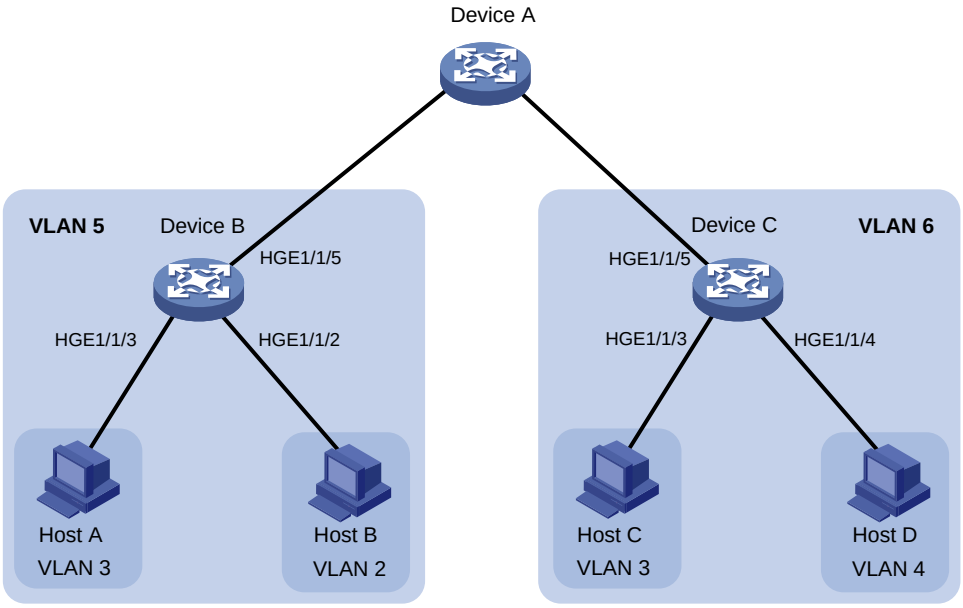

Device A
VLAN 5
Device B
WGE1/0/5
WGE1/0/3
WGE1/0/2
Device C
VLAN 6
WGE1/0/5
WGE1/0/3
WGE1/0/4
Host A
VLAN 3
Host B
VLAN 2
Host C
VLAN 3
Host D
VLAN 4


Device A
VLAN 5
Device B
FGE1/0/5
FGE1/0/3
FGE1/0/2
Device C
VLAN 6
FGE1/0/5
FGE1/0/3
FGE1/0/4
Host A
VLAN 3
Host B
VLAN 2
Host C
VLAN 3
Host D
VLAN 4


Device A
VLAN 5
Device B
Eth1/0/5
Eth1/0/3
Eth1/0/2
Device C
VLAN 6
Eth1/0/5
Eth1/0/3
Eth1/0/4
Host A
VLAN 3
Host B
VLAN 2
Host C
VLAN 3
Host D
VLAN 4


Device A
VLAN 5
Device B
HGE3/0/5
HGE3/0/3
HGE3/0/2
Device C
VLAN 6
HGE3/0/5
HGE3/0/3
HGE3/0/4
Host A
VLAN 3
Host B
VLAN 2
Host C
VLAN 3
Host D
VLAN 4





Procedure

This example describes the configurations on Device B and Device C.

1. Configure Device B:

Configure VLAN 5 as a primary VLAN.

```
<DeviceB> system-view
[DeviceB] vlan 5
[DeviceB-vlan5] private-vlan primary
[DeviceB-vlan5] quit
```

Create VLANs 2 and 3.

```
[DeviceB] vlan 2 to 3
```

Associate secondary VLANs 2 and 3 with primary VLAN 5.

```
[DeviceB] vlan 5
```

```
[DeviceB-vlan5] private-vlan secondary 2 to 3
[DeviceB-vlan5] quit
```

Configure the uplink port (GigabitEthernet 1/0/5) as a promiscuous port of VLAN 5.

```
[DeviceB] interface gigabitethernet 1/0/5
[DeviceB-GigabitEthernet1/0/5] port private-vlan 5 promiscuous
[DeviceB-GigabitEthernet1/0/5] quit
```

Assign downlink port GigabitEthernet 1/0/2 to VLAN 2, and configure the port as a host port.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port access vlan 2
[DeviceB-GigabitEthernet1/0/2] port private-vlan host
[DeviceB-GigabitEthernet1/0/2] quit
```

Assign downlink port GigabitEthernet 1/0/3 to VLAN 3, and configure the port as a host port.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port access vlan 3
[DeviceB-GigabitEthernet1/0/3] port private-vlan host
[DeviceB-GigabitEthernet1/0/3] quit
```

2. Configure Device C:

Configure VLAN 6 as a primary VLAN.

```
<DeviceC> system-view
[DeviceC] vlan 6
[DeviceC–vlan6] private-vlan primary
[DeviceC–vlan6] quit
```

Create VLANs 3 and 4.

```
[DeviceC] vlan 3 to 4
```

Associate secondary VLANs 3 and 4 with primary VLAN 6.

```
[DeviceC] vlan 6
[DeviceC-vlan6] private-vlan secondary 3 to 4
[DeviceC-vlan6] quit
```

Configure the uplink port (GigabitEthernet 1/0/5) as a promiscuous port of VLAN 6.

```
[DeviceC] interface gigabitethernet 1/0/5
[DeviceC-GigabitEthernet1/0/5] port private-vlan 6 promiscuous
[DeviceC-GigabitEthernet1/0/5] quit
```

Assign downlink port GigabitEthernet 1/0/3 to VLAN 3, and configure the port as a host port.

```
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] port access vlan 3
[DeviceC-GigabitEthernet1/0/3] port private-vlan host
[DeviceC-GigabitEthernet1/0/3] quit
```

Assign downlink port GigabitEthernet 1/0/4 to VLAN 4, and configure the port as a host port.

```
[DeviceC] interface gigabitethernet 1/0/4
[DeviceC-GigabitEthernet1/0/4] port access vlan 4
[DeviceC-GigabitEthernet1/0/4] port private-vlan host
[DeviceC-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Verify the private VLAN configurations on the devices, for example, on Device B.

```
[DeviceB] display private-vlan
 Primary VLAN ID: 5
 Secondary VLAN ID: 2-3
```

```
VLAN ID: 5
VLAN type: Static
Private VLAN type: Primary
Route interface: Not configured
Description: VLAN 0005
Name: VLAN 0005
Tagged ports:   None
Untagged ports:
    GigabitEthernet1/0/2
    GigabitEthernet1/0/3
    GigabitEthernet1/0/5

VLAN ID: 2
VLAN type: Static
Private VLAN type: Secondary
Route interface: Not configured
Description: VLAN 0002
Name: VLAN 0002
Tagged ports:   None
Untagged ports:
    GigabitEthernet1/0/2
    GigabitEthernet1/0/5

VLAN ID: 3
VLAN type: Static
Private VLAN type: Secondary
Route interface: Not configured
Description: VLAN 0003
Name: VLAN 0003
Tagged Ports:   None
Untagged Ports:
    GigabitEthernet1/0/3
    GigabitEthernet1/0/5
```

The output shows that:

- The promiscuous port (GigabitEthernet 1/0/5) is an untagged member of primary VLAN 5 and secondary VLANs 2 and 3.
- Host port GigabitEthernet 1/0/2 is an untagged member of primary VLAN 5 and secondary VLAN 2.
- Host port GigabitEthernet 1/0/3 is an untagged member of primary VLAN 5 and secondary VLAN 3.

Example: Configuring trunk promiscuous ports

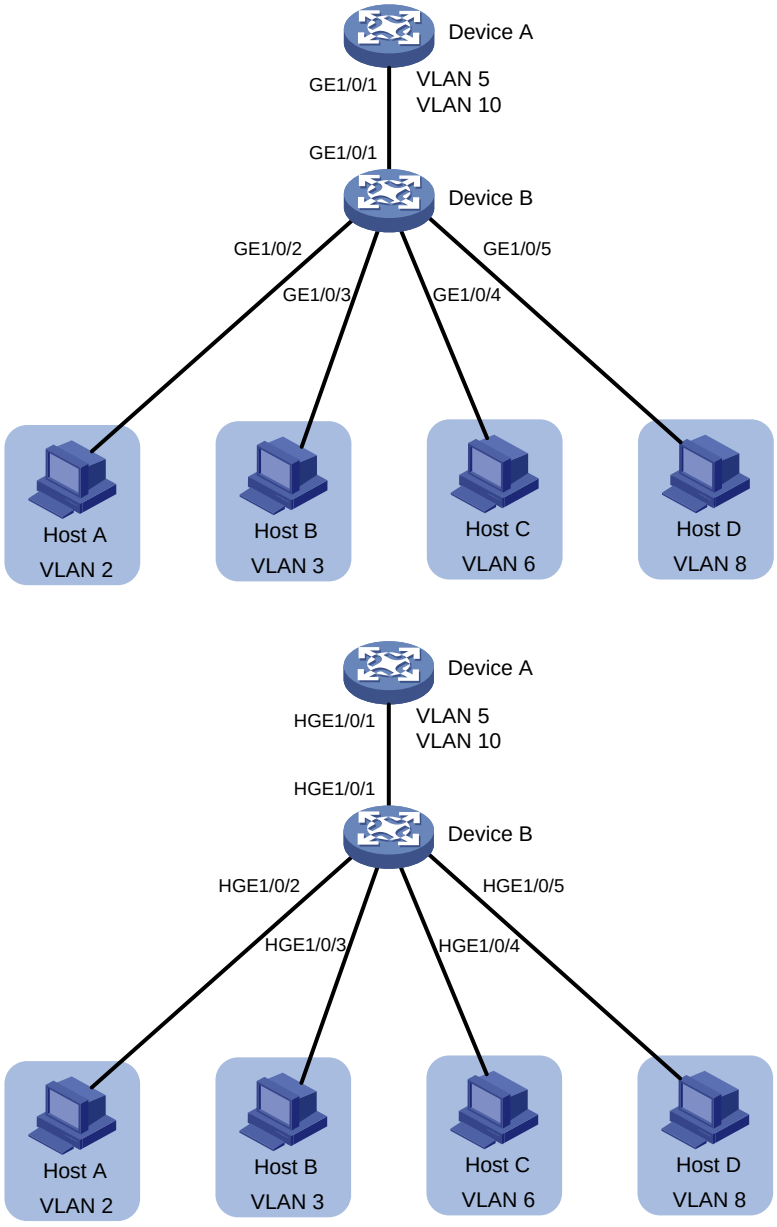
Network configuration

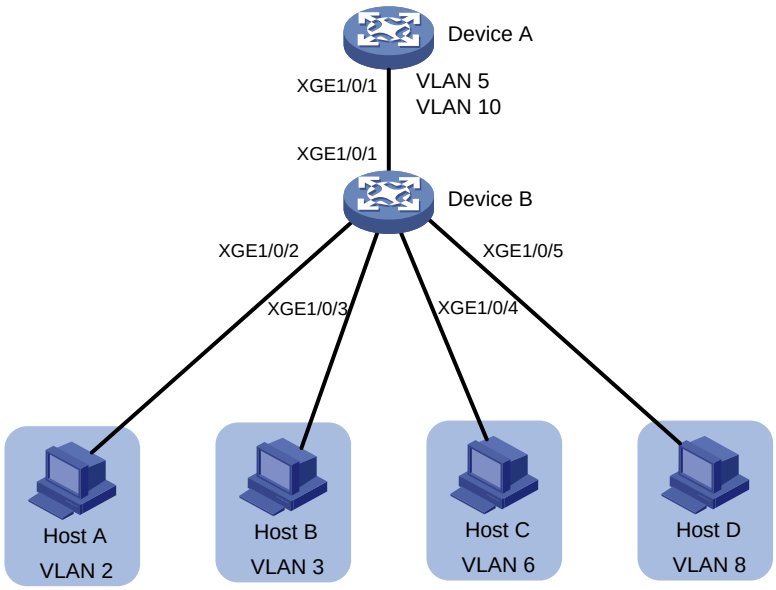
As shown in Figure 12, configure the private VLAN feature to meet the following requirements:

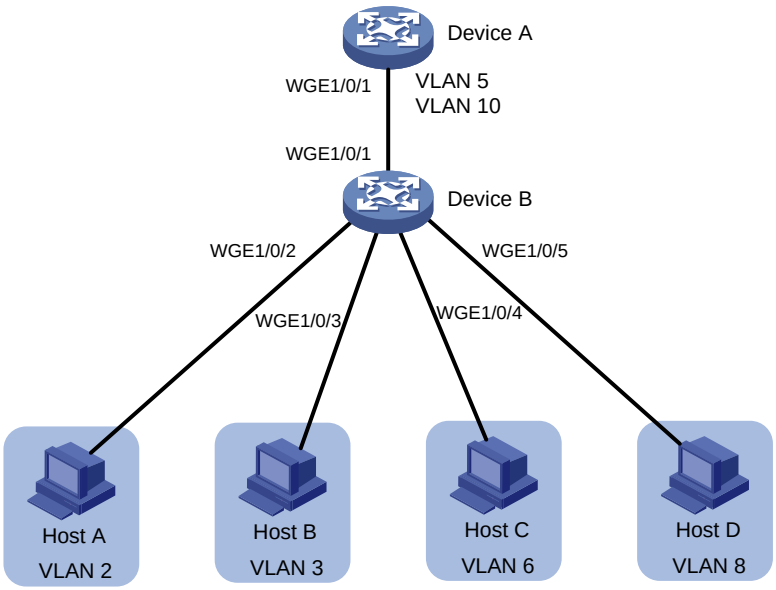
- VLANs 5 and 10 are primary VLANs on Device B. The uplink port (GigabitEthernet 1/0/1) on Device B permits the packets from VLANs 5 and 10 to pass through tagged.

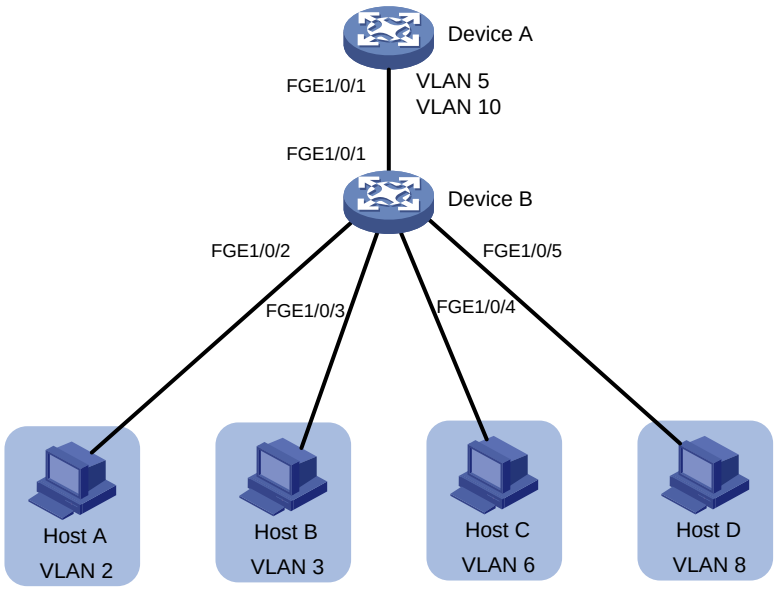
- On Device B, downlink port GigabitEthernet 1/0/2 permits secondary VLAN 2. Downlink port GigabitEthernet 1/0/3 permits secondary VLAN 3. Secondary VLANs 2 and 3 are associated with primary VLAN 5.
- On Device B, downlink port GigabitEthernet 1/0/4 permits secondary VLAN 6. Downlink port GigabitEthernet 1/0/5 permits secondary VLAN 8. Secondary VLANs 6 and 8 are associated with primary VLAN 10.
- Device A is aware of only VLANs 5 and 10 on Device B.

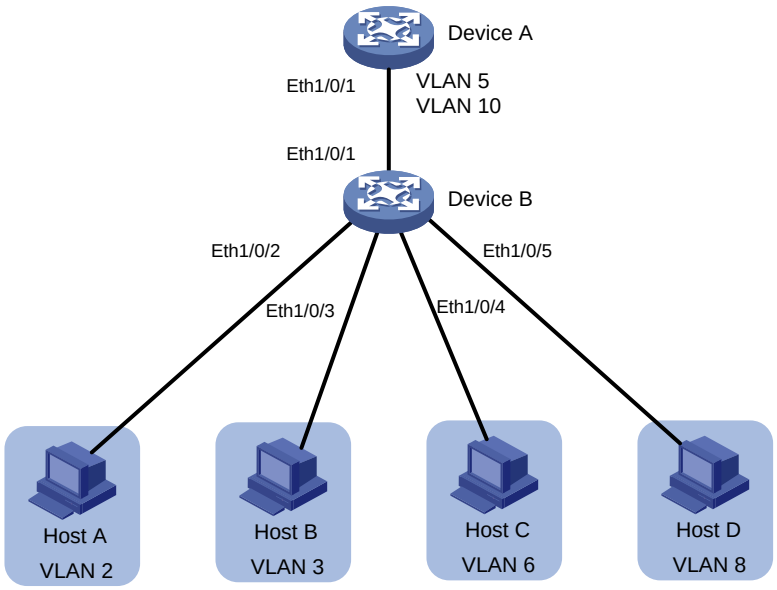
Figure 12 Network diagram

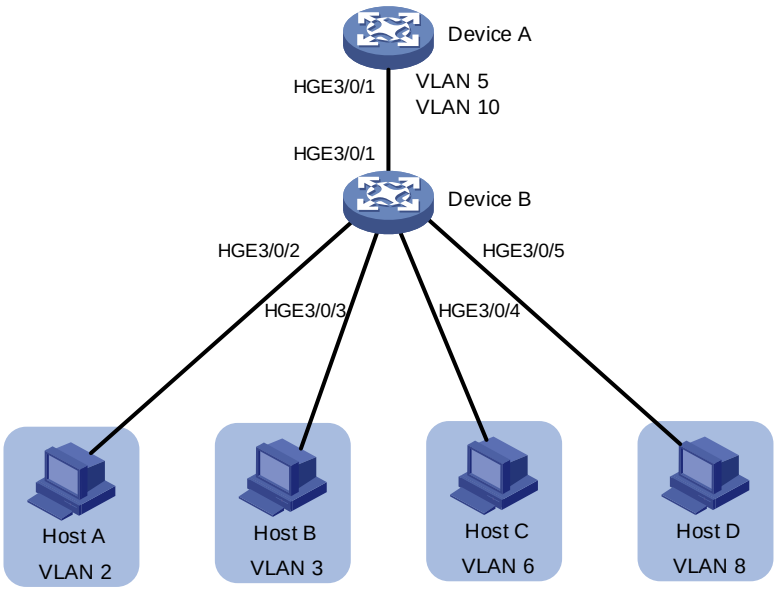


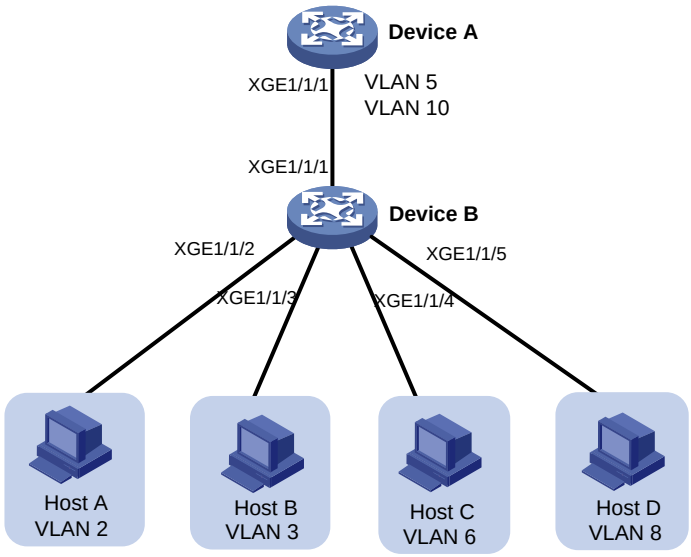

Device A
XGE1/0/1
VLAN 5
VLAN 10
XGE1/0/1
Device B
XGE1/0/2
XGE1/0/3
XGE1/0/4
XGE1/0/5
Host A
VLAN 2
Host B
VLAN 3
Host C
VLAN 6
Host D
VLAN 8

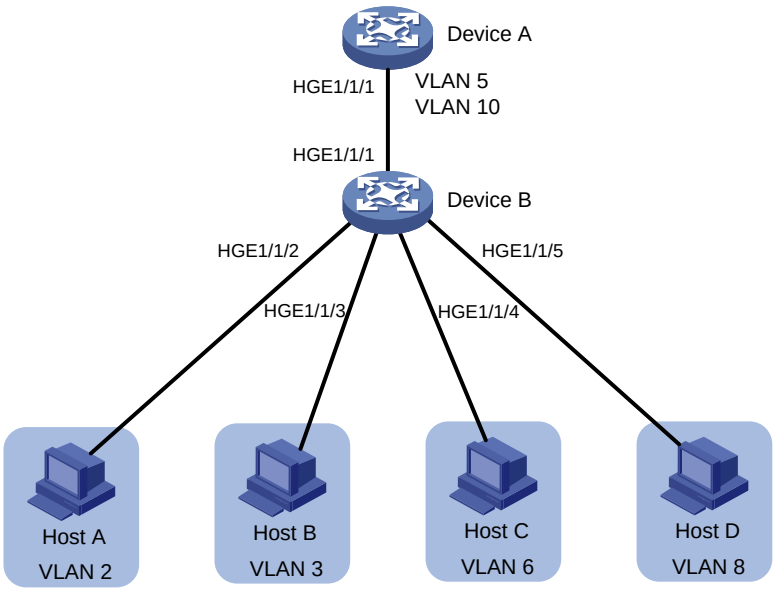

Device A
WGE1/0/1
VLAN 5
VLAN 10
WGE1/0/1
Device B
WGE1/0/2
WGE1/0/3
WGE1/0/4
WGE1/0/5
Host A
VLAN 2
Host B
VLAN 3
Host C
VLAN 6
Host D
VLAN 8


Device A
FGE1/0/1
VLAN 5
VLAN 10
FGE1/0/1
Device B
FGE1/0/2
FGE1/0/3
FGE1/0/4
FGE1/0/5
Host A
VLAN 2
Host B
VLAN 3
Host C
VLAN 6
Host D
VLAN 8


Device A
Eth1/0/1
VLAN 5
VLAN 10
Eth1/0/1
Device B
Eth1/0/2
Eth1/0/3
Eth1/0/4
Eth1/0/5
Host A
VLAN 2
Host B
VLAN 3
Host C
VLAN 6
Host D
VLAN 8


Device A
HGE3/0/1
VLAN 5
VLAN 10
HGE3/0/1
Device B
HGE3/0/2
HGE3/0/3
HGE3/0/4
HGE3/0/5
Host A
VLAN 2
Host B
VLAN 3
Host C
VLAN 6
Host D
VLAN 8


Device A
XGE1/1/1
VLAN 5
VLAN 10
XGE1/1/1
Device B
XGE1/1/2
XGE1/1/3
XGE1/1/4
XGE1/1/5
Host A
VLAN 2
Host B
VLAN 3
Host C
VLAN 6
Host D
VLAN 8



Procedure

1. Configure Device B:

   # Configure VLANs 5 and 10 as primary VLANs.

```
<DeviceB> system-view
[DeviceB] vlan 5
[DeviceB-vlan5] private-vlan primary
[DeviceB-vlan5] quit
[DeviceB] vlan 10
[DeviceB-vlan10] private-vlan primary
[DeviceB-vlan10] quit
```

   # Create VLANs 2, 3, 6, and 8.

```
[DeviceB] vlan 2 to 3
[DeviceB] vlan 6
[DeviceB-vlan6] quit
[DeviceB] vlan 8
[DeviceB-vlan8] quit
```

   # Associate secondary VLANs 2 and 3 with primary VLAN 5.

```
[DeviceB] vlan 5
[DeviceB-vlan5] private-vlan secondary 2 to 3
[DeviceB-vlan5] quit
```

   # Associate secondary VLANs 6 and 8 with primary VLAN 10.

```
[DeviceB] vlan 10
[DeviceB-vlan10] private-vlan secondary 6 8
[DeviceB-vlan10] quit
```

   # Configure the uplink port (GigabitEthernet 1/0/1) as a trunk promiscuous port of VLANs 5 and 10.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port private-vlan 5 10 trunk promiscuous
[DeviceB-GigabitEthernet1/0/1] quit
```

   # Assign downlink port GigabitEthernet 1/0/2 to VLAN 2, and configure the port as a host port.

```
[DeviceB] interface gigabitethernet 1/0/2
```

```
[DeviceB-GigabitEthernet1/0/2] port access vlan 2
[DeviceB-GigabitEthernet1/0/2] port private-vlan host
[DeviceB-GigabitEthernet1/0/2] quit
```

Assign downlink port GigabitEthernet 1/0/3 to VLAN 3, and configure the port as a host port.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port access vlan 3
[DeviceB-GigabitEthernet1/0/3] port private-vlan host
[DeviceB-GigabitEthernet1/0/3] quit
```

Assign downlink port GigabitEthernet 1/0/4 to VLAN 6, and configure the port as a host port.

```
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port access vlan 6
[DeviceB-GigabitEthernet1/0/4] port private-vlan host
[DeviceB-GigabitEthernet1/0/4] quit
```

Assign downlink port GigabitEthernet 1/0/5 to VLAN 8, and configure the port as a host port.

```
[DeviceB] interface gigabitethernet 1/0/5
[DeviceB-GigabitEthernet1/0/5] port access vlan 8
[DeviceB-GigabitEthernet1/0/5] port private-vlan host
[DeviceB-GigabitEthernet1/0/5] quit
```

2. Configure Device A:

Create VLANs 5 and 10.

```
[DeviceA] vlan 5
[DeviceA-vlan5] quit
[DeviceA] vlan 10
[DeviceA-vlan10] quit
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign it to VLANs 5 and 10 as a tagged VLAN member.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type hybrid
[DeviceA-GigabitEthernet1/0/1] port hybrid vlan 5 10 tagged
[DeviceA-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Verify the primary VLAN configurations on Device B. The following output uses primary VLAN 5 as an example.

```
[DeviceB] display private-vlan 5
 Primary VLAN ID: 5
 Secondary VLAN ID: 2-3

 VLAN ID: 5
 VLAN type: Static
 Private VLAN type: Primary
 Route interface: Not configured
 Description: VLAN 0005
 Name: VLAN 0005
 Tagged ports:
     GigabitEthernet1/0/1
 Untagged ports:
     GigabitEthernet1/0/2
     GigabitEthernet1/0/3
```

```
VLAN ID: 2
VLAN type: Static
Private VLAN type: Secondary
Route interface: Not configured
Description: VLAN 0002
Name: VLAN 0002
 Tagged ports:
    GigabitEthernet1/0/1
 Untagged ports:
    GigabitEthernet1/0/2

VLAN ID: 3
VLAN type: Static
Private VLAN type: Secondary
Route interface: Not configured
Description: VLAN 0003
Name: VLAN 0003
 Tagged ports:
    GigabitEthernet1/0/1
 Untagged ports:
    GigabitEthernet1/0/3
```

The output shows that:

- The trunk promiscuous port (GigabitEthernet 1/0/1) is a tagged member of primary VLAN 5 and secondary VLANs 2 and 3.
- Host port GigabitEthernet 1/0/2 is an untagged member of primary VLAN 5 and secondary VLAN 2.
- Host port GigabitEthernet 1/0/3 is an untagged member of primary VLAN 5 and secondary VLAN 3.

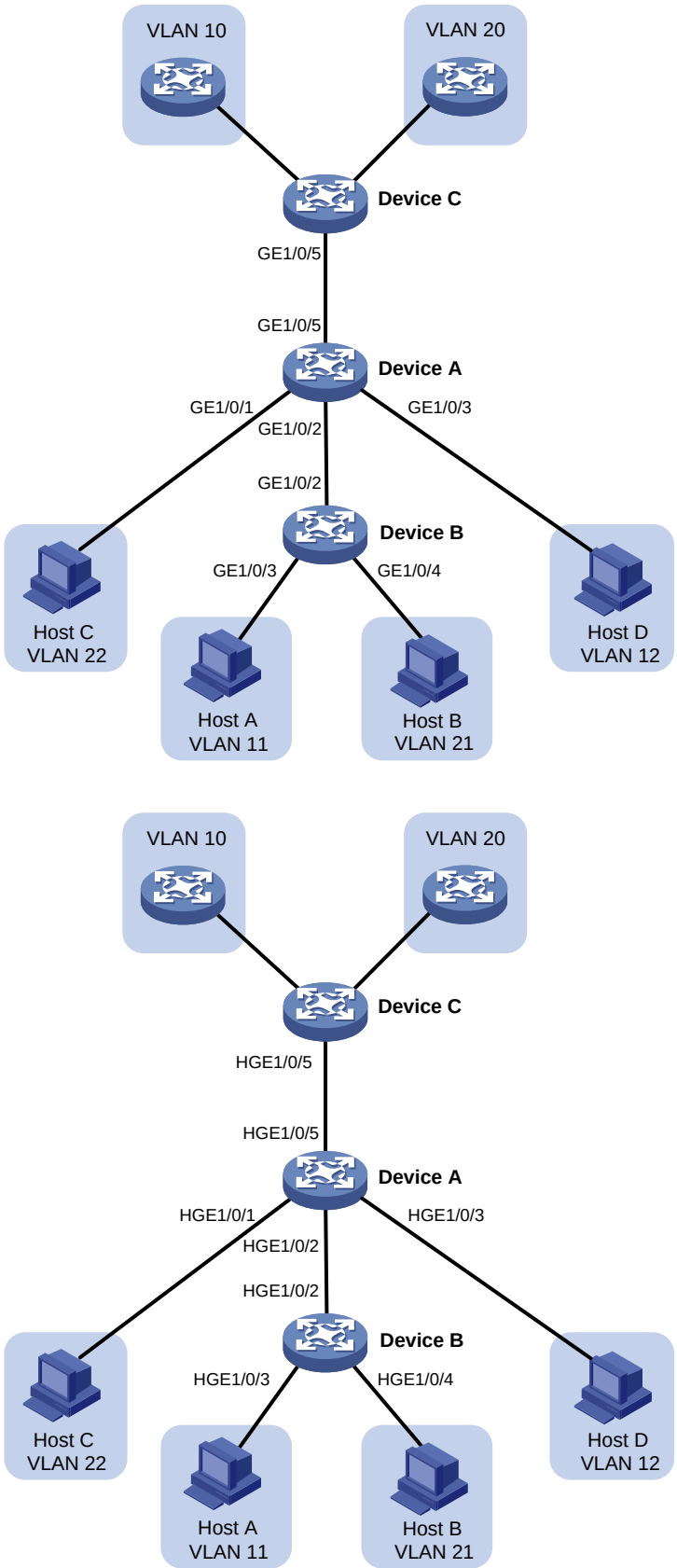
Example: Configuring trunk promiscuous and trunk secondary ports

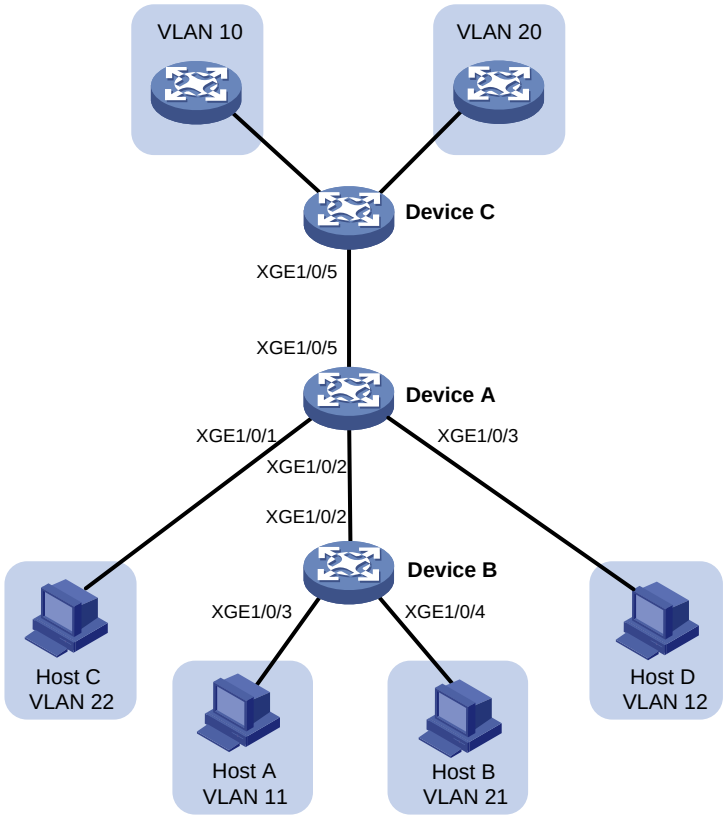
Network configuration

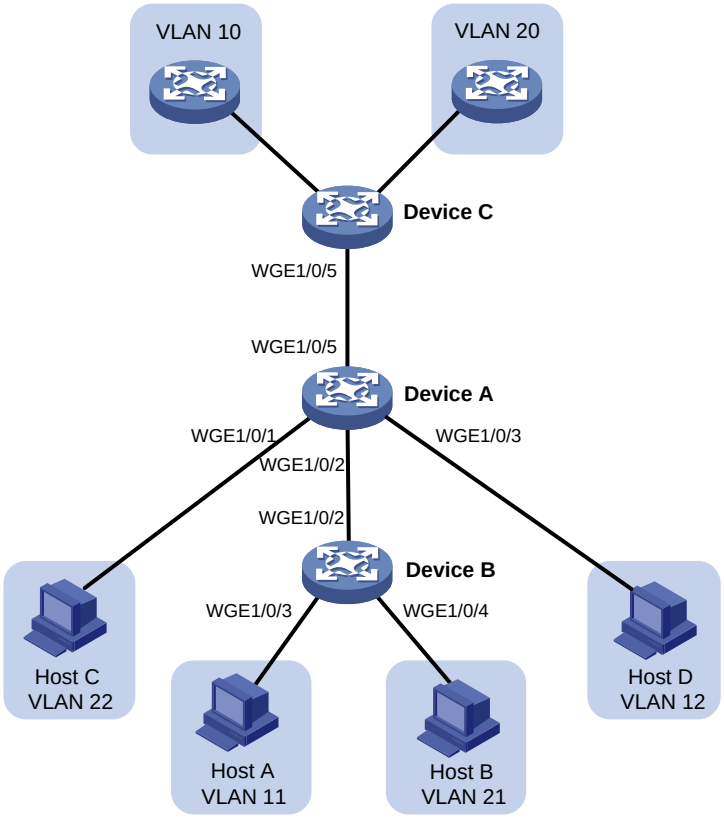
As shown in Figure 13, configure the private VLAN feature to meet the following requirements:

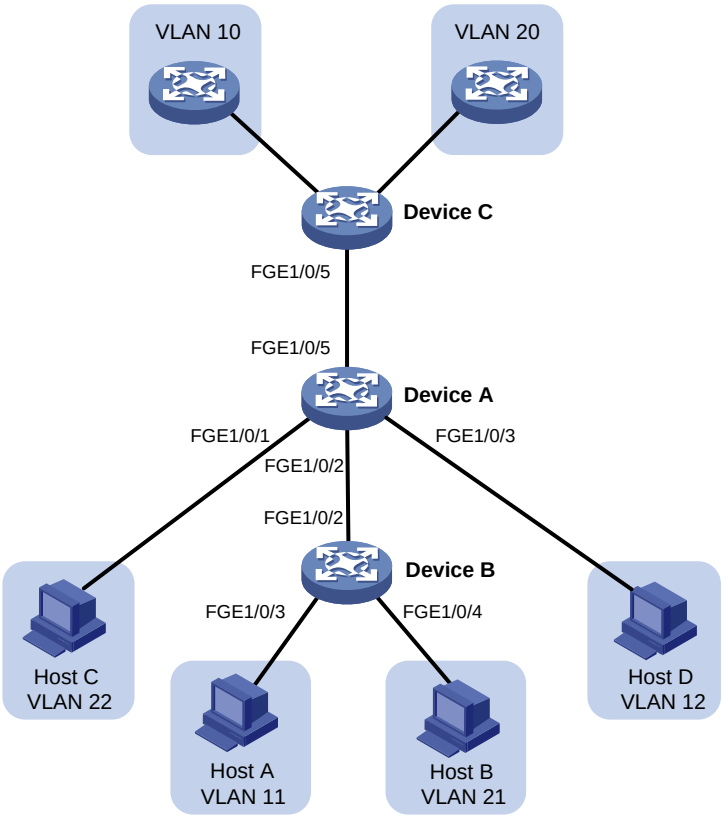
- VLANs 10 and 20 are primary VLANs on Device A. The uplink port (GigabitEthernet 1/0/5) on Device A permits the packets from VLANs 10 and 20 to pass through tagged.
- VLANs 11, 12, 21, and 22 are secondary VLANs on Device A.
  - Downlink port GigabitEthernet 1/0/2 permits the packets from secondary VLANs 11 and 21 to pass through tagged.
  - Downlink port GigabitEthernet 1/0/1 permits secondary VLAN 22.
  - Downlink port GigabitEthernet 1/0/3 permits secondary VLAN 12.
- Secondary VLANs 11 and 12 are associated with primary VLAN 10.
- Secondary VLANs 21 and 22 are associated with primary VLAN 20.

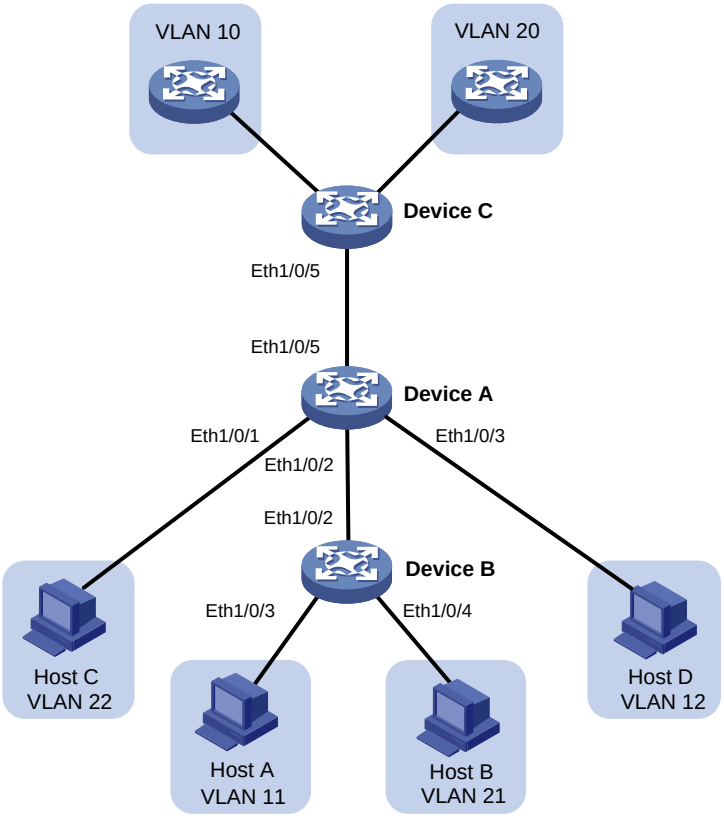
Figure 13 Network diagram

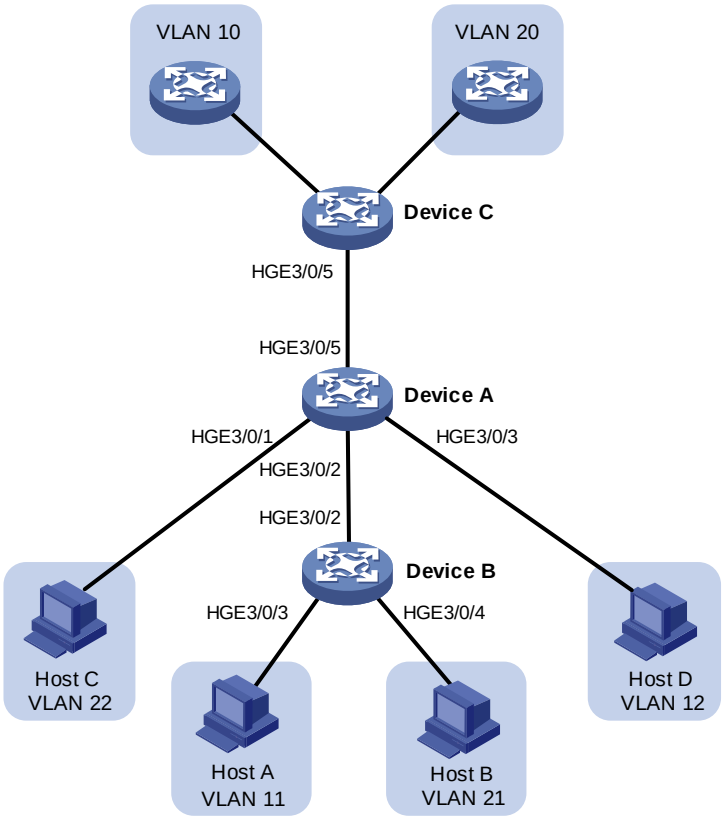


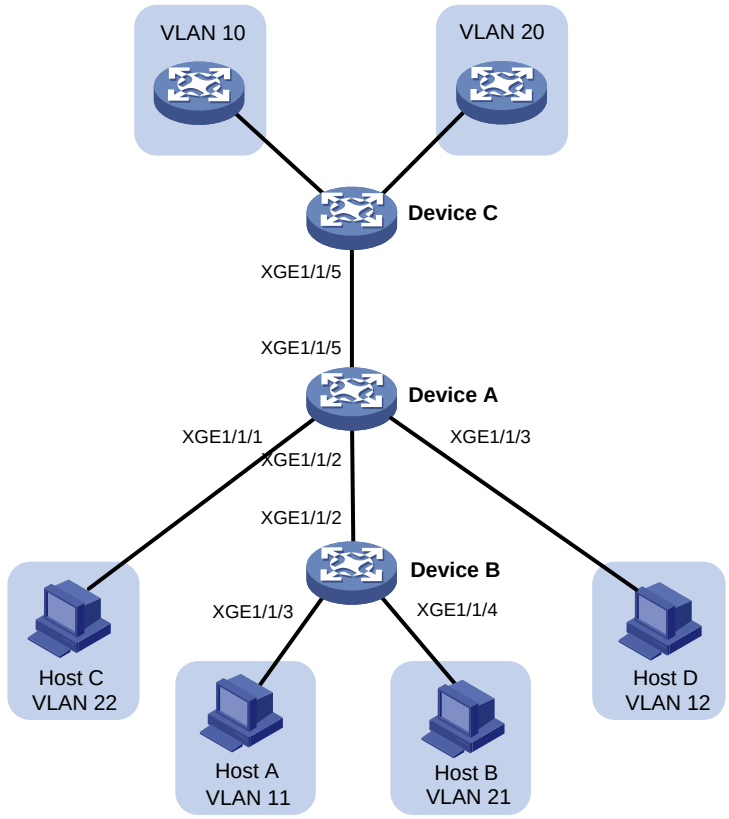

VLAN 10
VLAN 20
Device C
XGE1/0/5
XGE1/0/5
Device A
XGE1/0/1
XGE1/0/2
XGE1/0/3
XGE1/0/2
Device B
XGE1/0/3
XGE1/0/4
Host C
VLAN 22
Host A
VLAN 11
Host B
VLAN 21
Host D
VLAN 12

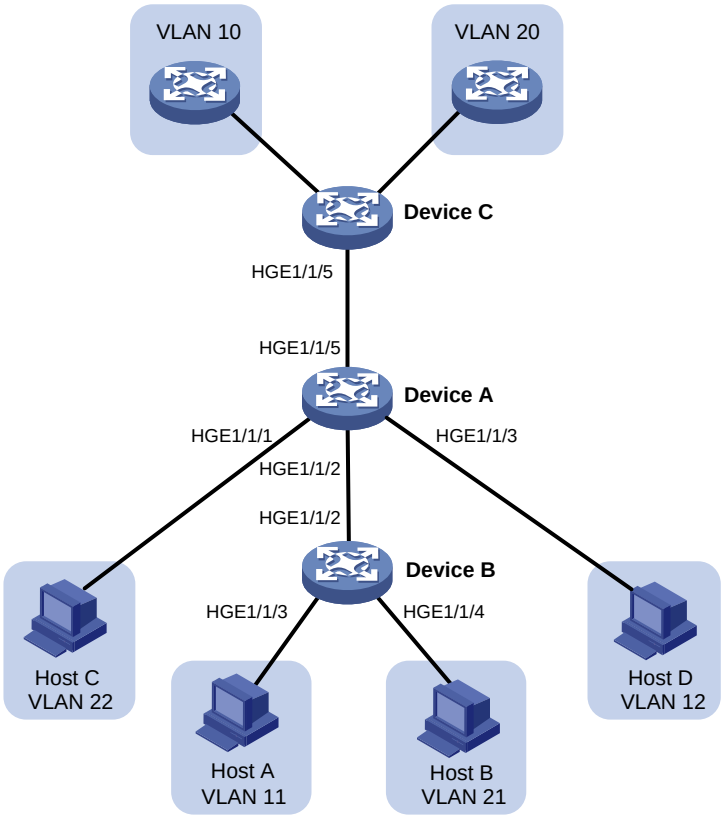

VLAN 10
VLAN 20
Device C
WGE1/0/5
WGE1/0/5
Device A
WGE1/0/1
WGE1/0/2
WGE1/0/3
WGE1/0/2
Device B
WGE1/0/3
WGE1/0/4
Host C
VLAN 22
Host A
VLAN 11
Host B
VLAN 21
Host D
VLAN 12


VLAN 10
VLAN 20
Device C
FGE1/0/5
FGE1/0/5
Device A
FGE1/0/1
FGE1/0/2
FGE1/0/3
FGE1/0/2
Device B
FGE1/0/3
FGE1/0/4
Host C
VLAN 22
Host A
VLAN 11
Host B
VLAN 21
Host D
VLAN 12


VLAN 10
VLAN 20
Device C
Eth1/0/5
Eth1/0/5
Device A
Eth1/0/1
Eth1/0/2
Eth1/0/3
Eth1/0/2
Device B
Eth1/0/3
Eth1/0/4
Host C
VLAN 22
Host A
VLAN 11
Host B
VLAN 21
Host D
VLAN 12


VLAN 10
VLAN 20
Device C
HGE3/0/5
HGE3/0/5
Device A
HGE3/0/1
HGE3/0/3
HGE3/0/2
HGE3/0/2
Device B
HGE3/0/3
HGE3/0/4
Host C
VLAN 22
Host D
VLAN 12
Host A
VLAN 11
Host B
VLAN 21


VLAN 10
VLAN 20
Device C
XGE1/1/5
XGE1/1/5
Device A
XGE1/1/1
XGE1/1/3
XGE1/1/2
XGE1/1/2
Device B
XGE1/1/3
XGE1/1/4
Host C
VLAN 22
Host D
VLAN 12
Host A
VLAN 11
Host B
VLAN 21



Procedure

1. Configure Device A:

   # Configure VLANs 10 and 20 as primary VLANs.

```
<DeviceA> system-view
[DeviceA] vlan 10
[DeviceA-vlan10] private-vlan primary
[DeviceA-vlan10] quit
[DeviceA] vlan 20
[DeviceA-vlan20] private-vlan primary
[DeviceA-vlan20] quit
```

   # Create VLANs 11, 12, 21, and 22.

```
[DeviceA] vlan 11 to 12
[DeviceA] vlan 21 to 22
```

   # Associate secondary VLANs 11 and 12 with primary VLAN 10.

```
[DeviceA] vlan 10
[DeviceA-vlan10] private-vlan secondary 11 12
[DeviceA-vlan10] quit
```

   # Associate secondary VLANs 21 and 22 with primary VLAN 20.

```
[DeviceA] vlan 20
[DeviceA-vlan20] private-vlan secondary 21 22
[DeviceA-vlan20] quit
```

   # Configure the uplink port (GigabitEthernet 1/0/5) as a trunk promiscuous port of VLANs 10 and 20.

```
[DeviceA] interface gigabitethernet 1/0/5
```

```
[DeviceA-GigabitEthernet1/0/5] port private-vlan 10 20 trunk promiscuous
[DeviceA-GigabitEthernet1/0/5] quit
```

Assign downlink port GigabitEthernet 1/0/1 to VLAN 22 and configure the port as a host port.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port access vlan 22
[DeviceA-GigabitEthernet1/0/1] port private-vlan host
[DeviceA-GigabitEthernet1/0/1] quit
```

Assign downlink port GigabitEthernet 1/0/3 to VLAN 12 and configure the port as a host port.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port access vlan 12
[DeviceA-GigabitEthernet1/0/3] port private-vlan host
[DeviceA-GigabitEthernet1/0/3] quit
```

Configure downlink port GigabitEthernet 1/0/2 as a trunk secondary port of VLANs 11 and 21.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port private-vlan 11 21 trunk secondary
[DeviceA-GigabitEthernet1/0/2] quit
```

2. Configure Device B:

Create VLANs 11 and 21.

```
<DeviceB> system-view
[DeviceB] vlan 11
[DeviceB-vlan11] quit
[DeviceB] vlan 21
[DeviceB-vlan21] quit
```

Configure GigabitEthernet 1/0/2 as a hybrid port, and assign it to VLANs 11 and 21 as a tagged VLAN member.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type hybrid
[DeviceB-GigabitEthernet1/0/2] port hybrid vlan 11 21 tagged
[DeviceB-GigabitEthernet1/0/2] quit
```

Assign GigabitEthernet 1/0/3 to VLAN 11.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port access vlan 11
[DeviceB-GigabitEthernet1/0/3] quit
```

Assign GigabitEthernet 1/0/4 to VLAN 21.

```
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port access vlan 21
[DeviceB-GigabitEthernet1/0/4] quit
```

3. Configure Device C:

Create VLANs 10 and 20.

```
<DeviceC> system-view
[DeviceC] vlan 10
[DeviceC-vlan10] quit
[DeviceC] vlan 20
[DeviceC-vlan20] quit
```

Configure GigabitEthernet 1/0/5 as a hybrid port, and assign it to VLANs 10 and 20 as a tagged VLAN member.

```
[DeviceC] interface gigabitethernet 1/0/5
[DeviceC-GigabitEthernet1/0/5] port link-type hybrid
```

```
[DeviceC-GigabitEthernet1/0/5] port hybrid vlan 10 20 tagged
[DeviceC-GigabitEthernet1/0/5] quit
```

Verifying the configuration

Verify the primary VLAN configurations on Device A. The following output uses primary VLAN 10 as an example.

```
[DeviceA] display private-vlan 10
 Primary VLAN ID: 10
 Secondary VLAN ID: 11-12

 VLAN ID: 10
 VLAN type: Static
 Private-vlan type: Primary
 Route interface: Not configured
 Description: VLAN 0010
 Name: VLAN 0010
 Tagged ports:
    GigabitEthernet1/0/2
    GigabitEthernet1/0/5
 Untagged ports:
    GigabitEthernet1/0/3

 VLAN ID: 11
 VLAN type: Static
 Private-vlan type: Secondary
 Route interface: Not configured
 Description: VLAN 0011
 Name: VLAN 0011
 Tagged ports:
    GigabitEthernet1/0/2
    GigabitEthernet1/0/5
 Untagged ports: None

 VLAN ID: 12
 VLAN type: Static
 Private-vlan type: Secondary
 Route interface: Not configured
 Description: VLAN 0012
 Name: VLAN 0012
 Tagged ports:
    GigabitEthernet1/0/5
 Untagged ports:
    GigabitEthernet1/0/3
```

The output shows that:

- The trunk promiscuous port (GigabitEthernet 1/0/5) is a tagged member of primary VLAN 10 and secondary VLANs 11 and 12.
- The trunk secondary port (GigabitEthernet 1/0/2) is a tagged member of primary VLAN 10 and secondary VLAN 11.

- The host port (GigabitEthernet 1/0/3) is an untagged member of primary VLAN 10 and secondary VLAN 12.

Configuring voice VLANs

About voice VLANs

A voice VLAN is used for transmitting voice traffic. The device can configure QoS parameters for voice packets to ensure higher transmission priority of the voice packets.

Common voice devices include IP phones and integrated access devices (IADs). This chapter uses IP phones as an example.

Working mechanism

When an IP phone accesses a device, the device performs the following operations:

1. Identifies the IP phone in the network and obtains the MAC address of the IP phone.
2. Advertises the voice VLAN information to the IP phone.

After receiving the voice VLAN information, the IP phone performs automatic configuration. Voice packets sent from the IP phone can then be transmitted within the voice VLAN.

Methods of identifying IP phones

Devices can use the OUI addresses or LLDP to identify IP phones.

Identifying IP phones through OUI addresses

A device identifies voice packets based on their source MAC addresses. A packet whose source MAC address complies with an Organizationally Unique Identifier (OUI) address of the device is regarded as a voice packet.

You can use system default OUI addresses (see Table 1) or configure OUI addresses for the device. You cannot delete the system default OUI addresses.

Table 1 Default OUI addresses

Number	OUI address	Vendor
1	0001-e300-0000	Siemens phone
2	0003-6b00-0000	Cisco phone
3	0004-0d00-0000	Avaya phone
4	000f-e200-0000	H3C Aolynk phone
5	0060-b900-0000	Philips/NEC phone
6	00d0-1e00-0000	Pingtel phone
7	00e0-7500-0000	Polycom phone
8	00e0-bb00-0000	3Com phone

Typically, an OUI address refers to the first 24 bits of a MAC address (in binary notation) and is a globally unique identifier that IEEE assigns to a vendor. However, OUI addresses in this chapter are addresses that the system uses to identify voice packets. They are the logical AND results of the *mac-address* and *oui-mask* arguments in the **voice-vlan mac-address** command.

Automatically identifying IP phones through LLDP

If IP phones support LLDP, configure LLDP for automatic IP phone discovery on the device. The device can then automatically discover the peer through LLDP, and exchange LLDP TLVs with the peer.

If the LLDP System Capabilities TLV received on a port indicates that the peer can act as a telephone, the device performs the following operations:

1. Sends an LLDP TLV with the voice VLAN configuration to the peer.
2. Assigns the receiving port to the voice VLAN.
3. Increases the transmission priority of the voice packets sent from the IP phone.
4. Adds the MAC address of the IP phone to the MAC address table to ensure that the IP phone can pass authentication.

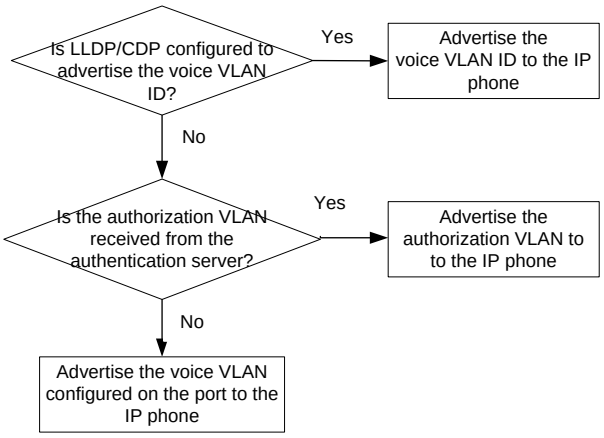
Use LLDP instead of the OUI list to identify IP phones if the network has more IP phone categories than the maximum number of OUI addresses supported on the device. LLDP has higher priority than the OUI list.

For more information about LLDP, see "Configuring LLDP."

Advertising the voice VLAN information to IP phones

Figure 14 shows the workflow of advertising the voice VLAN information to IP phones.

Figure 14 Workflow of advertising the voice VLAN information to IP phones



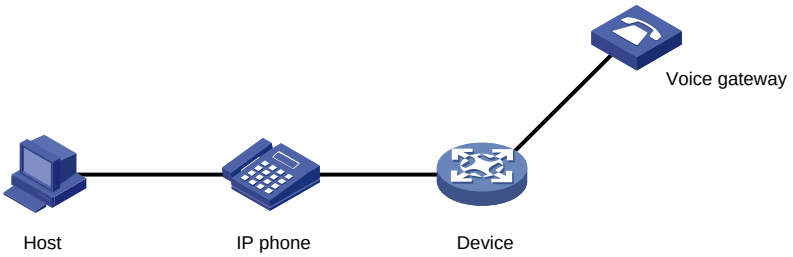
IP phone access methods

Connecting the host and the IP phone in series

As shown in Figure 15, the host is connected to the IP phone, and the IP phone is connected to the device. In this scenario, the following requirements must be met:

- The host and the IP phone use different VLANs.
- The IP phone is able to send out VLAN-tagged packets, so that the device can differentiate traffic from the host and the IP phone.
- The port connecting to the IP phone forwards packets from the voice VLAN and the PVID.

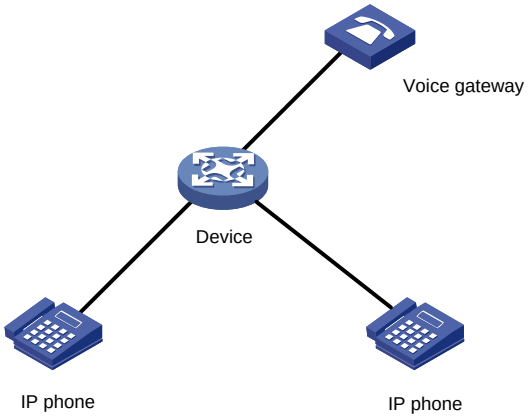
Figure 15 Connecting the host and IP phone in series



Connecting the IP phone to the device

As shown in Figure 16, IP phones are connected to the device without the presence of the host. Use this connection method when IP phones sends out untagged voice packets. In this scenario, you must configure the voice VLAN as the PVID of the access port of the IP phone, and configure the port to forward the packets from the PVID.

Figure 16 Connecting the IP phone to the device



Voice VLAN assignment modes

A port can be assigned to a voice VLAN automatically or manually.

Automatic mode

Use automatic mode when PCs and IP phones are connected in series to access the network through the device, as shown in the diagram for connecting the host and IP phone in series in "Connecting the host and the IP phone in series." Ports on the device transmit both voice traffic and data traffic.

When an IP phone is powered on, it sends out protocol packets. After receiving these protocol packets, the device uses the source MAC address of the protocol packets to match its OUI addresses. If the match succeeds, the device performs the following operations:

- Assigns the receiving port of the protocol packets to the voice VLAN.
- Issues ACL rules to set the packet precedence.
- Starts the voice VLAN aging timer.

If no voice packet is received from the port before the aging timer expires, the device will remove the port from the voice VLAN. The aging timer is also configurable.

When the IP phone reboots, the port is reassigned to the voice VLAN to ensure the correct operation of the existing voice connections. The reassignment occurs automatically without being triggered by voice traffic as long as the voice VLAN operates correctly.

Manual mode

Use manual mode when only IP phones access the network through the device, as shown in the diagram for connecting the IP phone to the device in "Connecting the IP phone to the device." In this mode, ports are assigned to a voice VLAN that transmits voice traffic exclusively. No data traffic affects the voice traffic transmission.

You must manually assign the port that connects to the IP phone to a voice VLAN. The device uses the source MAC address of the received voice packets to match its OUI addresses. If the match succeeds, the device issues ACL rules to set the packet precedence.

To remove the port from the voice VLAN, you must manually remove it.

Cooperation of voice VLAN assignment modes and IP phones

Some IP phones send out VLAN-tagged packets, and others send out only untagged packets. For correct packet processing, ports of different link types must meet specific configuration requirements in different voice VLAN assignment modes.

If an IP phone sends out tagged voice traffic, and its access port is configured with 802.1X authentication, guest VLAN, Auth-Fail VLAN, or critical VLAN, VLAN IDs must be different for the following VLANs:

- Voice VLAN.
- PVID of the access port.
- 802.1X guest, Auth-Fail, or critical VLAN.

If an IP phone sends out untagged voice traffic, the PVID of the access port must be the voice VLAN. In this scenario, 802.1X authentication is not supported.

Access ports do not transmit tagged packets.

Configuration requirements for transmitting tagged voice traffic

Port link type	Voice VLAN assignment mode	Configuration requirements
Trunk	Automatic	The PVID of the port cannot be the voice VLAN.
	Manual	The PVID of the port cannot be the voice VLAN. The port must forward packets from the voice VLAN.
Hybrid	Automatic	The PVID of the port cannot be the voice VLAN.
	Manual	The PVID of the port cannot be the voice VLAN. The port must forward packets from the voice VLAN with VLAN tags.

Configuration requirements for transmitting untagged voice traffic

When IP phones send out untagged packets, you must set the voice VLAN assignment mode to manual.

Table 2 Configuration requirements for ports in manual mode to support untagged voice traffic

Port link type	Configuration requirements
Access	The voice VLAN must be the PVID of the port.

Port link type	Configuration requirements
Trunk	The voice VLAN must be the PVID of the port. The port must forward packets from the voice VLAN.
Hybrid	The voice VLAN must be the PVID of the port. The port must forward packets from the voice VLAN without VLAN tags.

Security mode and normal mode of voice VLANs

Depending on the filtering mechanisms to incoming packets, a voice VLAN-enabled port can operate in one of the following modes:

- **Normal mode**—The port receives voice-VLAN-tagged packets and forwards them in the voice VLAN without examining their MAC addresses. If the PVID of the port is the voice VLAN and the port operates in manual VLAN assignment mode, the port forwards all the received untagged packets in the voice VLAN.

  In this mode, voice VLANs are vulnerable to traffic attacks. Malicious users might send a large number of forged voice-VLAN-tagged or untagged packets to affect voice communication.
- **Security mode**—The port uses the source MAC addresses of voice packets to match the OUI addresses of the device. Packets that fail the match will be dropped.

In a safe network, you can configure the voice VLANs to operate in normal mode. This mode reduces system resource consumption in source MAC address checking.

In either mode, the device modifies the transmission priority only for voice VLAN packets whose source MAC addresses match OUI addresses of the device.

As a best practice, do not transmit both voice traffic and non-voice traffic in a voice VLAN. If you must transmit different traffic in a voice VLAN, make sure the voice VLAN security mode is disabled.

Table 3 Packet processing on a voice VLAN-enabled port in normal or security mode

Voice VLAN mode	Packet type	Packet processing
Normal	• Untagged packets • Packets with the voice VLAN tags	The port does not examine their source MAC addresses. The device forwards voice packets and non-voice packets in the voice VLAN and modifies the priority settings for packets with source MAC addresses matching OUI addresses.
	Packets with other VLAN tags	The port forwards or drops them depending on whether the port permits packets from these VLANs to pass through.
Security	• Untagged packets • Packets with the voice VLAN tags	• If the source MAC address of a packet matches an OUI address on the device, the device modifies the priority settings for the packet as configured, and forwards the packet in the voice VLAN. • If the source MAC address of a packet does not match an OUI address on the device, the packet is dropped.
	Packets with other VLAN tags	The port forwards or drops them depending on whether the port permits packets from these VLANs to pass through.

Restrictions and guidelines: Voice VLAN configuration

The aging timer of a voice VLAN starts only when the dynamic MAC address entry of the voice VLAN ages out. The aging period for the voice VLAN equals the sum of the voice VLAN aging timer and the aging timer for its dynamic MAC address entry. For more information about the aging timer for dynamic MAC address entries, see "Configuring the MAC address table."

As a best practice, do not both configure voice VLAN and disable MAC address learning on a port. If the two features are configured together on a port, the port forwards only packets exactly matching the OUI addresses and drops inexactly matching packets.

As a best practice, do not configure both voice VLAN and the MAC learning limit on a port. If the two features are configured together on a port and the port learns the configured maximum number of MAC address entries, the port processes packets as follows:

- Forwards only packets matching the MAC address entries learnt by the port and OUI addresses.
- Drops unmatching packets.

Voice VLAN tasks at a glance

To configure a voice VLAN, perform the following tasks:

1. Configuring the QoS priority settings for voice traffic
2. (Optional.) Configuring the ACL resource occupation mode of voice VLAN
3. Use one of the following methods:
   - Configuring a port to operate in automatic voice VLAN assignment mode
   - Configuring a port to operate in manual voice VLAN assignment mode
4. (Optional.) Enabling LLDP for automatic IP phone discovery
5. (Optional.) Use one of the following methods:
   - Configuring LLDP to advertise a voice VLAN
   - Configuring CDP to advertise a voice VLAN

Configuring the QoS priority settings for voice traffic

About this task

You can configure the device to modify the QoS priority settings for voice traffic.

The QoS priority settings carried in voice traffic include the CoS and DSCP values.

- **CoS priority**—Class of service priority defined by 802.1p, a 3-bit priority field in an 802.1Q frame. When the network is congested, a device preferentially sends packets with high 802.1p priority values.
- **DSCP priority**—Differentiated Services Codepoint priority. It is a 6-bit field in the ToS field of an IPv4 packet. On a DiffServ network, a device executes a forwarding action for a packet according to its DSCP priority value.

Restrictions and guidelines

You cannot configure the QoS priority settings on a voice VLAN-enabled port. Before you configure the QoS priority settings for voice traffic on a port, you must disable the voice VLAN feature on it.

If you execute the **voice-vlan qos** and **voice-vlan qos trust** commands multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
3. Configure QoS priority settings for incoming voice VLAN packets.
   - Configure the port to trust the QoS priority settings and not to modify the priority settings for packets.
     voice-vlan qos trust
   - Configure the port to modify the CoS and DSCP values.
     voice-vlan qos *cos-value dscp-value*

   By default, a port modifies the CoS and DSCP values for voice VLAN packets to 6 and 46, respectively.

   If a port trusts the QoS priority settings in incoming voice VLAN packets, the port does not modify their CoS and DSCP values.

Configuring the ACL resource occupation mode of voice VLAN

About this task

With voice VLAN enabled on ports, you can configure the ACL resource occupation mode of voice VLAN as the global mode or the port mode.

- In global mode, ports in the same voice VLAN share one ACL resource. The global mode sharply saves ACL resources.
- In port mode, each port uses its dedicated ACL resource.

Restrictions and guidelines

Before switching the ACL resource occupation mode of voice VLAN from global mode to port mode, make sure the remaining ACL resource are sufficient. If you cannot do that, the system will prompt that the resources are insufficient and not allow the switching.

For the ACL resource occupation mode of voice VLAN to take effect, save the configuration on the device and restart the device. Before restarting the device, as a best practice, do not configure the voice VLAN related features. If you do that, the voice VLAN feature might have defects.

Software version and feature compatibility

This feature is supported only in Release 8307P08 and later.

Procedure

1. Enter system view.
   system-view
2. Configure the ACL resource occupation mode of voice VLAN.
   voice-vlan acl-resource-mode { **global** | **port** }
   By default, the ACL resource occupation mode of voice VLAN is not set.

Configuring voice VLAN assignment modes for a port

Configuring a port to operate in automatic voice VLAN assignment mode

Restrictions and guidelines

- Do not configure a VLAN as both a voice VLAN and a protocol-based VLAN.
  - A voice VLAN in automatic mode on a hybrid port processes only tagged incoming voice traffic.
  - A protocol-based VLAN on a hybrid port processes only untagged incoming packets. For more information about protocol-based VLANs, see "Configuring protocol-based VLANs."
- As a best practice, do not use this mode with MSTP. In MSTP mode, if a port is blocked in the MSTI of the target voice VLAN, the port drops the received packets instead of delivering them to the CPU. As a result, the port will not be dynamically assigned to the voice VLAN.
- As a best practice, do not use this mode with PVST. In PVST mode, if the target voice VLAN is not permitted on a port, the port is placed in blocked state. The port drops the received packets instead of delivering them to the CPU. As a result, the port will not be dynamically assigned to the voice VLAN.
- As a best practice, do not configure both dynamic MAC-based VLAN assignment and automatic voice VLAN assignment mode on a port. They can have a negative impact on each other.

Procedure

1. Enter system view.

   system-view
2. (Optional.) Set the voice VLAN aging timer.

   voice-vlan aging *minutes*

   By default, the aging timer of a voice VLAN is 1440 minutes.

   The voice VLAN aging timer takes effect only on ports in automatic voice VLAN assignment mode.
3. (Optional.) Enable the voice VLAN security mode.

   voice-vlan security enable

   By default, the voice VLAN security mode is enabled.
4. (Optional.) Add an OUI address for voice packet identification.

   voice-vlan mac-address *oui* **mask** *oui-mask* [**description** *text*]

   By default, system default OUI addresses exist. For more information, see the default OUI addresses table in "Identifying IP phones through OUI addresses."
5. Enter Layer 2 Ethernet interface view.

   interface *interface-type interface-number*
6. Configure the link type of the port.
   - **port link-type trunk**
   - **port link-type hybrid**
7. Configure the port to operate in automatic voice VLAN assignment mode.

   voice-vlan mode auto

   By default, the automatic voice VLAN assignment mode is enabled.

8. Enable the voice VLAN feature on the port.

   voice-vlan *vlan-id* **enable**

   By default, the voice VLAN feature is disabled.

   Before you execute this command, make sure the specified VLAN already exists.

Configuring a port to operate in manual voice VLAN assignment mode

Restrictions and guidelines

- You can configure different voice VLANs for different ports on the same device. Make sure the following requirements are met:
  - One port can be configured with only one voice VLAN.
  - Voice VLANs must be existing static VLANs.
- Do not enable voice VLAN on the member ports of a link aggregation group. For more information about link aggregation, see "Configuring Ethernet link aggregation."
- To make a voice VLAN take effect on a port operating in manual mode, you must manually assign the port to the voice VLAN.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Enable the voice VLAN security mode.

   voice-vlan security enable

   By default, the voice VLAN security mode is enabled.

3. (Optional.) Add an OUI address for voice packet identification.

   voice-vlan mac-address *oui* **mask** *oui-mask* [**description** *text*]

   By default, system default OUI addresses exist. For more information, see the default OUI addresses table in "Identifying IP phones through OUI addresses."

4. Enter Layer 2 Ethernet interface view.

   interface *interface-type interface-number*

5. Configure the port to operate in manual voice VLAN assignment mode.

   undo voice-vlan mode auto

   By default, a port operates in automatic voice VLAN assignment mode.

6. Assign the access, trunk, or hybrid port to the voice VLAN.
   - For the access port, see "Assigning an access port to a VLAN."
   - For the trunk port, see "Assigning a trunk port to a VLAN."
   - For the hybrid port, see "Assigning a hybrid port to a VLAN."

   After you assign an access port to the voice VLAN, the voice VLAN becomes the PVID of the port.

7. (Optional.) Configure the voice VLAN as the PVID of the trunk or hybrid port.
   - For the trunk port, see "Assigning a trunk port to a VLAN."
   - For the hybrid port, see "Assigning a hybrid port to a VLAN."

   This step is required for untagged incoming voice traffic and prohibited for tagged incoming voice traffic.

8. Enable the voice VLAN feature on the port.

   voice-vlan *vlan-id* **enable**

By default, the voice VLAN feature is disabled.

Before you execute this command, make sure the specified VLAN already exists.

Enabling LLDP for automatic IP phone discovery

Restrictions and guidelines

- Before you enable this feature, enable LLDP both globally and on access ports.

Procedure

1. Enter system view.
   `system-view`
2. Enable LLDP for automatic IP phone discovery.
   `voice-vlan track lldp`
   By default, this feature is disabled.

Configuring LLDP or CDP to advertise a voice VLAN

Configuring LLDP to advertise a voice VLAN

About this task

For IP phones that support LLDP, the device advertises the voice VLAN information to the IP phones through the LLDP-MED TLVs.

Prerequisites

Before you configure this feature, enable LLDP both globally and on access ports.

Procedure

1. Enter system view.
   `system-view`
2. Enter Layer 2 Ethernet interface view.
   `interface` *`interface-type interface-number`*
3. Configure an advertised voice VLAN ID.
   `lldp tlv-enable med-tlv network-policy` *`vlan-id`*
   By default, no advertised voice VLAN ID is configured.
   For more information about the command, see *Layer 2—LAN Switching Command Reference*.
4. (Optional.) Display the voice VLAN advertised by LLDP.
   `display lldp local-information`
   For more information about the command, see *Layer 2—LAN Switching Command Reference*.

Configuring CDP to advertise a voice VLAN

About this task

If an IP phone supports CDP but does not support LLDP, it will send out CDP packets to the device to request the voice VLAN ID. If the IP phone does not receive the voice VLAN ID within a time period,

it will send out untagged packets. The device cannot differentiate untagged voice packets from other types of packets.

You can configure CDP compatibility on the device to enable it to perform the following operations:

- Receive and identify CDP packets from the IP phone.
- Send CDP packets to the IP phone. The voice VLAN information is carried in the CDP packets.

After receiving the advertised VLAN information, the IP phone performs automatic voice VLAN configuration. Packets from the IP phone will be transmitted in the dedicated voice VLAN.

LLDP packets sent from the device carry the priority information. CDP packets sent from the device do not carry the priority information.

Prerequisites

Before you configure this feature, enable LLDP globally and on access ports.

Procedure

1. Enter system view.
   `system-view`
2. Enable CDP compatibility.
   `lldp compliance cdp`
   By default, CDP compatibility is disabled.
3. Enter Layer 2 Ethernet interface view.
   `interface` *`interface-type interface-number`*
4. Configure CDP-compatible LLDP to operate in TxRx mode.
   `lldp compliance admin-status cdp txrx`
   By default, CDP-compatible LLDP operates in Disable mode.
5. Configure an advertised voice VLAN ID.
   `cdp voice-vlan` *`vlan-id`*
   By default, no advertised voice VLAN ID is configured.
   For more information about the command, see *Layer 2—LAN Switching Command Reference*.

Display and maintenance commands for voice VLANs

Execute **`display`** commands in any view.

Task	Command
Display OUI addresses on a device.	**`display voice-vlan mac-address`**
Display the voice VLAN state.	**`display voice-vlan state`**

Voice VLAN configuration examples

Example: Configuring automatic voice VLAN assignment mode

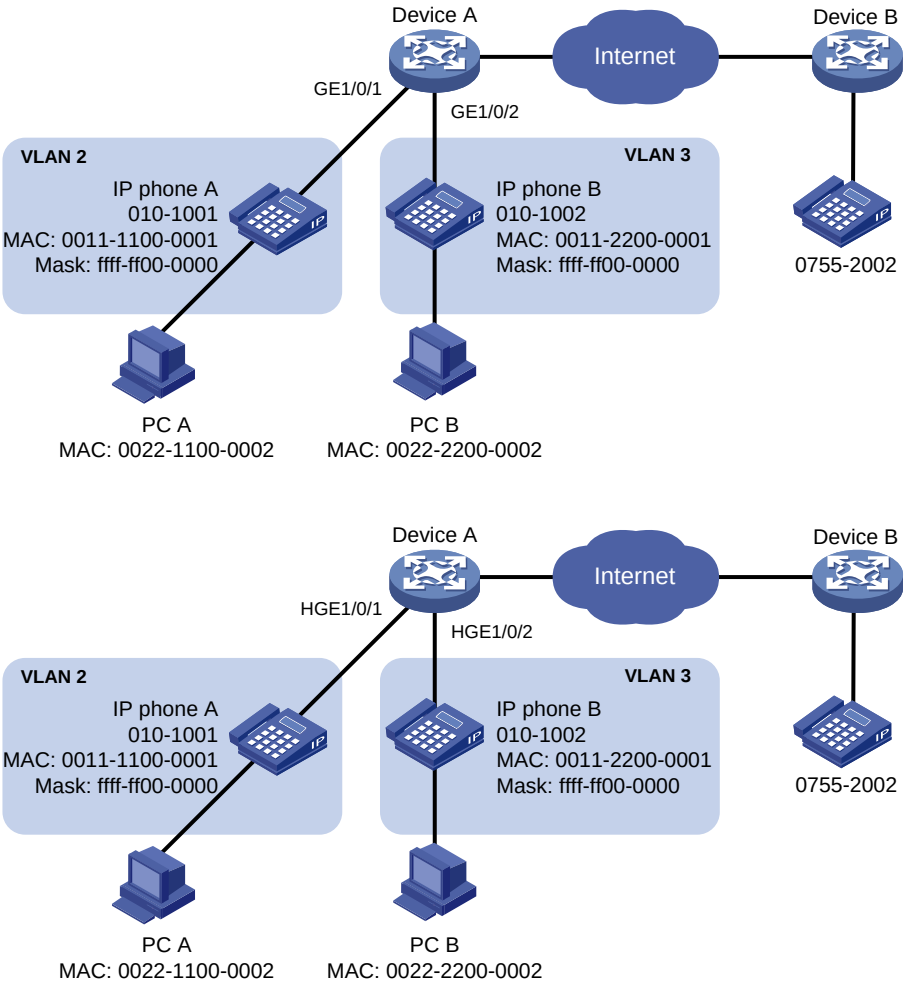
Network configuration

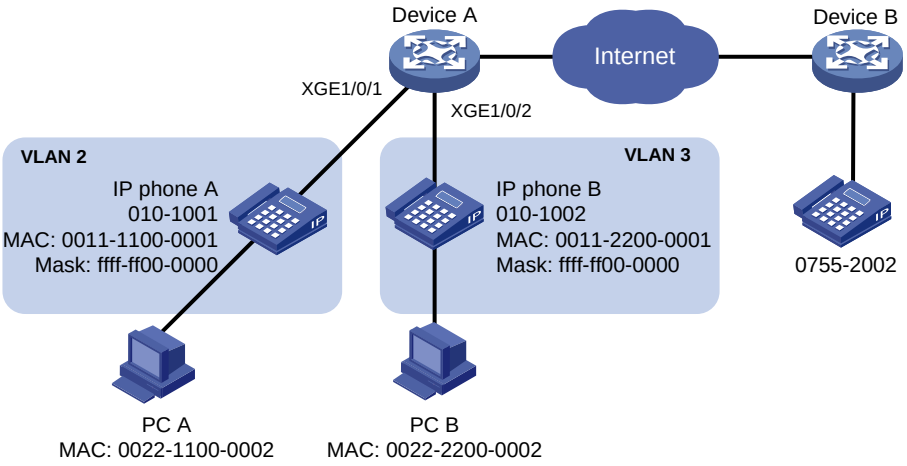
As shown in Figure 17, Device A transmits traffic from IP phones and hosts.

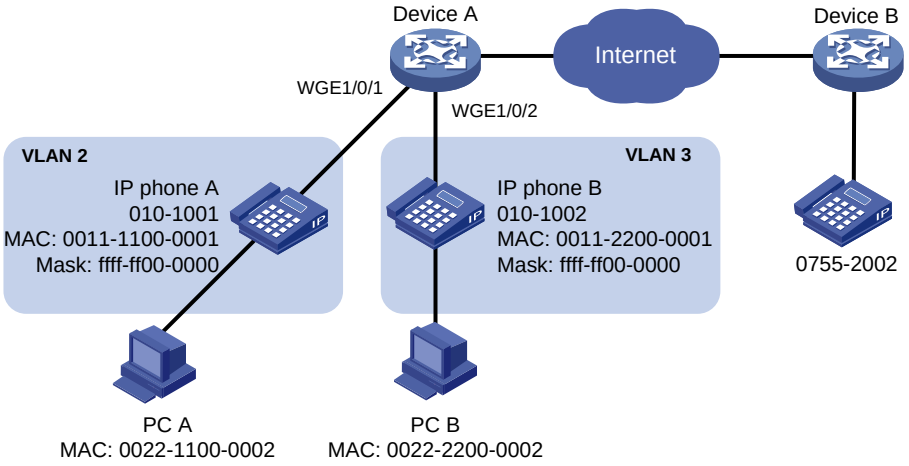
For correct voice traffic transmission, perform the following tasks on Device A:

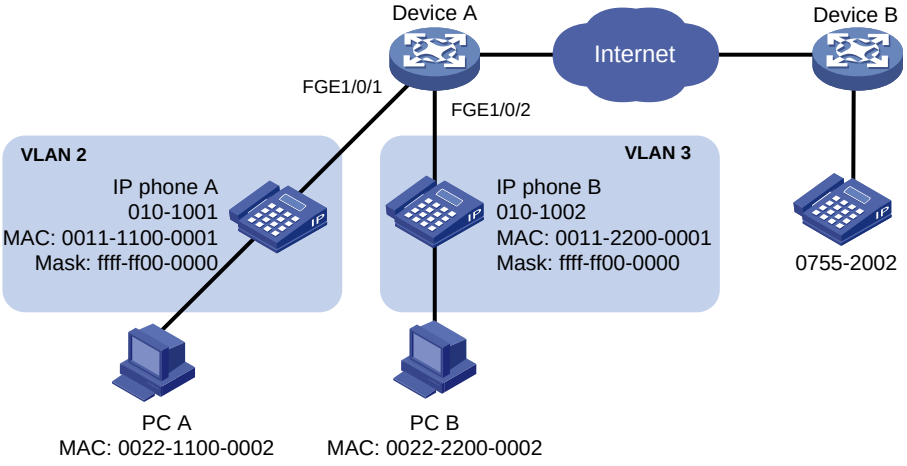
- Configure voice VLANs 2 and 3 to transmit voice packets from IP phone A and IP phone B, respectively.
- Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to operate in automatic voice VLAN assignment mode.
- Add MAC addresses of IP phones A and B to the device for voice packet identification. The mask of the two MAC addresses is FFFF-FF00-0000.
- Set an aging timer for voice VLANs.

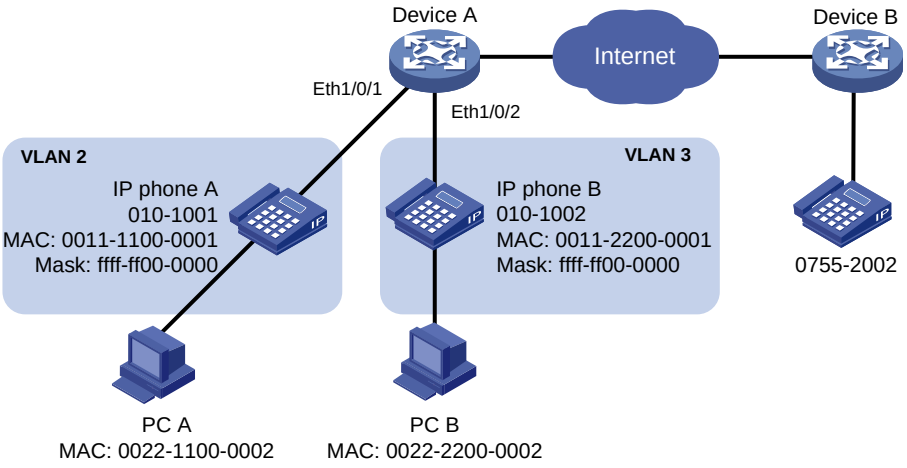
Figure 17 Network diagram

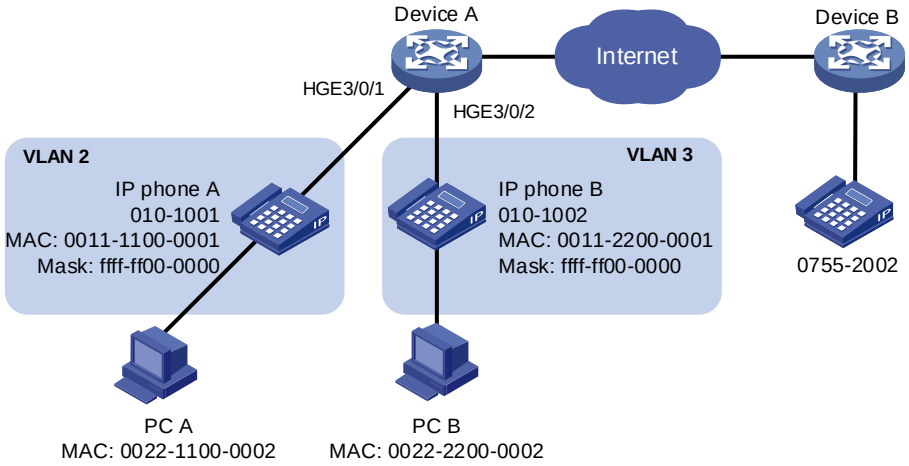


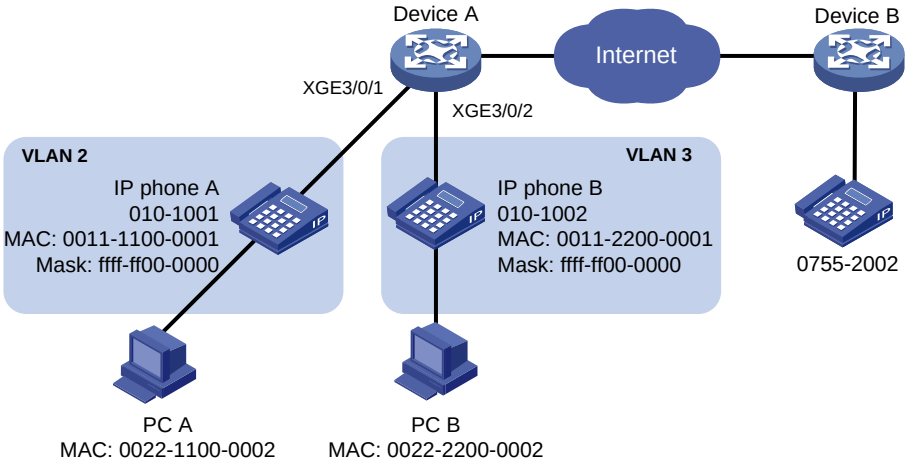

Device A
Internet
Device B
XGE1/0/1
XGE1/0/2
VLAN 2
IP phone A
010-1001
MAC: 0011-1100-0001
Mask: ffff-ff00-0000
VLAN 3
IP phone B
010-1002
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
0755-2002
PC A
MAC: 0022-1100-0002
PC B
MAC: 0022-2200-0002

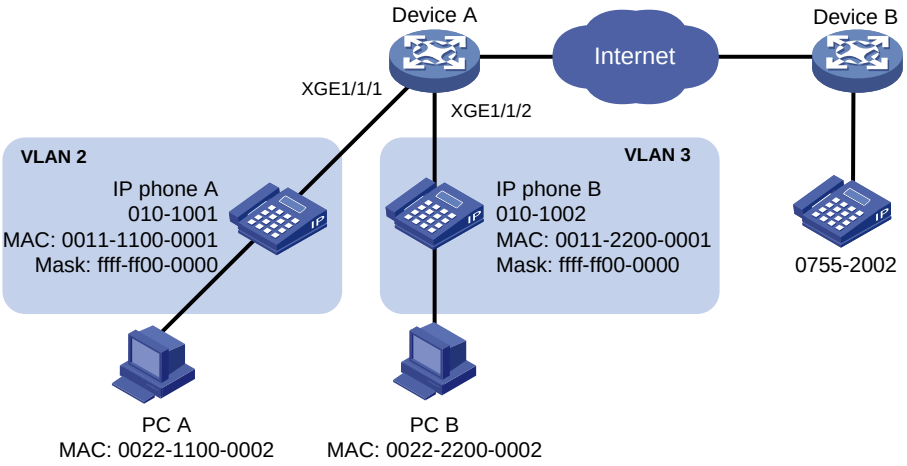

Device A
Internet
Device B
WGE1/0/1
WGE1/0/2
VLAN 2
IP phone A
010-1001
MAC: 0011-1100-0001
Mask: ffff-ff00-0000
VLAN 3
IP phone B
010-1002
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
0755-2002
PC A
MAC: 0022-1100-0002
PC B
MAC: 0022-2200-0002

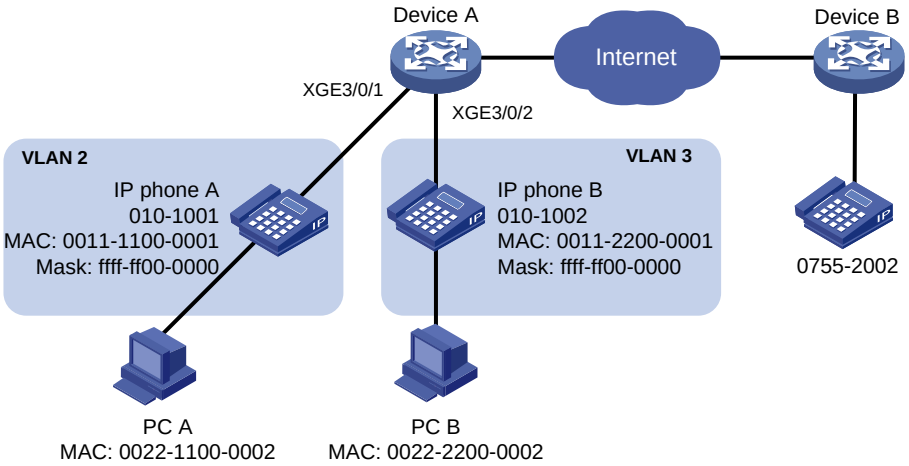

Device A
Internet
Device B
FGE1/0/1
FGE1/0/2
VLAN 2
IP phone A
010-1001
MAC: 0011-1100-0001
Mask: ffff-ff00-0000
VLAN 3
IP phone B
010-1002
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
0755-2002
PC A
MAC: 0022-1100-0002
PC B
MAC: 0022-2200-0002

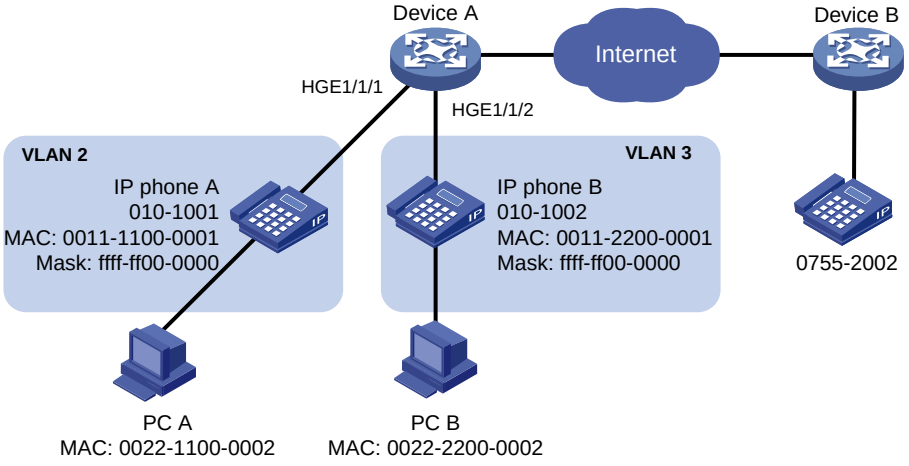

Device A
Internet
Device B
Eth1/0/1
Eth1/0/2
VLAN 2
IP phone A
010-1001
MAC: 0011-1100-0001
Mask: ffff-ff00-0000
VLAN 3
IP phone B
010-1002
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
0755-2002
PC A
MAC: 0022-1100-0002
PC B
MAC: 0022-2200-0002


Device A
Internet
Device B
HGE3/0/1
HGE3/0/2
VLAN 2
IP phone A
010-1001
MAC: 0011-1100-0001
Mask: ffff-ff00-0000
VLAN 3
IP phone B
010-1002
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
0755-2002
PC A
MAC: 0022-1100-0002
PC B
MAC: 0022-2200-0002


Device A
Internet
Device B
XGE3/0/1
XGE3/0/2
VLAN 2
IP phone A
010-1001
MAC: 0011-1100-0001
Mask: ffff-ff00-0000
VLAN 3
IP phone B
010-1002
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
0755-2002
PC A
MAC: 0022-1100-0002
PC B
MAC: 0022-2200-0002







Procedure

1. Configure voice VLANs:

   # Create VLANs 2 and 3.

   ```
   <DeviceA> system-view
   [DeviceA] vlan 2 to 3
   ```

   # Set the voice VLAN aging timer to 30 minutes.

```
[DeviceA] voice-vlan aging 30
```

Enable security mode for voice VLANs.

```
[DeviceA] voice-vlan security enable
```

Add MAC addresses of IP phones A and B to the device with mask FFFF-FF00-0000.

```
[DeviceA] voice-vlan mac-address 0011-1100-0001 mask ffff-ff00-0000 description IP
phone A
[DeviceA] voice-vlan mac-address 0011-2200-0001 mask ffff-ff00-0000 description IP
phone B
```

2. Configure GigabitEthernet 1/0/1:

Configure GigabitEthernet 1/0/1 as a hybrid port.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type hybrid
```

Configure GigabitEthernet 1/0/1 to operate in automatic voice VLAN assignment mode.

```
[DeviceA-GigabitEthernet1/0/1] voice-vlan mode auto
```

Enable voice VLAN on GigabitEthernet 1/0/1 and configure VLAN 2 as the voice VLAN for it.

```
[DeviceA-GigabitEthernet1/0/1] voice-vlan 2 enable
[DeviceA-GigabitEthernet1/0/1] quit
```

3. Configure GigabitEthernet 1/0/2:

Configure GigabitEthernet 1/0/2 as a hybrid port.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type hybrid
```

Configure GigabitEthernet 1/0/2 to operate in automatic voice VLAN assignment mode.

```
[DeviceA-GigabitEthernet1/0/2] voice-vlan mode auto
```

Enable voice VLAN on GigabitEthernet 1/0/2 and configure VLAN 3 as the voice VLAN for it.

```
[DeviceA-GigabitEthernet1/0/2] voice-vlan 3 enable
[DeviceA-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Display the OUI addresses supported on Device A.

```
[DeviceA] display voice-vlan mac-address
OUI Address     Mask            Description
0001-e300-0000  ffff-ff00-0000  Siemens phone
0003-6b00-0000  ffff-ff00-0000  Cisco phone
0004-0d00-0000  ffff-ff00-0000  Avaya phone
000f-e200-0000  ffff-ff00-0000  H3C Aolynk phone
0011-1100-0000  ffff-ff00-0000  IP phone A
0011-2200-0000  ffff-ff00-0000  IP phone B
0060-b900-0000  ffff-ff00-0000  Philips/NEC phone
00d0-1e00-0000  ffff-ff00-0000  Pingtel phone
00e0-7500-0000  ffff-ff00-0000  Polycom phone
00e0-bb00-0000  ffff-ff00-0000  3Com phone
```

Display the voice VLAN state.

```
[DeviceA] display voice-vlan state
 Current voice VLANs: 2
 Voice VLAN security mode: Security
 Voice VLAN aging time: 30 minutes
 Voice VLAN enabled ports and their modes:
 Port                        VLAN        Mode        CoS        DSCP
```

```
GE1/0/1                    2           Auto        6           46
GE1/0/2                    3           Auto        6           46
```

Example: Configuring manual voice VLAN assignment mode

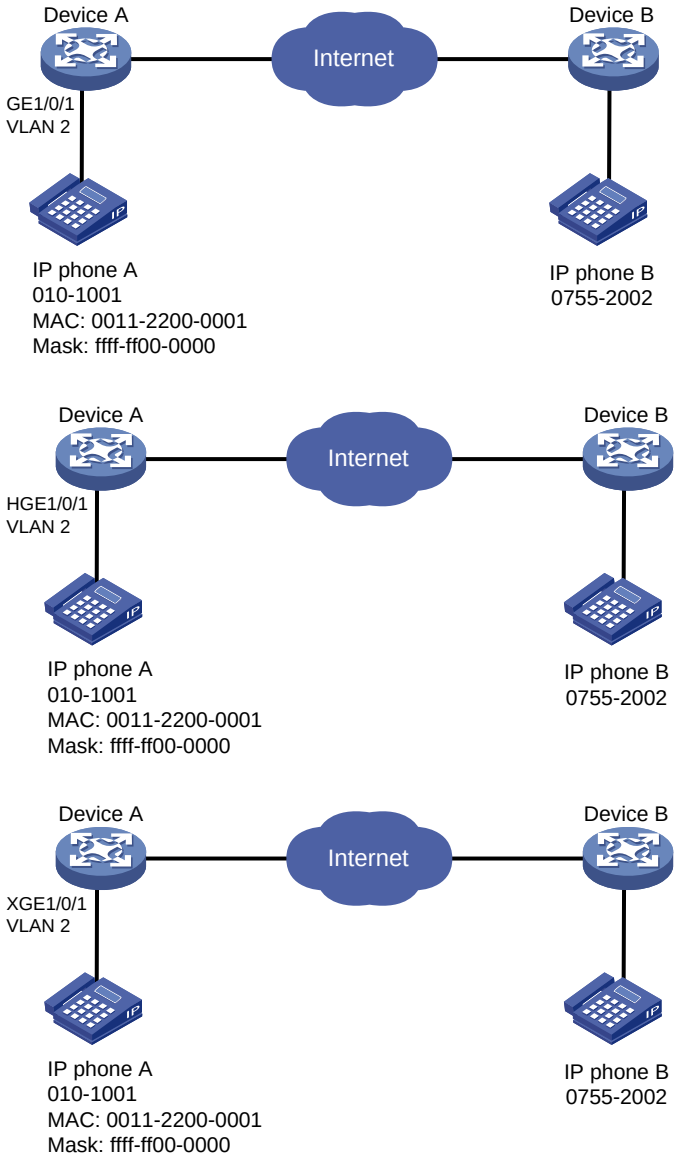
Network configuration

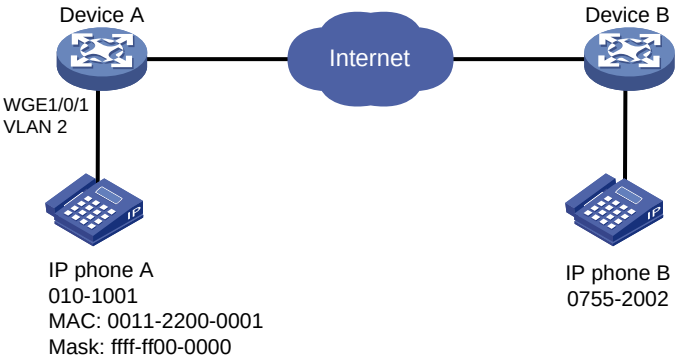
As shown in Figure 18, IP phone A send untagged voice traffic.

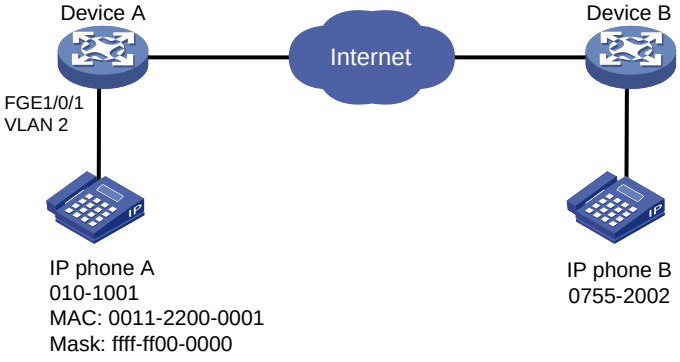
To enable GigabitEthernet 1/0/1 to transmit only voice packets, perform the following tasks on Device A:

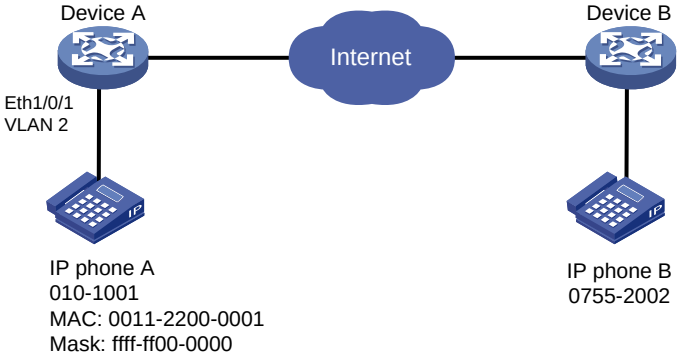
- Create VLAN 2. This VLAN will be used as a voice VLAN.
- Configure GigabitEthernet 1/0/1 to operate in manual voice VLAN assignment mode and add it to VLAN 2.
- Add the OUI address of IP phone A to the OUI list of Device A.

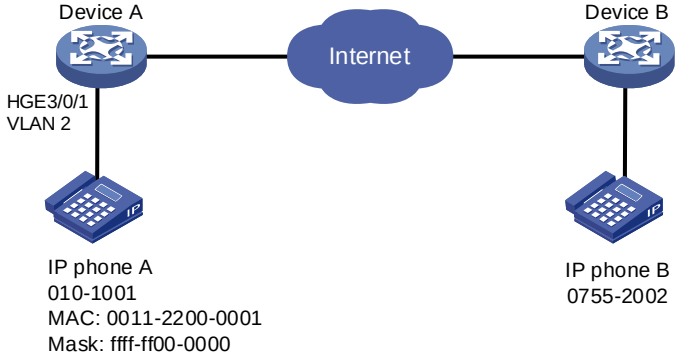
Figure 18 Network diagram

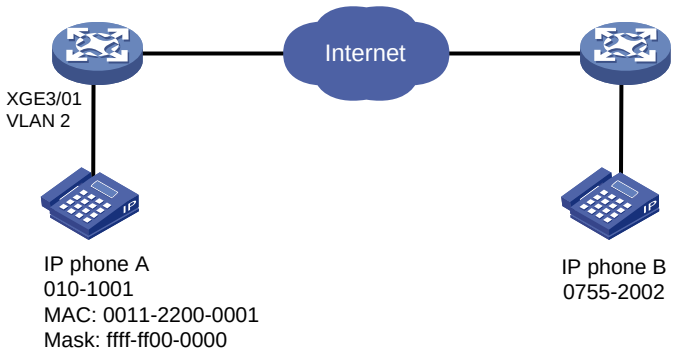


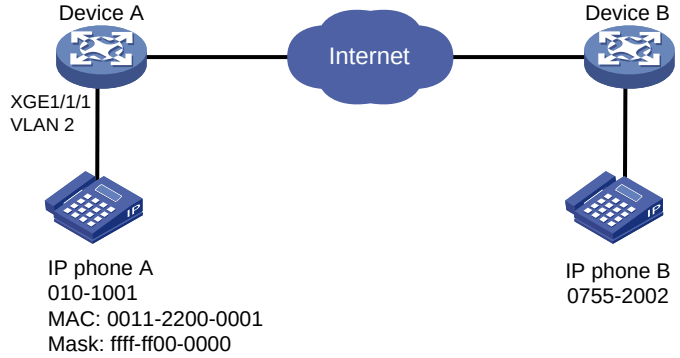

Device A
Internet
Device B
WGE1/0/1
VLAN 2
IP phone A
010-1001
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
IP phone B
0755-2002

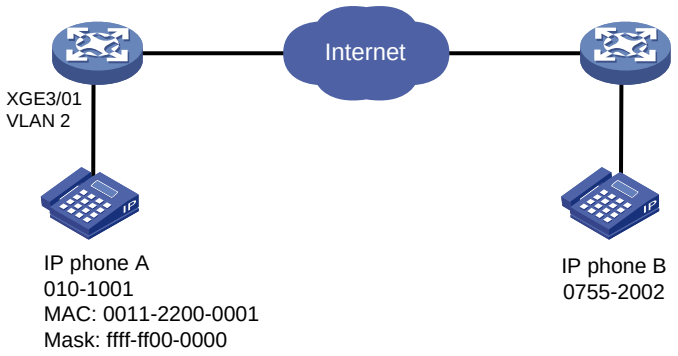

Device A
Internet
Device B
FGE1/0/1
VLAN 2
IP phone A
010-1001
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
IP phone B
0755-2002

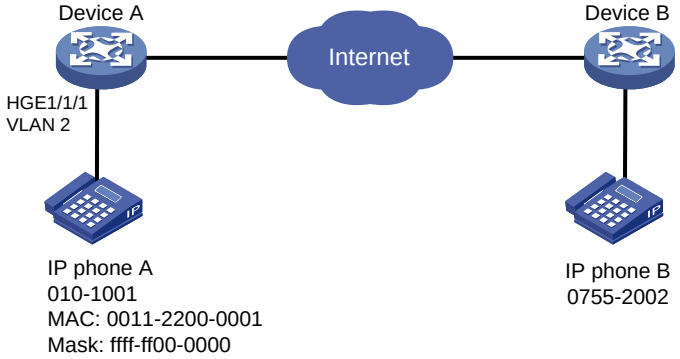

Device A
Internet
Device B
Eth1/0/1
VLAN 2
IP phone A
010-1001
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
IP phone B
0755-2002


Device A
Internet
Device B
HGE3/0/1
VLAN 2
IP phone A
010-1001
MAC: 0011-2200-0001
Mask: ffff-ff00-0000
IP phone B
0755-2002









Procedure

Enable security mode for voice VLANs.

```
<DeviceA> system-view
[DeviceA] voice-vlan security enable
```

Add MAC address 0011-2200-0001 with mask FFFF-FF00-0000.

```
[DeviceA] voice-vlan mac-address 0011-2200-0001 mask ffff-ff00-0000 description test
```

Create VLAN 2.

```
[DeviceA] vlan 2
[DeviceA-vlan2] quit
```

Configure GigabitEthernet 1/0/1 to operate in manual voice VLAN assignment mode.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo voice-vlan mode auto
```

Configure GigabitEthernet 1/0/1 as a hybrid port.

```
[DeviceA-GigabitEthernet1/0/1] port link-type hybrid
```

Set the PVID of GigabitEthernet 1/0/1 to VLAN 2.

```
[DeviceA-GigabitEthernet1/0/1] port hybrid pvid vlan 2
```

Assign GigabitEthernet 1/0/1 to VLAN 2 as an untagged VLAN member.

```
[DeviceA-GigabitEthernet1/0/1] port hybrid vlan 2 untagged
```

Enable voice VLAN and configure VLAN 2 as the voice VLAN on GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] voice-vlan 2 enable
[DeviceA-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Display the OUI addresses supported on Device A.

```
[DeviceA] display voice-vlan mac-address
OUI Address     Mask            Description
0001-e300-0000  ffff-ff00-0000  Siemens phone
0003-6b00-0000  ffff-ff00-0000  Cisco phone
0004-0d00-0000  ffff-ff00-0000  Avaya phone
000f-e200-0000  ffff-ff00-0000  H3C Aolynk phone
0011-2200-0000  ffff-ff00-0000  test
0060-b900-0000  ffff-ff00-0000  Philips/NEC phone
00d0-1e00-0000  ffff-ff00-0000  Pingtel phone
00e0-7500-0000  ffff-ff00-0000  Polycom phone
00e0-bb00-0000  ffff-ff00-0000  3Com phone
```

Display the voice VLAN state.

```
[DeviceA] display voice-vlan state
 Current voice VLANs: 1
 Voice VLAN security mode: Security
 Voice VLAN aging time: 1440 minutes
 Voice VLAN enabled ports and their modes:
 Port                        VLAN        Mode        CoS         DSCP
 GE1/0/1                     2           Manual      6           46
```

Contents

Configuring MVRP 1

About MVRP 1
MRP implementation 1
MRP messages 1
MRP timers 3
MVRP registration modes 3
Protocols and standards 4
Restrictions and guidelines: MVRP configuration 4
MVRP tasks at a glance 4
Prerequisites 4
Enabling MVRP 5
Setting an MVRP registration mode 5
Setting MRP timers 5
Enabling GVRP compatibility 6
Display and maintenance commands for MVRP 7
MVRP configuration examples 7
Example: Configuring basic MVRP functions 7

Configuring MVRP

About MVRP

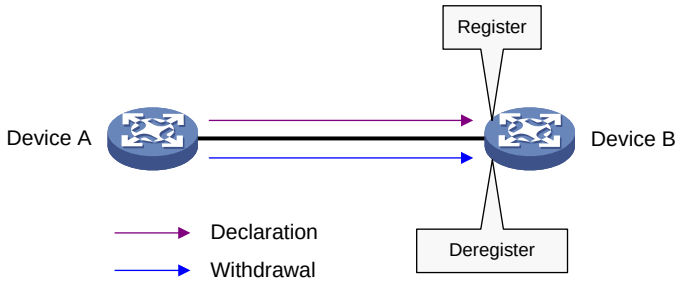
Multiple Registration Protocol (MRP) is an attribute registration protocol used to transmit attribute values. Multiple VLAN Registration Protocol (MVRP) is a typical MRP application. It synchronizes VLAN information among devices and greatly reduces the workload of network administrators.

MRP implementation

An MRP-enabled port is called an MRP participant. An MVRP-enabled port is called an MVRP participant.

As shown in Figure 1, an MRP participant sends declarations and withdrawals to notify other participants to register and deregister its attribute values. It also registers and deregisters the attribute values of other participants according to the received declarations and withdrawals. MRP rapidly propagates the configuration information of an MRP participant throughout the LAN.

Figure 1 MRP implementation



For example, MRP registers and deregisters VLAN attributes as follows:

- When a port receives a declaration for a VLAN, the port registers the VLAN and joins the VLAN.
- When a port receives a withdrawal for a VLAN, the port deregisters the VLAN and leaves the VLAN.

MRP allows devices in the same LAN to transmit attribute values on a per MSTI basis. Figure 1 shows a simple MRP implementation on an MSTI. In a network with multiple MSTIs, MRP performs attribute registration and deregistration on a per MSTI basis. For more information about MSTIs, see "Configuring spanning tree protocols."

MRP messages

MRP messages include the following types:

- **Declaration**—Includes Join and New messages.
- **Withdrawal**—Includes Leave and LeaveAll messages.

Join message

An MRP participant sends a Join message to request the peer participant to register attributes in the Join message.

When receiving a Join message from the peer participant, an MRP participant performs the following tasks:

- Registers the attributes in the Join message.

- Propagates the Join message to all other participants on the device.

After receiving the Join message, other participants send the Join message to their respective peer participants.

Join messages sent from a local participant to its peer participant include the following types:

- **JoinEmpty**—Declares an unregistered attribute. For example, when an MRP participant joins an unregistered static VLAN, it sends a JoinEmpty message.

  VLANs created manually and locally are called static VLANs. VLANs learned through MRP are called dynamic VLANs.
- **JoinIn**—Declares a registered attribute. A JoinIn message is used in one of the following situations:
  - An MRP participant joins an existing static VLAN and sends a JoinIn message after registering the VLAN.
  - The MRP participant receives a Join message propagated by another participant on the device and sends a JoinIn message after registering the VLAN.

New message

Similar to a Join message, a New message enables MRP participants to register attributes.

When the MSTP topology changes, an MRP participant sends a New message to the peer participant to declare the topology change.

Upon receiving a New message from the peer participant, an MRP participant performs the following tasks:

- Registers the attributes in the message.
- Propagates the New message to all other participants on the device.

After receiving the New message, other participants send the New message to their respective peer participants.

Leave message

An MRP participant sends a Leave message to the peer participant when it wants the peer participant to deregister attributes that it has deregistered.

When the peer participant receives the Leave message, it performs the following tasks:

- Deregisters the attribute in the Leave message.
- Propagates the Leave message to all other participants on the device.

After a participant on the device receives the Leave message, it determines whether to send the Leave message to its peer participant depending on the attribute status on the device.

- If the VLAN in the Leave message is a dynamic VLAN not registered by any participants on the device, both of the following events occur:
  - The VLAN is deleted on the device.
  - The participant sends the Leave message to its peer participant.
- If the VLAN in the Leave message is a static VLAN, the participant will not send the Leave message to its peer participant.

LeaveAll message

Each MRP participant starts its LeaveAll timer when starting up. When the timer expires, the MRP participant sends LeaveAll messages to the peer participant.

Upon sending or receiving a LeaveAll message, the local participant starts the Leave timer. The local participant determines whether to send a Join message depending on its attribute status. A participant can re-register the attributes in the received Join message before the Leave timer expires.

When the Leave timer expires, a participant deregisters all attributes that have not been re-registered to periodically clear useless attributes in the network.

MRP timers

MRP uses the following timers to control message transmission.

Periodic timer

The Periodic timer controls the transmission of MRP messages. An MRP participant starts its own Periodic timer upon startup, and stores MRP messages to be sent before the Periodic timer expires. When the Periodic timer expires, MRP sends stored MRP messages in as few MRP frames as possible and restarts the Periodic timer. This mechanism reduces the number of MRP frames sent.

You can enable or disable the Periodic timer. When the Periodic timer is disabled, MRP does not periodically send MRP messages. Instead, an MRP participant sends MRP messages when the LeaveAll timer expires or the participant receives a LeaveAll message from the peer participant.

Join timer

The Join timer controls the transmission of Join messages. An MRP participant starts the Join timer after sending a Join message to the peer participant. Before the Join timer expires, the participant does not resend the Join message when the following conditions exist:

- The participant receives a JoinIn message from the peer participant.
- The received JoinIn message has the same attributes as the sent Join message.

When both the Join timer and the Periodic timer expire, the participant resends the Join message.

Leave timer

The Leave timer controls the deregistration of attributes.

An MRP participant starts the Leave timer in one of the following conditions:

- The participant receives a Leave message from its peer participant.
- The participant receives or sends a LeaveAll message.

The MRP participant does not deregister the attributes in the Leave or LeaveAll message if the following conditions exist:

- The participant receives a Join message before the Leave timer expires.
- The Join message includes the attributes that have been encapsulated in the Leave or LeaveAll message.

If the participant does not receive a Join message for these attributes before the Leave timer expires, MRP deregisters the attributes.

LeaveAll timer

After startup, an MRP participant starts its own LeaveAll timer. When the LeaveAll timer expires, the MRP participant sends out a LeaveAll message and restarts the LeaveAll timer.

Upon receiving the LeaveAll message, other participants restart their LeaveAll timer. The value of the LeaveAll timer is randomly selected between the LeaveAll timer and 1.5 times the LeaveAll timer. This mechanism provides the following benefits:

- Effectively reduces the number of LeaveAll messages in the network.
- Prevents the LeaveAll timer of a particular participant from always expiring first.

MVRP registration modes

VLAN information propagated by MVRP includes dynamic VLAN information from other devices and local static VLAN information.

Based on how an MVRP participant handles registration of dynamic VLANs, MVRP has the following registration modes:

- **Normal**—An MVRP participant in normal registration mode registers and deregisters dynamic VLANs.
- **Fixed**—An MVRP participant in fixed registration mode disables deregistering dynamic VLANs and drops received MVRP frames. The MVRP participant does not deregister dynamic VLANs or register new dynamic VLANs.
- **Forbidden**—An MVRP participant in forbidden registration mode disables registering dynamic VLANs and drops received MVRP frames. When you set the forbidden registration mode for a port, VLAN 1 of the port retains and all dynamically registered VLANs of the port will be deleted.

Protocols and standards

IEEE 802.1ak, *IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks – Amendment 07: Multiple Registration Protocol*

Restrictions and guidelines: MVRP configuration

When you configure MVRP, follow these restrictions and guidelines:

- MVRP can work with STP, RSTP, or MSTP. Ports blocked by STP, RSTP, or MSTP can receive and send MVRP frames. Do not configure MVRP with other link layer topology protocols (such as PVST, RRPP, and Smart Link) on the same port.

  For more information about STP, RSTP, MSTP, and PVST, see "Configuring spanning tree protocols." For more information about RRPP and Smart Link, see *High Availability Configuration Guide*.
- Do not configure both MVRP and remote port mirroring on a port. Otherwise, MVRP might register the remote probe VLAN with incorrect ports, which would cause the monitor port to receive undesired copies. For more information about port mirroring, see *Network Management and Monitoring Configuration Guide*.
- Enabling MVRP on a Layer 2 aggregate interface takes effect on the aggregate interface and all Selected member ports in the link aggregation group.
- MVRP configuration made on an aggregation group member port takes effect only after the port is removed from the aggregation group.

MVRP tasks at a glance

To configure MVRP, perform the following tasks:

1. Enabling MVRP
2. Setting an MVRP registration mode
3. (Optional.) Setting MRP timers
4. (Optional.) Enabling GVRP compatibility

Prerequisites

Before you configure MVRP, complete the following tasks:

- Map each MSTI used by MVRP to an existing VLAN on each device in the network.
- Set the port link type of MVRP participants to trunk because MVRP takes effect only on trunk ports. For more information about trunk ports, see "Configuring VLANs."

Enabling MVRP

1. Enter system view.
   system-view
2. Enable MVRP globally.
   mvrp global enable
   By default, MVRP is globally disabled.
   For MVRP to take effect on a port, enable MVRP both on the port and globally.
3. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   interface *interface-type interface-number*
4. Configure the port as a trunk port.
   port link-type trunk
   By default, each port is an access port. For more information about the **port link-type trunk** command, see *Layer 2—LAN Switching Command Reference*.
5. Configure the trunk port to permit the specified VLANs.
   port trunk permit vlan { *vlan-id-list* | **all** }
   By default, a trunk port permits only VLAN 1.
   Make sure the trunk port permits all registered VLANs.
   For more information about the **port trunk permit vlan** command, see *Layer 2—LAN Switching Command Reference*.
6. Enable MVRP on the port.
   mvrp enable
   By default, MVRP is disabled on a port.

Setting an MVRP registration mode

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   interface *interface-type interface-number*
3. Set an MVRP registration mode for the port.
   mvrp registration { **fixed** | **forbidden** | **normal** }
   The default setting is normal registration mode.

Setting MRP timers

Restrictions and guidelines

When you set MVRP timers, follow these restrictions and guidelines:

- Follow the value range requirements for Join, Leave, and LeaveAll timers and their dependencies as described in Table 1. If you set a timer to a value beyond the allowed value range, your configuration fails. You can set a timer by tuning the value of any other timer. The value of each timer must be an integer multiple of 20 centiseconds.

Table 1 Dependencies of the Join, Leave, and LeaveAll timers

Timer	Lower limit	Upper limit
Join	20 centiseconds	Half the Leave timer
Leave	Twice the Join timer	LeaveAll timer
LeaveAll	Leave timer on each port	32760 centiseconds

- To avoid frequent VLAN registrations and deregistrations, use the same MRP timers throughout the network.
- Each port maintains its own Periodic, Join, and LeaveAll timers, and each attribute of a port maintains a Leave timer.
- As a best practice, restore the timers in the order of Join, Leave, and LeaveAll when you restore these timers to their default values.
- You can restore the Periodic timer to its default value at any time.

Procedure

1. Enter system view.
   `system-view`
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   `interface` *`interface-type interface-number`*
3. Set the LeaveAll timer.
   `mrp timer leaveall` *`timer-value`*
   The default setting is 1000 centiseconds.
4. Set the Join timer.
   `mrp timer join` *`timer-value`*
   The default setting is 20 centiseconds.
5. Set the Leave timer.
   `mrp timer leave` *`timer-value`*
   The default setting is 60 centiseconds.
6. Set the Periodic timer.
   `mrp timer periodic` *`timer-value`*
   The default setting is 100 centiseconds.

Enabling GVRP compatibility

About this task

Perform this task to enable the device to receive and send both MVRP and GVRP frames when the peer device supports GVRP. For more information about GVRP, see the IEEE 802.1Q standard.

Restrictions and guidelines

When you enable GVRP compatibility, follow these restrictions and guidelines:

- GVRP compatibility enables MVRP to work with STP or RSTP, but not MSTP.
- When the system is busy, disable the Period timer to prevent the participant from frequently registering or deregistering attributes.

Procedure

1. Enter system view.

system-view

2. Enable GVRP compatibility.

   mvrp gvrp-compliance enable

   By default, GVRP compatibility is disabled.

Display and maintenance commands for MVRP

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display MVRP running status.	**display mvrp running-status** [**interface** *interface-list*]
Display the MVRP state of a port in a VLAN.	**display mvrp state interface** *interface-type interface-number* **vlan** *vlan-id*
Display MVRP statistics.	**display mvrp statistics** [**interface** *interface-list*]
Clear MVRP statistics.	**reset mvrp statistics** [**interface** *interface-list*]

MVRP configuration examples

Example: Configuring basic MVRP functions

Network configuration

As shown in Figure 2:

- Create VLAN 10 on Device A and VLAN 20 on Device B.
- Configure MSTP, map VLAN 10 to MSTI 1, map VLAN 20 to MSTI 2, and map the other VLANs to MSTI 0 (VLAN 40 as an exmaple). Thus, Host A can reach Host C, and Host B can reach Host D.

Configure MVRP on Device A, Device B, Device C, and Device D to meet the following requirements:

- The devices can register and deregister dynamic VLANs.
- The devices can keep identical VLAN configurations for each MSTI.

Figure 2 Network diagram

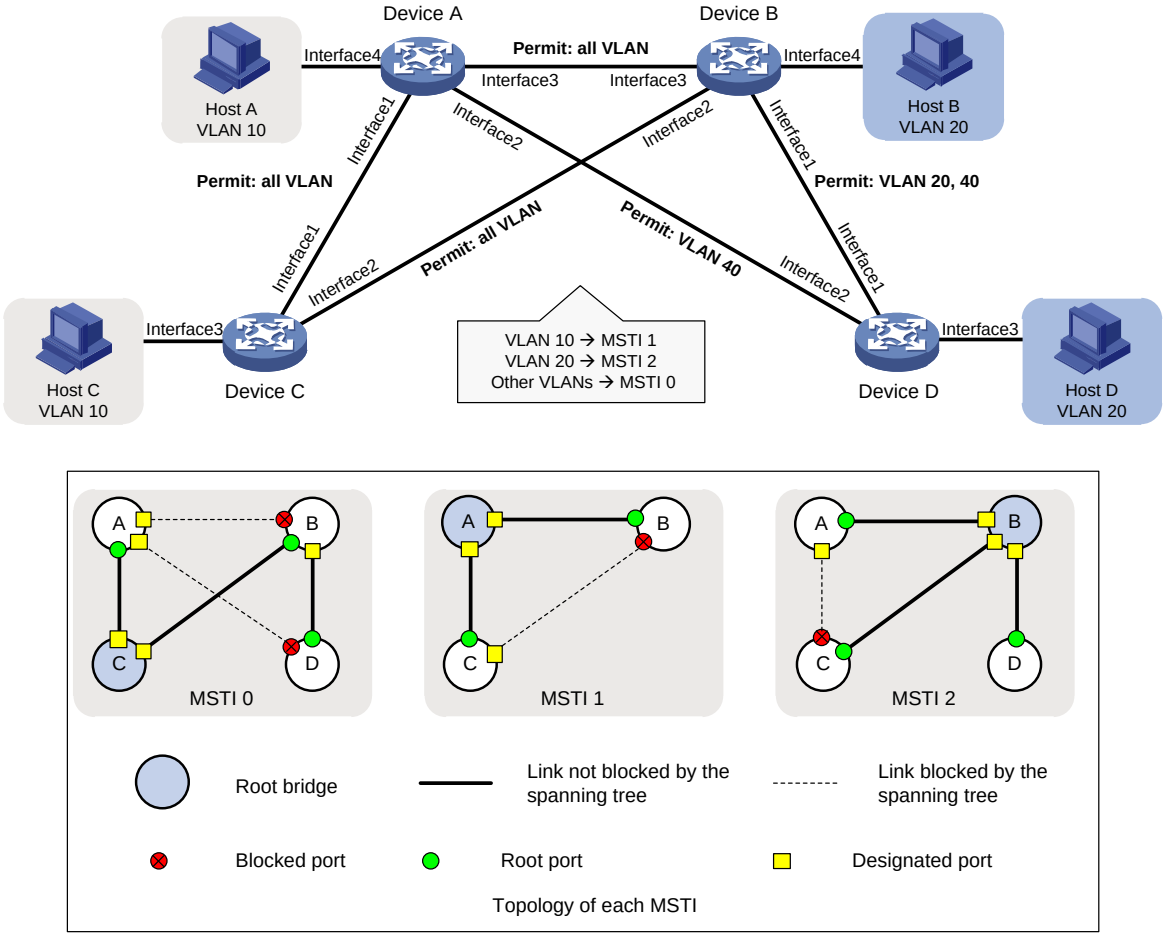


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Device A:
   - Configure port-based VLAN assignment.

     # Configure GigabitEthernet1/0/1 as a trunk port, and configure it to permit all VLANs.

     ```
     <DeviceA> system-view
     [DeviceA] interface gigabitethernet 1/0/1
     [DeviceA-GigabitEthernet1/0/1] port link-type trunk
     [DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
     [DeviceA-GigabitEthernet1/0/1] quit
     ```

     # Configure GigabitEthernet 1/0/2 as a trunk port, and configure it to permit VLAN 40.

     ```
     [DeviceA] interface gigabitethernet 1/0/2
     ```

```
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 40
[DeviceA-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port, and configure it to permit all VLANs.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-type trunk
[DeviceA-GigabitEthernet1/0/3] port trunk permit vlan all
[DeviceA-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet 1/0/4 as a trunk port, configure it to permit VLAN 10, and set the default port VLAN ID to 10.

```
[DeviceA] interface gigabitethernet 1/0/4
[DeviceA-GigabitEthernet1/0/4] port link-type trunk
[DeviceA-GigabitEthernet1/0/4] port trunk permit vlan 10
[DeviceA-GigabitEthernet1/0/4] port trunk pvid vlan 10
[DeviceA-GigabitEthernet1/0/4] quit
```

- Configure MSTP.

  # Enter MST region view.

  ```
  [DeviceA] stp region-configuration
  ```

  # Configure the MST region name, VLAN-to-instance mappings, and revision level.

  ```
  [DeviceA-mst-region] region-name example
  [DeviceA-mst-region] instance 1 vlan 10
  [DeviceA-mst-region] instance 2 vlan 20
  [DeviceA-mst-region] revision-level 0
  ```

  # Manually activate the MST region configuration.

  ```
  [DeviceA-mst-region] active region-configuration
  [DeviceA-mst-region] quit
  ```

  # Configure Device A as the primary root bridge of MSTI 1.

  ```
  [DeviceA] stp instance 1 root primary
  ```

  # Globally enable the spanning tree feature.

  ```
  [DeviceA] stp global enable
  ```

- Configure MVRP.

  # Globally enable MVRP.

  ```
  [DeviceA] mvrp global enable
  ```

  # Enable MVRP on ports GigabitEthernet1/0/1 through GigabitEthernet 1/0/4.

  ```
  [DeviceA] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/4
  [DeviceA-if-range] mvrp enable
  [DeviceA-if-range] quit
  ```

  # Create VLAN 10.

  ```
  [DeviceA] vlan 10
  [DeviceA-vlan10] quit
  ```

2. Configure Device B:

- Configure port-based VLAN assignment.

  # Configure GigabitEthernet1/0/1 as a trunk port, and configure it to permit VLAN 20 and VLAN 40.

  ```
  <DeviceB> system-view
  [DeviceB] interface gigabitethernet 1/0/1
  [DeviceB-GigabitEthernet1/0/1] port link-type trunk
  ```

```
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 20 40
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port, and configure it to permit all VLANs.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port, and configure it to permit all VLANs.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan all
[DeviceB-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet 1/0/4 as a trunk port, configure it to permit VLAN 20, and set the default port VLAN ID to 20.

```
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port link-type trunk
[DeviceB-GigabitEthernet1/0/4] port trunk permit vlan 20
[DeviceB-GigabitEthernet1/0/4] port trunk pvid vlan 20
[DeviceB-GigabitEthernet1/0/4] quit
```

- Configure MSTP.

Enter MST region view.

```
[DeviceB] stp region-configuration
```

Configure the MST region name, VLAN-to-instance mappings, and revision level.

```
[DeviceB-mst-region] region-name example
[DeviceB-mst-region] instance 1 vlan 10
[DeviceB-mst-region] instance 2 vlan 20
[DeviceB-mst-region] revision-level 0
```

Manually activate the MST region configuration.

```
[DeviceB-mst-region] active region-configuration
[DeviceB-mst-region] quit
```

Configure Device B as the primary root bridge of MSTI 2.

```
[DeviceB] stp instance 2 root primary
```

Globally enable the spanning tree feature.

```
[DeviceB] stp global enable
```

- Configure MVRP.

Globally enable MVRP.

```
[DeviceB] mvrp global enable
```

Enable MVRP on GigabitEthernet1/0/1 through GigabitEthernet 1/0/4.

```
[DeviceB] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[DeviceB-if-range] mvrp enable
[DeviceB-if-range] quit
```

Create VLAN 20.

```
[DeviceB] vlan 20
[DeviceB-vlan20] quit
```

3. Configure Device C:

- Configure port-based VLAN assginment.

Configure GigabitEthernet1/0/1 as a trunk port, and configure it to permit all VLANs.

```
<DeviceC> system-view
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port, and configure it to permit all VLANs.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceC-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port, configure it to permit VLAN 10, and set the default port VLAN ID to 10.

```
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] port link-type trunk
[DeviceC-GigabitEthernet1/0/3] port trunk permit vlan 10
[DeviceC-GigabitEthernet1/0/3] port trunk pvid vlan 10
[DeviceC-GigabitEthernet1/0/3] quit
```

- Configure MSTP.

  # Enter MST region view.

  ```
  [DeviceC] stp region-configuration
  ```

  # Configure the MST region name, VLAN-to-instance mappings, and revision level.

  ```
  [DeviceC-mst-region] region-name example
  [DeviceC-mst-region] instance 1 vlan 10
  [DeviceC-mst-region] instance 2 vlan 20
  [DeviceC-mst-region] revision-level 0
  ```

  # Manually activate the MST region configuration.

  ```
  [DeviceC-mst-region] active region-configuration
  [DeviceC-mst-region] quit
  ```

  # Configure Device C as the root bridge of MSTI 0.

  ```
  [DeviceC] stp instance 0 root primary
  ```

  # Globally enable the spanning tree feature.

  ```
  [DeviceC] stp global enable
  ```

- Configure MVRP.

  # Globally enable MVRP.

  ```
  [DeviceC] mvrp global enable
  ```

  # Enable MVRP on GigabitEthernet1/0/1 through GigabitEthernet 1/0/3.

  ```
  [DeviceC] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/3
  [DeviceC-if-range] mvrp enable
  [DeviceC-if-range] quit
  ```

  # Create VLAN 40.

  ```
  [DeviceC] vlan 40
  [DeviceC-vlan40] quit
  ```

4. Configure Device D:

- Configure port-based VLAN assignment.

  # Configure GigabitEthernet1/0/1 as a trunk port, and configure it to permit VLAN 20 and VLAN 40.

  ```
  <DeviceD> system-view
  ```

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 20 40
[DeviceD-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port, and configure it to permit VLAN 40.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 40
[DeviceD-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port, configure it to permit VLAN 20, and set the default port VLAN ID to 20.

```
[DeviceD] interface gigabitethernet 1/0/3
[DeviceD-GigabitEthernet1/0/3] port link-type trunk
[DeviceD-GigabitEthernet1/0/3] port trunk permit vlan 20
[DeviceD-GigabitEthernet1/0/3] port trunk pvid vlan 20
[DeviceD-GigabitEthernet1/0/3] quit
```

- Configure MSTP.

  # Enter MST region view.

  ```
  [DeviceD] stp region-configuration
  ```

  # Configure the MST region name, VLAN-to-instance mappings, and revision level.

  ```
  [DeviceD-mst-region] region-name example
  [DeviceD-mst-region] instance 1 vlan 10
  [DeviceD-mst-region] instance 2 vlan 20
  [DeviceD-mst-region] revision-level 0
  ```

  # Manually activate the MST region configuration.

  ```
  [DeviceD-mst-region] active region-configuration
  [DeviceD-mst-region] quit
  ```

  # Globally enable the spanning tree feature.

  ```
  [DeviceD] stp global enable
  ```

- Configure MVRP.

  # Globally enable MVRP.

  ```
  [DeviceD] mvrp global enable
  ```

  # Enable MVRP on GigabitEthernet1/0/1 through GigabitEthernet 1/0/3.

  ```
  [DeviceD] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/3
  [DeviceD-if-range] mvrp enable
  [DeviceD-if-range] quit
  ```

5. Configure Host A, Host B, Host C, and Host D.

   Make sure the IP addresses of Host A and Host C are on 192.168.10.0/24, while the IP addresses of Host B and Host D are on 192.168.20.0/24.

Verifying the configuration

1. Use the **ping** command to verify the configuration.

   Verify that Host A and Host C can ping each other, and Host B and Host D can ping each other.

2. Verify the normal registration mode configuration.

   # Display local VLAN information on Device A.

   ```
   [DeviceA] display mvrp running-status
    -------[MVRP Global Info]-------
    Global Status     : Enabled
   ```

```
Compliance-GVRP   : False

----[GigabitEthernet1/0/1]----
Config  Status                 : Enabled
Running Status                 : Enabled
Join Timer                     : 20 (centiseconds)
Leave Timer                    : 60 (centiseconds)
Periodic Timer                 : 100 (centiseconds)
LeaveAll Timer                 : 1000 (centiseconds)
Registration Type              : Normal
Registered VLANs :
 1(default), 10, 40
Declared VLANs :
 1(default), 10
Propagated VLANs :
 1(default), 10, 40

----[GigabitEthernet1/0/2]----
Config  Status                 : Enabled
Running Status                 : Enabled
Join Timer                     : 20 (centiseconds)
Leave Timer                    : 60 (centiseconds)
Periodic Timer                 : 100 (centiseconds)
LeaveAll Timer                 : 1000 (centiseconds)
Registration Type              : Normal
Registered VLANs :
 1(default)
Declared VLANs :
 1(default), 40
Propagated VLANs :
 1(default)

----[GigabitEthernet1/0/3]----
Config  Status                 : Enabled
Running Status                 : Enabled
Join Timer                     : 20 (centiseconds)
Leave Timer                    : 60 (centiseconds)
Periodic Timer                 : 100 (centiseconds)
LeaveAll Timer                 : 1000 (centiseconds)
Registration Type              : Normal
Registered VLANs :
 20
Declared  VLANs :
 1(default), 10, 40
Propagated VLANs :
 20

----[GigabitEthernet1/0/4]----
```

```
 Config Status                  : Enabled
 Running Status                 : Enabled
 Join Timer                     : 20 (centiseconds)
 Leave Timer                    : 60 (centiseconds)
 Periodic Timer                 : 100 (centiseconds)
 LeaveAll Timer                 : 1000 (centiseconds)
 Registration Type              : Normal
 Registered VLANs :
  None
 Declared VLANs :
  1(default), 10
 Propagated VLANs :
  None
```

The output shows that the following events have occurred:

- GigabitEthernet1/0/1 has registered and propagated VLAN 1, VLAN 10, and VLAN 40, and declared VLAN 1 and VLAN 10.
- GigabitEthernet 1/0/2 has registered and propagated VLAN 1, and declared VLAN 1 and VLAN 40.
- GigabitEthernet 1/0/3 has registered and propagated VLAN 20, and declared VLAN 1, VLAN 10, and VLAN 40.
- GigabitEthernet 1/0/4 has not registered or propagated any VLANs, but has declared VLAN 1 and VLAN 10.

Display local VLAN information on Device B.

```
[DeviceB] display mvrp running-status
 -------[MVRP Global Info]-------
 Global Status     : Enabled
 Compliance-GVRP   : False

 ----[GigabitEthernet1/0/1]----
 Config  Status                 : Enabled
 Running Status                 : Enabled
 Join Timer                     : 20 (centiseconds)
 Leave Timer                    : 60 (centiseconds)
 Periodic Timer                 : 100 (centiseconds)
 LeaveAll Timer                 : 1000 (centiseconds)
 Registration Type              : Normal
 Registered VLANs :
  None
 Declared VLANs :
  1(default), 20, 40
 Propagated VLANs :
  None

 ----[GigabitEthernet1/0/2]----
 Config  Status                 : Enabled
 Running Status                 : Enabled
 Join Timer                     : 20 (centiseconds)
 Leave Timer                    : 60 (centiseconds)
 Periodic Timer                 : 100 (centiseconds)
```

```
 LeaveAll Timer                : 1000 (centiseconds)
 Registration Type             : Normal
 Registered VLANs :
  1(default), 10, 40
 Declared VLANs :
  1(default), 20
 Propagated VLANs :
  1(default), 40

 ----[GigabitEthernet1/0/3]----
 Config  Status                : Enabled
 Running Status                : Enabled
 Join Timer                    : 20 (centiseconds)
 Leave Timer                   : 60 (centiseconds)
 Periodic Timer                : 100 (centiseconds)
 LeaveAll Timer                : 1000 (centiseconds)
 Registration Type             : Normal
 Registered VLANs :
  1(default), 10, 40
 Declared VLANs :
  20
 Propagated VLANs :
  10

----[GigabitEthernet1/0/4]----
 Config Status                 : Enabled
 Running Status                : Disabled
 Join Timer                    : 20 (centiseconds)
 Leave Timer                   : 60 (centiseconds)
 Periodic Timer                : 100 (centiseconds)
 LeaveAll Timer                : 1000 (centiseconds)
 Registration Type             : Normal
 Registered VLANs :
  None
 Declared VLANs :
  1(default), 20
 Propagated VLANs :
  None
```

The output shows that the following events have occurred:

- GigabitEthernet 1/0/1 has not registered or propagated any VLANs, but has declared VLAN 1, VLAN 20, and VLAN 40.
- GigabitEthernet 1/0/2 has registered VLAN 1, VLAN 10, and VLAN 40, declared VLAN 1 and VLAN 20, and propagated VLAN 1 and VLAN 40.
- GigabitEthernet 1/0/3 has registered VLAN 1, VLAN 10, and VLAN 40, declared VLAN 20, and propagated VLAN 10.
- GigabitEthernet 1/0/4 has not registered or propagated any VLANs, but has declared VLAN 1 and VLAN 20.

Display local VLAN information on Device C.

```
[DeviceC] display mvrp running-status
```

```
-------[MVRP Global Info]-------
Global Status     : Enabled
Compliance-GVRP   : False

----[GigabitEthernet1/0/1]----
Config  Status                 : Enabled
Running Status                 : Enabled
Join Timer                     : 20 (centiseconds)
Leave Timer                    : 60 (centiseconds)
Periodic Timer                 : 100 (centiseconds)
LeaveAll Timer                 : 1000 (centiseconds)
Registration Type              : Normal
Registered VLANs :
 1(default), 10
Declared VLANs :
 1(default), 10, 40
Propagated VLANs :
 1(default), 10

----[GigabitEthernet1/0/2]----
Config  Status                 : Enabled
Running Status                 : Enabled
Join Timer                     : 20 (centiseconds)
Leave Timer                    : 60 (centiseconds)
Periodic Timer                 : 100 (centiseconds)
LeaveAll Timer                 : 1000 (centiseconds)
Registration Type              : Normal
Registered VLANs :
 1(default), 20
Declared VLANs :
 1(default), 40
Propagated VLANs :
 1(default), 20

----[GigabitEthernet1/0/3]----
Config Status                  : Enabled
Running Status                 : Enabled
Join Timer                     : 20 (centiseconds)
Leave Timer                    : 60 (centiseconds)
Periodic Timer                 : 100 (centiseconds)
LeaveAll Timer                 : 1000 (centiseconds)
Registration Type              : Normal
Registered VLANs :
 None
Declared VLANs :
 1(default), 10
Propagated VLANs :
 None
```

The output shows that the following events have occurred:

- GigabitEthernet 1/0/1 has registered and propagated VLAN 1 and VLAN 10, and declared VLAN 1, VLAN 10, and VLAN 40.
- GigabitEthernet 1/0/2 has registered and propagated VLAN 1 and VLAN 20, and declared VLAN 1 and VLAN 40.
- GigabitEthernet 1/0/3 has not registered or propagated any VLANs, and has declared VLAN 1 and VLAN 10.

Display local VLAN information on Device D.

```
[DeviceD] display mvrp running-status
 -------[MVRP Global Info]-------
 Global Status     : Enabled
 Compliance-GVRP   : False

 ----[GigabitEthernet1/0/1]----
 Config  Status                 : Enabled
 Running Status                 : Enabled
 Join Timer                     : 20 (centiseconds)
 Leave Timer                    : 60 (centiseconds)
 Periodic Timer                 : 100 (centiseconds)
 LeaveAll Timer                 : 1000 (centiseconds)
 Registration Type              : Normal
 Registered VLANs :
  1(default), 20, 40
 Declared VLANs :
  None
 Propagated VLANs :
  1(default), 20, 40

 ----[GigabitEthernet1/0/2]----
 Config  Status                 : Enabled
 Running Status                 : Enabled
 Join Timer                     : 20 (centiseconds)
 Leave Timer                    : 60 (centiseconds)
 Periodic Timer                 : 100 (centiseconds)
 LeaveAll Timer                 : 1000 (centiseconds)
 Registration Type              : Normal
 Registered VLANs :
  1(default), 40
 Declared  VLANs :
  1(default)
 Propagated VLANs :
  1(default)

----[GigabitEthernet1/0/3]----
 Config Status                  : Enabled
 Running Status                 : Enabled
 Join Timer                     : 20 (centiseconds)
 Leave Timer                    : 60 (centiseconds)
 Periodic Timer                 : 100 (centiseconds)
```

```
 LeaveAll Timer                : 1000 (centiseconds)
 Registration Type             : Normal
 Registered VLANs :
  None
 Declared VLANs :
  1(default), 20
 Propagated VLANs :
  None
```

The output shows that the following events have occurred:

- GigabitEthernet 1/0/1 has registered and propagated VLAN 10, VLAN 20, and VLAN 40, and has not declared any VLANs.
- GigabitEthernet 1/0/2 has registered and propagated VLAN 1 and VLAN 40, and declared VLAN 1.
- GigabitEthernet 1/0/3 has not registered or propagated VLANs, and has declared VLAN 1 and VLAN 20.

3. Verify the configuration after changing the registration mode.

When the network is stable, set the MVRP registration mode to **fixed** on the port of Device B connected to Device A. Then, verify that dynamic VLANs on the port will not be deregistered.

Set the MVRP registration mode to **fixed** on GigabitEthernet 1/0/3 of Device B.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] mvrp registration fixed
[DeviceB-GigabitEthernet1/0/3] quit
[DeviceB] commit
```

Display local MVRP VLAN information on GigabitEthernet 1/0/3.

```
[DeviceB] display mvrp running-status interface gigabitethernet 1/0/3
 -------[MVRP Global Info]-------
 Global Status     : Enabled
 Compliance-GVRP   : False

 ----[GigabitEthernet1/0/3]----
 Config  Status                : Enabled
 Running Status                : Enabled
 Join Timer                    : 20 (centiseconds)
 Leave Timer                   : 60 (centiseconds)
 Periodic Timer                : 100 (centiseconds)
 LeaveAll Timer                : 1000 (centiseconds)
 Registration Type             : Fixed
 Registered VLANs :
  1(default), 10, 40
 Declared VLANs :
  20
 Propagated VLANs :
  10
```

The output shows that VLAN information on GigabitEthernet 1/0/3 is not changed after you set its MVRP registration mode to **fixed**.

Delete VLAN 10 on Device A.

```
[DeviceA] undo vlan 10
[DeviceA] commit
```

Display local MVRP VLAN information on GigabitEthernet 1/0/3 of Device B.

```
[DeviceB] display mvrp running-status interface gigabitethernet 1/0/3
 -------[MVRP Global Info]-------
 Global Status     : Enabled
 Compliance-GVRP   : False

 ----[GigabitEthernet1/0/3]----
 Config  Status                 : Enabled
 Running Status                 : Enabled
 Join Timer                     : 20 (centiseconds)
 Leave Timer                    : 60 (centiseconds)
 Periodic Timer                 : 100 (centiseconds)
 LeaveAll Timer                 : 1000 (centiseconds)
 Registration Type              : Fixed
 Registered VLANs :
  1(default), 10, 40
 Declared VLANs :
  20
 Propagated VLANs :
  10
```

The output shows that dynamic VLAN information on GigabitEthernet 1/0/3 is not changed after you set its MVRP registration mode to **fixed**.

Configuration files

Device A

```
#
 mvrp global enable
#
stp region-configuration
 region-name example
 instance 1 vlan 10
 instance 2 vlan 20
 active region-configuration
#
 stp instance 1 root primary
 stp global enable
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan all
 mvrp enable
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 40
 mvrp enable
#
interface GigabitEthernet1/0/3
 port link-type trunk
```

```
 port trunk permit vlan all
 mvrp enable
#
interface GigabitEthernet1/0/4
 port link-type trunk
 port trunk permit vlan 1 10
 port trunk pvid vlan 10
 mvrp enable
#
```

Device B

```
#
 mvrp global enable
#
vlan 20
#
stp region-configuration
 region-name example
 instance 1 vlan 10
 instance 2 vlan 20
 active region-configuration
#
 stp instance 2 root primary
 stp global enable
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 20 40
 mvrp enable
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan all
 mvrp enable
#
interface GigabitEthernet1/0/3
 port link-type trunk
 port trunk permit vlan all
 mvrp registration fixed
 mvrp enable
#
interface GigabitEthernet1/0/4
 port link-type trunk
 port trunk permit vlan 1 20
 port trunk pvid vlan 20
 mvrp enable
#
```

Device C

```
#
```

```
 mvrp global enable
#
vlan 40
#
stp region-configuration
 region-name example
 instance 1 vlan 10
 instance 2 vlan 20
 active region-configuration
#
 stp instance 0 root primary
 stp global enable
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan all
 mvrp enable
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan all
 mvrp enable
#
interface GigabitEthernet1/0/3
 port link-type trunk
 port trunk permit vlan 1 10
 port trunk pvid vlan 10
 mvrp enable
#
```

Device D

```
#
 mvrp global enable
#
stp region-configuration
 region-name example
 instance 1 vlan 10
 instance 2 vlan 20
 active region-configuration
#
 stp global enable
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 20 40
 mvrp enable
#
interface GigabitEthernet1/0/2
 port link-type trunk
```

```
 port trunk permit vlan 1 40
 mvrp enable
#
interface GigabitEthernet1/0/3
 port link-type trunk
 port trunk permit vlan 1 20
 port trunk pvid vlan 20
 mvrp enable
#
```

Contents

Configuring loopback, null, and inloopback interfaces 1
About loopback, null, and inloopback interfaces 1
About loopback interfaces 1
About null interfaces 1
About inloopback interfaces 1
Configuring a loopback interface 1
Configuring a null interface 2
Restoring the default settings for an interface 2
Display and maintenance commands for loopback, null, and inloopback interfaces 3

Configuring loopback, null, and inloopback interfaces

This chapter describes how to configure a loopback interface, a null interface, and an inloopback interface.

About loopback, null, and inloopback interfaces

About loopback interfaces

A loopback interface is a virtual interface. The physical layer state of a loopback interface is always up unless the loopback interface is manually shut down. Because of this benefit, loopback interfaces are widely used in the following scenarios:

- **Configuring a loopback interface address as the source address of the IP packets that the device generates**—Because loopback interface addresses are stable unicast addresses, they are usually used as device identifications.

  When you configure a rule on an authentication or security server, you can configure it to permit or deny packets carrying the loopback interface address of a device. This simplifies your configuration and achieves the effect of permitting or denying packets that the device generates. To use a loopback interface address as the source address of IP packets, make sure the loopback interface is reachable from the peer by performing routing configuration. All data packets sent to the loopback interface are considered packets sent to the device itself, so the device does not forward these packets.
- **Using a loopback interface in dynamic routing protocols**—With no router ID configured for a dynamic routing protocol, the system selects the highest loopback interface IP address as the router ID. In BGP, to avoid interruption of BGP sessions due to physical port failure, you can use a loopback interface as the source interface of BGP packets.

About null interfaces

A null interface is a virtual interface and is always up, but you cannot use it to forward data packets or configure it with an IP address or link layer protocol. The null interface provides a simpler way to filter packets than ACL. You can filter undesired traffic by transmitting it to a null interface instead of applying an ACL. For example, if you specify a null interface as the next hop of a static route to a network segment, any packets routed to the network segment are dropped.

About inloopback interfaces

An inloopback interface is a virtual interface created by the system, which cannot be configured or deleted. The physical layer and link layer protocol states of an inloopback interface are always up. All IP packets sent to an inloopback interface are considered packets sent to the device itself and are not forwarded.

Configuring a loopback interface

1. Enter system view.

   ```
   system-view
   ```
2. Create a loopback interface and enter loopback interface view.

interface loopback *interface-number*

3. Configure the interface description.

   description *text*

   The default setting is *interface name* **Interface** (for example, **LoopBack1 Interface**).

4. Configure the expected bandwidth of the loopback interface.

   bandwidth *bandwidth-value*

   By default, the expected bandwidth of a loopback interface is 0 kbps.

5. Bring up the loopback interface.

   undo shutdown

   By default, a loopback interface is up.

Configuring a null interface

1. Enter system view.

   system-view

2. Enter null interface view.

   interface null 0

   Interface Null 0 is the default null interface on the device and cannot be manually created or removed.

   Only one null interface, Null 0, is supported on the device. The null interface number is always 0.

3. Configure the interface description.

   description *text*

   The default setting is NULL0 Interface.

Restoring the default settings for an interface

Restrictions and guidelines

CAUTION:

This feature might interrupt ongoing network services. Make sure you are fully aware of the impact of this feature when you use it on a live network.

This feature might fail to restore the default settings for some commands because of command dependencies or system restrictions. You can use the **display this** command in interface view to check for these commands and perform their **undo** forms or follow the command reference to restore their default settings. If your restoration attempt still fails, follow the error message to resolve the problem.

Procedure

1. Enter system view.

   system-view

2. Enter loopback interface view or null interface view.
   - **interface loopback** *interface-number*
   - **interface null 0**

3. Restore the default settings for the interface.

   default

Display and maintenance commands for loopback, null, and inloopback interfaces

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about the inloopback interface.	**display interface** [**inloopback** [**0**]] [**brief** [**description** | **down**]]
Display information about the specified or all loopback interfaces.	**display interface** [**loopback** [*interface-number*]] [**brief** [**description** | **down**]]
Display information about the null interface.	**display interface** [**null** [**0**]] [**brief** [**description** | **down**]]
Clear the statistics on the specified or all loopback interfaces.	**reset counters interface** [**loopback** [*interface-number*]]
Clear the statistics on the null interface.	**reset counters interface** [**null** [**0**]]

Contents

Configuring QinQ 1

About QinQ 1
QinQ benefits 1
How QinQ works 1
QinQ implementations 2
Protocols and standards 3
Restrictions and guidelines: QinQ configuration 3
Enabling QinQ 3
Configuring transmission for transparent VLANs 4
Configuring the TPID for VLAN tags 5
About TPID 5
Restrictions and guidelines 5
Configuring the TPID for CVLAN tags 5
Configuring the TPID for SVLAN tags 6
Setting the 802.1p priority in SVLAN tags 6
About the 802.1p priority in SVLAN tags 6
Prerequisites for setting the 802.1p priority in SVLAN tags 6
Tasks at a glance 6
Creating a traffic class and configuring CVLAN match criteria 6
Creating a traffic behavior and configuring a priority marking action for SVLAN tags 7
Creating a QoS policy 7
Applying the QoS policy 7
Display and maintenance commands for QinQ 7
QinQ configuration examples 8
Example: Configuring VLAN transparent transmission 8
Example: Configuring basic QinQ 13

Configuring QinQ

This document uses the following terms:

- **CVLAN**—Customer network VLANs, also called inner VLANs, refer to VLANs that a customer uses on the private network.
- **SVLAN**—Service provider network VLANs, also called outer VLANs, refer to VLANs that a service provider uses to transmit VLAN tagged traffic for customers.

About QinQ

802.1Q-in-802.1Q (QinQ) adds an 802.1Q tag to 802.1Q tagged customer traffic. It enables a service provider to extend Layer 2 connections across an Ethernet network between customer sites.

QinQ benefits

QinQ provides the following benefits:

- Enables a service provider to use a single SVLAN to convey multiple CVLANs for a customer.
- Enables customers to plan CVLANs without conflicting with SVLANs.
- Enables customers to keep their VLAN assignment schemes unchanged when the service provider changes its VLAN assignment scheme.
- Allows different customers to use overlapping CVLAN IDs. Devices in the service provider network make forwarding decisions based on SVLAN IDs instead of CVLAN IDs.

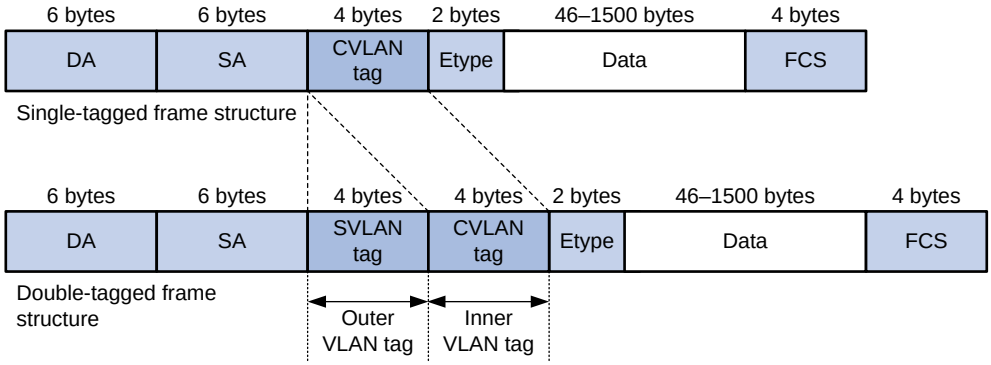
How QinQ works

As shown in Figure 1, a QinQ frame transmitted over the service provider network carries the following tags:

- **CVLAN tag**—Identifies the VLAN to which the frame belongs when it is transmitted in the customer network.
- **SVLAN tag**—Identifies the VLAN to which the QinQ frame belongs when it is transmitted in the service provider network. The service provider allocates the SVLAN tag to the customer.

The devices in the service provider network forward a tagged frame according to its SVLAN tag only. The CVLAN tag is transmitted as part of the frame's payload.

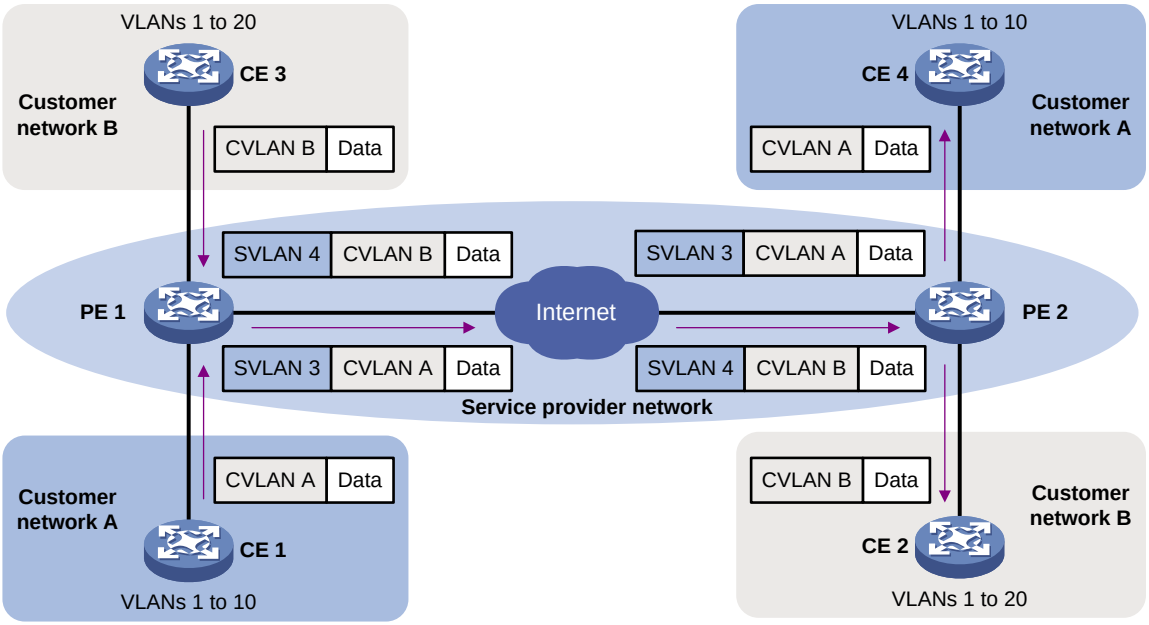
Figure 1 Single-tagged Ethernet frame header and double-tagged Ethernet frame header



As shown in Figure 2, customer A has remote sites CE 1 and CE 4. Customer B has remote sites CE 2 and CE 3. The CVLANs of the two customers overlap. The service provider assigns SVLANs 3 and 4 to customers A and B, respectively.

When a tagged Ethernet frame from CE 1 arrives at PE 1, the PE tags the frame with SVLAN 3. The double-tagged Ethernet frame travels over the service provider network until it arrives at PE 2. PE 2 removes the SVLAN tag of the frame, and then sends the frame to CE 4.

Figure 2 Typical QinQ application scenario



QinQ implementations

QinQ is enabled on a per-port basis. The link type of a QinQ-enabled port can be access, hybrid, or trunk. The QinQ tagging behaviors are the same across these types of ports.

A QinQ-enabled port tags all incoming frames (tagged or untagged) with the PVID tag.

- If an incoming frame already has one tag, it becomes a double-tagged frame.
- If the frame does not have any 802.1Q tags, it becomes a frame tagged with the PVID.

QinQ provides the most basic VLAN manipulation method to tag all incoming frames (tagged or untagged) with the PVID tag. To perform advanced VLAN manipulations, use VLAN mappings or QoS policies as follows:

- To replace the SVLAN ID, CVLAN ID, or both IDs for an incoming double-tagged frame, use two-to-two VLAN mappings.
- To use criteria other than the CVLAN ID to match packets for SVLAN tagging, use the QoS nest action. The QoS nest action can also be used with other actions in the same traffic behavior.
- To set the 802.1p priority in SVLAN tags, use the priority marking action as described in "Setting the 802.1p priority in SVLAN tags."

For more information about VLAN mappings, see "Configuring VLAN mapping." For more information about QoS, see *ACL and QoS Configuration Guide*.

Protocols and standards

- IEEE 802.1Q, *IEEE Standard for Local and Metropolitan Area Networks-Virtual Bridged Local Area Networks*
- IEEE 802.1ad, *IEEE Standard for Local and Metropolitan Area Networks-Virtual Bridged Local Area Networks-Amendment 4: Provider Bridges*

Restrictions and guidelines: QinQ configuration

When you configure QinQ, follow these restrictions and guidelines:

- The inner 802.1Q tag of QinQ frames is treated as part of the payload. As a best practice to ensure correct transmission of QinQ frames, set the MTU to a minimum of 1504 bytes for each port on their forwarding path. This value is the sum of the default Ethernet interface MTU (1500 bytes) and the length (4 bytes) of a VLAN tag.
- You can use a VLAN mapping and QinQ on a port for VLAN tag manipulation. If their settings conflict, the VLAN mapping takes precedence over QinQ.

QinQ and two-to-two mappings are mutually exclusive. The device does not support adding an SVLAN tag on a QinQ-enabled port and then modifying the CVLAN and SVLAN IDs.

For QinQ and QoS nest action to take effect, do not configure both features on the same port.

Enabling QinQ

About this task

Enable QinQ on customer-side ports of PEs. A QinQ-enabled port tags an incoming frame with its PVID.

Restrictions and guidelines

Before you enable or disable QinQ on a port, you must remove any VLAN mappings on the port. For more information about VLAN mapping, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.

   interface *interface-type interface-number*
3. Set the port link type.

   port link-type { **access** | **hybrid** | **trunk** }

   By default, the link type of a port is **access**.
4. Configure the port to allow packets from its PVID to pass through.
   - Assign the access port to the specified VLAN.

     port access vlan *vlan-id*

     By default, all access ports belong to VLAN 1.

     The PVID of an access port is the VLAN to which the port belongs. The port sends packets from the VLAN untagged.
   - Configure the hybrid port to send packets from its PVID untagged.

     port hybrid vlan *vlan-id-list* **untagged**

By default, the hybrid port is an untagged member of the VLAN to which the port belongs when its link type is access.

- Configure trunk port to allow packets from its PVID to pass through.

  port trunk permit vlan { *vlan-id-list* | **all** }

  By default, a trunk port allows packets only from VLAN 1 to pass through.

5. Enable QinQ on the port.

   qinq enable

   By default, QinQ is disabled on the port.

Configuring transmission for transparent VLANs

About this task

You can exclude a VLAN (for example, the management VLAN) from the QinQ tagging action on a customer-side port. This VLAN is called a transparent VLAN.

Restrictions and guidelines

- Make sure all ports on the traffic path permit the transparent VLAN to pass through.
- Do not configure any other VLAN manipulation actions for the transparent VLAN on the port.
- If you use both transparent VLANs and VLAN mappings on an interface, the transparent VLANs cannot be the following VLANs:
  - Original or translated VLANs of one-to-one, one-to-two, and many-to-one VLAN mappings.
  - Original or translated VLANs in two-to-one VLAN mappings.
  - Original or translated outer VLANs of two-to-two VLAN mappings.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.

   interface *interface-type interface-number*

3. Set the port link type.

   port link-type { **hybrid** | **trunk** }

   By default, the link type of a port is **access**.

4. Configure the port to allow packets from the transparent VLANs to pass through.
   - Configure the hybrid port to allow packets from the transparent VLANs to pass through.

     port hybrid vlan *vlan-id-list* { **tagged** | **untagged** }

     By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is **access**.
   - Configure the trunk port to allow packets from the transparent VLANs to pass through.

     port trunk permit vlan { *vlan-id-list* | **all** }

     By default, a trunk port allows packets only from VLAN 1 to pass through.

5. Specify transparent VLANs for the port.

   qinq transparent-vlan *vlan-id-list*

   By default, transparent transmission is not configured for any VLANs.

Configuring the TPID for VLAN tags

About TPID

TPID identifies a frame as an 802.1Q tagged frame. The TPID value varies by vendor. On an H3C device, the TPID in the 802.1Q tag added on a QinQ-enabled port is 0x8100 by default, in compliance with IEEE 802.1Q. In a multi-vendor network, make sure the TPID setting is the same between directly connected devices so 802.1Q tagged frames can be identified correctly.

TPID settings include CVLAN TPID and SVLAN TPID.

The TPID field is at the same position as the EtherType field in an untagged Ethernet frame. To ensure correct packet type identification, do not set the TPID value to any of the values listed in Table 1.

Table 1 Reserved EtherType values

Protocol type	Value
ARP	0x0806
PUP	0x0200
RARP	0x8035
IP	0x0800
IPv6	0x86dd
PPPoE	0x8863/0x8864
MPLS	0x8847/0x8848
IPX/SPX	0x8137
IS-IS	0x8000
LACP	0x8809
LLDP	0x88cc
802.1X	0x888e
802.1ag	0x8902
Cluster	0x88a7
Reserved	0xfffd/0xfffe/0xffff

Restrictions and guidelines

When a QinQ-enabled port adds VLAN tags to frames, the VLAN tags contain the TPID value specified in the **qinq ethernet-type service-tag** command.

Configuring the TPID for CVLAN tags

1. Enter system view.
   system-view
2. Set the TPID for CVLAN tags.
   qinq ethernet-type customer-tag *hex-value*
   By default, the TPID is 0x8100 for CVLAN tags.

Configuring the TPID for SVLAN tags

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   interface *interface-type interface-number*
3. Set the TPID for SVLAN tags.
   qinq ethernet-type service-tag *hex-value*
   By default, the TPID is 0x8100 for SVLAN tags.

Setting the 802.1p priority in SVLAN tags

About the 802.1p priority in SVLAN tags

By default, the 802.1p priority in the SVLAN tag added by a QinQ-enabled port depends on the priority trust mode on the port.

- If the 802.1p priority in frames is trusted, the device copies the 802.1p priority in the CVLAN tag to the SVLAN tag.
- If port priority is trusted, the port priority (0 by default) is used as the 802.1p priority in the SVLAN tag.

You can configure a QoS policy to modify the 802.1p priority in the SVLAN tag based on the 802.1p priority in the CVLAN tag or the CVLAN ID.

For more information about QoS policies and priority trust mode, see *ACL and QoS Configuration Guide*.

Prerequisites for setting the 802.1p priority in SVLAN tags

1. Enable QinQ. For more information, see "Enabling QinQ."
   To use the CVLAN ID or 802.1p priority of the CVLAN tag to set the 802.1p priority of the SVLAN tag, you must first enable QinQ on the port.
2. Use the **qos trust dot1p** command to configure the port to trust the 802.1p priority in incoming frames. For more information, see *ACL and QoS Configuration Guide*.
   This setting is required if the **remark dot1p** command is configured.

Tasks at a glance

To use QoS policies to set the 802.1p priority in SVLAN tags, perform the following tasks:

1. Creating a traffic class and configuring CVLAN match criteria
2. Creating a traffic behavior and configuring a priority marking action for SVLAN tags
3. Creating a QoS policy
4. Applying the QoS policy

Creating a traffic class and configuring CVLAN match criteria

1. Enter system view.
   system-view

2. Create a traffic class and enter its view.
   traffic classifier *classifier-name* [**operator** { **and** | **or** }]
3. Configure CVLAN match criteria.
   Choose one option as needed:
   - Match CVLAN IDs.
     if-match customer-vlan-id *vlan-id-list*
   - Match 802.1p priority.
     if-match customer-dot1p *dot1p-value*&<1-8>

Creating a traffic behavior and configuring a priority marking action for SVLAN tags

1. Enter system view.
   system-view
2. Create a traffic behavior and enter its view.
   traffic behavior *behavior-name*
3. Configure a priority marking action for SVLAN tags.
   remark dot1p *dot1p-value*

Creating a QoS policy

1. Enter system view.
   system-view
2. Create a QoS policy and enter its view.
   qos policy *policy-name*
3. Specify the traffic behavior for the traffic class in the QoS policy.
   classifier *classifier-name* **behavior** *behavior-name*

Applying the QoS policy

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
3. Apply the QoS policy to the inbound direction of the port.
   qos apply policy *policy-name* **inbound**

Display and maintenance commands for QinQ

Execute **display** commands in any view.

Task	Command
Display QinQ-enabled ports.	**display qinq** [**interface** *interface-type interface-number*]

QinQ configuration examples

Example: Configuring VLAN transparent transmission

Network configuration

As shown in Figure 3:

- The service provider assigns VLAN 100 to a company's VLANs 10 through 50.
- VLAN 3000 is the dedicated VLAN of the company on the service provider network.

Configure QinQ on PE 1 and PE 2 to provide Layer 2 connectivity for CVLANs 10 through 50 over the service provider network.

Configure VLAN transparent transmission for VLAN 3000 on PE 1 and PE 2 to enable the hosts in VLAN 3000 to communicate without using an SVLAN.

Figure 3 Network diagram

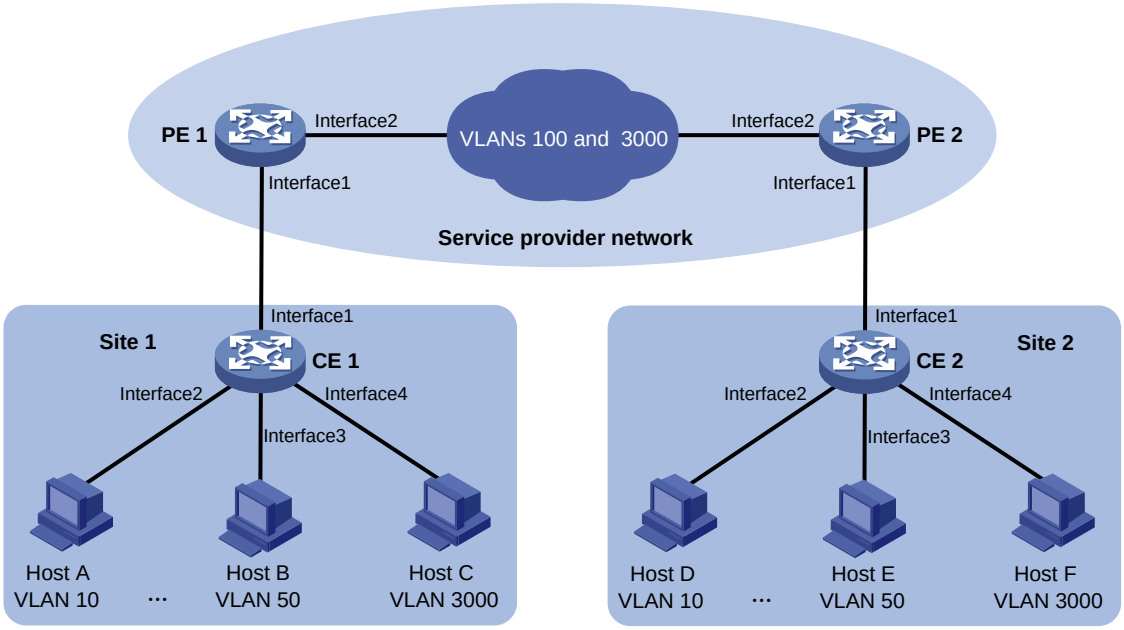


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure PE 1:

   # Create VLANs 10 through 50, VLAN 100, and VLAN 3000.

   ```
   <PE1> system-view
   [PE1] vlan 10 to 50
   [PE1] vlan 100
   ```

```
[PE1-vlan100] quit
[PE1] vlan 3000
[PE1-vlan3000] quit
```

Configure GigabitEthernet1/0/1 as a trunk port, and assign it to VLAN 100, VLAN 3000, and VLANs 10 through 50.

```
[PE1] interface gigabitethernet 1/0/1
[PE1-GigabitEthernet1/0/1] port link-type trunk
[PE1-GigabitEthernet1/0/1] port trunk permit vlan 100 3000 10 to 50
```

Set the PVID of GigabitEthernet1/0/1 to VLAN 100.

```
[PE1-GigabitEthernet1/0/1] port trunk pvid vlan 100
```

Enable QinQ on GigabitEthernet1/0/1.

```
[PE1-GigabitEthernet1/0/1] qinq enable
```

Enable transparent transmission for VLAN 3000 on GigabitEthernet1/0/1.

```
[PE1-GigabitEthernet1/0/1] qinq transparent-vlan 3000
[PE1-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet1/0/2 as a trunk port, and assign it to VLANs 100 and 3000.

```
[PE1] interface gigabitethernet 1/0/2
[PE1-GigabitEthernet1/0/2] port link-type trunk
[PE1-GigabitEthernet1/0/2] port trunk permit vlan 100 3000
[PE1-GigabitEthernet1/0/2] quit
```

2. Configure PE 2:

Create VLANs 10 through 50, VLAN 100, and VLAN 3000.

```
<PE2> system-view
[PE2] vlan 10 to 50
[PE2] vlan 100
[PE2-vlan100] quit
[PE2] vlan 3000
[PE2-vlan3000] quit
```

Configure GigabitEthernet1/0/1 as a trunk port, and assign it to VLAN 100, VLANs 10 through 50, and VLAN 3000.

```
[PE2] interface gigabitethernet 1/0/1
[PE2-GigabitEthernet1/0/1] port link-type trunk
[PE2-GigabitEthernet1/0/1] port trunk permit vlan 100 3000 10 to 50
```

Set the PVID of GigabitEthernet1/0/1 to VLAN 100.

```
[PE2-GigabitEthernet1/0/1] port trunk pvid vlan 100
```

Enable QinQ on GigabitEthernet1/0/1.

```
[PE2-GigabitEthernet1/0/1] qinq enable
```

Enable transparent transmission for VLAN 3000 on GigabitEthernet1/0/1.

```
[PE2-GigabitEthernet1/0/1] qinq transparent-vlan 3000
[PE2-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet1/0/2 as a trunk port, and assign it to VLANs 100 and 3000.

```
[PE2] interface gigabitethernet 1/0/2
[PE2-GigabitEthernet1/0/2] port link-type trunk
[PE2-GigabitEthernet1/0/2] port trunk permit vlan 100 3000
[PE2-GigabitEthernet1/0/2] quit
```

3. Configure CE 1:

Create VLANs 10 through 50.

```
<CE1> system-view
[CE1] vlan 10 to 50
```

Configure GigabitEthernet1/0/2 as an access port, and assign it to VLAN 10.

```
[CE1] interface gigabitethernet 1/0/2
[CE1-GigabitEthernet1/0/2] port link-type access
[CE1-GigabitEthernet1/0/2] port access vlan 10
[CE1-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet1/0/3 as an access port, and assign it to VLAN 50.

```
[CE1] interface gigabitethernet 1/0/3
[CE1-GigabitEthernet1/0/3] port link-type access
[CE1-GigabitEthernet1/0/3] port access vlan 50
[CE1-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet1/0/4 as an access port, and assign it to VLAN 3000.

```
[CE1] interface gigabitethernet 1/0/4
[CE1-GigabitEthernet1/0/4] port link-type access
[CE1-GigabitEthernet1/0/4] port access vlan 3000
[CE1-GigabitEthernet1/0/4] quit
```

Configure GigabitEthernet1/0/1 as a trunk port, and assign it to VLANs 10 through 50 and VLAN 3000.

```
[CE1] interface gigabitethernet 1/0/1
[CE1-GigabitEthernet1/0/1] port link-type trunk
[CE1-GigabitEthernet1/0/1] port trunk permit vlan 3000 10 to 50
[CE1-GigabitEthernet1/0/1] quit
```

4. Configure CE 2:

Create VLANs 10 through 50.

```
<CE2> system-view
[CE2] vlan 10 to 50
```

Configure GigabitEthernet1/0/2 as an access port, and assign it to VLAN 10.

```
[CE2] interface gigabitethernet 1/0/2
[CE2-GigabitEthernet1/0/2] port link-type access
[CE2-GigabitEthernet1/0/2] port access vlan 10
[CE2-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet1/0/3 as an access port, and assign it to VLAN 50.

```
[CE2] interface gigabitethernet 1/0/3
[CE2-GigabitEthernet1/0/3] port link-type access
[CE2-GigabitEthernet1/0/3] port access vlan 50
[CE2-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet1/0/4 as an access port, and assign it to VLAN 3000.

```
[CE2] interface gigabitethernet 1/0/4
[CE2-GigabitEthernet1/0/4] port link-type access
[CE2-GigabitEthernet1/0/4] port access vlan 3000
[CE2-GigabitEthernet1/0/4] quit
```

Configure GigabitEthernet1/0/1 as a trunk port, and assign it to VLANs 10 through 50 and VLAN 3000.

```
[CE2] interface gigabitethernet 1/0/1
[CE2-GigabitEthernet1/0/1] port link-type trunk
[CE2-GigabitEthernet1/0/1] port trunk permit vlan 3000 10 to 50
[CE2-GigabitEthernet1/0/1] quit
```

5. Configure the devices between PE 1 and PE 2:

   # Set the MTU to a minimum of 1504 bytes for each port on the path of QinQ frames. (Details not shown.)

   # Configure all ports on the forwarding path to allow frames from VLANs 100 and 3000 to pass through without removing the VLAN tag. (Details not shown.)

6. Configure Host A through Host F:

   # Configure the IP addresses of the hosts in the same VLAN to be on the same subnet, for example, configure the IP addresses of Host A and Host D to be on subnet 192.168.1.0/24.

Verifying the configuration

- Execute the **ping** command to verify the following information:

  At Site 1 and Site 2, hosts that belong to the same VLAN can ping each other successfully and learn the MAC address of each other. For example, Host A and Host D can communicate with each other.

- Execute the **display mac-address** command (Host A and Host D, and Host C and Host F ping each other).

  # Display MAC address entries on PE 1.

```
<PE1> display mac-address
MAC Address      VLAN ID    State           Port/Nickname           Aging
7e43-5b0a-0502   100        Learned         GE1/0/1                 Y
7e42-da03-0200   3000       Learned         GE1/0/1                 Y
7e43-5d90-0602   100        Learned         GE1/0/2                 Y
7e42-df16-0400   3000       Learned         GE1/0/2                 Y
```

  The output shows that PE 1 has learned the MAC addresses of Host A and Host C through GigabitEthernet1/0/1 and those of Host D and Host F through GigabitEthernet1/0/2, respectively. The MAC addresses of Host A and Host D belong to VLAN 100 and those of Host C and Host F belong to VLAN 3000.

  # Display MAC address entries on CE 1.

```
<CE1>display mac-address
MAC Address      VLAN ID    State           Port/Nickname           Aging
7e43-5d90-0602   10         Learned         GE1/0/1                 Y
7e42-df16-0400   3000       Learned         GE1/0/1                 Y
7e43-5b0a-0502   10         Learned         GE1/0/2                 Y
7e42-da03-0200   3000       Learned         GE1/0/4                 Y
```

  The output shows that CE 1 has learned the MAC addresses of Host D and Host F through GigabitEthernet1/0/1, the MAC address of Host A through GigabitEthernet1/0/2, and the MAC address of Host C through GigabitEthernet1/0/4. The MAC addresses of Host A and Host D belong to VLAN 10 and those of Host C and Host F belong to VLAN 3000.

Configuration files

PE 1

```
#
vlan 10 to 50
#
vlan 100
#
vlan 3000
#
interface GigabitEthernet1/0/1
 port link-type trunk
```

```
 port trunk permit vlan 10 to 50 100 3000
 port trunk pvid vlan 100
 qinq enable
 qinq transparent-vlan 3000
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100 3000
#
```

PE 2

```
#
vlan 10 to 50
#
vlan 100
#
vlan 3000
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 10 to 50 100 3000
 port trunk pvid vlan 100
 qinq enable
 qinq transparent-vlan 3000
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100 3000
#
```

CE 1

```
#
vlan 10 to 50
#
vlan 3000
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 10 to 50 3000
#
interface GigabitEthernet1/0/2
 port access vlan 10
#
interface GigabitEthernet1/0/3
 port access vlan 50
#
interface GigabitEthernet1/0/4
 port access vlan 3000
#
```

CE 2

```
#
vlan 10 to 50
#
vlan 3000
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 10 to 50 3000
#
interface GigabitEthernet1/0/2
 port access vlan 10
#
interface GigabitEthernet1/0/3
 port access vlan 50
#
interface GigabitEthernet1/0/4
 port access vlan 3000
#
```

Example: Configuring basic QinQ

Network configuration

As shown in Figure 4:

- The service provider assigns VLAN 100 to Company A's VLANs 10 through 70.
- The service provider assigns VLAN 200 to Company B's VLANs 30 through 90.
- The devices between PE 1 and PE 2 in the service provider network use a TPID value of 0x8200.

Configure QinQ on PE 1 and PE 2 to transmit traffic in VLANs 100 and 200 for Company A and Company B, respectively.

For the QinQ frames to be identified correctly, set the SVLAN TPID to 0x8200 on the service provider-side ports of PE 1 and PE 2.

Figure 4 Network diagram

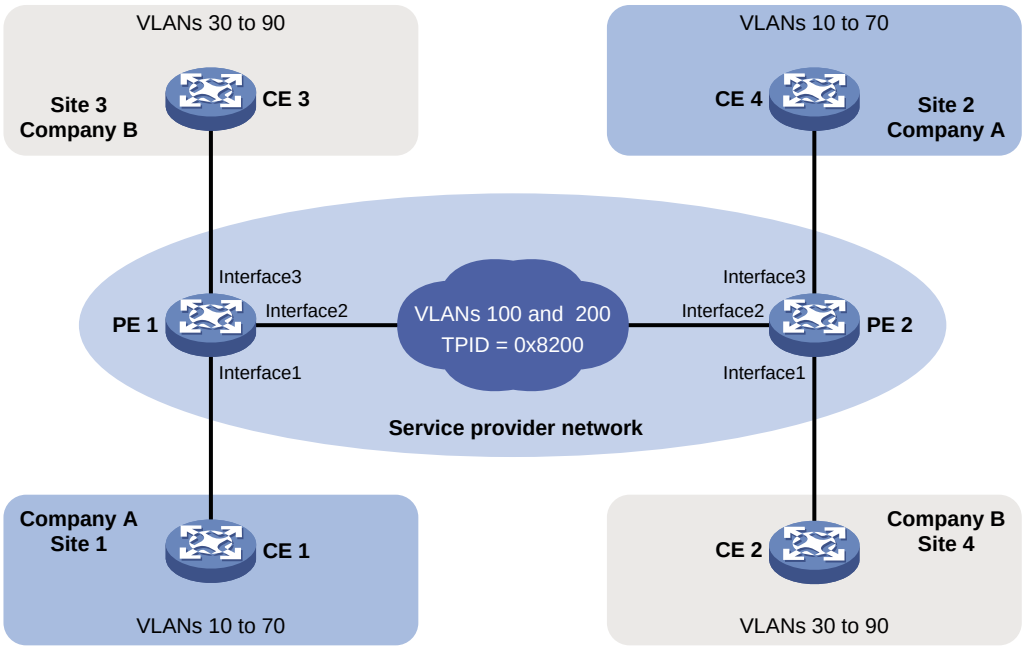


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure PE 1:

Configure GigabitEthernet 1/0/1 as a trunk port, and assign it to VLAN 100.

```
<PE1> system-view
[PE1] interface gigabitethernet 1/0/1
[PE1-GigabitEthernet1/0/1] port link-type trunk
[PE1-GigabitEthernet1/0/1] port trunk permit vlan 100
```

Set the PVID of GigabitEthernet 1/0/1 to VLAN 100.

```
[PE1-GigabitEthernet1/0/1] port trunk pvid vlan 100
```

Enable QinQ on GigabitEthernet 1/0/1.

```
[PE1-GigabitEthernet1/0/1] qinq enable
[PE1-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port, and assign it to VLANs 100 and 200.

```
[PE1] interface gigabitethernet 1/0/2
[PE1-GigabitEthernet1/0/2] port link-type trunk
[PE1-GigabitEthernet1/0/2] port trunk permit vlan 100 200
```

Set the TPID value in the SVLAN tags to 0x8200 on GigabitEthernet 1/0/2.

```
[PE1-GigabitEthernet1/0/2] qinq ethernet-type service-tag 8200
[PE1-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port, and assign it to VLAN 200.

```
[PE1] interface gigabitethernet 1/0/3
[PE1-GigabitEthernet1/0/3] port link-type trunk
[PE1-GigabitEthernet1/0/3] port trunk permit vlan 200
```

Set the PVID of GigabitEthernet 1/0/3 to VLAN 200.

```
[PE1-GigabitEthernet1/0/3] port trunk pvid vlan 200
```

Enable QinQ on GigabitEthernet 1/0/3.

```
[PE1-GigabitEthernet1/0/3] qinq enable
[PE1-GigabitEthernet1/0/3] quit
```

2. Configure PE 2:

Configure GigabitEthernet 1/0/1 as a trunk port, and assign it to VLAN 200.

```
<PE2> system-view
[PE2] interface gigabitethernet 1/0/1
[PE2-GigabitEthernet1/0/1] port link-type trunk
[PE2-GigabitEthernet1/0/1] port trunk permit vlan 200
```

Set the PVID of GigabitEthernet 1/0/1 to VLAN 200.

```
[PE2-GigabitEthernet1/0/1] port trunk pvid vlan 200
```

Enable QinQ on GigabitEthernet 1/0/1.

```
[PE2-GigabitEthernet1/0/1] qinq enable
[PE2-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port, and assign it to VLANs 100 and 200.

```
[PE2] interface gigabitethernet 1/0/2
[PE2-GigabitEthernet1/0/2] port link-type trunk
[PE2-GigabitEthernet1/0/2] port trunk permit vlan 100 200
```

Set the TPID value in the SVLAN tags to 0x8200 on GigabitEthernet 1/0/2.

```
[PE2-GigabitEthernet1/0/2] qinq ethernet-type service-tag 8200
[PE2-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port, and assign it to VLAN 100.

```
[PE2] interface gigabitethernet 1/0/3
[PE2-GigabitEthernet1/0/3] port link-type trunk
[PE2-GigabitEthernet1/0/3] port trunk permit vlan 100
```

Set the PVID of GigabitEthernet 1/0/3 to VLAN 100.

```
[PE2-GigabitEthernet1/0/3] port trunk pvid vlan 100
```

Enable QinQ on GigabitEthernet 1/0/3.

```
[PE2-GigabitEthernet1/0/3] qinq enable
[PE2-GigabitEthernet1/0/3] quit
```

3. Configure the devices between PE 1 and PE 2:

Set the MTU to a minimum of 1504 bytes for each port on the path of QinQ frames. (Details not shown.)

Configure all ports on the forwarding path to allow frames from VLANs 100 and 200 to pass through without removing the VLAN tag. (Details not shown.)

Contents

Configuring VLAN mapping 1

About VLAN mapping 1
VLAN mapping types 1
VLAN mapping application scenarios 1
VLAN mapping implementations 4
Restrictions and guidelines: VLAN mapping configuration 7
VLAN mapping tasks at a glance 7
Prerequisites 8
Configuring one-to-one VLAN mapping 8
Configuring many-to-one VLAN mapping 9
Configuring one-to-two VLAN mapping 10
Configuring zero-to-two VLAN mapping 11
Configuring two-to-one VLAN mapping 12
Configuring two-to-two VLAN mapping 13
Display and maintenance commands for VLAN mapping 14
VLAN mapping configuration examples 14
Example: Configuring one-to-one VLAN mapping 14
Example: Configuring many-to-one VLAN mapping 17
Example: Configuring one-to-two and two-to-two VLAN mapping 19

Configuring VLAN mapping

About VLAN mapping

VLAN mapping re-marks VLAN traffic with new VLAN IDs.

VLAN mapping types

H3C provides the following types of VLAN mapping:

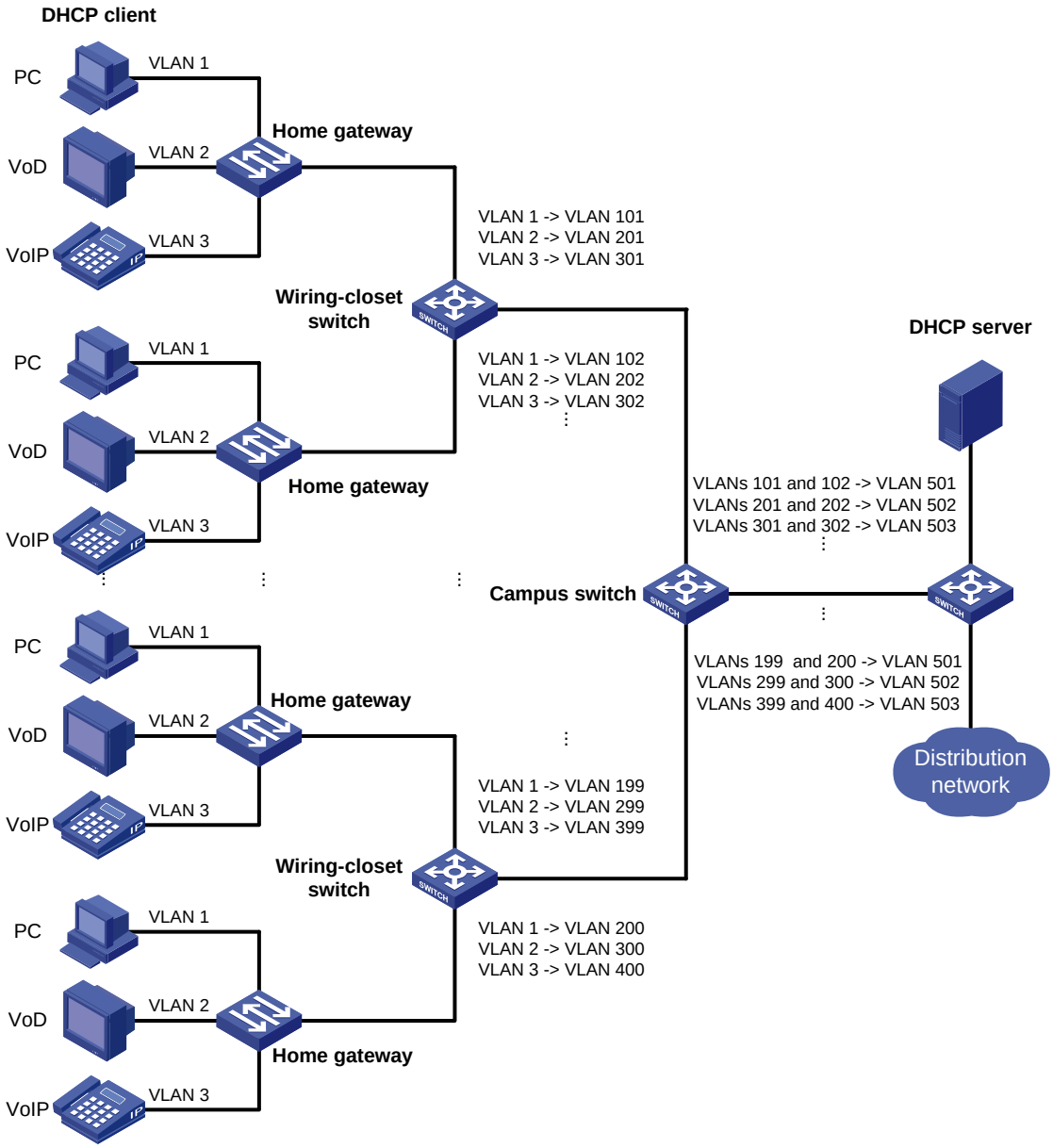
- **One-to-one VLAN mapping**—Replaces one VLAN tag with another.
- **Many-to-one VLAN mapping**—Replaces multiple VLAN tags with the same VLAN tag.
- **One-to-two VLAN mapping**—Tags single-tagged packets with an outer VLAN tag.
- **Zero-to-two VLAN mapping**—Adds double VLAN tags to untagged packets.
- **Two-to-one VLAN mapping**—Removes VLAN tags from double-tagged packets and adds a new VLAN tag to them.
- **Two-to-two VLAN mapping**—Replaces the outer and inner VLAN IDs of double tagged traffic with a new pair of VLAN IDs.

VLAN mapping application scenarios

One-to-one and many-to-one VLAN mapping

One-to-one and many-to-one VLAN mapping are typically used by a community for broadband Internet access, as shown in Figure 1.

Figure 1 Application scenario of one-to-one and many-to-one VLAN mapping



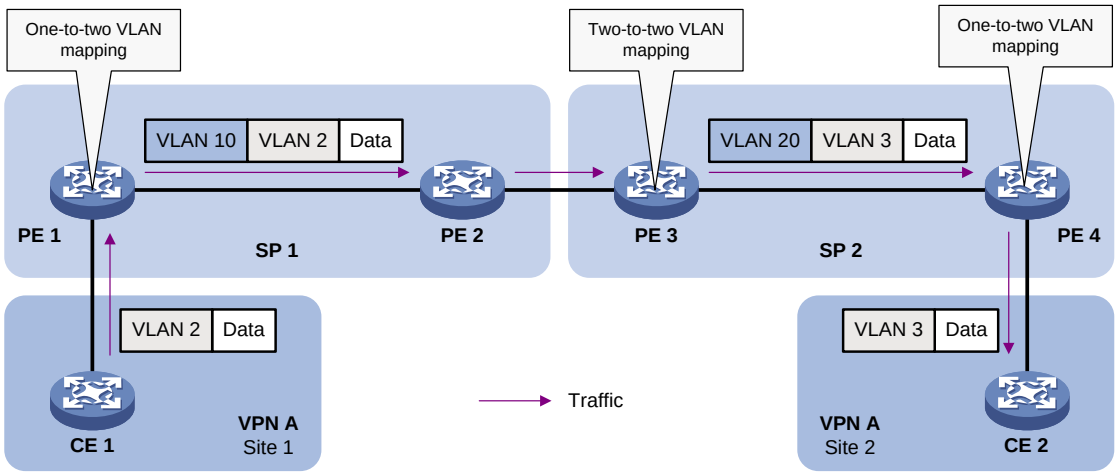
As shown in Figure 1, the network is implemented as follows:

- Each home gateway uses different VLANs to transmit the PC, VoD, and VoIP services.
- To further subclassify each type of traffic by customer, configure one-to-one VLAN mapping on the wiring-closet switches. This feature assigns a separate VLAN to each type of traffic from each customer. The required total number of VLANs in the network can be very large.
- To prevent the maximum number of VLANs from being exceeded on the distribution layer device, configure many-to-one VLAN mapping on the campus switch. This feature assigns the same VLAN to the same type of traffic from different customers.

One-to-two and two-to-two VLAN mapping

One-to-two and two-to-two VLAN mapping are typically used to implement communication across different SP networks, as shown in Figure 2.

Figure 2 Application scenario of one-to-two and two-to-two VLAN mapping



As shown in Figure 2:

Site 1 and Site 2 of VPN A are in VLAN 2 and VLAN 3, respectively. The SP 1 network assigns SVLAN 10 to Site 1. The SP 2 network assigns SVLAN 20 to Site 2. When the packet from Site 1 arrives at PE 1, PE 1 tags the packet with SVLAN 10 by using one-to-two VLAN mapping.

One-to-two VLAN mapping provides the following benefits:

- Enables a customer network to plan its CVLAN assignment without conflicting with SVLANs.
- Adds a VLAN tag to a tagged packet and expands the number of available VLANs to 4094 × 4094.
- Reduces the stress on the SVLAN resources, which were 4094 VLANs in the SP network before the mapping process was initiated.

When the double-tagged packet from the SP 1 network arrives at the SP 2 network interface, PE 3 processes the packet as follows:

- Replaces SVLAN tag 10 with SVLAN tag 20.
- Replaces CVLAN tag 2 with CVLAN tag 3.

Zero-to-two VLAN mapping

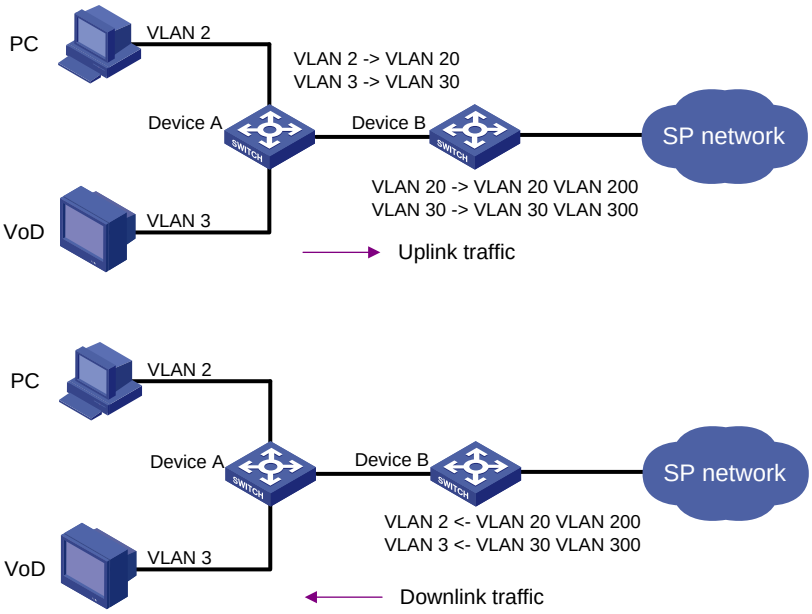
The application scenario of zero-to-two VLAN mapping is similar to the scenario of one-to-two VLAN mapping. Zero-to-two VLAN mapping is used on the customer-side port of a PE to add double tags to untagged packets.

Two-to-one VLAN mapping

As shown in Figure 3, configure VLANs and VLAN mappings to isolate traffic of different services and ensure the communication between the user network and the service provider network:

- On Device A, assign different types of service traffic to different VLANs.
- On Device B, configure one-to-one VLAN mappings and one-to-two VLAN mappings for uplink traffic.
- On Device B, configure two-to-one VLAN mappings for the downlink traffic. After receiving double-tagged reply packets, Device B removes their double VLAN tags and adds the original VLAN tags to them.

Figure 3 Application scenario of two-to-one VLAN mapping



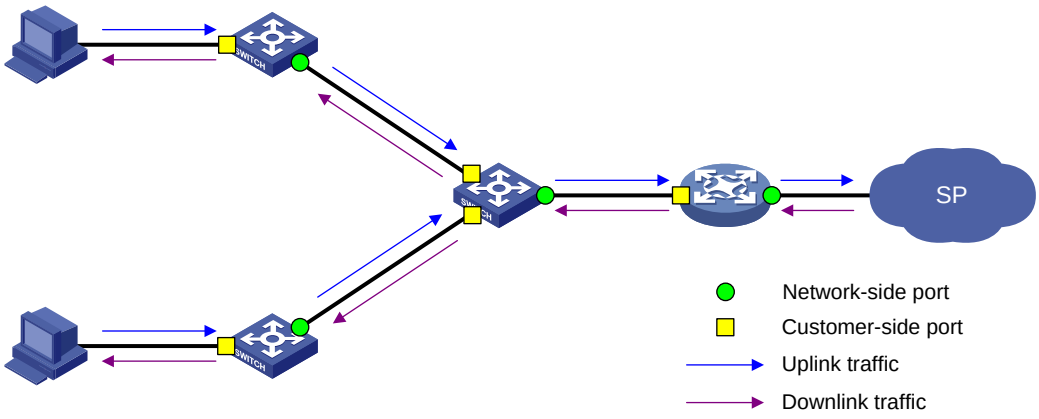
VLAN mapping implementations

Figure 4 shows a simplified network that illustrates basic VLAN mapping terms.

Basic VLAN mapping terms include the following:

- **Uplink traffic**—Traffic transmitted from the customer network to the service provider network.
- **Downlink traffic**—Traffic transmitted from the service provider network to the customer network.
- **Network-side port**—A port connected to or closer to the service provider network.
- **Customer-side port**—A port connected to or closer to the customer network.

Figure 4 Basic VLAN mapping terms

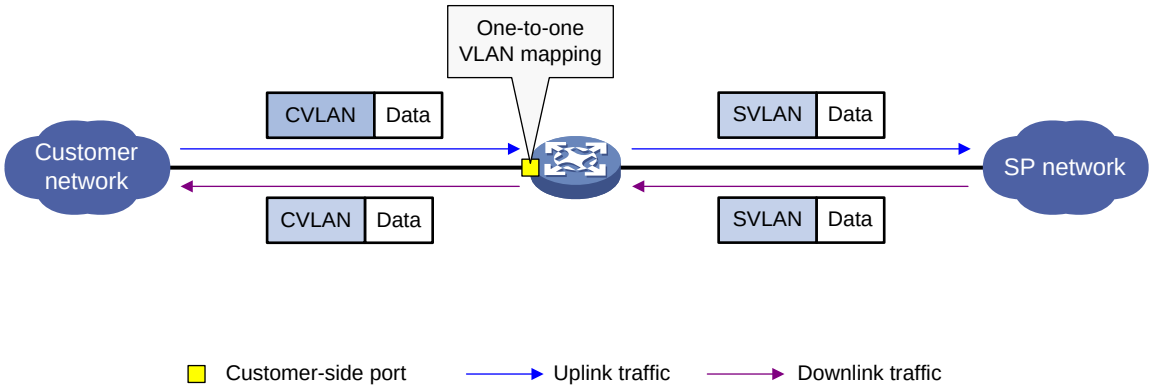


One-to-one VLAN mapping

As shown in Figure 5, one-to-one VLAN mapping is implemented on the customer-side port and replaces VLAN tags as follows:

- Replaces the CVLAN with the SVLAN for the uplink traffic.
- Replaces the SVLAN with the CVLAN for the downlink traffic.

Figure 5 One-to-one VLAN mapping implementation



Many-to-one VLAN mapping

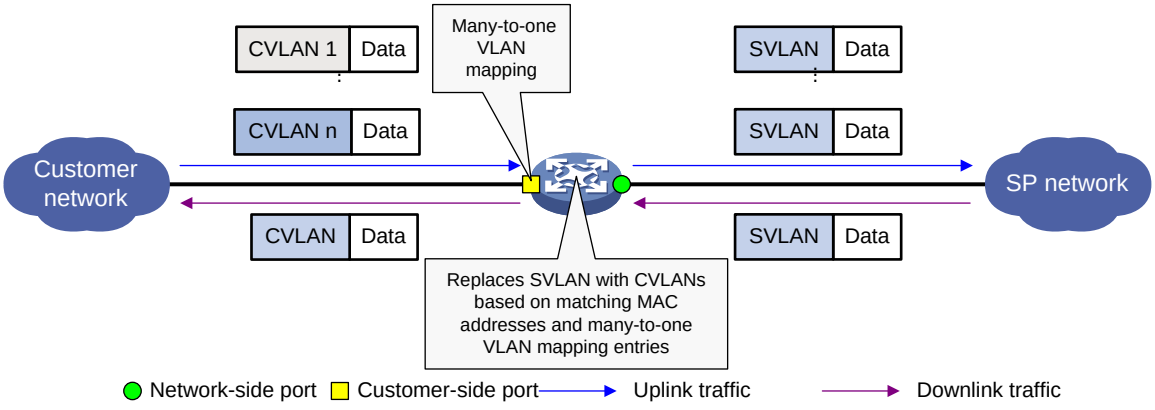
As shown in Figure 6, many-to-one VLAN mapping is implemented on both the customer-side and network-side ports as follows:

- For the uplink traffic, the customer-side many-to-one VLAN mapping replaces multiple CVLANs with the same SVLAN.
- For the downlink traffic, the device performs the following operations:
  a. Searches the MAC address table for an entry that matches the destination MAC address of the downlink packets.
  b. Replaces the SVLAN tag with a CVLAN tag based on the matching many-to-one mapping entry.

  The maximum number of the matching many-to-one mapping entries supported on the device is 170.

For more information about the MAC address table, see "Configuring the MAC address table."

Figure 6 Many-to-one VLAN mapping implementation (IPv4)



One-to-two VLAN mapping

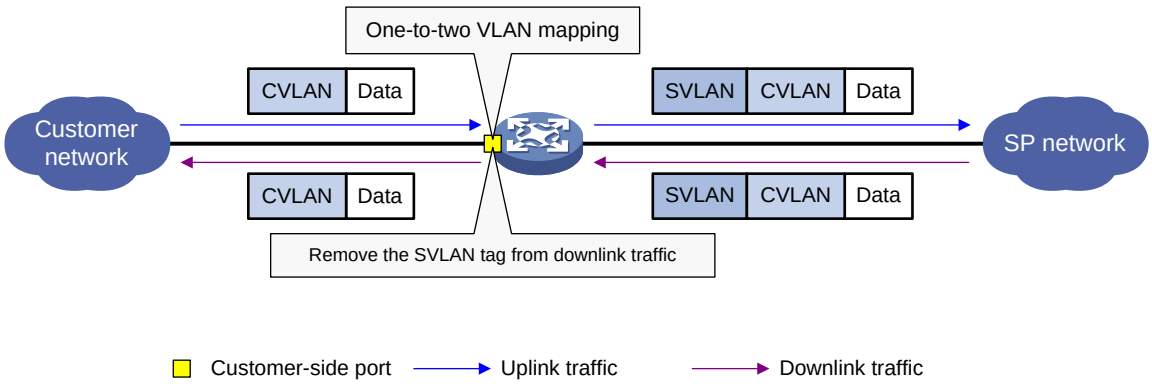
As shown in Figure 7, one-to-two VLAN mapping is implemented on the customer-side port to add the SVLAN tag for the uplink traffic.

For the downlink traffic to be correctly sent to the customer network, make sure the SVLAN tag is removed on the customer-side port before transmission. Use one of the following methods to remove the SVLAN tag from the downlink traffic:

- Configure the customer-side port as a hybrid port and assign the port to the SVLAN as an untagged member.

- Configure the customer-side port as a trunk port and set the port PVID to the SVLAN.

Figure 7 One-to-two VLAN mapping implementation



Zero-to-two VLAN mapping

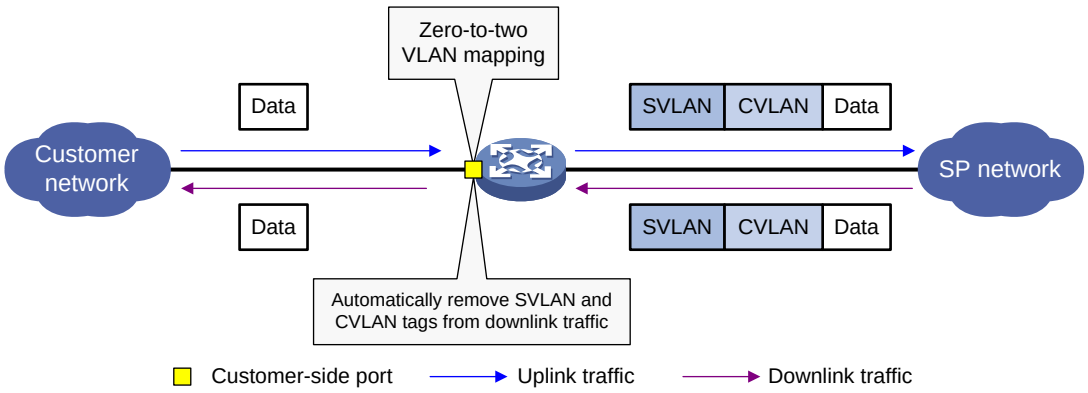
As shown in Figure 8, zero-to-two VLAN mapping is implemented on the customer-side port to add double tags to untagged uplink traffic. For zero-to-two VLAN mapping to take effect, the port PVID must be VLAN 1.

For correct downlink traffic transmission, the downlink traffic must be double-tagged. Then, the customer-side port removes both SVLAN and CVLAN tags from the traffic.

Use one of the following methods to ensure that the downlink traffic contains double tags on the customer-side port:

- Configure the port as a trunk port and assign it to the SVLAN, which must be different from the port PVID (VLAN 1).
- Configure the port as a hybrid port and assign it to the SVLAN as a tagged member.

Figure 8 Zero-to-two VLAN mapping implementation

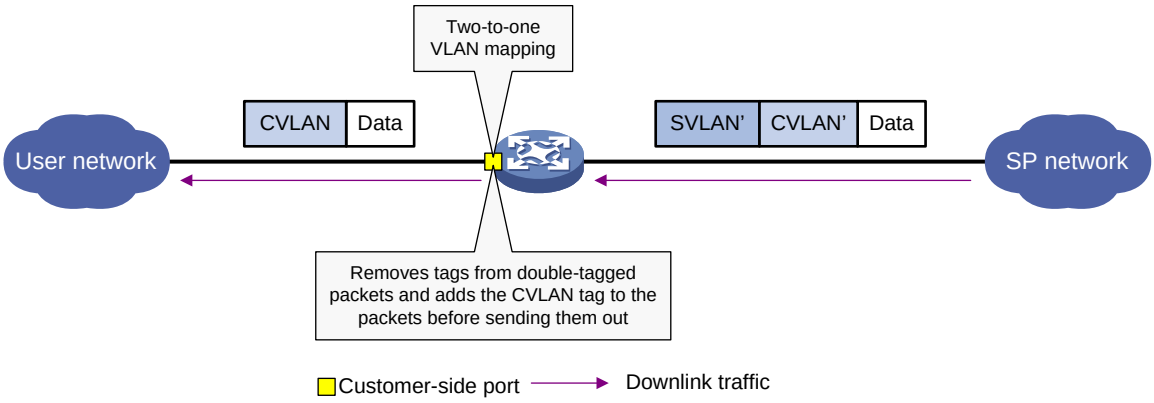


Two-to-one VLAN mapping

As shown in Figure 9, configure two-to-one VLAN mapping on the customer-side port to remove double VLAN tags from downstream packets and add the CVLAN tag to them.

Two-to-one VLAN mapping takes effect only on the outgoing downstream packets of the customer-side port and does not affect the incoming packets of the port.

Figure 9 Two-to-one VLAN mapping implementation

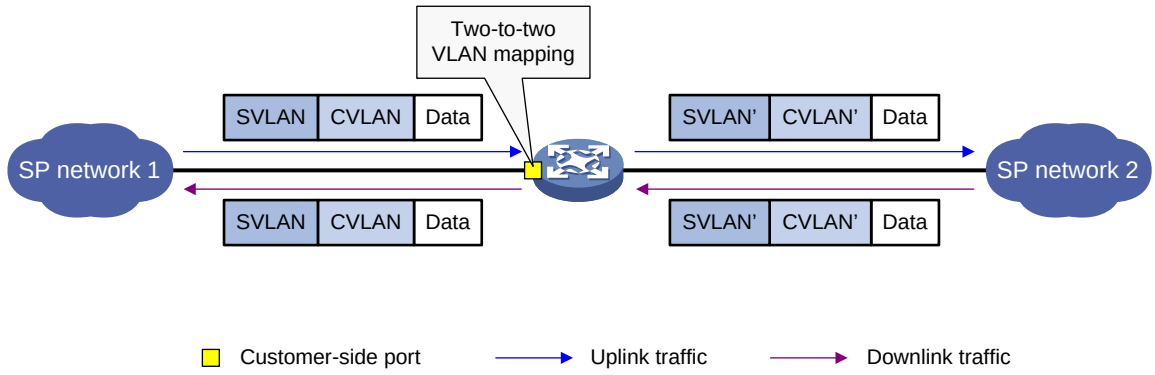


Two-to-two VLAN mapping

As shown in Figure 10, two-to-two VLAN mapping is implemented on the customer-side port and replaces VLAN tags as follows:

- Replaces the CVLAN and the SVLAN with the CVLAN' and the SVLAN' for the uplink traffic.
- Replaces the SVLAN' and CVLAN' with the SVLAN and the CVLAN for the downlink traffic.

Figure 10 Two-to-two VLAN mapping implementation



Restrictions and guidelines: VLAN mapping configuration

You cannot use VLAN mapping and Layer 3 forwarding together.

To add VLAN tags to packets, you can configure both VLAN mapping and QinQ. VLAN mapping takes effect if a configuration conflict occurs. For more information about QinQ, see "Configuring QinQ."

To add or replace VLAN tags for packets, you can configure both VLAN mapping and a QoS policy. The QoS policy takes effect if a configuration conflict occurs. For information about QoS policies, see *ACL and QoS Configuration Guide*.

VLAN mapping tasks at a glance

Use the appropriate VLAN mapping methods for the devices in the network.

To configure VLAN mapping, perform the following tasks:

- Configuring one-to-one VLAN mapping
- Configure one-to-one VLAN mapping on the wiring-closet switch, as shown in the network diagram in "One-to-one and many-to-one VLAN mapping."
- Configuring many-to-one VLAN mapping

  Configure many-to-one VLAN mapping on the campus switch, as shown in the network diagram in "One-to-one and many-to-one VLAN mapping."
- Configuring one-to-two VLAN mapping

  Configure one-to-two VLAN mapping on PE 1 and PE 4, as shown in the network diagram in "One-to-two and two-to-two VLAN mapping," through which traffic from customer networks enters the service provider networks.
- Configuring zero-to-two VLAN mapping

  Configure zero-to-two VLAN mapping on PE 1 and PE 4, as shown in the network diagram in "One-to-two and two-to-two VLAN mapping," through which traffic from customer networks enters the service provider networks.
- Configuring two-to-one VLAN mapping

  Configure two-to-one VLAN mapping on the customer-side port of Device B, as shown in the network diagram in "Two-to-one VLAN mapping."
- Configuring two-to-two VLAN mapping

  Configure two-to-two VLAN mapping on PE 3, as shown in the network diagram in "One-to-two and two-to-two VLAN mapping," which is an edge device of the SP 2 network.

Prerequisites

Before you configure VLAN mapping, create original and translated VLANs.

Configuring one-to-one VLAN mapping

About this task

Configure one-to-one VLAN mapping on the customer-side ports of wiring-closet switches (see the network diagram in "One-to-one and many-to-one VLAN mapping") to isolate traffic of the same service type from different homes.

Restrictions and guidelines

You cannot use one-to-one VLAN mapping together with zero-to-two, two-to-one, or two-to-two VLAN mapping.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
3. Set the link type of the port.

   port link-type { **hybrid** | **trunk** }

   By default, the link type of a port is **access**.
4. Assign the port to the original VLAN and the translated VLAN.

- Assign the trunk port to the original VLAN and the translated VLAN.

  port trunk permit vlan *vlan-id-list*

  By default, a trunk port is assigned to VLAN 1.
- Assign the hybrid port to the original VLAN and the translated VLAN as a tagged member.

  port hybrid vlan *vlan-id-list* **tagged**

  By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is access.

5. Configure a one-to-one VLAN mapping.

   vlan mapping *vlan-id* **translated-vlan** *vlan-id*

   By default, no VLAN mapping is configured on an interface.

Configuring many-to-one VLAN mapping

About this task

Configure many-to-one VLAN mapping on campus switches (see the network diagram in "One-to-one and many-to-one VLAN mapping") to transmit the same type of traffic from different users in one VLAN.

Restrictions and guidelines for many-to-one VLAN mapping

- The device supports a maximum of 170 many-to-one mapping entries.
- When you configure both many-to-one VLAN mapping and other types of mappings on an interface, only the many-to-one VLAN mapping takes effect.
- You cannot use many-to-one VLAN mapping together with zero-to-two, two-to-one, or two-to-two VLAN mapping.
- After you configure an interface as the network-side interface of many-to-one VLAN mapping, do not configure the other types of VLAN mapping on the interface
- You can configure only one many-to-one VLAN mapping on the same interface. To modify the many-to-one VLAN mapping, perform the following tasks:
  a. Execute the **undo vlan mapping** command to delete the existing many-to-one VLAN mapping.
  b. Configure another many-to-one VLAN mapping.
- To ensure correct traffic forwarding from the service provider network to the customer network, do not configure many-to-one VLAN mapping together with the following features:
  - uRPF.
  - Disabling MAC address learning.
  - Setting the MAC learning limit.

  For more information about uRPF, see *Security Configuration Guide*. For more information about MAC address learning, see "Configuring the MAC address table."

Many-to-one VLAN mapping tasks at a glance

To configure many-to-one VLAN mapping in dynamic IP address assignment environment, perform the following tasks:

1. Configuring the customer-side port
2. Configuring the network-side port

Configuring the customer-side port

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
3. Set the link type of the port.
   port link-type { **hybrid** | **trunk** }
   By default, the link type of a port is **access**.
4. Assign the port to the original VLANs.
   - Assign the trunk port to the original VLANs
     port trunk permit vlan *vlan-id-list*
     By default, a trunk port is assigned to VLAN 1.
   - Assign the hybrid port to the original VLANs as a tagged member.
     port hybrid vlan *vlan-id-list* **tagged**
     By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is access.
5. Configure a many-to-one VLAN mapping.
   vlan mapping uni { **range** *vlan-range-list* | **single** *vlan-id-list* } **translated-vlan** *vlan-id*
   By default, no VLAN mapping is configured on an interface.

Configuring the network-side port

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
3. Set the link type of the port.
   port link-type { **hybrid** | **trunk** }
   By default, the link type of a port is **access**.
4. Assign the port to the translated VLAN.
   - Assign the trunk port to the translated VLAN.
     port trunk permit vlan *vlan-id-list*
     By default, a trunk port is assigned to VLAN 1.
   - Assign the hybrid port to the translated VLAN as a tagged member.
     port hybrid vlan *vlan-id-list* **tagged**
     By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is access.

Configuring one-to-two VLAN mapping

About this task

Configure one-to-two VLAN mapping on the customer-side ports of edge devices from which customer traffic enters SP networks, for example, on PEs 1 and 4 in the network diagram in "One-to-two and two-to-two VLAN mapping." One-to-two VLAN mapping enables the edge devices to add an SVLAN tag to each incoming packet.

Restrictions and guidelines

Only one SVLAN tag can be added to packets from the same CVLAN. To add different SVLAN tags to different CVLAN packets on a port, set the port link type to hybrid and configure multiple one-to-two VLAN mappings.

The MTU of an interface is 1500 bytes by default. After a VLAN tag is added to a packet, the packet length is added by 4 bytes. As a best practice, set the MTU to a minimum of 1504 bytes for ports on the forwarding path of the packet in the service provider network.

Do not use one-to-two VLAN mapping together with zero-to-two, two-to-one, or two-to-two VLAN mapping.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Set the link type of the port.
   port link-type { **hybrid** | **trunk** }
   By default, the link type of a port is **access**.
4. Configure the port to allow packets from the SVLAN to pass through untagged.
   - Configure the SVLAN as the PVID of the trunk port and assign the trunk port to the SVLAN.
     port trunk pvid vlan *vlan-id*
     port trunk permit vlan { *vlan-id-list* | **all** }
   - Assign the hybrid port to the SVLAN as an untagged member.
     port hybrid vlan *vlan-id-list* **untagged**
5. Configure a one-to-two VLAN mapping.
   vlan mapping nest { **range** *vlan-range-list* | **single** *vlan-id-list* } **nested-vlan** *vlan-id*
   By default, no VLAN mapping is configured on an interface.

Configuring zero-to-two VLAN mapping

About this task

Configure zero-to-two VLAN mapping on PE 1 and PE 4, as shown in the network diagram in "One-to-two and two-to-two VLAN mapping," through which traffic from customer networks enters the service provider networks.

Configuration restrictions and guidelines

As a best practice, set the MTU to a minimum of 1504 bytes for ports on the forwarding path of the double-tagged packet in the service provider network.

Do not use zero-to-two VLAN mapping together with one-to-one, one-to-two, or many-to-one VLAN mapping.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*

- Enter Layer 2 aggregate interface view.
  interface bridge-aggregation *interface-number*
- Enter ONU subinterface service profile view.
  onu service-profile *profile-name*
- Execute the following commands in sequence to enter ONU line profile subinterface view.
  onu line-profile *profile-name*
  subinterface *subnumber*

3. Set the link type of the port.
   port link-type { **hybrid** | **trunk** }
   By default, the link type of a port is **access**.
4. Set the port PVID to VLAN 1.
   - Set the PVID to VLAN 1 for the trunk port.
     port trunk pvid vlan 1
   - Set the PVID to VLAN 1 for the hybrid port.
     port hybrid pvid vlan 1
5. Assign the port to the SVLAN and the port PVID (VLAN 1).
   - Assign the trunk port to the SVLAN and the port PVID (VLAN 1).
     port trunk permit vlan *vlan-id-list*
     By default, a trunk port is assigned to VLAN 1.
     The SVLAN of the trunk port must be different from the port PVID (VLAN 1).
   - Assign the hybrid port to the SVLAN and the port PVID (VLAN 1) as a tagged member.
     port hybrid vlan *vlan-id-list* **tagged**
     By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is access.
6. Configure a zero-to-two VLAN mapping.
   vlan mapping untagged nested-outer-vlan *outer-vlan-id* **nested-inner-vlan** *inner-vlan-id*
   By default, no VLAN mapping is configured on an interface.

Configuring two-to-one VLAN mapping

About this task

Configure two-to-one VLAN mapping on the customer-side port of Device B, as shown in the network diagram in "Two-to-one VLAN mapping." Device B will remove VLAN tags from double-tagged packets and add the CVLAN tag to them. When packets arrives Device A, Device A removes the CVLAN tag.

Restrictions and guidelines

On an interface, the original CVLAN and SVLAN of a two-to-one VLAN mapping cannot be the same as the translated CVLAN and SVLAN of a two-to-two or zero-to-two VLAN mapping.

You cannot specify multiple translated VLANs for the same original CVLAN and SVLAN on an interface. To modify an existing two-to-one VLAN mapping on an interface, you must execute the **undo vlan mapping egress** command to remove it first and then configure a new mapping.

Do not use two-to-one VLAN mapping together with one-to-one, one-to-two, or many-to-one VLAN mapping.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Set the port link type to hybrid or trunk.
   port link-type { **hybrid** | **trunk** }
   By default, the link type of a port is **access**.
4. Assign the port to the translated VLAN.
   - Assign the trunk port to the translated VLAN.
     port trunk permit vlan *vlan-id-list*
     By default, a trunk port is assigned to VLAN 1.
   - Assign the hybrid port to the translated VLAN.
     port hybrid vlan *vlan-id-list* **tagged**
     By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is access.
5. Configure a two-to-one VLAN mapping.
   vlan mapping egress outer-vlan *outer-vlan-id* **inner-vlan** *inner-vlan-id* **translated-vlan** *vlan-id*
   By default, no VLAN mapping is configured on an interface.

Configuring two-to-two VLAN mapping

About this task

Configure two-to-two VLAN mapping on the customer-side port of an edge device that connects two SP networks, for example, on PE 3 in the network diagram in "One-to-two and two-to-two VLAN mapping." Two-to-two VLAN mapping enables two sites in different VLANs to communicate at Layer 2 across two service provider networks that use different VLAN assignment schemes.

Restrictions and guidelines

Do not use two-to-two VLAN mapping together with one-to-one, one-to-two, or many-to-one VLAN mapping.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Set the link type of the port.
   port link-type { **hybrid** | **trunk** }

By default, the link type of a port is **access**.

4. Assign the port to the original VLANs and the translated VLANs.
   - Assign the trunk port to the original VLANs and the translated VLANs.

     port trunk permit vlan *vlan-id-list*

     By default, a trunk port is assigned to VLAN 1.
   - Assign the hybrid port to the original VLANs and the translated VLANs as a tagged member.

     port hybrid vlan *vlan-id-list* **tagged**

     By default, a hybrid port is an untagged member of the VLAN to which the port belongs when its link type is access.
5. Configure a two-to-two VLAN mapping.

   vlan mapping tunnel *outer-vlan-id inner-vlan-id* **translated-vlan** *outer-vlan-id inner-vlan-id*

   By default, no VLAN mapping is configured on an interface.

Display and maintenance commands for VLAN mapping

Execute **display** commands in any view.

Task	Command
Display VLAN mapping information.	**display vlan mapping** [**interface** *interface-type interface-number*]

VLAN mapping configuration examples

Example: Configuring one-to-one VLAN mapping

Network configuration

As shown in Figure 11:

- Each household subscribes to PC, VoD, and VoIP services.
- On the home gateways, VLANs 1, 2, and 3 are assigned to PC, VoD, and VoIP traffic, respectively.

To isolate traffic of the same service type from different households, configure one-to-one VLAN mappings on the wiring-closet switches. This feature assigns one VLAN to each type of traffic from each household.

Figure 11 Network diagram

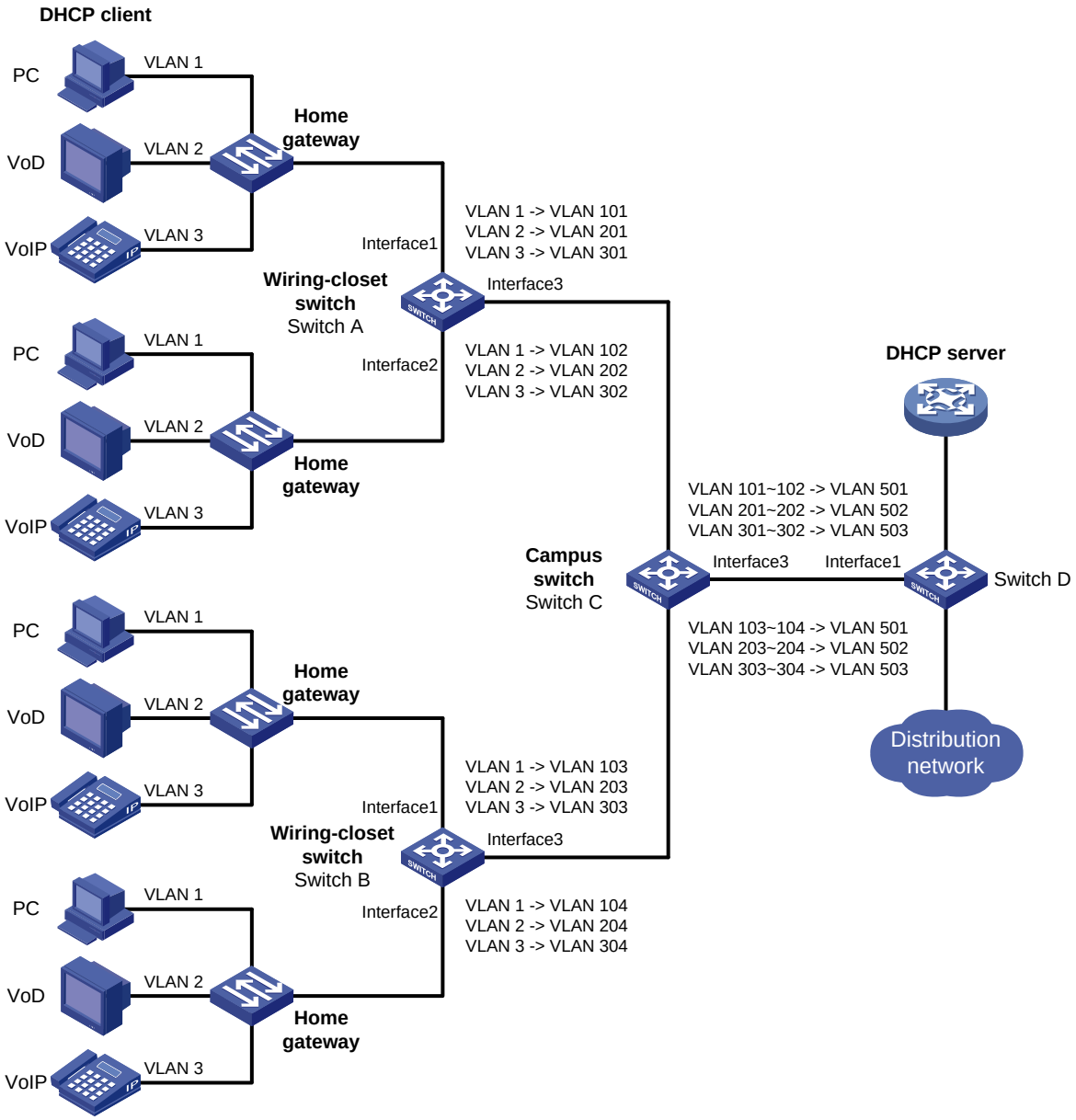


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Switch A:

   # Create the original VLANs.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 2 to 3
   ```

   # Create the translated VLANs.

```
[SwitchA] vlan 101 to 102
[SwitchA] vlan 201 to 202
[SwitchA] vlan 301 to 302
```

Configure customer-side port GigabitEthernet 1/0/1 as a trunk port.

```
<SwitchA> system-view
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-type trunk
```

Assign GigabitEthernet 1/0/1 to all original VLANs and translated VLANs.

```
[SwitchA-GigabitEthernet1/0/1] port trunk permit vlan 1 2 3 101 201 301
```

Configure one-to-one VLAN mappings on GigabitEthernet 1/0/1 to map VLANs 1, 2, and 3 to VLANs 101, 201, and 301, respectively.

```
[SwitchA-GigabitEthernet1/0/1] vlan mapping 1 translated-vlan 101
[SwitchA-GigabitEthernet1/0/1] vlan mapping 2 translated-vlan 201
[SwitchA-GigabitEthernet1/0/1] vlan mapping 3 translated-vlan 301
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure customer-side port GigabitEthernet 1/0/2 as a trunk port.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-type trunk
```

Assign GigabitEthernet 1/0/2 to all original VLANs and translated VLANs.

```
[SwitchA-GigabitEthernet1/0/2] port trunk permit vlan 1 2 3 102 202 302
```

Configure one-to-one VLAN mappings on GigabitEthernet 1/0/2 to map VLANs 1, 2, and 3 to VLANs 102, 202, and 302, respectively.

```
[SwitchA-GigabitEthernet1/0/2] vlan mapping 1 translated-vlan 102
[SwitchA-GigabitEthernet1/0/2] vlan mapping 2 translated-vlan 202
[SwitchA-GigabitEthernet1/0/2] vlan mapping 3 translated-vlan 302
[SwitchA-GigabitEthernet1/0/2] quit
```

Configure the network-side port (GigabitEthernet 1/0/3) as a trunk port.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] port link-type trunk
```

Assign GigabitEthernet 1/0/3 to the translated VLANs.

```
[SwitchA-GigabitEthernet1/0/3] port trunk permit vlan 101 201 301 102 202 302
[SwitchA-GigabitEthernet1/0/3] quit
```

2. Configure Switch B in the same way Switch A is configured. (Details not shown.)

Verifying the configuration

Verify VLAN mapping information on the wiring-closet switches, for example, Switch A.

```
[SwitchA] display vlan mapping
Interface GigabitEthernet1/0/1:
  Outer VLAN    Inner VLAN    Translated Outer VLAN    Translated Inner VLAN
  1             N/A           101                      N/A
  2             N/A           201                      N/A
  3             N/A           301                      N/A
Interface GigabitEthernet1/0/2:
  Outer VLAN    Inner VLAN    Translated Outer VLAN    Translated Inner VLAN
  1             N/A           102                      N/A
  2             N/A           202                      N/A
  3             N/A           302                      N/A
```

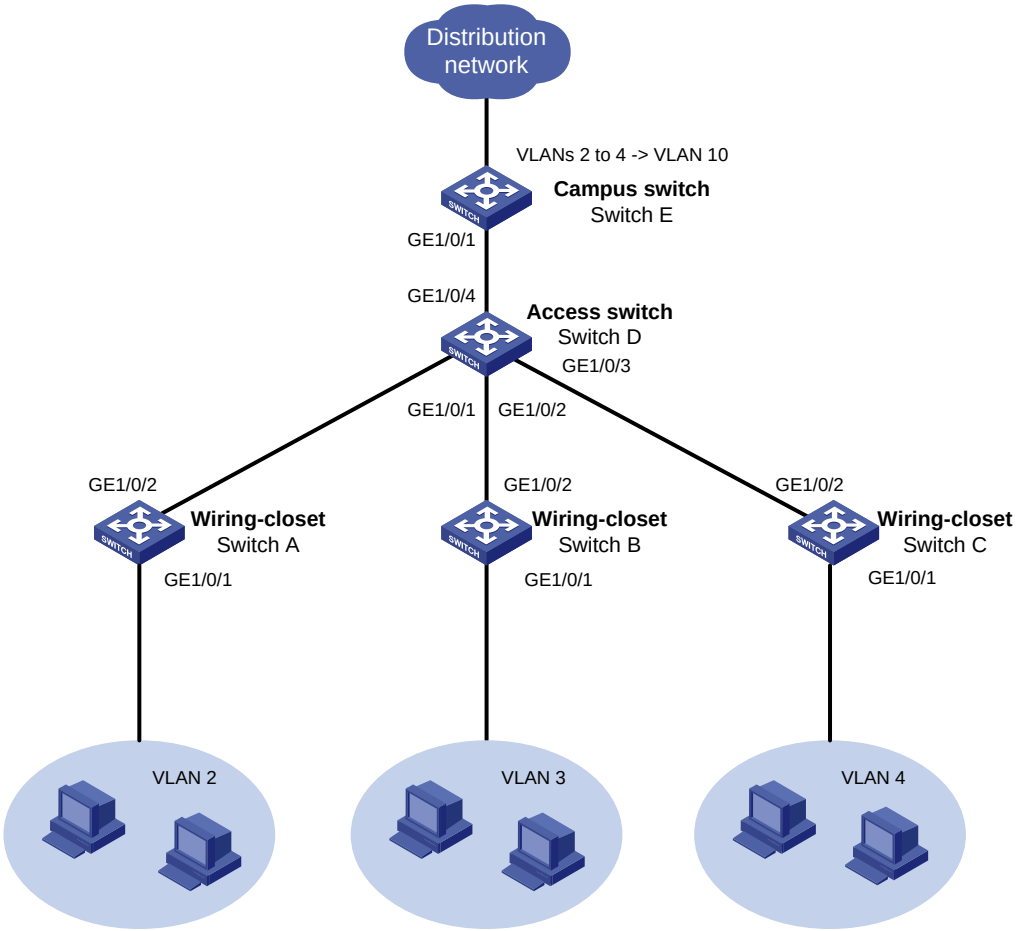
Example: Configuring many-to-one VLAN mapping

Network configuration

As shown in Figure 12:

- Create VLAN 2, VLAN 3, and VLAN 4 on the wiring-closet switches to isolate traffic of the same service type from different households.
- Configure many-to-one VLAN mappings on the campus switch. This feature assigns one VLAN to each type of traffic from different households.

Figure 12 Network diagram



Procedure

1. Configure Switch A:

   # Create VLAN 2 as an original VLAN.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 2
   [SwitchA-vlan2] quit
   ```

   # Assign ports GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to VLAN 2.

   ```
   [SwitchA] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/2
   [SwitchA-if-range] port access vlan 2
   [SwitchA-if-range] quit
   ```

2. Configure Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

3. Configure Switch D:

   # Create VLANs 2, 3, and 4 as original VLANs.

```
<SwitchD> system-view
[SwitchD] vlan 2 to 4
```

   # Assign ports GigabitEthernet 1/0/1 to VLAN 2, GigabitEthernet 1/0/2 to VLAN 3, and GigabitEthernet 1/0/3 to VLAN 4.

```
[SwitchD] interface gigabitethernet 1/0/1
[SwitchD-GigabitEthernet1/0/1] port access vlan 2
[SwitchD-GigabitEthernet1/0/1] quit
[SwitchD] interface gigabitethernet 1/0/2
[SwitchD-GigabitEthernet1/0/2] port access vlan 3
[SwitchD-GigabitEthernet1/0/2] quit
[SwitchD] interface gigabitethernet 1/0/3
[SwitchD-GigabitEthernet1/0/3] port access vlan 4
[SwitchD-GigabitEthernet1/0/3] quit
```

   # Configure port GigabitEthernet 1/0/4 as a trunk port.

```
[SwitchD] interface gigabitethernet 1/0/4
[SwitchD-GigabitEthernet1/0/4] port link-type trunk
```

   # Assign GigabitEthernet 1/0/4 to the original VLANs.

```
[SwitchD-GigabitEthernet1/0/4] port trunk permit vlan 2 to 4
[SwitchD-GigabitEthernet1/0/4] quit
```

4. Configure Switch E:

   # Create VLANs 2, 3, and 4 as original VLANs.

```
<SwitchE> system-view
[SwitchE] vlan 2 to 4
```

   # Configure the customer-side port (GigabitEthernet 1/0/1) as a trunk port.

```
<SwitchE> system-view
[SwitchE] interface gigabitethernet 1/0/1
[SwitchE-GigabitEthernet1/0/1] port link-type trunk
```

   # Assign GigabitEthernet 1/0/1 to the original VLANs.

```
[SwitchE-GigabitEthernet1/0/1] port trunk permit vlan 2 to 4
```

   # Configure many-to-one VLAN mapping on GigabitEthernet 1/0/1, which replaces VLAN tag 2 through VLAN tag 4 with VLAN tag 10.

```
[SwitchE-GigabitEthernet1/0/1] vlan mapping uni range 2 to 4 translated-vlan 10
[SwitchE-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Verify VLAN mapping information on Switch E.

```
[SwitchE] display vlan mapping
Interface GigabitEthernet1/0/1:
  Outer VLAN    Inner VLAN    Translated Outer VLAN    Translated Inner VLAN
  2-4           N/A           10                       N/A
```

Example: Configuring one-to-two and two-to-two VLAN mapping

Network configuration

As shown in Figure 13:

- Two VPN A branches, Site 1 and Site 2, are in VLAN 5 and VLAN 6, respectively.
- The two sites use different VPN access services from different service providers, SP 1 and SP 2.
- SP 1 assigns VLAN 100 to Site 1 and Site 2. SP 2 assigns VLAN 200 to Site 1 and Site 2.

Configure one-to-two VLAN mappings and two-to-two VLAN mappings to enable the two branches to communicate across networks SP 1 and SP 2.

Figure 13 Network diagram

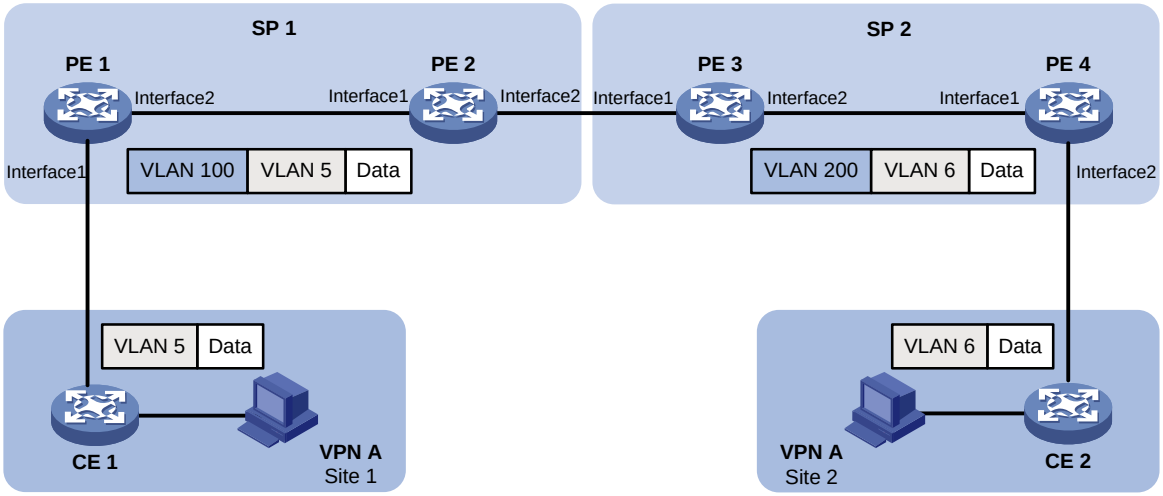


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure PE 1:

   # Create VLANs 5 and 100.

   ```
   <PE1> system-view
   [PE1] vlan 5
   [PE1-vlan5] quit
   [PE1] vlan 100
   [PE1-vlan100] quit
   ```

   # Configure a one-to-two VLAN mapping on the customer-side port (GigabitEthernet 1/0/1) to add SVLAN tag 100 to packets from VLAN 5.

   ```
   [PE1] interface gigabitethernet 1/0/1
   [PE1-GigabitEthernet1/0/1] vlan mapping nest single 5 nested-vlan 100
   ```

   # Configure GigabitEthernet 1/0/1 as a hybrid port.

   ```
   [PE1-GigabitEthernet1/0/1] port link-type hybrid
   ```

Assign GigabitEthernet 1/0/1 to VLAN 100 as an untagged member.

```
[PE1-GigabitEthernet1/0/1] port hybrid vlan 100 untagged
[PE1-GigabitEthernet1/0/1] quit
```

Configure the network-side port (GigabitEthernet 1/0/2) as a trunk port.

```
[PE1] interface gigabitethernet 1/0/2
[PE1-GigabitEthernet1/0/2] port link-type trunk
```

Assign GigabitEthernet 1/0/2 to VLAN 100.

```
[PE1-GigabitEthernet1/0/2] port trunk permit vlan 100
[PE1-GigabitEthernet1/0/2] quit
```

2. Configure PE 2:

Create VLAN 100.

```
<PE2> system-view
[PE2] vlan 100
[PE2-vlan100] quit
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[PE2] interface gigabitethernet 1/0/1
[PE2-GigabitEthernet1/0/1] port link-type trunk
```

Assign GigabitEthernet 1/0/1 to VLAN 100.

```
[PE2-GigabitEthernet1/0/1] port trunk permit vlan 100
[PE2-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port.

```
[PE2] interface gigabitethernet 1/0/2
[PE2-GigabitEthernet1/0/2] port link-type trunk
```

Assign GigabitEthernet 1/0/2 to VLAN 100.

```
[PE2-GigabitEthernet1/0/2] port trunk permit vlan 100
[PE2-GigabitEthernet1/0/2] quit
```

3. Configure PE 3:

Create VLANs 5, 6, 100, and 200.

```
<PE3> system-view
[PE3] vlan 5 to 6
[PE3] vlan 100
[PE3-vlan100] quit
[PE3] vlan 200
[PE3-vlan200] quit
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[PE3] interface gigabitethernet 1/0/1
[PE3-GigabitEthernet1/0/1] port link-type trunk
```

Assign GigabitEthernet 1/0/1 to VLANs 100 and 200.

```
[PE3-GigabitEthernet1/0/1] port trunk permit vlan 100 200
```

Configure a two-to-two VLAN mapping on GigabitEthernet 1/0/1 to map SVLAN 100 and CVLAN 5 to SVLAN 200 and CVLAN 6.

```
[PE3-GigabitEthernet1/0/1] vlan mapping tunnel 100 5 translated-vlan 200 6
[PE3-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port.

```
[PE3] interface gigabitethernet 1/0/2
[PE3-GigabitEthernet1/0/2] port link-type trunk
```

Assign GigabitEthernet 1/0/2 to VLAN 200.

```
[PE3-GigabitEthernet1/0/2] port trunk permit vlan 200
[PE3-GigabitEthernet1/0/2] quit
```

4. Configure PE 4:

Create VLANs 6 and 200.

```
<PE4> system-view
[PE4] vlan 6
[PE4-vlan6] quit
[PE4] vlan 200
[PE4-vlan200] quit
```

Configure the network-side port (GigabitEthernet 1/0/1) as a trunk port.

```
[PE4] interface gigabitethernet 1/0/1
[PE4-GigabitEthernet1/0/1] port link-type trunk
```

Assign GigabitEthernet 1/0/1 to VLAN 200.

```
[PE4-GigabitEthernet1/0/1] port trunk permit vlan 200
[PE4-GigabitEthernet1/0/1] quit
```

Configure the customer-side port (GigabitEthernet 1/0/2) as a hybrid port.

```
[PE4] interface gigabitethernet 1/0/2
[PE4-GigabitEthernet1/0/2] port link-type hybrid
```

Assign GigabitEthernet 1/0/2 to VLAN 200 as an untagged member.

```
[PE4-GigabitEthernet1/0/2] port hybrid vlan 200 untagged
```

Configure a one-to-two VLAN mapping on GigabitEthernet 1/0/2 to add SVLAN tag 200 to packets from VLAN 6.

```
[PE4-GigabitEthernet1/0/2] vlan mapping nest single 6 nested-vlan 200
[PE4-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Verify VLAN mapping information on PE 1.

```
[PE1] display vlan mapping
Interface GigabitEthernet1/0/1:
  Outer VLAN    Inner VLAN    Translated Outer VLAN    Translated Inner VLAN
  5             N/A           100                      5
```

Verify VLAN mapping information on PE 3.

```
[PE3] display vlan mapping
Interface GigabitEthernet1/0/1:
  Outer VLAN    Inner VLAN    Translated Outer VLAN    Translated Inner VLAN
  100           5             200                      6
```

Verify VLAN mapping information on PE 4.

```
[PE4] display vlan mapping
Interface GigabitEthernet1/0/2:
  Outer VLAN    Inner VLAN    Translated Outer VLAN    Translated Inner VLAN
  6             N/A           200                      6
```

Contents

Configuring loop detection 1

About loop detection 1
Loop detection mechanism 1
Loop detection interval 2
Loop protection actions 2
Loop protection delay 2
Interface status auto recovery 3
Restriction and guidelines: Loop detection configuration 3
Loop detection tasks at a glance 3
Enabling loop detection 4
Restrictions and guidelines for loop detection configuration 4
Enabling loop detection globally 4
Enabling loop detection on an interface 4
Configuring protected VLANs 4
Setting the loop protection action 5
Restrictions and guidelines for loop protection action configuration 5
Setting the global loop protection action 5
Setting the loop protection action on an interface 5
Setting the loop detection interval 5
Setting the loop protection delay timer 6
Restrictions and guidelines for loop protection delay timer configuration 6
Setting the global loop protection delay timer 6
Setting the loop protection delay timer on an interface 6
Display and maintenance commands for loop detection 6
Loop detection configuration examples 7
Example: Configuring basic loop detection functions 7
Example: Configuring loop detection for protected VLANs 10
Example: Configuring loop detection on an M-LAG system 14

Configuring loop detection

About loop detection

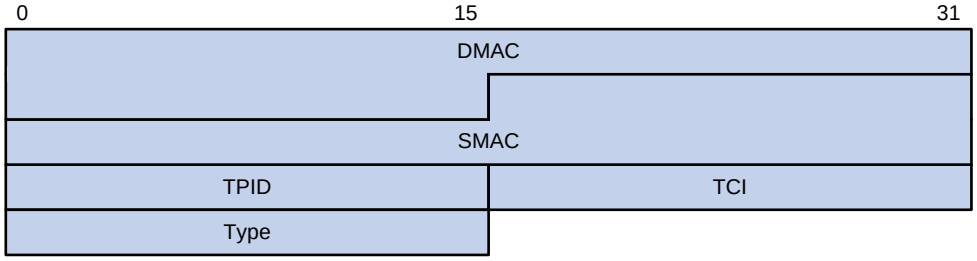
The loop detection mechanism performs periodic checking for Layer 2 loops. The mechanism immediately generates a log when a loop occurs so that you are promptly notified to adjust network connections and configurations. You can configure loop detection to shut down the looped interface. Logs are maintained in the information center. For more information, see *Network Management and Monitoring Configuration Guide*.

Loop detection mechanism

The device detects loops by sending detection frames and then checking whether these frames return to any interface on the device. If they do, the device considers that the interface is on a looped link.

Loop detection usually works within a VLAN. If a detection frame is returned with a different VLAN tag than it was sent out with, an inter-VLAN loop has occurred. To remove the loop, examine the QinQ or VLAN mapping configuration for incorrect settings. For more information about QinQ and VLAN mapping, see "Configuring QinQ" and "Configuring VLAN mapping."

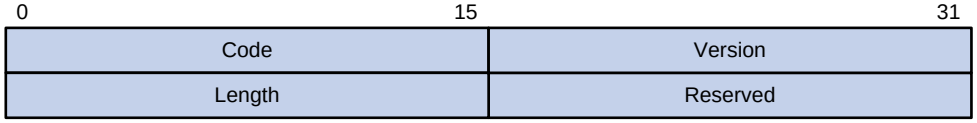
Figure 1 Ethernet frame header for loop detection



The Ethernet frame header of a loop detection packet contains the following fields:

- **DMAC**—Destination MAC address of the frame, which is the multicast MAC address 010f-e200-0007. When a loop detection-enabled device receives a frame with this destination MAC address, it performs the following operations:
  - Sends the frame to the CPU.
  - Floods the frame in the VLAN from which the frame was originally received.
- **SMAC**—Source MAC address of the frame, which is the bridge MAC address of the sending device.
- **TPID**—Type of the VLAN tag, with the value of 0x8100.
- **TCI**—Information of the VLAN tag, including the priority and VLAN ID.
- **Type**—Protocol type, with the value of 0x8918.

Figure 2 Inner frame header for loop detection



The inner frame header of a loop detection packet contains the following fields:

- **Code**—Protocol sub-type, which is 0x0001, indicating the loop detection protocol.
- **Version**—Protocol version, which is always 0x0000.
- **Length**—Length of the frame. The value includes the inner header, but excludes the Ethernet header.
- **Reserved**—This field is reserved.

Frames for loop detection are encapsulated as TLV triplets.

Table 1 TLVs supported by loop detection

TLV	Description	Remarks
End of PDU	End of a PDU.	Optional.
Device ID	Bridge MAC address of the sending device.	Required.
Port ID	ID of the PDU sending interface.	Optional.
Port Name	Name of the PDU sending interface.	Optional.
System Name	Device name.	Optional.
Chassis ID	Chassis ID of the sending interface.	Optional.
Slot ID	Slot ID of the sending interface.	Optional.
Sub Slot ID	Sub-slot ID of the sending interface.	Optional.

Loop detection interval

Loop detection is a continuous process as the network changes. Loop detection frames are sent at the loop detection interval to determine whether loops occur on interfaces and whether loops are removed.

Loop protection actions

When the device detects a loop on an interface, it generates a log but performs no action on the interface by default. You can configure the device to take one of the following actions:

- **Block**—Disables the interface from learning MAC addresses and blocks the interface.
- **No-learning**—Disables the interface from learning MAC addresses.
- **Shutdown**—Shuts down the interface to disable it from receiving and sending any frames.
- **VLAN-block**—Blocks traffic of the VLAN where the loop is detected on the interface.

  Loop detection allows you to use protected VLANs to block traffic on a per-VLAN basis. After you configure a VLAN as a protected VLAN, the device periodically sends loop detection packets in the VLAN. If the device receives a self-sent loop detection packet with the protected VLAN ID included on any interface, the device generates a log and blocks traffic of the VLAN on all interfaces configured with the VLAN-block action.

Loop protection delay

The loop protection delay enables the device to wait for a period of time before taking loop protection actions on looped interfaces. On a network protected by loop detection, the shutdown or block action taken on a looped interface of an upstream device might cause traffic interruption to its downstream devices. To minimize the impacts of loop detection on downstream traffic, set the loop protection

delay on the upstream devices. This delay allows the downstream devices to remove loops before the upstream devices take an action.

Interface status auto recovery

When the device configured with the block or no-learning loop action detects a loop on an interface, it performs the action and waits three loop detection intervals. If the device does not receive a loop detection frame within three loop detection intervals, it performs the following operations:

- Automatically sets the interface to the forwarding state.
- Notifies the user of the event.

When the device configured with the shutdown action detects a loop on an interface, the following events occur:

1. The device automatically shuts down the interface.
2. The device automatically sets the interface to the forwarding state after the detection timer set by using the **`shutdown-interval`** command expires. For more information about the **`shutdown-interval`** command, see *Fundamentals Command Reference*.
3. The device shuts down the interface again if a loop is still detected on the interface when the detection timer expires.

This process is repeated until the loop is removed.

The device performs the following operations if it has not received any loop detection packets in a blocked protected VLAN before three loop detection intervals expire:

- Enables interfaces to forward traffic from the VLAN.
- Generates a log message to report that the loop has been removed.

NOTE:

Incorrect recovery can occur when loop detection frames are discarded to reduce the load. To avoid this, use the shutdown action, or manually remove the loop.

Restriction and guidelines: Loop detection configuration

Member devices in an M-LAG system must have the same loop detection configuration.

Loop detection tasks at a glance

To configure loop detection, perform the following tasks:

1. Enabling loop detection
   - Enabling loop detection globally
   - Enabling loop detection on an interface
2. Configuring protected VLANs

   To use the VLAN-block action, you must first configure protected VLANs. Do not perform this task if you use any other loop protection action.
3. (Optional) Setting the loop protection action
   - Setting the global loop protection action
   - Setting the loop protection action on an interface

4. (Optional) Setting the loop detection interval
5. (Optional) Setting the loop protection delay timer
    - Setting the global loop protection delay timer
    - Setting the loop protection delay timer on an interface

Enabling loop detection

Restrictions and guidelines for loop detection configuration

You can enable loop detection globally or on a per-interface basis. When an interface receives a detection frame in any VLAN, the loop protection action is triggered on that interface, regardless of whether loop detection is enabled on it.

Enabling loop detection globally

1. Enter system view.
   system-view
2. Globally enable loop detection.
   loopback-detection global enable vlan { *vlan-id--list* | **all** }
   By default, loop detection is globally disabled.

Enabling loop detection on an interface

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   interface *interface-type interface-number*
3. Enable loop detection on the interface.
   loopback-detection enable vlan { *vlan-id-list* | **all** }
   By default, loop detection is disabled on interfaces.

Configuring protected VLANs

1. Enter system view.
   system-view
2. Configure protected VLANs.
   loopback-detection protected-vlan *vlan-id-list*
   By default, no protected VLANs are configured for loop detection.

Setting the loop protection action

Restrictions and guidelines for loop protection action configuration

You can set the loop protection action globally or on a per-interface basis. The global action applies to all interfaces. An interface-specific action applies to the interface on which it is configured. On an interface, the interface-specific action takes precedence over the global action.

Setting the global loop protection action

1. Enter system view.

   system-view

2. Set the global loop protection action.

   loopback-detection global action shutdown

   By default, the device generates a log but performs no action on the interface on which a loop is detected.

Setting the loop protection action on an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the loop protection action on the interface.

   loopback-detection action { **block** | **no-learning** | **shutdown** | **vlan-block** }

   By default, the device generates a log but performs no action on the interface on which a loop is detected.

   Support for the keywords of this command varies by interface type. For more information, see *Layer 2—LAN Switching Command Reference*.

Setting the loop detection interval

About this task

With loop detection enabled, the device sends loop detection frames at the loop detection interval. A shorter interval offers more sensitive detection but consumes more resources. Consider the system performance and loop detection speed when you set the loop detection interval.

Procedure

1. Enter system view.

   system-view

2. Set the loop detection interval.

   loopback-detection interval-time *interval*

   The default setting is 30 seconds.

Setting the loop protection delay timer

Restrictions and guidelines for loop protection delay timer configuration

The global delay timer applies to all interfaces enabled with loop detection. An interface-specific delay timer applies to the interface on which it is configured. On an interface, the interface-specific delay timer takes precedence over the global delay timer.

To prevent interfaces from going down or coming up concurrently, set the loop protection delay timer to different values on different devices or on different interfaces of the same device.

Setting the global loop protection delay timer

1. Enter system view.
   system-view
2. Set the global loop protection delay timer.
   loopback-detection global delay-timer *time*
   By default, loop protection delay is disabled globally.

Setting the loop protection delay timer on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the loop protection delay timer on the interface.
   loopback-detection delay-timer *time*
   By default, loop protection delay is disabled on an interface.

Display and maintenance commands for loop detection

Execute **display** commands in any view.

Task	Command
Display the loop detection configuration and status.	**display loopback-detection [loop-info]**

Loop detection configuration examples

Example: Configuring basic loop detection functions

Network configuration

As shown in Figure 3, configure loop detection on Device A to meet the following requirements:

- Device A generates a log as a notification.
- Device A automatically shuts down the interface on which a loop is detected.

Figure 3 Network diagram

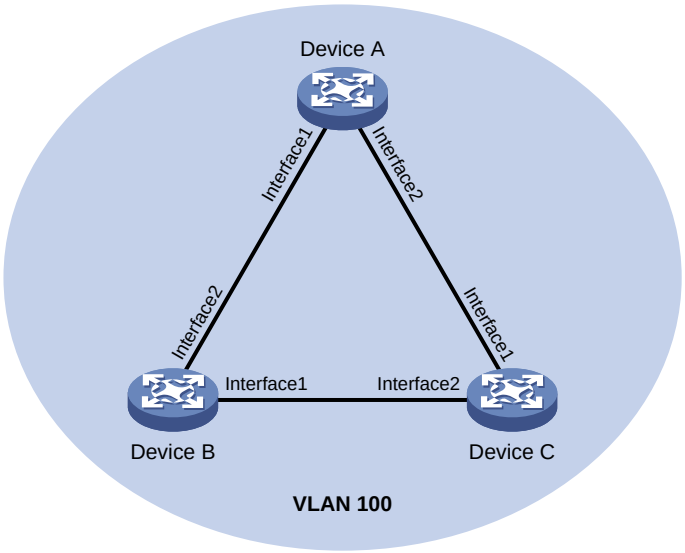


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

   # Create VLAN 100, and globally enable loop detection for the VLAN.

   ```
   <Sysname> system-view
   [Sysname] sysname DeviceA
   [DeviceA] vlan 100
   [DeviceA-vlan100] quit
   [DeviceA] loopback-detection global enable vlan 100
   ```

   # Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as trunk ports, and assign them to VLAN 100.

   ```
   [DeviceA] interface GigabitEthernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] port link-type trunk
   [DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 100
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 100
[DeviceA-GigabitEthernet1/0/2] quit
```

Set the global loop protection action to shutdown.

```
[DeviceA] loopback-detection global action shutdown
```

Set the loop detection interval to 35 seconds.

```
[DeviceA] loopback-detection interval-time 35
```

2. Configure Device B:

Create VLAN 100.

```
<Sysname> system-view
[Sysname] sysname DeviceB
[DeviceB] vlan 100
[DeviceB–vlan100] quit
```

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as trunk ports, and assign them to VLAN 100.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 100
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 100
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device C:

Create VLAN 100.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] vlan 100
[DeviceC–vlan100] quit
```

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as trunk ports, and assign them to VLAN 100.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 100
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 100
[DeviceC-GigabitEthernet1/0/2] quit
```

Verifying the configuration

View the system logs on devices, for example, Device A.

```
[DeviceA]
%Feb 24 15:04:29:663 2019 DeviceA LPDT/4/LPDT_LOOPED:  A loop was detected on
GigabitEthernet1/0/1.
%Feb 24 15:04:29:664 2019 DeviceA LPDT/4/LPDT_VLAN_LOOPED: A loop was detected on
GigabitEthernet1/0/1 in VLAN 100.
```

```
%Feb 24 15:04:29:667 2019 DeviceA LPDT/4/LPDT_LOOPED: A loop was detected on
GigabitEthernet1/0/2.
%Feb 24 15:04:29:668 2019 DeviceA LPDT/4/LPDT_VLAN_LOOPED: A loop was detected on
GigabitEthernet1/0/2 in VLAN 100.
%Feb 24 15:04:44:243 2019 DeviceA LPDT/5/LPDT_VLAN_RECOVERED: A loop was removed on
GigabitEthernet1/0/1 in VLAN 100.
%Feb 24 15:04:44:243 2019 DeviceA LPDT/5/LPDT_RECOVERED: All loops were removed on
GigabitEthernet1/0/1.
%Feb 24 15:04:44:248 2019 DeviceA LPDT/5/LPDT_VLAN_RECOVERED: A loop was removed on
GigabitEthernet1/0/2 in VLAN 100.
%Feb 24 15:04:44:248 2019 DeviceA LPDT/5/LPDT_RECOVERED: All loops were removed on
GigabitEthernet1/0/2.
```

The output shows the following information:

- Device A detected loops on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 within a loop detection interval.
- Loops on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 were removed.

Use the **`display loopback-detection`** command to display the loop detection configuration and status on devices, for example, Device A.

```
[DeviceA] display loopback-detection
Loop detection is enabled.
Global loop detection interval is 35 second(s).
Loop is detected on following interfaces:
  Interface                     Action mode     VLANs/VSI
  GigabitEthernet1/0/1          Shutdown        100
  GigabitEthernet1/0/2          Shutdown        100
```

The output shows that the device has removed the loops from GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 according to the shutdown action.

Display the status of GigabitEthernet 1/0/1 on devices, for example, Device A.

```
[DeviceA] display interface gigabitethernet 1/0/1
GigabitEthernet1/0/1 current state: DOWN (Loop detection down)
...
```

The output shows that GigabitEthernet 1/0/1 is already shut down by the loop detection module.

Display the status of GigabitEthernet 1/0/2 on devices, for example, Device A.

```
[DeviceA] display interface gigabitethernet 1/0/2
GigabitEthernet1/0/2 current state: DOWN (Loop detection down)
...
```

The output shows that GigabitEthernet 1/0/2 is already shut down by the loop detection module.

Configuration files

Device A

```
#
 sysname DeviceA
#
 loopback-detection global enable vlan 100
 loopback-detection global action shutdown
 loopback-detection interval-time 35
#
vlan 100
```

```
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 100
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100
#
```

Device B

```
#
 sysname DeviceB
#
vlan 100
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 100
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100
#
```

Device C

```
#
 sysname DeviceC
#
vlan 100
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 100
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100
#
```

Example: Configuring loop detection for protected VLANs

Network configuration

As shown in Figure 4, configure loop detection on Device A to meet the following requirements:

- Device A generates a log as a notification.
- Device A automatically blocks a protected VLAN on an interface if a loop is detected in the VLAN on the interface.

Figure 4 Network diagram

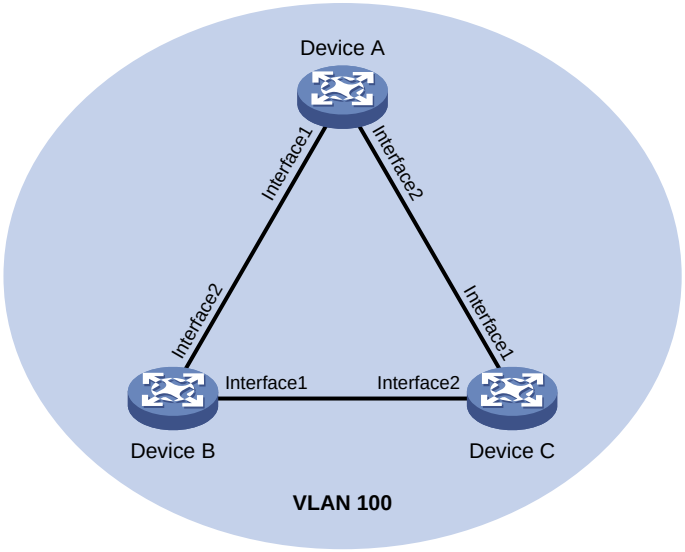


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

Create VLAN 100, configure it as a protected VLAN, and then globally enable loop detection for the VLAN.

```
<Sysname> system-view
[Sysname] sysname DeviceA
[DeviceA] vlan 100
[DeviceA–vlan100] quit
[DeviceA] loopback-detection protected-vlan 100
[DeviceA] loopback-detection global enable vlan 100
```

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as trunk ports, and assign them to VLAN 100.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 100
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 100
[DeviceA-GigabitEthernet1/0/2] quit
```

Enable VLAN-block mode on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo stp enable
[DeviceA-GigabitEthernet1/0/1] loopback-detection action vlan-block
```

```
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/1] undo stp enable
[DeviceA-GigabitEthernet1/0/1] loopback-detection action vlan-block
[DeviceA-GigabitEthernet1/0/1] quit
```

Set the loop detection interval to 35 seconds.

```
[DeviceA] loopback-detection interval-time 35
```

2. Configure Device B:

Create VLAN 100.

```
<Sysname> system-view
[Sysname] sysname DeviceB
[DeviceB] vlan 100
[DeviceB–vlan100] quit
```

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as trunk ports, and assign them to VLAN 100.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 100
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 100
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device C:

Create VLAN 100.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] vlan 100
[DeviceC–vlan100] quit
```

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as trunk ports, and assign them to VLAN 100.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 100
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 100
[DeviceC-GigabitEthernet1/0/2] quit
```

Verifying the configuration

View the system logs on Device A.

```
[DeviceA]
%Feb 24 15:04:29:663 2019 DeviceA LPDT/4/LPDT_LOOPED:  A loop was detected on
GigabitEthernet1/0/1.
%Feb 24 15:04:29:664 2019 DeviceA LPDT/4/LPDT_VLAN_LOOPED: A loop was detected on
GigabitEthernet1/0/1 in VLAN 100.
%Feb 24 15:04:29:667 2019 DeviceA LPDT/4/LPDT_LOOPED: A loop was detected on
GigabitEthernet1/0/2.
```

```
%Feb 24 15:04:29:668 2019 DeviceA LPDT/4/LPDT_VLAN_LOOPED: A loop was detected on
GigabitEthernet1/0/2 in VLAN 100.
```

The output shows that Device A detected loops in VLAN 100 on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 within a loop detection interval.

Use the **display loopback-detection** command to display the loop detection configuration and status on Device A.

```
[DeviceA] display loopback-detection
Loop detection is enabled.
Global loop detection interval is 35 second(s).
Loop is detected on following interfaces:
  Interface                     Action mode     VLANs/VSI
  GigabitEthernet1/0/1          VLAN-block      100
  GigabitEthernet1/0/2          VLAN-block      100
```

The output shows that the device has taken the VLAN-block action to remove the loops in VLAN 100 from GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

Configuration files

Device A

```
#
 sysname DeviceA
#
 loopback-detection global enable vlan 100
 loopback-detection interval-time 35
 loopback-detection protected-vlan 100
#
vlan 100
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 100
 undo stp enable
 loopback-detection action vlan-block
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100
 undo stp enable
 loopback-detection action vlan-block
#
```

Device B

```
#
 sysname DeviceB
#
vlan 100
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 100
```

```
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100
#
```

Device C

```
#
 sysname DeviceC
#
vlan 100
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 100
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 100
#
```

Example: Configuring loop detection on an M-LAG system

Network configuration

As shown in Figure 5, configure loop detection on the M-LAG system formed by Device A and Device B to meet the following requirements:

- Generates a log as a notification.
- Automatically shuts down the interface on which a loop is detected.

Figure 5 Network diagram

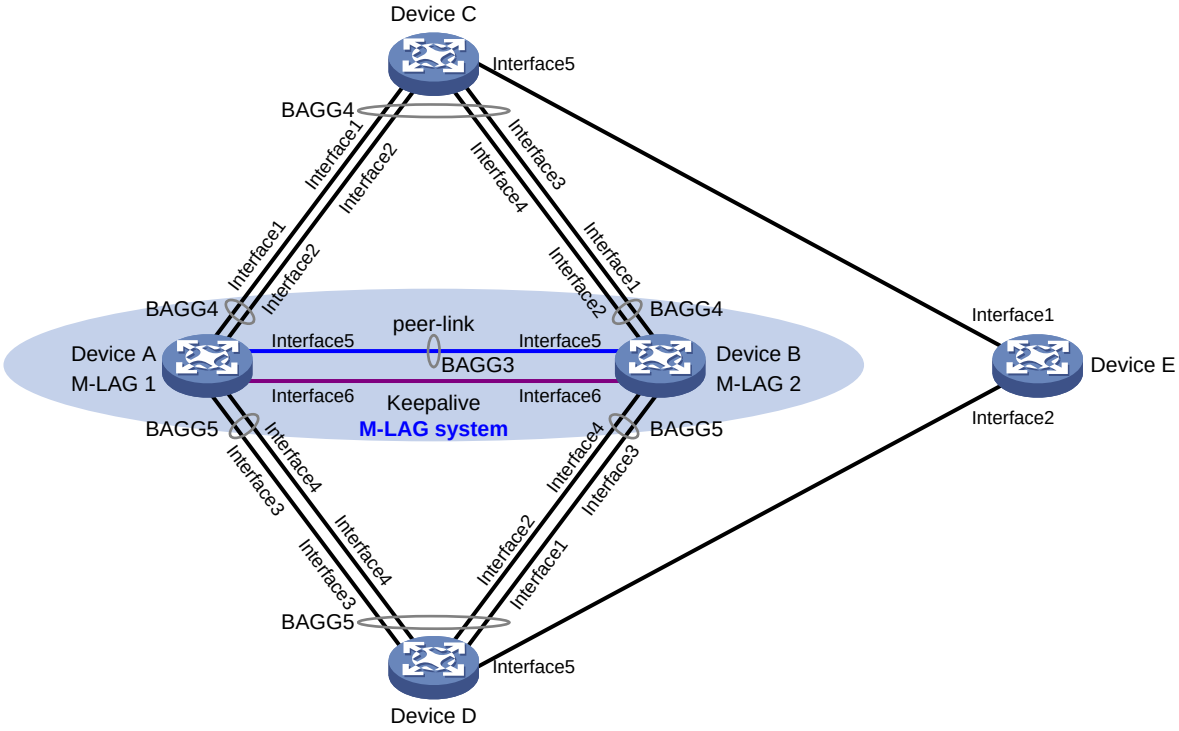


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5
Interface6	GigabitEthernet1/0/6

Procedure

1. Configure Device A:

   # Create VLAN 100.

   ```
   <DeviceA> system-view
   [DeviceA] vlan 100
   [DeviceA-vlan100] quit
   ```

   # Configure M-LAG system settings.

   ```
   [DeviceA] m-lag system-mac 1-1-1
   [DeviceA] m-lag system-number 1
   [DeviceA] m-lag system-priority 123
   ```

   # Configure M-LAG keepalive packet parameters.

   ```
   [DeviceA] m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
   ```

   # Set the link mode of GigabitEthernet 1/0/6 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[DeviceA] interface gigabitethernet 1/0/6
[DeviceA-GigabitEthernet1/0/6] port link-mode route
[DeviceA-GigabitEthernet1/0/6] ip address 1.1.1.2 24
[DeviceA-GigabitEthernet1/0/6] quit
```

Exclude the interface used for M-LAG keepalive detection (GigabitEthernet 1/0/6) from the shutdown action by M-LAG MAD.

```
[DeviceA] drni m-lag exclude interface gigabitethernet 1/0/6
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3, and specify it as the M-LAG peer-link interface.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation3] port m-lag peer-link 1
[DeviceA-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/5 to aggregation group 3.

```
[DeviceA] interface gigabitethernet 1/0/5
[DeviceA-GigabitEthernet1/0/5] port link-aggregation group 3
[DeviceA-GigabitEthernet1/0/5] quit
```

Set the link type of Bridge-Aggregation 3 to trunk, and assign it to VLAN 100.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] port link-type trunk
[DeviceA-Bridge-Aggregation3] port trunk permit vlan 100
[DeviceA-Bridge-Aggregation3] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4, and assign it to M-LAG group 4.

```
[DeviceA] interface bridge-aggregation 4
[DeviceA-Bridge-Aggregation4] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation4] port m-lag group 4
[DeviceA-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 4.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 4
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 4
[DeviceA-GigabitEthernet1/0/2] quit
```

Set the link type of Bridge-Aggregation 4 to trunk, and assign it to VLAN 100.

```
[DeviceA] interface bridge-aggregation 4
[DeviceA-Bridge-Aggregation4] port link-type trunk
[DeviceA-Bridge-Aggregation4] port trunk permit vlan 100
[DeviceA-Bridge-Aggregation4] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5, and assign it to M-LAG group 5.

```
[DeviceA] interface bridge-aggregation 5
[DeviceA-Bridge-Aggregation5] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation5] port m-lag group 5
[DeviceA-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 5.

```
[DeviceA] interface gigabitethernet 1/0/3
```

```
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 5
[DeviceA-GigabitEthernet1/0/3] quit
[DeviceA] interface gigabitethernet 1/0/4
[DeviceA-GigabitEthernet1/0/4] port link-aggregation group 5
[DeviceA-GigabitEthernet1/0/4] quit
```

Set the link type of Bridge-Aggregation 5 to trunk, and assign it to VLAN 100.

```
[DeviceA] interface bridge-aggregation 5
[DeviceA-Bridge-Aggregation5] port link-type trunk
[DeviceA-Bridge-Aggregation5] port trunk permit vlan 100
[DeviceA-Bridge-Aggregation5] quit
```

Disable the spanning tree feature.

```
[DeviceA] undo stp global enable
```

Enable loop detection for VLAN 100 globally, set the global loop protection action to shutdown, and set the loop detection interval to 35 seconds.

```
[DeviceA] loopback-detection global enable vlan 100
[DeviceA] loopback-detection global action shutdown
[DeviceA] loopback-detection interval-time 35
```

2. Configure Device B in the same way Device A is configured. Set the M-LAG system number to 2, and set the source and destination IP addresses of keepalive packets to 1.1.1.1 and 1.1.1.2, respectively. (Details not shown.)

3. Configure Device C:

Disable the spanning tree feature.

```
<DeviceC> system-view
[DeviceC] undo stp global enable
```

Create VLAN 100.

```
[DeviceC] vlan 100
[DeviceC-vlan100] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
[DeviceC] interface bridge-aggregation 4
[DeviceC-Bridge-Aggregation4] link-aggregation mode dynamic
[DeviceC-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to aggregation group 4.

```
[DeviceC] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[DeviceC-if-range] port link-aggregation group 4
[DeviceC-if-range] quit
```

Set the link type of Bridge-Aggregation 4 to trunk, and assign it to VLAN 100.

```
[DeviceC] interface bridge-aggregation 4
[DeviceC-Bridge-Aggregation4] port link-type trunk
[DeviceC-Bridge-Aggregation4] port trunk permit vlan 100
[DeviceC-Bridge-Aggregation4] quit
```

Set the link type of GigabitEthernet 1/0/5 to trunk, and assign it to VLAN 100.

```
[DeviceC] interface gigabitethernet 1/0/5
[DeviceC-GigabitEthernet1/0/5] port link-type trunk
[DeviceC-GigabitEthernet1/0/5] port trunk permit vlan 100
[DeviceC-GigabitEthernet1/0/5] quit
```

4. Configure Device D:

Disable the spanning tree feature.

```
<DeviceD> system-view
```

```
[DeviceD] undo stp global enable
```

Create VLAN 100.

```
[DeviceD] vlan 100
[DeviceD-vlan100] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5.

```
[DeviceD] interface bridge-aggregation 5
[DeviceD-Bridge-Aggregation5] link-aggregation mode dynamic
[DeviceD-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to aggregation group 5.

```
[DeviceD] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[DeviceD-if-range] port link-aggregation group 5
[DeviceD-if-range] quit
```

Set the link type of Bridge-Aggregation 5 to trunk, and assign it to VLAN 100.

```
[DeviceD] interface bridge-aggregation 5
[DeviceD-Bridge-Aggregation5] port link-type trunk
[DeviceD-Bridge-Aggregation5] port trunk permit vlan 100
[DeviceD-Bridge-Aggregation5] quit
```

Set the link type of GigabitEthernet 1/0/5 to trunk, and assign it to VLAN 100.

```
[DeviceD] interface gigabitethernet 1/0/5
[DeviceD-GigabitEthernet1/0/5] port link-type trunk
[DeviceD-GigabitEthernet1/0/5] port trunk permit vlan 100
[DeviceD-GigabitEthernet1/0/5] quit
```

5. Configure Device E:

Disable the spanning tree feature.

```
<DeviceE> system-view
[DeviceE] undo stp global enable
```

Create VLAN 100.

```
[DeviceE] vlan 100
[DeviceE-vlan100] quit
```

Set the link type of GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to trunk, and assign them to VLAN 100.

```
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] port link-type trunk
[DeviceE-GigabitEthernet1/0/1] port trunk permit vlan 100
[DeviceE-GigabitEthernet1/0/1] quit
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] port link-type trunk
[DeviceE-GigabitEthernet1/0/2] port trunk permit vlan 100
[DeviceE-GigabitEthernet1/0/2] quit
```

Verifying the configuration

View the system logs on Device A.

```
[DeviceA]
%Aug  1 03:28:48:110 2019 DeviceA LPDT/4/LPDT_LOOPED: A loop was detected on
 Bridge-Aggregation4.
%Aug  1 03:28:48:191 2019 DeviceA LPDT/4/LPDT_VLAN_LOOPED: A loop was detect
ed on Bridge-Aggregation4 in VLAN 100.
%Aug  1 03:28:48:194 2019 DeviceA LPDT/4/LPDT_LOOPED: A loop was detected on
```

```
 Bridge-Aggregation5.
%Aug  1 03:28:48:288 2019 DeviceA LPDT/4/LPDT_VLAN_LOOPED: A loop was detect
ed on Bridge-Aggregation5 in VLAN 100.
%Aug  1 03:28:48:290 2019 DeviceA LPDT/5/LPDT_VLAN_RECOVERED: A loop was rem
oved on Bridge-Aggregation4 in VLAN 100.
%Aug  1 03:28:48:291 2019 DeviceA LPDT/5/LPDT_RECOVERED: All loops were remo
ved on Bridge-Aggregation4.
%Aug  1 03:28:48:302 2019 DeviceA LPDT/5/LPDT_VLAN_RECOVERED: A loop was rem
oved on Bridge-Aggregation5 in VLAN 100.
%Aug  1 03:28:48:304 2019 DeviceA LPDT/5/LPDT_RECOVERED: All loops were remo
ved on Bridge-Aggregation5.
```

The output shows the following information:

- Device A detected loops on Bridge-Aggregation 4 and Bridge-Aggregation 5 within a loop detection interval.
- Loops on Bridge-Aggregation 4 and Bridge-Aggregation 5 were removed.

Use the **`display loopback-detection`** command to display the loop detection configuration and status on Device A.

```
[DeviceA] display loopback-detection
Loop detection is enabled.
Global loop detection interval is 35 second(s).
Loop is detected on following interfaces:
  Interface                     Action mode     VLANs/VSI
  Bridge-Aggregation4           Shutdown        100
  Bridge-Aggregation5           Shutdown        100
```

The output shows that the device has removed the loops from Bridge-Aggregation 4 and Bridge-Aggregation 5 according to the shutdown action.

Verify that Bridge-Aggregation 4 has been shut down by loop detection.

```
[DeviceA] display interface Bridge-Aggregation 4
Bridge-Aggregation4
The interface has assigned a M-LAG group.
Current state: DOWN (Loopback detection down)
…
```

Verify that Bridge-Aggregation 5 has been shut down by loop detection.

```
[DeviceA] display interface Bridge-Aggregation 5
Bridge-Aggregation5
The interface has assigned a M-LAG group.
Current state: DOWN (Loopback detection down)
…
```

Verify that loops have been removed on Device B. (Details not shown.)

Contents

Spanning tree protocol overview 1

About STP 1
STP protocol frames 1
Basic concepts in STP 3
Calculation process of the STP algorithm 4
Example of STP calculation 5
The configuration BPDU forwarding mechanism of STP 9
STP timers 9
About RSTP 10
RSTP protocol frames 10
Basic concepts in RSTP 10
How RSTP works 11
RSTP BPDU processing 11
About PVST 12
PVST protocol frames 12
How PVST works 13
About MSTP 13
MSTP features 13
MSTP protocol frames 13
Basic concepts in MSTP 15
How MSTP works 18
MSTP implementation on devices 19
Rapid transition mechanism 19
Edge port rapid transition 19
Root port rapid transition 20
P/A transition 20
Protocols and standards 21

Configuring spanning tree protocols 23

Restrictions and guidelines: spanning tree protocol configuration 23
Restrictions: Compatibility with other features 23
Restrictions: Interface configuration 23
Spanning tree protocol tasks at a glance 24
STP tasks at a glance 24
RSTP tasks at a glance 25
PVST tasks at a glance 26
MSTP tasks at a glance 27
Setting the spanning tree mode 29
Configuring an MST region 29
Configuring the root bridge or a secondary root bridge 30
Restrictions and guidelines 30
Configuring the device as the root bridge of a spanning tree 31
Configuring the device as a secondary root bridge of a spanning tree 31
Configuring the device priority 31
Configuring the maximum hops of an MST region 32
Configuring the network diameter of a switched network 32
Setting spanning tree timers 33
Setting the timeout factor 34
Configuring the BPDU transmission rate 35
Configuring edge ports 35
Configuring path costs of ports 36
About path cost 36
Specifying a standard for the default path cost calculation 36
Configuring path costs of ports 39
Configuring the port priority 40
Configuring the port link type 40
Configuring the mode a port uses to recognize and send MSTP frames 41

Enabling outputting port state transition information ····41
Enabling the spanning tree feature ····42
Restrictions and guidelines ····42
Enabling the spanning tree feature in STP/RSTP/MSTP mode ····42
Enabling the spanning tree feature in PVST mode ····43
Performing mCheck ····43
About mCheck ····43
Restrictions and guidelines ····43
Performing mCheck globally ····43
Performing mCheck in interface view ····44
Disabling inconsistent PVID protection ····44
Ignoring the specified VLANs ····44
Configuring Digest Snooping ····45
Configuring No Agreement Check ····46
Configuring TC Snooping ····48
Configuring protection features ····49
Spanning tree protection tasks at a glance ····49
Configuring BPDU guard ····50
Configuring BPDU filter ····51
Enabling root guard ····51
Enabling loop guard ····52
Configuring port role restriction ····53
Configuring TC-BPDU transmission restriction ····53
Enabling TC-BPDU guard ····54
Enabling BPDU drop ····54
Enabling PVST BPDU guard ····55
Disabling dispute guard ····55
Disabling loopback guard ····57
Enabling the device to log events of detecting or receiving TC BPDUs ····58
Disabling the device from reactivating edge ports shut down by BPDU guard ····58
Enabling BPDU transparent transmission on a port ····59
Configuring TC groups ····59
Configuring spanning tree blackhole detection ····61
Enabling SNMP notifications for new-root election and topology change events ····63
Display and maintenance commands for the spanning tree protocols ····64
Spanning tree configuration examples ····65
Example: Configuring MSTP ····65
Example: Configuring PVST ····69
Example: Configuring M-LAG with PVST ····72

Spanning tree protocol overview

Spanning tree protocols eliminate loops in a physical link-redundant network by selectively blocking redundant links and putting them in a standby state.

The recent versions of STP include the Rapid Spanning Tree Protocol (RSTP), the Per-VLAN Spanning Tree (PVST), and the Multiple Spanning Tree Protocol (MSTP).

About STP

STP was developed based on the 802.1d standard of IEEE to eliminate loops at the data link layer in a LAN. Networks often have redundant links as backups in case of failures, but loops are a very serious problem. Devices running STP detect loops in the network by exchanging information with one another. They eliminate loops by selectively blocking certain ports to prune the loop structure into a loop-free tree structure. This avoids proliferation and infinite cycling of packets that would occur in a loop network.

In a narrow sense, STP refers to IEEE 802.1d STP. In a broad sense, STP refers to the IEEE 802.1d STP and various enhanced spanning tree protocols derived from that protocol.

STP protocol frames

STP uses bridge protocol data units (BPDUs), also known as configuration messages, as its protocol frames. This chapter uses BPDUs to represent all types of spanning tree protocol frames.

STP-enabled devices exchange BPDUs to establish a spanning tree. BPDUs contain sufficient information for the devices to complete spanning tree calculation.

STP uses two types of BPDUs, configuration BPDUs and topology change notification (TCN) BPDUs.

Configuration BPDUs

Devices exchange configuration BPDUs to elect the root bridge and determine port roles. Figure 1 shows the configuration BPDU format.

Figure 1 Configuration BPDU format



DMA: Destination MAC address
SMA: Source MAC address
L/T: Frame length
LLC header: Logical link control header
Payload: BPDU data

Fields	Byte
Protocol ID	2
Protocol version ID	1
BPDU type	1
Flags	1
Root ID	8
Root path cost	4
Bridge ID	8
Port ID	2
Message age	2
Max age	2
Hello time	2
Forward delay	2

The payload of a configuration BPDU includes the following fields:

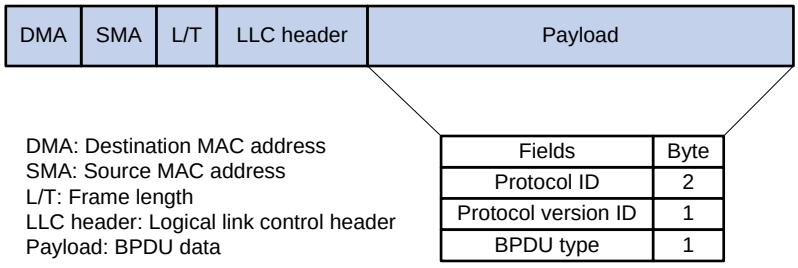
- **Protocol ID**—Fixed at 0x0000, which represents IEEE 802.1d.
- **Protocol version ID**—Spanning tree protocol version ID. The protocol version ID for STP is 0x00.
- **BPDU type**—Type of the BPDU. The value is 0x00 for a configuration BPDU.
- **Flags**—An 8-bit field indicates the purpose of the BPDU. The lowest bit is the Topology Change (TC) flag. The highest bit is the Topology Change Acknowledge (TCA) flag. All other bits are reserved.
- **Root ID**—Root bridge ID formed by the priority and MAC address of the root bridge.
- **Root path cost**—Cost of the path to the root bridge.
- **Bridge ID**—Designated bridge ID formed by the priority and MAC address of the designated bridge.
- **Port ID**—Designated port ID formed by the priority and global port number of the designated port.
- **Message age**—Age of the configuration BPDU while it propagates in the network.
- **Max age**—Maximum age of the configuration BPDU stored on the switch.
- **Hello time**—Configuration BPDU transmission interval.
- **Forward delay**—Delay for STP bridges to transit port state.

Devices use the root bridge ID, root path cost, designated bridge ID, designated port ID, message age, max age, hello time, and forward delay for spanning tree calculation.

TCN BPDUs

Devices use TCN BPDUs to announce changes in the network topology. Figure 2 shows the TCN BPDU format.

Figure 2 TCN BPDU format



The payload of a TCN BPDU includes the following fields:

- **Protocol ID**—Fixed at 0x0000, which represents IEEE 802.1d.
- **Protocol version ID**—Spanning tree protocol version ID. The protocol version ID for STP is 0x00.
- **BPDU type**—Type of the BPDU. The value is 0x80 for a TCN BPDU.

A non-root bridge sends TCN BPDUs when one of the following events occurs on the bridge:

- A port transits to the forwarding state, and the bridge has a minimum of one designated port.
- A port transits from the forwarding or learning state to the blocking state.

The non-root bridge uses TCN BPDUs to notify the root bridge once the network topology changes. The root bridge then sets the TC flag in its configuration BPDU and propagates it to other bridges.

Basic concepts in STP

Root bridge

A tree network must have a root bridge. The entire network contains only one root bridge, and all the other bridges in the network are called leaf nodes. The root bridge is not permanent, but can change with changes of the network topology.

Upon initialization of a network, each device generates and periodically sends configuration BPDUs, with itself as the root bridge. After network convergence, only the root bridge generates and periodically sends configuration BPDUs. The other devices only forward the BPDUs.

Root port

On a non-root bridge, the port nearest to the root bridge is the root port. The root port communicates with the root bridge. Each non-root bridge has only one root port. The root bridge has no root port.

Designated bridge and designated port

Classification	Designated bridge	Designated port
For a device	Device directly connected to the local device and responsible for forwarding BPDUs to the local device.	Port through which the designated bridge forwards BPDUs to this device.
For a LAN	Device responsible for forwarding BPDUs to this LAN segment.	Port through which the designated bridge forwards BPDUs to this LAN segment.

As shown in Figure 3, Device B and Device C are directly connected to a LAN.

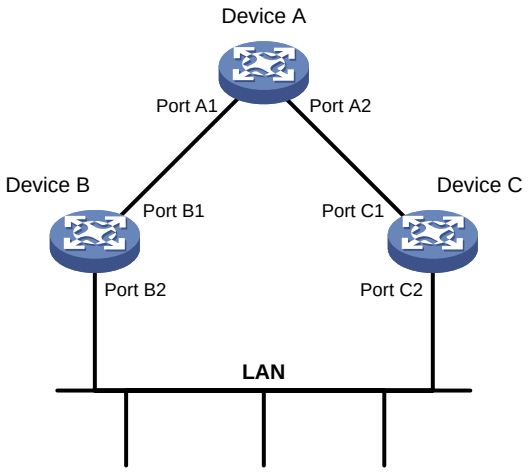
If Device A forwards BPDUs to Device B through port A1, the designated bridge and designated port are as follows:

- The designated bridge for Device B is Device A.
- The designated port for Device B is port A1 on Device A.

If Device B forwards BPDUs to the LAN, the designated bridge and designated port are as follows:

- The designated bridge for the LAN is Device B.
- The designated port for the LAN is port B2 on Device B.

Figure 3 Designated bridges and designated ports



Port states

Table 1 lists the port states in STP.

Table 1 STP port states

State	Receives/sends BPDUs	Learns MAC addresses	Forwards user data
Disabled	No	No	No
Listening	Yes	No	No
Learning	Yes	Yes	No
Forwarding	Yes	Yes	Yes
Blocking	Receive	No	No

Path cost

Path cost is a reference value used for link selection in STP. To prune the network into a loop-free tree, STP calculates path costs to select the most robust links and block redundant links that are less robust.

Calculation process of the STP algorithm

In STP calculation, a device compares the priorities of the received configuration BPDUs from different ports, and elects the root bridge, root ports and designated ports. When the spanning tree calculation is completed, a tree-shape topology forms.

The spanning tree calculation process described in the following sections is an example of a simplified process.

Network initialization

Upon initialization of a device, each port generates a BPDU with the following contents:

- The port as the designated port.
- The device as the root bridge.
- 0 as the root path cost.
- The device ID as the designated bridge ID.

Root bridge selection

The root bridge can be selected in the following methods:

- **Automatic election**—Initially, each STP-enabled device on the network assumes itself to be the root bridge, with its own device ID as the root bridge ID. By exchanging configuration BPDUs, the devices compare their root bridge IDs to elect the device with the smallest root bridge ID as the root bridge.
- **Manual assignment**—You can configure a device as the root bridge or a secondary root bridge of a spanning tree.
  - A spanning tree can have only one root bridge. If you configure multiple devices as the root bridge for a spanning tree, the device with the lowest MAC address is selected.
  - You can configure one or multiple secondary root bridges for a spanning tree. When the root bridge fails or is shut down, a secondary root bridge can take over. If multiple secondary root bridges are configured, the one with the lowest MAC address is selected. However, if a new root bridge is configured, the secondary root bridge is not selected.

Root port and designated ports selection on the non-root bridges

Step	Description
1	A non-root-bridge device regards the port on which it received the optimum configuration BPDU as the root port. Table 2 describes how the optimum configuration BPDU is selected.

Step	Description
2	Based on the configuration BPDU and the path cost of the root port, the device calculates a designated port configuration BPDU for each of the other ports. • The root bridge ID is replaced with that of the configuration BPDU of the root port. • The root path cost is replaced with that of the configuration BPDU of the root port plus the path cost of the root port. • The designated bridge ID is replaced with the ID of this device. • The designated port ID is replaced with the ID of this port.
3	The device compares the calculated configuration BPDU with the configuration BPDU on the port whose port role will be determined. Then, the device acts depending on the result of the comparison: • If the calculated configuration BPDU is superior, the device performs the following operations: ○ Considers this port as the designated port. ○ Replaces the configuration BPDU on the port with the calculated configuration BPDU. ○ Periodically sends the calculated configuration BPDU. • If the configuration BPDU on the port is superior, the device blocks this port without updating its configuration BPDU. The blocked port can receive BPDUs, but cannot send BPDUs or forward data traffic.

When the network topology is stable, only the root port and designated ports forward user traffic. Other ports are all in the blocking state to receive BPDUs but not to forward BPDUs or user traffic.

Table 2 Selecting the optimum configuration BPDU

Step	Actions
1	Upon receiving a configuration BPDU on a port, the device compares the priority of the received configuration BPDU with that of the configuration BPDU generated by the port. • If the former priority is lower, the device discards the received configuration BPDU and keeps the configuration BPDU the port generated. • If the former priority is higher, the device replaces the content of the configuration BPDU generated by the port with the content of the received configuration BPDU.
2	The device compares the configuration BPDUs of all the ports and chooses the optimum configuration BPDU.

The following are the principles of configuration BPDU comparison:

1. The configuration BPDU with the lowest root bridge ID has the highest priority.
2. If configuration BPDUs have the same root bridge ID, their root path costs are compared. For example, the root path cost in a configuration BPDU plus the path cost of a receiving port is S. The configuration BPDU with the smallest S value has the highest priority.
3. If all configuration BPDUs have the same root bridge ID and S value, the following attributes are compared in sequence:
   a. Designated bridge IDs.
   b. Designated port IDs.
   c. IDs of the receiving ports.

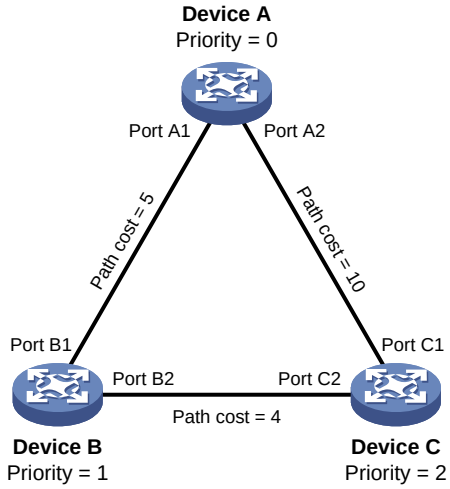
   The configuration BPDU that contains a smaller designated bridge ID, designated port ID, or receiving port ID is selected.

A tree-shape topology forms when the root bridge, root ports, and designated ports are selected.

Example of STP calculation

Figure 4 provides an example showing how the STP algorithm works.

Figure 4 The STP algorithm



As shown in Figure 4, the priority values of Device A, Device B, and Device C are 0, 1, and 2, respectively. The path costs of links among the three devices are 5, 10, and 4.

Device state initialization

In Table 3, each configuration BPDU contains the following fields: root bridge ID, root path cost, designated bridge ID, and designated port ID.

Table 3 Initial state of each device

Device	Port name	Configuration BPDU on the port
Device A	Port A1	{0, 0, 0, Port A1}
	Port A2	{0, 0, 0, Port A2}
Device B	Port B1	{1, 0, 1, Port B1}
	Port B2	{1, 0, 1, Port B2}
Device C	Port C1	{2, 0, 2, Port C1}
	Port C2	{2, 0, 2, Port C2}

Configuration BPDUs comparison on each device

In Table 4, each configuration BPDU contains the following fields: root bridge ID, root path cost, designated bridge ID, and designated port ID.

Table 4 Comparison process and result on each device

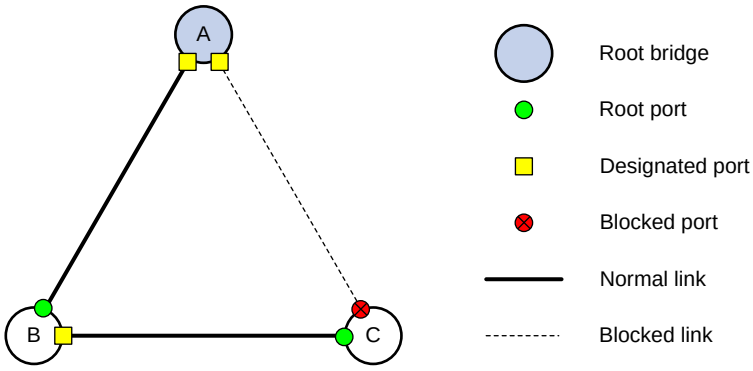
Device	Comparison process	Configuration BPDU on ports after comparison
Device A	Port A1 performs the following operations: **1.** Receives the configuration BPDU of Port B1 {1, 0, 1, Port B1}. **2.** Determines that its existing configuration BPDU {0, 0, 0, Port A1} is superior to the received configuration BPDU. **3.** Discards the received one. Port A2 performs the following operations: **4.** Receives the configuration BPDU of Port C1 {2, 0, 2, Port C1}. **5.** Determines that its existing configuration BPDU {0, 0, 0, Port A2} is superior to the received configuration BPDU. **6.** Discards the received one. Device A determines that it is both the root bridge and designated bridge in the configuration BPDUs of all its ports. It considers itself as the root bridge. It does not change the configuration BPDU of any port and starts to periodically send configuration BPDUs.	• Port A1: {0, 0, 0, Port A1} • Port A2: {0, 0, 0, Port A2}
Device B	Port B1 performs the following operations: **7.** Receives the configuration BPDU of Port A1 {0, 0, 0, Port A1}. **8.** Determines that the received configuration BPDU is superior to its existing configuration BPDU {1, 0, 1, Port B1}. **9.** Updates its configuration BPDU. Port B2 performs the following operations: **10.** Receives the configuration BPDU of Port C2 {2, 0, 2, Port C2}. **11.** Determines that its existing configuration BPDU {1, 0, 1, Port B2} is superior to the received configuration BPDU. **12.** Discards the received BPDU.	• Port B1: {0, 0, 0, Port A1} • Port B2: {1, 0, 1, Port B2}
	Device B performs the following operations: **13.** Compares the configuration BPDUs of all its ports. **14.** Decides that the configuration BPDU of Port B1 is the optimum. **15.** Selects Port B1 as the root port with the configuration BPDU unchanged. Based on the configuration BPDU and path cost of the root port, Device B calculates a designated port configuration BPDU for Port B2 {0, 5, 1, Port B2}. Device B compares it with the existing configuration BPDU of Port B2 {1, 0, 1, Port B2}. Device B determines that the calculated one is superior, and determines that Port B2 is the designated port. It replaces the configuration BPDU on Port B2 with the calculated one, and periodically sends the calculated configuration BPDU.	• Root port (Port B1): {0, 0, 0, Port A1} • Designated port (Port B2): {0, 5, 1, Port B2}
Device C	Port C1 performs the following operations: **16.** Receives the configuration BPDU of Port A2 {0, 0, 0, Port A2}. **17.** Determines that the received configuration BPDU is superior to its existing configuration BPDU {2, 0, 2,	• Port C1: {0, 0, 0, Port A2} • Port C2: {1, 0, 1, Port B2}

	Port C1}. **18.** Updates its configuration BPDU. Port C2 performs the following operations: **19.** Receives the original configuration BPDU of Port B2 {1, 0, 1, Port B2}. **20.** Determines that the received configuration BPDU is superior to the existing configuration BPDU {2, 0, 2, Port C2}. **21.** Updates its configuration BPDU.	
	Device C performs the following operations: **22.** Compares the configuration BPDUs of all its ports. **23.** Decides that the configuration BPDU of Port C1 is the optimum. **24.** Selects Port C1 as the root port with the configuration BPDU unchanged. Based on the configuration BPDU and path cost of the root port, Device C calculates the configuration BPDU of Port C2 {0, 10, 2, Port C2}. Device C compares it with the existing configuration BPDU of Port C2 {1, 0, 1, Port B2}. Device C determines that the calculated configuration BPDU is superior to the existing one, selects Port C2 as the designated port, and replaces the configuration BPDU of Port C2 with the calculated one.	• Root port (Port C1): {0, 0, 0, Port A2} • Designated port (Port C2): {0, 10, 2, Port C2}
	Port C2 performs the following operations: **25.** Receives the updated configuration BPDU of Port B2 {0, 5, 1, Port B2}. **26.** Determines that the received configuration BPDU is superior to its existing configuration BPDU {0, 10, 2, Port C2}. **27.** Updates its configuration BPDU. Port C1 performs the following operations: **28.** Receives a periodic configuration BPDU {0, 0, 0, Port A2} from Port A2. **29.** Determines that it is the same as the existing configuration BPDU. **30.** Discards the received BPDU.	• Port C1: {0, 0, 0, Port A2} • Port C2: {0, 5, 1, Port B2}
	Device C determines that the root path cost of Port C1 is larger than that of Port C2. The root path cost of Port C1 is 10, root path cost of the received configuration BPDU (0) plus path cost of Port C1 (10). The root path cost of Port C2 is 9, root path cost of the received configuration BPDU (5) plus path cost of Port C2 (4). Device C determines that the configuration BPDU of Port C2 is the optimum, and selects Port C2 as the root port with the configuration BPDU unchanged. Based on the configuration BPDU and path cost of the root port, Device C performs the following operations: **31.** Calculates a designated port configuration BPDU for Port C1 {0, 9, 2, Port C1}. **32.** Compares it with the existing configuration BPDU of Port C1 {0, 0, 0, Port A2}. **33.** Determines that the existing configuration BPDU is superior to the calculated one and blocks Port C1 with the configuration BPDU unchanged. Port C1 does not forward data until a new event triggers a spanning tree calculation process: for example, the link between Device B and Device C is down.	• Blocked port (Port C1): {0, 0, 0, Port A2} • Root port (Port C2): {0, 5, 1, Port B2}

Final calculated spanning tree

After the comparison processes described in Table 4, a spanning tree with Device A as the root bridge is established, as shown in Figure 5.

Figure 5 The final calculated spanning tree



The configuration BPDU forwarding mechanism of STP

The configuration BPDUs of STP are forwarded according to these guidelines:

- Upon network initiation, every device regards itself as the root bridge and generates configuration BPDUs with itself as the root. Then it sends the configuration BPDUs at a regular hello interval.
- If the root port receives a configuration BPDU superior to the configuration BPDU of the port, the device performs the following operations:
  - Increases the message age carried in the configuration BPDU.
  - Starts a timer to time the configuration BPDU.
  - Sends this configuration BPDU through the designated port.
- If a designated port receives a configuration BPDU with a lower priority than its configuration BPDU, the port immediately responds with its configuration BPDU.
- If a path fails, the root port on this path no longer receives new configuration BPDUs and the old configuration BPDUs will be discarded due to timeout. The device generates a configuration BPDU with itself as the root and sends the BPDUs and TCN BPDUs. This triggers a new spanning tree calculation process to establish a new path to restore the network connectivity.

However, the newly calculated configuration BPDU cannot be propagated throughout the network immediately. As a result, the old root ports and designated ports that have not detected the topology change continue forwarding data along the old path. If the new root ports and designated ports begin to forward data as soon as they are elected, a temporary loop might occur.

STP timers

The most important timing parameters in STP calculation are forward delay, hello time, and max age.

- Forward delay

  Forward delay is the delay time for port state transition. By default, the forward delay is 15 seconds.

  A path failure can cause spanning tree re-calculation to adapt the spanning tree structure to the change. However, the resulting new configuration BPDU cannot propagate throughout the network immediately. If the newly elected root ports and designated ports start to forward data immediately, a temporary loop will likely occur.

The newly elected root ports or designated ports must go through the listening and learning states before they transit to the forwarding state. This requires twice the forward delay time and allows the new configuration BPDU to propagate throughout the network.

- Hello time

  The device sends configuration BPDUs at the hello time interval to the neighboring devices to ensure that the paths are fault-free. By default, the hello time is 2 seconds. If the device does not receive configuration BPDUs within the timeout period, it recalculates the spanning tree. The formula for calculating the timeout period is timeout period = timeout factor × 3 × hello time.

- Max age

  The device uses the max age to determine whether a stored configuration BPDU has expired and discards it if the max age is exceeded. By default, the max age is 20 seconds. In the CIST of an MSTP network, the device uses the max age timer to determine whether a configuration BPDU received by a port has expired. If it is expired, a new spanning tree calculation process starts. The max age timer does not take effect on MSTIs.

If a port does not receive any configuration BPDUs within the timeout period, the port transits to the listening state. The device will recalculate the spanning tree. It takes the port 50 seconds to transit back to the forwarding state. This period includes 20 seconds for the max age, 15 seconds for the listening state, and 15 seconds for the learning state.

To ensure a fast topology convergence, make sure the timer settings meet the following formulas:

- 2 × (forward delay – 1 second) ≥ max age
- Max age ≥ 2 × (hello time + 1 second)

About RSTP

RSTP achieves rapid network convergence by allowing a newly elected root port or designated port to enter the forwarding state much faster than STP.

RSTP protocol frames

An RSTP BPDU uses the same format as an STP BPDU except that a Version1 length field is added to the payload of RSTP BPDUs. The differences between an RSTP BPDU and an STP BPDU are as follows:

- **Protocol version ID**—The value is 0x02 for RSTP.
- **BPDU type**—The value is 0x02 for RSTP BPDUs.
- **Flags**—All 8 bits are used.
- **Version1 length**—The value is 0x00, which means no version 1 protocol information is present.

RSTP does not use TCN BPDUs to advertise topology changes. RSTP floods BPDUs with the TC flag set in the network to advertise topology changes.

Basic concepts in RSTP

Port roles

In addition to root port and designated port, RSTP also uses the following port roles:

- **Alternate port**—Acts as the backup port for a root port. When the root port is blocked, the alternate port takes over.
- **Backup port**—Acts as the backup port of a designated port. When the designated port is invalid, the backup port becomes the new designated port. A loop occurs when two ports of the

same spanning tree device are connected, so the device blocks one of the ports. The blocked port is the backup port.

- **Edge port**—Directly connects to a user host rather than a network device or network segment.

Port states

RSTP uses the discarding state to replace the disabled, blocking, and listening states in STP. Table 5 shows the differences between the port states in RSTP and STP.

Table 5 Port state differences between RSTP and STP

STP port state	RSTP port state	Sends BPDU	Learns MAC addresses	Forwards user data
Disabled	Discarding	No	No	No
Blocking	Discarding	No	No	No
Listening	Discarding	Yes	No	No
Learning	Learning	Yes	Yes	No
Forwarding	Forwarding	Yes	Yes	Yes

How RSTP works

During RSTP calculation, the following events occur:

- If a port in discarding state becomes an alternate port, it retains its state.
- If a port in discarding state is elected as the root port or designated port, it enters the learning state after the forward delay. The port learns MAC addresses, and enters the forwarding state after another forward delay.
  - A newly elected RSTP root port rapidly enters the forwarding state if the following requirements are met:
    - The old root port on the device has stopped forwarding data.
    - The upstream designated port has started forwarding data.
  - A newly elected RSTP designated port rapidly enters the forwarding state if one of the following requirements is met:
    - The designated port is configured as an edge port which directly connects to a user terminal.
    - The designated port connects to a point-to-point link and receives a handshake response from the directly connected device.

RSTP BPDU processing

In RSTP, a non-root bridge actively sends RSTP BPDUs at the hello time through designated ports without waiting for the root bridge to send RSTP BPDUs. This enables RSTP to quickly detect link failures. If a device fails to receive any RSTP BPDUs on a port within triple the hello time, the device considers that a link failure has occurred. After the stored configuration BPDU expires, the device floods RSTP BPDUs with the TC flag set to initiate a new RSTP calculation.

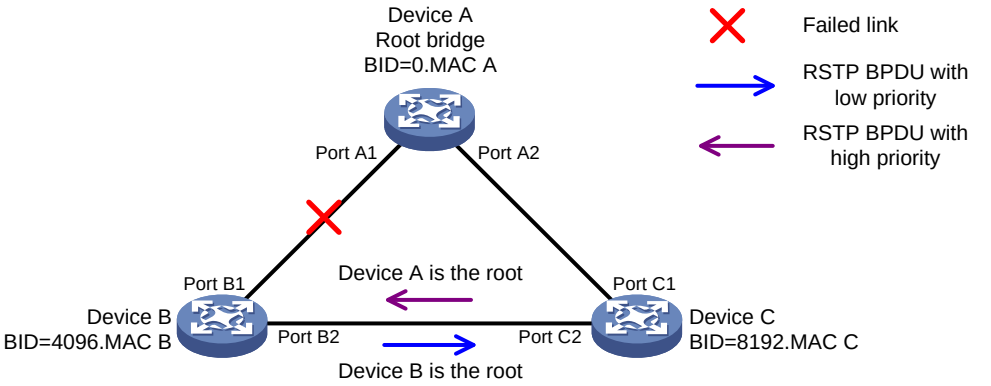
In RSTP, a port in blocking state can immediately respond to an RSTP BPDU with a lower priority than its own BPDU.

As shown in Figure 6, Device A is the root bridge. The priority of Device B is higher than the priority of Device C. Port C2 on Device C is blocked.

When the link between Device A and Device B fails, the following events occur:

1. Device B sends an RSTP BPDU with itself as the root bridge to Device C.
2. Device C compares the RSTP BPDU with its own BPDU.
3. Because the RSTP BPDU from Device B has a lower priority, Device C sends its own BPDU to Device B.
4. Device B considers that Port B2 is the root port and stops sending RSTP BPDUs to Device C.

Figure 6 BPDU processing in RSTP



About PVST

In an STP- or RSTP-enabled LAN, all bridges share one spanning tree. Traffic from all VLANs is forwarded along the spanning tree, and ports cannot be blocked on a per-VLAN basis to prune loops.

PVST allows every VLAN to have its own spanning tree, which increases usage of links and bandwidth. Because each VLAN runs RSTP independently, a spanning tree only serves its VLAN.

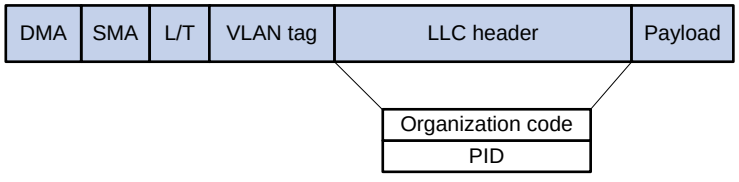
A PVST-enabled H3C device can communicate with a third-party device that is running Rapid PVST or PVST. The PVST-enabled H3C device supports fast network convergence like RSTP when connected to PVST-enabled H3C devices or third-party devices enabled with Rapid PVST.

PVST protocol frames

As shown in Figure 7, a PVST BPDU uses the same format as an RSTP BPDU except the following differences:

- The destination MAC address of a PVST BPDU is 01-00-0c-cc-cc-cd, which is a private MAC address.
- Each PVST BPDU carries a VLAN tag. The VLAN tag identifies the VLAN to which the PVST BPDU belongs.
- The organization code and PID fields are added to the LLC header of the PVST BPDU.

Figure 7 PVST BPDU format



A port's link type determines the type of BPDUs the port sends.

- An access port sends RSTP BPDUs.

- A trunk or hybrid port sends RSTP BPDUs in the default VLAN and sends PVST BPDUs in other VLANs.

How PVST works

PVST implements per-VLAN spanning tree calculation by mapping each VLAN to an MSTI. In PVST, each VLAN runs RSTP independently to maintain its own spanning tree without affecting the spanning trees of other VLANs. In this way, loops in each VLAN are eliminated and traffic of different VLANs is load shared over links. PVST uses RSTP BPDUs in the default VLAN and PVST BPDUs in other VLANs for spanning tree calculation.

PVST uses the same port roles and port states as RSTP for rapid transition. For more information, see "Basic concepts in RSTP."

About MSTP

MSTP features

Developed based on IEEE 802.1s, MSTP overcomes the limitations of STP, RSTP, and PVST. In addition to supporting rapid network convergence, it allows data flows of different VLANs to be forwarded along separate paths. This provides a better load sharing mechanism for redundant links.

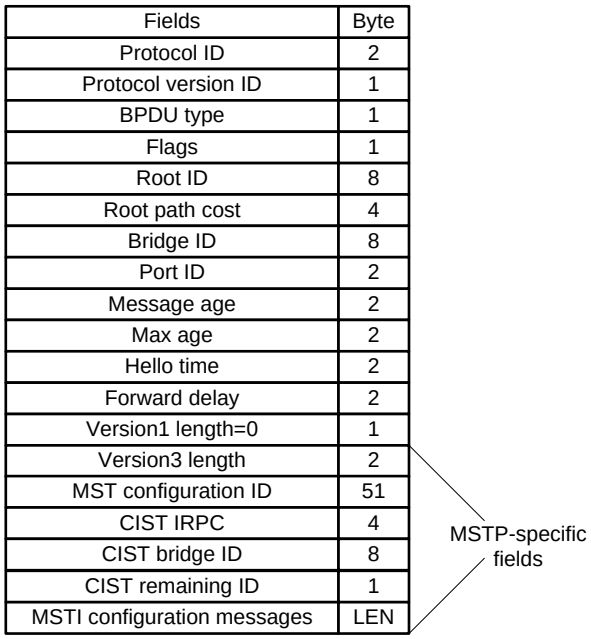
MSTP provides the following features:

- MSTP divides a switched network into multiple regions, each of which contains multiple spanning trees that are independent of one another.
- MSTP supports mapping VLANs to spanning tree instances by means of a VLAN-to-instance mapping table. MSTP can reduce communication overheads and resource usage by mapping multiple VLANs to one instance.
- MSTP prunes a loop network into a loop-free tree, which avoids proliferation and endless cycling of frames in a loop network. In addition, it supports load balancing of VLAN data by providing multiple redundant paths for data forwarding.
- MSTP is compatible with STP and RSTP, and partially compatible with PVST.

MSTP protocol frames

Figure 8 shows the format of an MSTP BPDU.

Figure 8 MSTP BPDU format



The first 13 fields of an MSTP BPDU are the same as an RSTP BPDU. The other six fields are unique to MSTP.

- **Protocol version ID**—The value is 0x03 for MSTP.
- **BPDU type**—The value is 0x02 for RSTP/MSTP BPDUs.
- **Root ID**—ID of the common root bridge.
- **Root path cost**—CIST external path cost.
- **Bridge ID**—ID of the regional root for the IST or an MSTI.
- **Port ID**—ID of the designated port in the CIST.
- **Version3 length**—Length of the MSTP-specific fields. Devices use this field for verification upon receiving an MSTP BPDU.
- **MST configuration ID**—Includes the format selector, configuration name, revision level, and configuration digest. The value for format selector is fixed at 0x00. The other parameters are used to identify the MST region for the originating bridge.
- **CIST IRPC**—Internal root path cost (IRPC) from the originating bridge to the root of the MST region.
- **CIST bridge ID**—ID of the bridge that sends the MSTP BPDU.
- **CIST remaining ID**—Remaining hop count. This field limits the scale of the MST region. The regional root sends a BPDU with the remaining hop count set to the maximum value. Each device that receives the BPDU decrements the hop count by one. When the hop count reaches zero, the BPDU is discarded. Devices beyond the maximum hops of the MST region cannot participate in spanning tree calculation. The default remaining hop count is 20.
- **MSTI configuration messages**—Contains MSTI configuration messages. Each MSTI configuration message is 16 bytes. This field can contain 0 to 64 MSTI configuration messages. The number of the MSTI configuration messages is determined by the number of MSTIs in the MST region.

Basic concepts in MSTP

Figure 9 shows a switched network that contains four MST regions, each MST region containing four MSTP devices. Figure 10 shows the networking topology of MST region 3.

Figure 9 Basic concepts in MSTP

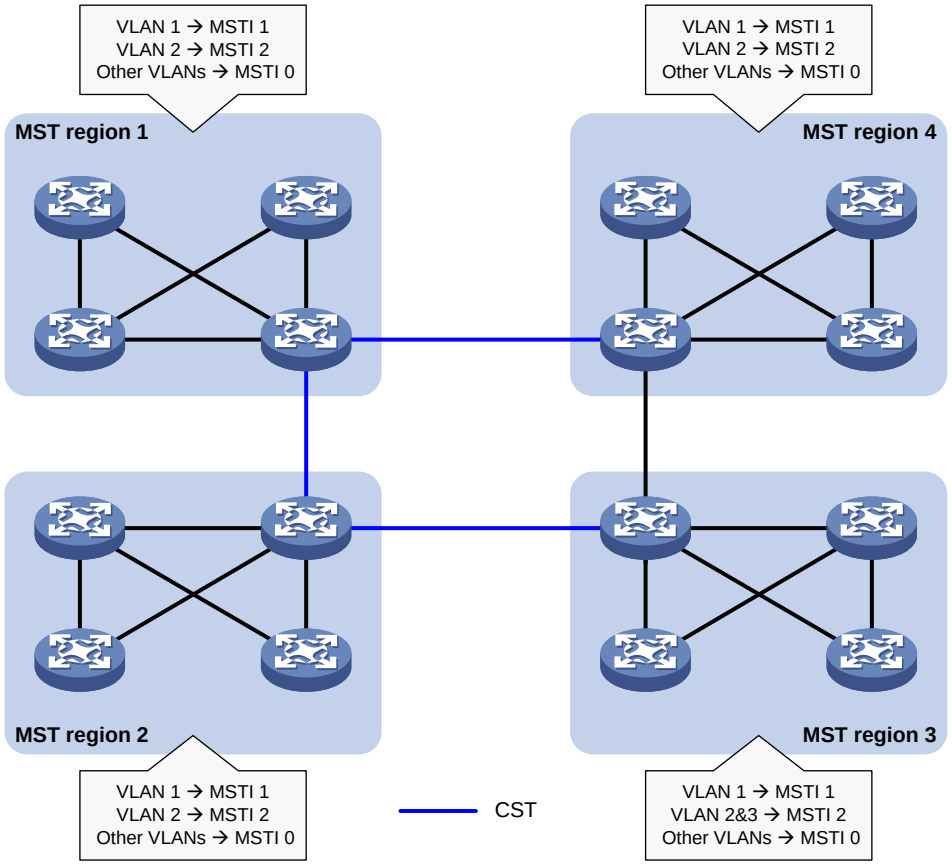
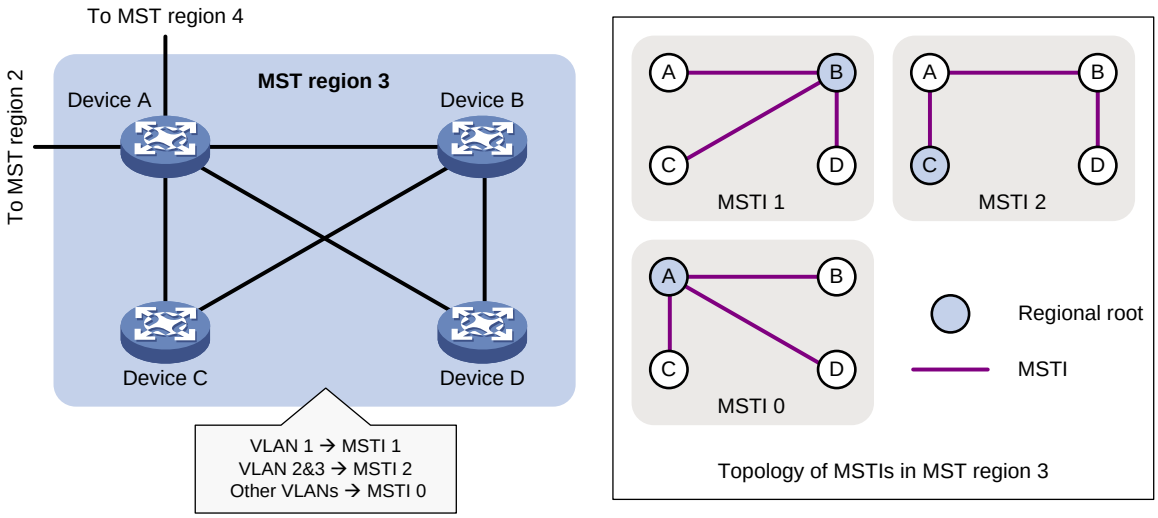


Figure 10 Network diagram and topology of MST region 3



MST region

A multiple spanning tree region (MST region) consists of multiple devices in a switched network and the network segments among them. All these devices have the following characteristics:

- A spanning tree protocol enabled
- Same region name
- Same VLAN-to-instance mapping configuration
- Same MSTP revision level
- Physically linked together

Multiple MST regions can exist in a switched network. You can assign multiple devices to the same MST region, as shown in Figure 9.

- The switched network contains four MST regions, MST region 1 through MST region 4.
- All devices in each MST region have the same MST region configuration.

MSTI

MSTP can generate multiple independent spanning trees in an MST region, and each spanning tree is mapped to the specific VLANs. Each spanning tree is referred to as a multiple spanning tree instance (MSTI).

In Figure 10, MST region 3 contains three MSTIs, MSTI 1, MSTI 2, and MSTI 0.

VLAN-to-instance mapping table

As an attribute of an MST region, the VLAN-to-instance mapping table describes the mapping relationships between VLANs and MSTIs.

In Figure 10, the VLAN-to-instance mapping table of MST region 3 is as follows:

- VLAN 1 to MSTI 1.
- VLAN 2 and VLAN 3 to MSTI 2.
- Other VLANs to MSTI 0.

MSTP achieves load balancing by means of the VLAN-to-instance mapping table.

CST

The common spanning tree (CST) is a single spanning tree that connects all MST regions in a switched network. If you regard each MST region as a device, the CST is a spanning tree calculated by these devices through STP or RSTP.

The blue lines in Figure 9 represent the CST.

IST

An internal spanning tree (IST) is a spanning tree that runs in an MST region. It is also called MSTI 0, a special MSTI to which all VLANs are mapped by default.

In Figure 9, MSTI 0 is the IST in MST region 3.

CIST

The common and internal spanning tree (CIST) is a single spanning tree that connects all devices in a switched network. It consists of the ISTs in all MST regions and the CST.

In Figure 9, the ISTs (MSTI 0) in all MST regions plus the inter-region CST constitute the CIST of the entire network.

Regional root

The root bridge of the IST or an MSTI within an MST region is the regional root of the IST or MSTI. Based on the topology, different spanning trees in an MST region might have different regional roots, as shown in MST region 3 in Figure 10.

- The regional root of MSTI 1 is Device B.
- The regional root of MSTI 2 is Device C.
- The regional root of MSTI 0 (also known as the IST) is Device A.

Common root bridge

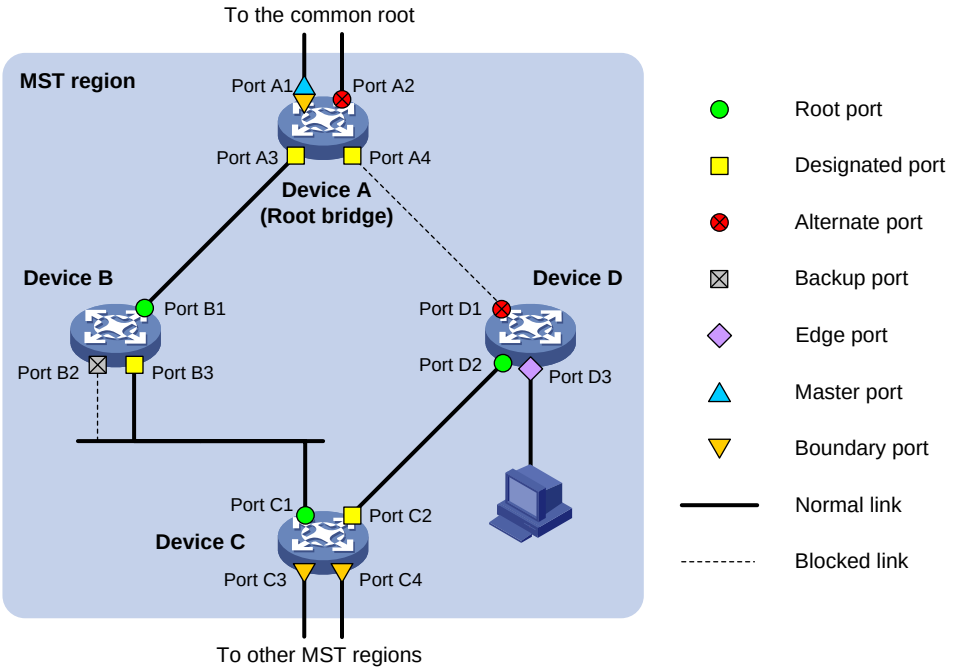
The common root bridge is the root bridge of the CIST.

In Figure 9, the common root bridge is a device in MST region 1.

Port roles

A port can play different roles in different MSTIs. As shown in Figure 11, an MST region contains Device A, Device B, Device C, and Device D. Port A1 and port A2 of Device A connect to the common root bridge. Port B2 and Port B3 of Device B form a loop. Port C3 and Port C4 of Device C connect to other MST regions. Port D3 of Device D directly connects to a host.

Figure 11 Port roles



MSTP calculation involves the following port roles:

- **Root port**—Forwards data for a non-root bridge to the root bridge. The root bridge does not have any root port.
- **Designated port**—Forwards data to the downstream network segment or device.
- **Alternate port**—Acts as the backup port for a root port or master port. When the root port or master port is blocked, the alternate port takes over.
- **Backup port**—Acts as the backup port of a designated port. When the designated port is invalid, the backup port becomes the new designated port. A loop occurs when two ports of the same spanning tree device are connected, so the device blocks one of the ports. The blocked port acts as the backup.
- **Edge port**—Directly connects to a user host rather than a network device or network segment.
- **Master port**—Acts as a port on the shortest path from the local MST region to the common root bridge. The master port is not always located on the regional root. It is a root port on the IST or CIST and still a master port on the other MSTIs.
- **Boundary port**—Connects an MST region to another MST region or to an STP/RSTP-running device. In MSTP calculation, a boundary port's role on an MSTI is consistent with its role on the

CIST. However, that is not true with master ports. A master port on MSTIs is a root port on the CIST.

Port states

In MSTP, a port can be in one of the following states:

- **Forwarding**—The port receives and sends BPDUs, learns MAC addresses, and forwards user traffic.
- **Learning**—The port receives and sends BPDUs, learns MAC addresses, but does not forward user traffic. Learning is an intermediate port state.
- **Discarding**—The port receives and sends BPDUs, but does not learn MAC addresses or forward user traffic.

NOTE:

When in different MSTIs, a port can be in different states.

A port state is not exclusively associated with a port role. Table 6 lists the port states that each port role supports. (A check mark [√] indicates that the port supports this state, while a dash [—] indicates that the port does not support this state.)

Table 6 Port states that different port roles support

Port role (right) Port state (below)	Root port/master port	Designated port	Alternate port	Backup port
Forwarding	√	√	—	—
Learning	√	√	—	—
Discarding	√	√	√	√

How MSTP works

MSTP divides an entire Layer 2 network into multiple MST regions, which are connected by a calculated CST. Inside an MST region, multiple spanning trees, called MSTIs, are calculated. Among these MSTIs, MSTI 0 is the IST.

Like STP, MSTP uses configuration BPDUs to calculate spanning trees. An important difference is that an MSTP BPDU carries the MSTP configuration of the bridge from which the BPDU is sent.

CIST calculation

During the CIST calculation, the following process takes place:

- The device with the highest priority is elected as the root bridge of the CIST.
- MSTP generates an IST within each MST region through calculation.
- MSTP regards each MST region as a single device and generates a CST among these MST regions through calculation.

The CST and ISTs constitute the CIST of the entire network.

MSTI calculation

Within an MST region, MSTP generates different MSTIs for different VLANs based on the VLAN-to-instance mappings. For each spanning tree, MSTP performs a separate calculation process similar to spanning tree calculation in STP. For more information, see "Calculation process of the STP algorithm."

In MSTP, a VLAN frame is forwarded along the following paths:

- Within an MST region, the frame is forwarded along the corresponding MSTI.
- Between two MST regions, the frame is forwarded along the CST.

MSTP implementation on devices

MSTP is compatible with STP and RSTP. Devices that are running MSTP and that are used for spanning tree calculation can identify STP and RSTP protocol frames.

In addition to basic MSTP features, the following features are provided for ease of management:

- Root bridge hold.
- Root bridge backup.
- Root guard.
- BPDU guard.
- Loop guard.
- TC-BPDU guard.
- Port role restriction.
- TC-BPDU transmission restriction.

Rapid transition mechanism

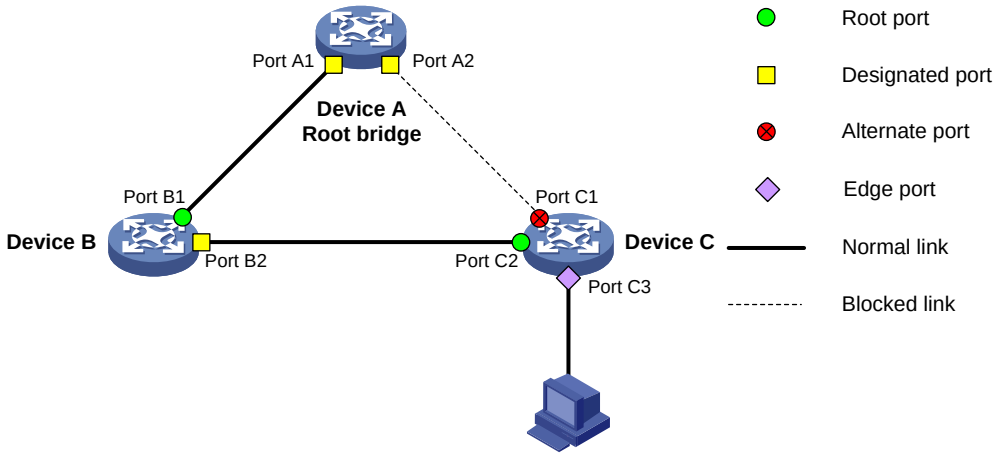
In STP, a port must wait twice the forward delay (30 seconds by default) before it transits from the blocking state to the forwarding state. The forward delay is related to the hello time and network diameter. If the forward delay is too short, loops might occur. This affects the stability of the network.

RSTP, PVST, and MSTP all use the rapid transition mechanism to speed up port state transition for edge ports, root ports, and designated ports. The rapid transition mechanism for designated ports is also known as the proposal/agreement (P/A)_transition.

Edge port rapid transition

As shown in Figure 12, Port C3 is an edge port connected to a host. When a network topology change occurs, the port can immediately transit from the blocking state to the forwarding state because no loop will be caused.

Figure 12 Edge port rapid transition

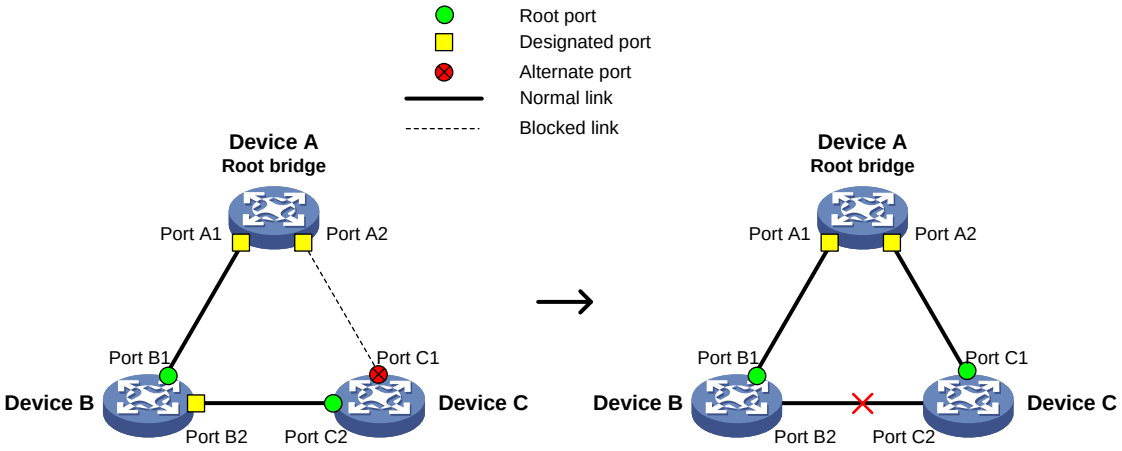


Root port rapid transition

When a root port is blocked, the bridge will elect the alternate port with the highest priority as the new root port. If the new root port's peer is in the forwarding state, the new root port immediately transits to the forwarding state.

As shown in Figure 13, Port C2 on Device C is a root port and Port C1 is an alternate port. When Port C2 transits to the blocking state, Port C1 is elected as the root port and immediately transits to the forwarding state.

Figure 13 Root port rapid transition



P/A transition

The P/A transition enables a designated port to rapidly transit to the forwarding state after a handshake with its peer. The P/A transition applies only to point-to-point links.

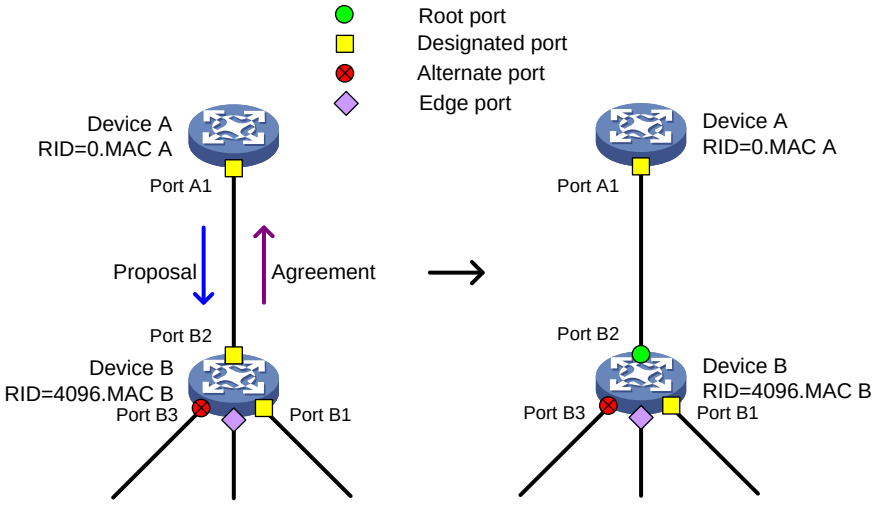
P/A transition for RSTP and PVST

In RSTP or PVST, the ports on a new link or recovered link are designated ports in blocking state. When one of the designated ports transits to the discarding or learning state, it sets the proposal flag in its BPDU. Its peer bridge receives the BPDU and determines whether the receiving port is the root port. If it is the root port, the bridge blocks the other ports except edge ports. The bridge then replies an agreement BPDU to the designated port. The designated port immediately transits to the forwarding state upon receiving the agreement BPDU. If the designated port does not receive the agreement BPDU, it waits for twice the forward delay to transit to the forwarding state.

As shown in Figure 14, the P/A transition operates as follows:

1. Device A sends a proposal BPDU to Device B through Port A1.
2. Device B receives the proposal BPDU on Port B2. Port B2 is elected as the root port.
3. Device B blocks its designated port Port B1 and alternate port Port B3 to eliminate loops.
4. The root port Port B2 transits to the forwarding state and sends an agreement BPDU to Device A.
5. The designated port Port A1 on Device A immediately transits to the forwarding state after receiving the agreement BPDU.

Figure 14 P/A transition for RSTP and PVST



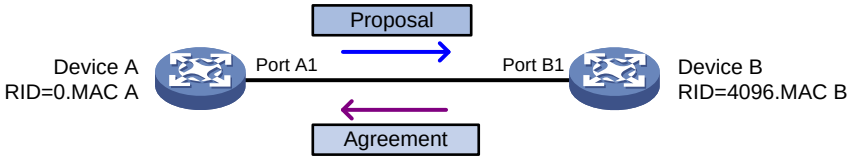
P/A transition for MSTP

In MSTP, an upstream bridge sets both the proposal and agreement flags in its BPDU. If a downstream bridge receives the BPDU and its receiving port is elected as the root port, the bridge blocks all the other ports except edge ports. The downstream bridge then replies an agreement BPDU to the upstream bridge. The upstream port immediately transits to the forwarding state upon receiving the agreement BPDU. If the upstream port does not receive the agreement BPDU, it waits for twice the forward delay to transit to the forwarding state.

As shown in Figure 15, the P/A transition operates as follows:

1. Device A sets the proposal and agreement flags in its BPDU and sends it to Device B through Port A1.
2. Device B receives the BPDU. Port B1 of Device B is elected as the root port.
3. Device B then blocks all its ports except the edge ports.
4. The root port Port B1 of Device B transits to the forwarding state and sends an agreement BPDU to Device A.
5. Port A1 of Device A immediately transits to the forwarding state upon receiving the agreement BPDU.

Figure 15 P/A transition for MSTP



Protocols and standards

MSTP is documented in the following protocols and standards:

- IEEE 802.1d, *Media Access Control (MAC) Bridges*
- IEEE 802.1w, *Part 3: Media Access Control (MAC) Bridges—Amendment 2: Rapid Reconfiguration*
- IEEE 802.1s, *Virtual Bridged Local Area Networks—Amendment 3: Multiple Spanning Trees*

- IEEE 802.1Q-REV/D1.3, *Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks —Clause 13: Spanning tree Protocols*

Configuring spanning tree protocols

Restrictions and guidelines: spanning tree protocol configuration

Restrictions: Compatibility with other features

- When the spanning tree protocol is enabled for an M-LAG system, following these restrictions and guidelines:
  - Make sure the M-LAG member devices have the same spanning tree configuration, including:
    - Global spanning tree configuration.
    - Spanning tree configuration on the peer-link interface.
    - Spanning tree configuration on M-LAG interfaces.

    Violation of this rule might cause network flapping. Peer-link interfaces in the M-LAG system do not participate in spanning tree calculation.
  - If spanning tree is enabled on an M-LAG system, the M-LAG member devices still use the M-LAG system MAC address after the M-LAG system splits, which will cause spanning tree calculation issues. To avoid the issues, enable M-LAG standalone mode on the M-LAG member devices before the M-LAG system splits.

  For more information about the M-LAG system, M-LAG interfaces, peer-link interfaces, and M-LAG standalone mode, see "Configuring M-LAG."
- If both MVRP and a spanning tree protocol are enabled on a device, MVRP packets are forwarded along MSTIs. To advertise a specific VLAN within the network through MVRP, make sure this VLAN is mapped to an MSTI when you configure the VLAN-to-instance mapping table. For more information about MVRP, see "Configuring MVRP."
- The spanning tree configurations are mutually exclusive with any of the following features on a port: RRPP and Smart Link

Restrictions: Interface configuration

- Some spanning tree features are supported in Layer 2 Ethernet interface view and Layer 2 aggregate interface view. Unless otherwise stated, these views are collectively referred to as interface view in this document. BPDU drop can be configured only in Layer 2 Ethernet interface view.
- Configurations made in system view take effect globally. Configurations made in Layer 2 Ethernet interface view take effect only on the interface. Configurations made in Layer 2 aggregate interface view take effect only on the aggregate interface. Configurations made on an aggregation member port can take effect only after the port is removed from the aggregation group.
- After you enable a spanning tree protocol on a Layer 2 aggregate interface, the system performs spanning tree calculation on the Layer 2 aggregate interface. It does not perform spanning tree calculation on the aggregation member ports. The spanning tree protocol enable state and forwarding state of each selected member port are consistent with those of the corresponding Layer 2 aggregate interface.
- The member ports of an aggregation group do not participate in spanning tree calculation. However, the ports still reserve their spanning tree configurations for participating in spanning tree calculation after leaving the aggregation group.

- In an M-LAG system, you can use the **display** command to view the spanning tree information of M-LAG interfaces on both the primary M-LAG device and the member M-LAG devices.

Spanning tree protocol tasks at a glance

STP tasks at a glance

Configuring the root bridge

To configure the root bridge in STP mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to STP.
3. (Optional.) Configuring the root bridge or a secondary root bridge
4. (Optional.) Configuring the device priority
5. (Optional.) Configuring parameters that affects STP topology convergence
   - Configuring the network diameter of a switched network
   - Setting spanning tree timers
   - Setting the timeout factor
   - Configuring the BPDU transmission rate
6. (Optional.) Enabling outputting port state transition information
7. Enabling the spanning tree feature
8. (Optional.) Configuring advanced spanning tree features
   - Ignoring the specified VLANs
   - Configuring TC Snooping
   - Configuring protection features
   - Disabling the device from reactivating edge ports shut down by BPDU guard
   - Enabling BPDU transparent transmission on a port
   - Configuring TC Snooping
   - Configuring spanning tree blackhole detection
   - Enabling SNMP notifications for new-root election and topology change events

Configuring the leaf nodes

To configure the leaf nodes in STP mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to STP.
3. (Optional.) Configuring the device priority
4. (Optional.) Configuring parameters that affects STP topology convergence
   - Setting the timeout factor
   - Configuring the BPDU transmission rate
   - Configuring path costs of ports
   - Configuring the port priority
5. (Optional.) Enabling outputting port state transition information
6. Enabling the spanning tree feature
7. (Optional.) Configuring advanced spanning tree features
   - Ignoring the specified VLANs

- Configuring TC Snooping
- Configuring protection features
- Disabling the device from reactivating edge ports shut down by BPDU guard
- Enabling BPDU transparent transmission on a port
- Configuring TC groups
- Configuring spanning tree blackhole detection
- Enabling SNMP notifications for new-root election and topology change events

RSTP tasks at a glance

Configuring the root bridge

To configure the root bridge in RSTP mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to RSTP.
3. (Optional.) Configuring the root bridge or a secondary root bridge
4. (Optional.) Configuring the device priority
5. (Optional.) Configuring parameters that affects RSTP topology convergence
   - Configuring the network diameter of a switched network
   - Setting spanning tree timers
   - Setting the timeout factor
   - Configuring the BPDU transmission rate
   - Configuring edge ports
   - Configuring the port link type
6. (Optional.) Enabling outputting port state transition information
7. Enabling the spanning tree feature
8. (Optional.) Configuring advanced spanning tree features
   - Performing mCheck
   - Ignoring the specified VLANs
   - Configuring TC Snooping
   - Configuring protection features
   - Disabling the device from reactivating edge ports shut down by BPDU guard
   - Enabling BPDU transparent transmission on a port
   - Configuring TC groups
   - Configuring spanning tree blackhole detection
   - Enabling SNMP notifications for new-root election and topology change events

Configuring the leaf nodes

To configure the leaf nodes in RSTP mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to RSTP.
3. (Optional.) Configuring the device priority
4. (Optional.) Configuring parameters that affects RSTP topology convergence
   - Setting the timeout factor
   - Configuring the BPDU transmission rate
   - Configuring edge ports

- Configuring path costs of ports
- Configuring the port priority
- Configuring the port link type

5. (Optional.) Enabling outputting port state transition information
6. Enabling the spanning tree feature
7. (Optional.) Configuring advanced spanning tree features
   - Performing mCheck
   - Ignoring the specified VLANs
   - Configuring TC Snooping
   - Configuring protection features
   - Disabling the device from reactivating edge ports shut down by BPDU guard
   - Enabling BPDU transparent transmission on a port
   - Configuring TC groups
   - Configuring spanning tree blackhole detection
   - Enabling SNMP notifications for new-root election and topology change events

PVST tasks at a glance

Configuring the root bridge

To configure the root bridge in PVST mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to PVST.
3. (Optional.) Configuring the root bridge or a secondary root bridge
4. (Optional.) Configuring the device priority
5. (Optional.) Configuring parameters that affects PVST topology convergence
   - Configuring the network diameter of a switched network
   - Setting spanning tree timers
   - Setting the timeout factor
   - Configuring the BPDU transmission rate
   - Configuring edge ports
   - Configuring the port link type
6. (Optional.) Enabling outputting port state transition information
7. Enabling the spanning tree feature
8. (Optional.) Configuring advanced spanning tree features
   - Performing mCheck
   - Disabling inconsistent PVID protection
   - Ignoring the specified VLANs
   - Configuring protection features
   - Enabling the device to log events of detecting or receiving TC BPDUs
   - Disabling the device from reactivating edge ports shut down by BPDU guard
   - Enabling BPDU transparent transmission on a port
   - Configuring TC groups
   - Configuring spanning tree blackhole detection
   - Enabling SNMP notifications for new-root election and topology change events

Configuring the leaf nodes

To configure the leaf nodes in PVST mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to PVST.
3. (Optional.) Configuring the device priority
4. (Optional.) Configuring parameters that affects PVST topology convergence
   - Setting the timeout factor
   - Configuring the BPDU transmission rate
   - Configuring edge ports
   - Configuring path costs of ports
   - Configuring the port priority
   - Configuring the port link type
5. (Optional.) Enabling outputting port state transition information
6. Enabling the spanning tree feature
7. (Optional.) Configuring advanced spanning tree features
   - Performing mCheck
   - Disabling inconsistent PVID protection
   - Ignoring the specified VLANs
   - Configuring protection features
   - Enabling the device to log events of detecting or receiving TC BPDUs
   - Disabling the device from reactivating edge ports shut down by BPDU guard
   - Enabling BPDU transparent transmission on a port
   - Configuring TC groups
   - Configuring spanning tree blackhole detection
   - Enabling SNMP notifications for new-root election and topology change events

MSTP tasks at a glance

Configuring the root bridge

To configure the root bridge in MSTP mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to MSTP.
3. Configuring an MST region
4. (Optional.) Configuring the root bridge or a secondary root bridge
5. (Optional.) Configuring the device priority
6. (Optional.) Configuring parameters that affects MSTP topology convergence
   - Configuring the maximum hops of an MST region
   - Configuring the network diameter of a switched network
   - Setting spanning tree timers
   - Setting the timeout factor
   - Configuring the BPDU transmission rate
   - Configuring edge ports
   - Configuring the port link type
7. (Optional.) Configuring the mode a port uses to recognize and send MSTP frames

8. (Optional.) Enabling outputting port state transition information
9. Enabling the spanning tree feature
10. (Optional.) Configuring advanced spanning tree features
    - Performing mCheck
    - Ignoring the specified VLANs
    - Configuring Digest Snooping
    - Configuring No Agreement Check
    - Configuring TC Snooping
    - Configuring protection features
    - Disabling the device from reactivating edge ports shut down by BPDU guard
    - Enabling BPDU transparent transmission on a port
    - Configuring TC groups
    - Configuring spanning tree blackhole detection
    - Enabling SNMP notifications for new-root election and topology change events

Configuring the leaf nodes

To configure the leaf nodes in MSTP mode, perform the following tasks:

1. Setting the spanning tree mode
2. Set the spanning tree mode to MSTP.
3. Configuring an MST region
4. (Optional.) Configuring the device priority
5. (Optional.) Configuring parameters that affects MSTP topology convergence
    - Setting the timeout factor
    - Configuring the BPDU transmission rate
    - Configuring edge ports
    - Configuring path costs of ports
    - Configuring the port priority
    - Configuring the port link type
6. (Optional.) Configuring the mode a port uses to recognize and send MSTP frames
7. (Optional.) Enabling outputting port state transition information
8. Enabling the spanning tree feature
9. (Optional.) Configuring advanced spanning tree features
    - Performing mCheck
    - Ignoring the specified VLANs
    - Configuring Digest Snooping
    - Configuring No Agreement Check
    - Configuring TC Snooping
    - Configuring protection features
    - Disabling the device from reactivating edge ports shut down by BPDU guard
    - Enabling BPDU transparent transmission on a port
    - Configuring TC groups
    - Configuring spanning tree blackhole detection
    - Enabling SNMP notifications for new-root election and topology change events

Setting the spanning tree mode

About this task

The spanning tree modes include:

- **STP mode**—All ports of the device send STP BPDUs. Select this mode when the peer device of a port supports only STP.
- **RSTP mode**—All ports of the device send RSTP BPDUs. A port in this mode automatically transits to the STP mode when it receives STP BPDUs from the peer device. A port in this mode does not transit to the MSTP mode when it receives MSTP BPDUs from the peer device.
- **PVST mode**—All ports of the device send PVST BPDUs. Each VLAN maintains a spanning tree. In a network, the amount of spanning trees maintained by all devices equals the number of PVST-enabled VLANs multiplied by the number of PVST-enabled ports. If the amount of spanning trees exceeds the capacity of the network, device CPUs will be overloaded. Packet forwarding is interrupted, and the network becomes unstable. The device can maintain spanning trees for 128 VLANs.
- **MSTP mode**—All ports of the device send MSTP BPDUs. A port in this mode automatically transits to the STP mode when receiving STP BPDUs from the peer device. A port in this mode does not transit to the RSTP mode when receiving RSTP BPDUs from the peer device.

Restrictions and guidelines

The MSTP mode is compatible with the RSTP mode, and the RSTP mode is compatible with the STP mode.

Compatibility of the PVST mode depends on the link type of a port.

- On an access port, the PVST mode is compatible with other spanning tree modes in all VLANs.
- On a trunk port or hybrid port, the PVST mode is compatible with other spanning tree modes only in the default VLAN.

Procedure

1. Enter system view.

   `system-view`
2. Set the spanning tree mode.

   `stp mode` { **`mstp`** | **`pvst`** | **`rstp`** | **`stp`** }

   The default setting is the MSTP mode.

Configuring an MST region

About this task

Spanning tree devices belong to the same MST region if they are both connected through a physical link and configured with the following details:

- Format selector (0 by default, not configurable).
- MST region name.
- MST region revision level.
- VLAN-to-instance mapping entries in the MST region.

The configuration of MST region-related parameters (especially the VLAN-to-instance mapping table) might cause MSTP to begin a new spanning tree calculation. To reduce the possibility of topology instability, the MST region configuration takes effect only after you activate it by doing one of the following:

- Use the **`active region-configuration`** command.

- Enable a spanning tree protocol by using the **stp global enable** command if the spanning tree protocol is disabled.

Restrictions and guidelines

In STP, RSTP, or PVST mode, MST region configurations do not take effect.

Procedure

1. Enter system view.
   system-view
2. Enter MST region view.
   stp region-configuration
3. Configure the MST region name.
   region-name *name*
   The default setting is the MAC address.
4. Configure the VLAN-to-instance mapping table. Choose one option as needed:
   - Map a list of VLANs to an MSTI.
   - **instance** *instance-id* **vlan** *vlan-id-list*
   - Quickly create a VLAN-to-instance mapping table.
   - **vlan-mapping modulo** *modulo*

   By default, all VLANs in an MST region are mapped to the CIST (or MSTI 0).
5. Configure the MSTP revision level of the MST region.
   revision-level *level*
   The default setting is 0.
6. (Optional.) Display the MST region configurations that are not activated yet.
   check region-configuration
7. Manually activate MST region configuration.
   active region-configuration

Configuring the root bridge or a secondary root bridge

Restrictions and guidelines

You can have the spanning tree protocol determine the root bridge of a spanning tree through calculation. You can also specify a device as the root bridge or as a secondary root bridge.

When you specify a device as the root bridge or as a secondary root bridge, follow these restrictions and guidelines:

- A device has independent roles in different spanning trees. It can act as the root bridge in one spanning tree and as a secondary root bridge in another. However, one device cannot be the root bridge and a secondary root bridge in the same spanning tree.
- If you specify the root bridge for a spanning tree, no new root bridge is elected according to the device priority settings. Once you specify a device as the root bridge or a secondary root bridge, you cannot change the priority of the device.
- You can configure a device as the root bridge by setting the device priority to 0. For the device priority configuration, see "Configuring the device priority."

Configuring the device as the root bridge of a spanning tree

1. Enter system view.
   system-view
2. Configure the device as the root bridge.
   - In STP/RSTP mode:
     stp root primary
   - In PVST mode:
     stp vlan *vlan-id-list* **root primary**
   - In MSTP mode:
     stp [**instance** *instance-list*] **root primary**

   By default, the device is not a root bridge.

Configuring the device as a secondary root bridge of a spanning tree

1. Enter system view.
   system-view
2. Configure the device as a secondary root bridge.
   - In STP/RSTP mode:
     stp root secondary
   - In PVST mode:
     stp vlan *vlan-id-list* **root secondary**
   - In MSTP mode:
     stp [**instance** *instance-list*] **root secondary**

   By default, the device is not a secondary root bridge.

Configuring the device priority

About this task

Device priority is a factor in calculating the spanning tree. The priority of a device determines whether the device can be elected as the root bridge of a spanning tree. A lower value indicates a higher priority. You can set the priority of a device to a low value to specify the device as the root bridge of the spanning tree. A spanning tree device can have different priorities in different spanning trees.

During root bridge selection, if all devices in a spanning tree have the same priority, the one with the lowest MAC address is selected. You cannot change the priority of a device after it is configured as the root bridge or as a secondary root bridge.

Procedure

1. Enter system view.
   system-view
2. Configure the priority of the device.
   - In STP/RSTP mode:
     stp priority *priority*

- In PVST mode:
  stp vlan *vlan-id-list* **priority** *priority*
- In MSTP mode:
  stp [**instance** *instance-list*] **priority** *priority*

The default setting is 32768.

Configuring the maximum hops of an MST region

About this task

Restrict the region size by setting the maximum hops of an MST region. The hop limit configured on the regional root bridge is used as the hop limit for the MST region.

Configuration BPDUs sent by the regional root bridge always have a hop count set to the maximum value. When a device receives this configuration BPDU, it decrements the hop count by one, and uses the new hop count in the BPDUs that it propagates. When the hop count of a BPDU reaches zero, it is discarded by the device that received it. Devices beyond the reach of the maximum hops can no longer participate in spanning tree calculations, so the size of the MST region is limited.

Restrictions and guidelines

Make this configuration only on the root bridge. All other devices in the MST region use the maximum hop value set for the root bridge.

You can configure the maximum hops of an MST region based on the STP network size. As a best practice, set the maximum hops to a value that is greater than the maximum hops of each edge device to the root bridge.

Procedure

1. Enter system view.
   system-view
2. Configure the maximum hops of the MST region.
   stp max-hops *hops*
   The default setting is 20.

Configuring the network diameter of a switched network

About this task

Any two terminal devices in a switched network can reach each other through a specific path, and there are a series of devices on the path. The switched network diameter is the maximum number of devices on the path for an edge device to reach another one in the switched network through the root bridge. The network diameter indicates the network size. The bigger the diameter, the larger the network size.

Based on the network diameter you configured, the system automatically sets an optimal hello time, forward delay, and max age for the device.

In STP, RSTP, or MSTP mode, each MST region is considered a device. The configured network diameter takes effect only on the CIST (or the common root bridge) but not on other MSTIs.

In PVST mode, the configured network diameter takes effect only on the root bridges of the specified VLANs.

Procedure

1. Enter system view.
   system-view
2. Configure the network diameter of the switched network.
   - In STP/RSTP/MSTP mode:
     stp bridge-diameter *diameter*
   - In PVST mode:
     stp vlan *vlan-id-list* **bridge-diameter** *diameter*

   The default setting is 7.

Setting spanning tree timers

About this task

The following timers are used for spanning tree calculation:

- **Forward delay**—Delay time for port state transition. To prevent temporary loops on a network, the spanning tree feature sets an intermediate port state (the learning state) before it transits from the discarding state to the forwarding state. The feature also requires that the port transit its state after a forward delay timer. This ensures that the state transition of the local port stays synchronized with the peer.
- **Hello time**—Interval at which the device sends configuration BPDUs to detect link failures. If the device does not receive configuration BPDUs within the timeout period, it recalculates the spanning tree. The formula for calculating the timeout period is timeout period = timeout factor × 3 × hello time.
- **Max age**—In the CIST of an MSTP network, the device uses the max age timer to determine whether a configuration BPDU received by a port has expired. If it is expired, a new spanning tree calculation process starts. The max age timer does not take effect on MSTIs.

To ensure a fast topology convergence, make sure the timer settings meet the following formulas:

- 2 × (forward delay – 1 second) ≥ max age
- Max age ≥ 2 × (hello time + 1 second)

As a best practice, specify the network diameter and letting spanning tree protocols automatically calculate the timers based on the network diameter instead of manually setting the spanning tree timers. If the network diameter uses the default value, the timers also use their default values.

Set the timers only on the root bridge. The timer settings on the root bridge apply to all devices on the entire switched network.

Restrictions and guidelines

- The length of the forward delay is related to the network diameter of the switched network. The larger the network diameter is, the longer the forward delay time should be. As a best practice, use the automatically calculated value because inappropriate forward delay setting might cause temporary redundant paths or increase the network convergence time.
- An appropriate hello time setting enables the device to promptly detect link failures on the network without using excessive network resources. If the hello time is too long, the device mistakes packet loss for a link failure and triggers a new spanning tree calculation process. If the hello time is too short, the device frequently sends the same configuration BPDUs, which wastes device and network resources. As a best practice, use the automatically calculated value.
- If the max age timer is too short, the device frequently begins spanning tree calculations and might mistake network congestion as a link failure. If the max age timer is too long, the device might fail to promptly detect link failures and quickly launch spanning tree calculations, reducing

the auto-sensing capability of the network. As a best practice, use the automatically calculated value.

Procedure

1. Enter system view.
   system-view
2. Set the forward delay timer.
   - In STP/RSTP/MSTP mode:
     stp timer forward-delay *time*
   - In PVST mode:
     stp vlan *vlan-id-list* **timer forward-delay** *time*

   The default setting is 15 seconds.
3. Set the hello timer.
   - In STP/RSTP/MSTP mode:
     stp timer hello *time*
   - In PVST mode:
     stp vlan *vlan-id-list* **timer hello** *time*

   The default setting is 2 seconds.
4. Set the max age timer.
   - In STP/RSTP/MSTP mode:
     stp timer max-age *time*
   - In PVST mode:
     stp vlan *vlan-id-list* **timer max-age** *time*

     The default setting is 20 seconds.

Setting the timeout factor

About this task

The timeout factor is a parameter used to decide the timeout period. The formula for calculating the timeout period is: *timeout period* = *timeout factor* × 3 × *hello time*.

In a stable network, each non-root-bridge device forwards configuration BPDUs to the downstream devices at the hello time interval to detect link failures. If a device does not receive a BPDU from the upstream device within nine times the hello time, it assumes that the upstream device has failed. Then, it starts a new spanning tree calculation process.

Restrictions and guidelines

As a best practice, set the timeout factor to 5, 6, or 7 in the following situations:

- To prevent undesired spanning tree calculations. An upstream device might be too busy to forward configuration BPDUs in time, for example, many Layer 2 interfaces are configured on the upstream device. In this case, the downstream device fails to receive a BPDU within the timeout period and then starts an undesired spanning tree calculation.
- To save network resources on a stable network.

Procedure

1. Enter system view.
   system-view
2. Set the timeout factor of the device.

stp timer-factor *factor*

The default setting is 3.

Configuring the BPDU transmission rate

About this task

The maximum number of BPDUs a port can send within each hello time equals the BPDU transmission rate plus the hello timer value.

The higher the BPDU transmission rate, the more BPDUs are sent within each hello time, and the more system resources are used. By setting an appropriate BPDU transmission rate, you can limit the rate at which the port sends BPDUs. Setting an appropriate rate also prevents spanning tree protocols from using excessive network resources when the network topology changes.

Restrictions and guidelines

The BPDU transmission rate depends on the physical status of the port and the network structure. As a best practice, use the default setting.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the BPDU transmission rate of the ports.

   stp transmit-limit *limit*

   The default setting is 10.

Configuring edge ports

About this task

If a port directly connects to a user terminal rather than another device or a shared LAN segment, this port is regarded as an edge port. When network topology change occurs, an edge port will not cause a temporary loop. Because a device does not determine whether a port is directly connected to a terminal, you must manually configure the port as an edge port. After that, the port can rapidly transit from the blocking state to the forwarding state.

Restrictions and guidelines

- If BPDU guard is disabled on a port configured as an edge port, the port becomes a non-edge port again if it receives a BPDU from another port. To restore the edge port, re-enable it.
- If a port directly connects to a user terminal, configure it as an edge port and enable BPDU guard for it. This enables the port to quickly transit to the forwarding state when ensuring network security.
- After spanning tree is enabled on an interface, the system checks whether the interface receives BPDUs within twice the hello time plus 1 second. If no BPDUs are received, the device automatically assigns the edge port role to the interface. The edge port role assigned by using the **stp edged-port** command takes precedence over the system assigned edge port role.
- On a port, the loop guard feature and the edge port setting are mutually exclusive.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the port as an edge port.
   stp edged-port
   By default, all ports are non-edge ports.

Configuring path costs of ports

About path cost

Path cost is a parameter related to the link speed of a port. On a spanning tree device, a port can have different path costs in different MSTIs. Setting appropriate path costs allows VLAN traffic flows to be forwarded along different physical links, achieving VLAN-based load balancing.

You can have the device automatically calculate the default path cost, or you can configure the path cost for ports.

Specifying a standard for the default path cost calculation

About this task

You can specify a standard for the device to use in automatic calculation for the default path cost. The device supports the following standards:

- **dot1d-1998**—The device calculates the default path cost for ports based on IEEE 802.1d-1998.
- **dot1t**—The device calculates the default path cost for ports based on IEEE 802.1t.
- **legacy**—The device calculates the default path cost for ports based on a private standard.

Table 7 Mappings between the link speed (100M and below) and the path cost

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
0	N/A	65535	200000000	200000
10 Mbps	Single port	100	2000000	2000
	Aggregate interface containing two Selected ports		1000000	1800
	Aggregate interface containing three Selected ports		666666	1600
	Aggregate interface containing four Selected ports		500000	1400
100 Mbps	Single port	19	200000	200
	Aggregate interface containing two Selected ports		100000	180
	Aggregate interface containing three Selected ports		66666	160

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
	Aggregate interface containing four Selected ports		50000	140

Table 8 Mappings between the link speed (1000M) and the path cost

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
1000 Mbps	Single port	4	20000	20
	Aggregate interface containing two Selected ports		10000	18
	Aggregate interface containing three Selected ports		6666	16
	Aggregate interface containing four Selected ports		5000	14

Table 9 Mappings between the link speed (10G) and the path cost

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
10 Gbps	Single port	2	2000	2
	Aggregate interface containing two Selected ports		1000	1
	Aggregate interface containing three Selected ports		666	1
	Aggregate interface containing four Selected ports		500	1

Table 10 Mappings between the link speed (20G) and the path cost

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
20 Gbps	Single port	1	1000	1
	Aggregate interface containing two Selected ports		500	1
	Aggregate interface containing three Selected		333	1

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
	ports			
	Aggregate interface containing four Selected ports		250	1

Table 11 Mappings between the link speed (25G) and the path cost

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
25 Gbps	Single port	1	800	1
	Aggregate interface containing two Selected ports		400	1
	Aggregate interface containing three Selected ports		266	1
	Aggregate interface containing four Selected ports		200	1

Table 12 Mappings between the link speed (40G) and the path cost

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
40 Gbps	Single port	1	500	1
	Aggregate interface containing two Selected ports		250	1
	Aggregate interface containing three Selected ports		166	1
	Aggregate interface containing four Selected ports		125	1

Table 13 Mappings between the link speed (100G) and the path cost

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
100 Gbps	Single port	1	200	1
	Aggregate interface containing two Selected ports		100	1
	Aggregate interface		66	1

Link speed	Port type	Path cost		
		IEEE 802.1d-1998	IEEE 802.1t	Private standard
	containing three Selected ports			
	Aggregate interface containing four Selected ports		50	1

Restrictions and guidelines

If you change the standard for the default path cost calculation, you restore the path costs to the default.

When the device calculates the path cost for an aggregate interface, IEEE 802.1t takes into account the number of Selected ports in its aggregation group. However, IEEE 802.1d-1998 does not take into account the number of Selected ports. The calculation formula of IEEE 802.1t is: Path cost = 200,000,000/link speed (in 100 kbps). The link speed is the sum of the link speed values of the Selected ports in the aggregation group.

IEEE 802.1d-1998 or the private standard always assigns the smallest possible value to a single port or aggregate interface with a speed exceeding 10 Gbps. The forwarding path selected based on this criterion might not be the best one. To solve this problem, perform one of the following tasks:

- Use **dot1t** as the standard for default path cost calculation.
- Manually set the path cost for the port (see "Configuring path costs of ports").

Procedure

1. Enter system view.
   system-view
2. Specify a standard for the default path costs calculation.
   stp pathcost-standard { **dot1d-1998** | **dot1t** | **legacy** }
   By default, the device uses **legacy** to calculate the default path costs of its ports.

Configuring path costs of ports

Restrictions and guidelines

When the path cost of a port changes, the system recalculates the port role and initiates a state transition.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the path cost of the ports.
   - In STP/RSTP mode:
     stp cost *cost-value*
   - In PVST mode:
     stp vlan *vlan-id-list* **cost** *cost-value*
   - In MSTP mode:
     stp [**instance** *instance-list*] **cost** *cost-value*

By default, the system automatically calculates the path cost of each port.

Configuring the port priority

About this task

The priority of a port is a factor that determines whether the port can be elected as the root port of a device. If all other conditions are the same, the port with the highest priority is elected as the root port.

On a spanning tree device, a port can have different priorities and play different roles in different spanning trees. As a result, data of different VLANs can be propagated along different physical paths, implementing per-VLAN load balancing. You can set port priority values based on the actual networking requirements.

Restrictions and guidelines

When the priority of a port changes, the system recalculates the port role and initiates a state transition. Prepare for the network topology change before configuring the port priority.

Procedure

1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Configure the port priority.
   - In STP/RSTP mode:
     `stp port priority` *`priority`*
   - In PVST mode:
     `stp vlan` *`vlan-id-list`* **`port priority`** *`priority`*
   - In MSTP mode:
     `stp` [**`instance`** *`instance-list`*] **`port priority`** *`priority`*

   The default setting is 128 for all ports.

Configuring the port link type

About this task

A point-to-point link directly connects two devices. If two root ports or designated ports are connected over a point-to-point link, they can rapidly transit to the forwarding state after a proposal-agreement handshake process.

Restrictions and guidelines

- You can configure the link type as point-to-point for a Layer 2 aggregate interface or a port that operates in full duplex mode. As a best practice, use the default setting and let the device automatically detect the port link type.
- In PVST or MSTP mode, the **`stp point-to-point force-false`** or **`stp point-to-point force-true`** command configured on a port takes effect on all VLANs or all MSTIs.
- Before you set the link type of a port to point-to-point, make sure the port is connected to a point-to-point link. Otherwise, a temporary loop might occur.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the port link type.
   stp point-to-point { **auto** | **force-false** | **force-true** }
   By default, the link type is **auto** where the port automatically detects the link type.

Configuring the mode a port uses to recognize and send MSTP frames

About this task

A port can receive and send MSTP frames in the following formats:

- **dot1s**—802.1s-compliant standard format
- **legacy**—Compatible format

By default, the frame format recognition mode of a port is **auto**. The port automatically distinguishes the two MSTP frame formats, and determines the format of frames that it will send based on the recognized format.

You can configure the MSTP frame format on a port. Then, the port sends only MSTP frames of the configured format to communicate with devices that send frames of the same format.

By default, a port in **auto** mode sends 802.1s MSTP frames. When the port receives an MSTP frame of a legacy format, the port starts to send frames only of the legacy format. This prevents the port from frequently changing the format of sent frames. To configure the port to send 802.1s MSTP frames, shut down and then bring up the port.

Restrictions and guidelines

When the number of existing MSTIs exceeds 48, the port can send only 802.1s MSTP frames.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the mode that the port uses to recognize/send MSTP frames.
   stp compliance { **auto** | **dot1s** | **legacy** }
   The default setting is **auto**.

Enabling outputting port state transition information

About this task

In a large-scale spanning tree network, you can enable devices to output the port state transition information. Then, you can monitor the port states in real time.

Procedure

1. Enter system view.
   system-view
2. Enable outputting port state transition information.
   - In STP/RSTP mode:
     stp port-log instance 0
   - In PVST mode:
     stp port-log vlan *vlan-id-list*
   - In MSTP mode:
     stp port-log { **all** | **instance** *instance-list* }

   When the device starts up with initial settings, outputting port state transition information is disabled by default.

   When the device starts up with factory defaults, outputting port state transition information is enabled by default.

   For more information about the initial settings and factory defaults, see *Fundamentals Configuration Guide*.

Enabling the spanning tree feature

Restrictions and guidelines

You must enable the spanning tree feature for the device before any other spanning tree related configurations can take effect. In STP, RSTP, or MSTP mode, make sure the spanning tree feature is enabled globally and on the desired ports. In PVST mode, make sure the spanning tree feature is enabled globally, in the desired VLANs, and on the desired ports.

To exclude specific ports from spanning tree calculation and save CPU resources, disable the spanning tree feature for these ports with the **undo stp enable** command. Make sure no loops occur in the network after you disable the spanning tree feature on these ports.

Enabling the spanning tree feature in STP/RSTP/MSTP mode

1. Enter system view.
   system-view
2. Enable the spanning tree feature.
   stp global enable

   When the device starts up with initial settings, the spanning tree feature is globally disabled by default.

   When the device starts up with factory defaults, the spanning tree feature is globally enabled by default.

   For more information about the initial settings and factory defaults, see *Fundamentals Configuration Guide*.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable the spanning tree feature for the port.
   stp enable

By default, the spanning tree feature is enabled on all ports.

Enabling the spanning tree feature in PVST mode

1. Enter system view.

   system-view

2. Enable the spanning tree feature.

   stp global enable

   When the device starts up with initial settings, the spanning tree feature is globally disabled by default.

   When the device starts up with factory defaults, the spanning tree feature is globally enabled by default.

   For more information about the initial settings and factory defaults, see *Fundamentals Configuration Guide*.

3. Enable the spanning tree feature in VLANs.

   stp vlan *vlan-id-list* **enable**

   By default, the spanning tree feature is enabled in VLANs.

4. Enter interface view.

   interface *interface-type interface-number*

5. Enable the spanning tree feature on the port.

   stp enable

   By default, the spanning tree feature is enabled on all ports.

Performing mCheck

About mCheck

The mCheck feature enables user intervention in the port state transition process.

When a port on an MSTP, RSTP, or PVST device connects to an STP device and receives STP BPDUs, the port automatically transits to the STP mode. However, the port cannot automatically transit back to the original mode when the following conditions exist:

- The peer STP device is shut down or removed.
- The port cannot detect the change.

To forcibly transit the port to operate in the original mode, you can perform an mCheck operation.

For example, Device A, Device B, and Device C are connected in sequence. Device A runs STP, Device B does not run any spanning tree protocol, and Device C runs RSTP, PVST, or MSTP. In this case, when Device C receives an STP BPDU transparently transmitted by Device B, the receiving port transits to the STP mode. If you configure Device B to run RSTP, PVST, or MSTP with Device C, you must perform mCheck operations on the ports interconnecting Device B and Device C.

Restrictions and guidelines

The mCheck operation takes effect on devices operating in MSTP, PVST, or RSTP mode.

Performing mCheck globally

1. Enter system view.

system-view

2. Perform mCheck.
   stp global mcheck

Performing mCheck in interface view

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Perform mCheck.
   stp mcheck

Disabling inconsistent PVID protection

About this task

In PVST, if two connected ports use different PVIDs, PVST calculation errors might occur. By default, inconsistent PVID protection is enabled to avoid PVST calculation errors. If PVID inconsistency is detected on a port, the system blocks the port.

Restrictions and guidelines

If different PVIDs are required on two connected ports, disable inconsistent PVID protection on the devices that host the ports. To avoid PVST calculation errors, make sure the following requirements are met:

- Make sure the VLANs on one device do not use the same ID as the PVID of its peer port (except the default VLAN) on another device.
- If the local port or its peer is a hybrid port, do not configure the local and peer ports as untagged members of the same VLAN.
- Disable inconsistent PVID protection on both the local device and the peer device.

This feature takes effect only when the device is operating in PVST mode.

Procedure

1. Enter system view.
   system-view
2. Disable the inconsistent PVID protection feature.
   stp ignore-pvid-inconsistency
   By default, the inconsistent PVID protection feature is enabled.

Ignoring the specified VLANs

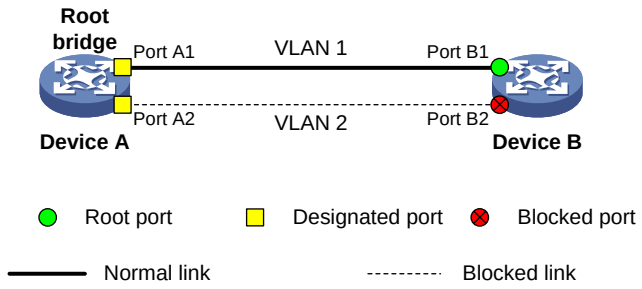
About this task

Traffic of a VLAN on a complex network might be blocked by the spanning tree, as shown in Figure 16.

- Port A1 on Device A allows the traffic of VLAN 1 to pass through.
- Port A2 on Device A allows the traffic of VLAN 2 to pass through.
- Port B1 on Device B allows the traffic of VLAN 1 to pass through.

- Port B2 on Device B allows the traffic of VLAN 2 to pass through.
- Device A and Device B run a spanning tree protocol. Device A is the root bridge, and Port A1 and Port A2 are designated ports. On Device B, Port B1 is the root port, and Port B2 is the blocked port. Traffic of VLAN 2 is blocked.

Figure 16 VLAN connectivity blocked by MSTP



Ignoring a VLAN can make ports of the VLAN forward frames correctly rather than comply with the spanning tree calculation result.

Procedure

1. Enter system view.

   system-view

2. Ignore the specified VLANs.

   stp ignored vlan *vlan-id-list*

   By default, no VLAN is ignored.

Configuring Digest Snooping

About this task

As defined in IEEE 802.1s, connected devices are in the same region only when they have the same MST region-related configurations, including:

- Region name.
- Revision level.
- VLAN-to-instance mappings.

A spanning tree device identifies devices in the same MST region by determining the configuration ID in BPDUs. The configuration ID includes the region name, revision level, and configuration digest. It is 16-byte long and is the result calculated through the HMAC-MD5 algorithm based on VLAN-to-instance mappings.

Because spanning tree implementations vary by vendor, the configuration digests calculated through private keys are different. The devices of different vendors in the same MST region cannot communicate with each other.

To enable communication between an H3C device and a third-party device in the same MST region, enable Digest Snooping on the H3C device port connecting them.

Restrictions and guidelines

CAUTION:

Use caution with global Digest Snooping in the following situations:

- When you modify the VLAN-to-instance mappings.
- When you restore the default MST region configuration.

If the local device has different VLAN-to-instance mappings than its neighboring devices, loops or traffic interruption will occur.

- Before you enable Digest Snooping, make sure associated devices of different vendors are connected and run spanning tree protocols.
- With Digest Snooping enabled, in-the-same-region verification does not require comparison of configuration digest. The VLAN-to-instance mappings must be the same on associated ports.
- To make Digest Snooping take effect, you must enable Digest Snooping both globally and on associated ports. As a best practice, enable Digest Snooping on all associated ports first and then enable it globally. This will make the configuration take effect on all configured ports and reduce impact on the network.
- To prevent loops, do not enable Digest Snooping on MST region edge ports.
- As a best practice, enable Digest Snooping first and then enable the spanning tree feature. To avoid traffic interruption, do not configure Digest Snooping when the network is already working well.

Prerequisites

Before configuring Digest Snooping, you need to make sure your H3C device and the third-party device both run spanning tree protocols properly.

Procedure

1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Enable Digest Snooping on the interface.
   `stp config-digest-snooping`
   By default, Digest Snooping is disabled on ports.
4. Return to system view.
   `quit`
5. Enable Digest Snooping globally.
   `stp global config-digest-snooping`
   By default, Digest Snooping is disabled globally.

Configuring No Agreement Check

About this task

In RSTP and MSTP, the following types of messages are used for rapid state transition on designated ports:

- **Proposal**—Sent by designated ports to request rapid transition
- **Agreement**—Used to acknowledge rapid transition requests

Both RSTP and MSTP devices can perform rapid transition on a designated port only when the port receives an agreement packet from the downstream device. RSTP and MSTP devices have the following differences:

- For MSTP, the root port of the downstream device sends an agreement packet only after it receives an agreement packet from the upstream device.
- For RSTP, the downstream device sends an agreement packet whether or not an agreement packet from the upstream device is received.

Figure 17 Rapid state transition of an MSTP designated port

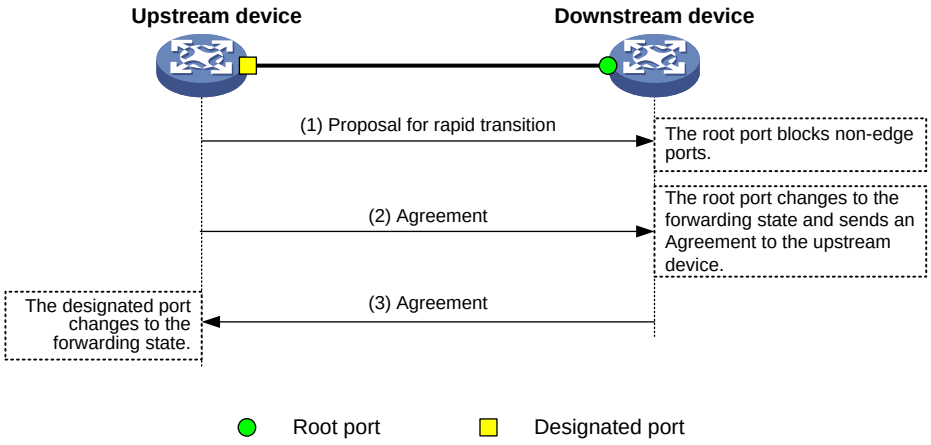
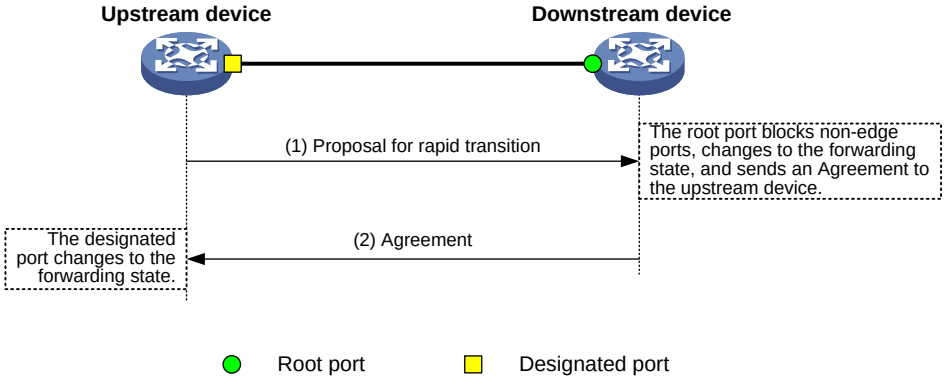


Figure 18 Rapid state transition of an RSTP designated port



If the upstream device is a third-party device, the rapid state transition implementation might be limited as follows:

- The upstream device uses a rapid transition mechanism similar to that of RSTP.
- The downstream device runs MSTP and does not operate in RSTP mode.

In this case, the following occurs:

1. The root port on the downstream device receives no agreement from the upstream device.
2. It sends no agreement to the upstream device.

As a result, the designated port of the upstream device can transit to the forwarding state only after a period twice the forward delay.

To enable the designated port of the upstream device to transit its state rapidly, enable No Agreement Check on the downstream device's port.

Restrictions and guidelines

Configure No Agreement Check on the root port of your device, because this feature takes effect only if it's configured on root ports.

Prerequisites

Before you configure the No Agreement Check feature, complete the following tasks:

- Connect a device to a third-party upstream device that supports spanning tree protocols through a point-to-point link.
- Configure the same region name, revision level, and VLAN-to-instance mappings on the two devices.

Procedure

Enable the No Agreement Check feature on the root port.

1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Enable No Agreement Check.
   `stp no-agreement-check`
   By default, No Agreement Check is disabled.

Configuring TC Snooping

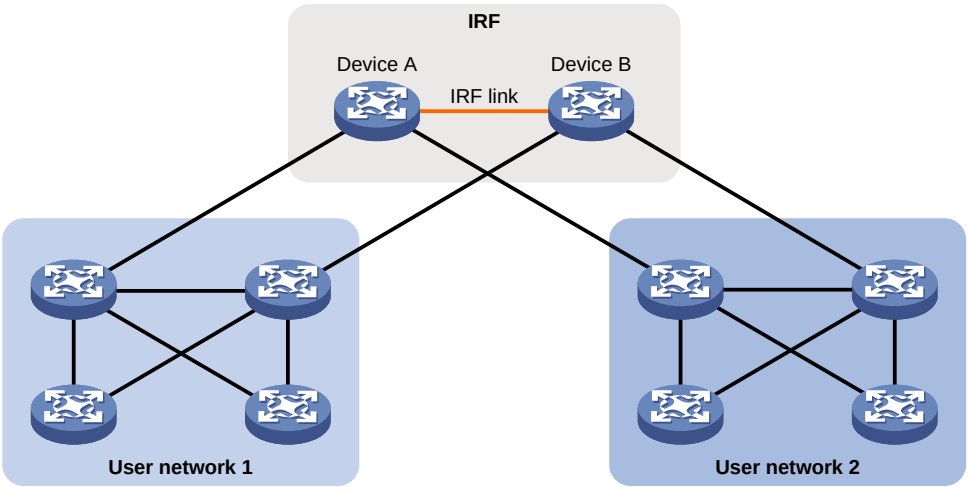
About this task

As shown in Figure 19, an IRF fabric connects to two user networks through double links.

- Device A and Device B form the IRF fabric.
- The spanning tree feature is disabled on Device A and Device B and enabled on all devices in user network 1 and user network 2.
- The IRF fabric transparently transmits BPDUs for both user networks and is not involved in the calculation of spanning trees.

When the network topology changes, it takes time for the IRF fabric to update its MAC address table and ARP table. During this period, traffic in the network might be interrupted.

Figure 19 TC Snooping application scenario



To avoid traffic interruption, you can enable TC Snooping on the IRF fabric. After receiving a TC-BPDU through a port, the IRF fabric updates MAC address table and ARP table entries associated with the port's VLAN. In this way, TC Snooping prevents topology change from interrupting traffic forwarding in the network. For more information about the MAC address table and the ARP table, see "Configuring the MAC address table" and *Layer 3—IP Services Configuration Guide*.

Restrictions and guidelines

- TC Snooping and the spanning tree feature are mutually exclusive. You must globally disable the spanning tree feature before enabling TC Snooping.
- TC Snooping does not support the PVST mode.

Procedure

1. Enter system view.

   `system-view`
2. Globally disable the spanning tree feature.

   `undo stp global enable`

   When the device starts up with initial settings, the spanning tree feature is globally disabled by default.

   When the device starts up with factory defaults, the spanning tree feature is globally enabled by default.

   For more information about the initial settings and factory defaults, see *Fundamentals Configuration Guide*.
3. Enable TC Snooping.

   `stp tc-snooping`

   By default, TC Snooping is disabled.

Configuring protection features

Spanning tree protection tasks at a glance

All spanning tree protection tasks are optional.

- Configuring BPDU guard
- Configuring BPDU filter
- Enabling root guard
- Enabling loop guard
- Configuring port role restriction
- Configuring TC-BPDU transmission restriction
- Enabling TC-BPDU guard
- Enabling BPDU drop
- Enabling PVST BPDU guard
- Disabling dispute guard
- Disabling loopback guard

Configuring BPDU guard

About this task

For access layer devices, the access ports can directly connect to the user terminals (such as PCs) or file servers. The access ports are configured as edge ports to allow rapid transition. When these ports receive configuration BPDUs, the system automatically sets the ports as non-edge ports and starts a new spanning tree calculation process. This causes a change of network topology. Under normal conditions, these ports should not receive configuration BPDUs. However, if someone uses configuration BPDUs maliciously to attack the devices, the network will become unstable.

The spanning tree protocol provides the BPDU guard feature to protect the system against such attacks. When ports with BPDU guard enabled receive configuration BPDUs on a device, the device performs the following operations:

- Shuts down these ports.
- Notifies the NMS that these ports have been shut down by the spanning tree protocol.

The device reactivates the ports that have been shut down when the port status detection timer expires. You can set this timer by using the **shutdown-interval** command. For more information about this command, see device management commands in *Fundamentals Command Reference*.

Restrictions and guidelines

You can configure the BPDU guard feature in system view or on a per-edge port basis. An edge port preferentially uses the port-specific BPDU guard setting. If the port-specific BPDU guard setting is not available, the edge port uses the global BPDU guard setting.

Global BPDU guard takes effect only on the edge ports configured by using the **stp edged-port** command. For the BPDU guard feature to take effect on a non-edge port, use the port-specific BPDU guard setting.

Configure BPDU guard on edge ports which directly connect to a user terminal rather than other device or shared LAN segment.

BPDU guard does not take effect on loopback-testing-enabled ports. For more information about loopback testing, see Ethernet interface configuration in *Layer 2—LAN Switching Configuration Guide*.

Enabling BPDU guard in system view

1. Enter system view.

   system-view
2. Enable BPDU guard globally.

   stp bpdu-protection

   By default, BPDU guard is globally disabled.

Configuring BPDU guard in interface view

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure BPDU guard.

   stp port bpdu-protection { **enable** | **disable** }

   By default, the enabling status of BPDU guard on an interface is the same as that of global BPDU guard. BPDU guard is not configured on non-edge ports.

Configuring BPDU filter

About this task

To prevent undesired spanning tree calculation and network flapping in the networks attached to an edge device, enable BPDU filter on the device to disable its ports from sending BPDUs.

Restrictions and guidelines

IMPORTANT:

With BPDU filter enabled, a port does not send any BPDUs and ignores all incoming BPDUs regardless of whether it is an edge port. Make sure you are fully aware of the impacts of this feature before you enable it on a port.

You can configure the BPDU filter feature globally or on a per-port basis. A port preferentially uses the port-specific BPDU filter setting. If the port-specific BPDU filter setting is not available, the port uses the global BPDU filter setting.

The global BPDU filter setting takes effect on all edge ports configured by using the **stp edged-port** command. With BPDU filter enabled globally, edge ports no longer send BPDUs, and they will become non-edge ports to participate in spanning tree calculation after receiving BPDUs.

Enabling BPDU filter globally

1. Enter system view.

   system-view
2. Enable BPDU filter globally.

   stp bpdu-filter

   By default, BPDU filter is disabled globally.

Configuring BPDU filter on a port

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure BPDU filter on the interface.

   stp port bpdu-filter { **disable** | **enable** }

   By default, the enabling status of BPDU filter on an edge port is the same as that of global BPDU filter. The BPDU filter feature is disabled on non-edge ports.

Enabling root guard

About this task

Configure root guard on a designated port.

The root bridge and secondary root bridge of a spanning tree should be located in the same MST region. Especially for the CIST, the root bridge and secondary root bridge are put in a high-bandwidth core region during network design. However, due to possible configuration errors or malicious attacks in the network, the legal root bridge might receive a configuration BPDU with a higher priority. Another device supersedes the current legal root bridge, causing an undesired change of the network topology. The traffic that should go over high-speed links is switched to low-speed links, resulting in network congestion.

To prevent this situation, MSTP provides the root guard feature. If root guard is enabled on a port of a root bridge, this port plays the role of designated port on all MSTIs. After this port receives a configuration BPDU with a higher priority from an MSTI, it performs the following operations:

- Immediately sets that port to the listening state in the MSTI.
- Does not forward the received configuration BPDU.

This is equivalent to disconnecting the link connected to this port in the MSTI. If the port receives no BPDUs with a higher priority within twice the forwarding delay, it reverts to its original state.

Restrictions and guidelines

On a port, the loop guard feature and the root guard feature are mutually exclusive.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enter interface view.
   ```
   interface interface-type interface-number
   ```
3. Enable the root guard feature.
   ```
   stp root-protection
   ```
   By default, root guard is disabled.

Enabling loop guard

About this task

Configure loop guard on the root port and alternate ports of a device.

By continuing to receive BPDUs from the upstream device, a device can maintain the state of the root port and blocked ports. However, link congestion or unidirectional link failures might cause these ports to fail to receive BPDUs from the upstream devices. In this situation, the device reselects the following port roles:

- Those ports in forwarding state that failed to receive upstream BPDUs become designated ports.
- The blocked ports transit to the forwarding state.

As a result, loops occur in the switched network. The loop guard feature can suppress the occurrence of such loops.

The initial state of a loop guard-enabled port is **discarding** in every MSTI. When the port receives BPDUs, it transits its state. Otherwise, it stays in the discarding state to prevent temporary loops.

Restrictions and guidelines

Do not enable loop guard on a port that connects user terminals. Otherwise, the port stays in the discarding state in all MSTIs because it cannot receive BPDUs.

On a port, the loop guard feature is mutually exclusive with the root guard feature or the edge port setting.

A loop guard-enabled interface can receive BPDUs and transit from the discarding state to the forwarding state after two forward delays if one of the following events occurs:

- The state of the interface changes from down to up.
- The spanning tree feature is enabled on the up interface.

Procedure

1. Enter system view.
   ```
   system-view
   ```

2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the loop guard feature.
   stp loop-protection
   By default, loop guard is disabled.

Configuring port role restriction

About this task

Make this configuration on the port that connects to the user access network.

The bridge ID change of a device in the user access network might cause a change to the spanning tree topology in the core network. To avoid this problem, you can enable port role restriction on a port. With this feature enabled, when the port receives a superior BPDU, it becomes an alternate port rather than a root port.

Restrictions and guidelines

Use this feature with caution, because enabling port role restriction on a port might affect the connectivity of the spanning tree topology.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable port role restriction.
   stp role-restriction
   By default, port role restriction is disabled.

Configuring TC-BPDU transmission restriction

About this task

Make this configuration on the port that connects to the user access network.

The topology change to the user access network might cause the forwarding address changes to the core network. When the user access network topology is unstable, the user access network might affect the core network. To avoid this problem, you can enable TC-BPDU transmission restriction on a port. With this feature enabled, when the port receives a TC-BPDU, it does not forward the TC-BPDU to other ports.

Restrictions and guidelines

Enabling TC-BPDU transmission restriction on a port might cause the previous forwarding address table to fail to be updated when the topology changes.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable TC-BPDU transmission restriction.
   stp tc-restriction

By default, TC-BPDU transmission restriction is disabled.

Enabling TC-BPDU guard

About this task

When a device receives topology change (TC) BPDUs (the BPDUs that notify devices of topology changes), it flushes its forwarding address entries. If someone uses TC-BPDUs to attack the device, the device will receive a large number of TC-BPDUs within a short time. Then, the device is busy with forwarding address entry flushing. This affects network stability.

TC-BPDU guard allows you to set the maximum number of immediate forwarding address entry flushes performed within 10 seconds after the device receives the first TC-BPDU. For TC-BPDUs received in excess of the limit, the device performs a forwarding address entry flush when the time period expires. This prevents frequent flushing of forwarding address entries.

Restrictions and guidelines

As a best practice, enable TC-BPDU guard.

Procedure

1. Enter system view.

   system-view

2. Enable the TC-BPDU guard feature.

   stp tc-protection

   By default, TC-BPDU guard is enabled.

3. (Optional.) Configure the maximum number of forwarding address entry flushes that the device can perform every 10 seconds.

   stp tc-protection threshold *number*

   The default setting is 6.

Enabling BPDU drop

About this task

In a spanning tree network, every BPDU arriving at the device triggers an STP calculation process and is then forwarded to other devices in the network. Malicious attackers might use the vulnerability to attack the network by forging BPDUs. By continuously sending forged BPDUs, they can make all devices in the network continue performing STP calculations. As a result, problems such as CPU overload and BPDU protocol status errors occur.

To avoid this problem, you can enable BPDU drop on ports. A BPDU drop-enabled port does not receive any BPDUs and is invulnerable to forged BPDU attacks.

Restrictions and guidelines

With this feature enabled, the packets with destination MAC addresses in the range of 0x01-80-c2-00-00-00 to 0x01-80-c2-00-00-2f will be dropped, including but not limited to packets from the following protocols: STP, RSTP, MSTP, LACP, LLDP, 802.1X, and GARP. Make sure you are fully aware of the impact of this feature when you use it on a live network.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet interface view.

   interface *interface-type interface-number*

3. Enable BPDU drop on the interface.

   bpdu-drop any

   By default, BPDU drop is disabled.

Enabling PVST BPDU guard

About this task

This feature takes effect only when the device is operating in MSTP mode.

An MSTP-enabled device forwards PVST BPDUs as data traffic because it cannot recognize PVST BPDUs. If a PVST-enabled device in another independent network receives the PVST BPDUs, a PVST calculation error might occur. To avoid PVST calculation errors, enable PVST BPDU guard on the MSTP-enabled device. The device shuts down a port if the port receives PVST BPDUs.

Procedure

1. Enter system view.

   system-view

2. Enable PVST BPDU guard.

   stp pvst-bpdu-protection

   By default, PVST BPDU guard is disabled.

Disabling dispute guard

About this task

Dispute guard can be triggered by unidirectional link failures. If an upstream port receives inferior BPDUs from a downstream designated port in forwarding or learning state because of a unidirectional link failure, a loop appears. Dispute guard blocks the upstream designated port to prevent the loop.

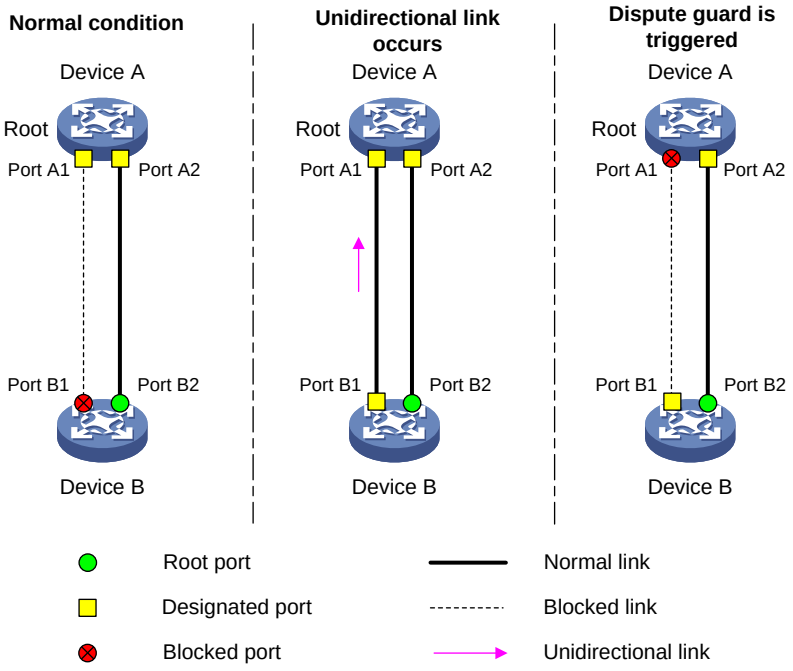
As shown in Figure 20, in normal conditions, the spanning tree calculation result is as follows:

- Device A is the root bridge, and Port A1 is a designated port.
- Port B1 is blocked.

When the link between Port A1 and Port B1 fails in the direction of Port A1 to Port B1 and becomes unidirectional, the following events occur:

1. Port A1 can only receive BPDUs and cannot send BPDUs to Port B1.
2. Port B1 does not receive BPDUs from Port A1 for a certain period of time.
3. Device B determines itself as the root bridge.
4. Port B1 sends its BPDUs to Port A1.
5. Port A1 determines the received BPDUs are inferior to its own BPDUs. A dispute is detected.
6. Dispute guard is triggered and blocks Port A1 to prevent a loop.

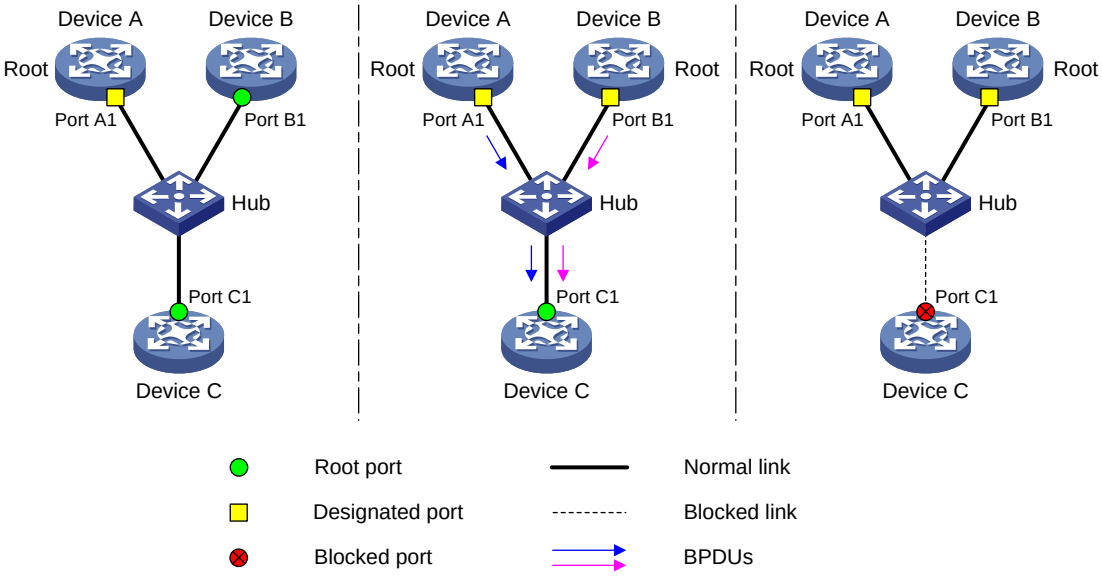
Figure 20 Dispute guard triggering scenario (on a designated port)



As shown in Figure 21, in normal conditions, Device A is the root bridge, and Port B1 and Port C1 are root ports. When the links between Device A and Device B become unidirectional (the links fail in the direction of Port A1 to Port B1), the following events occur:

7. Device B cannot receive BPDUs from Device A.
8. Device B determines itself as the root bridge.
9. Port B1 sends BPDUs in which the root bridge is Device B to Port C1.
10. Port C1 receives BPDUs from two root bridges, Device A and Device B. A dispute is detected.
11. Dispute guard is triggered and blocks Port C1 to avoid a loop.

Figure 21 Dispute guard triggering scenario (on a root port)



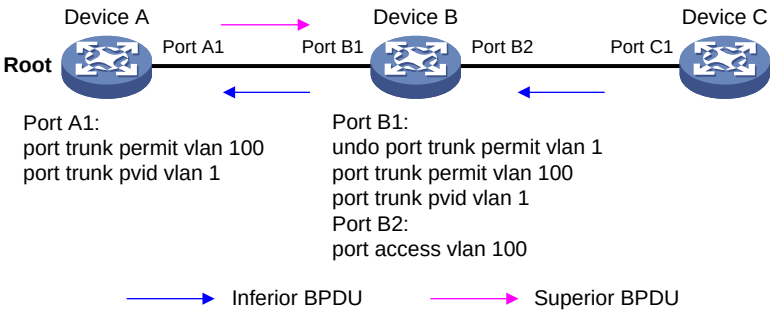
However, dispute guard might disrupt the network connectivity. You can disable dispute guard to avoid connectivity loss in VLAN networks. As shown in Figure 22, the spanning tree feature is

disabled on Device B and enabled on Device A and device C. Device B transparently transmits BPDUs.

Device C cannot receive superior BPDUs of VLAN 1 from Device A because Port B1 of Device B is configured to deny packets of VLAN 1. Device C determines itself as the root bridge after a certain period of time. Then, Port C1 sends an inferior BPDU of VLAN 100 to Device A.

When Device A receives the inferior BPDU, dispute guard blocks Port A1, which causes traffic interruption. To ensure service continuity, you can disable dispute guard on Device A to prevent the link from being blocked.

Figure 22 Disabling dispute guard application scenario



Restrictions and guidelines

You can disable dispute guard if the network does not have unidirectional link failures.

Procedure

1. Enter system view.
   - **`system-view`**
2. Disable dispute guard.
   - **`undo stp dispute-protection`**
   - By default, dispute guard is enabled.

Disabling loopback guard

About this task

By default, loopback guard is enabled on interfaces. The device blocks a loopback guard-enabled interface to remove loops when the interface receives a BPDU sent by itself. In some scenarios, an interface can receive the BPDUs sent by itself when no loop exists. For example, a fiber-to-copper converter might reflect back BPDUs. To prevent the device from blocking loopless interfaces in those scenarios and ensure a stable network topology, you can disable loopback guard.

Restrictions and guidelines

Disabling loopback guard might cause loops. Disable this feature unless necessary. Make sure you are fully aware of the impact when you disable loopback guard on a live network.

In Layer 2 Ethernet interface view, loopback guard configuration takes effect only on that interface.

In Layer 2 aggregate interface view, loopback guard configuration takes effect only on the aggregate interface.

On a member port in an aggregation group, loopback guard configuration takes effect only after the port leaves the aggregation group.

Procedure

1. Enter system view.

system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Enable loopback guard.
   undo stp loopback-protection
   - By default, loopback guard is enabled on interfaces.

Enabling the device to log events of detecting or receiving TC BPDUs

About this task

This feature allows the device to generate logs when it detects or receives TC BPDUs. This feature applies only to PVST mode.

Procedure

1. Enter system view.
   system-view
2. Enable the device to log events of receiving or detecting TC BPDUs.
   stp log enable tc
   By default, the device does not generate logs when it detects or receives TC BPDUs.

Disabling the device from reactivating edge ports shut down by BPDU guard

About this task

BPDU guard shuts down edge ports that have received configuration BPDUs and notifies the NMS of the shutdown event.

The device reactivates the ports that have been shut down when the port status detection timer expires. You can set this timer by using the **shutdown-interval** command. For more information about this command, see device management commands in *Fundamentals Command Reference*.

Restrictions and guidelines

This feature prevents the device from reactivating edge ports shut down by BPDU guard after this feature is configured. The device does not bring up the shutdown ports if you execute the **undo stp port shutdown permanent** command. To bring up these ports, use the **undo shutdown** command.

Procedure

1. Enter system view.
   system-view
2. Disable the device from reactivating edge ports shut down by BPDU guard.
   stp port shutdown permanent
   By default, the device reactivates an edge port shut down by BPDU guard after the port status detection timer expires.

Enabling BPDU transparent transmission on a port

Restrictions and guidelines

Perform this task to enable a port to transmit BPDUs transparently. Whether the spanning tree protocols are enabled on a port does not affect the BPDU transparent transmission feature.

If this feature and the spanning tree protocol are enabled on a port which is inferior to its downstream port, the downstream port can receive BPDUs from that port. To prevent network flapping caused by this problem, disable the spanning tree protocol before you enable BPDU transparent transmission on the port.

Procedure

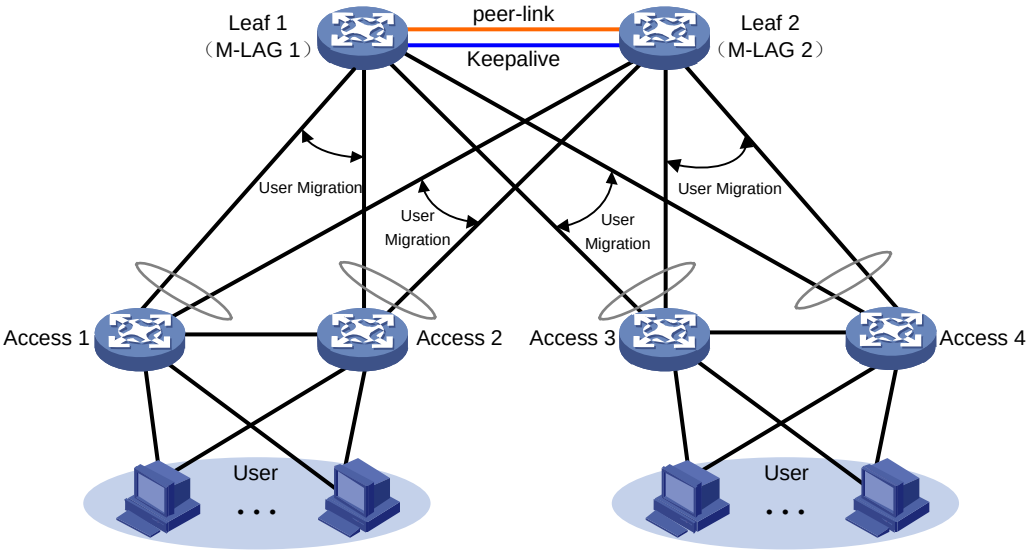
1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Enable BPDU transparent transmission.
   `stp transparent enable`
   By default, the BPDU transparent transmission feature is disabled on a port.

Configuring TC groups

About this task

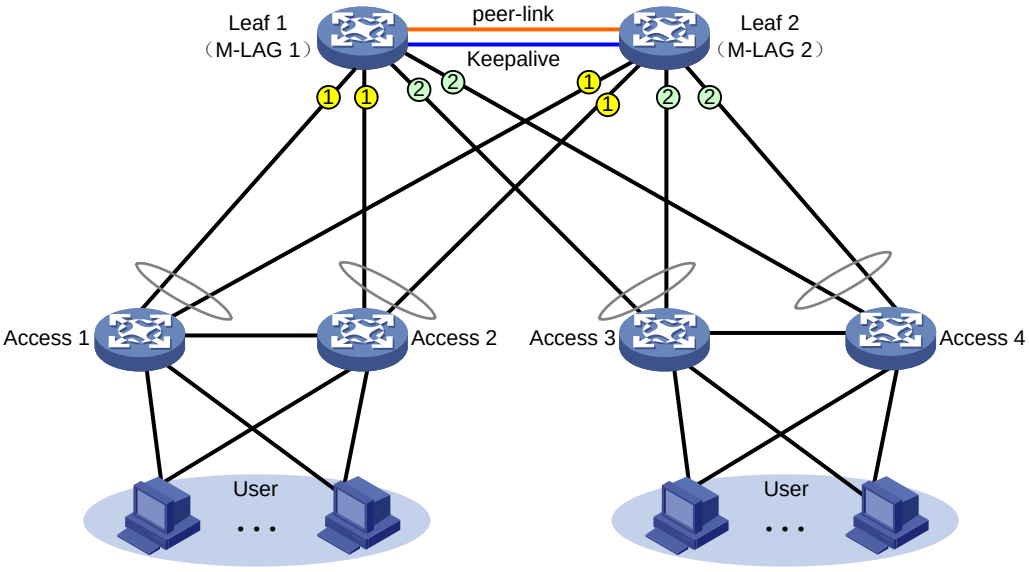
As shown in Figure 23, two devices at the leaf tier form an M-LAG system. The M-LAG system is connected to two loops (Access 1 — Access 2 — M-LAG and Access 3 — Access 4 — M-LAG), and STP runs on all devices. By default, any topology change by a device causes TC-BPDUs to be sent to both loops and create excessive TC-BPDUs in the network, which impacts system stability. If the topology changes, the interfaces for the user onboarding service on the leaf devices might change. In this case, users must be re-authenticated on other interfaces of the leaf devices to come online. During the re-authentication process, service traffic might be lost.

Figure 23 Network diagram



To address this issue, you can configure topology change (TC) groups. A TC group is a logical combination of multiple interfaces on a device, and supports a maximum of two interfaces. As shown in Figure 24, the two M-LAG interfaces connecting to Access 1 and Access 2 are added to TC group 1, and the two M-LAG interfaces connecting to Access 3 and Access 4 are added to TC group 2 on the leaf devices.

Figure 24 Network diagram



After TC groups are configured, the network can achieve the following functions:

- When an M-LAG interface in a TC group receives a TC-BPDU, it forwards it only to the other M-LAG interface in the same TC group. This mechanism reduces the number of TC-BPDUs in the entire network, isolating the TC-BPDUs of the Access 1—Access 2—M-LAG loop and those of the Access 3—Access 4—M-LAG loop. When leaf devices are connected to many loops, this mechanism reduces the impact of exchanging TC-BPDU packets among loops.
- When the network topology changes, static users, MAC authentication users, and 802.1X users in the same TC group can migrate between the two M-LAG interfaces in the TC group without authentication, enabling users to quickly come online again. For more information about the configuration procedures, migration conditions, and migration rules for authentication-free migration, see port security configuration in *Security Configuration Guide*.

Restrictions and guidelines

You can add interfaces to a TC group in either of the following methods:

- Execute the **port stp tc-group** command in interface view.
- Execute the **port** command in TC group view.

Both methods have the same effect, but their configurations cannot conflict with each other. Once an interface is added to a TC group in one method, it can be removed from the TC group in the other method. However, you cannot directly modify the TC group to which the interface belongs in the other method.

Procedure

1. Enter system view.

   system-view

2. Create a TC group and enter its view.

   stp tc-group *group-id*

3. Add interfaces to the TC group in one of the following methods:

- Add interfaces to the TC group in TC group view.

  port *interface-type interface-number*

- Execute the following commands in sequence to add interfaces to the TC group in Layer 2 aggregate interface view.

  quit

  interface *interface-type interface-number*

  port stp tc-group *group-id*

  By default, an interface is not added to any TC group.

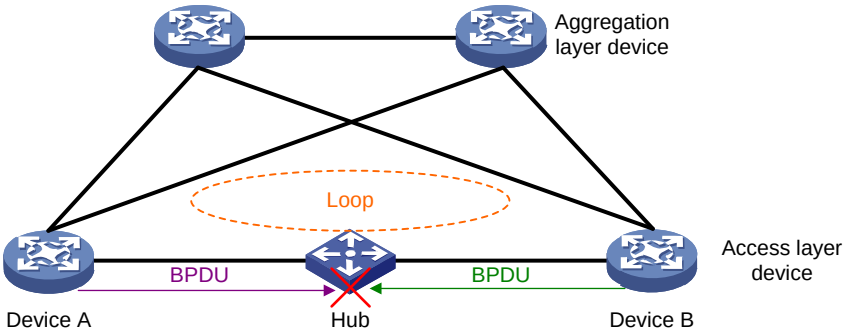
  A maximum of two interfaces can be added to a TC group.

Configuring spanning tree blackhole detection

About this task

As shown in Figure 25, Device A and Device B are connected via a HUB, which creates a loop in the network. Since BPDUs are point-to-point frames, once they are transmitted from a device and received by the next node, their transmission is terminated. The HUB acts as a blackhole for BPDUs. BPDUs cannot be transmitted between Device A and Device B through the HUB, and Device A and Device B cannot eliminate the loop through correct spanning tree topology calculation. Therefore, a method to detect BPDU blackholes is crucial for devices to block links to blackholes when detecting them, effectively eliminating potential loop risks.

Figure 25 Network diagram



Spanning tree blackhole detection feature can detect the presence of BPDU blackholes in the port links. Its working mechanism is as follows.

1. Blackhole verification phase.

   A port enabled with spanning tree blackhole detection will start the BPDU reception timer of the blackhole detection feature. If the port receives a BPDU before the BPDU reception timer expires, the port resets the BPDU reception timer. If the BPDU reception timer expires, BPDU blackholes might exist on the network. As a result, the device fails to receive BPDUs for a long time, and the port starts to send spanning tree blackhole detection packets continuously.

   Spanning tree blackhole detection packet is a type of Layer 2 packets defined by H3C. Different from the BPDU, the spanning tree blackhole detection packet can be transparently transmitted by the HUB, and will be sent continuously at a fixed time interval after being triggered to send.

   Use the **stp global blackhole-detection rx-bpdu timeout** command to set the BPDU reception timer. Use the **stp timer blackhole-detection-interval** or **stp global timer blackhole-detection-interval** command to set the interval for sending spanning tree blackhole detection packets.

2. Blackhole confirmation phase.

If the port is triggered to send spanning tree blackhole detection packets and receives a BPDU, it proves that no BPDU blackhole exist between the devices. The port stops transmitting the spanning tree blackhole detection packets and resets the BPDU reception timer.

If a port with the spanning tree blackhole detection feature enabled receives a blackhole detection frame while its BPDU reception timer is in a timeout state, it indicates that the link between the local and remote devices is connected. Both devices are capable of sending BPDUs, but neither device receives BPDUs. The local device determines that a BPDU blackhole exists between the two ends of the link.

3. Blackhole handling phase.

   After the device discovers a BPDU blackhole, it blocks the port that receives the spanning tree blackhole detection packet. Loops in the network are eliminated by preventing ports from sending and receiving any user data packets. You can use the **stp timer rx-blackhole-timeout** or **stp global timer rx-blackhole-timeout** command to specify the timeout period of the blocked port for spanning tree blackhole detection. If the port does not receive the spanning tree blackhole detection packet again within the timeout period, it means that the BPDU blackhole is eliminated, and the port returns to the normal forwarding state. Otherwise, the port continues to retain the blocking state.

The **stp timer blackhole-detection-interval** and **stp global timer blackhole-detection-interval** commands affect the frequency at which the specified port sends spanning tree blackhole detection packets. Set an appropriate packet sending interval as needed.

- When the sending interval is short, the port sends spanning tree blackhole packets more frequently. The advantage is that the network is more sensitive in detection and response to BPDU blackholes. The disadvantage is that it takes up more device CPU resources. Set a short sending interval in scenarios with strong device performance or a strong need to prevent loops.
- When the sending interval is long, the port sends spanning tree blackhole packets slowly. The advantage is that it occupies less CPU resources of the device. The disadvantage is that the network detects and responds more slowly to BPDU blackholes. Set a long sending interval in the scenario where the device performance is weak or tolerant of loops.

Restrictions and guidelines

The spanning tree blackhole detection feature takes effect when the following conditions are met:

- On the devices at both ends of a link, the spanning tree blackhole detection is enabled both globally and on ports.
- The link status of the port on which the spanning tree blackhole detection is enabled is up, and the spanning tree feature is enabled.

If the spanning tree blackhole detection is disabled on the port, or the BPDU reception timer on the port has not expired, the port does not perform any processing after receiving the spanning tree blackhole detection packet.

On a device with a long interval for sending BPDUs, set a longer value for the BPDU reception timer to prevent the device port from being blocked when no BPDU blackhole exists on the network. On the device with a short interval for sending BPDUs, set a shorter value for the BPDU reception timer to improve the network's detection speed of BPDU blackholes.

You can use the **stp timer blackhole-detection-interval** and **stp global timer blackhole-detection-interval** commands to modify the interval for sending spanning tree blackhole detection packets on a port. When both commands are configured at the same time, the configuration of the **stp timer blackhole-detection-interval** command takes effect. If no command is configured on the port, the port inherits the configuration of the **stp global timer blackhole-detection-interval** command.

You can use the **stp timer rx-blackhole-timeout** and **stp global timer rx-blackhole-timeout** commands to modify the timeout timer for receiving spanning tree blackhole detection packets on a port. When both commands are configured at the same time, the

configuration of the **stp timer rx-blackhole-timeout** command takes effect. If no command is configured on the port, the port inherits the configuration of the **stp global timer rx-blackhole-timeout** command.

Procedure

1. Enter system view.
   system-view
2. Enable spanning tree blackhole detection globally.
   stp global blackhole-detection enable
   By default, spanning tree blackhole detection is disabled globally.
3. (Optional.) Set the BPDU reception timer for the blackhole detection feature.
   stp global blackhole-detection rx-bpdu timeout *timeout*
   By default, the BPDU reception timer of the blackhole detection feature is 18 seconds.
4. (Optional.) Set the interval for sending spanning tree blackhole detection packets globally.
   stp global timer blackhole-detection-interval *interval*
   By default, the interval for sending spanning tree blackhole detection packets is 2 seconds.
5. (Optional.) Set the detection packet reception timer for spanning tree blackhole detection globally.
   stp global timer rx-blackhole-timeout *timeout*
   By default, the formula for calculating the timeout period (seconds) for receiving spanning tree blackhole packets is: Multiply the sending interval of spanning tree blackhole detection packets on a port by 3 and add 10.
6. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
7. (Optional.) Enable spanning tree blackhole detection on the port.
   stp blackhole-detection enable
   By default, spanning tree blackhole detection on the port is enabled.
8. (Optional.) Set the interval for sending spanning tree blackhole detection packets on the port.
   stp timer blackhole-detection-interval *interval*
   By default, the interval for sending spanning tree blackhole detection packets is 2 seconds.
9. (Optional.) Set the detection packet reception timer for spanning tree blackhole detection on the port.
   stp timer rx-blackhole-timeout *timeout*
   By default, the formula for calculating the timeout period (seconds) for receiving spanning tree blackhole packets on the port is: Multiply the sending interval of spanning tree blackhole detection packets on a port by 3 and add 10.

Enabling SNMP notifications for new-root election and topology change events

About this task

This task enables the device to generate logs and report new-root election events or spanning tree topology changes to SNMP. For the event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

When you use the **snmp-agent trap enable stp** [**new-root** | **tc**] command, follow these guidelines:

- The **new-root** keyword applies only to STP, MSTP, and RSTP modes.
- The **tc** keyword applies only to PVST mode.
- In STP, MSTP, or RSTP mode, the **snmp-agent trap enable stp** command enables SNMP notifications for new-root election events.
- In PVST mode, the **snmp-agent trap enable stp** command enables SNMP notifications for spanning tree topology changes.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for new-root election and topology change events.
   snmp-agent trap enable stp [**new-root** | **tc**]
   The default settings are as follows:
   - SNMP notifications are disabled for new-root election events.
   - In MSTP mode, SNMP notifications are enabled in MSTI 0 and disabled in other MSTIs for spanning tree topology changes.
   - In PVST mode, SNMP notifications are disabled for spanning tree topology changes in all VLANs.

Display and maintenance commands for the spanning tree protocols

Execute **display** commands in any view and **reset** command in user view.

Task	Command
Display the spanning tree status and statistics.	**display stp** [**instance** *instance-list* | **vlan** *vlan-id-list*] [**interface** *interface-list* | **slot** *slot-number*] [**brief**]
Display the port role calculation history for the specified MSTI or all MSTIs.	**display stp** [**instance** *instance-list* | **vlan** *vlan-id-list*] **history** [**slot** *slot-number*]
Display the incoming and outgoing TC/TCN BPDU statistics by all ports in the specified MSTI or all MSTIs.	**display stp** [**instance** *instance-list* | **vlan** *vlan-id-list*] **tc** [**slot** *slot-number*]
Display history about ports blocked by spanning tree protection features.	**display stp abnormal-port**
Display BPDU statistics on ports.	**display stp bpdu-statistics** [**interface** *interface-type interface-number* [**instance** *instance-list*]]
Display information about ports shut down by spanning tree protection features.	**display stp down-port**
Display the ignored VLANs.	**display stp ignored-vlan**

Task	Command
Display the MST region configuration information that has taken effect.	**display stp region-configuration**
Display the root bridge information of all MSTIs.	**display stp root**
Display information about TC groups.	**display stp tc-group** [*group-id*]
Display information about ports blocked by the spanning tree blackhole detection feature.	**display stp blackhole-detection blocked-port**
Clear the spanning tree statistics.	**reset stp** [**interface** *interface-list*]

Spanning tree configuration examples

Example: Configuring MSTP

Network configuration

As shown in Figure 26, all devices on the network are in the same MST region. Device A and Device B work at the distribution layer. Device C and Device D work at the access layer.

Configure MSTP so that frames of different VLANs are forwarded along different spanning trees.

- VLAN 10 frames are forwarded along MSTI 1.
- VLAN 30 frames are forwarded along MSTI 3.
- VLAN 40 frames are forwarded along MSTI 4.
- VLAN 20 frames are forwarded along MSTI 0.

VLAN 10 and VLAN 30 are terminated on the distribution layer devices, and VLAN 40 is terminated on the access layer devices. The root bridges of MSTI 1 and MSTI 3 are Device A and Device B, respectively, and the root bridge of MSTI 4 is Device C.

Figure 26 Network diagram

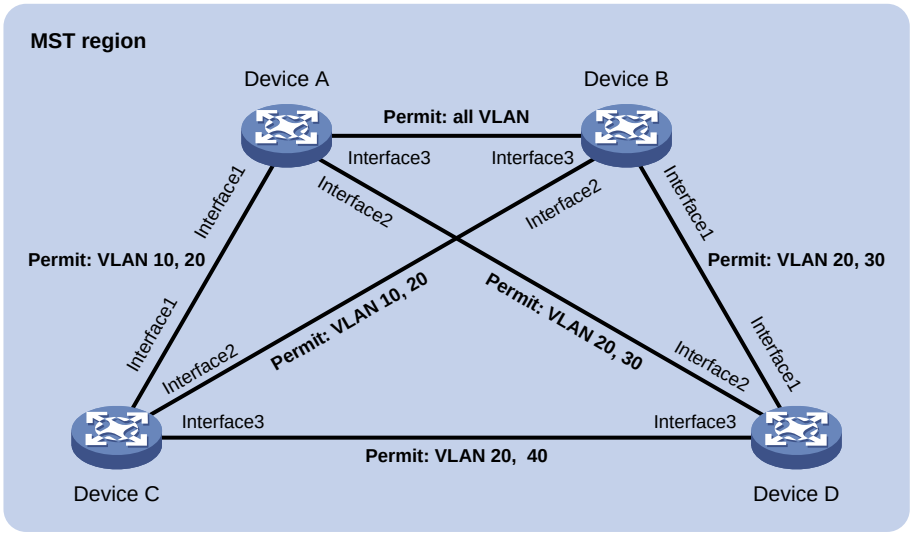


Table 14 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure VLANs and VLAN member ports. (Details not shown.)
   - Create VLAN 10, VLAN 20, and VLAN 30 on both Device A and Device B.
   - Create VLAN 10, VLAN 20, and VLAN 40 on Device C.
   - Create VLAN 20, VLAN 30, and VLAN 40 on Device D.
   - Configure the ports on these devices as trunk ports and assign them to related VLANs.
2. Configure Device A:

   # Enter MST region view, and configure the MST region name as **example**.

   ```
   <DeviceA> system-view
   [DeviceA] stp region-configuration
   [DeviceA-mst-region] region-name example
   ```

   # Map VLAN 10, VLAN 30, and VLAN 40 to MSTI 1, MSTI 3, and MSTI 4, respectively.

   ```
   [DeviceA-mst-region] instance 1 vlan 10
   [DeviceA-mst-region] instance 3 vlan 30
   [DeviceA-mst-region] instance 4 vlan 40
   ```

   # Configure the revision level of the MST region as 0.

   ```
   [DeviceA-mst-region] revision-level 0
   ```

   # Activate MST region configuration.

   ```
   [DeviceA-mst-region] active region-configuration
   [DeviceA-mst-region] quit
   ```

   # Configure the Device A as the root bridge of MSTI 1.

   ```
   [DeviceA] stp instance 1 root primary
   ```

   # Enable the spanning tree feature globally.

   ```
   [DeviceA] stp global enable
   ```

3. Configure Device B:

   # Enter MST region view, and configure the MST region name as **example**.

   ```
   <DeviceB> system-view
   [DeviceB] stp region-configuration
   [DeviceB-mst-region] region-name example
   ```

   # Map VLAN 10, VLAN 30, and VLAN 40 to MSTI 1, MSTI 3, and MSTI 4, respectively.

   ```
   [DeviceB-mst-region] instance 1 vlan 10
   [DeviceB-mst-region] instance 3 vlan 30
   [DeviceB-mst-region] instance 4 vlan 40
   ```

   # Configure the revision level of the MST region as 0.

   ```
   [DeviceB-mst-region] revision-level 0
   ```

   # Activate MST region configuration.

   ```
   [DeviceB-mst-region] active region-configuration
   [DeviceB-mst-region] quit
   ```

   # Configure Device B as the root bridge of MSTI 3.

```
[DeviceB] stp instance 3 root primary
```

Enable the spanning tree feature globally.

```
[DeviceB] stp global enable
```

4. Configure Device C:

Enter MST region view, and configure the MST region name as **example**.

```
<DeviceC> system-view
[DeviceC] stp region-configuration
[DeviceC-mst-region] region-name example
```

Map VLAN 10, VLAN 30, and VLAN 40 to MSTI 1, MSTI 3, and MSTI 4, respectively.

```
[DeviceC-mst-region] instance 1 vlan 10
[DeviceC-mst-region] instance 3 vlan 30
[DeviceC-mst-region] instance 4 vlan 40
```

Configure the revision level of the MST region as 0.

```
[DeviceC-mst-region] revision-level 0
```

Activate MST region configuration.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Configure the Device C as the root bridge of MSTI 4.

```
[DeviceC] stp instance 4 root primary
```

Enable the spanning tree feature globally.

```
[DeviceC] stp global enable
```

5. Configure Device D:

Enter MST region view, and configure the MST region name as **example**.

```
<DeviceD> system-view
[DeviceD] stp region-configuration
[DeviceD-mst-region] region-name example
```

Map VLAN 10, VLAN 30, and VLAN 40 to MSTI 1, MSTI 3, and MSTI 4, respectively.

```
[DeviceD-mst-region] instance 1 vlan 10
[DeviceD-mst-region] instance 3 vlan 30
[DeviceD-mst-region] instance 4 vlan 40
```

Configure the revision level of the MST region as 0.

```
[DeviceD-mst-region] revision-level 0
```

Activate MST region configuration.

```
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

Enable the spanning tree feature globally.

```
[DeviceD] stp global enable
```

Verifying the configuration

In this example, Device B has the lowest root bridge ID. As a result, Device B is elected as the root bridge in MSTI 0.

When the network is stable, you can use the **`display stp brief`** command to display brief spanning tree information on each device.

Display brief spanning tree information on Device A.

```
[DeviceA] display stp brief
 MST ID      Port                        Role  STP State     Protection
 0           GigabitEthernet1/0/1        ALTE  DISCARDING    NONE
```

```
 0           GigabitEthernet1/0/2         DESI  FORWARDING    NONE
 0           GigabitEthernet1/0/3         ROOT  FORWARDING    NONE
 1           GigabitEthernet1/0/1         DESI  FORWARDING    NONE
 1           GigabitEthernet1/0/3         DESI  FORWARDING    NONE
 3           GigabitEthernet1/0/2         DESI  FORWARDING    NONE
 3           GigabitEthernet1/0/3         ROOT  FORWARDING    NONE
```

Display brief spanning tree information on Device B.

```
[DeviceB] display stp brief
 MST ID      Port                         Role  STP State     Protection
 0           GigabitEthernet1/0/1         DESI  FORWARDING    NONE
 0           GigabitEthernet1/0/2         DESI  FORWARDING    NONE
 0           GigabitEthernet1/0/3         DESI  FORWARDING    NONE
 1           GigabitEthernet1/0/2         DESI  FORWARDING    NONE
 1           GigabitEthernet1/0/3         ROOT  FORWARDING    NONE
 3           GigabitEthernet1/0/1         DESI  FORWARDING    NONE
 3           GigabitEthernet1/0/3         DESI  FORWARDING    NONE
```

Display brief spanning tree information on Device C.

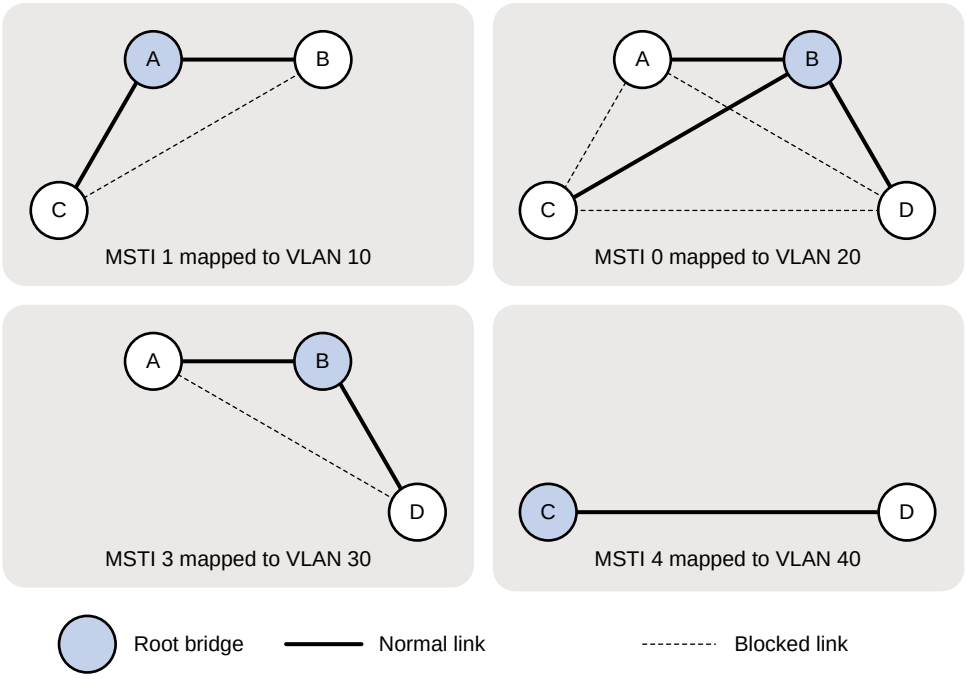
```
[DeviceC] display stp brief
 MST ID      Port                         Role  STP State     Protection
 0           GigabitEthernet1/0/1         DESI  FORWARDING    NONE
 0           GigabitEthernet1/0/2         ROOT  FORWARDING    NONE
 0           GigabitEthernet1/0/3         DESI  FORWARDING    NONE
 1           GigabitEthernet1/0/1         ROOT  FORWARDING    NONE
 1           GigabitEthernet1/0/2         ALTE  DISCARDING    NONE
 4           GigabitEthernet1/0/3         DESI  FORWARDING    NONE
```

Display brief spanning tree information on Device D.

```
[DeviceD] display stp brief
 MST ID      Port                         Role  STP State     Protection
 0           GigabitEthernet1/0/1         ROOT  FORWARDING    NONE
 0           GigabitEthernet1/0/2         ALTE  DISCARDING    NONE
 0           GigabitEthernet1/0/3         ALTE  DISCARDING    NONE
 3           GigabitEthernet1/0/1         ROOT  FORWARDING    NONE
 3           GigabitEthernet1/0/2         ALTE  DISCARDING    NONE
 4           GigabitEthernet1/0/3         ROOT  FORWARDING    NONE
```

Based on the output, you can draw each MSTI mapped to each VLAN, as shown in Figure 27.

Figure 27 MSTIs mapped to different VLANs



Example: Configuring PVST

Network configuration

As shown in Figure 28, Device A and Device B work at the distribution layer, and Device C and Device D work at the access layer.

Configure PVST to meet the following requirements:

- Frames of a VLAN are forwarded along the spanning trees of the VLAN.
- VLAN 10, VLAN 20, and VLAN 30 are terminated on the distribution layer devices, and VLAN 40 is terminated on the access layer devices.
- The root bridge of VLAN 10 and VLAN 20 is Device A.
- The root bridge of VLAN 30 is Device B.
- The root bridge of VLAN 40 is Device C.

Figure 28 Network diagram

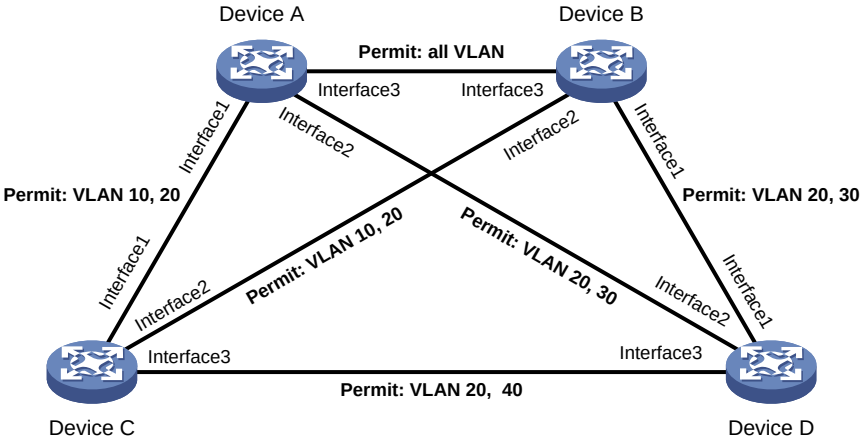


Table 15 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure VLANs and VLAN member ports. (Details not shown.)
   - Create VLAN 10, VLAN 20, and VLAN 30 on both Device A and Device B.
   - Create VLAN 10, VLAN 20, and VLAN 40 on Device C.
   - Create VLAN 20, VLAN 30, and VLAN 40 on Device D.
   - Configure the ports on these devices as trunk ports and assign them to related VLANs.
2. Configure Device A:

   # Set the spanning tree mode to PVST.

   ```
   <DeviceA> system-view
   [DeviceA] stp mode pvst
   ```

   # Configure the device as the root bridge of VLAN 10 and VLAN 20.

   ```
   [DeviceA] stp vlan 10 20 root primary
   ```

   # Enable the spanning tree feature globally and in VLAN 10, VLAN 20, and VLAN 30.

   ```
   [DeviceA] stp global enable
   [DeviceA] stp vlan 10 20 30 enable
   ```
3. Configure Device B:

   # Set the spanning tree mode to PVST.

   ```
   <DeviceB> system-view
   [DeviceB] stp mode pvst
   ```

   # Configure the device as the root bridge of VLAN 30.

   ```
   [DeviceB] stp vlan 30 root primary
   ```

   # Enable the spanning tree feature globally and in VLAN 10, VLAN 20, and VLAN 30.

   ```
   [DeviceB] stp global enable
   [DeviceB] stp vlan 10 20 30 enable
   ```
4. Configure Device C:

Set the spanning tree mode to PVST.

```
<DeviceC> system-view
[DeviceC] stp mode pvst
```

Configure the device as the root bridge of VLAN 40.

```
[DeviceC] stp vlan 40 root primary
```

Enable the spanning tree feature globally and in VLAN 10, VLAN 20, and VLAN 40.

```
[DeviceC] stp global enable
[DeviceC] stp vlan 10 20 40 enable
```

5. Configure Device D:

Set the spanning tree mode to PVST.

```
<DeviceD> system-view
[DeviceD] stp mode pvst
```

Enable the spanning tree feature globally and in VLAN 20, VLAN 30, and VLAN 40.

```
[DeviceD] stp global enable
[DeviceD] stp vlan 20 30 40 enable
```

Verifying the configuration

When the network is stable, you can use the **`display stp brief`** command to display brief spanning tree information on each device.

Display brief spanning tree information on Device A.

```
[DeviceA] display stp brief
 VLAN ID     Port                          Role  STP State     Protection
 10          GigabitEthernet1/0/1          DESI  FORWARDING    NONE
 10          GigabitEthernet1/0/3          DESI  FORWARDING    NONE
 20          GigabitEthernet1/0/1          DESI  FORWARDING    NONE
 20          GigabitEthernet1/0/2          DESI  FORWARDING    NONE
 20          GigabitEthernet1/0/3          DESI  FORWARDING    NONE
 30          GigabitEthernet1/0/2          DESI  FORWARDING    NONE
 30          GigabitEthernet1/0/3          ROOT  FORWARDING    NONE
```

Display brief spanning tree information on Device B.

```
[DeviceB] display stp brief
 VLAN ID     Port                          Role  STP State     Protection
 10          GigabitEthernet1/0/2          DESI  FORWARDING    NONE
 10          GigabitEthernet1/0/3          ROOT  FORWARDING    NONE
 20          GigabitEthernet1/0/1          DESI  FORWARDING    NONE
 20          GigabitEthernet1/0/2          DESI  FORWARDING    NONE
 20          GigabitEthernet1/0/3          ROOT  FORWARDING    NONE
 30          GigabitEthernet1/0/1          DESI  FORWARDING    NONE
 30          GigabitEthernet1/0/3          DESI  FORWARDING    NONE
```

Display brief spanning tree information on Device C.

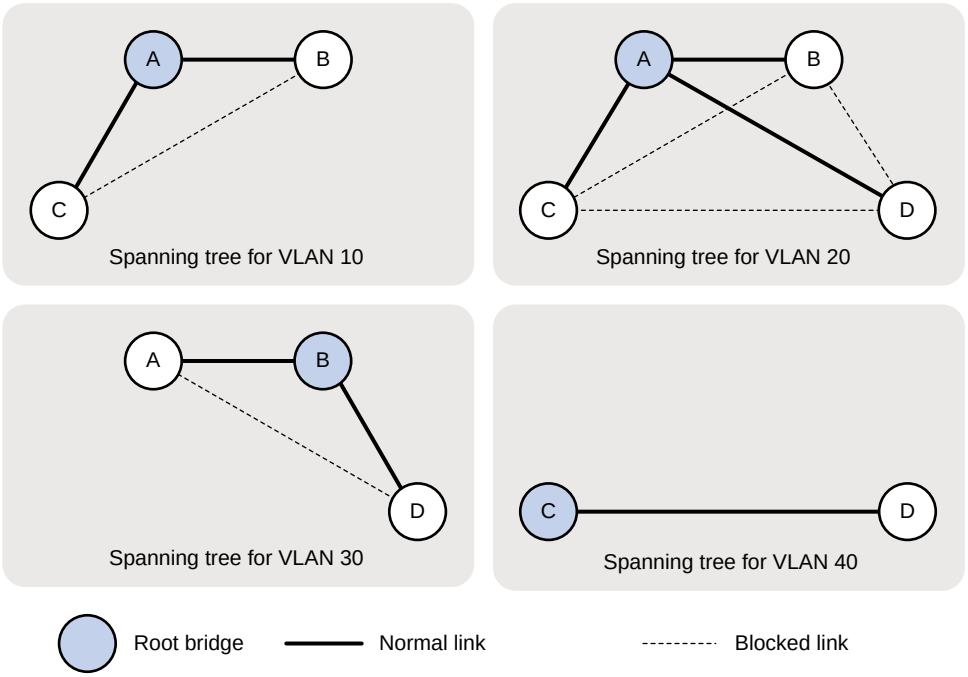
```
[DeviceC] display stp brief
 VLAN ID     Port                          Role  STP State     Protection
 10          GigabitEthernet1/0/1          ROOT  FORWARDING    NONE
 10          GigabitEthernet1/0/2          ALTE  DISCARDING    NONE
 20          GigabitEthernet1/0/1          ROOT  FORWARDING    NONE
 20          GigabitEthernet1/0/2          ALTE  DISCARDING    NONE
 20          GigabitEthernet1/0/3          DESI  FORWARDING    NONE
 40          GigabitEthernet1/0/3          DESI  FORWARDING    NONE
```

Display brief spanning tree information on Device D.

```
[DeviceD] display stp brief
 VLAN ID     Port                          Role  STP State     Protection
 20          GigabitEthernet1/0/1          ALTE  DISCARDING    NONE
 20          GigabitEthernet1/0/2          ROOT  FORWARDING    NONE
 20          GigabitEthernet1/0/3          ALTE  DISCARDING    NONE
 30          GigabitEthernet1/0/1          ROOT  FORWARDING    NONE
 30          GigabitEthernet1/0/2          ALTE  DISCARDING    NONE
 40          GigabitEthernet1/0/3          ROOT  FORWARDING    NONE
```

Based on the output, you can draw a topology for each VLAN spanning tree, as shown in Figure 29.

Figure 29 VLAN spanning tree topologies



Example: Configuring M-LAG with PVST

Network configuration

As shown in Figure 30, Device A and Device B work at the distribution layer, and Device C and Device D work at the access layer.

Configure M-LAG on Device A and Device B. In the M-LAG system, Device A is the primary M-LAG device, and Device B is the secondary M-LAG device.

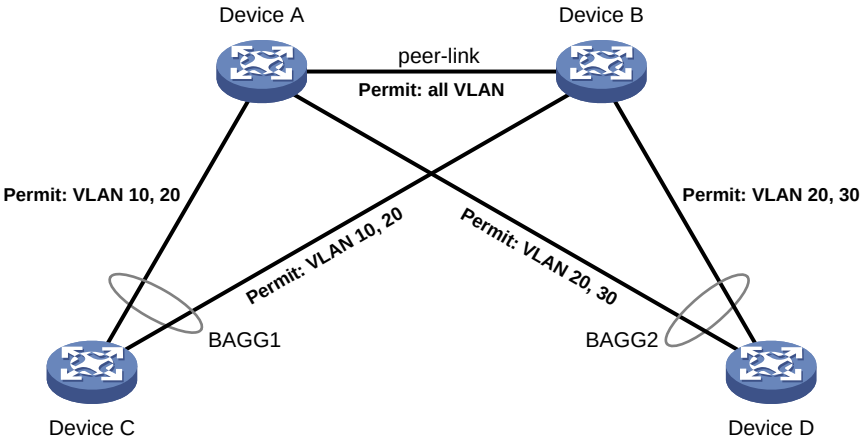
Configure PVST on the devices to meet the following requirements:

- Frames of a VLAN are forwarded along the spanning trees of the VLAN.
- VLAN 10, VLAN 20, and VLAN 30 are terminated on the distribution layer devices.
- The root bridge of VLAN 10, VLAN 20, and VLAN 30 is the M-LAG system formed by Device A and Device B.

NOTE:

- As a best practice, do not connect ports on Device A and Device B that have the same port ID with each other, for example Layer 2 aggregate ports. Otherwise, when Device A and Device B communicate through the link, the spanning tree protocol determines that the device receives its own BPDUs. Loop guard will block the link, though spanning tree features are not affected.
- You can view port IDs of interfaces on the device by using the **`display stp interface`** command.
- The preceding restrictions do not apply to peer-link interfaces and their member ports.

Figure 30 Network diagram



Procedure

1. Configure VLANs and VLAN member ports. (Details not shown.)
   - Create VLAN 10, VLAN 20, and VLAN 30 on both Device A and Device B.
   - Create VLAN 10, and VLAN 20 on Device C.
   - Create VLAN 20, and VLAN 30 on Device D.
   - Configure the ports on these devices as trunk ports and assign them to related VLANs.
2. Configure M-LAG on Device A and Device B. (Details not shown.)

   For more information about M-LAG, see "Configuring M-LAG."
3. Configure Device A:

   # Set the spanning tree mode to PVST.

   ```
   <DeviceA> system-view
   [DeviceA] stp mode pvst
   ```

   # Configure the device as the root bridge of VLAN 10 and VLAN 20.

   ```
   [DeviceA] stp vlan 10 20 root primary
   ```

   # Enable the spanning tree feature globally and in VLAN 10, VLAN 20, and VLAN 30.

   ```
   [DeviceA] stp global enable
   [DeviceA] stp vlan 10 20 30 enable
   ```
4. Configure Device B in the same way Device A is configured. (Details not shown.)
5. Configure Device C:

   # Set the spanning tree mode to PVST.

   ```
   <DeviceC> system-view
   [DeviceC] stp mode pvst
   ```

   # Enable the spanning tree feature globally and in VLAN 10, and VLAN 20.

   ```
   [DeviceC] stp global enable
   [DeviceC] stp vlan 10 20 enable
   ```

6. Configure Device D:

   # Set the spanning tree mode to PVST.

```
<DeviceD> system-view
[DeviceD] stp mode pvst
```

   # Enable the spanning tree feature globally and in VLAN 20, and VLAN 30.

```
[DeviceD] stp global enable
[DeviceD] stp vlan 20 30 enable
```

Verifying the configuration

When the network is stable, you can use the **display stp brief** command to display brief spanning tree information on each device.

Display brief spanning tree information of the M-LAG system on the primary M-LAG device, Device A.

```
[DeviceA] display stp brief
VLAN ID     Port                          Role  STP State     Protection
 10         Bridge-Aggregation1 (M-LAG)   DESI  FORWARDING    NONE
 20         Bridge-Aggregation1 (M-LAG)   DESI  FORWARDING    NONE
 20         Bridge-Aggregation2 (M-LAG)   DESI  FORWARDING    NONE
 30         Bridge-Aggregation2 (M-LAG)   DESI  FORWARDING    NONE
```

Display brief spanning tree information on Device C.

```
[DeviceC] display stp brief
VLAN ID     Port                          Role  STP State     Protection
 10         Bridge-Aggregation1           ROOT  FORWARDING    NONE
 20         Bridge-Aggregation1           ROOT  FORWARDING    NONE
```

Display brief spanning tree information on Device D.

```
[DeviceD] display stp brief
VLAN ID     Port                          Role  STP State     Protection
 20         Bridge-Aggregation2           ROOT  FORWARDING    NONE
 30         Bridge-Aggregation2           ROOT  FORWARDING    NONE
```

Contents

Configuring LLDP 1

About LLDP 1
LLDP agents and bridge modes 1
LLDP frame formats 2
LLDPDUs 3
TLVs 3
Management address 6
LLDP operating modes 6
Transmitting and receiving LLDP frames 7
Collaboration with Track 7
Protocols and standards 7
Restrictions and guidelines: LLDP configuration 8
LLDP tasks at a glance 8
Enabling LLDP 9
Setting the LLDP bridge mode 9
Setting the LLDP operating mode 10
Setting the LLDP reinitialization delay 10
Configuring the advertisable TLVs 10
Configuring advertisement of the management address TLV 13
Setting the encapsulation format for LLDP frames 14
Setting LLDP frame transmission parameters 15
Setting the timeout for receiving LLDP frames 16
Configuring port ID subtype of port ID TLVs advertised by LLDP 16
Enabling LLDP polling 17
Disabling LLDP PVID inconsistency check 17
Configuring CDP compatibility 17
Configuring LLDP trapping and LLDP-MED trapping 19
Configuring MAC address learning for DCN 20
About MAC address learning for DCN 20
Setting the source MAC address of LLDP frames 20
Enabling generation of ARP or ND entries for received management address TLVs 21
Enabling transparent transmission of LLDP frames 22
Configuring LLDP black hole detection 22
Configuring LLDP cross-domain detection 25
Disabling AP restart through LLDP 27
Disabling LLDP control over PoE 28
Display and maintenance commands for LLDP 28
LLDP configuration examples 29
Example: Configuring basic LLDP functions 29
Example: Configuring LLDP cross-domain detection 33

Configuring DCBX 39

About DCBX 39
DCBX versions 39
DCBX functions 39
DCBX application scenario 39
Protocols and standards 40
DCBX tasks at a glance 40
Enabling LLDP and DCBX TLV advertising 40
Setting the DCBX version 41
Configuring PFC parameters 41
DCBX configuration examples 42
Example: Configuring DCBX 42

Configuring LLDP

About LLDP

The Link Layer Discovery Protocol (LLDP) is a standard link layer protocol that allows network devices from different vendors to discover neighbors and exchange system and configuration information.

In an LLDP-enabled network, a device advertises local device information in LLDP Data Units (LLDPDUs) to the directly connected devices. The information distributed through LLDP is stored by its recipients in standard MIBs, making it possible for the information to be accessed by a Network Management System (NMS) through SNMP.

Information that can be distributed through LLDP includes (but is not limited to):

- Major capabilities of the system.
- Management IP address of the system.
- Device ID.
- Port ID.

LLDP agents and bridge modes

An LLDP agent is a mapping of a protocol entity that implements LLDP. Multiple LLDP agents can run on the same interface.

LLDP agents are classified into the following types:

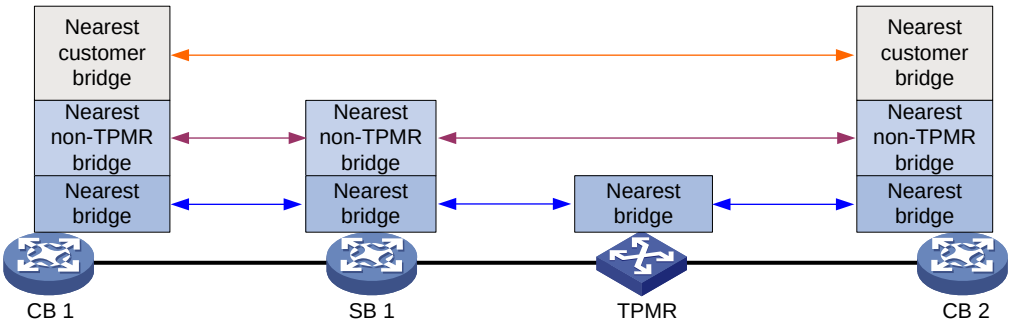
- Nearest bridge agent.
- Nearest customer bridge agent.
- Nearest non-TPMR bridge agent.

  A Two-port MAC Relay (TPMR) is a type of bridge that has only two externally-accessible bridge ports. It supports a subset of the features of a MAC bridge. A TPMR is transparent to all frame-based media-independent protocols except for the following protocols:

  - Protocols destined for the TPMR.
  - Protocols destined for reserved MAC addresses that the relay feature of the TPMR is configured not to forward.

LLDP exchanges packets between neighbor agents and creates and maintains neighbor information for them. Figure 1 shows the neighbor relationships for these LLDP agents.

Figure 1 LLDP neighbor relationships



The types of supported LLDP agents vary with the bridge mode in which LLDP operates. LLDP supports the following bridge modes: customer bridge (CB) and service bridge (SB).

- **Customer bridge mode**—LLDP supports nearest bridge agent, nearest non-TPMR bridge agent, and nearest customer bridge agent. LLDP processes the LLDP frames with destination MAC addresses for these agents and transparently transmits the LLDP frames with other destination MAC addresses in VLANs.
- **Service bridge mode**—LLDP supports nearest bridge agent and nearest non-TPMR bridge agent. LLDP processes the LLDP frames with destination MAC addresses for these agents and transparently transmits the LLDP frames with other destination MAC addresses in VLANs.

LLDP frame formats

LLDP sends device information in LLDP frames. LLDP frames are encapsulated in Ethernet II or Subnetwork Access Protocol (SNAP) format.

LLDP frame encapsulated in Ethernet II

Figure 2 Ethernet II-encapsulated LLDP frame

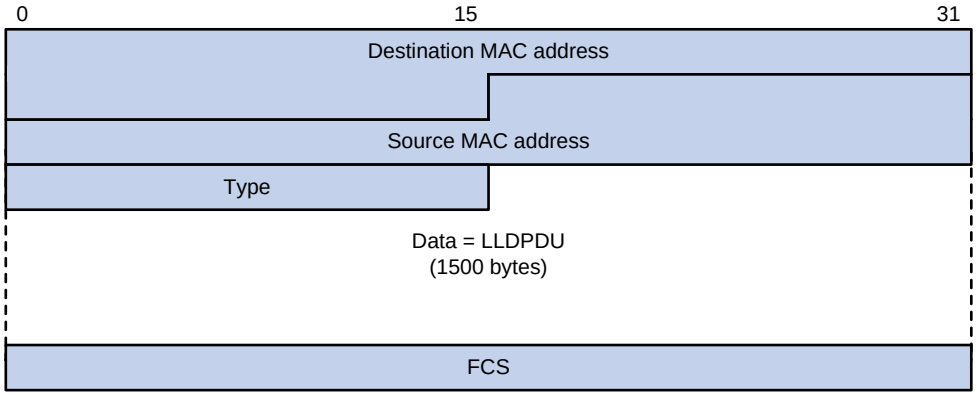


Table 1 Fields in an Ethernet II-encapsulated LLDP frame

Field	Description
Destination MAC address	MAC address to which the LLDP frame is advertised. LLDP specifies different multicast MAC addresses as destination MAC addresses for LLDP frames destined for agents of different types. This helps distinguish between LLDP frames sent and received by different agent types on the same interface. The destination MAC address is fixed to one of the following multicast MAC addresses: • 0x0180-c200-000E for LLDP frames destined for nearest bridge agents. • 0x0180-c200-0000 for LLDP frames destined for nearest customer bridge agents. • 0x0180-c200-0003 for LLDP frames destined for nearest non-TPMR bridge agents.
Source MAC address	MAC address of the sending port.
Type	Ethernet type for the upper-layer protocol. This field is 0x88CC for LLDP.
Data	LLDPDU.
FCS	Frame check sequence, a 32-bit CRC value used to determine the validity of the received Ethernet frame.

LLDP frame encapsulated in SNAP

Figure 3 SNAP-encapsulated LLDP frame

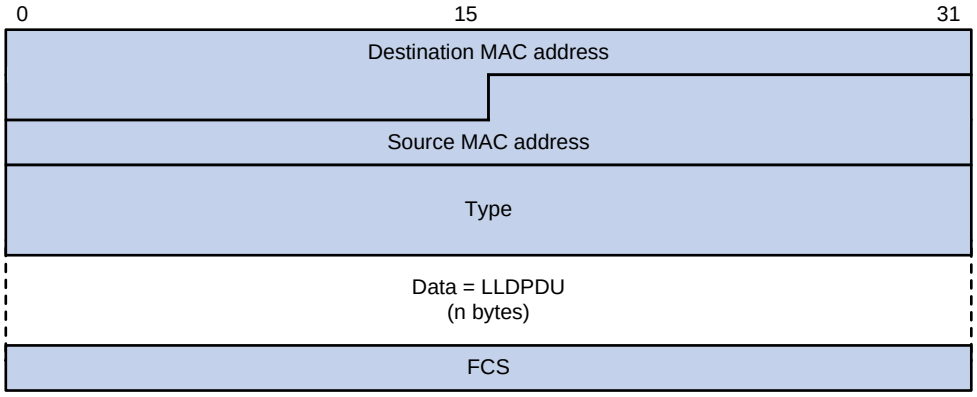


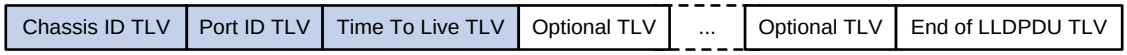
Table 2 Fields in a SNAP-encapsulated LLDP frame

Field	Description
Destination MAC address	MAC address to which the LLDP frame is advertised. It is the same as that for Ethernet II-encapsulated LLDP frames.
Source MAC address	MAC address of the sending port.
Type	SNAP type for the upper-layer protocol. This field is 0xAAAA-0300-0000-88CC for LLDP.
Data	LLDPDU.
FCS	Frame check sequence, a 32-bit CRC value used to determine the validity of the received Ethernet frame.

LLDPDUs

Each LLDP frame contains one LLDPDU. Each LLDPDU is a sequence of type-length-value (TLV) structures.

Figure 4 LLDPDU encapsulation format



As shown in Figure 4, each LLDPDU starts with the following mandatory TLVs: Chassis ID TLV, Port ID TLV, and Time to Live TLV. The mandatory TLVs are followed by a maximum of 29 optional TLVs.

TLVs

A TLV is an information element that contains the type, length, and value fields.

LLDPDU TLVs include the following categories:

- Basic management TLVs.
- Organizationally (IEEE 802.1 and IEEE 802.3) specific TLVs.
- LLDP-MED (media endpoint discovery) TLVs.

Basic management TLVs are essential to device management.

Organizationally specific TLVs and LLDP-MED TLVs are used for enhanced device management. They are defined by standardization or other organizations and are optional for LLDPDUs.

Basic management TLVs

Table 3 lists the basic management TLV types. Some of them are mandatory for LLDPDUs.

Table 3 Basic management TLVs

Type	Description	Remarks
Chassis ID	Specifies the bridge MAC address of the sending device.	Mandatory.
Port ID	Specifies the ID of the sending port: • If the LLDPDU carries LLDP-MED TLVs, the port ID TLV carries the MAC address of the sending port. • Otherwise, the port ID TLV carries the port name.	
Time to Live	Specifies the life of the transmitted information on the receiving device.	
End of LLDPDU	Marks the end of the TLV sequence in the LLDPDU.	
Port Description	Specifies the description for the sending port.	Optional.
System Name	Specifies the assigned name of the sending device.	
System Description	Specifies the description for the sending device.	
System Capabilities	Identifies the primary features of the sending device and the enabled primary features.	
Management Address	Specifies the following elements: • The management address of the local device. • The interface number and object identifier (OID) associated with the address.	

IEEE 802.1 organizationally specific TLVs

Table 4 lists the IEEE 802.1 organizationally specific TLVs.

The device can receive protocol identity TLVs and VID usage digest TLVs, but it cannot send these TLVs.

Layer 3 Ethernet ports support only link aggregation TLVs.

Table 4 IEEE 802.1 organizationally specific TLVs

Type	Description
Port VLAN ID (PVID)	Specifies the port PVID.
Port And Protocol VLAN ID (PPVID)	Indicates whether the device supports protocol VLANs and, if so, what VLAN IDs these protocols will be associated with.
VLAN Name	Specifies the textual name of any VLAN to which the port belongs.
Protocol Identity	Indicates protocols supported on the port.
DCBX	Data center bridging exchange protocol.
EVB module	Edge Virtual Bridging module, including EVB TLV and CDCP TLV. EVB module TLVs are not supported in the current software version.
Link Aggregation	Indicates whether the port supports link aggregation, and if yes, whether link aggregation is enabled.

Type	Description
Management VID	Management VLAN ID.
VID Usage Digest	VLAN ID usage digest.
ETS Configuration	Enhanced Transmission Selection configuration. ETS configuration TLVs are not supported in the current software version.
ETS Recommendation	ETS recommendation. ETS recommendation TLVs are not supported in the current software version.
PFC	Priority-based Flow Control. PFC TLVs are not supported in the current software version.
APP	Application protocol. APP TLVs are not supported in the current software version.
QCN	Quantized Congestion Notification. QCN TLVs are not supported in the current software version.

IEEE 802.3 organizationally specific TLVs

Table 5 shows the IEEE 802.3 organizationally specific TLVs.

The Power Stateful Control TLV is defined in IEEE P802.3at D1.0 and is not supported in later versions. The device sends this type of TLVs only after receiving them.

Table 5 IEEE 802.3 organizationally specific TLVs

Type	Description
MAC/PHY Configuration/Status	Contains the bit-rate and duplex capabilities of the port, support for autonegotiation, enabling status of autonegotiation, and the current rate and duplex mode.
Link Aggregation	Indicates whether the port supports link aggregation, and if yes, whether link aggregation is enabled.
Power Via MDI	Contains the power supply capabilities of the port: Power Via MDI TLVs are not supported in the current software version.
Maximum Frame Size	Indicates the supported maximum frame size.
Power Stateful Control	Indicates the power state control configured on the sending port, including the following: • Power supply mode of the PSE/PD. • PSE/PD priority. • PSE/PD power. Power Stateful Control TLVs are not supported in the current software version.
Energy-Efficient Ethernet	Indicates Energy Efficient Ethernet (EEE). Energy-Efficient Ethernet TLVs are not supported in the current software version.

LLDP-MED TLVs

LLDP-MED TLVs provide multiple advanced applications for voice over IP (VoIP), such as basic configuration, network policy configuration, and address and directory management. LLDP-MED

TLVs provide a cost-effective and easy-to-use solution for deploying voice devices in Ethernet. LLDP-MED TLVs are shown in Table 6.

If the MAC/PHY configuration/status TLV is not advertisable, none of the LLDP-MED TLVs will be advertised even if they are advertisable.

If the LLDP-MED capabilities TLV is not advertisable, the other LLDP-MED TLVs will not be advertised even if they are advertisable.

Table 6 LLDP-MED TLVs

Type	Description
LLDP-MED Capabilities	Allows a network device to advertise the LLDP-MED TLVs that it supports.
Network Policy	Allows a network device or terminal device to advertise the VLAN ID of a port, the VLAN type, and the Layer 2 and Layer 3 priorities for specific applications.
Extended Power-via-MDI	Allows a network device or terminal device to advertise power supply capability. This TLV is an extension of the Power Via MDI TLV.
Hardware Revision	Allows a terminal device to advertise its hardware version.
Firmware Revision	Allows a terminal device to advertise its firmware version.
Software Revision	Allows a terminal device to advertise its software version.
Serial Number	Allows a terminal device to advertise its serial number.
Manufacturer Name	Allows a terminal device to advertise its vendor name.
Model Name	Allows a terminal device to advertise its model name.
Asset ID	Allows a terminal device to advertise its asset ID. The typical case is that the user specifies the asset ID for the endpoint to facilitate directory management and asset tracking.
Location Identification	Allows a network device to advertise the appropriate location identifier information for a terminal device to use in the context of location-based applications.

Management address

The network management system uses the management address of a device to identify and manage the device for topology maintenance and network management. The management address is encapsulated in the management address TLV.

LLDP operating modes

An LLDP agent can operate in one of the following modes:

- **TxRx mode**—An LLDP agent in this mode can send and receive LLDP frames.
- **Tx mode**—An LLDP agent in this mode can only send LLDP frames.
- **Rx mode**—An LLDP agent in this mode can only receive LLDP frames.
- **Disable mode**—An LLDP agent in this mode cannot send or receive LLDP frames.

Each time the operating mode of an LLDP agent changes, its LLDP protocol state machine reinitializes. A configurable reinitialization delay prevents frequent initializations caused by frequent

changes to the operating mode. If you configure the reinitialization delay, an LLDP agent must wait the specified amount of time to initialize LLDP after the LLDP operating mode changes.

Transmitting and receiving LLDP frames

Transmitting LLDP frames

An LLDP agent operating in TxRx mode or Tx mode sends LLDP frames to its directly connected devices both periodically and when the local configuration changes. To prevent LLDP frames from overwhelming the network during times of frequent changes to local device information, LLDP uses the token bucket mechanism to rate limit LLDP frames. For more information about the token bucket mechanism, see *ACL and QoS Configuration Guide*.

LLDP automatically enables the fast LLDP frame transmission mechanism in either of the following cases:

- A new LLDP frame is received and carries device information new to the local device.
- The LLDP operating mode of the LLDP agent changes from Disable or Rx to TxRx or Tx.

The fast LLDP frame transmission mechanism successively sends the specified number of LLDP frames at a configurable fast LLDP frame transmission interval. The mechanism helps LLDP neighbors discover the local device as soon as possible. Then, the normal LLDP frame transmission interval resumes.

Receiving LLDP frames

An LLDP agent operating in TxRx mode or Rx mode confirms the validity of TLVs carried in every received LLDP frame. If the TLVs are valid, the LLDP agent saves the information and starts an aging timer. The initial value of the aging timer is equal to the TTL value in the Time To Live TLV carried in the LLDP frame. When the LLDP agent receives a new LLDP frame, the aging timer restarts. When the aging timer decreases to zero, all saved information ages out.

Collaboration with Track

You can configure a track entry and associate it with an LLDP interface. The LLDP module checks the neighbor availability of the LLDP interface and reports the check result to the Track module. The Track module changes the track entry status accordingly so the associated application module can take correct actions.

The Track module changes the track entry status based on the neighbor availability of a monitored LLDP interface as follows:

- If the neighbor of the LLDP interface is available, the Track module sets the track entry to Positive state.
- If the neighbor of the LLDP interface is unavailable, the Track module sets the track entry to Negative state.

For more information about collaboration between Track and LLDP, see the track configuration in *High Availability Configuration Guide.*

Protocols and standards

- IEEE 802.1AB-2005, *Station and Media Access Control Connectivity Discovery*
- IEEE 802.1AB-2009, *Station and Media Access Control Connectivity Discovery*
- ANSI/TIA-1057, *Link Layer Discovery Protocol for Media Endpoint Devices*
- IEEE Std 802.1Qaz-2011, *Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks-Amendment 18: Enhanced Transmission Selection for Bandwidth Sharing Between Traffic Classes*

Restrictions and guidelines: LLDP configuration

When you configure LLDP, follow these restrictions and guidelines:

- Some of the LLDP configuration tasks are available in different interface views (see Table 7).

  Table 7 Support of LLDP configuration tasks in different views

Tasks	Supported views
Enabling LLDP	Layer 2 Ethernet interface view Layer 3 Ethernet interface view Management Ethernet interface view Layer 2 aggregate interface view Layer 3 aggregate interface view IRF physical interface view
Setting the LLDP operating mode	
Configuring the advertisable TLVs	
Configuring advertisement of the management address TLV	
Setting the encapsulation format for LLDP frames	
Enabling LLDP polling	
Configuring LLDP trapping	
Configuring port ID subtype of port ID TLVs advertised by LLDP (not supported in Layer 2 aggregate interface view and Layer 3 aggregate interface view)	

- To use LLDP together with OpenFlow, you must enable LLDP globally on OpenFlow switches. To prevent LLDP from affecting topology discovery of OpenFlow controllers, disable LLDP on ports of OpenFlow instances. For more information about OpenFlow, see *OpenFlow Configuration Guide*.
- You can configure LLDP on an IRF physical interface to monitor the connection and link status of the IRF physical link. An LLDP-enabled IRF physical interface supports only the nearest bridge agent.

For transparent transmission of LLDP frames in an EVPN-capable M-LAG system, if the **evpn m-lag local** command is executed on an M-LAG member device, the remote peer will forward multiple LLDP frames to the M-LAG member device. To resolve this issue, you can execute the **undo evpn m-lag local** command on the M-LAG member device. Identify whether the **undo evpn m-lag local** command can be executed according to networking requirements. For more information about support of EVPN for M-LAG, see EVPN VXLAN configuration in *EVPN Configuration Guide*.

LLDP tasks at a glance

To configure LLDP, perform the following tasks:

1. Enabling LLDP
2. Setting the LLDP bridge mode
3. Setting the LLDP operating mode
4. (Optional.) Setting the LLDP reinitialization delay
5. (Optional.) Configuring LLDP packet-related settings
   - Configuring the advertisable TLVs
   - Configuring advertisement of the management address TLV
   - Setting the encapsulation format for LLDP frames
   - Setting LLDP frame transmission parameters

- Setting the timeout for receiving LLDP frames
- Configuring port ID subtype of port ID TLVs advertised by LLDP

6. (Optional.) Configuring LLDP extended features
   - Enabling LLDP polling
   - Disabling LLDP PVID inconsistency check
   - Configuring CDP compatibility
   - Configuring LLDP trapping and LLDP-MED trapping
   - Configuring MAC address learning for DCN
   - Enabling transparent transmission of LLDP frames
   - Configuring LLDP black hole detection
   - Configuring LLDP cross-domain detection
   - Disabling AP restart through LLDP
   - Disabling LLDP control over PoE

Enabling LLDP

Restrictions and guidelines

For LLDP to take effect on a port, you must enable LLDP both globally and on the port.

The enabling status of LLDP on an aggregate interface will not be synchronized to its member ports.

Procedure

7. Enter system view.
   `system-view`
8. Enable LLDP globally.
   `lldp global enable`
   By default:
   - If the device starts up with the initial configuration, LLDP is disabled globally.
   - If the device starts up with the factory defaults, LLDP is enabled globally.

   For more information about device startup with the initial configuration or factory defaults, see configuration file management in *Fundamentals Configuration Guide*.
9. Enter interface view.
   `interface` *`interface-type interface-number`*
10. Enable LLDP.
    `lldp enable`
    By default, LLDP is disabled on ONU subinterfaces and is enabled on the other interfaces.

Setting the LLDP bridge mode

11. Enter system view.
    `system-view`
12. Set the LLDP bridge mode.
    - Set the LLDP bridge mode to service bridge.
      `lldp mode service-bridge`
      By default, LLDP operates in customer bridge mode.

- Set the LLDP bridge mode to customer bridge.

  undo lldp mode

By default, LLDP operates in customer bridge mode.

Setting the LLDP operating mode

13. Enter system view.

    system-view

14. Enter interface view.

    interface *interface-type interface-number*

15. Set the LLDP operating mode.
    - In Layer 2/Layer 3 Ethernet interface view or management Ethernet interface view:

      lldp [**agent** { **nearest-customer** | **nearest-nontpmr** }] **admin-status** { **disable** | **rx** | **tx** | **txrx** }

      In Ethernet interface view, if you do not specify an agent type, the command sets the operating mode for the nearest bridge agent.
    - In Layer 2/Layer 3 aggregate interface view:

      lldp agent { **nearest-customer** | **nearest-nontpmr** } **admin-status** { **disable** | **rx** | **tx** | **txrx** }

      In aggregate interface view, you can set the operating mode only for the nearest customer bridge agent and nearest non-TPMR bridge agent.
    - In IRF physical interface view:

      lldp admin-status { **disable** | **rx** | **tx** | **txrx** }

      In IRF physical interface view, you can set the operating mode only for the nearest bridge agent.

    By default, the nearest bridge agent operates in **TxRx** mode, and the other agents operate in **Disable** mode.

Setting the LLDP reinitialization delay

About this task

When the LLDP operating mode changes on a port, the port initializes the protocol state machines after an LLDP reinitialization delay. By adjusting the delay, you can avoid frequent initializations caused by frequent changes to the LLDP operating mode on a port.

Procedure

16. Enter system view.

    system-view

17. Set the LLDP reinitialization delay.

    lldp timer reinit-delay *delay*

    The default LLDP reinitialization delay is 2 seconds.

Configuring the advertisable TLVs

18. Enter system view.

    system-view

19. Enter interface view.

    interface *interface-type interface-number*

20. Configure the advertisable TLVs.

    - In Layer 2 Ethernet interface view:

      lldp tlv-enable { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** | **management-address-tlv** [**ipv6**] [*ip-address* | **interface loopback** *interface-number*] } | **dot1-tlv** { **all** | **port-vlan-id** | **link-aggregation** | **dcbx** | **protocol-vlan-id** [*vlan-id*] | **vlan-name** [*vlan-id-list*] | **management-vid** [*mvlan-id*] } | **dot3-tlv** { **all** | **link-aggregation** | **mac-physic** | **max-frame-size** | **power** } | **med-tlv** { **all** | **capability** | **inventory** | **network-policy** [*vlan-id*] | **power-over-ethernet** | **location-id** { **civic-address** *device-type country-code* { *ca-type ca-value* }&<1-10> | **elin-address** *tel-number* } } }

      By default, the nearest bridge agent advertises all supported TLVs except the following TLVs:

      - DCBX TLVs.
      - Location identification TLVs.
      - Port and protocol VLAN ID TLVs.
      - VLAN name TLVs.
      - Management VLAN ID TLVs.
      - Link aggregation TLVs.

      lldp agent nearest-nontpmr tlv-enable { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** | **management-address-tlv** [**ipv6**] [*ip-address*] } | **dot1-tlv** { **all** | **port-vlan-id** | **link-aggregation** } | **dot3-tlv** { **all** | **link-aggregation** } }

      lldp tlv-enable dot1-tlv { **protocol-vlan-id** [*vlan-id*] | **vlan-name** [*vlan-id-list*] | **management-vid** [*mvlan-id*] }

      By default, the nearest non-TPMR bridge agent does not advertise any TLVs.

      lldp agent nearest-customer tlv-enable { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** | **management-address-tlv** [**ipv6**] [*ip-address*] } | **dot1-tlv** { **all** | **port-vlan-id** | **link-aggregation** } | **dot3-tlv** { **all** | **link-aggregation** } }

      lldp tlv-enable dot1-tlv { **protocol-vlan-id** [*vlan-id*] | **vlan-name** [*vlan-id-list*] | **management-vid** [*mvlan-id*] }

      By default, the nearest customer bridge agent advertises all the supported basic management TLVs and port VLAN ID TLVs and link aggregation TLVs in the IEEE 802.1 organizationally specific TLV set.

    - In Layer 3 Ethernet interface view:

      lldp tlv-enable { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** | **management-address-tlv** [**ipv6**] [*ip-address* | **interface loopback** *interface-number*] } | **dot1-tlv** { **all** | **link-aggregation** } | **dot3-tlv** { **all** | **link-aggregation** | **mac-physic** | **max-frame-size** | **power** } | **med-tlv** { **all** | **capability** | **inventory** | **power-over-ethernet** | **location-id** { **civic-address** *device-type country-code* { *ca-type ca-value* }&<1-10> | **elin-address** *tel-number* } } }

By default, the nearest bridge agent advertises the following TLVs:

- Link aggregation TLVs in the 802.1 organizationally specific TLV set.
- All supported 802.3 organizationally specific TLVs except the Energy-Efficient Ethernet TLVs.
- All supported LLDP-MED TLVs except the network policy TLVs.

lldp agent { **nearest-nontpmr** | **nearest-customer** } **tlv-enable** { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** | **management-address-tlv** [**ipv6**] [*ip-address*] } | **dot1-tlv** { **all** | **link-aggregation** } | **dot3-tlv** { **all** | **link-aggregation** } }

By default:

- The nearest non-TPMR bridge agent does not advertise any TLVs.
- The nearest customer bridge agent advertises all supported basic management TLVs and link aggregation TLVs in the IEEE 802.1 organizationally specific TLV set.

- In management Ethernet interface view:

  lldp tlv-enable { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** | **management-address-tlv** [**ipv6**] [*ip-address*] } | **dot1-tlv** { **all** | **link-aggregation** } | **dot3-tlv** { **all** | **link-aggregation** | **mac-physic** | **max-frame-size** | **power** } | **med-tlv** { **all** | **capability** | **inventory** | **power-over-ethernet** | **location-id** { **civic-address** *device-type country-code* { *ca-type ca-value* }&<1-10> | **elin-address** *tel-number* } } }

  By default, the nearest bridge agent advertises the following TLVs:

  - All supported basic management TLVs.
  - Link aggregation TLVs in the 802.1 organizationally specific TLV set.
  - All supported 802.3 organizationally specific TLVs except the link aggregation TLVs.
  - All supported LLDP-MED TLVs except the network policy TLVs and location Identification TLVs.

  lldp agent { **nearest-nontpmr** | **nearest-customer** } **tlv-enable** { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** | **management-address-tlv** [**ipv6**] [*ip-address*] } | **dot1-tlv** { **all** | **link-aggregation** } | **dot3-tlv** { **all** | **link-aggregation** } }

  By default:

  - The nearest non-TPMR bridge agent does not advertise any TLVs.
  - The nearest customer bridge agent advertises all supported basic management TLVs and link aggregation TLVs in the IEEE 802.1 organizationally specific TLV set.

- In Layer 2 aggregate interface view:

  lldp tlv-enable dot1-tlv { **protocol-vlan-id** [*vlan-id*] | **vlan-name** [*vlan-id-list*] | **management-vid** [*mvlan-id*]

  lldp agent nearest-nontpmr tlv-enable { **basic-tlv** { **all** | **management-address-tlv** [**ipv6**] [*ip-address* | **interface loopback** *interface-number*] | **port-description** | **system-capability** | **system-description** | **system-name** } | **dot1-tlv** { **all** | **port-vlan-id** } }

  By default, the nearest non-TPMR bridge agent does not advertise any TLVs.

  lldp agent nearest-customer tlv-enable { **basic-tlv** { **all** | **management-address-tlv** [**ipv6**] [*ip-address*] | **port-description** |

system-capability | **system-description** | **system-name** } | **dot1-tlv** { **all** | **port-vlan-id** } }

By default, the nearest customer bridge agent advertises all supported basic management TLVs and port VLAN ID TLVs in the IEEE 802.1 organizationally specific TLV set.

The nearest bridge agent is not supported.

- In Layer 3 aggregate interface view:

  lldp agent { **nearest-customer** | **nearest-nontpmr** } **tlv-enable basic-tlv** { **all** | **management-address-tlv** [**ipv6**] [*ip-address*] | **port-description** | **system-capability** | **system-description** | **system-name** }

  By default:

  - The nearest non-TPMR bridge agent does not advertise any TLVs.
  - The nearest customer bridge agent advertises all supported basic management TLVs.

  The nearest bridge agent is not supported.

- In IRF physical interface view:

  lldp tlv-enable basic-tlv { **port-description** | **system-capability** | **system-description** | **system-name** }

  By default, the nearest bridge agent advertises all supported basic management TLVs.

  Only the nearest bridge agent is supported.

Configuring advertisement of the management address TLV

About this task

LLDP encodes management addresses in numeric or string format in management address TLVs.

If a neighbor encodes its management address in string format, set the encoding format of the management address to **string** on the connecting port. This guarantees normal communication with the neighbor.

You can configure advertisement of the management address TLV globally or on a per-interface basis. The device selects the management address TLV advertisement setting for an interface in the following order:

21. Interface-based setting, configured by using the **lldp tlv-enable** command with the **management-address-tlv** keyword.
22. Global setting, configured by using the **lldp global tlv-enable basic-tlv management-address-tlv** command.
23. Default setting for the interface.

    By default:

    - The nearest bridge agent and nearest customer bridge agent advertise the management address TLV.
    - The nearest non-TPMR bridge agent does not advertise the management address TLV.

Procedure

24. Enter system view.

    system-view

25. Enable advertisement of the management address TLV globally and set the management address to be advertised.

lldp [**agent** { **nearest-customer** | **nearest-nontpmr** }] **global tlv-enable basic-tlv management-address-tlv** [**ipv6**] { *ip-address* | **interface loopback** *interface-number* | **interface m-gigabitethernet** *interface-number* | **interface vlan-interface** *interface-number* }

By default, advertisement of the management address TLV is disabled globally.

26. Enter interface view.

interface *interface-type interface-number*

27. Enable advertisement of the management address TLV on the interface and set the management address to be advertised.

- In Layer 2 Ethernet interface view:

  lldp tlv-enable basic-tlv management-address-tlv [**ipv6**] [*ip-address* | **interface loopback** *interface-number*]

  lldp agent { **nearest-customer** | **nearest-nontpmr** } **tlv-enable basic-tlv management-address-tlv** [**ipv6**] [*ip-address*]

- In management interface view:

  lldp [**agent** { **nearest-customer** | **nearest-nontpmr** }] **tlv-enable basic-tlv management-address-tlv** [**ipv6**] [*ip-address*]

- In Layer 3 Ethernet interface view:

  lldp tlv-enable basic-tlv management-address-tlv [**ipv6**] [*ip-address* | **interface loopback** *interface-number*]

  lldp agent { **nearest-customer** | **nearest-nontpmr** } **tlv-enable basic-tlv management-address-tlv** [**ipv6**] [*ip-address*]

- In Layer 2/Layer 3 aggregate interface view:

  lldp agent { **nearest-customer** | **nearest-nontpmr** } **tlv-enable basic-tlv management-address-tlv** [**ipv6**] [*ip-address*]

By default:

- The nearest bridge agent and nearest customer bridge agent advertise the management address TLVs.
- The nearest non-TPMR bridge agent does not advertise the management address TLV.

28. Set the encoding format of the management address to string.

- In Layer 2/Layer 3 Ethernet interface view or management Ethernet interface view:

  lldp [**agent** { **nearest-customer** | **nearest-nontpmr** }] **management-address-format string**

- In Layer 2/Layer 3 aggregate interface view:

  lldp agent { **nearest-customer** | **nearest-nontpmr** } **management-address-format string**

The default management address encoding format is numeric.

The device supports only the numeric encoding format for IPv6 management addresses.

Setting the encapsulation format for LLDP frames

About this task

Earlier versions of LLDP require the same encapsulation format on both ends to process LLDP frames. To successfully communicate with a neighboring device running an earlier version of LLDP, the local device must be set with the same encapsulation format.

Procedure

29. Enter system view.

system-view

30. Enter interface view.
    interface *interface-type interface-number*
31. Set the encapsulation format for LLDP frames to SNAP.
    - In Layer 2/Layer 3 Ethernet interface view or management Ethernet interface view:
      lldp [**agent** { **nearest-customer** | **nearest-nontpmr** }] **encapsulation snap**
    - In Layer 2/Layer 3 aggregate interface view:
      lldp agent { **nearest-customer** | **nearest-nontpmr** } **encapsulation snap**
    - In IRF physical interface view:
      lldp encapsulation snap

    By default, the Ethernet II encapsulation format is used.

Setting LLDP frame transmission parameters

About this task

The Time to Live TLV carried in an LLDPDU determines how long the device information carried in the LLDPDU can be saved on a recipient device.

By setting the TTL multiplier, you can configure the TTL of locally sent LLDPDUs. The TTL is expressed by using the following formula:

TTL = Min (65535, (TTL multiplier × LLDP frame transmission interval + 1))

As the expression shows, the TTL can be up to 65535 seconds. TTLs greater than 65535 will be rounded down to 65535 seconds.

Procedure

32. Enter system view.
    system-view
33. Set the TTL multiplier.
    lldp hold-multiplier *value*
    The default setting is 4.
34. Set the LLDP frame transmission interval.
    lldp timer tx-interval *interval*
    The default setting is 30 seconds.
35. Set the token bucket size for sending LLDP frames.
    lldp max-credit *credit-value*
    The default setting is 5.
36. Set the number of LLDP frames sent each time fast LLDP frame transmission is triggered.
    lldp fast-count *count*
    The default setting is 4.
37. Set the fast LLDP frame transmission interval.
    lldp timer fast-interval *interval*
    The default setting is 1 second.

Setting the timeout for receiving LLDP frames

About this task

This feature allows the device to detect the presence of directly connected neighbors by setting the timeout timer for receiving LLDP frames. If an interface has not received any frames when the timeout timer expires, the device reports a no LLDP neighbor event to the NETCONF module.

Restrictions and guidelines

To avoid misdetection, make sure the timeout for receiving LLDP frames is greater than the LLDP frame transmission interval.

Procedure

38. Enter system view.

system-view

39. Set the timeout for receiving LLDP frames.

lldp timer rx-timeout *timeout*

By default, no timeout is set for receiving LLDP frames, and the device does not report no LLDP neighbor events.

Configuring port ID subtype of port ID TLVs advertised by LLDP

About this task

This task enables an H3C device to advertise only port ID TLVs that contain interface names. By default, an H3C device advertises port ID TLVs that contain interface MAC addresses or interface names. The media devices from some vendors can obtain interface information from H3C devices only through LLDP. For the media devices to obtain interface names, you must configure H3C devices to generate port ID TLVs that use interface names as the port ID subtype.

Restrictions and guidelines

Perform this task only when LLDP neighbors must obtain interface names from LLDPDUs. Do not perform this task in any other scenarios.

Procedure

40. Enter system view.

system-view

41. Enter Layer 2 Ethernet interface view/Layer 3 Ethernet interface view/management Ethernet interface view/IRF physical interface view.

interface *interface-type interface-number*

42. Configure the port ID subtype of port ID TLVs advertised by LLDP.

lldp tlv-config basic-tlv port-id *type-id*

By default, an interface advertises port ID TLVs that contain interface MAC addresses if it receives LLDP-MED TLVs, and advertises port ID TLVs that contain interface names if no LLDP-MED TLVs are received.

Enabling LLDP polling

About this task

With LLDP polling enabled, a device periodically searches for local configuration changes. When the device detects a configuration change, it sends LLDP frames to inform neighboring devices of the change.

Procedure

43. Enter system view.

system-view

44. Enter interface view.

interface *interface-type interface-number*

45. Enable LLDP polling and set the polling interval.

- In Layer 2/Layer 3 Ethernet interface view or management Ethernet interface view:

  lldp [**agent** { **nearest-customer** | **nearest-nontpmr** }] **check-change-interval** *interval*

- In Layer 2/Layer 3 aggregate interface view:

  lldp agent { **nearest-customer** | **nearest-nontpmr** } **check-change-interval** *interval*

- In IRF physical interface view:

  lldp check-change-interval *interval*

By default, LLDP polling is disabled.

Disabling LLDP PVID inconsistency check

About this task

By default, when the system receives an LLDP packet, it compares the PVID value contained in the packet with the PVID configured on the receiving interface. If the two PVIDs do not match, a log message will be printed to notify the user.

You can disable PVID inconsistency check if different PVIDs are required on a link.

Procedure

46. Enter system view.

system-view

47. Disable LLDP PVID inconsistency check.

lldp ignore-pvid-inconsistency

By default, LLDP PVID inconsistency check is enabled.

Configuring CDP compatibility

About this task

To enable your device to exchange information with a directly connected Cisco device that supports only CDP, you must enable CDP compatibility.

CDP compatibility enables your device to receive and recognize CDP packets from the neighboring CDP device and send CDP packets to the neighboring device. The CDP packets sent to the neighboring CDP device carry the following information:

- Device ID.
- ID of the port connecting to the neighboring device.
- Port IP address.
- TTL.

The port IP address is the primary IP address of a VLAN interface in up state. The VLAN ID of the VLAN interface must be the lowest among the VLANs permitted on the port. If no VLAN interfaces of the permitted VLANs are assigned an IP address or all VLAN interfaces are down, no port IP address will be advertised.

You can view the neighboring CDP device information that can be recognized by the device in the output of the **display lldp neighbor-information** command. For more information about the **display lldp neighbor-information** command, see LLDP commands in *Layer 2—LAN Switching Command Reference*.

To make your device work with Cisco IP phones, you must enable CDP compatibility.

If your LLDP-enabled device cannot recognize CDP packets, it does not respond to the requests of Cisco IP phones for the voice VLAN ID configured on the device. As a result, a requesting Cisco IP phone sends voice traffic without any tag to your device. Your device cannot differentiate the voice traffic from other types of traffic.

CDP compatibility enables your device to receive and recognize CDP packets from a Cisco IP phone and respond with CDP packets carrying TLVs with the configured voice VLAN. If no voice VLAN is configured for CDP packets, CDP packets carry the voice VLAN of the port or the voice VLAN assigned by the RADIUS server. The assigned voice VLAN has a higher priority. According to TLVs with the voice VLAN configuration, the IP phone automatically configures the voice VLAN. As a result, the voice traffic is confined in the configured voice VLAN and is differentiated from other types of traffic.

For more information about voice VLANs, see "Configuring voice VLANs."

When the device is connected to a Cisco IP phone that has a host attached to its data port, the host must access the network through the Cisco IP phone. If the data port goes down, the IP phone will send a CDP packet to the device so the device can log out the user.

CDP-compatible LLDP operates in one of the following modes:

- **TxRx**—CDP packets can be transmitted and received.
- **Rx**—CDP packets can be received but cannot be transmitted.
- **Disable**—CDP packets cannot be transmitted or received.

Restrictions and guidelines

When you configure CDP compatibility for LLDP, follow these restrictions and guidelines:

- To make CDP-compatible LLDP take effect on a port, follow these steps:
  a. Enable CDP-compatible LLDP globally.
  b. Configure CDP-compatible LLDP to operate in TxRx mode on the port.
- The maximum TTL value that CDP allows is 255 seconds. To make CDP-compatible LLDP work correctly with Cisco IP phones, configure the LLDP frame transmission interval to be no more than 1/3 of the TTL value.

Prerequisites

Before you configure CDP compatibility, complete the following tasks:

- Globally enable LLDP.
- Enable LLDP on the port connecting to a CDP device.
- Configure LLDP to operate in TxRx mode on the port.

Procedure

48. Enter system view.
    system-view
49. Enable CDP compatibility globally.
    lldp compliance cdp
    By default, CDP compatibility is disabled globally.
50. Enter Layer 2/Layer 3 Ethernet interface view or management Ethernet interface view.
    interface *interface-type interface-number*
51. Configure CDP-compatible LLDP to operate in TxRx mode.
    lldp compliance admin-status cdp txrx
    By default, CDP-compatible LLDP operates in **disable** mode.
52. Set the voice VLAN ID carried in CDP packets.
    cdp voice-vlan *vlan-id*
    By default, no voice VLAN ID is configured to be carried in CDP packets.

Configuring LLDP trapping and LLDP-MED trapping

About this task

LLDP trapping or LLDP-MED trapping notifies the network management system of events such as newly detected neighboring devices and link failures.

Before you set a trap transmission interval for LLDP, make sure you understand the following influences of the LLDP trap and LLDP-MED trap transmission interval:

- If the interval value is large, the trap sending frequency is low, which occupies fewer system resources. This is suitable for scenarios where the network topology on the device changes frequently. However, do not set the interval to a too large value. If the interval value is too large, the network management system cannot identify neighbor changes in time, which will cause the network topology on the device not to update in time.
- If the interval value is small, the trap sending frequency is high. This is suitable for topology-sensitive scenarios. However, do not set the interval to a too small value. If the interval value is too small, the network management system will frequently refresh the neighbor status information. This will not only occupies too many system resources but also cause network topology vibration on the device.

Restrictions and guidelines

In the early stage of networking, the network topology changes frequently. If LLDP trapping or LLDP-MED trapping is enabled, the device will frequently send trap information to the network management system, which increases the system burden and wastes resources. As a best practice, do not enable LLDP trapping or LLDP-MED trapping in the early stage of networking.

Procedure

53. Enter system view.
    system-view
54. Enter interface view.
    interface *interface-type interface-number*
55. Enable LLDP trapping.
    - In Layer 2/Layer 3 Ethernet interface view or management Ethernet interface view:

lldp [**agent** { **nearest-customer** | **nearest-nontpmr** }] **notification remote-change enable**

- In Layer 2/Layer 3 aggregate interface view:

  lldp agent { **nearest-customer** | **nearest-nontpmr** } **notification remote-change enable**

- In IRF physical interface view:

  lldp notification remote-change enable

By default, LLDP trapping is disabled.

56. (In Layer 2/Layer 3 Ethernet interface view or management Ethernet interface view.) Enable LLDP-MED trapping.

lldp notification med-topology-change enable

By default, LLDP-MED trapping is disabled.

57. Return to system view.

quit

58. (Optional.) Set the LLDP trap transmission interval.

lldp timer notification-interval *interval*

The default setting is 30 seconds.

Configuring MAC address learning for DCN

About MAC address learning for DCN

For the data communication network (DCN) to implement operation, administration, and maintenance on network elements (NEs), you must configure the NEs to learn their neighbors' MAC addresses through LLDP.

An NE learns the MAC address of its neighbor through the ARP or ND entry generated based on the management address TLV in LLDP frames received from the neighbor. For more information about DCN, see *Layer 3—IP Routing Configuration Guide*.

Setting the source MAC address of LLDP frames

About this task

This feature must be configured with generation of ARP or ND entries for received management address TLVs to meet the following requirements:

- The source MAC address of outgoing LLDP frames is the MAC address of a VLAN interface instead of the MAC address of the egress interface.
- The neighbor device can generate correct ARP or ND entries for the local device.

In Layer 2 Ethernet interface view or Layer 2 aggregate interface view, this feature sets the source MAC address of outgoing LLDP frames to the MAC address of a VLAN interface to which the specified VLAN ID belongs. The source MAC address of outgoing LLDP frames is the MAC address of the Layer 2 interface in the following situations:

- The specified VLAN or the corresponding VLAN interface does not exist.
- The VLAN interface to which the VLAN ID belongs is physically down.

In Layer 3 Ethernet interface view, the MAC address of the Layer 3 Ethernet interface is always used as the source MAC address of outgoing LLDP frames.

Procedure

59. Enter system view.

    system-view

60. Enter Layer 2 Ethernet interface view, Layer 2 aggregate interface view, or Layer 3 Ethernet interface view.

    interface *interface-type interface-number*

61. Set the source MAC address of LLDP frames to the MAC address of a VLAN interface or the Layer 3 Ethernet interface.

    lldp source-mac vlan *vlan-id*

    By default, the source MAC address of LLDP frames is the MAC address of the egress interface.

Enabling generation of ARP or ND entries for received management address TLVs

About this task

This feature enables the device to generate an ARP or ND entry after receiving an LLDP frame containing a management address TLV on an interface. The ARP or ND entry maps the advertised management address to the source MAC address of the frame.

You can enable generation of both ARP and ND entries on an interface. If the management address TLV contains an IPv4 address, the device generates an ARP entry. If the management address TLV contains an IPv6 address, the device generates an ND entry.

In Layer 2 Ethernet interface view or Layer 2 aggregate interface view, this feature sets the Layer 2 interface to the output interface in the generated entries. The VLAN to which the entries belong is the VLAN specified by this feature. The device cannot generate ARP or ND entries in one of the following situations:

- The specified VLAN or the corresponding VLAN interface does not exist.
- The VLAN interface to which the VLAN ID belongs is physically down.

In Layer 3 Ethernet interface view, the Layer 3 Ethernet interface is always recorded as the output interface, no matter whether you specify a VLAN ID.

Restrictions and guidelines

This feature must be configured with the feature of setting the source MAC address of LLDP frames to meet the following requirements:

- The device uses the MAC address of a VLAN interface instead of the egress interface's MAC address as the source MAC address of LLDP frames.
- The neighbor device can generate correct ARP or ND entries.

Procedure

62. Enter system view.

    system-view

63. Enter Layer 2 Ethernet interface view, Layer 2 aggregate interface view, or Layer 3 Ethernet interface view.

    interface *interface-type interface-number*

64. Enable generation of ARP or ND entries for management address TLVs received on the interface.

    - In Layer 2 Ethernet interface view or Layer 2 aggregate interface view:

      lldp management-address { **arp-learning** | **nd-learning** } **vlan** *vlan-id*

- In Layer 3 Ethernet interface view:

  lldp management-address { **arp-learning** | **nd-learning** } [**vlan** *vlan-id*]

By default, generation of ARP or ND entries for received management address TLVs is disabled on an interface.

In Layer 2 Ethernet interface view, the **vlan** *vlan-id* option specifies the ID of the VLAN to which the generated ARP or ND entry belongs. To prevent the ARP or ND entries from overwriting each other, do not specify the same VLAN ID for different Layer 2 Ethernet interfaces.

You can enable generation of both ARP and ND entries on an interface.

Enabling transparent transmission of LLDP frames

About this task

On an MPLS L2VPN network, you can enable transparent LLDP frame transmission on the PE interfaces connected to CEs so the CEs can establish neighbor relationships across the PEs.

An interface configured with this feature can transmit LLDP frames transparently, no matter whether or not LLDP is enabled on it.

Restrictions and guidelines

Before you enable transparent LLDP frame transmission on a PE interface connected to a CE, disable LLDP on the interface to prevent the CE from establishing neighbor relationships with the PE and the resulting topology errors.

Procedure

65. Enter system view.

system-view

66. Enter interface view.

interface *interface-type interface-number*

67. Enable transparent LLDP frame transmission on the interface.

lldp transparent enable

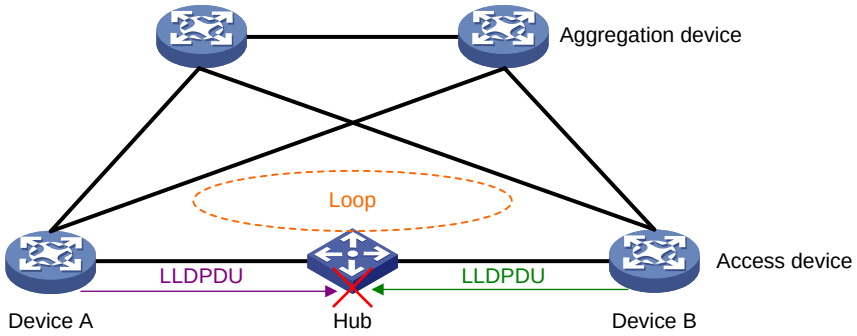
By default, transparent LLDP frame transmission is disabled on an interface.

Configuring LLDP black hole detection

About this task

As shown in Figure 5, Device A and Device B are connected through a hub, causing a loop in the network. However, since LLDP packets are point-to-point packets, they are terminated at the next node after being sent from the device. Therefore, the hub acts as a black hole for LLDP packets and LLDP packets cannot be transmitted between Device A and Device B through the HUB. As a result, Device A and Device B cannot use LLDP to learn about the physical link between them and the formed loop, and therefore the loop cannot be eliminated in time. In order to eliminate potential loop risks, it required to detect LLDP packet black holes and block the links leading to the black holes.

Figure 5 LLDP black hole detection application scenario



The LLDP black hole detection feature can detect the presence of LLDP black holes in links on ports. The operating mechanism is as follows:

68. Triggering the sending of black hole detection packets

The ports enabled with LLDP black hole detection will carry black hole detection TLVs in the transmitted LLDP packets. The black hole detection TLV is a type of TLV defined by H3C. If an LLDP packet carrying the black hole detection TLV is received, it indicates that there is no LLDP black hole between the device and its neighbor, and LLDP packets can be sent and received successfully. Only H3C devices can send and recognize the black hole detection TLV carried in LLDP packets.

After LLDP black hole detection is enabled on a port, the port will activate the LLDP packet receiving timer for black hole detection. If no LLDP packets are received on the port until the timer expires, it indicates a potential LLDP packet black hole in the network, causing the device to not receive LLDP packets. The port then starts sending LLDP black hole detection packets continuously. If the port receives an LLDP packet carrying the black hole detection TLV before the timer expires, the port terminates the timer and takes no further action.

After LLDP black hole detection is enabled, if a port receives an LLDP packet without the black hole detection TLV, the port will start sending LLDP black hole detection packets continuously.

LLDP black hole detection packet is an H3C proprietary Layer 2 packet type. Unlike LLDP packets, LLDP black hole detection packets can be transparently transmitted by a hub and are continuously sent at fixed intervals after being triggered.

The timeout for receiving LLDP packets is configured through the **`lldp global blackhole-detection rx-lldpdu timeout`** command. The interval for sending LLDP black hole detection packets is configured through the **`lldp timer blackhole-detection-interval`** command or the **`lldp global timer blackhole-detection-interval`** command.

69. Black hole confirmation

After a port starts to send LLDP black hole detection packets, if it receives an LLDP packet carrying the black hole detection TLV, it determines that no LLDP black hole exists between the port and its neighbor, stops sending LLDP black hole detection packets, and resets the LLDP packet receiving timer.

If a port enabled with LLDP black hole detection receives an LLDP black hole detection packet and the LLDP packet receiving timer on this port has timed out, it indicates that the local device and the peer device are connected with a link and both devices have the ability to send LLDP packets. However, neither of the devices has received LLDP packets, and therefore the local device concludes that there is an LLDP black hole between the two ends of the link.

70. Black hole processing

After detecting an LLDP black hole, the device blocks the port receiving LLDP black hole detection packets to eliminate network loops and prevent the transmission of user data packets. If the blocked port receives no LLDP black hole detection packets within the timeout period specified by the **`lldp timer rx-blackhole-timeout`** or **`lldp global timer`**

rx-blackhole-timeout command, it means that the LLDP black hole is eliminated and the port resumes normal forwarding. Otherwise, it remains blocked.

The **lldp timer blackhole-detection-interval** and **lldp global timer blackhole-detection-interval** commands affect the interval at which the specified ports send LLDP black hole detection packets. Set a suitable packet sending interval based on actual requirements.

- When the sending interval is short, the port sends LLDP black hole detection packets frequently. The advantage is that the network detects and responds to LLDP black holes more sensitively. The disadvantage is that it occupies more device CPU resources. Configure a short sending interval in scenarios where the device performance is high or there is a strong demand for loop prevention.
- When the sending interval is long, the port sends LLDP black hole detection packets slowly. The advantage is that it occupies less CPU resources of the device; while the disadvantage is that the device cannot detect and respond to LLDP black holes quickly. Configure longer sending intervals in scenarios where device performance is weaker or loop tolerance is higher.

Restrictions and guidelines

The LLDP black hole detection feature takes effect only if the following conditions are met:

- On both ends of the LLDP connection, both global and port-level LLDP black hole detection functions are enabled.
- On the ports with LLDP black hole detection enabled, the link status is UP, LLDP is enabled, and the operating mode is TxRx.

If LLDP black hole detection is not enabled on a port or the LLDP packet receiving timer of the port has not timed out, the port will not process any LLDP black hole detection packets received.

On devices with a long LLDP packet sending interval, configure a longer LLDP packet receiving timeout. Otherwise, it might result in device ports being blocked even if no LLDP black hole exists in the network. Configure a shorter LLDP receiving timeout on devices with a shorter LLDP packet sending interval for more efficient LLDP black hole detection.

The **lldp timer blackhole-detection-interval** and **lldp global timer blackhole-detection-interval** commands can both set the interval at which ports send LLDP black hole detection packets. When both commands are configured, the **lldp timer blackhole-detection-interval** command takes precedence. If the interval is not set on a port, the port uses the global interval set by the **lldp global timer blackhole-detection-interval** command.

The timeout for receiving LLDP black hole detection packets on a port can be configured by using the **lldp timer rx-blackhole-timeout** and **lldp global timer rx-blackhole-timeout** commands. When both commands are used, the **lldp timer rx-blackhole-timeout** command takes precedence. If a port is not configured with a timeout, it uses the global timeout set by the **lldp global timer rx-blackhole-timeout** command.

The timeout for receiving LLDP black hole detection packets on the local device should not be less than the interval for sending LLDP black hole detection packets on the peer device. Otherwise, LLDP black hole detection will not take effect.

As a best practice, do not enable LLDP black hole detection on a member port of an aggregate interface. If you fail to do so, LLDP black hole detection might fail to take effect.

Procedure

71. Enter system view.

```
system-view
```

72. Enable global LLDP black hole detection.

```
lldp global blackhole-detection enable
```

By default, global LLDP black hole detection is disabled.

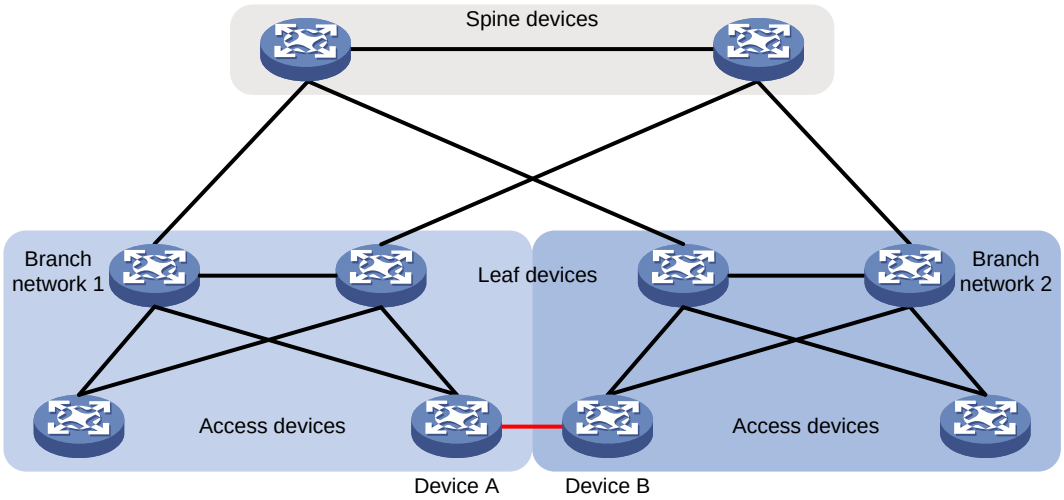
73. (Optional.) Configure the LLDP packet receiving timeout for black hole detection.
    lldp global blackhole-detection rx-lldpdu timeout *timeout*
    By default, the LLDP packet receiving timeout for black hole detection is 120 seconds.
74. (Optional.) Set the global interval for sending LLDP black hole detection packets.
    lldp global timer blackhole-detection-interval *interval*
    By default, the interval for sending LLDP black hole detection packets is 2 seconds.
75. (Optional.) Set the global timeout for receiving LLDP black hole detection packets.
    lldp global timer rx-blackhole-timeout *timeout*
    By default, the timeout for receiving LLDP black hole detection packets is (the global interval for sending LLDP black hole detection packets × 3 + 10) seconds.
76. (Optional.) Enter Layer 2 Ethernet interface view.
    interface *interface-type interface-number*
77. (Optional.) Enable LLDP black hole detection for a port.
    lldp blackhole-detection enable
    By default, LLDP black hole detection is enabled for a port.
78. (Optional.) Set the interval at which the port sends LLDP black hole detection packets.
    lldp timer blackhole-detection-interval *interval*
    By default, the interval for sending LLDP black hole detection packets is 2 seconds.
79. (Optional.) Set the timeout for receiving LLDP black hole detection packets on the port.
    lldp timer rx-blackhole-timeout *timeout*
    By default, the timeout for receiving LLDP black hole detection packets on a port is (the port's sending interval for LLDP black hole detection packets × 3 + 10) seconds.

Configuring LLDP cross-domain detection

About this task

As shown in Figure 6, Device A and Device B belongs to different branch networks. If a link exists between Device A and Device B, a loop might occur in the network. After configuring LLDP cross-domain detection, you can manually assign the devices to different domains, and use LLDP to detect whether the specified LLDP neighbor is in the local domain. In addition, you can block the link to neighbors in other domains to eliminate the potential loop risk.

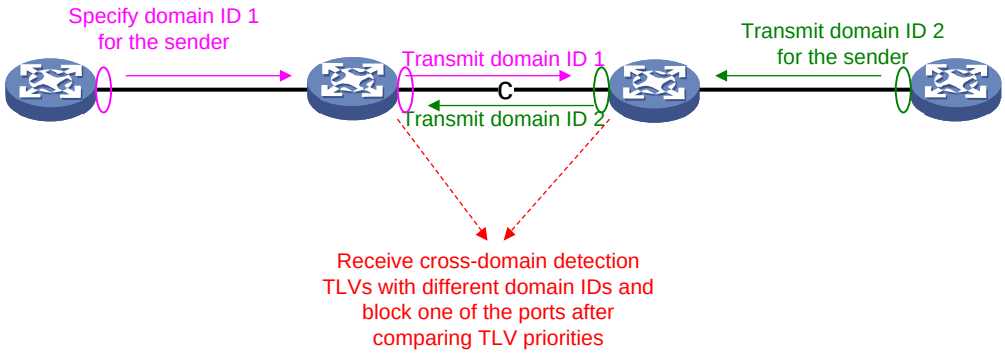
Figure 6 LLDP cross-domain detection application scenario



The LLDP cross-domain detection feature defines a private type of TLV for LLDP, that is, the cross-domain detection TLV. This TLV is used to advertise the local domain to the LLDP neighbor at the peer end. The LLDP cross-domain detection feature takes effect on a port only after you execute the **lldp global cross-domain-detection enable** or **lldp cross-domain-detection** command to enable LLDP cross-domain detection. Then, LLDP packets sent by the device will carry cross-domain detection TLVs.

The domain ID in a cross-domain detection TLV is key information, because it indicates the domain to which the port that sends the cross-domain detection TLV belongs. If you execute the **lldp cross-domain-detection domain-id** command to specify a domain ID for a port, the cross-domain detection TLVs sent by this port must carry the specified domain ID. If you do not specify a domain ID for a port, the cross-domain detection TLVs sent by this port carry a domain ID in the cross-domain detection TLVs received by the device. If the device has received multiple cross-domain detection TLVs with different domain IDs, it will compare the priorities of these TLVs. When sending a cross-domain detection TLV, the device will carry only the domain ID of the optimal TLV it received, and block ports that receive non-optimal TLVs with different domain IDs. This will help eliminating potential loops in the network, as shown in Figure 7. A blocked port can restore normal forwarding status when LLDP cross-domain detection is disabled or it does not receive any cross-domain detection TLV that carries a different domain ID.

Figure 7 Domain ID transmission



Restrictions and guidelines

As a best practice, enable LLDP cross-domain detection only on a port of the start device of a domain ID (typically the spine device in a spine-leaf architecture), and do not enable LLDP cross-domain detection globally. If you fail to do so, LLDP cross-domain detection might fail to take effect.

If the domain ID is not configured for LLDP cross-domain detection, or the port is not enabled with the cross-domain detection TLV advertisement capability, the cross-domain detection feature cannot take effect on the port. In this case, the device cannot identify cross-domain detection TLVs carried in LLDP packets or add cross-domain detection TLVs to LLDP packets.

You can use the **lldp global cross-domain-detection enable** or **lldp cross-domain-detection** command to enable LLDP cross-domain detection for a port. If the **lldp cross-domain-detection** command is not used for the port, the global setting (configured with the **lldp global cross-domain-detection enable** command) applies. If both commands are configured for the port, the port-specific setting (configured with the **lldp cross-domain-detection** command) applies.

Procedure

80. Enter system view.

```
system-view
```

81. Enable LLDP cross-domain detection globally.

```
lldp global cross-domain-detection enable
```

By default, LLDP cross-domain detection is disabled globally.

82. Enter Layer 2 Ethernet interface view.
    interface *interface-type interface-number*
83. Configure the domain ID for LLDP cross-domain detection.
    lldp cross-domain-detection domain-id *domain-id*
    By default, the domain ID is not configured for LLDP cross-domain detection.
84. Enable or disable LLDP cross-domain detection for the port.
    lldp cross-domain-detection { **enable** | **disable** }
    By default, the global setting (configured with the **lldp global cross-domain-detection enable** command) applies.

Disabling AP restart through LLDP

About this task

When the device acts as Power Sourcing Equipment (PSE), it can use the LLDP aging mechanism to determine the status of the directly connected AP it powers. If the AP fails and cannot send LLDP packets, the device calls the PoE module to restart the interface, thus achieving the function of rebooting the connected AP. This function is known as AP restart through LLDP.

AP restart through LLDP identifies abnormal APs as follows:

- When the device receives an LLDP packet from a neighbor, if the System Capabilities TLV contains the WlanAccessPoint field, the neighbor is considered to be an AP neighbor.
- The device determines if the directly connected AP neighbor is abnormal by aging out the LLDP neighbor. The device considers the AP neighbor as abnormal only when the LLDP packet reception times out and the physical status of the interface directly connected to the AP neighbor is UP. If the neighbor ages out and the physical status of the interface directly connected to the AP neighbor is DOWN, the device will not trigger the restart mechanism but will directly delete the LLDP neighbor.
- The AP restart through LLDP mechanism can only be triggered when only one AP neighbor is connected to the interface.

After confirming that the AP is abnormal, the device invokes the PoE module to restart the interface, causing the AP connected to that interface to be rebooted through power-off and re-powering.

By default, the AP restart through LLDP feature is enabled. To avoid accidental restarts of the AP, you can disable this feature.

Restrictions and guidelines

The device configured with this feature must be a PSE that supports PoE, and the AP neighbor that needs to be restarted must be a Powered Device (PD). The interface that is directly connected to the AP neighbor must be a PoE interface with the PoE feature enabled. For more information about PSE, PD, and PoE interfaces, see *Network Management and Monitoring Configuration Guide*.

Procedure

85. Enter system view.
    system-view
86. Disable AP restart through LLDP.
    undo lldp ap-timeout-reset enable
    By default, the AP restart through LLDP feature is enabled.

Disabling LLDP control over PoE

About this task

In scenarios where a switch is connected to a large number of APs and provides PoE power to these APs, LLDP neighbors must be disconnected during batch maintenance of APs. At such times, if LLDP still maintains control over PoE, this might cause power disruption or reboot of APs, resulting in maintenance failure. In this scenario, it is necessary for LLDP not to control the switch's PoE power supply to ensure the smooth maintenance of AP devices.

After you disable LLDP control over PoE, LLDP no longer participates in the PoE power provisioning process and does not handle the Power Via MDI TLV defined by IEEE 802.3. This achieves the decoupling of LLDP functionality and PoE capability.

Procedure

87. Enter system view.

system-view

88. Disabling LLDP control over PoE.

undo lldp process-poe enable

By default, LLDP control over PoE is enabled.

Display and maintenance commands for LLDP

Execute **display** commands in any view.

Task	Command
Display local LLDP information.	**display lldp local-information** [**global** | **interface** *interface-type interface-number*]
Display the information contained in the LLDP TLVs sent from neighboring devices.	**display lldp neighbor-information** [[[**interface** *interface-type interface-number*] [**agent** { **nearest-bridge** | **nearest-customer** | **nearest-nontpmr** }] [**verbose**]] | **list** [**system-name** *system-name*]]
Display LLDP statistics.	**display lldp statistics** [**global** | [**interface** *interface-type interface-number*] [**agent** { **nearest-bridge** | **nearest-customer** | **nearest-nontpmr** }]]
Display LLDP status of a port.	**display lldp status** [**interface** *interface-type interface-number*] [**agent** { **nearest-bridge** | **nearest-customer** | **nearest-nontpmr** }]
Display types of advertisable optional LLDP TLVs.	**display lldp tlv-config** [**interface** *interface-type interface-number*] [**agent** { **nearest-bridge** | **nearest-customer** | **nearest-nontpmr** }]
Clear LLDP statistics on ports.	**reset lldp statistics** [**interface** *interface-type interface number*] [**agent** { **nearest-bridge** | **nearest-customer** | **nearest-nontpmr** }]

LLDP configuration examples

Example: Configuring basic LLDP functions

Network configuration

As shown in Figure 8, enable LLDP globally on Switch A and Switch B to perform the following tasks:

- Monitor the link between Switch A and Switch B on the NMS.
- Monitor the link between Switch A and the MED device on the NMS.

Figure 8 Network diagram

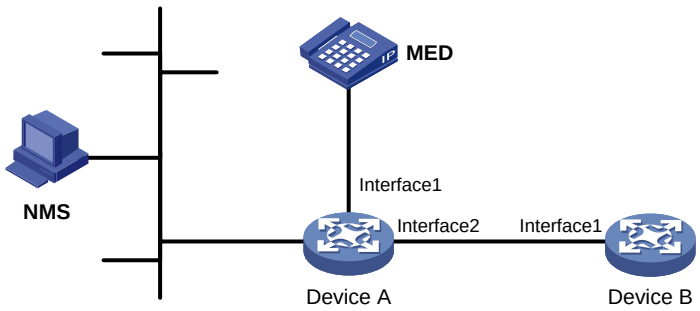


Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

89. Configure Switch A:

Enable LLDP globally.

```
<SwitchA> system-view
[SwitchA] lldp global enable
```

Enable LLDP on GigabitEthernet 1/0/1. By default, LLDP is enabled on ports.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] lldp enable
```

Set the LLDP operating mode to Rx on GigabitEthernet 1/0/1.

```
[SwitchA-GigabitEthernet1/0/1] lldp admin-status rx
[SwitchA-GigabitEthernet1/0/1] quit
```

Enable LLDP on GigabitEthernet 1/0/2. By default, LLDP is enabled on ports.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] lldp enable
```

Set the LLDP operating mode to Rx on GigabitEthernet 1/0/2.

```
[SwitchA-GigabitEthernet1/0/2] lldp admin-status rx
[SwitchA-GigabitEthernet1/0/2] quit
```

90. Configure Switch B:

Enable LLDP globally.

```
<SwitchB> system-view
```

```
[SwitchB] lldp global enable
```

Enable LLDP on GigabitEthernet 1/0/1. By default, LLDP is enabled on ports.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] lldp enable
```

Set the LLDP operating mode to Tx on GigabitEthernet 1/0/1.

```
[SwitchB-GigabitEthernet1/0/1] lldp admin-status tx
[SwitchB-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Verify the following items:

- GigabitEthernet 1/0/1 of Switch A connects to a MED device.
- GigabitEthernet 1/0/2 of Switch A connects to a non-MED device.
- Both ports operate in Rx mode, and they can receive LLDP frames but cannot send LLDP frames.

```
[SwitchA] display lldp status
Global status of LLDP: Enable
Bridge mode of LLDP: customer-bridge
The current number of LLDP neighbors: 2
The current number of CDP neighbors: 0
LLDP neighbor information last changed time: 0 days, 0 hours, 4 minutes, 40 seconds
Transmit interval              : 30s
Fast transmit interval         : 1s
Transmit credit max            : 5
Hold multiplier                : 4
Reinit delay                   : 2s
Trap interval                  : 30s
Fast start times               : 4

LLDP status information of port 1 [GigabitEthernet1/0/1]:
LLDP agent nearest-bridge:
Port status of LLDP            : Enable
Admin status                   : Rx_Only
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 1
Number of MED neighbors        : 1
Number of CDP neighbors        : 0
Number of sent optional TLV    : 21
Number of received unknown TLV : 0

LLDP agent nearest-customer:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
```

```
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 16
Number of received unknown TLV : 0

LLDP status information of port 2 [GigabitEthernet1/0/2]:
LLDP agent nearest-bridge:
Port status of LLDP            : Enable
Admin status                   : Rx_Only
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 1
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 21
Number of received unknown TLV : 3

LLDP agent nearest-nontpmr:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 1
Number of received unknown TLV : 0

LLDP agent nearest-customer:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 16
Number of received unknown TLV : 0
```

Remove the link between Switch A and Switch B.

Verify that GigabitEthernet 1/0/2 of Switch A does not connect to any neighboring devices.

```
[SwitchA] display lldp status
Global status of LLDP: Enable
The current number of LLDP neighbors: 1
The current number of CDP neighbors: 0
LLDP neighbor information last changed time: 0 days, 0 hours, 5 minutes, 20 seconds
```

```
Transmit interval              : 30s
Fast transmit interval         : 1s
Transmit credit max            : 5
Hold multiplier                : 4
Reinit delay                   : 2s
Trap interval                  : 30s
Fast start times               : 4

LLDP status information of port 1 [GigabitEthernet1/0/1]:
LLDP agent nearest-bridge:
Port status of LLDP            : Enable
Admin status                   : Rx_Only
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 1
Number of MED neighbors        : 1
Number of CDP neighbors        : 0
Number of sent optional TLV    : 0
Number of received unknown TLV : 5

LLDP agent nearest-nontpmr:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 1
Number of received unknown TLV : 0

LLDP status information of port 2 [GigabitEthernet1/0/2]:
LLDP agent nearest-bridge:
Port status of LLDP            : Enable
Admin status                   : Rx_Only
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 0
Number of received unknown TLV : 0

LLDP agent nearest-nontpmr:
Port status of LLDP            : Enable
```

```
Admin status                  : Disable
Trap flag                     : No
MED trap flag                 : No
Polling interval              : 0s
Number of LLDP neighbors      : 0
Number of MED neighbors       : 0
Number of CDP neighbors       : 0
Number of sent optional TLV   : 1
Number of received unknown TLV : 0

LLDP agent nearest-customer:
Port status of LLDP           : Enable
Admin status                  : Disable
Trap flag                     : No
MED trap flag                 : No
Polling interval              : 0s
Number of LLDP neighbors      : 0
Number of MED neighbors       : 0
Number of CDP neighbors       : 0
Number of sent optional TLV   : 16
Number of received unknown TLV : 0
```

Example: Configuring LLDP cross-domain detection

Network configuration

As shown in Figure 9, in a spine-leaf architecture, Device A, Device C, Device D, and Device G belong to the same branch, and Device B, Device E, Device F, and Device H belong to the other. To prevent loops caused by incorrect wiring between devices, make sure only the devices within the same branch can exchange data. To meet these requirements, configure LLDP cross-domain detection. Assign Device A and its downstream devices to one domain and Device B and its downstream devices to another. Then, the ports that connect different domains will be blocked to prevent data traffic between different branches.

Figure 9 Network diagram

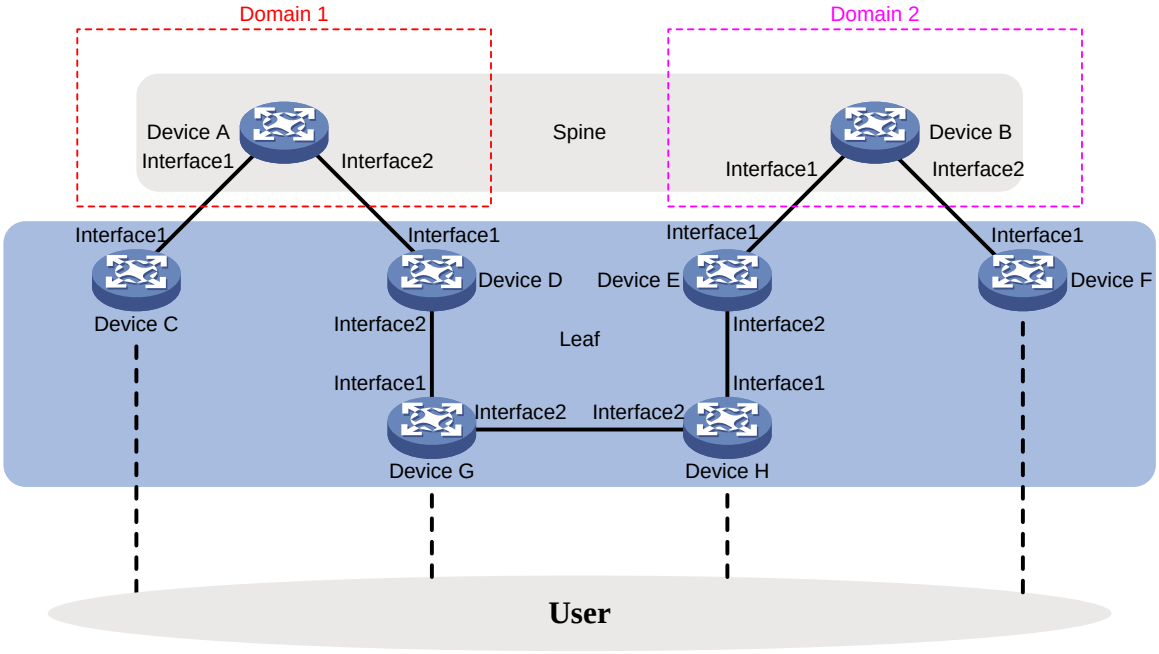


Table 9 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

91. Configure Device A:

Enable LLDP globally.

```
<DeviceA> system-view
[DeviceA] lldp global enable
```

Enable LLDP on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 separately. By default, LLDP is enabled on ports.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] lldp enable
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] lldp enable
[DeviceA-GigabitEthernet1/0/2] quit
```

Specify domain ID 1 for GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 and enable LLDP cross-domain detection.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] lldp cross-domain-detection domain-id 1
[DeviceA-GigabitEthernet1/0/1] lldp cross-domain-detection enable
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] lldp cross-domain-detection domain-id 1
[DeviceA-GigabitEthernet1/0/2] lldp cross-domain-detection enable
```

```
[DeviceA-GigabitEthernet1/0/2] quit
```

92. Configure Device B:

Enable LLDP globally.

```
<DeviceB> system-view
[DeviceB] lldp global enable
```

Enable LLDP on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 separately. By default, LLDP is enabled on ports.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] lldp enable
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] lldp enable
[DeviceB-GigabitEthernet1/0/2] quit
```

Specify domain ID 2 for GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 and enable LLDP cross-domain detection.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] lldp cross-domain-detection domain-id 2
[DeviceB-GigabitEthernet1/0/1] lldp cross-domain-detection enable
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] lldp cross-domain-detection domain-id 2
[DeviceB-GigabitEthernet1/0/2] lldp cross-domain-detection enable
[DeviceB-GigabitEthernet1/0/2] quit
```

93. Configure Device C:

Enable LLDP globally.

```
<DeviceC> system-view
[DeviceC] lldp global enable
```

Enable LLDP cross-domain detection globally.

```
[DeviceC] lldp global cross-domain-detection enable
```

Enable LLDP on GigabitEthernet 1/0/1. By default, LLDP is enabled on ports.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] lldp enable
[DeviceC-GigabitEthernet1/0/1] quit
```

94. Configure Device D

Enable LLDP globally.

```
<DeviceD> system-view
[DeviceD] lldp global enable
```

Enable LLDP cross-domain detection globally.

```
[DeviceD] lldp global cross-domain-detection enable
```

Enable LLDP on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2. By default, LLDP is enabled on ports.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] lldp enable
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] lldp enable
[DeviceD-GigabitEthernet1/0/2] quit
```

95. Configure Device G in the same steps as Device D.

96. Configure Device E:

Enable LLDP globally.

```
<DeviceE> system-view
[DeviceE] lldp global enable
```

Enable LLDP cross-domain detection globally.

```
[DeviceE] lldp global cross-domain-detection enable
```

Enable LLDP on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2. By default, LLDP is enabled on ports.

```
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] lldp enable
[DeviceE-GigabitEthernet1/0/1] quit
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] lldp enable
[DeviceE-GigabitEthernet1/0/2] quit
```

97. Configure Device H in the same steps as Device E.

98. Configure F:

Enable LLDP globally.

```
<DeviceF> system-view
[DeviceF] lldp global enable
```

Enable LLDP cross-domain detection globally.

```
[DeviceF] lldp global cross-domain-detection enable
```

Enable LLDP on GigabitEthernet 1/0/1. By default, LLDP is enabled on ports.

```
[DeviceF] interface gigabitethernet 1/0/1
[DeviceF-GigabitEthernet1/0/1] lldp enable
[DeviceF-GigabitEthernet1/0/1] quit
```

Verifying the configuration

On Device H, obtain the global LLDP status and LLDP status on all ports.

```
[DeviceG] display lldp status
Global status of LLDP: Enable
Bridge mode of LLDP: customer-bridge
The current number of LLDP neighbors: 2
The current number of CDP neighbors: 0
LLDP neighbor information last changed time: 0 days, 0 hours, 2 minutes, 28 seconds
Transmit interval              : 30s
Fast transmit interval         : 1s
Transmit max credit            : 5
Hold multiplier                : 4
Reinit delay                   : 2s
Trap interval                  : 30s
Fast start times               : 4

LLDP status information of port 2 [GigabitEthernet1/0/1]:
LLDP agent nearest-bridge:
Port status of LLDP            : Enable
Admin status                   : TX_RX
Trap flag                      : No
MED trap flag                  : No
```

```
Polling interval               : 0s
Number of LLDP neighbors       : 1
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 12
Number of received unknown TLV : 16
Neighbor protection status     : Port not protected

LLDP agent nearest-nontpmr:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 0
Number of received unknown TLV : 0
Neighbor protection status     : Port not protected

LLDP agent nearest-customer:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 0
Number of received unknown TLV : 0
Neighbor protection status     : Port not protected

LLDP status information of port 3 [GigabitEthernet1/0/2]:
LLDP agent nearest-bridge:
Port status of LLDP            : Enable
Admin status                   : TX_RX
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 1
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 12
Number of received unknown TLV : 17
Neighbor protection status     : Port blocked (different domain)
```

```
LLDP agent nearest-nontpmr:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 0
Number of received unknown TLV : 0

LLDP agent nearest-customer:
Port status of LLDP            : Enable
Admin status                   : Disable
Trap flag                      : No
MED trap flag                  : No
Polling interval               : 0s
Number of LLDP neighbors       : 0
Number of MED neighbors        : 0
Number of CDP neighbors        : 0
Number of sent optional TLV    : 0
Number of received unknown TLV : 0
```

Interface GigabitEthernet 1/0/2 on Device H is blocked by cross-domain detection.

Configuring DCBX

About DCBX

Data Center Ethernet (DCE), also known as Converged Enhanced Ethernet (CEE), is enhancement and expansion of traditional Ethernet local area networks for use in data centers. DCE uses the Data Center Bridging Exchange Protocol (DCBX) to negotiate and remotely configure the bridge capability of network elements.

DCBX versions

DCBX has the following self-adaptable versions:

- DCB Capability Exchange Protocol Specification Rev 1.00.
- DCB Capability Exchange Protocol Base Specification Rev 1.01.
- IEEE Std 802.1Qaz-2011 (Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks-Amendment 18: Enhanced Transmission Selection for Bandwidth Sharing Between Traffic Classes).

DCBX functions

DCBX offers the following functions:

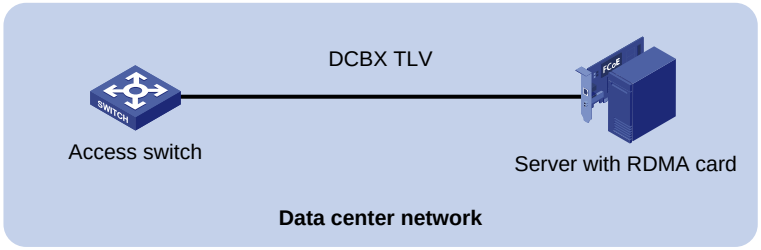
- Discovers the peer devices' capabilities and determines whether devices at both ends support these capabilities.
- Detects configuration errors on peer devices.
- Remotely configures the peer device if the peer device accepts the configuration.

NOTE:

H3C devices support only the remote configuration feature.

DCBX application scenario

Figure 10 DCBX application scenario



DCBX enables lossless packet transmission on DCE networks.

As shown in Figure 10, DCBX applies to an RoCE-based data center network, and operates on an access switch. DCBX enables the switch to control the server or storage adapter, and simplifies the configuration and guarantees configuration consistency. DCBX extends LLDP by using the IEEE 802.1 organizationally specific TLVs (DCBX TLVs) to transmit DCBX data, including:

- Priority-based Flow Control (PFC) in DCBX Rev 1.00 and DCBX Rev 1.01.

- PFC in IEEE Std 802.1Qaz-2011.

H3C devices can send these types of DCBX information to a server or storage adapter. However, H3C devices cannot accept these types of DCBX information.

Protocols and standards

- *DCB Capability Exchange Protocol Specification Rev 1.00*
- *DCB Capability Exchange Protocol Base Specification Rev 1.01*

DCBX tasks at a glance

To configure DCBX, perform the following tasks:

99. Enabling LLDP and DCBX TLV advertising

100. Setting the DCBX version

101. Configuring PFC parameters

Enabling LLDP and DCBX TLV advertising

Restrictions and guidelines

To enable the device to advertise PFC data through an interface, perform the following tasks:

- Enable LLDP globally.
- Enable LLDP and DCBX TLV advertising on the interface.

Procedure

102. Enter system view.

```
system-view
```

103. Enable LLDP globally.

```
lldp global enable
```

By default:

- If the device starts up with the initial configuration, LLDP is disabled globally.
- If the device starts up with the factory defaults, LLDP is enabled globally.

For more information about device startup with the initial configuration or factory defaults, see configuration file management in *Fundamentals Configuration Guide*.

104. Enter Layer 2 Ethernet interface view.

```
interface interface-type interface-number
```

105. Enable LLDP on the interface.

```
lldp enable
```

By default, LLDP is enabled on an interface.

106. Enable the interface to advertise DCBX TLVs.

```
lldp tlv-enable dot1-tlv dcbx
```

By default, DCBX TLV advertisement is disabled on an interface.

Setting the DCBX version

Restrictions and guidelines

When you set the DCBX version, follow these restrictions and guidelines:

- For DCBX to work correctly, configure the same DCBX version on the local port and peer port. As a best practice, configure the highest version supported on both ends. IEEE Std 802.1Qaz-2011, DCBX Rev 1.01, and DCBX Rev 1.00 are in descending order.
- After the configuration, LLDP frames sent by the local port carry information about the configured DCBX version. The local port and peer port do not negotiate the DCBX version.
- When the DCBX version is autonegotiated, the version IEEE Std 802.1Qaz-2011 is preferably negotiated.

Procedure

107. Enter system view.

system-view

108. Enter Layer 2 Ethernet interface view.

interface *interface-type interface-number*

109. Set the DCBX version.

dcbx version { **rev100** | **rev101** | **standard** }

By default, the DCBX version is not configured. It is autonegotiated by the local port and peer port.

Configuring PFC parameters

About this task

To prevent packets with an 802.1p priority value from being dropped, enable PFC for the 802.1p priority value. This feature reduces the sending rate of packets carrying this priority when network congestion occurs.

The device uses PFC parameters to negotiate with the server adapter and to enable PFC for the specified 802.1p priorities on the server adapter.

For more information about PFC commands, see Ethernet interface commands in *Layer 2—LAN Switching Command Reference.*

Procedure

110. Enter system view.

system-view

111. Enter Layer 2 Ethernet interface view.

interface *interface-type interface-number*

112. Enable PFC in auto mode on the Ethernet interface.

priority-flow-control auto

By default, PFC is disabled.

To advertise the PFC data, you must enable PFC in auto mode.

113. Enable PFC for the specified 802.1p priorities.

priority-flow-control no-drop dot1p *dot1p-list*

By default, PFC is disabled for all 802.1p priorities.

DCBX configuration examples

Example: Configuring DCBX

Network configuration

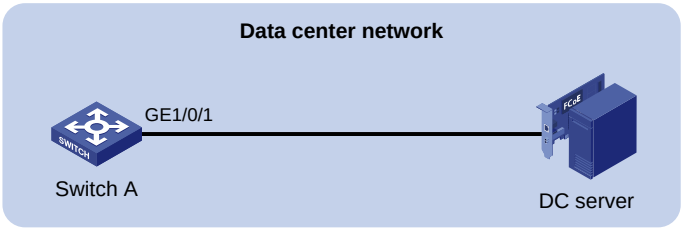
As shown in Figure 11, GigabitEthernet 1/0/1 of the access switch (Switch A) connects to the RDMA adapter of the data center server (DC server).

Configure Switch A to implement lossless RoCE frame transmission to DC server.

NOTE:

In this example, both Switch A and the DC server support DCBX Rev 1.01.

Figure 11 Network diagram



Procedure

114. Enable LLDP and DCBX TLV advertising:

Enable LLDP globally.

```
<SwitchA> system-view
[SwitchA] lldp global enable
```

Enable LLDP and DCBX TLV advertising on GigabitEthernet 1/0/1.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] lldp enable
[SwitchA-GigabitEthernet1/0/1] lldp tlv-enable dot1-tlv dcbx
```

115. Configure PFC:

Enable PFC in auto mode on GigabitEthernet 1/0/1.

```
[SwitchA-GigabitEthernet1/0/1] priority-flow-control auto
```

Enable PFC for 802.1 priority 3.

```
[SwitchA-GigabitEthernet1/0/1] priority-flow-control no-drop dot1p 3
```

Verifying the configuration

The DC server does not detect any packet loss during RoCE frame transmission.

Contents

Configuring L2PT 1

About L2PT 1
L2PT application scenario 1
Supported protocols 1
L2PT operating mechanism 2
L2PT tasks at a glance 3
Enabling L2PT 3
Restrictions and guidelines for L2PT 3
Enabling L2PT for a protocol in Layer 2 Ethernet interface view 4
Enabling L2PT for a protocol in Layer 2 aggregate interface view 4
Setting the destination multicast MAC address for tunneled packets 5
Configuring the maximum rate of transparently transmitting packets of the specified protocol on an L2TP-enabled interface 5
Display and maintenance commands for L2PT 6
L2PT configuration examples 6
Example: Configuring L2PT for STP 6
Example: Configuring L2PT for LACP 8

Configuring L2PT

About L2PT

Layer 2 Protocol Tunneling (L2PT) can transparently send Layer 2 protocol packets from geographically dispersed customer networks across a service provider network or drop them.

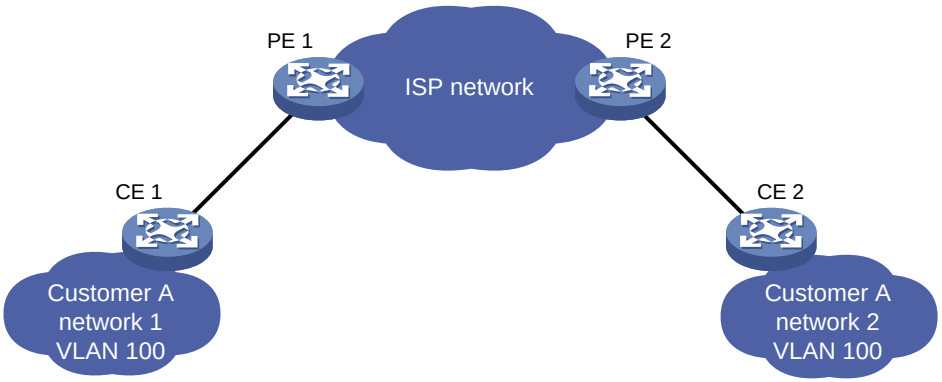
L2PT application scenario

Dedicated lines are used in a service provider network to build user-specific Layer 2 networks. As a result, a customer network contains sites located at different sides of the service provider network.

As shown in Figure 1, Customer A's network is divided into network 1 and network 2, which are connected by the service provider network. For Customer A's network to implement Layer 2 protocol calculations, the Layer 2 protocol packets must be transmitted across the service provider network.

Upon receiving a Layer 2 protocol packet, the PEs cannot determine whether the packet is from the customer network or the service provider network. They must deliver the packet to the CPU for processing. In this case, the Layer 2 protocol calculation in Customer A's network is mixed with the Layer 2 protocol calculation in the service provider network. Neither the customer network nor the service provider network can implement independent Layer 2 protocol calculations.

Figure 1 L2PT application scenario



L2PT is introduced to resolve the problem. L2PT provides the following functions:

- Multicasts Layer 2 protocol packets from a customer network in a VLAN. Dispersed customer networks can complete an independent Layer 2 protocol calculation, which is transparent to the service provider network.
- Isolates Layer 2 protocol packets from different customer networks through different VLANs.

Supported protocols

H3C devices support L2PT for the following protocols:

- CDP.
- DLDP.
- EOAM.
- GVRP.
- LACP.

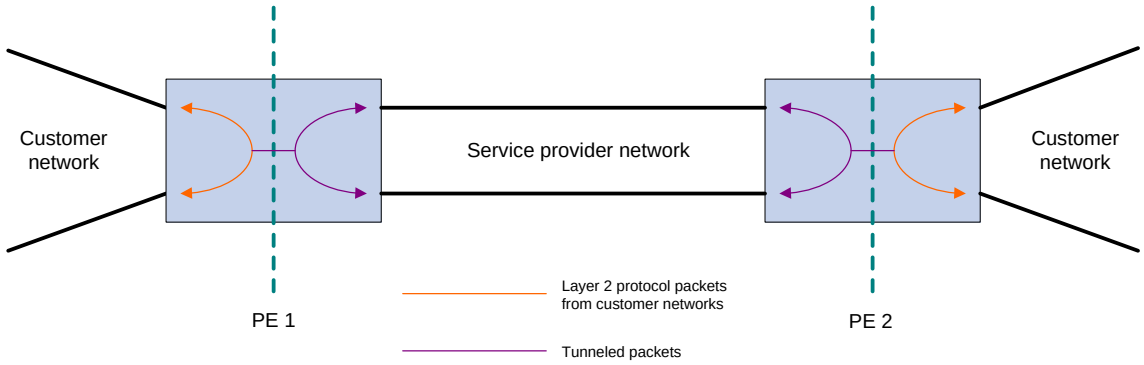
- LLDP.
- MVRP.
- PVST.
- STP (including STP, RSTP, and MSTP).

L2PT operating mechanism

As shown in Figure 2, L2PT operates as follows:

- When a port of PE 1 receives a Layer 2 protocol packet from the customer network in a VLAN, it performs the following operations:
  - Multicasts the packet out of all customer-facing ports in the VLAN except the receiving port.
  - Encapsulates the packet with a specified destination multicast address, and multicasts it out of all ISP-facing ports in the VLAN. The encapsulated packet is called the BPDU tunneled packet.
- When a port of PE 2 in the VLAN receives the tunneled packet from the service provider network, it performs the following operations:
  - Multicasts the packet out of all ISP-facing ports in the VLAN except the receiving port.
  - Decapsulates the packet and multicasts the decapsulated packet out of all customer-facing ports in the VLAN.

Figure 2 L2PT operating mechanism



For example, as shown in Figure 3, PE 1 receives an STP packet (BPDU) from network 1 to network 2. CEs are the edge devices on the customer network, and PEs are the edge devices on the service provider network. L2PT processes the packet as follows:

1. At the ingress of the ISP network, when an interface of PE 1 receives a BPDU from CE 1, the system matches the protocol packet based on locally configured L2PT protocol entries in a specific order. If a match is found, PE 1 replaces the destination MAC address of the original Layer 2 protocol packet with a fixed multicast MAC address (010f-e200-0003 by default) and forwards the protocol packet in the same broadcast domain.
2. At the egress of the ISP network, after PE 2 receives the packet with destination MAC address 010f-e200-0003, the system matches the packet based on locally configured L2PT protocol entries in a specific order. When a match is found, PE 2 restores the destination MAC address of the original packet according to the matching protocol and forwards the BPDU to CE 2.

Through L2PT, both the ISP network and Customer A's network can perform independent spanning tree calculations.

The L2PT protocol entries follow this matching order:

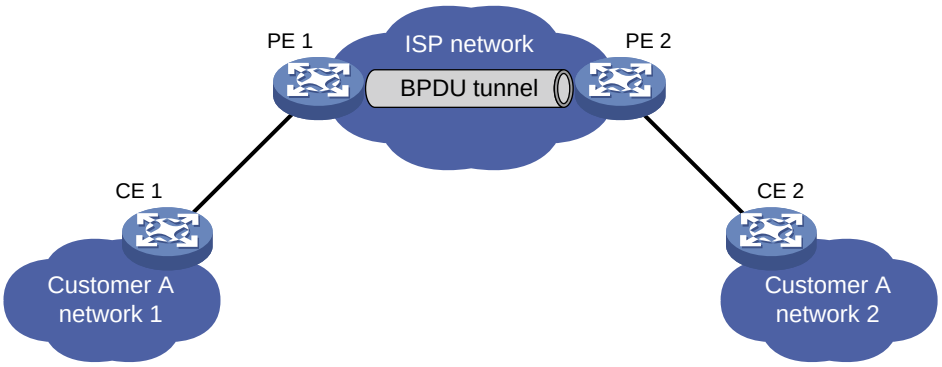
3. First, the system preferentially matches the L2PT protocols specified on an interface.

4. If no match is found, the system then matches the user-defined Layer 2 protocols. If an L2PT protocol on the interface is matched, the system does not continue to match the user-defined Layer 2 protocols.

NOTE:

When the ingress matches a packet with user-defined Layer 2 protocols for L2TP, the system compares the protocol type, encapsulation type, and destination MAC address. A match succeeds only if all three values are the same as those of a user-defined Layer 2 protocol. However, the protocol packet received at the egress has its destination MAC address replaced with a unified one. In this case, the egress cannot match the destination MAC address information. A match succeeds as long as the received packet matches the protocol type and encapsulation type of a user-defined Layer 2 protocol.

Figure 3 L2PT network diagram



L2PT tasks at a glance

To configure L2PT, perform the following tasks:

1. Enabling L2PT

   This feature is applicable only to customer-facing ports.
2. (Optional.) Setting the destination multicast MAC address for tunneled packets
3. (Optional.) Configuring the maximum rate of transparently transmitting packets of the specified protocol on an L2TP-enabled interface

Enabling L2PT

Restrictions and guidelines for L2PT

- To enable L2PT for a Layer 2 protocol on a port, perform the following tasks:
  - Enable the protocol on the connected CE, and disable the protocol on the port.
  - When a PE establishes a connection to a network device within the service provider network through CDP, you must enable CDP compatibility for LLDP on the PE. CDP compatibility for LLDP can be enabled only globally, and cannot be disabled separately on customer-facing interfaces. As a result, the CDP packets from the CE cannot be transparently transmitted within the service provider network. In this case, as a best practice, do not enable L2PT for CDP on the PE. For L2PT to take effect on CDP on the PE, you must disable CDP compatibility for LLDP globally on the PE, which will cause the PE to fail to communicate with the network devices within the service provider network through CDP. Before you disable CDP compatibility for LLDP on the PE, make sure you know its

influence on the network. For more information about CDP compatibility of LLDP, see "Configuring LLDP."

- Disable the protocol (for example, STP) on the PE ports connecting to an aggregate interface on a CE when the following conditions exist:
  - The protocol is running on the aggregate interface on the CE.
  - The aggregate interface on the CE connects to an L2PT-enabled port on the PE.
- Enable L2PT on PE ports connected to a customer network. If you enable L2PT on ports connected to the service provider network, L2PT determines that the ports are connected to a customer network.
- Make sure the VLAN tags of Layer 2 protocol packets are not changed or deleted for the tunneled packets to be transmitted correctly across the service provider network.

- L2PT for LLDP supports LLDP packets from only nearest bridge agents.
- You can enable L2PT on a member port of a Layer 2 aggregation group, but the configuration does not take effect.

Enabling L2PT for a protocol in Layer 2 Ethernet interface view

Restrictions and guidelines

LACP and EOAM require point-to-point transmission. If you enable L2PT on a Layer 2 Ethernet interface for LACP or EOAM, L2PT multicasts LACP or EOAM packets out of customer-facing ports. As a result, the transmission between two CEs is not point-to-point. To ensure point-to-point transmission for the LACP or EOAM packets, you must configure other features (for example, VLAN).

Procedure

1. Enter system view.

   system-view
2. Enter Layer 2 Ethernet interface view.

   interface *interface-type interface-number*
3. Enable L2PT for a protocol.

   l2protocol { **cdp** | **dldp** | **eoam** | **gvrp** | **lacp** | **lldp** | **mvrp** | **pvst** | **stp** } **tunnel dot1q**

   By default, L2PT is disabled for all protocols.

Enabling L2PT for a protocol in Layer 2 aggregate interface view

1. Enter system view.

   system-view
2. Enter Layer 2 aggregate interface view.

   interface bridge-aggregation *interface-type interface-number*
3. Enable L2PT for a protocol.

   l2protocol { **cdp** | **gvrp** | **lacp** | **lldp** | **mvrp** | **pvst** | **stp** } **tunnel dot1q**

   l2protocol { **all** | **cdp** | **gvrp** | **lacp** | **lldp** | **mvrp** | **pagp** | **pvst** | **stp** | **udld** | **vtp** } **tunnel dot1q** [**vlan**]

   By default, L2PT is disabled for all protocols.

Setting the destination multicast MAC address for tunneled packets

About this task

The default destination multicast MAC address for tunneled packets is 010f-e200-0003. You can modify the destination multicast MAC address for tunneled packets of the specific protocol or all protocols.

Restrictions and guidelines

The **l2protocol tunnel-dmac** command sets the destination multicast MAC address for tunneled packets of all protocols. The **l2protocol type tunnel-dmac** command sets the destination multicast MAC address for tunneled packets of the specified protocol. If both commands are executed, the **l2protocol type tunnel-dmac** command takes priority.

For tunneled packets to be recognized, set the same destination multicast MAC addresses for packets of the same protocol on PEs that are connected to the same customer network.

As a best practice, set different destination multicast MAC addresses on PEs connected to different customer networks. It prevents L2PT from sending packets of a customer network to another customer network.

Procedure

1. Enter system view.
   system-view
2. Set the destination multicast MAC address for tunneled packets of all protocols.
   l2protocol tunnel-dmac *mac-address*
   By default, 010f-e200-0003 is used for tunneled packets.

Configuring the maximum rate of transparently transmitting packets of the specified protocol on an L2TP-enabled interface

About this task

When there are a large number of protocol packets on the user side, the bandwidth resources of an L2TP-enabled interface might be all occupied by few protocols. As a result, packets of the other protocols will be congested and cannot be transparently transmitted. By configuring the maximum rate of transparently transmitting packets of each protocol on an interface, you can easily plan the bandwidth ratio for each protocol and fully utilize the interface bandwidth resources.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*

3. Configure the maximum rate of transparently transmitting packets of the specified protocol on the L2TP-enabled interface.
   - In Layer 2 Ethernet interface view:

     l2protocol rate-limit *rate* [**cdp** | **dldp** | **eoam** | **gvrp** | **lacp** | **lldp** | **mvrp** | **pvst** | **stp**]
   - In Layer 2 aggregate interface view:

     l2protocol rate-limit *rate* [**cdp** | **gvrp** | **lacp** | **lldp** | **mvrp** | **pvst** | **stp**]

   By default, the device does not limit the rate of transparently transmitting packets of a protocol on an L2TP-enabled interface.

   If you do not specify any protocol, this command configures the maximum rate of transparently transmitting packets of each protocol.

Display and maintenance commands for L2PT

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display L2PT statistics.	**display l2protocol statistics** [**interface** *interface-type interface-number*]
Clear L2PT statistics.	**reset l2protocol statistics** [**interface** *interface-type interface-number*]

L2PT configuration examples

Example: Configuring L2PT for STP

Network configuration

As shown in Figure 4, the MAC addresses of CE 1 and CE 2 are 00e0-fc02-5800 and 00e0-fc02-5802, respectively. MSTP is enabled in Customer A's network, and default MSTP settings are used.

Perform the following tasks on the PEs:

- Configure the ports that connect to CEs as access ports, and configure the ports in the service provider network as trunk ports. Configure ports in the service provider network to allow packets from any VLAN to pass.
- Enable L2PT for STP to enable Customer A's network to implement independent spanning tree calculation across the service provider network.
- Set the destination multicast MAC address to 0100-0ccd-cdd0 for tunneled packets.

Figure 4 Network diagram

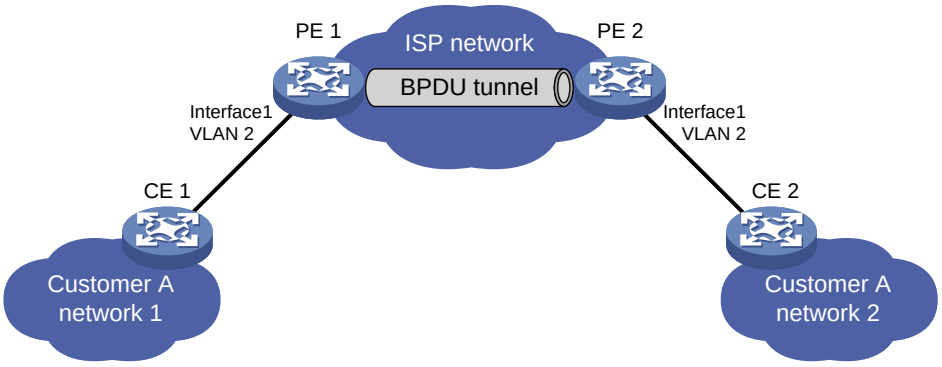


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure PE 1:

   # Set the destination multicast address to 0100-0ccd-cdd0 for tunneled packets.

   ```
   <PE1> system-view
   [PE1] l2protocol tunnel-dmac 0100-0ccd-cdd0
   ```

   # Create VLAN 2.

   ```
   [PE1] vlan 2
   [PE1-vlan2] quit
   ```

   # Configure GigabitEthernet 1/0/1 as an access port and assign the port to VLAN 2.

   ```
   [PE1] interface gigabitethernet 1/0/1
   [PE1-GigabitEthernet1/0/1] port access vlan 2
   ```

   # Disable STP and enable L2PT for STP on GigabitEthernet 1/0/1.

   ```
   [PE1-GigabitEthernet1/0/1] undo stp enable
   [PE1-GigabitEthernet1/0/1] l2protocol stp tunnel dot1q
   [PE1-GigabitEthernet1/0/1] quit
   ```

   # Configure GigabitEthernet 1/0/2 connected to the service provider network as a trunk port, and assign the port to all VLANs.

   ```
   [PE1] interface gigabitethernet 1/0/2
   [PE1-GigabitEthernet1/0/2] port link-type trunk
   [PE1-GigabitEthernet1/0/2] port trunk permit vlan all
   [PE1-GigabitEthernet1/0/2] quit
   ```

2. Configure PE 2 in the same way PE 1 is configured. (Details not shown.)

Verifying the configuration

Verify that the root bridge of Customer A's network is CE 1.

```
<CE2> display stp root
 MST ID   Root Bridge ID       ExtPathCost IntPathCost Root Port
 0        32768.00e0-fc02-5800 0           0
```

Verify that the root bridge of the service provider network is not CE 1.

```
[PE1] display stp root
 MST ID   Root Bridge ID       ExtPathCost IntPathCost Root Port
 0        32768.0cda-41c5-ba50 0           0
```

Example: Configuring L2PT for LACP

Network configuration

As shown in Figure 5, the MAC addresses of CE 1 and CE 2 are 0001-0000-0000 and 0004-0000-0000, respectively.

Perform the following tasks:

- Configure Ethernet link aggregation on CE 1 and CE 2.
- Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 on CE 1 to form aggregate links with GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 on CE 2, respectively.
- Enable L2PT for LACP to enable CE 1 and CE 2 to implement Ethernet link aggregation across the service provider network.

Figure 5 Network diagram

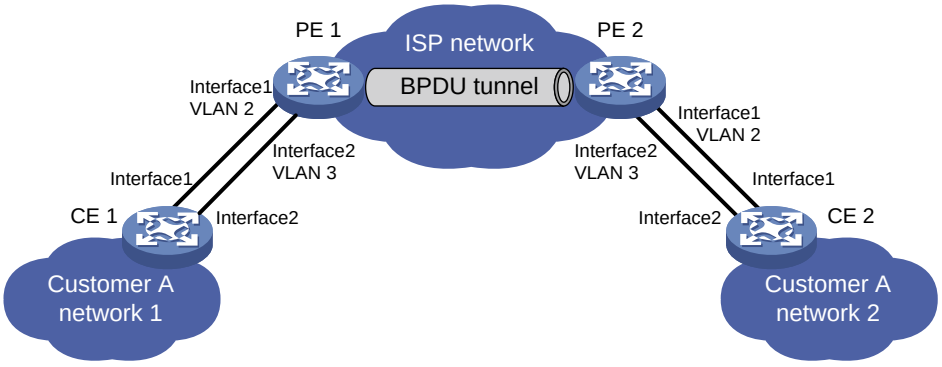


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Requirements analysis

To meet the network requirements, perform the following tasks:

- For Ethernet link aggregation to operate correctly, configure VLANs on the PEs to ensure point-to-point transmission between CE 1 and CE 2 in an aggregation group.
  - Set the PVIDs to VLAN 2 and VLAN 3 for GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 on PE 1, respectively.
  - Configure PE 2 in the same way PE 1 is configured.
  - Configure ports that connect to the CEs as trunk ports.
- To retain the VLAN tag of the customer network, enable QinQ on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 on both PE 1 and PE 2.
- For packets from any VLAN to be transmitted, configure all ports in the service provider network as trunk ports.

Procedure

1. Configure CE 1:

   # Configure Layer 2 aggregation group Bridge-Aggregation 1 to operate in dynamic aggregation mode.

```
<CE1> system-view
[CE1] interface bridge-aggregation 1
[CE1-Bridge-Aggregation1] port link-type access
[CE1-Bridge-Aggregation1] link-aggregation mode dynamic
[CE1-Bridge-Aggregation1] quit
```

   # Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to Bridge-Aggregation 1.

```
[CE1] interface gigabitethernet 1/0/1
[CE1-GigabitEthernet1/0/1] port link-aggregation group 1
[CE1-GigabitEthernet1/0/1] quit
[CE1] interface gigabitethernet 1/0/2
[CE1-GigabitEthernet1/0/2] port link-aggregation group 1
[CE1-GigabitEthernet1/0/2] quit
```

2. Configure CE 2 in the same way CE 1 is configured. (Details not shown.)
3. Configure PE 1:

   # Create VLANs 2 and 3.

```
<PE1> system-view
[PE1] vlan 2
[PE1-vlan2] quit
[PE1] vlan 3
[PE1-vlan3] quit
```

   # Configure GigabitEthernet 1/0/1 as a trunk port, assign the port to VLAN 2, and set the PVID to VLAN 2.

```
[PE1] interface gigabitethernet 1/0/1
[PE1-GigabitEthernet1/0/1] port link-mode bridge
[PE1-GigabitEthernet1/0/1] port link-type trunk
[PE1-GigabitEthernet1/0/1] port trunk permit vlan 2
[PE1-GigabitEthernet1/0/1] port trunk pvid vlan 2
```

   # Enable QinQ on GigabitEthernet 1/0/1.

```
[PE1-GigabitEthernet1/0/1] qinq enable
```

   # Enable L2PT for LACP on GigabitEthernet 1/0/1.

```
[PE1-GigabitEthernet1/0/1] l2protocol lacp tunnel dot1q
[PE1-GigabitEthernet1/0/1] quit
```

   # Configure GigabitEthernet 1/0/2 as a trunk port, assign the port to VLAN 3, and set the PVID to VLAN 3.

```
[PE1] interface gigabitethernet 1/0/2
[PE1-GigabitEthernet1/0/2] port link-mode bridge
[PE1-GigabitEthernet1/0/2] port link-type trunk
[PE1-GigabitEthernet1/0/2] port trunk permit vlan 3
[PE1-GigabitEthernet1/0/2] port trunk pvid vlan 3
```

   # Enable QinQ on GigabitEthernet 1/0/2.

```
[PE1-GigabitEthernet1/0/2] qinq enable
```

   # Enable L2PT for LACP on GigabitEthernet 1/0/2.

```
[PE1-GigabitEthernet1/0/2] l2protocol lacp tunnel dot1q
[PE1-GigabitEthernet1/0/2] quit
```

4. Configure PE 2 in the same way PE 1 is configured. (Details not shown.)

Verifying the configuration

Verify that CE 1 and CE 2 have completed Ethernet link aggregation successfully.

```
[CE1] display link-aggregation member-port
Flags: A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
       D -- Synchronization, E -- Collecting, F -- Distributing,
       G -- Defaulted, H -- Expired

GigabitEthernet1/0/1:
Aggregate Interface: Bridge-Aggregation1
Local:
   Port Number: 3
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Remote:
   System ID: 0x8000, 0004-0000-0000
   Port Number: 3
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Received LACP Packets: 23 packet(s)
Illegal: 0 packet(s)
Sent LACP Packets: 26 packet(s)

GigabitEthernet1/0/2:
Aggregate Interface: Bridge-Aggregation1
Local:
   Port Number: 4
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Remote:
   System ID: 0x8000, 0004-0000-0000
   Port Number: 4
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Received LACP Packets: 10 packet(s)
Illegal: 0 packet(s)
Sent LACP Packets: 13 packet(s)
[CE2] display link-aggregation member-port
Flags: A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
       D -- Synchronization, E -- Collecting, F -- Distributing,
       G -- Defaulted, H -- Expired

GigabitEthernet1/0/1:
Aggregate Interface: Bridge-Aggregation1
```

```
Local:
   Port Number: 3
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Remote:
   System ID: 0x8000, 0001-0000-0000
   Port Number: 3
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Received LACP Packets: 23 packet(s)
Illegal: 0 packet(s)
Sent LACP Packets: 26 packet(s)

GigabitEthernet1/0/2:
Aggregate Interface: Bridge-Aggregation1
Local:
   Port Number: 4
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Remote:
   System ID: 0x8000, 0001-0000-0000
   Port Number: 4
   Port Priority: 32768
   Oper-Key: 1
   Flag: {ACDEF}
Received LACP Packets: 10 packet(s)
Illegal: 0 packet(s)
Sent LACP Packets: 13 packet(s)
```

Contents

Configuring cut-through Layer 2 forwarding 1
About cut-through Layer 2 forwarding 1
Restrictions and guidelines for cut-through Layer 2 forwarding configuration 1
Procedure 1

Configuring cut-through Layer 2 forwarding

About cut-through Layer 2 forwarding

A cut-through forwarding-enabled device forwards a frame after it receives the first 64 bytes of the frame. This feature reduces the transmission time of a frame and enhances forwarding performance.

Restrictions and guidelines for cut-through Layer 2 forwarding configuration

With cut-through forwarding, the device forwards CRC-error frames because it starts forwarding frames before their CRC field is received.

In an IRF fabric, cut-through Layer 2 forwarding does not take effect for traffic where the ingress and egress interfaces are not on the same member device.

Cut-through Layer 2 forwarding takes effect only on multicast packets.

Procedure

1. Enter system view.
   `system-view`
2. Enable cut-through forwarding.
   `cut-through enable`
   By default, cut-through forwarding is disabled.

Contents

Configuring Ethernet interfaces ····· 1

About Ethernet interface ····· 1
Configuring a management Ethernet interface ····· 1
Ethernet interface naming conventions ····· 2
Configuring common Ethernet interface settings ····· 2
Configuring the physical type for a combo interface (single combo interface) ····· 2
Configuring basic settings of an Ethernet interface ····· 3
Configuring basic settings of an Ethernet subinterface ····· 4
Enabling automatic negotiation for speed downgrading ····· 4
Configuring the link mode of an Ethernet interface ····· 5
Configuring jumbo frame support ····· 5
Configuring physical state change suppression on an Ethernet interface ····· 6
Configuring dampening on an Ethernet interface ····· 7
Enabling link flapping protection on an interface ····· 8
Configuring FEC ····· 10
Configuring broadcast/multicast/unknown unicast storm suppression ····· 10
Configuring generic flow control on an Ethernet interface ····· 11
Enabling energy saving features on an Ethernet interface ····· 12
Setting the statistics polling interval ····· 12
Enabling interframe gap and preamble statistics in the traffic statistics ····· 13
Enabling loopback testing on an Ethernet interface ····· 13
Forcibly bringing up a fiber port ····· 14
Configuring interface alarm functions ····· 15
Configuring SNMP notifications for cards ····· 19
Enabling SNMP notifications for the ETH module ····· 20
Shutting down all physical interfaces ····· 20
Configuring the port speed mode for expansion slots ····· 21
Configuring the port speed mode for the device ····· 22
Restoring the default settings for an interface ····· 23
Enabling packet latency detection ····· 23
Configuring a Layer 2 Ethernet interface ····· 24
Setting speed options for autonegotiation on an Ethernet interface ····· 24
Setting the MDIX mode of an Ethernet interface ····· 25
Configuring storm control on an Ethernet interface ····· 26
Testing the cable connection of an Ethernet interface ····· 27
Enabling bridging on an Ethernet interface ····· 28
Configuring a Layer 3 Ethernet interface or subinterface ····· 28
Setting the MTU for an Ethernet interface or subinterface ····· 28
Setting the MAC address of an Ethernet interface or subinterface ····· 29
Display and maintenance commands for Ethernet interfaces ····· 29

Configuring Ethernet interfaces

About Ethernet interface

The Switch Series supports Ethernet interfaces, management Ethernet interfaces, Console interfaces, and USB interfaces. For the interface types and the number of interfaces supported by a switch model, see the installation guide and hardware manuals.

This chapter describes how to configure management Ethernet interfaces and Ethernet interfaces.

Configuring a management Ethernet interface

About this task

You can connect a management Ethernet interface to a PC for software loading and system debugging, or connect it to a remote NMS for remote system management.

Each member device in an IRF system has a management Ethernet interface. For management link backup, perform the following tasks:

1. Connect your PC to the management Ethernet interface on the master device.
2. Connect the PC to a management Ethernet interface with the same interface number on a subordinate device.

The two management Ethernet interfaces operate as follows:

- When the IRF system has multiple management Ethernet interfaces, only the management Ethernet interface on the master device processes management traffic.
- When the management Ethernet interface on the master device fails, the management Ethernet interface on the subordinate device takes over to process management traffic.
- When the management Ethernet interface on the master device recovers, it takes over to process management traffic again.

Procedure

1. Enter system view.

   system-view
2. Enter management Ethernet interface view.

   interface M-GigabitEthernet *interface-number*
3. (Optional.) Set the interface description.

   description *text*

   The default setting is **M-GigabitEthernet0/0/0 Interface**.
4. (Optional.) Set the duplex mode for the management Ethernet interface.

   duplex { **auto** | **full** | **half** }

   By default, the duplex mode is **auto** for a management Ethernet interface.
5. (Optional.)_Set the speed for the management Ethernet interface.

   speed { **10** | **100** | **1000** | **auto** }

   By default, the speed is **auto** for a management Ethernet interface.
6. (Optional.) Shut down the interface.

   shutdown

   By default, the management Ethernet interface is up.

CAUTION:

Executing the **shutdown** command on an interface will disconnect the link of the interface and interrupt communication. Use this command with caution.

Ethernet interface naming conventions

The Ethernet interfaces are named in the format of **interface type A/B/C**. The letters that follow the interface type represent the following elements:

- **A**—IRF member ID. If the switch is not in an IRF fabric, A is 1 by default.
- **B**—Card slot number. **0** indicates the interface is a fixed interface of the switch. **1** indicates the interface is on expansion interface card 1. **2** indicates the interface is on expansion interface card 2.
- **C**—Port index.

Configuring common Ethernet interface settings

This section describes the settings common to Layer 2 Ethernet interfaces, Layer 3 Ethernet interfaces, and Layer 3 Ethernet subinterfaces. For more information about the settings specific to Layer 2 Ethernet interfaces, see "Configuring a Layer 2 Ethernet interface." For more information about the settings specific to Layer 3 Ethernet interfaces or subinterfaces, see "Configuring a Layer 3 Ethernet interface or subinterface."

Configuring the physical type for a combo interface (single combo interface)

About this task

A combo interface is a logical interface that physically comprises one fiber combo port and one copper combo port. The two ports share one forwarding channel and one interface view. As a result, they cannot work simultaneously. When you activate one port, the other port is automatically disabled. If you execute the **combo enable auto** command on a combo interface, the interface automatically identifies the media inserted and activates the corresponding combo port. In the interface view, you can activate the fiber or copper combo port, and configure other port attributes such as the interface rate and duplex mode.

Prerequisites

Before you configure combo interfaces, complete the following tasks:

- Determine the combo interfaces on your device. Identify the two physical interfaces that belong to each combo interface according to the marks on the device panel.
- Use the **display interface** command to determine which port (fiber or copper) of each combo interface is active:
  - If the copper port is active, the output includes "Media type is twisted pair."
  - If the fiber port is active, the output does not include this information.

Procedure

1. Enter system view.

   system-view
2. Enter Ethernet interface view.

   interface *interface-type interface-number*

3. Activate the copper combo port or fiber combo port.

   combo enable { **auto** | **copper** | **fiber** }

   By default, an interface automatically identifies the media inserted and activates the corresponding combo port.

Configuring basic settings of an Ethernet interface

About this task

You can configure an Ethernet interface to operate in one of the following duplex modes:

- **Full-duplex mode**—The interface can send and receive packets simultaneously.
- **Half-duplex mode**—The interface can only send or receive packets at a given time.
- **Autonegotiation mode**—The interface negotiates a duplex mode with its peer.

You can set the speed of an Ethernet interface or enable it to automatically negotiate a speed with its peer. For a 100-Mbps or 1000-Mbps Layer 2 Ethernet interface, you can also set speed options for autonegotiation. The two ends can select a speed only from the available options. For more information, see "Setting speed options for autonegotiation on an Ethernet interface."

Restrictions and guidelines

The **shutdown**, **port up-mode**, and **loopback** commands are mutually exclusive.

Procedure

1. Enter system view.

   system-view

2. Configure the device to display the default description for interfaces when it displays the interface information.

   interface default-description output

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Set the description for the Ethernet interface.

   description *text*

   The default setting is *interface-name* **Interface**. For example, **GigabitEthernet1/0/1 Interface**.

5. Set the duplex mode for the Ethernet interface.

   duplex { **auto** | **full** | **half** }

   By default, the duplex mode is **auto** for Ethernet interfaces.

   Ethernet copper ports that operate in 1000 Mbps or 10000 Mbps and fiber ports do not support the **half** keyword.

6. Set the speed for the Ethernet interface.

   speed { **10** | **100** | **1000** | **10000** | **25000** | **40000** | **100000** | **auto** }

   By default, an Ethernet interface negotiates a speed with its peer.

7. Set the expected bandwidth for the Ethernet interface.

   bandwidth *bandwidth-value*

   By default, the expected bandwidth (in kbps) is the interface baud rate divided by 1000.

8. Bring up the Ethernet interface.

   undo shutdown

   By default, Ethernet interfaces are in up state.

Configuring basic settings of an Ethernet subinterface

About this task

By default, a Layer 3 Ethernet subinterface processes packets for only the VLAN whose ID is the same as the subinterface number.

Restrictions and guidelines for Ethernet subinterface basic settings

- The **shutdown**, **port up-mode**, and **loopback** commands are mutually exclusive.
- The **shutdown** command cannot be executed on an Ethernet interface in a loopback test.
- When an Ethernet subinterface is used as the egress interface of the public network, it cannot forward traffic of the public network.

Procedure

1. Enter system view.

   system-view
2. Configure the device to display the default description for interfaces when it displays the interface information.

   interface default-description output
3. Create an Ethernet subinterface.

   interface *interface-type interface-number.subnumber*
4. Set the description for the Ethernet subinterface.

   description *text*

   The default setting is *interface-name* **Interface**. For example, **GigabitEthernet1/0/1.1 Interface**.
5. Set the expected bandwidth for the Ethernet subinterface.

   bandwidth *bandwidth-value*

   By default, the expected bandwidth (in kbps) is the interface baud rate divided by 1000.
6. Bring up the Ethernet subinterface.

   undo shutdown

   By default, Ethernet subinterfaces are in up state.

Enabling automatic negotiation for speed downgrading

About this task

Perform this task to enable GE or 10-GE interfaces at two ends of a link to automatically negotiate about downgrading their speed when the following conditions exist:

- The interfaces automatically negotiate a speed of 1000 Mbps or 10000 Mbps.
- The interfaces cannot operate at 1000 Mbps or 10000 Mbps because of link restrictions.

Take a GE Ethernet interface as an example. When a GE interface can only operate at 100 Mbps because of the aging of the network cable, packet loss occurs. You can enable automatic negotiation for speed downgrading for the interface to automatically negotiate a speed of 100 Mbps to ensure normal operation of the interface.

Restrictions and guidelines

This feature is available only on GE and 10GE interfaces.

Procedure

1. Enter system view.

system-view

2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Enable automatic negotiation for speed downgrading.
   speed auto downgrade
   The default settings vary by device model.

Configuring the link mode of an Ethernet interface

About this task

Interfaces on the device can operate either as Layer 2 or Layer 3 Ethernet interfaces. You can use commands to set the link mode to bridge or route.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Configure the link mode of the Ethernet interface.
   port link-mode { **bridge** | **route** }
   By default, all Ethernet interfaces on the device operate in bridge mode.

CAUTION:

After you change the link mode of an Ethernet interface, all commands (except the **description**, **duplex**, **jumboframe enable**, **speed**, **shutdown**, **port fec mode** , and **combo enable** commands) on the Ethernet interface are restored to their defaults in the new link mode.

Configuring jumbo frame support

About this task

Jumbo frames are frames larger than 1536 bytes and are typically received by an Ethernet interface during high-throughput data exchanges, such as file transfers.

The Ethernet interface processes jumbo frames in the following ways:

- When the Ethernet interface is configured to deny jumbo frames (by using the **undo jumboframe enable** command), the Ethernet interface discards jumbo frames.
- When the Ethernet interface is configured with jumbo frame support, the Ethernet interface performs the following operations:
  - Processes jumbo frames within the specified length.
  - Discards jumbo frames that exceed the specified length.

Restrictions and guidelines

If you execute this command multiple times on the same interface, the most recent configuration takes effect.

If you execute this command on multiple interfaces in the same slot, the most recent configuration takes effect on all interfaces of the slot.

To assign a physical interface to an aggregate interface, make sure the physical interface is configured with the same maximum jumbo frame size (13312) as the aggregate interface. After assigning a physical interface to an aggregate interface, you cannot modify the maximum jumbo frame size of the physical interface.

If a physical interface in a slot is configured as an IRF physical interface, the maximum jumbo frame size is restored to the default for all physical interfaces in the slot and cannot be modified.

You cannot modify the maximum jumbo frame size of an aggregate interface.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Configure jumbo frame support.
   jumboframe enable [*size*]
   By default, the device allows jumbo frames within 13312 bytes to pass through.

Configuring physical state change suppression on an Ethernet interface

About this task

The physical link state of an Ethernet interface is either up or down. Each time the physical link of an interface comes up or goes down, the interface immediately reports the change to the CPU. The CPU then performs the following operations:

- Notifies the upper-layer protocol modules (such as routing and forwarding modules) of the change for guiding packet forwarding.
- Automatically generates traps and logs to inform users to take the correct actions.

To prevent frequent physical link flapping from affecting system performance, configure physical state change suppression. You can configure this feature to suppress only link-down events, only link-up events, or both. If an event of the specified type still exists when the suppression interval expires, the system reports the event to the CPU.

Restrictions and guidelines

Do not enable this feature on an interface that has RRPP, spanning tree protocols, or Smart Link enabled.

You can configure different suppression intervals for link-up and link-down events.

If you execute the **link-delay** command multiple times on an interface, the following rules apply:

- You can configure the suppression intervals for link-up and link-down events separately.
- If you configure the suppression interval multiple times for link-up or link-down events, the most recent configuration takes effect.

The **link-delay**, **dampening**, and **port link-flap protect enable** commands are mutually exclusive on an Ethernet interface.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*

3. Configure physical state change suppression.

   link-delay { **down** | **up** } [**msec**] *delay-time*

   By default, each time the physical link of an interface goes up or comes down, the interface immediately reports the change to the CPU.

Configuring dampening on an Ethernet interface

About this task

The interface dampening feature uses an exponential decay mechanism to prevent excessive interface flapping events from adversely affecting routing protocols and routing tables in the network. Suppressing interface state change events protects the system resources.

If an interface is not dampened, its state changes are reported. For each state change, the system also generates an SNMP trap and log message.

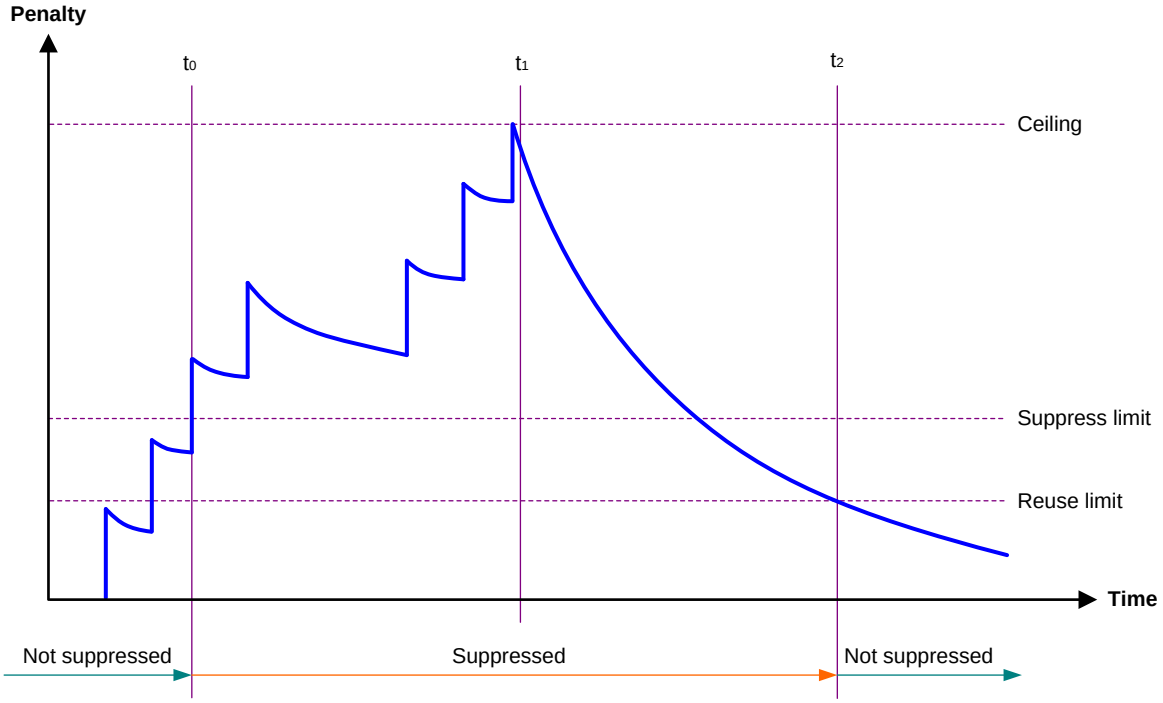
After a flapping interface is dampened, it does not report its state changes to the CPU. For state change events, the interface only generates SNMP trap and log messages.

Parameters

- **Penalty**—The interface has an initial penalty of 0. When the interface flaps, the penalty increases by 1000 for each down event until the ceiling is reached. It does not increase for up events. When the interface stops flapping, the penalty decreases by half each time the half-life timer expires until the penalty drops to the reuse threshold.
- **Ceiling**—The penalty stops increasing when it reaches the ceiling.
- **Suppress-limit**—The accumulated penalty that triggers the device to dampen the interface. In dampened state, the interface does not report its state changes to the CPU. For state change events, the interface only generates SNMP traps and log messages.
- **Reuse-limit**—When the accumulated penalty decreases to this reuse threshold, the interface is not dampened. Interface state changes are reported to the upper layers. For each state change, the system also generates an SNMP trap and log message.
- **Decay**—The amount of time (in seconds) after which a penalty is decreased.
- **Max-suppress-time**—The maximum amount of time the interface can be dampened. If the penalty is still higher than the reuse threshold when this timer expires, the penalty stops increasing for down events. The penalty starts to decrease until it drops below the reuse threshold.
- When configuring the **dampening** command, follow these rules to set the values mentioned above:
- The ceiling is equal to $2^{(\text{Max-suppress-time/Decay})}$ × reuse-limit. It is not user configurable.
- The configured suppress limit is lower than or equal to the ceiling.
- The ceiling is lower than or equal to the maximum suppress limit supported.

Figure 1 shows the change rule of the penalty value. The lines t_0 and t_2 indicate the start time and end time of the suppression, respectively. The period from t_0 to t_2 indicates the suppression period, t_0 to t_1 indicates the max-suppress-time, and t_1 to t_2 indicates the complete decay period.

Figure 1 Change rule of the penalty value



Restrictions and guidelines

- The **dampening**, **link-delay**, and **port link-flap protect enable** commands are mutually exclusive on an interface.
- The **dampening** command does not take effect on the administratively down events. When you execute the **shutdown** command, the penalty restores to 0, and the interface reports the down event to the upper-layer protocols.
- Do not enable the dampening feature on an interface with RRPP, MSTP, or Smart Link enabled.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Enable dampening on the interface.
   dampening [*half-life reuse suppress max-suppress-time*]
   By default, interface dampening is disabled on Ethernet interfaces.

Enabling link flapping protection on an interface

About this task

Link flapping on an interface changes network topology and increases the system overhead. For example, in an active/standby link scenario, when interface status on the active link changes between **UP** and **DOWN**, traffic switches between active and standby links. To solve this problem, configure this feature on the interface.

With this feature enabled on an interface, when the interface goes down, the system enables link flapping detection. During the link flapping detection interval, if the number of detected flaps reaches or exceeds the link flapping detection threshold, the system shuts down the interface.

When a port is shut down automatically by a feature, the device starts a port status detection timer. When the timer expires, the device restores the port status to its actual physical status automatically.

Restrictions and guidelines

This feature takes effect only if it is configured in both the system view and interface view.

If you specify the **down-auto-recovery** keyword in both system view and interface view, an interface comes up automatically when the port status detection timer for the ports that are shut down by the link flapping protection feature expires.

When you configure this feature on an interface, how link flapping protection is triggered depends on the parameters specified:

- If you specify level-1 link flapping protection parameters but do not specify level-2 parameters, link flapping protection will be triggered and the interface will be shut down when the level-1 link flapping protection conditions are met.
- If you specify both level-1 and level-2 link flapping protection parameters, link flapping protection will be triggered and the interface will be shut down when either of level-1 and level-2 link flapping protection conditions are met.

IRF system stability might be affected by IRF physical link flapping. For IRF system stability, this feature is enabled by default on IRF physical interfaces and the enabling status of this feature is not affected by the status of global link flapping protection. When the number of flaps detected on an IRF physical interface exceeds the threshold within the detection interval, the device outputs a log rather than shuts down the IRF physical interface.

The **dampening**, **link-delay**, and **port link-flap protect enable** commands are mutually exclusive on an Ethernet interface.

To bring up an interface that has been shut down by link flapping protection, execute the **undo shutdown** command.

In the **display interface** command output, the **Link-Flap DOWN** value of the **Current state** field indicates that the interface has been shut down by link flapping protection.

If you change the timer setting during port detection, the device compares the new setting (T1) with the time that elapsed since the port was shut down (T).

- If T < T1, the port will be brought up after T1 – T seconds.
- If T ≥ T1, the port is brought up immediately.

For example, the timer setting is 30 seconds. If you change it to 10 seconds 2 seconds after the port is shut down, the port will come up 8 seconds later. If you change the timer setting to 2 seconds 10 seconds after the port is shut down, the port comes up immediately.

Procedure

1. Enter system view.

   system-view

2. Enable link flapping protection globally.

   link-flap protect enable [**down-auto-recovery**]

   By default, link flapping protection is disabled globally.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Enable link flapping protection on the Ethernet interface.

   port link-flap protect enable [**down-auto-recovery**] [**interval** *interval* | **threshold** *threshold*] * [**second-interval** *second-interval* **second-threshold** *second-threshold*]

   By default, link flapping protection is disabled on an Ethernet interface.

5. Return to system view.

quit

6. Set the port status detection timer for the ports that are shut down by the link flapping protection feature.

   shutdown-interval [**link-flap**] *interval*

   By default, the timer is 30 seconds.

Configuring FEC

About this task

The forward error correction (FEC) feature corrects packet errors to improve transmission quality. It attaches correction information to a packet at the sending end, and corrects error codes generated during transmission at the receiving end based on the correction information. You can set the FEC mode as needed.

Restrictions and guidelines

This feature is supported only on 25-GE interfaces, and 100-GE interfaces.

Make sure you set the same FEC mode for both interfaces of a link.

Procedure

1. Enter system view.

   system-view

2. Enter Ethernet interface view.

   interface *interface-type interface-number*

3. Set the FEC mode of the Ethernet interface.

   port fec mode { auto | base-r | none | rs-fec }

   By default, the FEC mode of an Ethernet interface is autonegotiation.

   The **rs-fec** keyword is not supported on 25-GE interfaces.

   The **base-r** keyword is not supported on 100-GE interfaces.

Configuring broadcast/multicast/unknown unicast storm suppression

About this task

The storm suppression feature ensures that the size of a particular type of traffic (broadcast, multicast, or unknown unicast traffic) does not exceed the threshold on an interface. When the broadcast, multicast, or unknown unicast traffic on the interface exceeds this threshold, the system discards packets until the broadcast, multicast, or unknown unicast traffic drops below this threshold.

Both storm suppression and storm control can suppress storms on an interface. Storm suppression uses the chip to suppress traffic. Storm suppression has less impact on the device performance than storm control, which uses software to suppress traffic.

Restrictions and guidelines

- For the traffic suppression result to be determined, do not configure storm control together with storm suppression for the same type of traffic (broadcast/multicast/unknown unicast traffic). For more information about storm control, see "Configuring storm control on an Ethernet interface."
- The configured suppression threshold value in pps or kbps will be converted into a multiple of the step value supported by the chip. As a result, the effective suppression threshold might be different from the configured one. For information about the suppression threshold that takes effect, see the prompt on the device.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Enable broadcast suppression and set the broadcast suppression threshold.
   broadcast-suppression { *ratio* | **pps** *max-pps* | **kbps** *max-kbps* }
   By default, broadcast suppression is disabled.
4. Enable multicast suppression and set the multicast suppression threshold.
   multicast-suppression { *ratio* | **pps** *max-pps* | **kbps** *max-kbps* }
   By default, multicast suppression is disabled.
5. Enable unknown unicast suppression and set the unknown unicast suppression threshold.
   unicast-suppression { *ratio* | **pps** *max-pps* | **kbps** *max-kbps* }
   By default, unknown unicast suppression is disabled.

Configuring generic flow control on an Ethernet interface

About this task

To avoid dropping packets on a link, you can enable generic flow control at both ends of the link. When traffic congestion occurs at the receiving end, the receiving end sends a flow control (Pause) frame to ask the sending end to suspend sending packets. Generic flow control includes the following types:

- **TxRx-mode generic flow control**—Enabled by using the **flow-control** command. With TxRx-mode generic flow control enabled, an interface can both send and receive flow control frames:
  - When congestion occurs, the interface sends a flow control frame to its peer.
  - When the interface receives a flow control frame from its peer, it suspends sending packets to its peer.
- **Rx-mode generic flow control**—Enabled by using the **flow-control receive enable** command. With Rx-mode generic flow control enabled, an interface can receive flow control frames, but it cannot send flow control frames:
  - When congestion occurs, the interface cannot send flow control frames to its peer.
  - When the interface receives a flow control frame from its peer, it suspends sending packets to its peer.

To handle unidirectional traffic congestion on a link, configure the **flow-control receive enable** command at one end and the **flow-control** command at the other end. To enable both ends of a link to handle traffic congestion, configure the **flow-control** command at both ends.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Enable generic flow control.
   - Enable TxRx-mode generic flow control.
     flow-control

- Enable Rx-mode generic flow control.

  flow-control receive enable

By default, generic flow control is disabled on an Ethernet interface.

Enabling energy saving features on an Ethernet interface

About this task

With Energy Efficient Ethernet (EEE) enabled, a link-up interface enters low power state if it has not received any packet for a period of time. The time period depends on the chip specifications and is not configurable. When a packet arrives later, the device automatically restores power supply to the interface and the interface restores to the normal state.

Restrictions and guidelines

Fiber ports do not support this feature.

Configuring EEE on an Ethernet interface

1. Enter system view.

   system-view
2. Enter Ethernet interface view.

   interface *interface-type interface-number*
3. Enable EEE on the Ethernet interface.

   eee enable

   By default, EEE is disabled on an Ethernet interface.

Setting the statistics polling interval

About this task

By setting the statistics polling interval, you can collect statistics of packets and analyze packets at the specified interval. Based on the interface traffic statistics, you can take traffic control measures promptly to avoid network congestion and service interruption.

- When network congestion is detected, you can set the statistics polling interval to be smaller than 300 seconds (30 seconds when congestion deteriorates). Then, check traffic distribution on interfaces within a short period of time. For data packets that cause congestion, take traffic control measures.
- When the network bandwidth is sufficient and services are operating normally, you can set the statistics polling interval to be greater than 300 seconds. Once traffic parameter anomalies occur, modify the statistics polling interval promptly so that you can observe the traffic parameter trend in real time.

To display the interface statistics collected in the last statistics polling interval, use the **display interface** command. To clear the interface statistics, use the **reset counters interface** command.

Setting the statistics polling interval in Ethernet interface view

1. Enter system view.

   system-view
2. Enter Ethernet interface view.

   interface *interface-type interface-number*
3. Set the statistics polling interval for the Ethernet interface.

   flow-interval *interval*

By default, the statistics polling interval is 300 seconds.

Enabling interframe gap and preamble statistics in the traffic statistics

About this task

With the **display interface** command executed, the **Last 300 seconds input rate** or **Last 300 seconds output rate** field in the command output displays the average outbound or inbound traffic rate in the last 300 seconds.

By default, traffic rate = native frame length × packet count per second. Execute the **traffic-statistic include-interframe** command if you need the total traffic statistics, including the native frame length, interframe gap length and preamble length, for a specific time period. With this command executed, traffic rate = (native frame length + interframe gap length + preamble length) × packet count per second.

By default, the interframe gap length and preamble length are fixed. However, you can execute the **itf number** command to adjust the interframe gap length. Adjusting the interframe gap length causes inaccuracy in only the traffic statistics being collected.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Enable interframe gap and preamble statistics in the traffic statistics.
   traffic-statistic include-interframe
   By default, interframe gap and preamble statistics in the traffic statistics are disabled.

Enabling loopback testing on an Ethernet interface

About this task

Perform this task to determine whether an Ethernet link works correctly.

Internal loopback testing tests the device where the Ethernet interface resides. The Ethernet interface sends outgoing packets back to the local device. If the device fails to receive the packets, the device fails.

Restrictions and guidelines

- After you enable this feature on an Ethernet interface, the interface does not forward data traffic.

The **shutdown**, **port up-mode**, and **loopback** commands are mutually exclusive.

- After you enable this feature on an Ethernet interface, the Ethernet interface switches to full duplex mode. After you disable this feature, the Ethernet interface restores to its duplex setting.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Configure loopback testing on the Ethernet interface.

- Enable loopback testing on the Ethernet interface.
  `loopbackinternal`
  By default, loopback testing is disabled on an Ethernet interface.

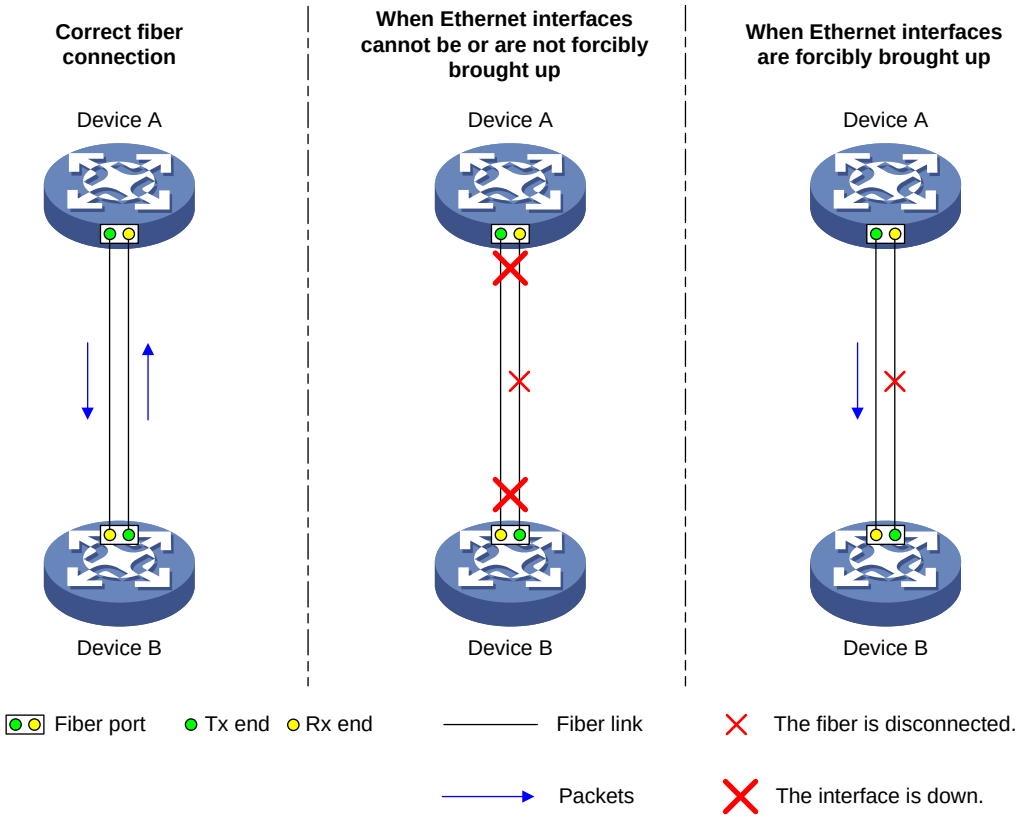
Forcibly bringing up a fiber port

About this task

As shown in Figure 2, a fiber port uses separate fibers for transmitting and receiving packets. The physical state of the fiber port is up only when both transmit and receive fibers are physically connected. If one of the fibers is disconnected, the fiber port does not work.

To enable a fiber port to forward traffic over a single link, you can use the **`port up-mode`** command. This command forcibly brings up a fiber port, even when no fiber links or transceiver modules are present for the fiber port. When one fiber link is present and up, the fiber port can forward packets over the link unidirectionally.

Figure 2 Forcibly bring up a fiber port



Restrictions and guidelines

- Copper ports do not support this feature.
- The **`port up-mode`**, **`shutdown`**, and **`loopback`** commands are mutually exclusive.
- A fiber port does not support this feature if the port is shut down by a protocol or by using the **`shutdown`** command.
- A fiber port forcibly brought up stays physically up whether or not a transceiver module or a fiber link is present for the port.

- A GE fiber port forcibly brought up cannot correctly forward traffic if it is installed with a fiber-to-copper converter, 100/1000-Mbps transceiver module, or 100-Mbps transceiver module. To solve the problem, use the **undo port up-mode** command on the fiber port.

Procedure

1. Enter system view.

   system-view

2. Enter Ethernet interface view.

   interface *interface-type interface-number*

3. Forcibly bring up the fiber port.

   port up-mode

   By default, a fiber port is not forcibly brought up, and the physical state of a fiber port depends on the physical state of the fibers.

Configuring interface alarm functions

About this task

With the interface alarm functions enabled, when the number of error packets on an interface in normal state within the specified interval exceeds the upper threshold, the interface generates an upper threshold exceeding alarm and enters the alarm state. When the number of error packets on an interface in the alarm state within the specified interval drops below the lower threshold, the interface generates a recovery alarm and restores to the normal state.

When a port is shut down automatically by a feature, the device starts a port status detection timer. When the timer expires, the device restores the port status to its actual physical status automatically.

Restrictions and guidelines

You can configure the interface alarm parameters in system view and interface view.

- The configuration in system view takes effect on all interfaces of the specified slot. The configuration in interface view takes effect only on the current interface.
- For an interface, the configuration in interface view takes priority, and the configuration in system view is used only when no configuration is made in interface view.

An interface that is shut down because of error packet alarms cannot automatically recover. To bring up the interface, execute the **undo shutdown** command on the interface.

If you change the timer setting during port detection, the device compares the new setting (T1) with the time that elapsed since the port was shut down (T).

- If T < T1, the port will be brought up after T1 – T seconds.
- If T ≥ T1, the port is brought up immediately.

For example, the timer setting is 30 seconds. If you change it to 10 seconds 2 seconds after the port is shut down, the port will come up 8 seconds later. If you change the timer setting to 2 seconds 10 seconds after the port is shut down, the port comes up immediately.

Configuring CRC error packet parameters

1. Enter system view.

   system-view

2. Configure global CRC error packet alarm parameters.

   ifmonitor crc-error slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**down-auto-recovery** | **shutdown**]

By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for CRC error packets.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure CRC error packet alarm parameters for the interface.

   port ifmonitor crc-error [**ratio**] **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**down-auto-recovery** | **shutdown**]

   By default, an interface uses the global CRC error packet alarm parameters.

5. Return to system view.

   quit

6. Set the port status detection timer for the ports that are shut down by the CRC error packet alarm feature.

   shutdown-interval [**crc-error**] *interval*

   By default, the timer is 0 seconds.

Configuring input error packet alarm parameters

1. Enter system view.

   system-view

2. Configure global input error packet alarm parameters.

   ifmonitor input-error slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for input error packets.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure input error packet alarm parameters for the interface.

   port ifmonitor input-error high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the global input error packet alarm parameters.

Configuring output error packet alarm parameters

1. Enter system view.

   system-view

2. Configure global output error packet alarm parameters.

   ifmonitor output-error slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for output error packets.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure output error packet alarm parameters.

   port ifmonitor output-error high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the global output error packet alarm parameters.

Configuring input bandwidth usage alarm parameters

1. Enter system view.

system-view

2. Configure global input bandwidth usage alarm parameters.

   ifmonitor input-usage slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value*

   By default, the upper threshold is 90 and the lower threshold is 80 for input bandwidth usage alarms.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure input bandwidth usage alarm parameters.

   port ifmonitor input-usage high-threshold *high-value* **low-threshold** *low-value*

   By default, an interface uses the global input bandwidth usage alarm parameters.

Configuring output bandwidth usage alarm parameters

1. Enter system view.

   system-view

2. Configure global output bandwidth usage alarm parameters.

   ifmonitor output-usage slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value*

   By default, the upper threshold is 90 and the lower threshold is 80 for output bandwidth usage alarms.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure output bandwidth usage alarm parameters.

   port ifmonitor output-usage high-threshold *high-value* **low-threshold** *low-value*

   By default, an interface uses the global output bandwidth usage alarm parameters.

Configuring received pause frame alarm parameters

1. Enter system view.

   system-view

2. Configure global received pause frame alarm parameters.

   ifmonitor rx-pause slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval*

   By default, the upper threshold is 500, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for received pause frames.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure received pause frame alarm parameters.

   port ifmonitor rx-pause high-threshold *high-value* **low-threshold** *low-value* **interval** *interval*

   By default, an interface uses the global received pause frame alarm parameters.

Configuring sent pause frame alarm parameters

1. Enter system view.

   system-view

2. Configure global sent pause frame alarm parameters.

ifmonitor tx-pause slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

By default, the upper threshold is 500, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for sent pause frames.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure sent pause frame alarm parameters.

   port ifmonitor tx-pause high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the global sent pause frame alarm parameters.

Configuring SDH error packet alarm parameters

1. Enter system view.

   system-view

2. Configure global SDH error packet alarm parameters.

   ifmonitor sdh-error slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for SDH error packets.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure SDH error packet alarm parameters for the interface.

   port ifmonitor sdh-error high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the global SDH error packet alarm parameters.

Configuring SDH-B1 error packet alarm parameters

1. Enter system view.

   system-view

2. Configure global SDH-B1 error packet alarm parameters.

   ifmonitor sdh-b1-error slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for SDH-B1 error packets.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure SDH-B1 error packet alarm parameters for the interface.

   port ifmonitor sdh-b1-error high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the global SDH-B1 error packet alarm parameters.

Configuring SDH-B2 error packet alarm parameters

1. Enter system view.

   system-view

2. Configure global SDH-B2 error packet alarm parameters.

   ifmonitor sdh-b2-error slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds for SDH-B2 error packets.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure SDH-B2 error packet alarm parameters for the interface.

   port ifmonitor sdh-b2-error high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the global SDH-B2 error packet alarm parameters.

Configuring packet-drop alarm parameters for the ingress buffer

1. Enter system view.

   system-view

2. Configure packet-drop alarm parameters for the ingress buffer globally.

   ifmonitor input-buffer-drop slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure packet-drop alarm parameters for the ingress buffer.

   port ifmonitor input-buffer-drop high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the packet-drop alarm parameters globally configured for the ingress buffer.

Configuring packet-drop alarm parameters for the egress buffer

1. Enter system view.

   system-view

2. Configure packet-drop alarm parameters for the egress buffer globally.

   ifmonitor output-buffer-drop slot *slot-number* **high-threshold** *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, the upper threshold is 1000, the lower threshold is 100, and the statistics collection and comparison interval is 10 seconds.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Configure packet-drop alarm parameters for the egress buffer.

   port ifmonitor output-buffer-drop high-threshold *high-value* **low-threshold** *low-value* **interval** *interval* [**shutdown**]

   By default, an interface uses the packet-drop alarm parameters globally configured for the egress buffer.

Configuring SNMP notifications for cards

About this task

With this feature configured for a card, the card triggers an SNMP notification when the number of dropped packets received within the specified time exceeds the upper threshold. When the number of dropped packets received within the specified time drops below the lower threshold, the card triggers a recovery notification and restores to normal state.

Enabling SNMP notifications for cards

1. Enter system view.

   system-view

2. Enable SNMP notifications for cards.

   snmp-agent trap enable slot-monitor [**packet-drop**] *

   By default, SNMP notifications are enabled for cards.

Configuring packet-drop alarm parameters for all cards

1. Enter system view.

   system-view

2. Configure packet-drop alarm parameters for all cards.

   slot-monitor packet-drop high-threshold *high-threshold-value* **low-threshold** *low-threshold-value* **interval** *interval* [**shutdown** | **reboot**]

   By default, the upper threshold is 1000000, the lower threshold is 100, and the statistics collection and comparison interval is 30 seconds.

Configuring packet-drop alarm parameters for a card

1. Enter system view.

   system-view

2. Configure packet-drop alarm parameters for a card.

   slot-monitor packet-drop slot *slot-number* **high-threshold** *high-threshold-value* **low-threshold** *low-threshold-value* **interval** *interval* [**shutdown** | **reboot**]

   By default, the upper threshold is 1000000, the lower threshold is 100, and the statistics collection and comparison interval is 30 seconds.

Enabling SNMP notifications for the ETH module

About this task

To report critical ETH events to an SNMP NMS, enable SNMP notifications for the ETH module. For ETH SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for the ETH module.

   snmp-agent trap enable eth

   By default, SNMP notifications are disabled for the ETH module.

Shutting down all physical interfaces

About this task

This feature shuts down all physical interfaces except the management Ethernet interfaces on the device. Physical interfaces shut down by using this command are in ADM state.

Restrictions and guidelines

To shut down all physical interfaces or the specified interface, execute the **shutdown all-physical-interfaces** command in system view or execute the **shutdown** command in interface view.

To bring up interfaces shut down by using the **shutdown all-physical-interfaces** command, execute its **undo** form in system view. To bring up an interface shut down by using the **shutdown** command, execute its **undo** form in interface view.

If you execute the **shutdown all-physical-interfaces** command with the **include irf-physical-interface** keyword multiple times, this command shuts down all physical interfaces except the management Ethernet interfaces on the device.

Procedure

1. Enter system view.
   system-view
2. Shut down all physical interfaces except management Ethernet interfaces.
   shutdown all-physical-interfaces [**include irf-physical-interface**]
   By default, physical interfaces are up.

Configuring the port speed mode for expansion slots

Hardware and feature compatibility

This feature is supported only on the S5590-HI switch series.

Restrictions and guidelines

For this feature to take effect, reboot the device after you configure the feature.

To set up an IRF fabric successfully, make sure the same port speed mode for expansion slots is configured on the member devices.

When the port speed mode is 2, only the first 44 GE interfaces are available for the S5590-48T4XC-HI and S5590-48S4XC-HI switches.

In different port speed modes, different expansion slots support different expansion cards. For more information, see Table 1 and Table 2.

Table 1 Expansion slot 1

Expansion card	mode 0	mode 1	mode 2
LSPM6FWD	Yes	No	Yes
LSWM2EC	No	Yes	Yes
LSWM2QP2P	Yes	No	Yes
LSWM2SP8P	No	No	Yes
LSWM2ZQP2P	No	No	No
LSWM2ZSP8P	No	No	No

Table 2 Expansion slot 2

Expansion card	mode 0	mode 1	mode 2
LSPM6FWD	Yes		
LSWM2EC	Yes		
LSWM2QP2P	Yes		
LSWM2SP8P	Yes		
LSWM2ZQP2P	Yes		
LSWM2ZSP8P	Yes		

Procedure

1. Enter system view.
   `system-view`
2. Configure the port speed mode for expansion slots.
   `port-speed-mode mode { 0 | 1 | 2 }`
   By default, the port speed mode is 0 for expansion slots.

Configuring the port speed mode for the device

Hardware and feature compatibility

This feature is available only on the S5590-32UN4X4Y2CC-EI and S5590-48UN4Y2CC-EI switches.

Restrictions and guidelines

For this feature to take effect, reboot the device after you configure the feature.

For the S5590-32UN4X4Y2CC-EI switch:

- When you execute the **`port-speed-mode mode 0`** command, among ports 37 through 42 on the S5590-32UN4X4Y2CC-EI front panel, only ports 37 through 40 and port 42 are available.
- When you execute the **`port-speed-mode mode 1`** command, among ports 37 through 42 on the S5590-32UN4X4Y2CC-EI front panel, only QSFP28 ports 41 and 42 are available and cannot be split.

For the S5590-48UN4Y2CC-EI switch:

- When you execute the **`port-speed-mode mode 0`** command, among ports 49 through 54 on the S5590-48UN4Y2CC-EI front panel, only ports 49 through 52 and port 54 are available.
- When you execute the **`port-speed-mode mode 1`** command, among ports 49 through 54 on the S5590-48UN4Y2CC-EI front panel, only QSFP28 ports 53 and 54 are available and cannot be split.

Procedure

1. Enter system view.
   `system-view`
2. Configure the port speed mode for the device.
   `port-speed-mode mode { 0 | 1 }`
   By default, the port speed mode is 0 for the device.

Restoring the default settings for an interface

Restrictions and guidelines

CAUTION:

This feature might interrupt ongoing network services. Make sure you are fully aware of the impacts of this feature when you use it in a live network.

This feature might fail to restore the default settings for some commands because of command dependencies or system restrictions. You can use the **display this** command in interface view to check for these commands and perform their **undo** forms or follow the command reference to restore their default settings. If your restoration attempt still fails, follow the error message to resolve the problem.

Procedure

1. Enter system view.

   system-view
2. Enter Ethernet interface view or Ethernet subinterface view.

   interface *interface-type* { *interface-number* | *interface-number.subnumber* }
3. Restore the default settings for the interface.

   default

Enabling packet latency detection

About this task

With this feature enabled on an interface, the device continuously monitors forwarding latency of packets transmitted through the interface within the device and assigns latency levels to the packets. You can determine the quality of packet forwarding accordingly.

Restrictions and guidelines

A higher level threshold cannot be lower than a lower level threshold.

Procedure

1. Enter system view.

   system-view
2. (Optional.) Configure the packet latency level threshold.

   ifmonitor latency { **level0** *level0-value* | **level1** *level1-value* | **level2** *level2-number* | **level3** *level3-number* | **level4** *level4-number* | **level5** *level5-number* | **level6** *level6-number* | **level7** *level7-number* }*

   The default for this command depends on the device model.
3. Enter Ethernet interface view.

   interface *interface-type interface-number*
4. Enable packet latency detection.

   port ifmonitor buffer flow-information enable

   By default, packet latency detection is not enabled.

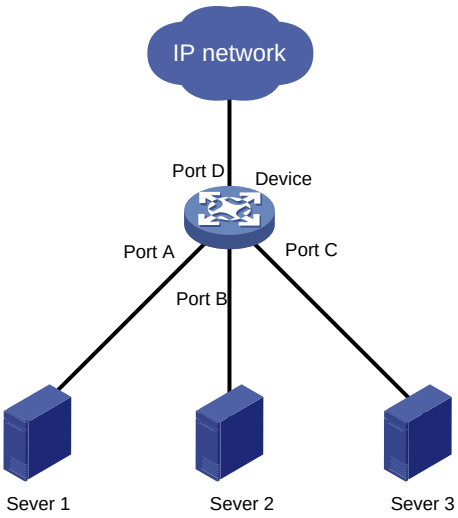
Configuring a Layer 2 Ethernet interface

Setting speed options for autonegotiation on an Ethernet interface

About this task

By default, speed autonegotiation enables an Ethernet interface to negotiate with its peer for the highest speed that both ends support. You can narrow down the speed option list for negotiation.

Figure 3 Speed autonegotiation application scenario



As shown in Figure 3:

- All interfaces on the device are operating in speed autonegotiation mode, with the highest speed of 1000 Mbps.
- Port D provides access to the Internet for the servers.

If the transmission rate of each server in the server cluster is 1000 Mbps, their total transmission rate exceeds the capability of Port D.

To avoid congestion on Port D, configure 100 Mbps as the only option available for speed negotiation on interfaces Port A, Port B, and Port C. As a result, the transmission rate on each interface connected to a server is limited to 100 Mbps.

Restrictions and guidelines

The **speed** and **speed auto** commands supersede each other, and whichever is configured last takes effect.

Procedure

1. Enter system view.
   system-view
2. Enter 100-Mbps or 1000-Mbps Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
3. Set speed options for autonegotiation.
   speed auto { **10** | **100** | **1000** } *
   By default, no speed options are set for autonegotiation.

Setting the MDIX mode of an Ethernet interface

IMPORTANT:

Fiber ports do not support the MDIX mode setting.

About this task

A physical Ethernet interface has eight pins, each of which plays a dedicated role. For example, pins 1 and 2 transmit signals, and pins 3 and 6 receive signals. You can use both crossover and straight-through Ethernet cables to connect copper Ethernet interfaces. To accommodate these types of cables, a copper Ethernet interface can operate in one of the following Medium Dependent Interface-Crossover (MDIX) modes:

- **MDIX mode**—Pins 1 and 2 are receive pins and pins 3 and 6 are transmit pins.
- **MDI mode**—Pins 1 and 2 are transmit pins and pins 3 and 6 are receive pins.
- **AutoMDIX mode**—The interface negotiates pin roles with its peer.

NOTE:

This feature does not take effect on pins 4, 5, 7, and 8 of physical Ethernet interfaces.

- Pins 4, 5, 7, and 8 of interfaces operating at 10 Mbps or 100 Mbps do not receive or transmit signals.
- Pins 4, 5, 7, and 8 of interfaces operating at 1000 Mbps or higher rates receive and transmit signals.

Restrictions and guidelines

To enable a copper Ethernet interface to communicate with its peer, set the MDIX mode of the interface by following these guidelines:

- Typically, set the MDIX mode of the interface to AutoMDIX. Set the MDIX mode of the interface to MDI or MDIX only when the device cannot determine the cable type.
- When a straight-through cable is used, configure the interface to operate in an MDIX mode different than its peer.
- When a crossover cable is used, perform one of the following tasks:
  - Configure the interface to operate in the same MDIX mode as its peer.
  - Configure either end to operate in AutoMDIX mode.

Procedure

1. Enter system view.

   system-view
2. Enter Ethernet interface view.

   interface *interface-type interface-number*
3. Set the MDIX mode of the Ethernet interface.

   mdix-mode { **automdix** | **mdi** | **mdix** }

   By default, a copper Ethernet interface operates in auto mode to negotiate pin roles with its peer.

Configuring storm control on an Ethernet interface

About this task

Storm control compares broadcast, multicast traffic regularly with their respective traffic thresholds on an Ethernet interface. For each type of traffic, storm control provides a lower threshold and an upper threshold.

Depending on your configuration, when a particular type of traffic exceeds its upper threshold, the interface performs either of the following operations:

- **Blocks this type of traffic and forwards other types of traffic**—Even though the interface does not forward the blocked traffic, it still counts the traffic. When the blocked traffic drops below the lower threshold, the interface begins to forward the traffic.
- **Goes down automatically**—The interface goes down automatically and stops forwarding any traffic. When the blocked traffic drops below the lower threshold, the interface does not automatically come up. To bring up the interface, use the **undo shutdown** command or disable the storm control feature.

You can configure an Ethernet interface to output threshold event traps and log messages when monitored traffic meets one of the following conditions:

- Exceeds the upper threshold.
- Drops below the lower threshold.

Both storm suppression and storm control can suppress storms on an interface. Storm suppression uses the chip to suppress traffic. Storm suppression has less impact on the device performance than storm control, which uses software to suppress traffic. For more information about storm suppression, see "Configuring broadcast/multicast/unknown unicast storm suppression."

Storm control uses a complete polling cycle to collect traffic data, and analyzes the data in the next cycle. An interface takes one to two polling intervals to take a storm control action.

Restrictions and guidelines

For the traffic suppression result to be determined, do not configure storm control together with storm suppression for the same type of traffic.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Set the statistics polling interval of the storm control module.

   storm-constrain interval *interval*

   The default setting is 10 seconds.

   For network stability, use the default or set a longer statistics polling interval.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Enable storm control, and set the lower and upper thresholds for broadcast, multicast, or unknown unicast traffic.

   storm-constrain { **broadcast** | **multicast** } { **pps** | **kbps** | **ratio** } *upperlimit lowerlimit*

   By default, storm control is disabled.

5. Set the control action to take when monitored traffic exceeds the upper threshold.

   storm-constrain control { **block** | **shutdown** }

   By default, storm control is disabled.

6. Enable the Ethernet interface to output log messages when it detects storm control threshold events.

   storm-constrain enable log

   By default, the Ethernet interface outputs log messages when monitored traffic exceeds the upper threshold or drops below the lower threshold from a value above the upper threshold.

7. Enable the Ethernet interface to send storm control threshold event traps.

   storm-constrain enable trap

   By default, the Ethernet interface sends traps when monitored traffic exceeds the upper threshold or drops below the lower threshold from the upper threshold from a value above the upper threshold.

Testing the cable connection of an Ethernet interface

IMPORTANT:

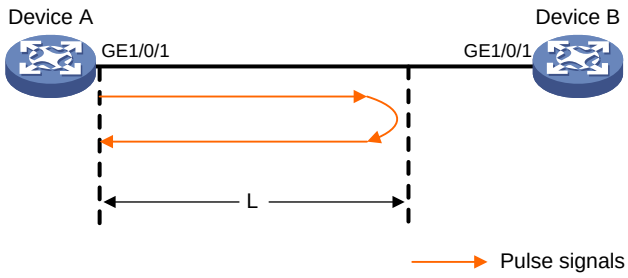
If the link of an Ethernet interface is up, testing its cable connection will cause the link to go down and then come up.

About this task

The virtual cable test (VCT) technology uses time domain reflectometry (TDR) to detect the cable status. When pulse signals are transmitted in a cable, some energy of the signals is reflected at the end or a failure point on the cable. This phenomenon is called TDR. The VCT algorithm measures the time spent in transmitting pulse signals over a cable, reaching the failure point, and returning the pulse signals. The measured time is converted into the distance.

As shown in Figure 4, GE 1/0/1 on Device A is connected to GE 1/0/1 on Device B through a cable. A failure point exists on the cable. After VCT is configured on GE 1/0/1 on Device A, the system will generate pulse signals. Some energy of the signals is reflected at the failure point on the cable. Suppose the distance from Device A to the failure point is L, the interval between sending signals and receiving reflected signals is T, and the signal transmission speed on the cable is V. Then, the distance between GE 1/0/1 on Device A and the failure point can be calculated as follows: L=(V×T)/2.

Figure 4 Schematic diagram for VCT



VCT detects the network cable failure type and identifies the failure point, facilitating failure location on network cables.

Restrictions and guidelines

Fiber ports do not support this feature.

Procedure

1. Enter system view.

   system-view

2. Enter Ethernet interface view.

interface *interface-type interface-number*

3. Perform a test for the cable connected to the Ethernet interface.
   - **virtual-cable-test**

Enabling bridging on an Ethernet interface

About this task

By default, the device drops packets whose outgoing interface and incoming interface are the same.

To enable the device to forward such packets rather than drop them, enable the bridging feature in Ethernet interface view.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Enable bridging on the Ethernet interface.
   port bridge enable
   By default, bridging is disabled on an Ethernet interface.

Configuring a Layer 3 Ethernet interface or subinterface

Setting the MTU for an Ethernet interface or subinterface

About this task

The maximum transmission unit (MTU) determines the maximum number of bytes in a single IP packet that can be sent. The length of an IP packet refers to the number of bytes starting from the IP header to the payload.

When the IP layer receives an IP data packet to be sent, the IP layer determines the local destination interface of the packet and obtains the MTU of the interface. The IP layer compares the MTU with the length of the data packet to be sent. If the length is greater than the MTU, the IP layer fragments the packet. The length of a fragment can be smaller than or equal to the MTU to ensure that big packets are not lost on the network.

As a best practice, use the default MTU. When the packet length or the packet receiver changes, you can adjust the MTU as needed. When configuring the MTU, follow these restrictions and guidelines:

- If the configured MTU is small but the packet length is large, the following events might occur:
  - Packets will be dropped when they are forwarded by hardware.
  - Packets will be fragmented into too many fragments when packets are forwarded through the CPUs, which affects normal data transmission.
- If the configured MTU is too large, the MTU will exceed the receiving capabilities of the receiver or a device along the transmission path. As a result, packets will be fragmented or even dropped, which increases the network transmission load and affects data transmission.

Restrictions and guidelines

The MTU of an Ethernet interface or subinterface affects the fragmentation and reassembly of IP packets on the interface. Typically, you do not need to modify the MTU of an interface.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type* { *interface-number* | *interface-number.subnumber* }
3. Set the MTU for the interface.
   mtu *size*
   The default setting is 1500 bytes.

Setting the MAC address of an Ethernet interface or subinterface

About this task

In a network, when the Layer 3 Ethernet interfaces or subinterfaces of different devices have the same MAC address, the devices might fail to communicate correctly. To eliminate the MAC address conflicts, use the **mac-address** command to modify the MAC addresses of Layer 3 Ethernet interfaces or subinterfaces.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type* { *interface-number* | *interface-number.subnumber* }
3. Set the interface MAC address.
   mac-address *mac-address*
   By default, no MAC address is set for an Ethernet interface.
   As a best practice, do not set a MAC address in the VRRP-reserved MAC address range for a Layer 3 Ethernet subinterface.

Display and maintenance commands for Ethernet interfaces

IMPORTANT:
The **display port-speed-mode** command is supported only on the S5590-HI switch series.

For more information about the **display diagnostic-information** command, see device management in *Fundamentals Command Reference*.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display interface traffic statistics.	**display counters** { **inbound** | **outbound** } **interface** [*interface-type* [*interface-number*]]
Display traffic rate statistics of interfaces in up state over the last statistics polling interval.	**display counters rate** { **inbound** | **outbound** } **interface** [*interface-type* [*interface-number*]]
Display the operating information for the IFMGR module.	**display diagnostic-information ifmgr** [**key-info**] [*filename*]
Display the Ethernet module statistics.	**display ethernet statistics slot** *slot-number*
Display the operational and status information of the specified interfaces.	**display interface** [*interface-type* [*interface-number* | *interface-number.subnumber*]] [**brief** [**description** | **down**]]
Display the status and packet statistics of interfaces.	**display interface link-info** [**main**]
Display the operational and status information of interfaces except subinterfaces.	**display interface** [*interface-type*] [**brief** [**description** | **down**]] **main**
Display information about link flapping protection on interfaces.	**display link-flap protection** [**interface** *interface-type* [*interface-number*]]
Display the physical link state change statistics of interfaces.	**display link-state-change statistics interface** [*interface-type* [*interface-number*]]
Display the state of the port functionality license.	**display port license state** [**slot** *slot-number*]
Display information about storm control on the specified interfaces.	**display storm-constrain** [**broadcast** | **multicast** | **unicast**] [**interface** *interface-type interface-number*]
Clear cards statistics and interface statistics on cards.	**reset counters slot** [*slot-number*]
Clear interface statistics.	**reset counters interface** [*interface-type* [*interface-number* | *interface-number.subnumber*]]
Clear the interface statistics on MIB.	**reset counters mib interface** [*interface-type* [*interface-number* | *interface-number.subnumber*]]
Clear the Ethernet module statistics.	**reset ethernet statistics** [**slot** *slot-number*]
Clear the physical link state change statistics of interfaces.	**reset link-state-change statistics interface** [*interface-type* [*interface-number*]]
Clear the statistics of dropped packets on the specified interfaces.	**reset packet-drop interface** [*interface-type* [*interface-number*]]

Contents

Configuring Ethernet link aggregation 1

About Ethernet link aggregation 1
Ethernet link aggregation application scenario 1
Aggregate interface, aggregation group, and member port 1
Operational key 2
Configuration types 2
Link aggregation modes 2
How static link aggregation works 3
Dynamic link aggregation 4
How dynamic link aggregation works 7
Edge aggregate interface 10
Load sharing modes for link aggregation groups 10
S-MLAG 10
Restrictions and guidelines: Mixed use of manual and automatic link aggregation configuration 11
Ethernet link aggregation tasks at a glance 11
Configuring the system ID 12
Configuring a manual link aggregation 13
Restrictions and guidelines for aggregation group configuration 13
Configuring a Layer 2 aggregation group 14
Configuring a Layer 3 aggregation group 15
Configuring automatic link aggregation 16
Configuring S-MLAG 17
Configuring an aggregate interface 18
Configuring the description of an aggregate interface 18
Setting the MAC address for an aggregate interface 19
Setting the MTU for a Layer 3 aggregate interface 19
Setting the expected bandwidth for an aggregate interface 19
Configuring an edge aggregate interface 20
Configuring physical state change suppression on an aggregate interface 20
Shutting down an aggregate interface 21
Enabling interframe gap and preamble statistics in the traffic statistics 22
Restoring the default settings for an aggregate interface 22
Enabling transparent LACPDU transmission 23
Setting the minimum and maximum numbers of Selected ports for an aggregation group 24
Disabling the default action of selecting a Selected port for dynamic aggregation groups that have not received LACPDUs 25
Configuring a dynamic aggregation group to use port speed as the prioritized criterion for reference port selection 26
Ignoring port speed in setting the aggregation states of member ports 26
Enabling port selection preemption on an aggregate interface 27
Configuring load sharing for link aggregation groups 28
Setting static load sharing modes for link aggregation groups 28
Enabling local-first load sharing for link aggregation 29
Configuring link aggregation load sharing algorithm settings 30
Specifying link aggregation management VLANs and link aggregation management port 31
Enabling link-aggregation traffic redirection 31
About link-aggregation traffic redirection 31
Restrictions and guidelines for link-aggregation traffic redirection 31
Enabling link-aggregation traffic redirection globally 32
Enabling link-aggregation traffic redirection for an aggregation group 32
Isolating aggregate interfaces on the device 32
Enabling BFD for an aggregation group 33
Display and maintenance commands for Ethernet link aggregation 34
Ethernet link aggregation configuration examples 35
Example: Configuring a Layer 2 static aggregation group 35
Example: Configuring a Layer 2 dynamic aggregation group 37
Example: Configuring a Layer 2 edge aggregate interface 39

Example: Configuring a Layer 3 static aggregation group ····41
Example: Configuring a Layer 3 dynamic aggregation group ····42
Example: Configuring S-MLAG ····44

Configuring Ethernet link aggregation

About Ethernet link aggregation

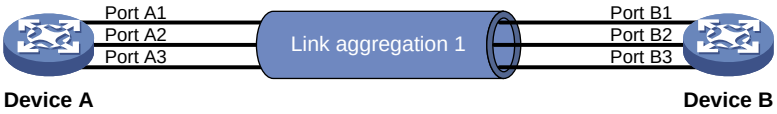
Ethernet link aggregation bundles multiple physical Ethernet links into one logical link (called an aggregate link). Link aggregation provides the following benefits:

- Increased bandwidth beyond the limits of a single individual link. In an aggregate link, traffic is distributed across the member ports.
- Improved link reliability. The member ports dynamically back up one another. When a member port fails, its traffic is automatically switched to other member ports.

Ethernet link aggregation application scenario

As shown in Figure 1, Device A and Device B are connected by three physical Ethernet links. These physical Ethernet links are combined into an aggregate link called link aggregation 1. The bandwidth of this aggregate link can reach up to the total bandwidth of the three physical Ethernet links. At the same time, the three Ethernet links back up one another. When a physical Ethernet link fails, the traffic transmitted on the failed link is switched to the other two links.

Figure 1 Ethernet link aggregation diagram



Aggregate interface, aggregation group, and member port

Each link aggregation is represented by a logical aggregate interface. Each aggregate interface has an automatically created aggregation group, which contains member ports to be used for aggregation. The type and number of an aggregation group are the same as its aggregate interface.

Supported aggregate interface types

An aggregate interface can be one of the following types:

- **Layer 2**—A Layer 2 aggregate interface is created manually. The member ports in a Layer 2 aggregation group can only be Layer 2 Ethernet interfaces.
- **Layer 3**—A Layer 3 aggregate interface is created manually. The member ports in its Layer 3 aggregation group can only be Layer 3 Ethernet interfaces.

  On a Layer 3 aggregate interface, you can create subinterfaces. A Layer 3 aggregate subinterface processes traffic only for the VLAN numbered with the same ID as the subinterface number.

The port rate of an aggregate interface equals the total rate of its Selected member ports. Its duplex mode is the same as that of the Selected member ports. For more information about Selected member ports, see "Aggregation states of member ports in an aggregation group."

Aggregation states of member ports in an aggregation group

A member port in an aggregation group can be in any of the following aggregation states:

- **Selected**—A Selected port can forward traffic.
- **Unselected**—An Unselected port cannot forward traffic.

- **Individual**—An Individual port can forward traffic as a normal physical port. This state is peculiar to the member ports of edge aggregate interfaces. A Selected or Unselected member port of an edge aggregate interface is placed in Individual state if the LACP timeout timer expires because it has not received LACPDUs.

  For more information about edge aggregate interfaces, see "Edge aggregate interface."

Operational key

When aggregating ports, the system automatically assigns each port an operational key based on port information, such as port rate and duplex mode. Any change to this information triggers a recalculation of the operational key.

In an aggregation group, all Selected ports have the same operational key.

Configuration types

Port configuration includes the attribute configuration and protocol configuration. Attribute configuration affects the aggregation state of the port but the protocol configuration does not.

Attribute configuration

To become a Selected port, a member port must have the same attribute configuration as the aggregate interface. Table 1 describes the attribute configuration.

Table 1 Attribute configuration

Feature	Attribute configuration
Port isolation	Membership of the port in an isolation group. Isolation group number.
QinQ	QinQ status (enabled/disabled)and TPID for VLAN tags. For information about QinQ, see "Configuring QinQ."
VLAN mapping	VLAN mapping configured on the port. For more information about VLAN mapping, see "Configuring VLAN mapping."
VLAN	VLAN attribute settings: • Permitted VLAN IDs. • PVID. • Link type (trunk, hybrid, or access). • VLAN tagging mode. For information about VLANs, see "Configuring VLANs."

Protocol configuration

Protocol configuration of a member port does not affect the aggregation state of the member port. MAC address learning and spanning tree settings are examples of the protocol configuration.

Link aggregation modes

An aggregation group operates in one of the following modes:

- **Static**—Static aggregation is stable. An aggregation group in static mode is called a static aggregation group. The aggregation states of the member ports in a static aggregation group are not affected by the peer ports.
- **Dynamic**—An aggregation group in dynamic mode is called a dynamic aggregation group. Dynamic aggregation is implemented through IEEE 802.3ad Link Aggregation Control Protocol

(LACP). The local system and the peer system automatically maintain the aggregation states of the member ports. Dynamic link aggregation reduces the administrators' workload.

How static link aggregation works

Reference port selection process

When setting the aggregation states of the ports in an aggregation group, the system automatically chooses a member port as the reference port. A Selected port must have the same operational key and attribute configurations as the reference port.

The system chooses a reference port from the member ports in up state.

The candidate reference ports are organized into different priority levels following these rules:

1. In descending order of port priority.
2. Full duplex.
3. In descending order of speed.
4. Half duplex.
5. In descending order of speed.

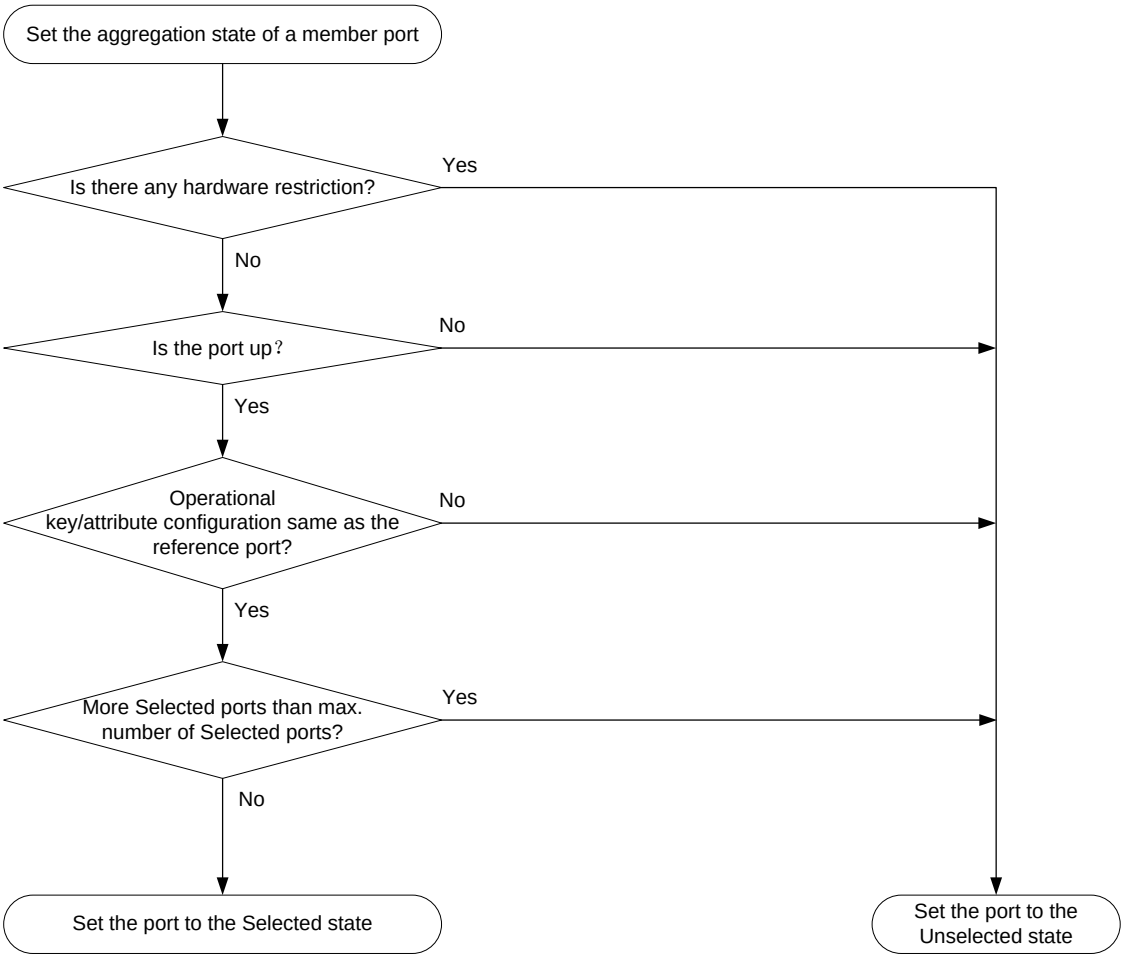
From the candidate ports with the same attribute configurations as the aggregate interface, the one with the highest priority level is chosen as the reference port.

- If multiple ports have the same priority level, the port that has been Selected (if any) is chosen. If multiple ports with the same priority level have been Selected, the one with the smallest port number is chosen.
- If multiple ports have the same priority level and none of them has been Selected, the port with the smallest port number is chosen.

Setting the aggregation state of each member port

After the reference port is chosen, the system sets the aggregation state of each member port in the static aggregation group.

Figure 2 Setting the aggregation state of a member port in a static aggregation group



After the limit on Selected ports is reached, the aggregation state of a new member port varies by following conditions:

- The port is placed in Unselected state if the port and the Selected ports have the same port priority. This mechanism prevents traffic interruption on the existing Selected ports. A device reboot can cause the device to recalculate the aggregation states of member ports.
- The port is placed in Selected state when the following conditions are met:
  - The port and the Selected ports have different port priorities, and the port has a higher port priority than a minimum of one Selected port.
  - The port has the same attribute configurations as the aggregate interface.

Any operational key or attribute configuration change might affect the aggregation states of link aggregation member ports.

Dynamic link aggregation

About LACP

Dynamic aggregation is implemented through IEEE 802.3ad Link Aggregation Control Protocol (LACP).

LACP uses LACPDUs to exchange aggregation information between LACP-enabled devices. Each member port in a dynamic aggregation group can exchange information with its peer. When a member port receives an LACPDU, it compares the received information with information received

on the other member ports. In this way, the two systems reach an agreement on which ports are placed in Selected state.

LACP functions

LACP offers basic LACP functions and extended LACP functions, as described in Table 2.

Table 2 Basic and extended LACP functions

Category	Description
Basic LACP functions	Implemented through the basic LACPDU fields, including the LACP system priority, system MAC address, port priority, port number, and operational key.
Extended LACP functions	Implemented by extending the LACPDU with new TLV fields. Extended LACP can implement LACP MAD for the IRF feature. For more information about IRF and the LACP MAD mechanism, see *Virtual Technologies Configuration Guide*. The device can participate in LACP MAD as either an IRF member device or an intermediate device.

LACP operating modes

LACP can operate in active or passive mode.

When LACP is operating in passive mode on a local member port and its peer port, both ports cannot send LACPDUs. When LACP is operating in active mode on either end of a link, both ports can send LACPDUs.

LACP priorities

LACP priorities include LACP system priority and port priority, as described in Table 3. The smaller the priority value, the higher the priority.

Table 3 LACP priorities

Type	Description
LACP system priority	Used by two peer devices (or systems) to determine which one is superior in link aggregation. In dynamic link aggregation, the system that has higher LACP system priority sets the Selected state of member ports on its side. The system that has lower priority sets the aggregation state of local member ports the same as their respective peer ports.
Port priority	Determines the likelihood of a member port to be a Selected port on a system. A port with a higher port priority is more likely to become Selected.

LACP timeout interval

The LACP timeout interval specifies how long a member port waits to receive LACPDUs from the peer port, and it can be either short or long.

- **Long timeout interval**—By default, the LACP timeout interval for a port is long (90 seconds). If a local member port does not receive an LACPDU from the peer within the LACP timeout interval plus 3 seconds (93 seconds), the system considers the peer member port down.
- **Short timeout interval**—Use the **`lacp period short`** command to set the LACP timeout interval to short for a port.
  - **Do not specify the timeout parameter**—In this method, the short timeout interval is set to a fixed value of 3 seconds. In this case, if a local member port does not receive LACPDUs from its peer member port within 6 seconds (3 seconds of LACP timeout interval + 3 seconds), the peer member port is considered to have failed.
  - **Specify the timeout parameter**—In this method, you can configure a short timeout interval in the range of 3 to 90 seconds. If a local member port does not receive LACPDUs from its

peer member port within the configured LACP short timeout interval, the peer member port is considered to have failed.

The LACP timeout interval also determines the LACPDU sending rate of the peer. LACP timeout intervals include the following types:

- **Short timeout interval**—The peer sends one LACPDU per second.
- **Long timeout interval**—The peer sends one LACPDU every 30 seconds.

Methods to assign interfaces to a dynamic link aggregation group

You can use one of the following methods to assign interfaces to a dynamic link aggregation group:

- **Manual assignment**—Manually assign interfaces to the dynamic link aggregation group.
- **Automatic assignment**—Enable automatic assignment on interfaces to have them automatically join a dynamic link aggregation group depending on the peer information in the received LACPDUs.

NOTE:

When you use automatic assignment on one end, you must use manual assignment on the other end.

Alternatively, you can use automatic link aggregation for two devices to automatically create a dynamic link aggregation between them by using LLDP.

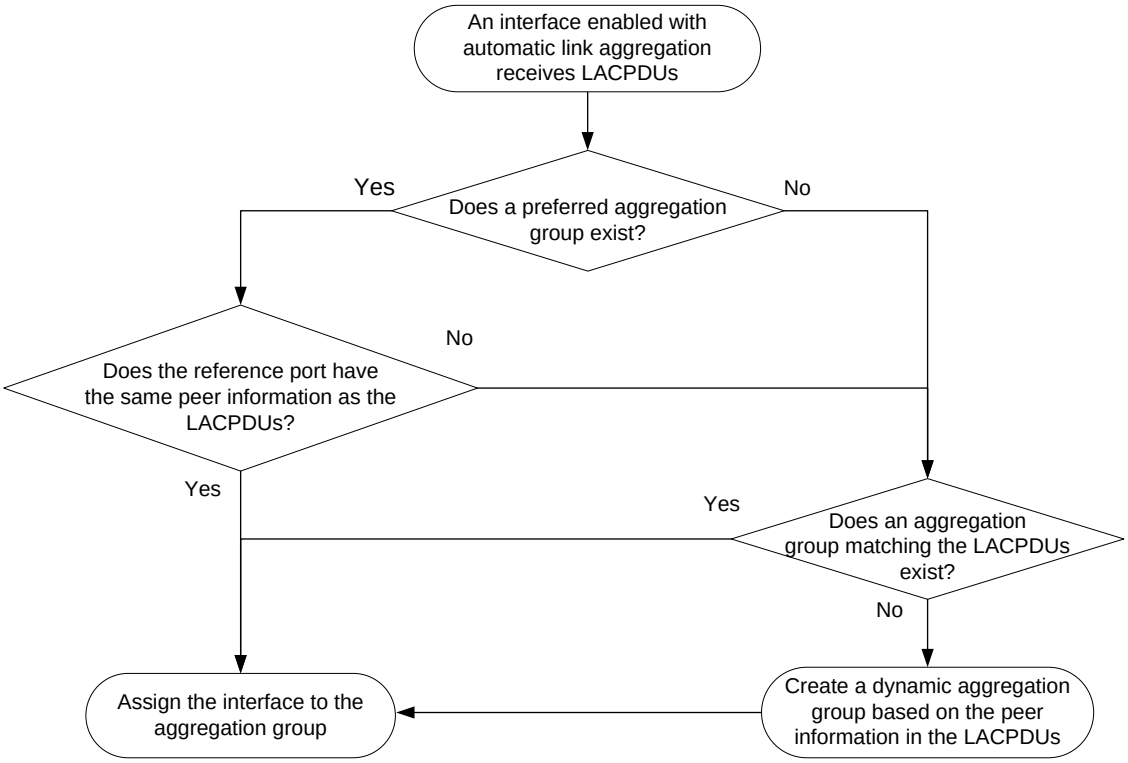
Automatic member port assignment

This feature automates the assignment of aggregation member ports to an aggregation group. You can use this feature when setting up an aggregate link to a server.

As shown in Figure 3, an interface enabled with automatic assignment joins a dynamic aggregation group based on the peer information in the LACPDUs received from the aggregation peer. If none of the existing dynamic aggregation groups is qualified, the device automatically creates a new dynamic aggregation group, Then, the device assigns the interface to that group and synchronizes the interface's attribute configurations to the aggregate interface.

A dynamic aggregation group that contains automatically assigned member ports selects a reference port and Selected ports as described in "How dynamic link aggregation works." The assignment methods of member ports do not change the processes of reference port selection and Selected port selection.

Figure 3 Automatic member port assignment process



Automatic link aggregation

Automatic link aggregation enables two devices to automatically create a dynamic link aggregation between them by using LLDP.

After you enable automatic link aggregation and LLDP on two connected devices, they automatically establish a dynamic link aggregation based on the information in incoming LLDP frames. The devices each automatically create a dynamic aggregate interface and assign the redundant ports connected to the peer to the aggregation group of that interface. When assigning the first member port to the aggregate group, the device synchronizes the member port's attribute configuration to the aggregate interface.

An automatically created dynamic aggregation group selects a reference port and Selected ports as described in "How dynamic link aggregation works." The aggregation group creation methods do not change the processes of reference port selection and Selected port selection.

IMPORTANT:

As a best practice to ensure correct operation of dynamic aggregation groups, do not use automatic link aggregation and automatic member port assignment together.

How dynamic link aggregation works

Choosing a reference port

The system chooses a reference port from the member ports in up state. A Selected port must have the same operational key and attribute configurations as the reference port.

The local system (the actor) and the peer system (the partner) negotiate a reference port by using the following workflow:

1. The two systems determine the system with the smaller system ID.

A system ID contains the LACP system priority and the system MAC address.

a. The two systems compare their LACP priority values.

The lower the LACP priority, the smaller the system ID. If the LACP priority values are the same, the two systems proceed to step b.

b. The two systems compare their MAC addresses.

The lower the MAC address, the smaller the system ID.

2. The system with the smaller system ID chooses the port with the smallest port ID as the reference port.

A port ID contains a port priority and a port number. The lower the port priority, the smaller the port ID.

a. The system chooses the port with the lowest priority value as the reference port.

If the ports have the same priority, the system proceeds to step b.

b. The system compares their port numbers.

The smaller the port number, the smaller the port ID.

The port with the smallest port number and the same attribute configurations as the aggregate interface is chosen as the reference port.

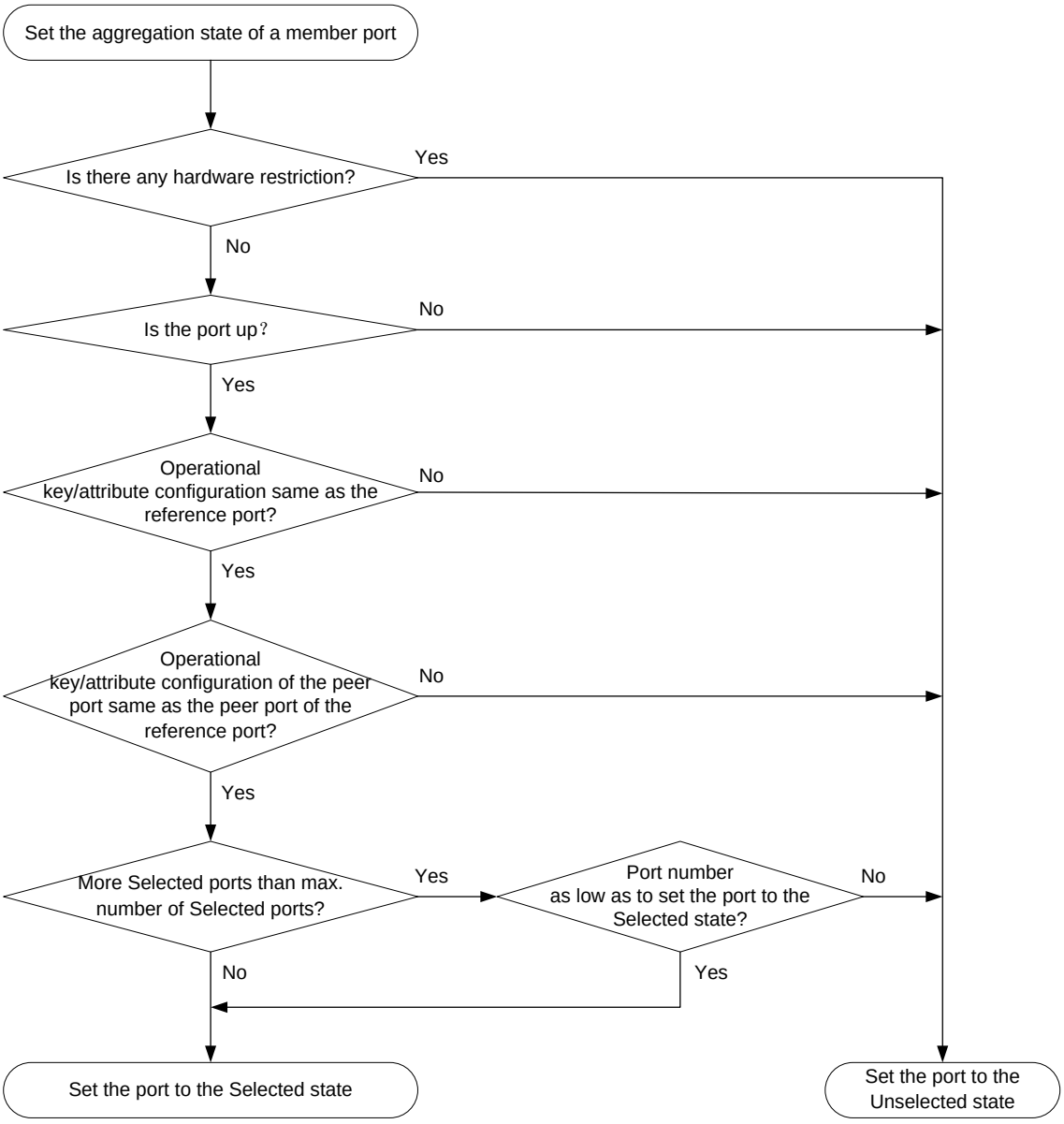
NOTE:

To identify the port numbers of aggregation member ports, execute the **`display link-aggregation verbose`** command and examine the **`Index`** field in the command output.

Setting the aggregation state of each member port

After the reference port is chosen, the system with the smaller system ID sets the state of each member port on its side.

Figure 4 Setting the state of a member port in a dynamic aggregation group



The system with the greater system ID can detect the aggregation state changes on the peer system. The system with the greater system ID sets the aggregation state of local member ports the same as their peer ports.

When you aggregate interfaces in dynamic mode, follow these guidelines:

- A dynamic link aggregation group chooses only full-duplex ports as the Selected ports.
- For stable aggregation and service continuity, do not change the operational key or attribute configurations on any member port.
- When a member port changes to the Selected or Unselected state, its peer port changes to the same aggregation state.
- After the Selected port limit is reached, a newly joining port becomes a Selected port if it is more eligible than a current Selected port.

Edge aggregate interface

Dynamic link aggregation fails on a server-facing aggregate interface if dynamic link aggregation is configured only on the device. The device forwards traffic by using only one of the physical ports that are connected to the server.

To improve link reliability, configure the aggregate interface as an edge aggregate interface. This feature enables all member ports of the aggregation group to forward traffic. When a member port fails, its traffic is automatically switched to other member ports.

After dynamic link aggregation is configured on the server, the device can receive LACPDUs from the server. Then, link aggregation between the device and the server operates correctly.

An edge aggregate interface takes effect only when it is configured on an aggregate interface corresponding to a dynamic aggregation group.

Load sharing modes for link aggregation groups

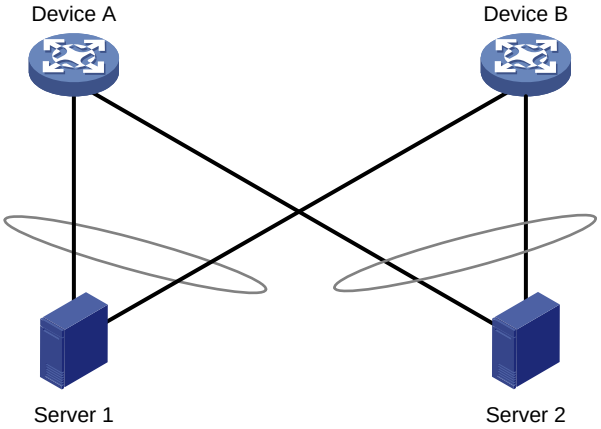
In a link aggregation group, traffic can be load shared across the Selected ports based on any of the following modes:

- **Per-flow load sharing**—Distributes traffic on a per-flow basis. The load sharing mode classifies packets into flows and forwards packets of the same flow on the same link. This mode can be one of or a combination of the following traffic classification criteria:
  - Ingress port.
  - Source or destination IP.
  - Source or destination MAC.
  - Source or destination port number.
- **Intelligent load sharing**—Distributes traffic based on the bandwidth usage of Selected member ports. In this mode, the device periodically samples the bandwidth usage of Selected member ports and adjusts traffic distribution if the bandwidth usage of a Selected member port exceeds a threshold. The threshold varies by device model. After you set the intelligent load sharing mode, the device distributes traffic based on the default load sharing mode and then optimizes traffic distribution based on the bandwidth usage.

S-MLAG

As shown in Figure 5, simple multichassis link aggregation (S-MLAG) aggregates multiple physical devices on the aggregation layer into one device to implement simple cross-device link aggregation and provide device-level redundancy protection and load sharing in server interconnect scenarios.

Figure 5 S-MLAG application scenario



Restrictions and guidelines: Mixed use of manual and automatic link aggregation configuration

To avoid unexpected aggregation issues, do not use manual assignment, automatic assignment, and automatic link aggregation in any combination. If you use any two of these features in combination, an automatically assigned member port might move between aggregation groups or undesirably change from Selected to Unselected in some situations.

To avoid unexpected aggregation issues, do not use the manual aggregation and semi-automatic in the same aggregation group.

For a mirroring group, do not assign the source port to an aggregation group other than the one that accommodates the destination port, egress port, or reflector port. If the source port is in a different aggregation group than the other ports, mirrored LACPDUs will be transmitted between the aggregation groups and cause aggregate interface flapping.

Ethernet link aggregation tasks at a glance

To configure Ethernet link aggregation, perform the following tasks:

1. Configuring the system ID
2. Configuring link aggregation modes
   - Configuring a manual link aggregation
   - Configuring automatic link aggregation
   - Configuring S-MLAG
3. (Optional.) Configuring an aggregate interface
   - Configuring the description of an aggregate interface
   - Setting the MAC address for an aggregate interface
   - Setting the MTU for a Layer 3 aggregate interface
   - Setting the expected bandwidth for an aggregate interface
   - Configuring an edge aggregate interface

     An edge aggregate interface uses all member ports to forward traffic when the aggregation peer is not enabled with dynamic link aggregation.
   - Configuring physical state change suppression on an aggregate interface

- Shutting down an aggregate interface
- Enabling interframe gap and preamble statistics in the traffic statistics
- Restoring the default settings for an aggregate interface

4. (Optional.) Enabling transparent LACPDU transmission
5. (Optional.) Adjusting aggregation states of link aggregation member ports
   - Setting the minimum and maximum numbers of Selected ports for an aggregation group
   - Disabling the default action of selecting a Selected port for dynamic aggregation groups that have not received LACPDUs
   - Configuring a dynamic aggregation group to use port speed as the prioritized criterion for reference port selection
   - Ignoring port speed in setting the aggregation states of member ports
   - Enabling port selection preemption on an aggregate interface
6. (Optional.) Configuring load sharing for link aggregation groups
   - Setting static load sharing modes for link aggregation groups
   - Enabling local-first load sharing for link aggregation
   - Configuring link aggregation load sharing algorithm settings
7. (Optional.) Optimizing traffic forwarding
   - Specifying link aggregation management VLANs and link aggregation management port

     Perform this task to enable an aggregation group to forward Layer 3 data traffic of some VLANs through a specific member port.
   - Enabling link-aggregation traffic redirection

     This feature redirects traffic on an unavailable Selected port to the remaining available Selected ports of an aggregation group to avoid traffic interruption.
   - Isolating aggregate interfaces on the device
8. (Optional.) Enabling BFD for an aggregation group

Configuring the system ID

About this task

The two ends of a dynamic aggregate link choose a reference port from the end with a smaller system ID.

The system ID contains the LACP system priority and LACP system MAC address. Two devices use the following rules to compare their system IDs:

- If their system IDs contain different LACP system priorities, the system ID with a smaller LACP system priority value is smaller.
- If their system IDs contain the same LACP system priority, the system ID with a lower LACP system MAC address is smaller.

To view the LACP system MAC address and LACP system priority, execute the **`display link-aggregation verbose`** command.

You can configure the system ID globally and in aggregate interface view. The global system ID takes effect on all aggregation groups, and an aggregate-interface-specific system ID takes precedence over the global system ID.

Restrictions and guidelines

Member devices in an S-MLAG system must use the same LACP system priority and LACP system MAC address.

On an M-LAG system, M-LAG interfaces in the same M-LAG group must use the same LACP system MAC address.

For member ports to be selected correctly, do not modify the LACP system priority and LACP system MAC address after a dynamic link aggregation is established.

Procedure

1. Enter system view.
   system-view
2. Set the LACP system MAC address globally.
   lacp system-mac *mac-address*
   By default, the LACP system MAC address is the bridge MAC address of the device.
3. Set the LACP system priority globally.
   lacp system-priority *priority*
   By default, the LACP system priority is 32768.
4. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
5. Set the LACP system MAC address on the aggregate interface.
   port lacp system-mac *mac-address*
   By default, the LACP system MAC address is the bridge MAC address of the device.
6. Set the LACP system priority on the aggregate interface.
   port lacp system-priority *priority*
   By default, the LACP system priority is 32768.

Configuring a manual link aggregation

Restrictions and guidelines for aggregation group configuration

Layer 2 aggregation group restrictions

You cannot assign an interface to a Layer 2 aggregation group if any features in Table 4 are configured on that interface.

Table 4 Features incompatible with Layer 2 aggregation member ports

Feature on the interface	Reference
MAC authentication	MAC authentication in *Security Configuration Guide*
Port security	Port security in *Security Configuration Guide*
802.1X	802.1X in *Security Configuration Guide*

Aggregation member port restrictions

Deleting an aggregate interface also deletes its aggregation group and causes all member ports to leave the aggregation group.

If an interface has different attribute configurations from the aggregate interface, the interface cannot join the aggregation group.

After joining an aggregation group, an interface inherits the attribute configurations on the aggregate interface. You can modify the attribute configurations only on the aggregate interface.

After an interface leaves an aggregation group, VLAN settings on the interface will not be restored to the default.

To avoid a loop, shut down an interface before removing the interface from an aggregation group.

Do not assign a reflector port for port mirroring to an aggregation group. For more information about reflector ports, see *Network Management and Monitoring Configuration Guide*.

You cannot assign monitor ports for port mirroring to an aggregation group.

Attribute and protocol configuration restrictions

For a link aggregation, attribute configurations are configurable only on the aggregate interface and are automatically synchronized to all member ports. You cannot configure attribute configurations on a member port until it is removed from the link aggregation group. After the aggregate interface is deleted, the configurations that have been synchronized from the aggregate interface are retained on the member ports.

If an attribute setting on the aggregate interface fails to be synchronized to a Selected member port, the port might change to the Unselected state.

The protocol configurations for an aggregate interface take effect only on the current aggregate interface. The protocol configurations for a member port take effect only when the port leaves its aggregation group.

Configuration consistency requirements

You must configure the same aggregation mode at the two ends of an aggregate link.

- For a successful static aggregation, make sure the ports at both ends of each link are in the same aggregation state.
- For a successful dynamic aggregation:
  - Make sure the ports at both ends of a link are assigned to the correct aggregation group. The two ends can automatically negotiate the aggregation state of each member port.
  - If you use automatic interface assignment on one end, you must use manual assignment on the other end.

Configuring a Layer 2 aggregation group

Configuring a Layer 2 dynamic aggregation group

1. Enter system view.

   `system-view`
2. Create a Layer 2 aggregate interface and enter Layer 2 aggregate interface view.

   `interface bridge-aggregation` *`interface-number`*

   When you create a Layer 2 aggregate interface, the system automatically creates a Layer 2 static aggregation group numbered the same as that interface.
3. Configure the aggregation group to operate in dynamic mode.

   `link-aggregation mode dynamic`

   By default, an aggregation group operates in static mode.
4. Return to system view.

   `quit`

5. (Optional.) Allow only interfaces with the default VLAN configuration to join aggregation groups and restore the default VLAN configuration when they leave aggregation groups.

   link-aggregation bind-mode bagg-vlan-default

   By default, the feature that allows only interfaces with the default VLAN configuration to join aggregation groups and restores the default VLAN configuration when they leave aggregation groups is not configured.

6. Assign an interface to the Layer 2 aggregation group:

   a. Enter Layer 2 Ethernet interface view.

      interface *interface-type interface-number*

   b. Assign the interface to the Layer 2 aggregation group or enable automatic assignment on that interface.

      port link-aggregation group { *group-id* [**force**] | **auto** [*group-id*] }

   Repeat these two substeps to assign more Layer 2 Ethernet interfaces to the aggregation group.

   To synchronize the attribute configurations from the aggregate interface when the current interface joins the aggregation group, specify the **force** keyword.

   To enable automatic assignment, specify the **auto** keyword. As a best practice, do not modify the configuration on an automatically created aggregate interface or its member ports.

7. Set the LACP operating mode for the interface.

   - Set the LACP operating mode to passive.

     lacp mode passive

   - Set the LACP operating mode to active.

     undo lacp mode

   By default, LACP is operating in active mode.

8. (Optional.) Set the port priority for the interface.

   link-aggregation port-priority *priority*

   The default setting is 32768.

9. (Optional.) Set the short LACP timeout interval for the interface.

   lacp period short [**timeout** *timeout*]

   By default, the long LACP timeout interval (90 seconds) is used by the interface.

   To avoid traffic interruption during an ISSU, do not set the short LACP timeout interval before performing the ISSU. For more information about ISSU, see *Fundamentals Configuration Guide*.

Configuring a Layer 3 aggregation group

Configuring a Layer 3 dynamic aggregation group

1. Enter system view.

   system-view

2. Create a Layer 3 aggregate interface and enter Layer 3 aggregate interface view.

   interface route-aggregation *interface-number*

   When you create a Layer 3 aggregate interface, the system automatically creates a Layer 3 static aggregation group numbered the same as that interface.

3. Configure the aggregation group to operate in dynamic mode.

   link-aggregation mode dynamic

   By default, an aggregation group operates in static mode.

4. Return to system view.
   quit
5. Assign an interface to the Layer 3 aggregation group:
   a. Enter Layer 3 Ethernet interface view.
      interface *interface-type interface-number*
   b. Assign the interface to the Layer 3 aggregation group or enable automatic assignment on that interface.
      port link-aggregation group { *group-id* | **auto** [*group-id*] }

   Repeat these two substeps to assign more Layer 3 Ethernet interfaces to the aggregation group.

   To enable automatic assignment, specify the **auto** keyword. As a best practice, do not modify the configuration on an automatically created aggregate interface or its member ports.
6. Set the LACP operating mode for the interface.
   - Set the LACP operating mode to passive.
     lacp mode passive
   - Set the LACP operating mode to active.
     undo lacp mode

   By default, LACP is operating in active mode.
7. (Optional.) Set the port priority of the interface.
   link-aggregation port-priority *priority*

   The default setting is 32768.
8. (Optional.) Set the short LACP timeout interval for the interface.
   lacp period short [**timeout** *timeout*]

   By default, the long LACP timeout interval (90 seconds) is used by the interface.

   To avoid traffic interruption during an ISSU, do not set the short LACP timeout interval before performing the ISSU. For more information about ISSU, see *Fundamentals Configuration Guide*.

Configuring a Layer 3 aggregate subinterface

1. Enter system view.
   system-view
2. Create a Layer 3 aggregate subinterface, and enter Layer 3 aggregate subinterface view.
   interface route-aggregation *interface-number.subnumber*

Configuring automatic link aggregation

About this task

You can use automatic link aggregation to aggregate the redundant physical links between devices to ease management and increase bandwidth and availability.

Restrictions and guidelines

On an interface, the **port link-aggregation group** setting takes precedence over automatic link aggregation. The interface will not be added to the aggregation group of an automatically created aggregate interface if it has been the member port of a manually created aggregate interface.

If automatic link aggregation is enabled, subinterface creation might fail on LLDP-enabled Layer 3 Ethernet interfaces. As a best practice, do not create subinterfaces on LLDP-enabled Layer 3 Ethernet interfaces.

To ensure correct operation of an automatically created aggregate interface ,do not modify the configuration on the member ports of its aggregation group. Doing so might cause the member ports to be removed from the aggregation group.

For M-LAG to operate correctly, do not configure automatic link aggregation on an M-LAG system.

As a best practice to ensure correct operation of dynamic aggregation groups, do not use automatic link aggregation and automatic member port assignment together.

Prerequisites

Before you configure automatic link aggregation, enable LLDP on the peer devices.

Procedure

1. Enter system view.

   `system-view`

2. Enable automatic link aggregation.

   `link-aggregation auto-aggregation enable`

   By default, automatic link aggregation is disabled.

Configuring S-MLAG

Restrictions and guidelines

Use S-MLAG only to establish aggregate links with servers.

S-MLAG is intended for a non-IRF environment. Do not configure it on an IRF fabric. For more formation about IRF, see *Virtual Technologies Configuration Guide*.

Each S-MLAG group can contain only one aggregate interface on each device.

The aggregate interfaces in an S-MLAG group cannot be used as M-LAG interfaces or M-LAG peer-link interfaces. For more information about M-LAG interfaces and peer-link interfaces, see M-LAG configuration in *Layer 2—LAN Switching Configuration Guide*.

On S-MLAG devices, make sure the member ports in an aggregation group have the same speed and duplex mode. Inconsistency in these settings might cause reference port reselection and interrupt traffic forwarding when new member ports join the aggregation group.

Do not configure the following settings on S-MLAG devices:

- LACP MAD.
- Link-aggregation traffic redirection.
- Maximum or minimum number of Selected ports.
- Automatic member port assignment.
- Ignoring port speed in setting the aggregation states of member ports.
- Spanning tree. For more information, see "Configuring spanning tree protocols."

As a best practice, maintain consistency across S-MLAG devices in service feature configuration.

Prerequisites

Configure the link aggregation settings other than S-MLAG settings on each S-MLAG device. Make sure the settings are consistent across the S-MLAG devices.

Procedure

1. Enter system view.

   `system-view`

2. Set the LACP system MAC address.

lacp system-mac *mac-address*

By default, the LACP system MAC address is the bridge MAC address of the device.

All S-MLAG devices must use the same LACP system MAC address.

3. Set the LACP system priority.

   lacp system-priority *priority*

   By default, the LACP system priority is 32768.

   All S-MLAG devices must use the same LACP system priority.

4. Set the LACP system number.

   lacp system-number *number*

   By default, the LACP system number is not set.

   You must assign a unique LACP system number to each S-MLAG device.

5. Enter Layer 2 aggregate interface view.

   interface bridge-aggregation *interface-number*

6. Set the link aggregation mode to dynamic.

   link-aggregation mode dynamic

   By default, an aggregation group operates in static mode.

7. Assign the aggregate interface to an S-MLAG group.

   port s-mlag group *group-id*

   By default, an aggregate interface is not assigned to any S-MLAG group.

Configuring an aggregate interface

Most settings that can be made on Layer 2 or Layer 3 Ethernet interfaces can also be made on Layer 2 or Layer 3 aggregate interfaces.

Configuring the description of an aggregate interface

About this task

You can configure the description of an aggregate interface for administration purposes, for example, describing the purpose of the interface.

Procedure

1. Enter system view.

   system-view

2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number*
   - Enter Layer 3 aggregate subinterface view.

     interface route-aggregation *interface-number.subnumber* }

3. Configure the interface description.

   description *text*

   By default, the description of an interface is *interface-name* **Interface**.

Setting the MAC address for an aggregate interface

About this task

Typically, all aggregate interfaces on a device use the same MAC address, and aggregate interfaces on different devices use different MAC addresses. However, you must set different MAC addresses for aggregate interfaces on a device in some situations.

Procedure

1. Enter system view.

   system-view

   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number*

   - Enter Layer 3 aggregate subinterface view.

     interface route-aggregation *interface-number.subnumber*

2. Set the MAC address for the aggregate interface.

   mac-address *mac-address*

   By default, MAC addresses are not configured for aggregate interfaces on the device.

Setting the MTU for a Layer 3 aggregate interface

About this task

The MTU of an interface affects IP packets fragmentation and reassembly on the interface.

Restrictions and guidelines

The **mtu** command applies only to protocol packets and does not affect forwarded data packets.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 3 aggregate interface or subinterface view.

   interface route-aggregation { *interface-number* | *interface-number.subnumber* }

3. Set the MTU.

   mtu *size*

   The default setting is 1500 bytes.

Setting the expected bandwidth for an aggregate interface

About this task

Expected bandwidth is an informational parameter used only by higher-layer protocols for calculation. You cannot adjust the actual bandwidth of an interface by performing this task.

Procedure

1. Enter system view.

   system-view

2. Enter aggregate interface view.

   - Enter Layer 2 aggregate interface view.

interface bridge-aggregation *interface-number*

- Enter Layer 3 aggregate interface view.
  interface route-aggregation *interface-number*
- Enter Layer 3 aggregate subinterface view.
  interface route-aggregation *interface-number.subnumber* }

3. Set the expected bandwidth for the interface.
   bandwidth *bandwidth-value*
   By default, the expected bandwidth (in kbps) is the interface baud rate divided by 1000.

Configuring an edge aggregate interface

Restrictions and guidelines

This configuration takes effect only on aggregate interfaces in dynamic mode.

Link-aggregation traffic redirection cannot operate correctly on an edge aggregate interface. For more information about link-aggregation traffic redirection, see "Enabling link-aggregation traffic redirection."

Procedure

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Configure the aggregate interface as an edge aggregate interface.
   lacp edge-port
   By default, an aggregate interface does not operate as an edge aggregate interface.

Configuring physical state change suppression on an aggregate interface

About this task

The physical link state of an aggregate interface is either up or down. Each time the physical link of an interface comes up or goes down, the system immediately reports the change to the CPU. The CPU then notifies the upper-layer protocol modules (such as routing and forwarding modules) of the change, and the device automatically generates traps and log messages and sends them to the SNMP and information center modules. You can configure SNMP and information center to output these messages.

To prevent frequent physical link flapping from affecting system performance, configure physical state change suppression. You can configure this feature to suppress link-down events, link-up events, or both. If an event of the specified type still exists when the suppression interval expires, the system reports the event to the CPU.

Restrictions and guidelines

When you configure suppression interval settings on an aggregate interface, make sure its peer interface is also an aggregate interface in the same aggregation mode. In addition, the suppression interval settings must be the same between the peer aggregate interfaces.

As a best practice, use the default setting in an S-MLAG environment.

On an interface, you can configure different suppression intervals for link-up and link-down events. If you execute the **link-delay** command multiple times for an event type, the most recent configuration takes effect on that event type.

Use this feature on an aggregate interface to reduce the impact of interface flapping on upper-layer services, for example, on an M-LAG peer-link interface. For more information about peer-link interfaces, see "Configuring M-LAG."

Procedure

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Configure physical state change suppression.
   link-delay { **down** | **up** } [**msec**] *delay-time*

   By default, each time the physical link of an aggregate interface goes up or comes down, the system immediately reports the change to the CPU.

Shutting down an aggregate interface

Restrictions and guidelines

Shutting down or bringing up an aggregate interface affects the aggregation states and link states of member ports in the corresponding aggregation group as follows:

- When an aggregate interface is shut down, all its Selected ports become Unselected and all member ports go down.
- When an aggregate interface is brought up, the aggregation states of all its member ports are recalculated.

When you shut down or bring up a Layer 3 aggregate interface, all its aggregate subinterfaces are also shut down or brought up. Shutting down or bringing up a Layer 3 aggregate subinterface does not affect the state of the main aggregate interface.

Procedure

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
   - Enter Layer 3 aggregate subinterface view.
     interface route-aggregation *interface-number.subnumber* }
3. Shut down the interface.
   shutdown

CAUTION:

The **shutdown** command will disconnect all links established on an interface. Make sure you are fully aware of the impacts of this command when you use it on a live network.

Enabling interframe gap and preamble statistics in the traffic statistics

About this task

If you execute the **display interface** command, the Last 300 seconds input rate or Last 300 seconds output rate field in the command output displays the average outbound or inbound traffic rate in the last 300 seconds.

By default, traffic rate = native frame length × packet throughput per second. Execute the **traffic-statistic include-interframe** command if you need the total traffic statistics, including the native frame length, interframe gap length and preamble length, for a specific time period. If you execute this command, traffic rate = (native frame length + interframe gap length + preamble length) × packet throughput per second.

Restrictions and guidelines

If you perform this task on an aggregate interface, the configuration takes effect on all Ethernet interfaces in the aggregation group of the aggregate interface. The interframe gap and preamble statistics setting configured in Ethernet interface view do not take effect. If an Ethernet interface is removed from the aggregation group, the interframe gap and preamble statistics setting configured in Ethernet interface view takes effect.

Procedure

1. Enter system view.

   system-view
2. Enter Ethernet interface view.
   - Enter Layer 2 Ethernet interface view.

     interface bridge-aggregation *interface-number*
   - Enter Layer 3 Ethernet interface view.

     interface route-aggregation *interface-number*
3. Enable interframe gap and preamble statistics in the traffic statistics.

   traffic-statistic include-interframe

   The default for this command varies by device model.

Restoring the default settings for an aggregate interface

Restrictions and guidelines

CAUTION:

The **default** command might interrupt ongoing network services. Make sure you are fully aware of the impacts of this command when you execute it on a live network.

The **default** command might fail to restore the default settings for some commands for reasons such as command dependencies and system restrictions.

To resolve this issue:

1. Use the **display this** command in interface view to identify these commands.

2. Use their **undo** forms or follow the command reference to restore their default settings.
3. If the restoration attempt still fails, follow the error message instructions to resolve the issue.

Procedure

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
   - Enter Layer 3 aggregate subinterface view.
     interface route-aggregation *interface-number.subnumber* }
3. Restore the default settings for the aggregate interface.
   Default

Enabling transparent LACPDU transmission

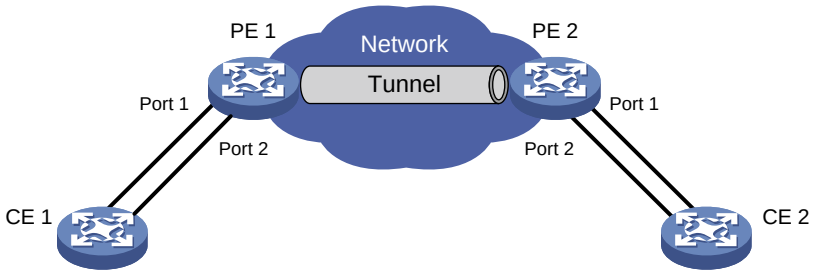
About this task

To establish a dynamic aggregation between two remote CEs in an L2VPN network, use transparent LACPDU transmission on the PEs to which the CEs are attached, as shown in Figure 6.

This feature enables the PEs to forward LACPDUs for the CEs to establish a dynamic aggregation.

If this feature is disabled, the PEs terminate the LACPDUs. The remote CEs cannot establish dynamic aggregations.

Figure 6 Application scenario of transparent LACPDU transmission



Prerequisites

Perform the following tasks on PEs:

1. Configure the **untagged** or **default** frame match criterion for the Ethernet service instances on the interfaces connected to CEs.
2. Map the Ethernet service instances to different VSIs and set the access mode to Ethernet for the VSIs.

For more information about L2VPN and Ethernet service instances, see L2VPN configuration in *MPLS Configuration Guide*.

Restrictions and guidelines

When you use this feature on the PEs, follow these guidelines:

- Enable transparent LACPDU transmission on the interfaces that transmit traffic between CEs and PEs and between PEs.

- Since transparent LACPDU transmission can affect the selection of member ports, this feature can only be configured on a single inbound and outbound port of the PEs, not on multiple inbound and outbound ports.

Procedure

1. Enter system view.

   `system-view`
2. Enter interface view.

   `interface` *`interface-type interface-number`*
3. Enable transparent LACPDU transmission.

   `lacp transparent enable`

   By default, transparent LACPDU transmission is disabled.

Setting the minimum and maximum numbers of Selected ports for an aggregation group

About this task

The bandwidth of an aggregate link increases as the number of Selected member ports increases. To avoid congestion, you can set the minimum number of Selected ports required for bringing up an aggregate interface.

This minimum threshold setting affects the aggregation states of aggregation member ports and the state of the aggregate interface.

- When the number of member ports eligible to be Selected ports is smaller than the minimum threshold, the following events occur:
  - The eligible member ports are placed in Unselected state.
  - The link layer state of the aggregate interface becomes down.
- When the number of member ports eligible to be Selected ports reaches or exceeds the minimum threshold, the following events occur:
  - The eligible member ports are placed in Selected state.
  - The link layer state of the aggregate interface becomes up.

The maximum number of Selected ports allowed in an aggregation group is limited by either manual configuration or hardware limitation, whichever value is smaller.

You can implement backup between two ports by performing the following tasks:

- Assigning two ports to an aggregation group.
- Setting the maximum number of Selected ports to 1 for the aggregation group.

Then, only one Selected port is allowed in the aggregation group, and the Unselected port acts as a backup port.

Restrictions and guidelines

IMPORTANT:

After you set the minimum percentage of Selected ports for an aggregation group, aggregate interface flapping might occur when ports join or leave an aggregation group. Make sure you are fully aware of the impacts of this setting when you configure it on a live network.

You can set either the minimum number or the minimum percentage of Selected ports for an aggregation group. If you configure both settings on an aggregate interface, the higher Selected port number limit takes effect.

The minimum and maximum numbers of Selected ports must be the same between the two ends of an aggregate link.

The minimum percentage of Selected ports must be the same between the two ends of an aggregate link.

For an aggregation group, the maximum number of Selected ports must be equal to or higher than the minimum number of Selected ports. If you set the minimum percentage of Selected ports for the aggregation group, do not set the maximum number of Selected ports. If you set the maximum number of Selected ports, the calculated minimum number of Selected ports might be larger than the maximum number of Selected ports.

Procedure

1. Enter system view.
   `system-view`
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     `interface bridge-aggregation` *`interface-number`*
   - Enter Layer 3 aggregate interface view.
     `interface route-aggregation` *`interface-number`*
3. Set the minimum number of Selected ports for the aggregation group. Choose one of the following methods:
   - Set the minimum number of Selected ports.
     `link-aggregation selected-port minimum` *`min-number`*
   - Set the minimum percentage of Selected ports.
     `link-aggregation selected-port minimum percentage` *`number`*

   By default, the minimum number of Selected ports is not specified for an aggregation group.
4. Set the maximum number of Selected ports for the aggregation group.
   `link-aggregation selected-port maximum` *`max-number`* [**`lacp-sync`**]

   By default, the maximum number of Selected ports for an aggregation group is 64.

   For a static aggregate link, you must set the maximum number of Selected ports to the same value at its two ends.

   For a dynamic aggregate link, you must set the maximum number of Selected ports to the same value at its two ends if you do not specify the **`lacp-sync`** keyword. If you specify this keyword, the two ends of the aggregate link compare their maximum Selected port number settings and use the smaller value.

Disabling the default action of selecting a Selected port for dynamic aggregation groups that have not received LACPDUs

About this task

The default port selection action applies to dynamic aggregation groups.

This action automatically chooses the port with the lowest ID from among all up member ports as a Selected port if none of them has received LACPDUs before the LACP timeout interval expires.

After this action is disabled, a dynamic aggregation group will not have any Selected ports to forward traffic if it has not received LACPDUs before the LACP timeout interval expires.

Procedure

1. Enter system view.
   system-view
2. Disable the default port selection action.
   lacp default-selected-port disable
   By default, the default port selection action is enabled for dynamic aggregation groups.

Configuring a dynamic aggregation group to use port speed as the prioritized criterion for reference port selection

About this task

Perform this task to ensure that a dynamic aggregation group selects a high-speed member port as the reference port. After you perform this task, the reference port will be selected based on the criteria in order of device ID, port speed, and port ID.

Restrictions and guidelines

Changing reference port selection criteria might cause transient traffic interruption. Make sure you understand the impact of this task on your network.

As a best practice, shut down the peer aggregate interfaces before you execute this command and bring up the interfaces after this command is executed on both of them.

Procedure

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Specify port speed as the prioritized criterion for reference port selection.
   lacp select speed
   By default, port ID is the prioritized criterion for reference port selection of a dynamic aggregation group.

Ignoring port speed in setting the aggregation states of member ports

About this task

This feature allows ports at a different speed than the reference port to become Selected by ignoring the port speed during operational key calculation.

You must configure the same port speed ignoring setting at the two ends of a static configuration to ensure that the peer ports are placed in the same aggregation state. This requirement does not apply to a dynamic aggregation, on which the two ends negotiate the aggregation state of the peer ports automatically.

Restriction and guidelines

Make sure you are fully aware of the impacts of this feature when you use it on a live network.

- When you execute the **link-aggregation ignore speed** command or its **undo** form, the operational key changes and the aggregate interface flaps temporarily.
- The **link-aggregation ignore speed** command might cause traffic loss on Selected ports operating at a speed lower than the other Selected ports in the aggregation group when traffic is distributed among the ports.

Procedure

1. Enter system view.

   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number*
3. Configure the aggregation group to ignore port speed in setting the aggregation states of member ports.

   link-aggregation ignore speed

   By default, an aggregation group does not ignore port speed in setting the aggregation states of member ports.

Enabling port selection preemption on an aggregate interface

About this task

When port A fails in a dynamic aggregation group, the device selects the highest priority member port, port B for example, as the new selected port. If port selection preemption is enabled and port A recovers from failure, it replaces port B. If port selection preemption is disabled, port B is still selected even though port A has higher priority, and traffic loss caused by selected port preemption is reduced.

If port selection preemption is enabled and port A recovers from failure, it immediately replaces port B. This might cause packet loss if port A's link status is unstable. To avoid this issue, configure a port selection preemption delay.

Procedure

1. Enter system view.

   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number*

3. Enable port selection preemption on an aggregate interface.

   lacp preempt enable

   By default, port selection preemption is enabled on an aggregate interface.

4. Set the port selection preemption delay.

   lacp preempt delay *delay-time*

   By default, the port selection preemption delay is 0 seconds on an aggregate interface, which means port selection preemption is performed without delay.

Configuring load sharing for link aggregation groups

Setting static load sharing modes for link aggregation groups

About this task

You can set the static global or group-specific load sharing mode. A link aggregation group preferentially uses the group-specific load sharing mode. If the group-specific load sharing mode is not available, the group uses the global load sharing mode.

Restrictions and guidelines

After you enable symmetric load sharing, link-aggregation load sharing also uses the symmetric load sharing mode. For more information about symmetric load sharing, see IP forwarding basics configuration in *Layer 3—IP Services Configuration Guide*.

The following are global load sharing modes supported on the device:

- Any combination of destination IP address, destination MAC address, destination port, source IP address, source MAC address, source port, GRE key, and VXLAN ID.
- Any combination of ingress port, source MAC address, and destination MAC address.

The device does not support load sharing IPv6 packets with extension headers based on the service port numbers.

As a best practice, configure GRE key, destination port, and source port simultaneously because they use the same hashing field.

Setting the global link-aggregation load sharing mode

1. Enter system view.

   system-view

2. Set the global link-aggregation load sharing mode.

   link-aggregation global load-sharing mode { **destination-ip** | **destination-mac** | **destination-port** | **gre-key** | **ingress-port** | **ip-protocol** | **source-ip** | **source-mac** | **source-port** | **vxlan-id** } *

   By default, packets are load shared depending on their packet type, as follows:

   - Layer 2 frames are load shared based on the source and destination MAC addresses.
   - Layer 3 packets are load shared based on the source and destination IP addresses.
   - Layer 4 packets are load shared based on the source and destination IP addresses, and source and destination port numbers.

Setting the group-specific load sharing mode

1. Enter system view.

   system-view

2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Set the load sharing mode for the aggregation group.
   link-aggregation load-sharing mode { { **destination-ip** | **destination-mac** | **mpls-label1** | **mpls-label2** | **source-ip** | **source-mac** } * | **per-packet** }

   By default, group-specific load sharing mode is the same as the global load sharing mode.

   The device supports per-packet load sharing only for the known unicast traffic forwarded by an aggregate interface.

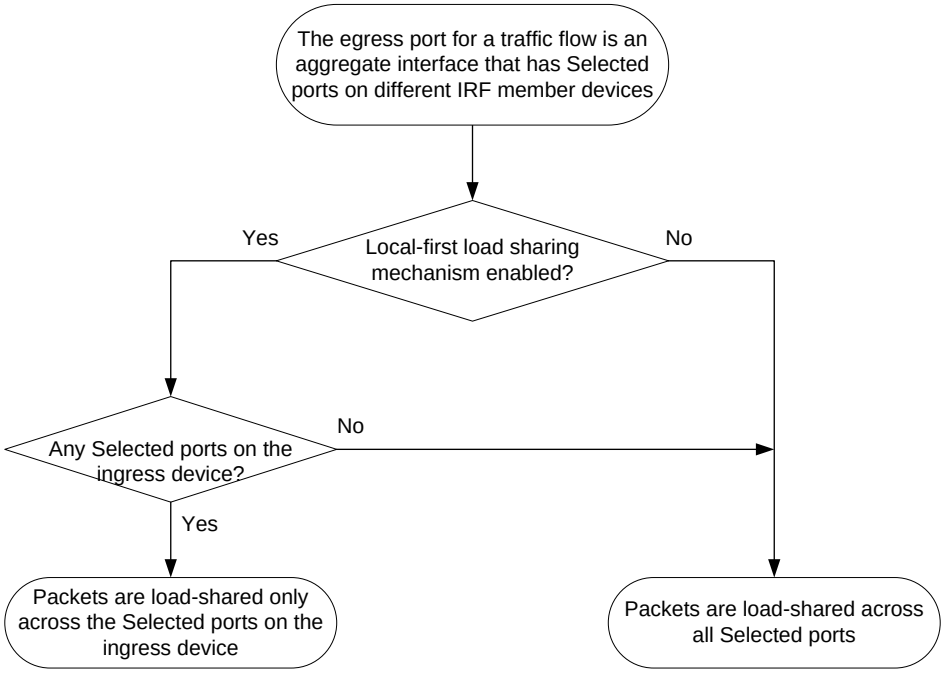
Enabling local-first load sharing for link aggregation

About this task

Use local-first load sharing in a multidevice link aggregation scenario to distribute traffic preferentially across member ports on the ingress slot.

When you aggregate ports on different member devices in an IRF fabric, you can use local-first load sharing to reduce traffic on IRF links, as shown in Figure 7. For more information about IRF, see *Virtual Technologies Configuration Guide*.

Figure 7 Load sharing for multidevice link aggregation in an IRF fabric



Restrictions and guidelines for local-first load sharing

This feature is available only in IRF mode.

You can use this feature only after physical interfaces are bound to an IRF port.

After you execute the **ip forward-broadcast** command, disable local-first load sharing on a Layer 2 aggregate interface if it is a traffic outgoing interface for directed broadcasts destined for the

directly connected network. For more information about this command, see IP performance optimization in *Layer 3—IP Services Configuration Guide*.

If you enable local-first load sharing on the device, an OpenFlow controller cannot create group entries on the device. For more information about OpenFlow, see *OpenFlow Configuration Guide*.

Enabling local-first load sharing for link aggregation globally

1. Enter system view.
   system-view
2. Enable local-first load sharing for link aggregation globally.
   link-aggregation load-sharing mode local-first
   By default, local-first load sharing is enabled globally.

Enabling local-first load sharing for an aggregate interface

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Enable local-first load sharing on the aggregate interface.
   link-aggregation group load-sharing mode local-first
   By default, local-first load sharing is not configured on aggregate interfaces. An aggregate interface uses the global local-first load-sharing setting.

Configuring link aggregation load sharing algorithm settings

About this task

Use the link aggregation load sharing algorithm to optimize traffic distribution on aggregate links when the default load sharing mode is used. Each algorithm represents a CRC calculation method.

Restrictions and guidelines

Change the link aggregation load sharing algorithm if different flows are not distributed to all member links of an aggregate link.

Procedure

1. Enter system view.
   system-view
2. Configure a link aggregation load sharing algorithm.
   link-aggregation global load-sharing algorithm *algorithm-number*
   By default, algorithm CRC is used.

Specifying link aggregation management VLANs and link aggregation management port

About this task

For an aggregation group to forward Layer 3 data traffic of some VLANs through a specific port, specify the VLANs as management VLANs and the port as a management port.

Procedure

1. Enter system view.
   system-view
2. Specify link aggregation management VLANs.
   link-aggregation management-vlan *vlan-id-list*
   By default, no link aggregation management VLANs are specified.
3. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
4. Configure the port as a management port for its aggregation group.
   link-aggregation management-port
   By default, a port does not act as a management port in its aggregation group.

Enabling link-aggregation traffic redirection

About link-aggregation traffic redirection

This feature operates on dynamic link aggregation groups. It redirects traffic on a Selected port to the remaining available Selected ports of an aggregation group if one of the following events occurs:

- The port is shut down by using the **shutdown** command.
- The slot that hosts the port reboots, and the aggregation group spans multiple slots.

NOTE:

The device does not redirect traffic to member ports that become Selected during the traffic redirection process.

This feature ensures zero packet loss for known unicast traffic, but does not protect unknown unicast traffic.

You can enable link-aggregation traffic redirection globally or for an aggregation group. Global link-aggregation traffic redirection settings take effect on all aggregation groups. A link aggregation group preferentially uses the group-specific link-aggregation traffic redirection settings. If group-specific link-aggregation traffic redirection is not configured, the group uses the global link-aggregation traffic redirection settings.

Restrictions and guidelines for link-aggregation traffic redirection

Link-aggregation traffic redirection applies only to dynamic link aggregation groups.

As a best practice, enable link-aggregation traffic redirection on a per-interface basis. If you enable this feature globally, communication with a third-party peer device might be affected if the peer is not compatible with this feature.

To prevent traffic interruption, enable link-aggregation traffic redirection at both ends of the aggregate link.

To prevent packet loss that might occur at a reboot, do not enable the spanning tree feature together with link-aggregation traffic redirection.

Link-aggregation traffic redirection does not operate correctly on an edge aggregate interface.

After the **link-aggregation lacp traffic-redirect-notification enable** command is executed, the device will add a proprietary field to the end of the LACP packets. Because the peer device cannot verify this proprietary field, LACP packets might fail to be verified and exchanged. As a result, the aggregation member ports cannot become Selected. To prevent this issue, you must disable link-aggregation traffic redirection on the H3C device when the H3C device connects to a third-party device.

Enabling link-aggregation traffic redirection globally

1. Enter system view.
   system-view
2. Enable link-aggregation traffic redirection globally.
   link-aggregation lacp traffic-redirect-notification enable
   By default, link-aggregation traffic redirection is disabled globally.

Enabling link-aggregation traffic redirection for an aggregation group

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Enable link-aggregation traffic redirection for the aggregation group.
   link-aggregation lacp traffic-redirect-notification enable

By default, link-aggregation traffic redirection is disabled for an aggregation group.

Isolating aggregate interfaces on the device

About this task

Aggregate interface isolation is applicable to dynamic aggregate interfaces when the device acts a M-LAG member device in an M-LAG system. It gracefully changes all dynamic aggregate interfaces on the device to the Unselected state and switch traffic over to their counterpart M-LAG interfaces on the other M-LAG member device.

Restrictions and guidelines

This feature takes effect only on dynamic aggregate interfaces. It cannot isolate static aggregate interfaces or M-LAG peer-link interfaces.

As a best practice, make sure no M-LAG interfaces are in M-LAG MAD DOWN state before you isolate them. If one of the M-LAG interfaces is in M-LAG MAD DOWN state when you isolate them, M-LAG interface will persist in that state and cannot forward traffic after the isolation is removed.

Procedure

1. Enter system view.

   system-view

2. Isolate aggregate interfaces.

   link-aggregation lacp isolate [**bagg** | **ragg**]

   By default, aggregate interfaces are not isolated.

   To remove DR interface isolation, execute the **undo** form of this command.

Enabling BFD for an aggregation group

About this task

You can use BFD to monitor member link status in an aggregation group. After you enable BFD on an aggregate interface, each Selected port in the aggregation group establishes a BFD session with its peer port. BFD operates differently depending on the aggregation mode.

- **BFD on a static aggregation**—When BFD detects a link failure, BFD notifies the Ethernet link aggregation module that the peer port is unreachable. The local port is then placed in Unselected state. However, the BFD session between the local and peer ports remains, and the local port keeps sending BFD packets. When BFD on the local port receives packets from the peer port upon link recovery, BFD notifies the Ethernet link aggregation module that the peer port is reachable. Then, the local port is placed in Selected state again. This mechanism ensures that the local and peer ports of a static aggregate link have the same aggregation state.
- **BFD on a dynamic aggregation**—When BFD detects a link failure, BFD notifies the Ethernet link aggregation module that the peer port is unreachable. At the same time, BFD clears the session and stops sending BFD packets. When the local port is placed in Selected state again upon link recovery, the local port establishes a new session with the peer port and BFD notifies the Ethernet link aggregation module that the peer port is reachable. Because BFD provides fast failure detection, the local and peer systems of a dynamic aggregate link can negotiate the aggregation state of their member ports faster.

For more information about BFD, see *High Availability Configuration Guide*.

Restrictions and guidelines

When you enable BFD for an aggregation group, follow these restrictions and guidelines:

- Make sure the source and destination IP addresses are reversed between the two ends of an aggregate link. For example, if you execute **link-aggregation bfd ipv4 source 1.1.1.1 destination 2.2.2.2** at the local end, execute **link-aggregation bfd ipv4 source 2.2.2.2 destination 1.1.1.1** at the peer end. The source and destination IP addresses cannot be the same.
- The BFD parameters configured on an aggregate interface take effect on all BFD sessions established by the member ports in its aggregation group. BFD on a link aggregation supports only control packet mode for session establishment and maintenance. The two ends of an established BFD session can only operate in **Asynchronous** mode.
- As a best practice, do not configure BFD for any protocols on a BFD-enabled aggregate interface.

- Make sure the number of member ports in a BFD-enabled aggregation group is less than or identical to the number of BFD sessions supported by the device. If the aggregation group contains more member ports than the supported sessions, some Selected ports might change to the Unselected state.
- If the number of BFD sessions differs between the two ends of an aggregate link, check their settings for inconsistency in the maximum number of Selected ports. You must make sure the two ends have the same setting for the maximum number of Selected ports.
- If both IPv4 and IPv6 BFD sessions exist on an aggregate interface, the device places the aggregation member ports in Unselected state when either IPv4 or IPv6 BFD detects a link failure.
- In the current software version, you cannot enable BFD for link aggregation on a Layer 2 aggregate interface.

Procedure

1. Enter system view.
   system-view
2. Enter aggregate interface view.
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. (Optional.) Enable BFD compatible mode for link aggregation.
   link-aggregation bfd-compatible enable
   By default, BFD compatible mode for link aggregation is disabled.
4. Enable BFD for the aggregation group.
   link-aggregation bfd ipv4 source *ipv4-address* **destination** *ipv4-address*
   link-aggregation bfd ipv6 source *ipv6-address* **destination** *ipv6-address*
   By default, BFD is disabled for an aggregation group.

Display and maintenance commands for Ethernet link aggregation

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about aggregate interfaces.	**display interface** [{ **bridge-aggregation** | **route-aggregation** } [*interface-number*]] [**brief** [**description** | **down**]]
Display the local system ID.	**display lacp system-id**
Display the global or group-specific link-aggregation load sharing modes.	**display link-aggregation load-sharing mode** [**interface** [{ **bridge-aggregation** | **route-aggregation** } *interface-number*]]

Task	Command
Display detailed link aggregation information about link aggregation member ports.	**display link-aggregation member-port** [*interface-list* | **auto**]
Display summary information about all aggregation groups.	**display link-aggregation summary**
Display the aggregation states of aggregation member ports and the reason why a port was placed in Unselected state.	**display link-aggregation troubleshooting** [{ **bridge-aggregation** | **route-aggregation** } [*interface-number*]]
Display detailed information about the specified aggregation groups.	**display link-aggregation verbose** [{ **bridge-aggregation** | **route-aggregation** } [*interface-number*]] [**all-configuration**]
Clear statistics for the specified aggregate interfaces.	**reset counters interface** [{ **bridge-aggregation** | **route-aggregation** } [*interface-number*]]
Clear LACP statistics for the specified link aggregation member ports.	**reset lacp statistics** [**interface** *interface-list*]

Ethernet link aggregation configuration examples

Example: Configuring a Layer 2 static aggregation group

Network configuration

On the network shown in Figure 8, perform the following tasks:

- Configure a Layer 2 static aggregation group on both Device A and Device B.
- Enable VLAN 10 at one end of the aggregate link to communicate with VLAN 10 at the other end.
- Enable VLAN 20 at one end of the aggregate link to communicate with VLAN 20 at the other end.

Figure 8 Network diagram

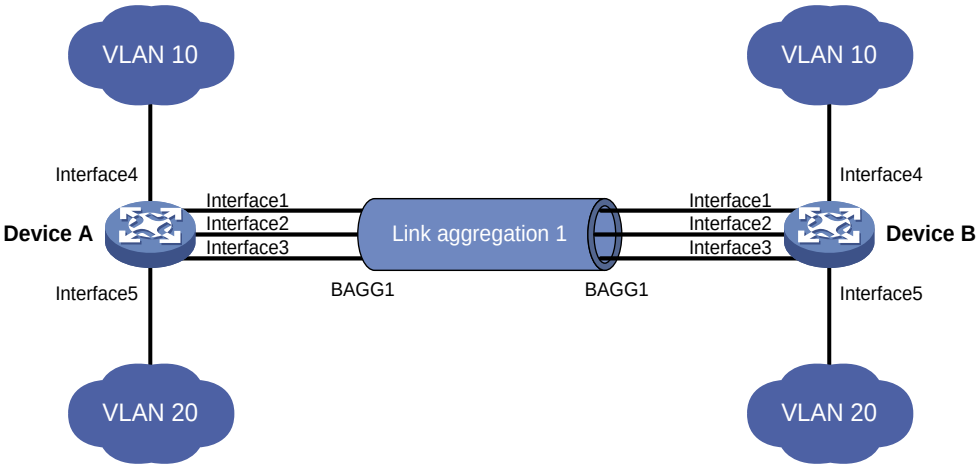


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

1. Configure Device A:

Create VLAN 10, and assign port GigabitEthernet 1/0/4 to VLAN 10.

```
<DeviceA> system-view
[DeviceA] vlan 10
[DeviceA-vlan10] port gigabitethernet 1/0/4
[DeviceA-vlan10] quit
```

Create VLAN 20, and assign port GigabitEthernet 1/0/5 to VLAN 20.

```
[DeviceA] vlan 20
[DeviceA-vlan20] port gigabitethernet 1/0/5
[DeviceA-vlan20] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 1.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] quit
```

Assign ports GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to link aggregation group 1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/2] quit
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/3] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 1 as a trunk port and assign it to VLANs 10 and 20.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] port link-type trunk
[DeviceA-Bridge-Aggregation1] port trunk permit vlan 10 20
[DeviceA-Bridge-Aggregation1] quit
```

2. Configure Device B in the same way Device A is configured. (Details not shown.)

Verifying the configuration

Display detailed information about all aggregation groups on Device A.

```
[DeviceA] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregate Interface: Bridge-Aggregation1
Aggregation Mode: Static
Loadsharing Type: Shar
Management VLANs: None
  Port             Status  Priority Oper-Key
  GE1/0/1(R)       S       32768    1
  GE1/0/2          S       32768    1
  GE1/0/3          S       32768    1
```

The output shows that link aggregation group 1 is a Layer 2 static aggregation group that contains three Selected ports.

Example: Configuring a Layer 2 dynamic aggregation group

Network configuration

On the network shown in Figure 9, perform the following tasks:

- Configure a Layer 2 dynamic aggregation group on both Device A and Device B.
- Enable VLAN 10 at one end of the aggregate link to communicate with VLAN 10 at the other end.
- Enable VLAN 20 at one end of the aggregate link to communicate with VLAN 20 at the other end.

Figure 9 Network diagram

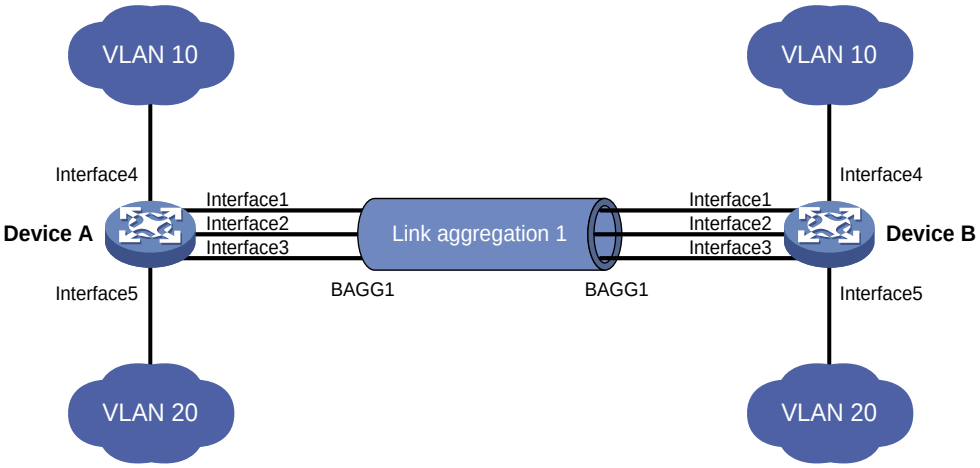


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

1. Configure Device A:

Create VLAN 10, and assign the port GigabitEthernet 1/0/4 to VLAN 10.

```
<DeviceA> system-view
[DeviceA] vlan 10
[DeviceA-vlan10] port gigabitethernet 1/0/4
[DeviceA-vlan10] quit
```

Create VLAN 20, and assign the port GigabitEthernet 1/0/5 to VLAN 20.

```
[DeviceA] vlan 20
[DeviceA-vlan20] port gigabitethernet 1/0/5
[DeviceA-vlan20] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 1, and set the link aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation1] quit
```

Assign ports GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to link aggregation group 1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/2] quit
[DeviceA] interface gigabitethernet 1/0/3
```

```
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/3] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 1 as a trunk port and assign it to VLANs 10 and 20.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] port link-type trunk
[DeviceA-Bridge-Aggregation1] port trunk permit vlan 10 20
[DeviceA-Bridge-Aggregation1] quit
```

2. Configure Device B in the same way Device A is configured. (Details not shown.)

Verifying the configuration

Display detailed information about all aggregation groups on Device A.

```
[DeviceA] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregate Interface: Bridge-Aggregation1
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0x8000, 000f-e267-6c6a
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/1(R)          S        32768    11       1                      {ACDEF}
  GE1/0/2             S        32768    12       1                      {ACDEF}
  GE1/0/3             S        32768    13       1                      {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/1             32768    81       1        0x8000, 000f-e267-57ad {ACDEF}
  GE1/0/2             32768    82       1        0x8000, 000f-e267-57ad {ACDEF}
  GE1/0/3             32768    83       1        0x8000, 000f-e267-57ad {ACDEF}
```

The output shows that link aggregation group 1 is a Layer 2 dynamic aggregation group that contains three Selected ports.

Example: Configuring a Layer 2 edge aggregate interface

Network configuration

As shown in Figure 10, a Layer 2 dynamic aggregation group is configured on the device. The server is not configured with dynamic link aggregation.

Configure an edge aggregate interface so that both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 can forward traffic to improve link reliability.

Figure 10 Network diagram



Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

Create Layer 2 aggregate interface Bridge-Aggregation 1, and set the link aggregation mode to dynamic.

```
<Device> system-view
[Device] interface bridge-aggregation 1
[Device-Bridge-Aggregation1] link-aggregation mode dynamic
```

Configure Layer 2 aggregate interface Bridge-Aggregation 1 as an edge aggregate interface.

```
[Device-Bridge-Aggregation1] lacp edge-port
[Device-Bridge-Aggregation1] quit
```

Assign ports GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to link aggregation group 1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] port link-aggregation group 1
[Device-GigabitEthernet1/0/1] quit
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] port link-aggregation group 1
[Device-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Display detailed information about all aggregation groups on the device when the server is not configured with dynamic link aggregation.

```
[Device] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregate Interface: Bridge-Aggregation1
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0x8000, 000f-e267-6c6a
Local:
  Port               Status   Priority Index    Oper-Key               Flag
```

```
  GE1/0/1             I        32768    11       1                     {AG}
  GE1/0/2             I        32768    12       1                     {AG}

Remote:
  Actor               Priority Index    Oper-Key SystemID              Flag
  GE1/0/1             32768    81       0        0x8000, 0000-0000-0000 {DEF}
  GE1/0/2             32768    82       0        0x8000, 0000-0000-0000 {DEF}
```

The output shows that GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are in Individual state when they do not receive LACPDUs from the server. Both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 can forward traffic. When one port fails, its traffic is automatically switched to the other port.

Example: Configuring a Layer 3 static aggregation group

Network configuration

On the network shown in Figure 11, perform the following tasks:

- Configure a Layer 3 static aggregation group on both Device A and Device B.
- Configure IP addresses and subnet masks for the corresponding Layer 3 aggregate interfaces.

Figure 11 Network diagram

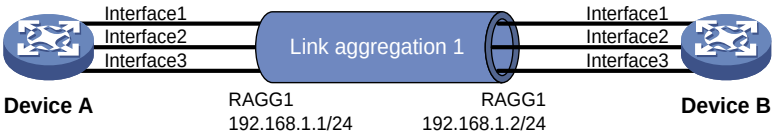


Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Device A:

   # Create Layer 3 aggregate interface Route-Aggregation 1, and configure an IP address and subnet mask for the aggregate interface.

   ```
   <DeviceA> system-view
   [DeviceA] interface route-aggregation 1
   [DeviceA-Route-Aggregation1] ip address 192.168.1.1 24
   [DeviceA-Route-Aggregation1] quit
   ```

   # Assign Layer 3 Ethernet interfaces GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to aggregation group 1.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] port link-aggregation group 1
   [DeviceA-GigabitEthernet1/0/1] quit
   [DeviceA] interface gigabitethernet 1/0/2
   [DeviceA-GigabitEthernet1/0/2] port link-aggregation group 1
   [DeviceA-GigabitEthernet1/0/2] quit
   [DeviceA] interface gigabitethernet 1/0/3
   ```

```
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/3] quit
```

2. Configure Device B in the same way Device A is configured. (Details not shown.)

Verifying the configuration

Display detailed information about all aggregation groups on Device A.

```
[DeviceA] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregate Interface: Route-Aggregation1
Aggregation Mode: Static
Loadsharing Type: Shar
Management VLANs: None
  Port            Status  Priority Oper-Key
  GE1/0/1(R)      S       32768    1
  GE1/0/2         S       32768    1
  GE1/0/3         S       32768    1
```

The output shows that link aggregation group 1 is a Layer 3 static aggregation group that contains three Selected ports.

Example: Configuring a Layer 3 dynamic aggregation group

Network configuration

On the network shown in Figure 12, perform the following tasks:

- Configure a Layer 3 dynamic aggregation group on both Device A and Device B.
- Configure IP addresses and subnet masks for the corresponding Layer 3 aggregate interfaces.

Figure 12 Network diagram

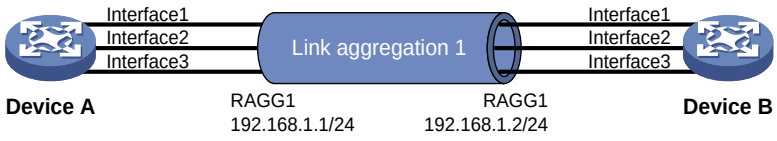


Table 9 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Device A:

   # Create Layer 3 aggregate interface Route-Aggregation 1.

```
<DeviceA> system-view
[DeviceA] interface route-aggregation 1
```

Set the link aggregation mode to dynamic.

```
[DeviceA-Route-Aggregation1] link-aggregation mode dynamic
```

Configure an IP address and subnet mask for Route-Aggregation 1.

```
[DeviceA-Route-Aggregation1] ip address 192.168.1.1 24
[DeviceA-Route-Aggregation1] quit
```

Assign Layer 3 Ethernet interfaces GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to aggregation group 1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/2] quit
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/3] quit
```

2. Configure Device B in the same way Device A is configured. (Details not shown.)

Verifying the configuration

Display detailed information about all aggregation groups on Device A.

```
[DeviceA] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregate Interface: Route-Aggregation1
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0x8000, 000f-e267-6c6a
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/1(R)          S        32768    11       1                      {ACDEF}
  GE1/0/2             S        32768    12       1                      {ACDEF}
  GE1/0/3             S        32768    13       1                      {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/1             32768    81       1        0x8000, 000f-e267-57ad {ACDEF}
  GE1/0/2             32768    82       1        0x8000, 000f-e267-57ad {ACDEF}
  GE1/0/3             32768    83       1        0x8000, 000f-e267-57ad {ACDEF}
```

The output shows that link aggregation group 1 is a Layer 3 dynamic aggregation group that contains three Selected ports.

Example: Configuring S-MLAG

Network configuration

To meet service high availability requirements, configure link aggregation between Device A/Device B and Server 1/Server 2. This provides device-level redundancy for the servers and implements traffic load sharing.

- Device A and Device B connect to Server 1 and Server 2 through Layer 2 Ethernet interfaces GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2, respectively.
- Configure S-MLAG on Device A and Device B. When both Device A and Device B run correctly, they forward traffic simultaneously. When one device fails, the other device takes over all service traffic.
- Server 1 and Server 2 can communicate at Layer 2.

Figure 13 Network diagram

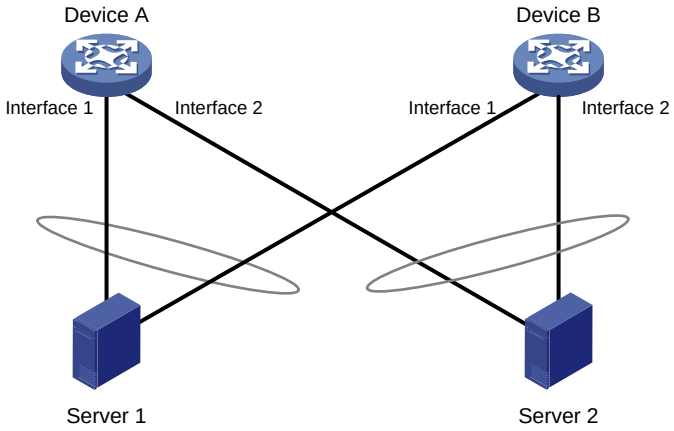


Table 10 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

Set the LACP system MAC address to 0000-5e00-0101.

```
<DeviceA> system-view
[DeviceA] lacp system-mac 0000-5e00-0101
```

Set the LACP system priority to 200.

```
[DeviceA] lacp system-priority 200
```

Set the LACP system number to 1.

```
[DeviceA] lacp system-number 1
```

Create Layer 2 aggregate interface Bridge-Aggregation 1, and set the link aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] link-aggregation mode dynamic
```

Assign Bridge-Aggregation 1 to S-MLAG group 100.

```
[DeviceA-Bridge-Aggregation1] port s-mlag group 100
```

Assign GigabitEthernet 1/0/1 to aggregation group 1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/1] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 2, and set the link aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 2
[DeviceA-Bridge-Aggregation2] link-aggregation mode dynamic
```

Assign Bridge-Aggregation 2 to S-MLAG group 200.

```
[DeviceA-Bridge-Aggregation2] port s-mlag group 200
```

Assign GigabitEthernet 1/0/2 to aggregation group 2.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 2
[DeviceA-GigabitEthernet1/0/2] quit
```

Create VLAN 100.

```
[DeviceA] vlan 100
[DeviceA-vlan100] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 1 as a trunk port, assign the trunk port to VLAN 100, and set the PVID to 100.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] port link-type trunk
[DeviceA-Bridge-Aggregation1] port trunk permit vlan 100
[DeviceA-Bridge-Aggregation1] port trunk pvid vlan 100
```

Remove the trunk port from VLAN 1.

```
[DeviceA-Bridge-Aggregation1] undo port trunk permit vlan 1
[DeviceA-Bridge-Aggregation1] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 2 as a trunk port, assign the trunk port to VLAN 100, and set the PVID to 100.

```
[DeviceA] interface bridge-aggregation 2
[DeviceA-Bridge-Aggregation2] port link-type trunk
[DeviceA-Bridge-Aggregation2] port trunk permit vlan 100
[DeviceA-Bridge-Aggregation2] port trunk pvid vlan 100
```

Remove the trunk port from VLAN 1.

```
[DeviceA-Bridge-Aggregation2] undo port trunk permit vlan 1
[DeviceA-Bridge-Aggregation2] quit
```

2. Configure Device B:

Set the LACP system MAC address to 0000-5e00-0101.

```
<DeviceB> system-view
[DeviceB] lacp system-mac 0000-5e00-0101
```

Set the LACP system priority to 200.

```
[DeviceB] lacp system-priority 200
```

Set the LACP system number to 2.

```
[DeviceB] lacp system-number 2
```

Create Layer 2 aggregate interface Bridge-Aggregation 1, and set the link aggregation mode to dynamic.

```
[DeviceB] interface bridge-aggregation 1
[DeviceB-Bridge-Aggregation1] link-aggregation mode dynamic
```

Assign Bridge-Aggregation 1 to S-MLAG group 100.

```
[DeviceB-Bridge-Aggregation1] port s-mlag group 100
```

Assign GigabitEthernet 1/0/1 to aggregation group 1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-aggregation group 1
[DeviceB-GigabitEthernet1/0/1] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 2, and set the link aggregation mode to dynamic.

```
[DeviceB] interface bridge-aggregation 2
[DeviceB-Bridge-Aggregation2] link-aggregation mode dynamic
```

Assign Bridge-Aggregation 2 to S-MLAG group 200.

```
[DeviceB-Bridge-Aggregation2] port s-mlag group 200
```

Assign GigabitEthernet 1/0/2 to aggregation group 2.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-aggregation group 2
[DeviceB-GigabitEthernet1/0/2] quit
```

Create VLAN 100.

```
[DeviceB] vlan 100
[DeviceB-vlan100] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 1 as a trunk port, assign the trunk port to VLAN 100, and set the PVID to 100.

```
[DeviceB] interface bridge-aggregation 1
[DeviceB-Bridge-Aggregation1] port link-type trunk
[DeviceB-Bridge-Aggregation1] port trunk permit vlan 100
[DeviceB-Bridge-Aggregation1] port trunk pvid vlan 100
```

Remove the trunk port from VLAN 1.

```
[DeviceB-Bridge-Aggregation1] undo port trunk permit vlan 1
[DeviceB-Bridge-Aggregation1] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 2 as a trunk port, assign the trunk port to VLAN 100, and set the PVID to 100.

```
[DeviceB] interface bridge-aggregation 2
[DeviceB-Bridge-Aggregation2] port link-type trunk
[DeviceB-Bridge-Aggregation2] port trunk permit vlan 100
[DeviceB-Bridge-Aggregation2] port trunk pvid vlan 100
```

Remove the trunk port from VLAN 1.

```
[DeviceB-Bridge-Aggregation2] undo port trunk permit vlan 1
[DeviceB-Bridge-Aggregation2] quit
```

Verifying the configuration

View detailed information about all aggregation groups on Device A to verify that the S-MLAG aggregation group members are in **Selected** state.

```
[DeviceA] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired
Role: P -- Primary, S -- Secondary
```

```
Aggregate Interface: Bridge-Aggregation1
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0xc8, 0000-5e00-0101
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/1(R)          S        32768    16385    50001                  {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/1             32768    1        1        0x8000, 1451-7eae-dafe {ACDEF}

Aggregate Interface: Bridge-Aggregation2
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0xc8, 0000-5e00-0101
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/2(R)          S        32768    16386    50002                  {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/2             32768    1        1        0x8000, 5cc9-995f-1122 {ACDEF}
```

View detailed information about all aggregation groups on Device B to verify that the S-MLAG aggregation group members are in **Selected** state.

```
[DeviceB] display link-aggregation verbose
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected, I -- Individual
Port: A -- Auto port, M -- Management port, R -- Reference port
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired
Role: P -- Primary, S -- Secondary

Aggregate Interface: Bridge-Aggregation1
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0xc8, 0000-5e00-0101
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/1(R)          S        32768    32769    50001                  {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/1             32768    2        1        0x8000, 1451-7eae-dafe {ACDEF}
```

```
Aggregate Interface: Bridge-Aggregation2
Creation Mode: Manual
Aggregation Mode: Dynamic
Loadsharing Type: Shar
Management VLANs: None
System ID: 0xc8, 0000-5e00-0101
Local:
  Port                Status   Priority Index    Oper-Key               Flag
  GE1/0/2(R)          S        32768    32770    50002                  {ACDEF}
Remote:
  Actor               Priority Index    Oper-Key SystemID               Flag
  GE1/0/2             32768    2        1        0x8000, 5cc9-995f-1122 {ACDEF}
```

Configuration files

- Device A:

```
#
 lacp system-mac 0000-5e00-0101
 lacp system-number 1
 lacp system-priority 200
#
vlan 100
#
interface Bridge-Aggregation1
 port link-type trunk
 undo port trunk permit vlan 1
 port trunk permit vlan 100
 port trunk pvid vlan 100
 link-aggregation mode dynamic
 port s-mlag group 100
#
interface Bridge-Aggregation2
 port link-type trunk
 undo port trunk permit vlan 1
 port trunk permit vlan 100
 port trunk pvid vlan 100
 link-aggregation mode dynamic
 port s-mlag group 200
#
interface GigabitEthernet1/0/1
 port link-type trunk
 undo port trunk permit vlan 1
 port trunk permit vlan 100
 port trunk pvid vlan 100
 port link-aggregation group 1
#
interface GigabitEthernet1/0/2
 port link-type trunk
 undo port trunk permit vlan 1
```

```
 port trunk permit vlan 100
 port trunk pvid vlan 100
 port link-aggregation group 2
#
```

- Device B:

```
#
 lacp system-mac 0000-5e00-0101
 lacp system-number 2
 lacp system-priority 200
#
vlan 100
#
interface Bridge-Aggregation1
 port link-type trunk
 undo port trunk permit vlan 1
 port trunk permit vlan 100
 port trunk pvid vlan 100
 link-aggregation mode dynamic
 port s-mlag group 100
#
interface Bridge-Aggregation2
 port link-type trunk
 undo port trunk permit vlan 1
 port trunk permit vlan 100
 port trunk pvid vlan 100
 link-aggregation mode dynamic
 port s-mlag group 200
#
interface GigabitEthernet1/0/1
 port link-type trunk
 undo port trunk permit vlan 1
 port trunk permit vlan 100
 port trunk pvid vlan 100
 port link-aggregation group 1
#
interface GigabitEthernet1/0/2
 port link-type trunk
 undo port trunk permit vlan 1
 port trunk permit vlan 100
 port trunk pvid vlan 100
 port link-aggregation group 2
#
```

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Layer 3—IP Services Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the IP services fundamentals and configuration procedures that help you complete the IP address configuration, including ARP configuration, IP addressing configuration, DHCP configuration, and DNS configuration.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** >

	Folder.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring ARP 3
About ARP 3
ARP message format 3
ARP operating mechanism 3
ARP entry types 4
ARP tasks at a glance 5
Configuring a static ARP entry 6
Configuring a short static ARP entry 6
Configuring a long static ARP entry 6
Configuring a multiport ARP entry 7
Configuring features for dynamic ARP entries 7
Setting the dynamic ARP learning limit for a device 7
Setting the dynamic ARP learning limit for an interface 8
Setting the aging timer for dynamic ARP entries 8
Setting the maximum number of probes for dynamic ARP entries 9
Setting the interval for probing dynamic ARP entries 9
Enabling dynamic ARP entry check 10
Synchronizing ARP entries across all member devices 11
Enabling recording user IP address conflicts 11
Enabling interface consistency check between ARP and MAC address entries 12
Enabling recording user port migrations 12
Enabling ARP entry consistency check 13
Enabling unknown source MAC-triggered ARP probing 13
Enabling ARP logging 14
Enabling error logging for ARP entry deployment to hardware 14
Enabling the device to send only ARP broadcast requests destined for a VLAN interface on it to the CPU 15
Disabling the device from aging or deleting ARP entries in response to network topology changes 15
Disabling sending ARP requests when data packets trigger ARP resolution 16
Display and maintenance commands for ARP 16
ARP configuration examples 17
Example: Configuring a long static ARP entry 17
Example: Configuring a short static ARP entry 18
Configuring gratuitous ARP 19
About gratuitous ARP 19
IP conflict detection 19
Gratuitous ARP packet learning 19
Periodic sending of gratuitous ARP packets 20
Gratuitous ARP tasks at a glance 20
Enabling notification for passive IP conflict detection 21
Enabling gratuitous ARP packet learning 21
Enabling periodic sending of gratuitous ARP packets 21
Enabling sending gratuitous ARP packets for ARP requests with sender IP address on a different subnet 22
Configuring gratuitous ARP packet retransmission for the device MAC address change 22
Configuring gratuitous ARP packet forwarding for a VSI 23
Configuring proxy ARP 25
About proxy ARP 25
Enabling common proxy ARP 25
Enabling local proxy ARP 25
Display and maintenance commands for proxy ARP 26
Common proxy ARP configuration example 26
Example: Configuring common proxy ARP 26
Configuring ARP snooping 28
About ARP snooping 28
Creation of ARP snooping entries 28

Aging of ARP snooping entries ····· 28
Protection for ARP snooping ····· 28
Restrictions and guidelines: ARP snooping configuration ····· 28
Enabling ARP snooping for a VLAN ····· 29
Enabling ARP snooping for a VSI ····· 29
Enabling ARP snooping for an Layer 2 interface ····· 29
Configuring an interface as an uplink interface to disable it from learning ARP snooping entries ····· 29
Display and maintenance commands for ARP snooping ····· 30
Configuring ARP fast-reply ····· 31
About ARP fast-reply ····· 31
Enabling ARP fast-reply ····· 31
Enabling ARP gateway proxy ····· 31
ARP fast-reply configuration example ····· 32
Example: Configuring ARP fast-reply ····· 32
Configuring ARP direct route advertisement ····· 34
Configuring ARP direct route advertisement ····· 34
About ARP direct route advertisement ····· 34
Restrictions and guidelines ····· 34
Application in Layer 3 access networks ····· 34
Procedure ····· 34
Enabling ARP direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state ····· 35
Display and maintenance commands for ARP direct route advertisement ····· 36
Configuring ARP ping ····· 37
Configuring ARP ping for IPv4 addresses ····· 37
Configuring ARP ping for MAC addresses ····· 37

Configuring ARP

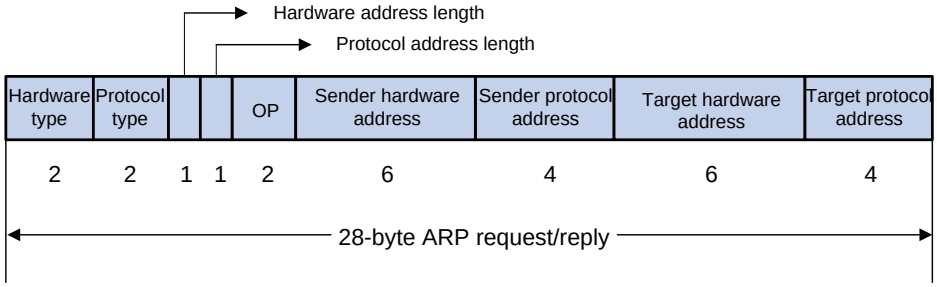
About ARP

ARP resolves IP addresses into MAC addresses on Ethernet networks.

ARP message format

ARP uses two types of messages: ARP request and ARP reply. Figure 1 shows the format of ARP request/reply messages. Numbers in the figure refer to field lengths.

Figure 1 ARP message format



- **Hardware type**—Hardware address type. The value 1 represents Ethernet.
- **Protocol type**—Type of the protocol address to be mapped. The hexadecimal value 0x0800 represents IP.
- **Hardware address length and protocol address length**—Length, in bytes, of a hardware address and a protocol address. For an Ethernet address, the value of the hardware address length field is 6. For an IPv4 address, the value of the protocol address length field is 4.
- **OP**—Operation code, which describes the type of ARP message. The value 1 represents an ARP request, and the value 2 represents an ARP reply.
- **Sender hardware address**—Hardware address of the device sending the message.
- **Sender protocol address**—Protocol address of the device sending the message.
- **Target hardware address**—Hardware address of the device to which the message is being sent.
- **Target protocol address**—Protocol address of the device to which the message is being sent.

ARP operating mechanism

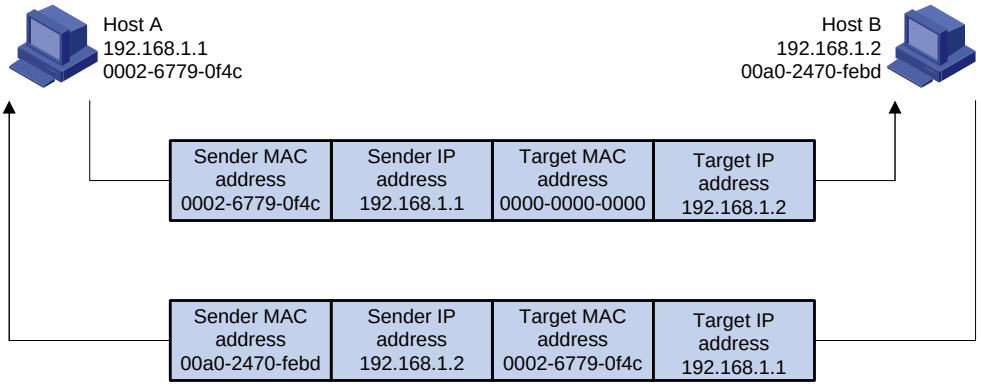
As shown in Figure 2, Host A and Host B are on the same subnet. Host A sends a packet to Host B as follows:

1. Host A looks through the ARP table for an ARP entry for Host B. If one entry is found, Host A uses the MAC address in the entry to encapsulate the IP packet into a data link layer frame. Then Host A sends the frame to Host B.
2. If Host A finds no entry for Host B, Host A buffers the packet and broadcasts an ARP request. The payload of the ARP request contains the following information:
   - **Sender IP address and sender MAC address**—Host A's IP address and MAC address.
   - **Target IP address**—Host B's IP address.
   - **Target MAC address**—An all-zero MAC address.

All hosts on this subnet can receive the broadcast request, but only the requested host (Host B) processes the request.

3. Host B compares its own IP address with the target IP address in the ARP request. If they are the same, Host B operates as follows:
   a. Adds the sender IP address and sender MAC address into its ARP table.
   b. Encapsulates its MAC address into an ARP reply.
   c. Unicasts the ARP reply to Host A.
4. After receiving the ARP reply, Host A operates as follows:
   a. Adds the MAC address of Host B into its ARP table.
   b. Encapsulates the MAC address into the packet and sends the packet to Host B.

Figure 2 ARP address resolution process



If Host A and Host B are on different subnets, Host A sends a packet to Host B as follows:

5. Host A broadcasts an ARP request where the target IP address is the IP address of the gateway.
6. The gateway responds with its MAC address in an ARP reply to Host A.
7. Host A uses the gateway's MAC address to encapsulate the packet, and then sends the packet to the gateway.
8. If the gateway has an ARP entry for Host B, it forwards the packet to Host B directly. If not, the gateway broadcasts an ARP request, in which the target IP address is the IP address of Host B.
9. After the gateway gets the MAC address of Host B, it sends the packet to Host B.

ARP entry types

An ARP table stores dynamic ARP entries, OpenFlow ARP entries, Rule ARP entries, and static ARP entries.

Dynamic ARP entry

ARP automatically creates and updates dynamic entries. A dynamic ARP entry is removed when its aging timer expires or the output interface goes down. In addition, a dynamic ARP entry can be overwritten by a static ARP entry.

Static ARP entry

A static ARP entry is manually configured and maintained. It does not age out and cannot be overwritten by any dynamic ARP entry.

Static ARP entries protect communication between devices because attack packets cannot modify the IP-to-MAC mapping in a static ARP entry.

The device supports the following types of static ARP entries:

- **Long static ARP entry**—It is directly used for forwarding packets. A long static ARP entry contains the IP address, MAC address, VLAN, and output interface.
- **Short static ARP entry**—It contains only the IP address and MAC address.

  If the output interface is a Layer 3 Ethernet interface, the short ARP entry can be directly used to forward packets.

  If the output interface is a VLAN interface, the device sends an ARP request whose target IP address is the IP address in the short entry. If the sender IP and MAC addresses in the received ARP reply match the short static ARP entry, the device performs the following operations:
  - Adds the interface that received the ARP reply to the short static ARP entry.
  - Uses the resolved short static ARP entry to forward IP packets.
- **Multiport ARP entry**—It contains the IP address, MAC address, and VLAN/VSI information.

  The device can use a multiport ARP entry that has the same MAC address and VLAN/VSI as a multicast or multiport unicast MAC address entry for packet forwarding. A multiport ARP entry is manually configured. It does not age out and cannot be overwritten by any dynamic ARP entry.

To communicate with a host by using a fixed IP-to-MAC mapping, configure a short static ARP entry on the device. To communicate with a host by using a fixed IP-to-MAC mapping through an interface in a VLAN, configure a long static ARP entry on the device.

OpenFlow ARP entry

ARP creates OpenFlow ARP entries by learning from the OpenFlow module. An OpenFlow ARP entry does not age out, and it cannot be updated. An OpenFlow ARP entry can be used directly to forward packets. For more information about OpenFlow, see *OpenFlow Configuration Guide*.

Rule ARP entry

Rule ARP entries can be directly used for packet forwarding. A Rule ARP entry does not age out, and it cannot be updated. It can be overwritten by a static ARP entry.

ARP creates Rule ARP entries by learning from the following modules:

- Portal. For more information about portal, see *Security Configuration Guide*.

ARP tasks at a glance

All ARP tasks are optional.

- Configuring a static ARP entry
  - Configuring a short static ARP entry
  - Configuring a long static ARP entry
  - Configuring a multiport ARP entry
- Configuring features for dynamic ARP entries
  - Setting the dynamic ARP learning limit for a device
  - Setting the dynamic ARP learning limit for an interface
  - Setting the aging timer for dynamic ARP entries
  - Setting the maximum number of probes for dynamic ARP entries
  - Setting the interval for probing dynamic ARP entries
  - Enabling dynamic ARP entry check
- Synchronizing ARP entries across all member devices
- Enabling user information checking for ARP entries
  - Enabling recording user IP address conflicts
  - Enabling interface consistency check between ARP and MAC address entries

    - Enabling recording user port migrations
    - Enabling ARP entry consistency check
- Enabling unknown source MAC-triggered ARP probing
- Enabling ARP logging
- Enabling error logging for ARP entry deployment to hardware
- Enabling the device to send only ARP broadcast requests destined for a VLAN interface on it to the CPU
- (Optional.) Disabling the device from aging or deleting ARP entries in response to network topology changes
- (Optional.) Disabling sending ARP requests when data packets trigger ARP resolution

Configuring a static ARP entry

Static ARP entries are effective when the device functions correctly.

Configuring a short static ARP entry

Restrictions and guidelines

A resolved short static ARP entry becomes unresolved upon certain events, for example, when the resolved output interface goes down, or the corresponding VLAN or VLAN interface is deleted.

Procedure

1. Enter system view.

   system-view
2. Configure a short static ARP entry.

   arp static *ip-address mac-address* [**vpn-instance** *vpn-instance-name*] [**description** *text*]

Configuring a long static ARP entry

About this task

Long static ARP entries can be effective or ineffective. Ineffective long static ARP entries cannot be used for packet forwarding. A long static ARP entry is ineffective when any of the following conditions exists:

- The IP address in the entry conflicts with a local IP address.
- No local interface has an IP address in the same subnet as the IP address in the ARP entry.

A long static ARP entry for a VLAN is deleted if the VLAN or VLAN interface is deleted.

Procedure

1. Enter system view.

   system-view
2. Configure a long static ARP entry.

   arp static *ip-address mac-address* [*vlan-id interface-type interface-number* | *interface-type interface-number interface-type interface-number* | **vsi-interface** *vsi-interface-id* **tunnel** *number* **vsi** *vsi-name* | **vsi-interface** *vsi-interface-id interface-type interface-number* **service-instance** *instance-id* **vsi** *vsi-name*] [**vpn-instance** *vpn-instance-name*] [**description** *text*]

Configuring a multiport ARP entry

About this task

A multiport ARP entry contains an IP address, MAC address, output interface, and VLAN ID/VSI information. The VLAN and output interfaces are specified by a multiport unicast MAC address entry or a multicast MAC address entry. For more information about multiport unicast MAC address entries, see *Layer 2—LAN Switching Configuration Guide*. For more information about multicast MAC address entries, see IGMP snooping configuration in *IP Multicast Configuration Guide*.

A multiport ARP entry can overwrite a dynamic, short static or long static ARP entry. Conversely, a short static or long static ARP entry can overwrite a multiport ARP entry.

Restrictions and guidelines

For a multiport ARP entry to be effective for packet forwarding, make sure the following conditions are met:

- A multiport unicast MAC address entry or a multicast MAC address entry exists.
- The multiport ARP entry must have the same MAC address and VLAN/VSI setting as the multiport unicast MAC address entry or the multicast MAC address entry.
- The IP address in the multiport ARP entry must reside on the same subnet as the VLAN interface of the specified VLAN.

Procedure

1. Enter system view.

   system-view
2. Configure a multiport unicast MAC address entry or a multicast MAC address entry.
   - Configure a multiport unicast MAC address entry.

     mac-address multiport *mac-address* **interface** *interface-list* **vlan** *vlan-id*
   - Configure a multicast MAC address entry.

     mac-address multicast *mac-address* **interface** *interface-list* **vlan** *vlan-id*
3. Configure a multiport ARP entry.

   arp multiport *ip-address mac-address vlan-id* [**vpn-instance** *vpn-instance-name*] [**description** *text*]

Configuring features for dynamic ARP entries

Setting the dynamic ARP learning limit for a device

About this task

A device can dynamically learn ARP entries. To prevent a device from holding too many ARP entries, you can set the maximum number of dynamic ARP entries that the device can learn. When the limit is reached, the device stops ARP learning.

If you set a value lower than the number of existing dynamic ARP entries, the device does not delete the existing entries unless they age out. You can use the **reset arp dynamic** command to clear dynamic ARP entries.

Procedure

1. Enter system view.

   system-view

2. Set the dynamic ARP learning limit for the device.

   arp max-learning-number *max-number* **slot** *slot-number*

   By default, the dynamic ARP learning limit for the device depends on the configuration of the **switch-mode** command. For information about the **switch-mode** command, see the device management in *Fundamentals Command Reference*.

   To disable the device from dynamic ARP learning, set the value to 0.

Setting the dynamic ARP learning limit for an interface

About this task

An interface can dynamically learn ARP entries. To prevent an interface from holding too many ARP entries, you can set the maximum number of dynamic ARP entries that the interface can learn. When the limit is reached, the interface stops ARP learning.

You can set limits for both a Layer 2 interface and the VLAN interface for a permitted VLAN on the Layer 2 interface. The Layer 2 interface learns an ARP entry only when neither limit is reached.

The total dynamic ARP learning limit for all interfaces will not be higher than the dynamic ARP learning limit for the device.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the dynamic ARP learning limit for the interface.

   arp max-learning-num *max-number* [**alarm** *alarm-threshold*]

   By default, the maximum number of dynamic ARP entries that an interface can learn depends on the configuration of the **switch-mode** command. For information about the **switch-mode** command, see the device management in *Fundamentals Command Reference*.

   To disable the interface from dynamic ARP learning, set the value to 0.

Setting the aging timer for dynamic ARP entries

About this task

Each dynamic ARP entry in the ARP table has a limited lifetime, called an aging timer. The aging timer of a dynamic ARP entry is reset each time the dynamic ARP entry is updated. A dynamic ARP entry that is not updated before its aging timer expires is deleted from the ARP table.

You can set the aging timer for dynamic ARP entries in system view or in interface view. The aging timer set in interface view takes precedence over the aging timer set in system view.

Procedure

1. Enter system view.

   system-view

2. Set the aging timer for dynamic ARP entries.
   - Set the aging timer for dynamic ARP entries in system view.

     arp timer aging { *aging-minutes* | **second** *aging-seconds* }

     By default, the aging timer for dynamic ARP entries in system view is 20 minutes.

- Execute the following commands in sequence to set the aging timer for dynamic ARP entries in interface view:

  interface *interface-type interface-number*

  arp timer aging { *aging-minutes* | **second** *aging-seconds* }

  By default, the aging timer for dynamic ARP entries in interface view is the aging timer set in system view.

Setting the maximum number of probes for dynamic ARP entries

About this task

This probe mechanism keeps legal dynamic ARP entries valid and avoids unnecessary ARP resolution during later traffic forwarding. It sends ARP requests for the IP address in a dynamic ARP entry.

- If the device receives an ARP reply before the entry aging timer expires, the device resets the aging timer.
- If the device does not receive any ARP reply after the maximum number of probes is made, the device deletes the entry when the entry aging timer expires.

You can set the maximum number of probes in system view or in interface view. The setting in interface view takes precedence over that in system view.

Procedure

1. Enter system view.

   system-view

2. Set the maximum number of probes for dynamic ARP entries.
   - Set the maximum number of probes for dynamic ARP entries in system view.

     arp timer aging probe-count *count*

     By default, the maximum number of probes for dynamic ARP entries is 3 in system view.
   - Execute the following commands in sequence to set the maximum number of probes for dynamic ARP entries on an interface:

     interface *interface-type interface-number*

     arp timer aging probe-count *count*

     By default, the maximum number of probes for dynamic ARP entries set on an interface is that set in system view.

Setting the interval for probing dynamic ARP entries

About this task

The probing feature keeps legal dynamic ARP entries valid and avoids unnecessary ARP resolution during later traffic forwarding.

Before a dynamic ARP entry is aged out, the device sends ARP requests for the IP address in the ARP entry.

- If the device receives an ARP reply during the probe interval, the device resets the aging timer.
- If the device does not receive any ARP reply during the probe interval, the device starts a new probe.
- If the maximum number probes are made, and still no ARP reply is received, the device deletes the entry.

You can set the probe interval in system view and in interface view. The probe interval in interface view takes precedence over the probe interval in system view.

Restrictions and guidelines

- If massive traffic exists in the network, set a long interval.
- During the dynamic ARP entry probing process, a dynamic ARP entry will not be deleted when its aging time expires. If a reply is received during the probe, the aging timer of the ARP entry is reset.
- To allow the device to perform the specified number of probes, make sure the following requirement is met:

  Aging time of the dynamic ARP entries > the maximum number of probes × probe interval

Procedure

1. Enter system view.

   system-view
2. Set the interval for probing dynamic ARP entries.
   - Set the interval for probing dynamic ARP entries in system view.

     arp timer aging probe-interval *interval*

     By default, the probe interval is 5 seconds.
   - Execute the following commands in sequence to set the interval for probing dynamic ARP entries on an interface:

     interface *interface-type interface-number*

     arp timer aging probe-interval *interval*

     By default, the probe interval equals the setting in system view.

Enabling dynamic ARP entry check

About this task

The dynamic ARP entry check feature disables the device from supporting dynamic ARP entries that contain multicast MAC addresses. The device cannot learn dynamic ARP entries containing multicast MAC addresses. You cannot manually add static ARP entries containing multicast MAC addresses.

When dynamic ARP entry check is disabled, ARP entries containing multicast MAC addresses are supported. The device can learn dynamic ARP entries containing multicast MAC addresses obtained from the ARP packets sourced from a unicast MAC address. You can also manually add static ARP entries containing multicast MAC addresses.

Procedure

1. Enter system view.

   system-view
2. Enable dynamic ARP entry check.

   arp check enable

   By default, dynamic ARP entry check is enabled.

Synchronizing ARP entries across all member devices

About this task

This task ensures that all IRF member devices in an IRF fabric have the same ARP entries.

Restrictions and guidelines

To synchronize ARP entries across all member devices in a timely manner, you can schedule the device to automatically execute the **arp smooth** command. For information about scheduling a task, see the device management configuration in *Fundamentals Configuration Guide*.

Procedure

To synchronize ARP entries from the master device to all subordinate devices, execute the following command in user view:

arp smooth

Enabling recording user IP address conflicts

About this task

This feature enables the device to detect and record user IP address conflicts. The device determines that a conflict occurs if an incoming ARP packet has the same sender IP address as an existing ARP entry but a different sender MAC address. The device generates a user IP address conflict record, logs the conflict, and sends the log to the information center. For information about the log destination and output rule configuration in the information center, see the information center in *Network Management and Monitoring Configuration Guide*.

After you enable SNMP notifications for user IP address conflict events, the device sends a notification to the SNMP module when a user IP address conflict occurs. The notification includes the sender IP and MAC addresses in the conflicting ARP packet, and MAC address in the corresponding local ARP entry. For ARP event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*. For more information about SNMP notifications for ARP, see ARP attack protection configuration in *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable recording user IP address conflicts.

   arp user-ip-conflict record enable

   By default, recording user IP address conflicts is disabled.

3. (Optional.) Enable SNMP notifications for ARP.

   snmp-agent trap enable arp [**user-ip-conflict**]

   By default, SNMP notifications for ARP are disabled.

   If you do not specify the **user-ip-conflict** keyword, this command enables all types of SNMP notifications for ARP. If you specify the **user-ip-conflict** keyword, this command enables SNMP notifications for user IP address conflict events.

Enabling interface consistency check between ARP and MAC address entries

About this task

In an unstable network, the receiving interface for packets from a user might change. The interface in the MAC address entry can be updated immediately while the interface in the ARP entry cannot. In this case, the packets matching the ARP entry will be sent out of an incorrect interface. To solve this problem, you can use this feature to periodically check the interface consistency between the ARP and MAC address entry for a user. If the interfaces are not the same, ARP sends ARP requests in the VLAN of the ARP entry and updates the entry with the ARP reply receiving interface.

Use **display mac-address** to display MAC address entries. For more information about this command, see MAC address table in *Layer 2—LAN Switching Command Reference.*

Procedure

1. Enter system view.
   system-view
2. Enable interface consistency check between ARP and MAC address entries.
   arp mac-interface-consistency check enable
   By default, interface consistency check between ARP and MAC address entries is disabled.

Enabling recording user port migrations

About this task

This feature enables the device to detect and record user port migration events. A user port migrates if an incoming ARP packet has the same sender IP address and sender MAC address as an existing ARP entry but a different ingress port. The device generates a user port migration record, logs the migration event, sends the log to the information center, and updates the interface for the ARP entry. For information about the log destination and output rule configuration in the information center, see the information center in *Network Management and Monitoring Configuration Guide*.

After you enable SNMP notifications for user port migration events, the device sends a notification to the SNMP module when a user port migration occurs. The notification includes the IP and MAC addresses of the user who has a different ingress port, original port name, and new port name. For ARP event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*. For more information about SNMP notifications for ARP, see ARP attack protection configuration in *Security Configuration Guide*.

Restrictions and guidelines

To avoid device performance degradation, disable recording user port migrations if too many user port migration logs are generated.

Procedure

1. Enter system view.
   system-view
2. Enable recording user port migrations.
   arp user-move record enable
   By default, recording user port migrations is disabled.
3. (Optional.) Enable SNMP notifications for ARP.
   snmp-agent trap enable arp [**user-move**]

By default, SNMP notifications for ARP are disabled.

If you do not specify the **user-move** keyword, this command enables all types of SNMP notifications for ARP. If you specify the **user-move** keyword, this command enables SNMP notifications for user port migration events.

Enabling ARP entry consistency check

About this task

This feature enables the device to periodically deploy all ARP software entries to the driver for ARP software and hardware entry consistency check. If an inconsistency is identified, the device generates a log, and refreshes the hardware entry based on the software entry.

Procedure

1. Enter system view.

   system-view

2. Enable ARP entry consistency check between software and hardware entries.

   arp consistency-check enable

   By default, ARP entry consistency check between software and hardware entries is disabled.

Enabling unknown source MAC-triggered ARP probing

About this task

By default, upon receiving a data packet with an unknown source MAC address from a user, the device learns a MAC address entry. It does not learn an ARP entry so the device cannot send packets destined for the user. To resolve this issue, you can enable unknown source MAC-triggered ARP probing.

This feature enables the device to perform ARP probing upon receiving a data packet with an unknown source MAC address. During ARP probing, the device sends an ARP request out of the receiving interface with the source IP address of the data packet as the target MAC address. If the device receives an ARP reply, the device creates an ARP entry.

In inter-AC roaming, a client roams among APs that are managed by different ACs with its IP address unchanged. For the client to receive packets after it roams to a new AC, enable this feature on the ACs.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable unknown source MAC-triggered ARP probing.

   arp unknown-source-mac-probing enable

   By default, unknown source MAC-triggered ARP probing is disabled.

Enabling ARP logging

About this task

This feature enables a device to log ARP events when ARP cannot resolve IP addresses correctly. The log information helps administrators locate and solve problems. The device can log the following ARP events:

- On a proxy ARP-disabled interface, the target IP address in a received ARP packet is not one of the following IP addresses:
  - The IP address of the receiving interface.
  - The virtual IP address of the VRRP group.
- The sender IP address in a received ARP reply conflicts with one of the following IP addresses:
  - The IP address of the receiving interface.
  - The virtual IP address of the VRRP group.

The device sends ARP log messages to the information center. You can use the **info-center source** command to specify the log output rules for the information center. For more information about information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable ARP logging.
   arp check log enable
   By default, ARP logging is disabled.

Enabling error logging for ARP entry deployment to hardware

About this task

After the device learns ARP entries, it deploys the ARP entries to hardware for packet forwarding. After you enable this feature, the device logs the errors that occur during the deployment. The logs are sent to the information center of the device. For the logs to be output correctly, you must also configure the information center on the device. For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Log generation occupies memory. As a best practice, enable this feature only when traffic forwarding is abnormal for troubleshooting.

If the maximum number of logs that can be output per second is reached, no new error logs will be output.

Procedure

1. Enter system view.
   system-view
2. Enable error logging for ARP entry deployment to hardware.
   arp hardware log enable [**count-limit** *count-limit-value*]
   By default, error logging for ARP entry deployment to hardware is disabled.

Enabling the device to send only ARP broadcast requests destined for a VLAN interface on it to the CPU

About this task

By default, the device sends all received ARP broadcast requests to the CPU, which consumes a lot of CPU resources. To resolve the issue, use this command to enable the device to send only ARP broadcast requests that have the same IP address as that of a VLAN interface on the device to the CPU.

Restrictions and guidelines

This feature is mutually exclusive with the proxy ARP and local proxy ARP features.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN interface view.

   interface vlan-interface *interface-number*

3. Enable the device to send only ARP broadcast requests destined for a VLAN interface on it to the CPU.

   arp broadcast pass-through enable

   By default, the device sends all received ARP broadcast requests to the CPU.

Disabling the device from aging or deleting ARP entries in response to network topology changes

About this task

After you configure STP for the device in a tree-shaped network topology, the system notifies the ARP module to age or delete the learned ARP entries when the loop protocol detects a network topology change. Then, the device must learn ARP entries again to obtain the latest ARP entry information. If the network topology changes frequently, a large number of ARP packets might occur when the device relearns ARP entries, which occupies too many system resources and affects the normal operation of other services. To avoid such an issue, you can execute the **undo arp topology-change enable** command. Even if the network topology changes, the device does not age or delete ARP entries.

Restrictions and guidelines

After you disable the device from aging or deleting ARP entries in response to network topology changes, the ARP entries saved by the device might not be the latest ones, which causes user traffic interruption. As a best practice, use this command only when necessary.

Procedure

1. Enter system view.

   system-view

2. Disable the device from aging or deleting ARP entries in response to network topology changes.

   undo arp topology-change enable

By default, the device ages or deletes ARP entries in response to network topology changes.

Disabling sending ARP requests when data packets trigger ARP resolution

About this task

By default, when the device receives a data packet not destined for it and cannot find a match for the next hop in the ARP table, it performs the following tasks:

1. Sends an ARP request to obtain the MAC address of the next hop.
2. Generates an ARP entry based on the obtained MAC address.

A large number of ARP requests consume too many network resources, affecting normal service operation. To resolve the issue, use this feature to disable the device from sending ARP requests for ARP learning when data packets trigger ARP resolution. This suppresses ARP flooding by reducing ARP packets on the network.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Disable the device from sending ARP requests for ARP learning when data packets trigger ARP resolution.

   arp fib-miss drop

   By default, the device sends ARP requests for ARP learning when data packets trigger ARP resolution.

Display and maintenance commands for ARP

CAUTION:

Clearing ARP entries from the ARP table might cause communication failures. Make sure the entries to be cleared do not affect current communications.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ARP entries.	**display arp** [[**all** | **dynamic** | **static**] [**slot** *slot-number*] | **vlan** *vlan-id* | **interface** *interface-type interface-number*] [**count** | **verbose**]
Display the maximum number of ARP entries that a device supports.	**display arp entry-limit**
Display the ARP entry for an IP address.	**display arp** *ip-address* [**slot** *slot-number*] [**verbose**]
Display the number of OpenFlow ARP entries.	**display arp openflow count** [**slot** *slot-number*]
Display the aging timer of dynamic ARP	**display arp timer aging**

Task	Command
entries.	
Display the ARP table usage.	**display arp usage**
Display user IP address conflicts.	**display arp user-ip-conflict record** [**slot** *slot-number*]
Display user port migrations.	**display arp user-move record** [**slot** *slot-number*]
Display the ARP entries for a VPN instance.	**display arp vpn-instance** *vpn-instance-name* [**count** | **verbose**]
Clear ARP entries from the ARP table.	**reset arp** { **all** | **dynamic** | **interface** *interface-type interface-number* | **slot** *slot-number* | **static** }

ARP configuration examples

Example: Configuring a long static ARP entry

Network configuration

As shown in Figure 3, hosts are connected to Device B. Device B is connected to Device A through interface GigabitEthernet 1/0/1 in VLAN 10.

To ensure secure communications between Device A and Device B, configure a long static ARP entry for Device A on Device B.

Figure 3 Network diagram

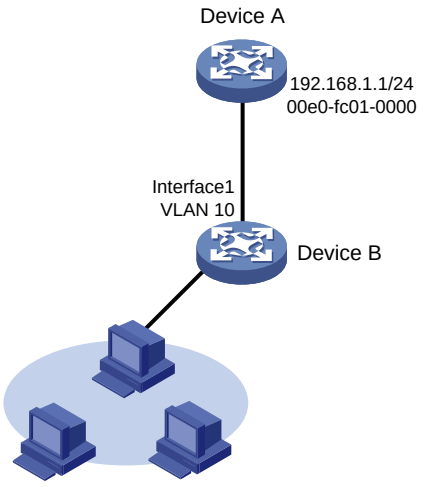


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Create VLAN 10.

```
<DeviceB> system-view
[DeviceB] vlan 10
[DeviceB-vlan10] quit
```

Add interface GigabitEthernet 1/0/1 to VLAN 10.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port access vlan 10
[DeviceB-GigabitEthernet1/0/1] quit
```

Create VLAN-interface 10 and configure its IP address.

```
[DeviceB] interface vlan-interface 10
[DeviceB-vlan-interface10] ip address 192.168.1.2 8
[DeviceB-vlan-interface10] quit
```

Configure a long static ARP entry that has IP address 192.168.1.1, MAC address 00e0-fc01-0000, and output interface GigabitEthernet 1/0/1 in VLAN 10.

```
[DeviceB] arp static 192.168.1.1 00e0-fc01-0000 10 gigabitethernet 1/0/1
```

Verifying the configuration

Verify that Device B has a long static ARP entry for Device A.

```
[DeviceB] display arp static
  Type: S-Static   D-Dynamic   O-Openflow   R-Rule   I-Invalid
IP address      MAC address    VLAN/VSI name Interface               Aging Type
192.168.1.1     00e0-fc01-0000 10            GE1/0/1                 --    S
```

Example: Configuring a short static ARP entry

Network configuration

As shown in Figure 4, hosts are connected to Device B. Device B is connected to Device A through interface GigabitEthernet 1/0/2.

To ensure secure communications between Device A and Device B, configure a short static ARP entry for Device A on Device B.

Figure 4 Network diagram

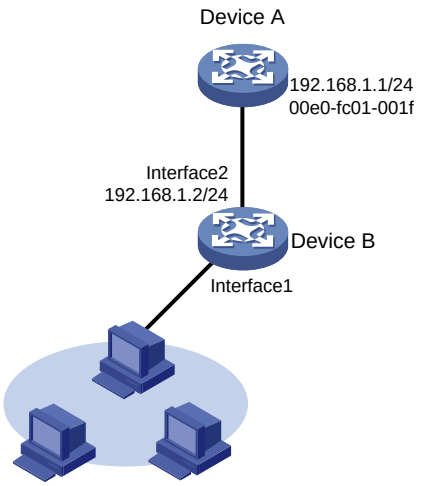


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

Configure an IP address for GigabitEthernet 1/0/2.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ip address 192.168.1.2 24
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure a short static ARP entry that has IP address 192.168.1.1 and MAC address 00e0-fc01-001f.

```
[DeviceB] arp static 192.168.1.1 00e0-fc01-001f
```

Verifying the configuration

Verify that Device B has a short static ARP entry for Device A

```
[DeviceB] display arp static
  Type: S-Static   D-Dynamic   O-Openflow   R-Rule   I-Invalid
IP address      MAC address    VLAN/VSI name Interface               Aging Type
192.168.1.1     00e0-fc01-001f --            --                      --    S
```

Configuring gratuitous ARP

About gratuitous ARP

In a gratuitous ARP packet, the sender IP address and the target IP address are the IP address of the sending device.

A device sends a gratuitous ARP packet for either of the following purposes:

- Determine whether its IP address is already used by another device. If the IP address is already used, the device is informed of the conflict by an ARP reply.
- Inform other devices of a MAC address change.

IP conflict detection

When an interface obtains an IP address, the device broadcasts gratuitous ARP packets in the LAN where the interface resides. If the device receives an ARP reply, its IP address conflicts with the IP address of another device in the LAN. The device displays a log message about the conflict and informs the administrator to change the IP address. The device will not use the conflicting IP address. If no ARP reply is received, the device uses the IP address.

Gratuitous ARP packet learning

This feature enables a device to create or update ARP entries by using the sender IP and MAC addresses in received gratuitous ARP packets.

When this feature is disabled, the device uses received gratuitous ARP packets to update existing ARP entries only. ARP entries are not created based on the received gratuitous ARP packets, which saves ARP table space.

Periodic sending of gratuitous ARP packets

Periodic sending of gratuitous ARP packets helps downstream devices update ARP entries or MAC entries in a timely manner.

This feature can implement the following functions:

- Prevent gateway spoofing.

  Gateway spoofing occurs when an attacker uses the gateway address to send gratuitous ARP packets to the hosts on a network. The traffic destined for the gateway from the hosts is sent to the attacker instead. As a result, the hosts cannot access the external network.

  To prevent such gateway spoofing attacks, you can enable the gateway to send gratuitous ARP packets at intervals. Gratuitous ARP packets contain the primary IP address and manually configured secondary IP addresses of the gateway, so hosts can learn correct gateway information.
- Prevent ARP entries from aging out.

  If network traffic is heavy or if the host CPU usage is high, received ARP packets can be discarded or are not promptly processed. Eventually, the dynamic ARP entries on the receiving host age out. The traffic between the host and the corresponding devices is interrupted until the host re-creates the ARP entries.

  To prevent this problem, you can enable the gateway to send gratuitous ARP packets periodically. Gratuitous ARP packets contain the primary IP address and manually configured secondary IP addresses of the gateway, so the receiving hosts can update ARP entries in a timely manner.
- Prevent the virtual IP address of a VRRP group from being used by a host.

  The master router of a VRRP group can periodically send gratuitous ARP packets to the hosts on the local network. The hosts can then update local ARP entries and avoid using the virtual IP address of the VRRP group. The sender MAC address in the gratuitous ARP packet is the virtual MAC address of the virtual router. For more information about VRRP, see *High Availability Configuration Guide*.

Gratuitous ARP tasks at a glance

All gratuitous ARP tasks are optional. If all of the following features are disabled, gratuitous ARP still provides the IP conflict detection function.

- Enabling notification for passive IP conflict detection
- Enabling gratuitous ARP packet learning
- Enabling periodic sending of gratuitous ARP packets
- Enabling sending gratuitous ARP packets for ARP requests with sender IP address on a different subnet
- Configuring gratuitous ARP packet retransmission for the device MAC address change
- Configuring gratuitous ARP packet forwarding for a VSI

Enabling notification for passive IP conflict detection

About this task

With this feature disabled, the device does not immediately generate an IP conflict notification if it receives an ARP request with the same sender IP address as its interface IP address, or an ARP request with sender IP address 0.0.0.0 and the same target IP address as its interface IP address. Instead, it actively sends a gratuitous ARP packet for confirmation and generates an IP conflict notification only after it receives an ARP reply from the peer.

With this feature enabled, the device immediately generates an IP conflict notification if it receives an ARP request with the same sender IP address as its interface IP address, or an ARP request with sender IP address 0.0.0.0 and the same target IP address as its interface IP address. This helps identify IP conflicts earlier and improves operations efficiency.

Procedure

1. Enter system view.
   `system-view`
2. Enable notification for passive IP conflict detection.
   `arp ip-conflict log prompt`
   By default, notification for passive IP conflict detection is disabled.

Enabling gratuitous ARP packet learning

1. Enter system view.
   `system-view`
2. Enable gratuitous ARP packet learning.
   `gratuitous-arp-learning enable`
   By default, gratuitous ARP packet learning is enabled.

Enabling periodic sending of gratuitous ARP packets

Restrictions and guidelines

- You can enable periodic sending of gratuitous ARP packets on a maximum of 1024 interfaces.
- Periodic sending of gratuitous ARP packets takes effect on an interface only when the following conditions are met:
  - The data link layer state of the interface is up.
  - The interface has an IP address.
- If you change the sending interval for gratuitous ARP packets, the configuration takes effect at the next sending interval.
- The sending interval for gratuitous ARP packets might be much longer than the specified sending interval in any of the following circumstances:
  - This feature is enabled on multiple interfaces.
  - Each interface is configured with multiple secondary IP addresses.
  - A small sending interval is configured when the previous two conditions exist.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable periodic sending of gratuitous ARP packets.
   arp send-gratuitous-arp [**interval** *interval*]
   By default, periodic sending of gratuitous ARP packets is disabled.

Enabling sending gratuitous ARP packets for ARP requests with sender IP address on a different subnet

1. Enter system view.
   system-view
2. Enable the device to send gratuitous ARP packets upon receiving ARP requests whose sender IP address belongs to a different subnet.
   gratuitous-arp-sending enable
   By default, a device does not send gratuitous ARP packets upon receiving ARP requests whose sender IP address belongs to a different subnet.

Configuring gratuitous ARP packet retransmission for the device MAC address change

About this task

The device sends a gratuitous ARP packet to inform other devices of its MAC address change. However, the other devices might fail to receive the packet because the device sends the gratuitous ARP packet once only by default. Configure the gratuitous ARP packet retransmission feature to ensure that the other devices can receive the packet.

Procedure

1. Enter system view.
   system-view
2. Set the times and the interval for retransmitting a gratuitous ARP packet for the device MAC address change.
   gratuitous-arp mac-change retransmit *times* **interval** *seconds*
   By default, the device sends a gratuitous packet to inform its MAC address change once only.

Configuring gratuitous ARP packet forwarding for a VSI

About this task

When an access user in an authentication domain initiates an authentication request, the device looks for an available authentication server. If all authentication servers in that authentication domain are detected unavailable, the device changes to the critical state (also known as the fail-permit state). In this state, the device assigns the access user to the critical domain (also known as the fail-permit domain) so that the user can come online in the critical domain without being authenticated. If the local proxy ARP feature is enabled by the **local-proxy-arp enable** command, the device discards the gratuitous ARP packets received from the VSI of the critical domain instead of forwarding them.

To resolve this issue, enable gratuitous ARP packets forwarding for the VSI of the critical domain. The device will then forward the gratuitous ARP packet received from the VSI of the critical domain.

Restrictions and guidelines

Enable this feature only for the VSIs associated with critical domains.

If you disable this feature for a VSI, the device discards the gratuitous ARP packets received from that VSI.

Procedure

1. Enter system view.
   system-view
2. Create a VSI and enter its view.
   vsi *vsi-name*
3. Enable gratuitous ARP packets forwarding for the VSI.
   arp local-proxy gratuitous-arp forward enable
   By default, gratuitous ARP packet forwarding is enabled for a VSI.

Configuring proxy ARP

About proxy ARP

Proxy ARP enables a device on one network to answer ARP requests for an IP address on another network. With proxy ARP, hosts on different broadcast domains can communicate with each other as they would on the same broadcast domain.

Proxy ARP includes common proxy ARP and local proxy ARP.

- **Common proxy ARP**—Allows communication between hosts that connect to different Layer 3 interfaces and reside in different broadcast domains.
- **Local proxy ARP**—Allows communication between hosts that connect to the same Layer 3 interface and reside in different broadcast domains.

Enabling common proxy ARP

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   The following interface types are supported:
   - VLAN interface.
   - Layer 3 Ethernet interface.
   - Layer 3 Ethernet subinterface.
   - Layer 3 aggregate interface.
   - Layer 3 aggregate subinterface.
3. Enable common proxy ARP.
   proxy-arp enable
   By default, common proxy ARP is disabled.

Enabling local proxy ARP

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   The following interface types are supported:
   - VLAN interface.
   - Layer 3 Ethernet interface.
   - Layer 3 Ethernet subinterface.
   - Layer 3 aggregate interface.
   - Layer 3 aggregate subinterface.
3. Enable local proxy ARP.
   local-proxy-arp enable [**ip-range** *start-ip-address* **to** *end-ip-address*]

By default, local proxy ARP is disabled.

Display and maintenance commands for proxy ARP

Execute **display** commands in any view.

Task	Command
Display common proxy ARP status.	**display proxy-arp** [**interface** *interface-type interface-number*]
Display local proxy ARP status.	**display local-proxy-arp** [**interface** *interface-type interface-number*]
Display statistics about proxy ARP reply packets.	**display proxy-arp statistics**

Common proxy ARP configuration example

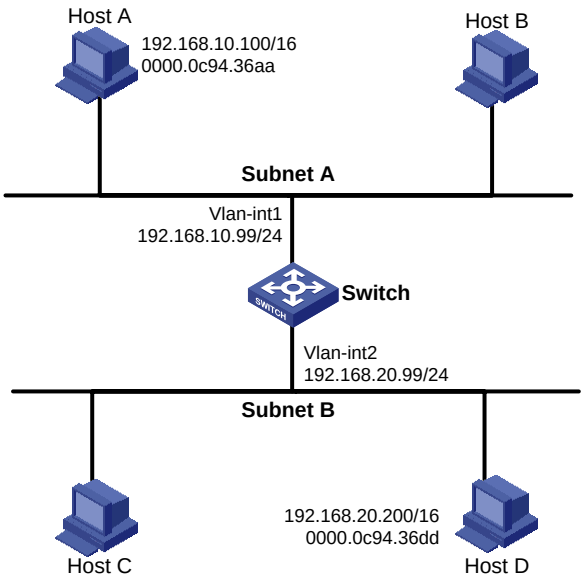
Example: Configuring common proxy ARP

Network configuration

As shown in Figure 5, Host A and Host D have the same IP prefix and mask, but they are located on different subnets separated by the switch. Host A belongs to VLAN 1, and Host D belongs to VLAN 2. No default gateway is configured on Host A and Host D.

Configure common proxy ARP on the switch to enable communication between the two hosts.

Figure 5 Network diagram



Procedure

Create VLAN 2.

```
<Switch> system-view
[Switch] vlan 2
[Switch-vlan2] quit
```

Configure the IP address of VLAN-interface 1.

```
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.10.99 255.255.255.0
```

Enable common proxy ARP on VLAN-interface 1.

```
[Switch-Vlan-interface1] proxy-arp enable
[Switch-Vlan-interface1] quit
```

Configure the IP address of VLAN-interface 2.

```
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.20.99 255.255.255.0
```

Enable common proxy ARP on VLAN-interface 2.

```
[Switch-Vlan-interface2] proxy-arp enable.
```

Verifying the configuration

Verify that Host A and Host D can ping each other.

Configuring ARP snooping

About ARP snooping

ARP snooping is used in Layer 2 switching networks. It creates ARP snooping entries by using information in ARP packets. ARP fast-reply and MFF can use the ARP snooping entries. For more information about MFF, see *Security Configuration Guide*.

Creation of ARP snooping entries

If you enable ARP snooping for a VLAN, ARP packets received in the VLAN are redirected to the CPU. For the VLAN, the CPU uses the sender IP and MAC addresses, VLAN, and input interface of the ARP packets to create ARP snooping entries. For the VXLAN, the CPU uses the sender IP and MAC addresses, VSI name, and link ID of the ARP packets to create ARP snooping entries.

If you enable ARP snooping on an interface, ARP packets received by the interface are redirected to the CPU. The CPU uses the sender IP and MAC addresses, VLAN, and input interface of the ARP packets to create ARP snooping entries.

Aging of ARP snooping entries

The aging timer and valid period of an ARP snooping entry are 25 minutes and 15 minutes. If an ARP snooping entry is not updated in 12 minutes, the device sends an ARP request. The ARP request uses the IP address of the entry as the target IP address. If an ARP snooping entry is not updated in 15 minutes, it becomes invalid and cannot be used. After that, if an ARP packet matching the entry is received, the entry becomes valid, and its aging timer restarts.

If the aging timer of an ARP snooping entry expires, the entry is removed.

Protection for ARP snooping

An attack occurs if an ARP packet has the same sender IP address as a valid ARP snooping entry but a different sender MAC address. The ARP snooping entry becomes invalid, and it is removed in 1 minute.

Restrictions and guidelines: ARP snooping configuration

A Layer 2 Ethernet interface is enabled with ARP snooping after you enable ARP snooping on the interface or in the VLAN to which the interface belongs.

A Layer 2 Ethernet interface is disabled with ARP snooping only after you disable ARP snooping both on the interface and in the VLAN to which the interface belongs.

If you associate a VLAN with a VXLAN, ARP snooping does not take effect when you enable both ARP snooping and ARP flood suppression in the VLAN. For more information about configuring ARP flood suppression in a VLAN, see VLAN configuration in *Layer 2—LAN Switching Configuration Guide*.

Enabling ARP snooping for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ARP snooping for the VLAN.
   arp snooping enable
   By default, ARP snooping is disabled for a VLAN.

Enabling ARP snooping for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable ARP snooping for the VSI.
   arp snooping enable
   By default, ARP snooping is disabled for a VSI.

Enabling ARP snooping for an Layer 2 interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   Only Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces are supported.
3. Enable ARP snooping for the interface.
   arp snooping enable
   By default, ARP snooping is disabled for an interface.

Configuring an interface as an uplink interface to disable it from learning ARP snooping entries

About this task

After you enable ARP snooping on an access device by using the **arp snooping enable** command, the access device will generate ARP snooping entries by listening to ARP packets. On a network where the aggregate device acts as the gateway, if you enable local proxy ARP on the gateway by using the **local-proxy-arp enable** command, the uplink interface of the access device will also learn ARP snooping entries. As a result, the input interface of an ARP snooping entry flaps between the uplink and downlink interfaces. To avoid such an issue, you can configure this feature on the access device.

After you configure this feature on an access device enabled with ARP snooping, the uplink interface no longer learns ARP snooping entries from incoming ARP packets.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   Supported interface types include Layer 2 Ethernet interface and Layer 2 aggregate interface.
3. Configure the interface as an uplink interface to disable it from learning ARP snooping entries.
   arp snooping uplink
   By default, an interface is not an uplink interface for ARP snooping. After you enable ARP snooping, the interface learns ARP snooping entries.

Display and maintenance commands for ARP snooping

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ARP snooping entries.	**display arp snooping vlan** [*vlan-id*] [**slot** *slot-number*] [**count**] **display arp snooping** { **interface** | **vlan** } **ip** *ip-address* [**slot** *slot-number*]
Delete ARP snooping entries.	**reset arp snooping** { **interface** [*interface-type interface-number*] | **vlan** [*vlan-id*] | **vsi** [*vsi-name*] } **reset arp snooping** { **interface** | **vlan** } **ip** *ip-address*

Configuring ARP fast-reply

About ARP fast-reply

ARP fast-reply enables a device to directly answer ARP requests according to IP source guard entries or ARP snooping entries. ARP fast-reply functions in a VLAN. For information about IP source guard, see IP source guard configuration in *Security Configuration Guide*.

If the target IP address of a received ARP request is the IP address of the VLAN interface, the device delivers the request to the ARP module. If not, the device takes the following steps to process the packet:

1. Search the IP source guard binding table for a match by using the target IP address.
2. If a match is found, whether the device returns a reply depends on the type of interface in the matching entry.
   - If the interface is the Ethernet interface that received the ARP request, the device does not return any reply.
   - If the interface is an Ethernet interface other than the receiving interface, the device returns a reply according to the matching entry.
3. If no matching IP source guard entry is found and ARP snooping is enabled, the device searches the ARP snooping table.
   - If the interface in the matching entry is the Ethernet interface that received the ARP request, the device does not return any reply.
   - If the interface is an Ethernet interface other than the receiving interface, the device returns a reply according to the ARP snooping entry.
4. If no match is found in both tables, the ARP request is forwarded to other interfaces except the receiving interface in the VLAN, or delivered to other modules.

Enabling ARP fast-reply

Restrictions and guidelines

To improve the availability of ARP fast-reply, enable ARP snooping at the same time.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ARP fast-reply.
   arp fast-reply enable
   By default, ARP fast-reply is disabled.

Enabling ARP gateway proxy

About this task

To connect hosts that require intercommunication to the same Layer 3 interface on a gateway and make them belong to different broadcast domains, you can enable local proxy ARP on the gateway.

This allows the gateway to respond to ARP requests and achieves Layer 2 isolation and Layer 3 communication.

However, in a large campus or enterprise network, a large number of endpoints connected to the access layer might generate substantial ARP requests per second. The gateway in the core layer might not efficiently process the requests. To resolve such an issue, you can enable ARP gateway proxy on a device in the access or distribution layer. Then, the device responds to the ARP requests, which alleviates the pressure on the gateway.

Enabled with ARP gateway proxy, the device actively records the gateway MAC address when it receives an ARP packet sourced from the gateway IP address. After the device receives an ARP request from an endpoint, it replies with the gateway MAC address. If the device does not learn the gateway MAC address, it does not respond to the ARP request.

Restrictions and guidelines

You must specify the gateway IP address when you configure ARP gateway proxy.

To avoid replied MAC address confusion, do not configure this feature and ARP fast-reply simultaneously on the device.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ARP gateway proxy.
   arp fast-reply l2-proxy gateway *ip-address*
   By default, ARP gateway proxy is disabled.

ARP fast-reply configuration example

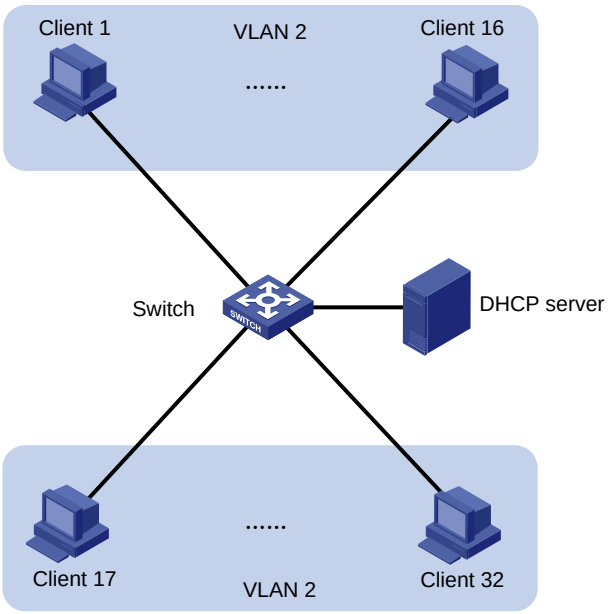
Example: Configuring ARP fast-reply

Network configuration

As shown in Figure 6, all clients are in VLAN 2, and access the network through the switch. They have obtained IP addresses through DHCP.

Enable ARP snooping and ARP fast-reply for VLAN 2. The switch directly returns an ARP reply without broadcasting received ARP requests in the VLAN.

Figure 6 Network diagram



Procedure

Enable ARP snooping for VLAN 2 on the switch.

```
<Switch> system-view
[Switch] vlan 2
[Switch-vlan2] arp snooping enable
```

Enable ARP fast-reply for VLAN 2 on the switch.

```
[Switch-vlan2] arp fast-reply enable
[Switch-vlan2] quit
```

Configuring ARP direct route advertisement

Configuring ARP direct route advertisement

About ARP direct route advertisement

This feature generates direct routes based on ARP entries for packet forwarding and route advertisement. The route preference value determines the match order of a route. Dynamic routing protocols use the tag value as the route identifier when redistributing the route.

After you enable ARP direct route advertisement, the device generates direct routes and adjacency entries based on ARP entries. If the direct routes are generated earlier than the adjacency entries, packet loss will occur due to lack of Layer 2 information during packet encapsulation. To avoid such an issue, you can set the delay time for ARP direct route generation.

After you enable ARP direct route advertisement and set a delay time for ARP direct route generation, the device starts a timer after it has learned an ARP entry. Then, it generates an ARP direct route when the timer value reaches the delay time. If you edit the direct route advertisement configuration within the delay time, the device advertises the direct route based on the new configuration immediately.

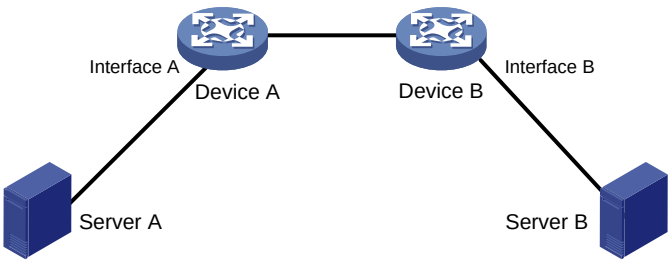
Restrictions and guidelines

The **`arp route-direct advertise`** command is mutually exclusive with the **`arp route-direct advertise mad-down-single-homed`** command. If you execute both of them, the most recent command takes effect.

Application in Layer 3 access networks

As shown in Figure 7, ARP direct route advertisement is enabled on Interface A and Interface B. This feature generates a direct route to Server A and a direct route to Server B for the routing protocols to advertise. So each device forwards only the traffic destined to the server within the network, which saves bandwidth.

Figure 7 Application in a Layer 3 access network



Procedure

1. Enter system view.
   `system-view`

2. Enter interface view.
   interface *interface-type interface-number*
3. (Optional.) Set the delay time for ARP direct route generation.
   arp route-direct advertise delay *delay-time*
   By default, ARP direct route generation is not delayed.
4. Enable the ARP direct route advertisement feature.
   arp route-direct advertise [**preference** *preference-value* | **tag** *tag-value*] *
   By default, the ARP direct route advertisement feature is disabled.

Enabling ARP direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state

About this task

If the peer link of an M-LAG system fails, M-LAG Multi-Active Detection (MAD) will shut down all interfaces on the secondary member device upon M-LAG system split, except for the interfaces excluded from the shutdown action by IRF MAD or M-LAG MAD. An interface in M-LAG MAD DOWN state cannot forward traffic. As a result, traffic cannot be forwarded for the devices single-homed to the secondary member device.

To ensure traffic forwarding for single-homed devices, enable ARP direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state. When the peer link fails, the ARP module on the secondary member device will send ARP entries to the routing management module, and the routing management module will generate direct routes based on the ARP entries. The direct routes are used to direct traffic forwarding or are advertised by routing protocols to the single-homed devices.

Restrictions and guidelines

The **arp route-direct advertise mad-down-single-homed** and **arp route-direct advertise** commands are mutually exclusive. If you execute both of them, the most recent command takes effect.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN interface view.
   interface vlan-interface *interface-number*
3. Enable ARP direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state.
   arp route-direct advertise mad-down-single-homed
   By default, ARP direct route advertisement is disabled for single-homing interfaces in M-LAG MAD DOWN state.

Display and maintenance commands for ARP direct route advertisement

Execute **display** commands in any view.

Task	Command
Display information about ARP direct route advertisement.	**display arp route-direct advertise interface** *interface-type interface-number*

Configuring ARP ping

Configuring ARP ping for IPv4 addresses

About this task

This feature enables you to test whether an IPv4 address in a LAN is being used by sending ARP requests.

After the device sends an ARP request to an IPv4 address, if it receives an ARP reply within the wait time, it determines that the IP address is being used. If an ARP reply is not received within the wait time, the device sends another ARP request. After the device sends the maximum number of ARP requests but no ARP reply is received, the device stops sending ARP requests and regards the IPv4 address as an unused address.

Compared with the **ping** command, the **ping arp ip** command avoids a wrong result when the destination host is enabled with a firewall that blocks ICMP packets. In addition, it consumes fewer network resources because an ARP request is shorter than an ICMP packet.

Restrictions and guidelines

If multiple devices exist on the LAN, executing the **ping arp ip** command will take some time. To stop an ongoing test, press **Ctrl** + **C**.

Prerequisites

To execute the **ping arp ip** command by specifying the destination host name, configure the DNS feature first. For more information about DNS, see "Configuring DNS."

Procedure

Execute the following command in any view:

ping arp ip *host* [**interface** *interface-type interface-number* [**vlan** *vlan-id*]] [**timeout** *timeout*] [**count** *count*]

Configuring ARP ping for MAC addresses

About this task

To obtain the corresponding IPv4 address of a MAC address, use the **ping arp mac** command to broadcast Layer 3 ICMP packets.

After the device sends an ICMP echo request, if an ICMP echo reply is not received within the wait time, the device sends another ICMP echo request. After the device sends the maximum number of ICMP echo requests but no ICMP echo reply is received, the device stops sending ICMP echo requests and regards the MAC address as a nonexistent address.

Restrictions and guidelines

If multiple devices exist in the network, executing the **ping arp mac** command will take some time. To stop an ongoing test, press **Ctrl** + **C**.

Procedure

Execute the following command in any view:

ping arp mac *mac-address* { **interface** *interface-type interface-number* |**ip** *ipv4-address* [**vpn-instance** *vpn-instance-name*] } [**timeout** *timeout*] [**count** *count*]

Contents

Configuring IP addressing 1

About IP addressing 1
IP address representation and classes 1
Special IP addresses 2
Subnetting and masking 2
IP address assignment 2
Assigning an IP address to an interface 3
Configuring IP unnumbered 4
Display and maintenance commands for IP addressing 4
IP addressing configuration examples 5
Example: Manually specifying an IP address 5

Configuring IP addressing

About IP addressing

The IP addresses in this chapter refer to IPv4 addresses unless otherwise specified.

IP address representation and classes

IP addressing uses a 32-bit address to identify each host on an IPv4 network. To make addresses easier to read, they are written in dotted decimal notation, each address being four octets in length. For example, address 00001010000000010000000100000001 in binary is written as 10.1.1.1.

Each IP address breaks down into the following sections:

- **Net ID**—Identifies a network. The first several bits of a net ID, known as the class field or class bits, identify the class of the IP address.
- **Host ID**—Identifies a host on a network.

IP addresses are divided into five classes, as shown in Figure 1. The shaded areas represent the address class. The first three classes are most commonly used.

Figure 1 IP address classes

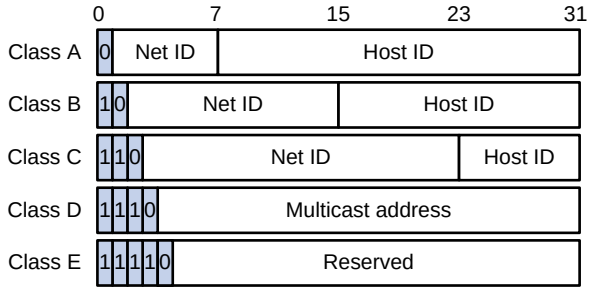


Table 1 IP address classes and ranges

Class	Address range	Remarks
A	0.0.0.0 to 127.255.255.255	The IP address 0.0.0.0 is used by a host at startup for temporary communication. This address is never a valid destination address. Addresses starting with 127 are reserved for loopback test. Packets destined to these addresses are processed locally as input packets rather than sent to the link.
B	128.0.0.0 to 191.255.255.255	N/A
C	192.0.0.0 to 223.255.255.255	N/A
D	224.0.0.0 to 239.255.255.255	Multicast addresses.
E	240.0.0.0 to 255.255.255.255	Reserved for future use, except for the broadcast address 255.255.255.255.

Special IP addresses

The following IP addresses are for special use and cannot be used as host IP addresses:

- **IP address with an all-zero net ID**—Identifies a host on the local network. For example, IP address 0.0.0.16 indicates the host with a host ID of 16 on the local network.
- **IP address with an all-zero host ID**—Identifies a network.
- **IP address with an all-one host ID**—Identifies a directed broadcast address. For example, a packet with the destination address of 192.168.1.255 will be broadcast to all the hosts on the network 192.168.1.0.

Subnetting and masking

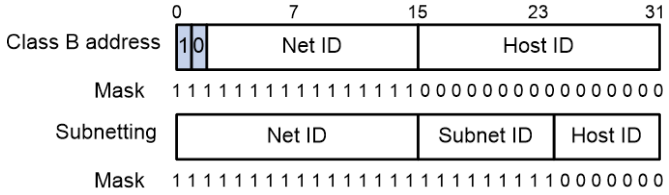
Subnetting divides a network into smaller networks called subnets by using some bits of the host ID to create a subnet ID.

Masking identifies the boundary between the host ID and the combination of net ID and subnet ID.

Each subnet mask comprises 32 bits that correspond to the bits in an IP address. In a subnet mask, consecutive ones represent the net ID and subnet ID, and consecutive zeros represent the host ID.

Before being subnetted, Class A, B, and C networks use these default masks (also called natural masks): 255.0.0.0, 255.255.0.0, and 255.255.255.0, respectively.

Figure 2 Subnetting a Class B network



Subnetting increases the number of addresses that cannot be assigned to hosts. Therefore, using subnets means accommodating fewer hosts.

For example, a Class B network without subnetting can accommodate 1022 more hosts than the same network subnetted into 512 subnets.

- **Without subnetting**—65534 (2^{16} – 2) hosts. (The two deducted addresses are the broadcast address, which has an all-one host ID, and the network address, which has an all-zero host ID.)
- **With subnetting**—Using the first nine bits of the host-id for subnetting provides 512 (2^9) subnets. However, only seven bits remain available for the host ID. This allows 126 (2^7 – 2) hosts in each subnet, a total of 64512 (512 × 126) hosts.

IP address assignment

The following are methods available for assigning an IP address to an interface:

- Manual assignment. This chapter describes only manual IP address assignment for interfaces.
- BOOTP. For information about BOOTP, see "Configuring the BOOTP client."
- DHCP. For information about DHCP, see "Configuring the DHCP client."

These methods are mutually exclusive. If you change the IP address assignment method, the new IP address will overwrite the previous address.

Assigning an IP address to an interface

About this task

An interface can have one primary address and multiple secondary addresses.

Typically, you need to configure a primary IP address for an interface. If the interface connects to multiple subnets, configure primary and secondary IP addresses on the interface so the subnets can communicate with each other through the interface.

In an IRF fabric, you can assign an IP address to the management Ethernet port of each member in the management Ethernet port view of the master. Only the IP address assigned to the management Ethernet port of the master takes effect. After an IRF fabric split, the IP addresses assigned to the management Ethernet ports of the new masters (original subordinates) take effect. Then you can use these IP addresses to log in to the new masters for troubleshooting.

Restrictions and guidelines

- An interface can have only one primary IP address. If you execute this command multiple times to specify different primary IP addresses on an interface, the most recent configuration takes effect.
- You cannot assign secondary IP addresses to an interface that obtains an IP address through IP unnumbered, BOOTP, or DHCP.
- The primary and secondary IP addresses assigned to the interface can be located on the same network segment. Different interfaces on your device must reside on different network segments.
- After an IRF split, the routing information on the original master might not be updated immediately. As a result, the management Ethernet port of the original master cannot be pinged from the master (original subordinate) in another IRF fabric. To resolve the problem, wait until route synchronization between the devices is completed or enable NSR for the routing protocol. For information about NSR, see *Layer 3—IP Routing Configuration Guide*.
- The following commands are mutually exclusive. You can configure only one of these commands to assign an IP address to the management Ethernet port of the IRF master.
  - The **ip address** command with the **irf-member** *member-id* option that specifies the master.
  - The **ip address** command that does not contain the **irf-member** *member-id* option.
  - The **mad ip address** command.
  - The **ip address dhcp-alloc** command.
- Exclude the management Ethernet port of the master from being shut down if MAD is enabled in the IRF fabric. The port can be kept in up state when the MAD status transits to Recovery.
- You can assign interfaces IP addresses that have different masks but the same network address if ANDed with the shortest mask. For example, 1.1.1.1/16 and 1.1.2.1/24 have the same network address 1.1.0.0 if ANDed with 255.255.0.0. You can assign the IP addresses to two interfaces on the device. By default, users connected to the two interfaces cannot communicate with each other. For the users to communicate, you must configure common proxy ARP on the device. For more information, see "Configuring proxy ARP."

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Assign an IP address to the interface.

ip address *ip-address* { *mask-length* | *mask* } [**irf-member** *member-id* | **sub**]

By default, no IP address is assigned to the interface.

To assign an IP address to the management Ethernet port of an IRF member device, enter the master's management Ethernet port view and specify the **irf-member** *member-id* option.

Configuring IP unnumbered

About this task

You can configure an interface to borrow an IP address from other interfaces. This is called IP unnumbered, and the interface borrowing the IP address is called IP unnumbered interface.

You can use IP unnumbered to save IP addresses when available IP addresses are inadequate or when an interface is used only occasionally.

Restrictions and guidelines

- Loopback interfaces cannot borrow IP addresses of other interfaces, but other interfaces can borrow IP addresses of loopback interfaces.
- An interface cannot borrow an IP address from an unnumbered interface.
- Multiple interfaces can use the same unnumbered IP address.
- If an interface has multiple manually configured IP addresses, only the manually configured primary IP address can be borrowed.
- A Layer 3 Ethernet subinterface or Layer 3 aggregate subinterface cannot learn the ARP entry for the IP address it borrows. As a best practice, do not configure IP unnumbered on Layer 3 Ethernet subinterfaces or Layer 3 aggregate subinterfaces.

Prerequisites

Configure an IP address for the interface from which you want to borrow the IP address. You can manually assign an IP address to the interface or configure the interface to obtain one through BOOTP or DHCP.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the interface to borrow the IP address of the specified interface.
   ip address unnumbered interface *interface-type interface-number*
   By default, the interface does not borrow IP addresses from other interfaces.

Display and maintenance commands for IP addressing

Execute **display** commands in any view.

Task	Command
Display IP configuration and statistics for the specified or all Layer 3 interfaces.	**display ip interface** [*interface-type* [*interface-number*]]

Task	Command
Display brief IP configuration for Layer 3 interfaces.	**display ip interface** [*interface-type* [*interface-number*]] **brief** [**description**]

IP addressing configuration examples

Example: Manually specifying an IP address

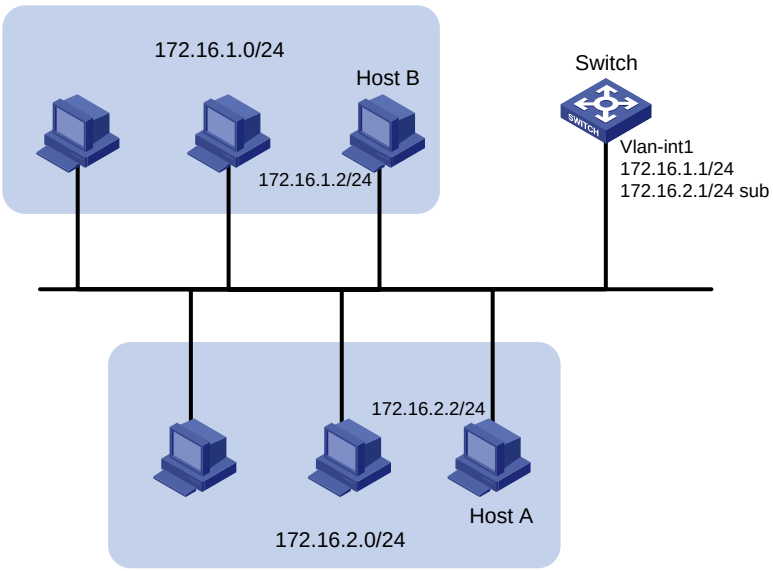
Network configuration

As shown in Figure 3, a port in VLAN 1 on a switch is connected to a LAN comprising two segments: 172.16.1.0/24 and 172.16.2.0/24.

To enable the hosts on the two network segments to communicate with the external network through the switch, and to enable the hosts on the LAN to communicate with each other:

- Assign a primary IP address and a secondary IP address to VLAN-interface 1 on the switch.
- Set the primary IP address of the switch as the gateway address of the PCs on subnet 172.16.1.0/24, and set the secondary IP address of the switch as the gateway address of the PCs on subnet 172.16.2.0/24.

Figure 3 Network diagram



Procedure

Assign a primary IP address and a secondary IP address to VLAN-interface 1.

```
<Sysname> system-view
[Sysname] sysname Switch
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 172.16.1.1 255.255.255.0
[Switch-Vlan-interface1] ip address 172.16.2.1 255.255.255.0 sub
[Switch-Vlan-interface1] quit
```

Set the gateway address to 172.16.1.1 on the PCs attached to subnet 172.16.1.0/24, and to 172.16.2.1 on the PCs attached to subnet 172.16.2.0/24.

Verifying the configuration

Verify the connectivity between a host on subnet 172.16.1.0/24 and the switch.

```
[Switch] ping 172.16.1.2
Ping 172.16.1.2 (172.16.1.2): 56 data bytes, press CTRL_C to break
56 bytes from 172.16.1.2: icmp_seq=0 ttl=128 time=7.000 ms
56 bytes from 172.16.1.2: icmp_seq=1 ttl=128 time=2.000 ms
56 bytes from 172.16.1.2: icmp_seq=2 ttl=128 time=1.000 ms
56 bytes from 172.16.1.2: icmp_seq=3 ttl=128 time=1.000 ms
56 bytes from 172.16.1.2: icmp_seq=4 ttl=128 time=2.000 ms
--- Ping statistics for 172.16.1.2 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/2.600/7.000/2.245 ms
```

Verify the connectivity between a host on subnet 172.16.2.0/24 and the switch.

```
[Switch] ping 172.16.2.2
Ping 172.16.2.2 (172.16.2.2): 56 data bytes, press CTRL_C to break
56 bytes from 172.16.2.2: icmp_seq=0 ttl=128 time=2.000 ms
56 bytes from 172.16.2.2: icmp_seq=1 ttl=128 time=7.000 ms
56 bytes from 172.16.2.2: icmp_seq=2 ttl=128 time=1.000 ms
56 bytes from 172.16.2.2: icmp_seq=3 ttl=128 time=2.000 ms
56 bytes from 172.16.2.2: icmp_seq=4 ttl=128 time=1.000 ms
--- Ping statistics for 172.16.2.2 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/2.600/7.000/2.245 ms
```

Verify the connectivity between a host on subnet 172.16.1.0/24 and a host on subnet 172.16.2.0/24. The ping operation succeeds.

Configuration files

```
#
 sysname Switch
#
interface Vlan-interface1
 port link-mode route
 ip address 172.16.1.1 255.255.255.0
 ip address 172.16.2.1 255.255.255.0 sub
#
```

Contents

DHCP overview 1

DHCP network model 1
DHCP address allocation 1
Allocation mechanisms 1
IP address allocation process 2
IP address lease extension 2
DHCP message format 3
DHCP options 4
Common DHCP options 4
Custom DHCP options 4
Vendor-specific option (Option 43) 5
Relay agent option (Option 82) 6
Option 184 7
Option 252 7
Option 253 7
Protocols and standards 8

Configuring the DHCP server 9

About DHCP server 9
DHCP address assignment mechanisms 9
Principles for selecting an address pool 10
IP address allocation sequence 11
DHCP server tasks at a glance 11
Creating a DHCP user class 12
Configuring an address pool on the DHCP server 12
DHCP address pool tasks at a glance 12
Creating a DHCP address pool 13
Specifying a primary subnet and multiple address ranges in a DHCP address pool 13
Specifying a primary subnet and multiple secondary subnets in a DHCP address pool 14
Configuring a static binding in a DHCP address pool 15
Specifying gateways for DHCP clients 16
Specifying a domain name suffix for DHCP clients 17
Specifying DNS servers for DHCP clients 17
Specifying WINS servers and NetBIOS node type for DHCP clients 17
Specifying BIMS server for DHCP clients 18
Specifying the configuration file for DHCP client automatic configuration 18
Specifying a server for DHCP clients 19
Configuring Option 184 parameters for DHCP clients 20
Customizing DHCP options 20
Applying a DHCP address pool to a VPN instance 22
Configuring the DHCP user class whitelist 22
Binding gateways to DHCP server's MAC address 23
Advertising subnets that are assigned to clients 24
Applying an address pool to an interface 24
Configuring a DHCP policy for dynamic assignment 25
Enabling DHCP 26
Enabling the DHCP server on an interface 26
Configuring IP address conflict detection 26
Enabling handling of Option 82 27
Configuring the DHCP server security features 27
Restrictions and guidelines 27
Configuring DHCP flood attack protection 27
Configuring DHCP starvation attack protection 28
Configuring DHCP server compatibility 29
Configuring the DHCP server to always broadcast responses 29
Enabling the DHCP server to return a DHCP-NAK message upon client notions of incorrect IP addresses 29

Configuring the DHCP server to ignore BOOTP requests ····30
Configuring the DHCP server to send BOOTP responses in RFC 1048 format ····30
Setting the DSCP value for DHCP packets sent by the DHCP server ····31
Configuring DHCP binding auto backup ····31
Enabling client offline detection on the DHCP server ····32
Configuring SNMP notifications for the DHCP server ····32
About SNMP DHCP server notifications ····32
Enabling IP exhaustion notifications ····32
Enabling IP usage alarm notifications ····33
Enabling DHCP logging on the DHCP server ····33
Enabling IP exhaustion event logging ····33
Display and maintenance commands for DHCP server ····34
DHCP server configuration examples ····35
Example: Configuring static IP address assignment ····35
Example: Configuring dynamic IP address assignment ····36
Example: Configuring DHCP user class ····38
Example: Configuring DHCP user class whitelist ····40
Example: Configuring primary and secondary subnets ····41
Example: Customizing DHCP option ····42
Troubleshooting DHCP server configuration ····44
Failure to obtain a non-conflicting IP address ····44
Configuring the DHCP relay agent ····45
About DHCP relay agent ····45
DHCP relay agent operation ····45
DHCP relay agent support for Option 82 ····46
DHCP relay agent support for MCE ····46
DHCP relay agent tasks at a glance ····46
Enabling DHCP ····47
Enabling the DHCP relay agent on an interface ····47
Specifying DHCP servers ····48
Specifying DHCP servers on a relay agent ····48
Specifying DHCP servers in a DHCP relay address pool ····48
Specifying the DHCP server selecting algorithm ····49
Specifying a DHCP relay address pool for DHCP clients ····50
Configuring the DHCP relay agent security features ····52
Restrictions and guidelines for DHCP relay agent security feature configuration ····52
Enabling the DHCP relay agent to record relay entries ····52
Enabling periodic refresh of dynamic relay entries ····52
Configuring DHCP flood attack protection ····53
Enabling DHCP starvation attack protection ····54
Enabling DHCP server proxy on the DHCP relay agent ····55
Enabling client offline detection on the DHCP relay agent ····55
Configuring the DHCP relay agent to release an IP address ····56
Disabling release notification ····56
Configuring DHCP relay agent support for Option 82 ····56
Enabling Option 60 insertion into DHCP requests ····58
Setting the DSCP value for DHCP packets sent by the DHCP relay agent ····58
Specifying the DHCP relay agent address for the **giaddr** field ····59
Manually specifying the DHCP relay agent address for the **giaddr** field ····59
Configuring smart relay to specify the DHCP relay agent address for the **giaddr** field ····59
Specifying the source IP address for relayed DHCP requests ····60
Discarding DHCP requests that are delivered from VXLAN tunnels ····60
Configuring support for forwarding DHCP replies based on MAC address table ····61
Display and maintenance commands for DHCP relay agent ····61
DHCP relay agent configuration examples ····62
Example: Configuring basic DHCP relay agent ····62
Example: Configuring Option 82 ····63
Example: Configuring DHCP server selection ····64
Troubleshooting DHCP relay agent configuration ····66
Failure of DHCP clients to obtain configuration parameters through the DHCP relay agent ····66

Configuring the DHCP client 67
About DHCP client 67
Support of DHCP Client for gateway availability detection 67
Restrictions and guidelines: DHCP client configuration 67
DHCP client tasks at a glance 67
Enabling the DHCP client on an interface 67
Configuring a DHCP client ID for an interface 68
Enabling duplicated address detection 69
Setting the DSCP value for DHCP packets sent by the DHCP client 69
Configuring SNMP notifications for the DHCP client 69
Display and maintenance commands for DHCP client 70
DHCP client configuration examples 70
Example: Configuring DHCP client 70
Configuring DHCP snooping 73
About DHCP snooping 73
Application of trusted and untrusted ports 73
DHCP snooping support for Option 82 74
DHCP snooping on an M-LAG network 75
Restrictions and guidelines: DHCP snooping configuration 76
DHCP snooping tasks at a glance 76
Configuring basic DHCP snooping features 77
Configuring basic DHCP snooping features in a common network 77
Configuring basic DHCP snooping features in a VXLAN network 78
Configuring DHCP snooping support for Option 82 80
Configuring DHCP snooping entry auto backup 81
Setting the maximum number of DHCP snooping entries 82
Configuring DHCP snooping security features 82
Enabling DHCP starvation attack protection 82
Enabling DHCP-REQUEST attack protection 83
Configuring a DHCP packet blocking port 84
Enabling recording untrusted DHCP servers 84
Enabling the giaddr field check in DHCP requests 84
Enabling client offline detection on the DHCP snooping device 85
Enabling DHCP snooping logging 85
Enabling DHCP snooping alarm notifications 86
About DHCP snooping alarm notifications 86
Enabling the binding-exhaust alarm 86
Enabling the binding-threshold alarm 87
Enabling the binding-mismatch or chaddr-mismatch packet drop alarm 87
Enabling the rate-limit or untrust-reply packet drop alarm 88
Enabling alarm log generation for packet drop events 89
Disabling DHCP snooping on an interface 89
Display and maintenance commands for DHCP snooping 90
DHCP snooping configuration examples 91
Example: Configuring basic DHCP snooping features globally 91
Example: Configuring basic DHCP snooping features for a VLAN 92
Example: Configuring DHCP snooping support for Option 82 93
Configuring the BOOTP client 95
About BOOTP client 95
BOOTP client application 95
Obtaining an IP address dynamically 95
Protocols and standards 95
Configuring an interface to use BOOTP for IP address acquisition 95
Display and maintenance commands for BOOTP client 96
BOOTP client configuration examples 96
Example: Configuring BOOTP client 96

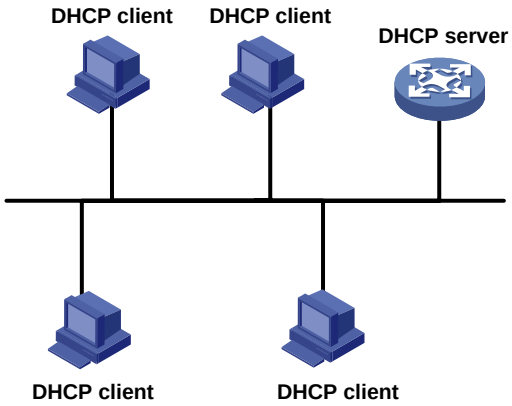
DHCP overview

DHCP network model

The Dynamic Host Configuration Protocol (DHCP) provides a framework to assign configuration information to network devices.

Figure 1 shows a typical DHCP application scenario where the DHCP clients and the DHCP server reside on the same subnet. The DHCP clients can also obtain configuration parameters from a DHCP server on another subnet through a DHCP relay agent. For more information about the DHCP relay agent, see "Configuring the DHCP relay agent."

Figure 1 A typical DHCP application



DHCP address allocation

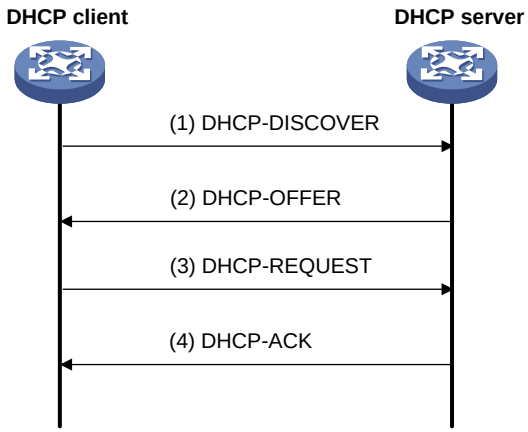
Allocation mechanisms

DHCP supports the following allocation mechanisms:

- **Static allocation**—The network administrator assigns an IP address to a client, such as a WWW server, and DHCP conveys the assigned address to the client.
- **Automatic allocation**—DHCP assigns a permanent IP address to a client.
- **Dynamic allocation**—DHCP assigns an IP address to a client for a limited period of time, which is called a lease. Most DHCP clients obtain their addresses in this way.

IP address allocation process

Figure 2 IP address allocation process



As shown in Figure 2, a DHCP server assigns an IP address to a DHCP client in the following process:

1. The client broadcasts a DHCP-DISCOVER message to locate a DHCP server.
2. Each DHCP server offers configuration parameters such as an IP address to the client in a DHCP-OFFER message. The sending mode of the DHCP-OFFER is determined by the flag field in the DHCP-DISCOVER message. For more information, see "DHCP message format."
3. If the client receives multiple offers, it accepts the first received offer, and broadcasts it in a DHCP-REQUEST message to formally request the IP address. (IP addresses offered by other DHCP servers can be assigned to other clients.)
4. All DHCP servers receive the DHCP-REQUEST message. However, only the server selected by the client does one of the following operations:
   - Returns a DHCP-ACK message to confirm that the IP address has been allocated to the client.
   - Returns a DHCP-NAK message to deny the IP address allocation.

After receiving the DHCP-ACK message, the client verifies the following details before using the assigned IP address:

- The assigned IP address is not in use. To verify this, the client broadcasts a gratuitous ARP packet. The assigned IP address is not in use if no response is received within the specified time.
- The assigned IP address is not on the same subnet as any IP address in use on the client.

Otherwise, the client sends a DHCP-DECLINE message to the server to request an IP address again.

IP address lease extension

A dynamically assigned IP address has a lease. When the lease expires, the IP address is reclaimed by the DHCP server. To continue using the IP address, the client must extend the lease duration.

When about half of the lease duration elapses, the DHCP client unicasts a DHCP-REQUEST to the DHCP server to extend the lease. Depending on the availability of the IP address, the DHCP server returns one of the following messages:

- A DHCP-ACK unicast confirming that the client's lease duration has been extended.
- A DHCP-NAK unicast denying the request.

If the client receives no reply, it broadcasts another DHCP-REQUEST message for lease extension when about seven-eighths of the lease duration elapses. Again, depending on the availability of the IP address, the DHCP server returns either a DHCP-ACK unicast or a DHCP-NAK unicast.

DHCP message format

Figure 3 shows the DHCP message format. DHCP uses some of the fields in significantly different ways. The numbers in parentheses indicate the size of each field in bytes.

Figure 3 DHCP message format

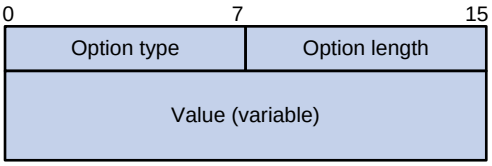


- **op**—Message type defined in options field. 1 = REQUEST, 2 = REPLY
- **htype**, **hlen**—Hardware address type and length of the DHCP client.
- **hops**—Number of relay agents a request message traveled.
- **xid**—Transaction ID, a random number chosen by the client to identify an IP address allocation.
- **secs**—Filled in by the client, the number of seconds elapsed since the client began address acquisition or renewal process. This field is reserved and set to 0.
- **flags**—The leftmost bit is defined as the BROADCAST (B) flag. If this flag is set to 0, the DHCP server sent a reply back by unicast. If this flag is set to 1, the DHCP server sent a reply back by broadcast. The remaining bits of the flags field are reserved for future use.
- **ciaddr**—Client IP address if the client has an IP address that is valid and usable. Otherwise, set to zero. (The client does not use this field to request an IP address to lease.)
- **yiaddr**—Your IP address. It is an IP address assigned by the DHCP server to the DHCP client.
- **siaddr**—Server IP address, from which the client obtained configuration parameters.
- **giaddr**—Gateway IP address. It is the IP address of the first relay agent to which a request message travels.
- **chaddr**—Client hardware address.
- **sname**—Server host name, from which the client obtained configuration parameters.
- **file**—Boot file (also called system software image) name and path information, defined by the server to the client.
- **options**—Optional parameters field that is variable in length. Optional parameters include the message type, lease duration, subnet mask, domain name server IP address, and WINS IP address.

DHCP options

DHCP extends the message format as an extension to BOOTP for compatibility. DHCP uses the options field to carry information for dynamic address allocation and provide additional configuration information for clients.

Figure 4 DHCP option format



Common DHCP options

The following are common DHCP options:

- **Option 3**—Router option. It specifies the gateway address to be assigned to the clients. If both Option 3 and Option 121 exist, Option 3 is ignored.
- **Option 6**—DNS server option. It specifies the DNS server IP address to be assigned to the clients.
- **Option 33**—Static route option. It specifies a list of classful static routes (the destination addresses in these static routes are classful) that a client should add into its routing table. If both Option 33 and Option 121 exist, Option 33 is ignored.
- **Option 51**—IP address lease option.
- **Option 53**—DHCP message type option. It identifies the type of the DHCP message.
- **Option 55**—Parameter request list option. It is used by a DHCP client to request specified configuration parameters. The option includes values that correspond to the parameters requested by the client.
- **Option 60**—Vendor class identifier option. A DHCP client uses this option to identify its vendor. A DHCP server uses this option to distinguish DHCP clients, and assigns IP addresses to them.
- **Option 66**—TFTP server name option. It specifies the TFTP server domain name to be assigned to the clients.
- **Option 67**—Boot file name option. It specifies the boot file name to be assigned to the client.
- **Option 121**—Classless route option. It specifies a list of classless static routes (the destination addresses in these static routes are classless) that a client should add into its routing table. If Option 3, Option 33, and Option 121 all exist, Option 3 and Option 33 are ignored.
- **Option 150**—TFTP server IP address option. It specifies the TFTP server IP address to be assigned to the clients.

For more information about DHCP options, see RFC 2132 and RFC 3442.

Custom DHCP options

Some options, such as Option 43, Option 82, and Option 184, have no standard definitions in RFC 2132.

Vendor-specific option (Option 43)

Option 43 function

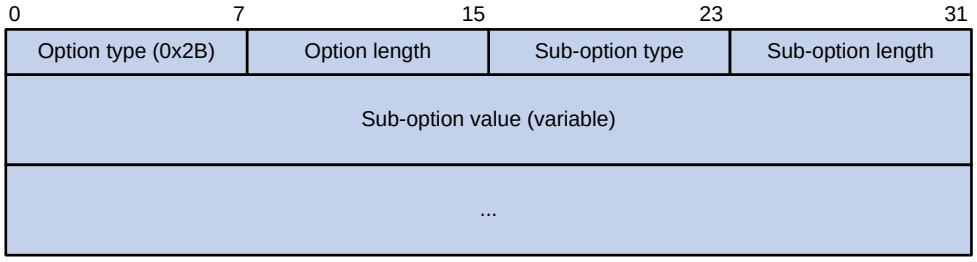
DHCP servers and clients use Option 43 to exchange vendor-specific configuration information.

The DHCP client can obtain the following information through Option 43:

- ACS parameters, including the ACS URL, username, and password.
- Service provider identifier, which is acquired by the CPE from the DHCP server and sent to the ACS for selecting vender-specific configurations and parameters. For more information about CPE and ACS, see *Network Management and Monitoring Configuration Guide*.
- PXE server address, which is used to obtain the boot file or other control information from the PXE server.

Option 43 format

Figure 5 Option 43 format



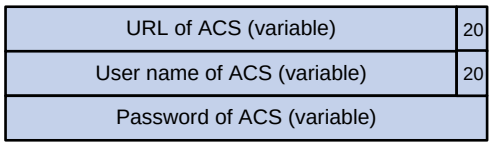
Network configuration parameters are carried in different sub-options of Option 43 as shown in Figure 5.

- **Sub-option type**—The field value can be 0x01 (ACS parameter sub-option), 0x02 (service provider identifier sub-option), or 0x80 (PXE server address sub-option).
- **Sub-option length**—Excludes the sub-option type and sub-option length fields.
- **Sub-option value**—The value format varies by sub-option.

Sub-option value field format

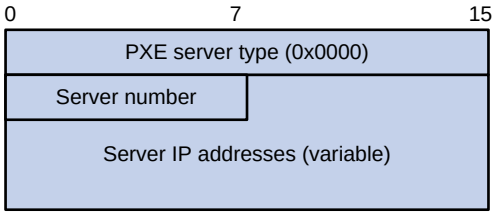
- **ACS parameter sub-option value field**—Includes the ACS URL, username, and password separated by spaces (hexadecimal number 20) as shown in Figure 6.

  Figure 6 ACS parameter sub-option value field

- **Service provider identifier sub-option value field**—Includes the service provider identifier.
- **PXE server address sub-option value field**—Includes the PXE server type that can only be 0, the server number that indicates the number of PXE servers contained in the sub-option, and server IP addresses, as shown in Figure 7.

Figure 7 PXE server address sub-option value field



Relay agent option (Option 82)

Option 82 is the relay agent option. It records the location information about the DHCP client. When a DHCP relay agent or DHCP snooping device receives a client's request, it adds Option 82 to the request and sends it to the server.

The administrator can use Option 82 to locate the DHCP client and further implement security control and accounting. The DHCP server can use Option 82 to provide individual configuration policies for the clients.

Option 82 can contain a maximum of 255 sub-options and must contain a minimum of one sub-option.

The device supports the following sub-options:

- Sub-option 1 (Circuit ID).
- Sub-option 2 (Remote ID).
- Sub-option 5 (Link Selection).
- Sub-option 9 (Vendor-Specific).

Option 82 does not have standard definitions. Its padding formats vary by vendor.

- Circuit ID has the following padding modes:
  - **String padding mode**—Includes a character string specified by the user.
  - **Normal padding mode**—Includes the VLAN ID and interface number of the interface that receives the client's request.
  - **Verbose padding mode**—Includes the access node identifier specified by the user, and the VLAN ID, interface number and interface type of the interface that receives the client's request.
- Remote ID has the following padding modes:
  - **String padding mode**—Includes a character string specified by the user.
  - **Normal padding mode**—Includes the MAC address of the DHCP relay agent interface or the MAC address of the DHCP snooping device that receives the client's request.
  - **Sysname padding mode**—Includes the name of the device. To set the device name, use the **sysname** command in system view.
- The Link Selection sub-option carries one of the following IP addresses:
  - Gateway address in the AAA-authorized address pool or the address pool bound to the relay interface.
  - IP address of the relay interface.

  To have the DHCP relay agent forward requests with Option 82 and sub-option 5, perform the following tasks first:
  - Enable the DHCP relay agent to support Option 82 by using the **dhcp relay information enable** command.

    - Specify a source IP address for relayed DHCP requests by using the **dhcp relay source-address** command.
- The Vendor-Specific sub-option supports only the bas padding mode. The padding content includes the user-configured access node identifier and the VLAN ID, interface number, and interface type of the interface that receives the client's request. This sub-option is supported only on DHCP snooping devices.

Option 184

Option 184 is a reserved option. You can define the parameters in the option as needed. The device supports Option 184 carrying voice related parameters, so a DHCP client with voice functions can get voice parameters from the DHCP server.

Option 184 has the following sub-options:

- **Sub-option 1**—Specifies the IP address of the primary network calling processor. The primary processor acts as the network calling control source and provides program download services. For Option 184, you must define sub-option 1 to make other sub-options take effect.
- **Sub-option 2**—Specifies the IP address of the backup network calling processor. DHCP clients contact the backup processor when the primary one is unreachable.
- **Sub-option 3**—Specifies the voice VLAN ID and the result whether the DHCP client takes this VLAN as the voice VLAN.
- **Sub-option 4**—Specifies the failover route that includes the IP address and the number of the target user. A SIP VoIP user uses this IP address and number to directly establish a connection to the target SIP user when both the primary and backup calling processors are unreachable.

Option 252

Option 252 is a vendor-specific DHCP option that records the verification password from the cloud platform server. On receipt of a reply from the DHCP server, a DHCP client automatically obtains the verification password by parsing Option 252. Then, the client can authenticate to the cloud platform server by using this password. For more information about cloud servers, see cloud connection configuration in *Network Management and Monitoring Configuration Guide*.

Option 253

Option 253 is a vendor-specific DHCP option that records the domain name of a cloud server. On receipt of a reply from the DHCP server, a DHCP client obtains the domain name of the cloud server by parsing Option 253. With the obtained domain name, the DHCP client can then remotely logs in to the cloud server for management purposes. For more information about cloud servers, see cloud connection configuration in *Network Management and Monitoring Configuration Guide*.

To have the device correctly obtain the IPv4 domain name in Option 253, make sure the value in Option 253 use one of the following formats:

- Domain name with a port number specified:
    - **Domain name format**—*domain name:port number*, for example, example.com:17443.
    - **IPv4 address format**—*X.X.X.X:port number*, for example, 192.168.1.1:17443.
    - **IPv6 address format**—*[X:X::X:X]:port number*, for example, [1::1]:64.
- Domain name without a port number specified (the default port number is used):
    - **Domain name format**—*domain name*, for example, example.com.
    - **IPv4 address format**—*X.X.X.X*, for example, 192.168.1.1.
    - **IPv6 address format**—*X:X::X:X* or *[X:X::X:X]*, for example, 1::1 or [1::1].

Protocols and standards

- RFC 2131, *Dynamic Host Configuration Protocol*
- RFC 2132, *DHCP Options and BOOTP Vendor Extensions*
- RFC 1542, *Clarifications and Extensions for the Bootstrap Protocol*
- RFC 3046, *DHCP Relay Agent Information Option*
- RFC 3442, *The Classless Static Route Option for Dynamic Host Configuration Protocol (DHCP) version 4*

Configuring the DHCP server

About DHCP server

A DHCP server manages a pool of IP addresses and client configuration parameters. It selects an IP address and configuration parameters from the address pool and allocates them to a requesting DHCP client.

DHCP address assignment mechanisms

Configure the following address assignment mechanisms as needed:

- **Static address allocation**—Manually bind the MAC address or ID of a client to an IP address in a DHCP address pool. When the client requests an IP address, the DHCP server assigns the IP address in the static binding to the client.
- **Dynamic address allocation**—Specify IP address ranges in a DHCP address pool. Upon receiving a DHCP request, the DHCP server dynamically selects an IP address from the matching IP address range in the address pool.

You can specify IP address ranges in an address pool by using either of the following methods:

- **Method 1**—A primary subnet being divided into multiple address ranges in an address pool
- **Method 2**—A primary subnet and multiple secondary subnets in an address pool

A primary subnet being divided into multiple address ranges in an address pool

An address range includes a common IP address range and IP address ranges for DHCP user classes.

Upon receiving a DHCP request, the DHCP server finds a user class matching the client and selects an IP address in the address range of the user class for the client. A user class can include multiple matching rules, and a client matches the user class as long as it matches any of the rules. In address pool view, you can specify different address ranges for different user classes.

The DHCP server selects an IP address for a client by performing the following steps:

1. DHCP server compares the client against DHCP user classes in the order they are configured.
2. If the client matches a user class, the DHCP server selects an IP address from the address range of the user class.
3. If the matching user class has no assignable addresses, the DHCP server compares the client against the next user class. If all the matching user classes have no assignable addresses, the DHCP server selects an IP address from the common address range.
4. If the DHCP client does not match any DHCP user class, the DHCP server selects an address in the IP address range specified by the **`address range`** command. If the address range has no assignable IP addresses or it is not configured, the address allocation fails.

NOTE:

All address ranges must belong to the primary subnet. If an address range does not reside on the primary subnet, DHCP cannot assign the addresses in the address range.

A primary subnet and multiple secondary subnets in an address pool

The DHCP server selects an IP address from the primary subnet first. If there is no assignable IP address on the primary subnet, the DHCP server selects an IP address from secondary subnets in the order they are configured.

Principles for selecting an address pool

The DHCP server observes the following principles to select an address pool for a client:

1. If the receiving interface has a DHCP policy and the DHCP client matches a user class, the DHCP server selects the address pool that is bound to the matching user class. If no matching user class is found, the server assigns an IP address and other parameters from the default DHCP address pool. If no default address pool is specified or the default address pool does not have assignable IP addresses, the address assignment fails.
2. If there is an address pool where an IP address is statically bound to the MAC address or ID of the client, the DHCP server selects this address pool and assigns the statically bound IP address and other configuration parameters to the client.
3. If the receiving interface has an address pool applied, the DHCP server selects an IP address and other configuration parameters from this address pool.
4. If the above conditions are not met, the DHCP server selects an address pool depending on the client location.
    - **Client on the same subnet as the server**—The DHCP server compares the IP address of the receiving interface with the primary subnets of all address pools.
        - If a match is found, the server selects the address pool with the longest-matching primary subnet.
        - If no match is found, the DHCP server compares the IP address with the secondary subnets of all address pools. The server selects the address pool with the longest-matching secondary subnet.
    - **Client on a different subnet than the server**—The DHCP server compares the IP address in the **giaddr** field of the DHCP request with the primary subnets of all address pools.
        - If a match is found, the server selects the address pool with the longest-matching primary subnet.
        - If no match is found, the DHCP server compares the IP address with the secondary subnets of all address pools. The server selects the address pool with the longest-matching secondary subnet.

For example, two address pools 1.1.1.0/24 and 1.1.1.0/25 are configured but not applied to any DHCP server's interfaces.

- If the IP address of the receiving interface is 1.1.1.1/25, the DHCP server selects the address pool 1.1.1.0/25. If the address pool has no available IP addresses, the DHCP server will not select the other pool and the address allocation will fail.
- If the IP address of the receiving interface is 1.1.1.130/25, the DHCP server selects the address pool 1.1.1.0/24.

To ensure correct address allocation, keep the IP addresses used for dynamic allocation on one of the subnets:

- **Clients on the same subnet as the server**—Subnet where the DHCP server receiving interface resides.
- **Clients on a different subnet than the server**—Subnet where the first DHCP relay interface that faces the clients resides.

NOTE:

As a best practice, configure a minimum of one matching primary subnet in your network. Otherwise, the DHCP server selects only the first matching secondary subnet for address allocation. If the network has more DHCP clients than the assignable IP addresses in the secondary subnet, not all DHCP clients can obtain IP addresses.

IP address allocation sequence

The DHCP server selects an IP address for a client in the following sequence:

1. IP address statically bound to the client's MAC address or ID.
2. IP address that was ever assigned to the client.
3. IP address designated by the Option 50 field in the DHCP-DISCOVER message sent by the client.

   Option 50 is the Requested IP Address option. The client uses this option to specify the wanted IP address in a DHCP-DISCOVER message. The content of Option 50 is user defined.
4. First assignable IP address found in the way discussed in "DHCP address assignment mechanisms" and "Principles for selecting an address pool."
5. IP address that was a conflict or passed its lease duration. If no IP address is assignable, the server does not respond.

NOTE:

- If a client moves to another subnet, the DHCP server selects an IP address in the address pool matching the new subnet. It does not assign the IP address that was once assigned to the client.
- Conflicted IP addresses can be assigned to other DHCP clients only after the addresses are in conflict for an hour.

DHCP server tasks at a glance

To configure the DHCP server, perform the following tasks:

1. (Optional.) Creating a DHCP user class
2. Configuring an address pool on the DHCP server
3. (Optional.) Modifying the address pool selection method on the DHCP server
   - Applying an address pool to an interface
   - Configuring a DHCP policy for dynamic assignment
4. Enabling DHCP
5. Enabling the DHCP server on an interface
6. (Optional.) Configuring advanced DHCP features
   - Configuring IP address conflict detection
   - Enabling handling of Option 82
   - Configuring the DHCP server security features
   - Configuring DHCP server compatibility
   - Setting the DSCP value for DHCP packets sent by the DHCP server
   - Configuring DHCP binding auto backup
   - Enabling client offline detection on the DHCP server
7. (Optional.) Configuring SNMP notification and logging
   - Configuring SNMP notifications for the DHCP server
   - Enabling DHCP logging on the DHCP server
   - Enabling IP exhaustion event logging

Creating a DHCP user class

About this task

The DHCP server classifies DHCP users into different user classes according to the hardware address, option information, or the **giaddr** field in the received DHCP requests. The server allocates IP addresses and configuration parameters to DHCP clients in different user classes.

Procedure

1. Enter system view.

   system-view
2. Create a DHCP user class and enter DHCP user class view.

   dhcp class *class-name*
3. Configure a match rule for the DHCP user class.

   if-match rule *rule-number* { **hardware-address** *hardware-address* **mask** *hardware-address-mask* | **option** *option-code* [**ascii** *ascii-string* [**offset** *offset* | **partial**] | **hex** *hex-string* [**mask** *mask* | **offset** *offset* **length** *length* | **partial**]] | **relay-agent** *gateway-address* }

   By default, no match rule is configured for a DHCP user class.

Configuring an address pool on the DHCP server

DHCP address pool tasks at a glance

To configure a DHCP address pool, perform the following tasks:

1. Creating a DHCP address pool
2. Specifying IP address ranges in a DHCP address pool

   In one DHCP address pool, the two dynamic allocation methods cannot be both configured, but static and dynamic address allocations can be both implemented.
   - Specifying a primary subnet and multiple address ranges in a DHCP address pool
   - Specifying a primary subnet and multiple secondary subnets in a DHCP address pool
   - Configuring a static binding in a DHCP address pool
3. Specifying other configuration parameters to be assigned to DHCP clients
   - Specifying gateways for DHCP clients
   - Specifying a domain name suffix for DHCP clients
   - Specifying DNS servers for DHCP clients
   - Specifying WINS servers and NetBIOS node type for DHCP clients
   - Specifying BIMS server for DHCP clients
   - Specifying the configuration file for DHCP client automatic configuration
   - Specifying a server for DHCP clients
   - Configuring Option 184 parameters for DHCP clients
   - Customizing DHCP options
4. (Optional.) Applying a DHCP address pool to a VPN instance
5. (Optional.) Configuring the DHCP user class whitelist
6. (Optional.) Binding gateways to DHCP server's MAC address
7. (Optional.) Advertising subnets that are assigned to clients

Creating a DHCP address pool

1. Enter system view.
   system-view
2. Create a DHCP address pool and enter its view.
   dhcp server ip-pool *pool-name*

Specifying a primary subnet and multiple address ranges in a DHCP address pool

About this task

Some scenarios need to classify DHCP clients on the same subnet into different address groups. To meet this need, you can configure DHCP user classes and specify different address ranges for the classes. The clients matching a user class can then get the IP addresses of an address range. In addition, you can specify a common address range for the clients that do not match any user class. If no common address range is specified, such clients fail to obtain IP addresses.

If there is no need to classify clients, you do not need to configure DHCP user classes or their address ranges.

Restrictions and guidelines

- If you execute the **network** or **address range** command multiple times for the same address pool, the most recent configuration takes effect.
- If you execute the **forbidden-ip** command multiple times, you exclude multiple address ranges from dynamic allocation.
- IP addresses specified by the **forbidden-ip** command are not assignable in the current address pool, but are assignable in other address pools. IP addresses specified by the **dhcp server forbidden-ip** command are not assignable in any address pool.
- You can use **class range** to modify an existing address range, and the new address range can include IP addresses that are being used by clients. Upon receiving a lease extension request for such an IP address, the DHCP server allocates a new IP address to the requesting client. But the original lease continues aging in the address pool, and will be released when the lease duration is reached. To release such lease without waiting for its timeout, execute the **reset dhcp server ip-in-use** command.
- The **address range** command is mutually exclusive with the **network secondary** command. You cannot use them together in the same address pool.
- The **class range** command is mutually exclusive with the **network secondary** command. You cannot use them together in the same address pool.

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Specify the primary subnet in the address pool.
   network *network-address* [*mask-length* | **mask** *mask*]
   By default, no primary subnet is specified.
4. (Optional.) Specify the common address range.
   address range *start-ip-address end-ip-address*

By default, no IP address range is specified.

5. (Optional.) Specify an IP address range for a DHCP user class.

   class *class-name* **range** *start-ip-address end-ip-address*

   By default, no IP address range is specified for a user class.

   The DHCP user class must already be created by using the **dhcp class** command.

6. (Optional.) Set the address lease duration.

   expired { **day** *day* [**hour** *hour* [**minute** *minute* [**second** *second*]]] | **unlimited** }

   The default setting is 1 day.

7. (Optional.) Exclude the specified IP addresses in the address pool from dynamic allocation.

   forbidden-ip *ip-address*&<1-8>

   By default, all IP addresses in the DHCP address pool are assignable.

8. (Optional.) Exclude the specified IP addresses from automatic allocation in system view.

   a. Return to system view.

      quit

   b. Exclude the specified IP addresses from automatic allocation globally.

      dhcp server forbidden-ip *start-ip-address* [*end-ip-address*] [**vpn-instance** *vpn-instance-name*]

      By default, except for the IP address of the DHCP server interface, IP addresses in all address pools are assignable.

Specifying a primary subnet and multiple secondary subnets in a DHCP address pool

About this task

If an address pool has a primary subnet and multiple secondary subnets, the server assigns IP addresses on a secondary subnet when the primary subnet has no assignable IP addresses.

Restrictions and guidelines

IP addresses specified by the **forbidden-ip** command are not assignable in the current address pool, but are assignable in other address pools. IP addresses specified by the **dhcp server forbidden-ip** command are not assignable in any address pool.

The **network secondary** command is mutually exclusive with the **address range** command or the **class range** command. You cannot use them together in the same address pool.

Specifying a primary subnet and multiple secondary subnets

1. Enter system view.

   system-view

2. Enter DHCP address pool view.

   dhcp server ip-pool *pool-name*

3. Specify the primary subnet.

   network *network-address* [*mask-length* | **mask** *mask*]

   By default, no primary subnet is specified.

   You can specify only one primary subnet in each address pool. If you execute the **network** command multiple times, the most recent configuration takes effect.

4. (Optional.) Specify a secondary subnet.

network *network-address* [*mask-length* | **mask** *mask*] **secondary**

By default, no secondary subnet is specified.

You can specify a maximum of 32 secondary subnets in one address pool.

5. (Optional.) Return to address pool view.

quit

Setting the lease duration for dynamically allocation IP addresses

1. Enter system view.

system-view

2. Enter DHCP address pool view.

dhcp server ip-pool *pool-name*

3. Set the address lease duration.

expired { **day** *day* [**hour** *hour* [**minute** *minute* [**second** *second*]]] | **unlimited** }

The default setting is 1 day.

Excluding IP addresses from dynamic allocation

1. Enter system view.

system-view

2. Enter DHCP address pool view.

dhcp server ip-pool *pool-name*

3. Exclude the specified IP addresses from dynamic allocation.

forbidden-ip *ip-address*&<1-8>

By default, all IP addresses in the DHCP address pool are assignable.

To exclude multiple address ranges from the address pool, repeat this step.

4. (Optional.) Exclude the specified IP addresses from dynamic allocation in system view.

   a. Return to system view.

   quit

   b. Exclude the specified IP addresses from dynamic allocation globally.

   dhcp server forbidden-ip *start-ip-address* [*end-ip-address*] [**vpn-instance** *vpn-instance-name*]

   By default, except for the IP address of the DHCP server interface, IP addresses in all address pools are assignable.

   To exclude multiple address ranges globally, repeat this step.

Configuring a static binding in a DHCP address pool

About this task

Some DHCP clients, such as a WWW server, need fixed IP addresses. To provide a fixed IP address for a client, you can statically bind the MAC address or ID of the client to an IP address in a DHCP address pool. When the client requests an IP address, the DHCP server assigns the IP address in the static binding to the client.

Restrictions and guidelines

- The IP address of a static binding cannot be the address of the DHCP server interface. Otherwise, an IP address conflict occurs and the bound client cannot obtain an IP address correctly.

- Multiple interfaces on the same device might all use DHCP to request a static IP address. In this case, use client IDs rather than the device's MAC address to identify the interfaces. Otherwise, IP address allocation will fail.

Procedure

1. Enter system view.

   system-view

2. Enter DHCP address pool view.

   dhcp server ip-pool *pool-name*

3. Configure a static binding.

   static-bind ip-address *ip-address* [*mask-length* | **mask** *mask*] { **client-identifier** *client-identifier* | **hardware-address** *hardware-address* [**ethernet** | **token-ring**] }

   By default, no static binding is configured.

   One IP address can be bound to only one client MAC or client ID. You cannot modify bindings that have been created. To change the binding for a DHCP client, you must delete the existing binding first.

4. (Optional.) Set the lease duration for the IP address.

   expired { **day** *day* [**hour** *hour* [**minute** *minute* [**second** *second*]]] | **unlimited** }

   By default, the lease duration is 1 day.

Specifying gateways for DHCP clients

About this task

DHCP clients send packets destined for other networks to a gateway. The DHCP server can assign the gateway address to the DHCP clients.

Restrictions and guidelines

You can specify gateway addresses in each address pool on the DHCP server. A maximum of 64 gateways can be specified in DHCP address pool view or secondary subnet view.

The DHCP server assigns gateway addresses to clients on a secondary subnet in the following ways:

- If gateways are specified in both address pool view and secondary subnet view, DHCP assigns those specified in the secondary subnet view.
- If gateways are specified in address pool view but not in secondary subnet view, DHCP assigns those specified in address pool view.

Procedure

1. Enter system view.

   system-view

2. Enter DHCP address pool view.

   dhcp server ip-pool *pool-name*

3. Specify gateways.

   gateway-list *ip-address*&<1-64>

   By default, no gateway is specified.

4. (Optional.) Specify gateways in secondary subnet view.

   a. Enter secondary subnet view.

      network *network-address* [*mask-length* | **mask** *mask*] **secondary**

b. Specify gateways.

gateway-list *ip-address*&<1-64>

By default, no gateway is specified.

Specifying a domain name suffix for DHCP clients

About this task

You can specify a domain name suffix in a DHCP address pool on the DHCP server. With this suffix assigned, the client only needs to input part of a domain name, and the system adds the domain name suffix for name resolution. For more information about DNS, see "Configuring DNS."

Procedure

1. Enter system view.

   system-view
2. Enter DHCP address pool view.

   dhcp server ip-pool *pool-name*
3. Specify a domain name suffix.

   domain-name *domain-name*

   By default, no domain name is specified.

Specifying DNS servers for DHCP clients

About this task

To access hosts on the Internet through domain names, a DHCP client must contact a DNS server to resolve names. You can specify up to eight DNS servers in a DHCP address pool.

Procedure

1. Enter system view.

   system-view
2. Enter DHCP address pool view.

   dhcp server ip-pool *pool-name*
3. Specify DNS servers.

   dns-list *ip-address*&<1-8>

   By default, no DNS server is specified.

Specifying WINS servers and NetBIOS node type for DHCP clients

About this task

A Microsoft DHCP client using NetBIOS protocol must contact a WINS server for name resolution.

In addition, you must specify one of the following NetBIOS node types to approach name resolution:

- **b (broadcast)-node**—A b-node client sends the destination name in a broadcast message. The destination returns its IP address to the client after receiving the message.
- **p (peer-to-peer)-node**—A p-node client sends the destination name in a unicast message to the WINS server. The WINS server returns the destination IP address.

- **m (mixed)-node**—An m-node client broadcasts the destination name. If it receives no response, it unicasts the destination name to the WINS server to get the destination IP address.
- **h (hybrid)-node**—An h-node client unicasts the destination name to the WINS server. If it receives no response, it broadcasts the destination name to get the destination IP address.

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
   By default, no DHCP address pool exists.
3. Specify WINS servers.
   nbns-list *ip-address*&<1-8>
   By default, no WINS server is specified.
   This step is optional for b-node. You can specify a maximum of eight WINS servers for such clients in one DHCP address pool.
4. Specify the NetBIOS node type.
   netbios-type { **b-node** | **h-node** | **m-node** | **p-node** }
   By default, no NetBIOS node type is specified.

Specifying BIMS server for DHCP clients

About this task

Perform this task to provide the BIMS server IP address, port number, and shared key for the clients. The DHCP clients contact the BIMS server to get configuration files and perform software upgrade and backup.

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Specify the BIMS server IP address, port number, and shared key.
   bims-server ip *ip-address* [**port** *port-number*] **sharekey** { **cipher** | **simple** } *string*
   By default, no BIMS server information is specified.

Specifying the configuration file for DHCP client automatic configuration

About this task

Automatic configuration enables a device to automatically obtain a set of configuration settings at startup. The server-based automatic configuration requires the cooperation of the DHCP server and file server (TFTP or HTTP server). The device uses the obtained parameters to contact the file server to get the configuration file. For more information about automatic configuration, see *Fundamentals Configuration Guide*.

Specifying the configuration file on a TFTP file server

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
   By default, no DHCP address pool exists.
3. Specify the IP address or the name of a TFTP server.
   - Specify the IP address of the TFTP server.
     tftp-server ip-address *ip-address*
     By default, no TFTP server IP address is specified.
   - Specify the name of the TFTP server.
     tftp-server domain-name *domain-name*
     By default, no TFTP server name is specified.
4. Specify the configuration file name.
   bootfile-name *bootfile-name*
   By default, no configuration file name is specified.

Specifying the URL of the configuration file on an HTTP file server

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Specify the URL of the configuration file.
   bootfile-name *url*
   By default, no configuration file URL is specified.

Specifying a server for DHCP clients

About this task

Some DHCP clients need to obtain configuration information from a server, such as a TFTP server. You can specify the IP address of that server. The DHCP server sends the server's IP address to DHCP clients along with other configuration information.

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Specify the IP address of a server.
   next-server *ip-address*
   By default, no server is specified.

Configuring Option 184 parameters for DHCP clients

About this task

To assign calling parameters to DHCP clients with voice service, you must configure Option 184 on the DHCP server. For more information about Option 184, see "Option 184."

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Specify the IP address of the primary network calling processor.
   voice-config ncp-ip *ip-address*
   By default, no primary network calling processor is specified.
   After you configure this command, the other Option 184 parameters take effect.
4. (Optional.) Specify the IP address of the backup server.
   voice-config as-ip *ip-address*
   By default, no backup network calling processor is specified.
5. (Optional.) Configure the voice VLAN.
   voice-config voice-vlan *vlan-id* { **disable** | **enable** }
   By default, no voice VLAN is configured.
6. (Optional.) Specify the failover IP address and dialer string.
   voice-config fail-over *ip-address dialer-string*
   By default, no failover IP address or dialer string is specified.

Customizing DHCP options

DHCP option customization applications

You can customize DHCP options for the following purposes:

- Add newly released options.
- Add options for which the vendor defines the contents, for example, Option 43.
- Add options for which the CLI does not provide a dedicated configuration command. For example, you can use the **option 4 ip-address 1.1.1.1** command to define the time server address 1.1.1.1 for DHCP clients.
- Add all option values if the actual requirement exceeds the limit for a dedicated option configuration command. For example, the **dns-list** command can specify up to eight DNS servers. To specify more than eight DNS servers, you must use the **option 6** command to define all DNS servers.

Common DHCP options

Table 1 lists common DHCP options and their parameters.

Table 1 Common DHCP options

Option	Option name	Corresponding command	Recommended parameter in the `option` command
3	Router Option	**gateway-list**	**ip-address**

Option	Option name	Corresponding command	Recommended parameter in the `option` command
6	Domain Name Server Option	**dns-list**	**ip-address**
15	Domain Name	**domain-name**	**ascii**
44	NetBIOS over TCP/IP Name Server Option	**nbns-list**	**ip-address**
46	NetBIOS over TCP/IP Node Type Option	**netbios-type**	**hex**
66	TFTP server name	**tftp-server**	**ascii**
67	Boot file name	**bootfile-name**	**ascii**
43	Vendor Specific Information	N/A	**hex**
252	Opt252 Password	N/A	**ascii**
253	Cloud Domain Name	N/A	**ascii**

Restrictions and guidelines

Use caution when customizing DHCP options because the configuration might affect DHCP operation.

You can customize a DHCP option in a DHCP address pool

You can customize a DHCP option in a DHCP option group, and specify the option group for a user class in an address pool. A DHCP client in the user class will obtain the option configuration.

Customizing a DHCP option in a DHCP address pool

1. Enter system view.

   system-view
2. Enter DHCP address pool view.

   dhcp server ip-pool *pool-name*
3. Customize a DHCP option.

   option *code* { **ascii** *ascii-string* | **hex** *hex-string* | **ip-address** *ip-address*&<1-8> }

   By default, no DHCP option is customized in a DHCP address pool.

   DHCP options specified in DHCP option groups take precedence over those specified in DHCP address pools.

Customizing a DHCP option in a DHCP option group

1. Enter system view.

   system-view
2. Create a DHCP option group and enter DHCP option group view.

   dhcp option-group *option-group-number*
3. Customize a DHCP option.

   option *code* { **ascii** *ascii-string* | **hex** *hex-string* | **ip-address** *ip-address*&<1-8> }

   By default, no DHCP option is customized in a DHCP option group.

   If multiple DHCP option groups have the same option, the server selects the option in the DHCP option group first matching the user class.

4. Return to system view.
   quit
5. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
6. Specify the DHCP option group for the DHCP user class.
   class *class-name* **option-group** *option-group-number*
   By default, no DHCP option group is specified for a DHCP user class.

Applying a DHCP address pool to a VPN instance

About this task

If a DHCP address pool is applied to a VPN instance, the DHCP server assigns IP addresses in this address pool to clients in the VPN instance. Addresses in this address pool will not be assigned to clients on the public network.

The DHCP server can obtain the VPN instance to which a DHCP client belongs from the following information:

- The client's VPN information stored in authentication modules, such as IPoE.
- The VPN information of the DHCP server's interface that receives DHCP packets from the client.

If both VPN instances can be obtained, the VPN information from authentication modules takes priority over the VPN information of the receiving interface.

An MCE acting as the DHCP server can assign IP addresses not only to clients on public networks, but also to clients on private networks. The IP address ranges of public and private networks or those of private networks on the DHCP server cannot overlap. For more information about MCE, see *MPLS Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Apply the address pool to a VPN instance.
   vpn-instance *vpn-instance-name*
   By default, the address pool is not applied to any VPN instance.

Configuring the DHCP user class whitelist

About this task

The DHCP user class whitelist functions as follows:

- When no online DHCP users exist, the DHCP server only processes the following requests:
  - Requests from users in the DHCP user class whitelist.
  - Lease renewal requests from users that are not in the DHCP user class whitelist. The DHCP server will reply DHCP-NAK messages to those requests.
- When online DHCP users exist, the DHCP server processes requests only from users in the DHCP user class whitelist.

Restrictions and guidelines

The whitelist does not take effect on clients who request static IP addresses, and the server always processes their requests.

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Enable the DHCP user class whitelist.
   verify class
   By default, the DHCP user class whitelist is disabled.
4. Add DHCP user classes to the DHCP user class whitelist.
   valid class *class-name*&<1-8>
   By default, no DHCP user class is on the DHCP user class whitelist.

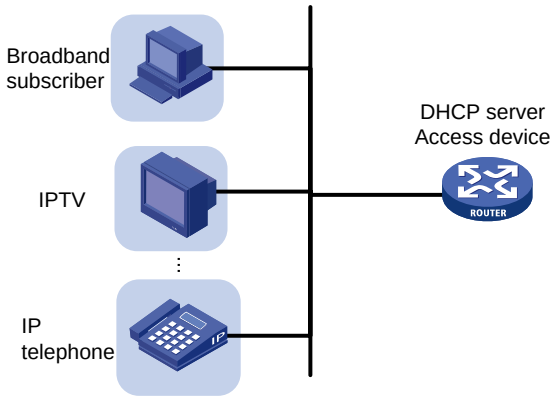
Binding gateways to DHCP server's MAC address

About this task

This feature enables the DHCP server to assign different gateway IP addresses to DHCP clients. In addition, the DHCP server uses the gateway IP addresses and the server's MAC address to reply to ARP requests from the clients.

As shown in Figure 8, the DHCP server is configured on the access device that provides access for clients of different service types, such as broadband, IPTV, and IP telephone. The clients of different types obtain IP addresses on different subnets. For the clients to access the network, the access interface typically has no IP address configured. You must bind the gateways to the DHCP server's MAC address when specifying gateways for the DHCP clients.

Figure 8 Network diagram



Restrictions and guidelines

The gateway binding feature on the master device takes effect if the DHCP address pool is bound to a VSRP instance.

If the address pool is applied to a VPN instance, make sure the VPN instance already exists.

Procedure

1. Enter system view.
   system-view

2. Enter DHCP address pool view.
   dhcp server ip-pool *pool-name*
3. Bind the gateways to the device's MAC address.
   gateway-list *ip-address*&<1-64> **export-route**
   By default, gateways are not bound to any MAC address.

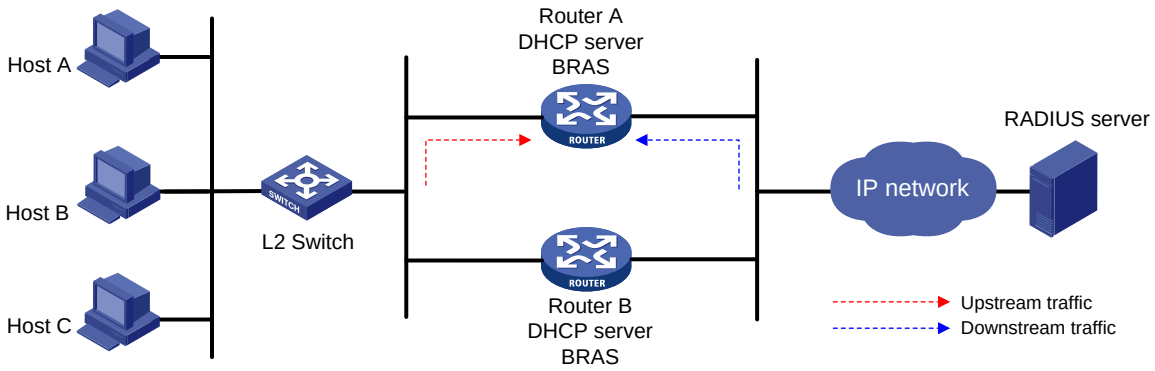
Advertising subnets that are assigned to clients

About this task

This feature enables the route management module to advertise subnets assigned to DHCP clients. This feature achieves symmetric routing for traffic of the same host.

As shown in Figure 9, Router A and Router B act as both the DHCP server and the BRAS device. The BRAS devices send accounting packets to the RADIUS server. To enable the BRAS devices to collect correct accounting information for each RADIUS user, configure the DHCP server to advertise subnets assigned to clients. The upstream and downstream traffic of a RADIUS user will pass through the same BRAS device.

Figure 9 Network diagram



Restrictions and guidelines

The subnet advertising on the master device takes effect if the DHCP address pool is bound to a VSRP instance. If the address pool is applied to a VPN instance, make sure the VPN instance already exists.

Procedure

1. Enter system view.
   system-view
2. Create a DHCP address pool and enter its view.
   dhcp server ip-pool *pool-name*
3. Advertise subnets assigned to DHCP clients.
   network *network-address* [*mask-length* | **mask** *mask*] [**secondary**] **export-route**
   By default, the subnets assigned to DHCP clients are not advertised.

Applying an address pool to an interface

About this task

Perform this task to apply a DHCP address pool to an interface.

Upon receiving a DHCP request from the interface, the DHCP server performs address allocation in the following ways:

- If a static binding is found for the client, the server assigns the static IP address and configuration parameters from the address pool that contains the static binding.
- If no static binding is found for the client, the server uses the address pool applied to the interface for address and configuration parameter allocation.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Apply an address pool to the interface.

   dhcp server apply ip-pool *pool-name*

   By default, no address pool is applied to an interface.

   If the applied address pool does not exist, the DHCP server fails to perform dynamic address allocation.

Configuring a DHCP policy for dynamic assignment

About this task

In a DHCP policy, each DHCP user class has a bound DHCP address pool. Clients matching different user classes obtain IP addresses and other parameters from different address pools. The DHCP policy must be applied to the interface that acts as the DHCP server. When receiving a DHCP request, the DHCP server compares the packet against the user classes in the order that they are configured.

- If a matching user class is found and the bound address pool has assignable IP addresses, the server assigns an IP address and other parameters from the address pool. If the address pool does not have assignable IP addresses, the address assignment fails.
- If no match is found, the server assigns an IP address and other parameters from the default DHCP address pool. If no default address pool is specified or the default address pool does not have assignable IP addresses, the address assignment fails.

For successful address assignment, make sure the applied DHCP policy and the bound address pools exist.

Restrictions and guidelines

A DHCP policy take effect only after it is applied to an interface.

Procedure

1. Enter system view.

   system-view
2. Create a DHCP policy and enter DHCP policy view.

   dhcp policy *policy-name*
3. Specify a DHCP address pool for a DHCP user class.

   class *class-name* **ip-pool** *pool-name*

   By default, no address pool is specified for a user class.
4. Specify the default DHCP address pool.

default ip-pool *pool-name*

By default, no default address pool is specified.

5. Return to system view.

   quit

6. Enter interface view.

   interface *interface-type interface-number*

7. Apply the DHCP policy to the interface.

   dhcp apply-policy *policy-name*

   By default, no DHCP policy is applied to an interface.

Enabling DHCP

Restrictions and guideline

You must enable DHCP to make other DHCP configurations take effect.

Procedure

1. Enter system view.

   system-view

2. Enable DHCP.

   dhcp enable

   By default, DHCP is disabled.

Enabling the DHCP server on an interface

About this task

Perform this task to enable the DHCP server on an interface. Upon receiving a DHCP request on the interface, the DHCP server assigns the client an IP address and other configuration parameters from a DHCP address pool.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the DHCP server on the interface.

   dhcp select server

   By default, the DHCP server is enabled on the interface.

Configuring IP address conflict detection

About this task

Before assigning an IP address, the DHCP server pings that IP address.

- If the server receives a response within the specified period, it selects and pings another IP address.
- If it receives no response, the server continues to ping the IP address until the maximum number of ping packets are sent. If still no response is received, the server assigns the IP

address to the requesting client. The DHCP client uses gratuitous ARP to perform IP address conflict detection.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. (Optional.) Set the maximum number of ping packets to be sent for conflict detection.
   ```
   dhcp server ping packets number
   ```
   The default setting is one.
   To disable IP address conflict detection, set the value to **0**.
3. (Optional.) Set the ping timeout time.
   ```
   dhcp server ping timeout milliseconds
   ```
   The default setting is 500 ms.
   To disable IP address conflict detection, set the value to **0**.

Enabling handling of Option 82

About this task

Perform this task to enable the DHCP server to handle Option 82. Upon receiving a DHCP request that contains Option 82, the DHCP server adds Option 82 into the DHCP response.

If you disable the DHCP to handle Option 82, it does not add Option 82 into the response message.

You must enable handling of Option 82 on both the DHCP server and the DHCP relay agent to ensure correct processing for Option 82. For information about enabling handling of Option 82 on the DHCP relay agent, see "Configuring DHCP relay agent support for Option 82."

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable the server to handle Option 82.
   ```
   dhcp server relay information enable
   ```
   By default, handling of Option 82 is enabled.

Configuring the DHCP server security features

Restrictions and guidelines

The DHCP server security features are not applicable if a DHCP relay agent exists in the network. This is because the MAC address of the DHCP relay agent is encapsulated as the source MAC address in the DHCP request received by the DHCP server. In this case, you must configure the DHCP relay agent security features. For more information, see "Configuring the DHCP relay agent security features."

Configuring DHCP flood attack protection

About this task

The DHCP flood attack protection enables the DHCP server to detect DHCP flood attacks according to the DHCP packet rate threshold on a per-MAC basis.

When the DHCP server receives a DHCP packet from a client (MAC address), it creates a DHCP flood attack entry in check state. If the number of DHCP packets from the same MAC address reaches the upper limit in the detection duration, the server determines that the client is launching a DHCP flood attack. The DHCP flood attack entry changes to the restrain state, and the DHCP server discards the DHCP packets from that client. When the aging time of the entry is reached, the DHCP server deletes the entry. If a DHCP packet from the MAC address arrives later, the DHCP server will create a flood attack entry and count the number of incoming DHCP packets for that client again.

If you enable this feature in VXLAN VSI view, the DHCP server will perform the following operations:

- Collect statistics about DHCP packets with the same source MAC address received on each VSI interface.
- Create a DHCP flood attack entry in check state for each source MAC address.

Configuring DHCP flood attack protection in a VXLAN network

1. Enter system view.

   system-view
2. (Optional.) Set the DHCP packet rate threshold for DHCP flood attack detection.

   dhcp flood-protection threshold *packet-number milliseconds*

   By default, the device allows a maximum of 6 DHCP packets per 5000 milliseconds from each DHCP client.
3. (Optional.) Set the DHCP flood attack entry aging time.

   dhcp flood-protection aging-time *time*

   The default setting is 300 seconds.
4. Enter VSI view.

   vsi *vsi-name*
5. Enable DHCP flood attack protection.

   dhcp flood-protection enable

   By default, DHCP flood attack protection is disabled.

Configuring DHCP starvation attack protection

About this task

A DHCP starvation attack occurs when an attacker constantly sends forged DHCP requests using different MAC addresses in the **chaddr** field to a DHCP server. This exhausts the IP address resources of the DHCP server so legitimate DHCP clients cannot obtain IP addresses. The DHCP server might also fail to work because of exhaustion of system resources. For information about the fields in the DHCP messages, see "DHCP message format."

The following methods are available to relieve or prevent such attacks.

- To relieve a DHCP starvation attack that uses DHCP packets encapsulated with different source MAC addresses, perform the following configuration on an interface:
  - Execute the **mac-address max-mac-count** command to set the MAC learning limit. For more information about this command, see *Layer 2—LAN Switching Command Reference*.
  - Disable unknown frame forwarding when the MAC learning limit is reached.
- To prevent a DHCP starvation attack that uses DHCP requests encapsulated with the same source MAC address, you can enable MAC address check on the DHCP server. The DHCP server compares the **chaddr** field of a received DHCP request with the source MAC address in the frame header. If they are the same, the DHCP server verifies this request as legal and processes it. If they are not the same, the server discards the DHCP request.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable MAC address check.
   dhcp server check mac-address
   By default, MAC address check is disabled.

Configuring DHCP server compatibility

Perform this task to enable the DHCP server to support DHCP clients that are incompliant with RFC.

Configuring the DHCP server to always broadcast responses

About this task

By default, the DHCP server broadcasts a response only when the broadcast flag in the DHCP request is set to 1. You can configure the DHCP server to ignore the broadcast flag and always broadcast a response. This feature is useful when some clients set the broadcast flag to 0 but do not accept unicast responses.

The DHCP server always unicasts a response in the following situations, regardless of whether this feature is configured or not:

- The DHCP request is from a DHCP client that has an IP address (the **ciaddr** field is not 0).
- The DHCP request is forwarded by a DHCP relay agent from a DHCP client (the **giaddr** field is not 0).

Procedure

1. Enter system view.
   system-view
2. Enable the DHCP server to always broadcast all responses.
   dhcp server always-broadcast
   By default, the DHCP server reads the broadcast flag to decide whether to broadcast or unicast a response.

Enabling the DHCP server to return a DHCP-NAK message upon client notions of incorrect IP addresses

About this task

A DHCP client can send a DHCP-REQUEST message directly or upon receiving a DHCP-OFFER message. Upon receiving the request, the DHCP server will check if the client notion of its IP address is correct. If the requested IP address is different from the allocated one or has no matching lease record, the DHCP server remains silent by default. After the allocated IP address lease for the client expires, the DHCP server will make response to request from the client.

This feature enables the DHCP server to return DHCP-NAK messages if the client notions of their IP addresses are incorrect. After receiving the DHCP-NAK message, the DHCP client will request an IP address again.

Procedure

1. Enter system view.

   ```
   system-view
   ```

2. Enable the DHCP server to return a DHCP-NAK message if the client notions of their IP addresses are incorrect.

   ```
   dhcp server request-ip-address check
   ```

   The DHCP server does not return a DHCP-NAK message if the client notions of their IP addresses are incorrect.

Configuring the DHCP server to ignore BOOTP requests

About this task

The lease duration of the IP addresses obtained by the BOOTP clients is unlimited. For some scenarios that do not allow unlimited leases, you can configure the DHCP server to ignore BOOTP requests.

Procedure

1. Enter system view.

   ```
   system-view
   ```

2. Configure the DHCP server to ignore BOOTP requests.

   ```
   dhcp server bootp ignore
   ```

   By default, the DHCP server processes BOOTP requests.

Configuring the DHCP server to send BOOTP responses in RFC 1048 format

About this task

Not all BOOTP clients can send requests that are compatible with RFC 1048. By default, the DHCP server does not process the Vend field of RFC 1048-incompliant requests but copies the Vend field into responses.

This feature enables the DHCP server to fill the Vend field in RFC 1048-compliant format in DHCP responses to RFC 1048-incompliant requests sent by BOOTP clients.

Procedure

1. Enter system view.

   ```
   system-view
   ```

2. Enable the DHCP server to send BOOTP responses in RFC 1048 format to the RFC 1048-incompliant BOOTP requests.

   ```
   dhcp server bootp reply-rfc-1048
   ```

   By default, the DHCP server directly copies the Vend field of such requests into the responses.

Setting the DSCP value for DHCP packets sent by the DHCP server

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet.

Procedure

1. Enter system view.

   system-view

2. Set the DSCP value for DHCP packets sent by the DHCP server.

   dhcp dscp *dscp-value*

   By default, the DSCP value in DHCP packets sent by the DHCP server is 56.

Configuring DHCP binding auto backup

About this task

The auto backup feature saves bindings to a backup file and allows the DHCP server to download the bindings from the backup file at the server reboot. The bindings include the lease bindings and conflicted IP addresses. They cannot survive a reboot on the DHCP server.

The DHCP server does not provide services during the download process. If a connection error occurs during the process and cannot be repaired in a short amount of time, you can terminate the download operation. Manual interruption allows the DHCP server to provide services without waiting for the connection to be repaired.

Procedure

1. Enter system view.

   system-view

2. Configure the DHCP server to back up the bindings to a file.

   dhcp server database filename { *filename* | **url** *url* [**username** *username* [**password** { **cipher** | **simple** } *string*]] }

   By default, the DHCP server does not back up the DHCP bindings.

   With this command executed, the DHCP server backs up its bindings immediately and runs auto backup.

3. (Optional.) Manually save the DHCP bindings to the backup file.

   dhcp server database update now

4. (Optional.) Set the waiting time after a DHCP binding change for the DHCP server to update the backup file.

   dhcp server database update interval *interval*

   By default, the DHCP server waits 300 seconds to update the backup file after a DHCP binding change. If no DHCP binding changes, the backup file is not updated.

5. (Optional.) Terminate the download of DHCP bindings from the backup file.

   dhcp server database update stop

   This command only triggers one termination.

Enabling client offline detection on the DHCP server

About this task

The client offline detection feature reclaims an assigned IP address and deletes the binding entry when the ARP entry for the IP address ages out.

Restrictions and guidelines

The feature does not function if an ARP entry is manually deleted.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable client offline detection.
   dhcp client-detect
   By default, client offline detection is disabled on the DHCP server.

Configuring SNMP notifications for the DHCP server

About SNMP DHCP server notifications

Perform this task to configure the device to send SNMP notifications of DHCP server alarms. The following are available DHCP server alarm notifications:

- **IP exhaustion notifications**—All assignable IP addresses in an IP pool are used up or an IP pool has recovered from the IP exhaustion condition.
- **IP usage alarm notifications**—The IP address usage of an IP pool has reached or exceeded the specified threshold, or an IP pool has recovered from an IP usage alarm condition.

The SNMP notifications are sent to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

When the IP pool usage is not lower than the threshold or drops below 90% of the threshold, the DHCP module also sends a log message to the information center. For log messages to be sent correctly, configure the information center to set log message filtering and output rules, including output destinations. For information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Enabling IP exhaustion notifications

1. Enter system view.
   system-view
2. Enable IP exhaustion notifications.
   snmp-agent trap enable dhcp server address-exhaust
   By default, IP exhaustion notifications are enabled.

3. Enter IP address pool view.
   dhcp server ip-pool *pool-name*
4. (Optional.) Enable IP exhaustion notifications for an IP pool.
   exhaustion trap enable
   By default, IP exhaustion notifications are enabled for an IP pool.

Enabling IP usage alarm notifications

1. Enter system view.
   system-view
2. Enable IP usage alarm notifications.
   snmp-agent trap enable dhcp server ip-in-use
   By default, IP usage alarm notifications are enabled.
3. Enter IP pool view.
   dhcp server ip-pool *pool-name*
4. (Optional.) Set the IP address usage threshold for the IP pool.
   ip-in-use threshold *threshold-value*
   The default threshold is 100%.

Enabling DHCP logging on the DHCP server

About this task

The DHCP logging feature enables the DHCP server to generate DHCP logs and send them to the information center. The information helps administrators locate and solve problems. For information about the log destination and output rule configuration in the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

As a best practice, disable this feature if the log generation affects the device performance or reduces the address allocation efficiency. For example, this situation might occur when a large number of clients frequently come online or go offline.

Procedure

1. Enter system view.
   system-view
2. Enable DHCP logging.
   dhcp log enable
   By default, DHCP logging is disabled.

Enabling IP exhaustion event logging

About this task

This feature enables the DHCP module to send IP exhaustion logs to the information center. IP exhaustion events include IP exhaustion alarms and recoveries from IP exhaustion alarm conditions.

An IP exhaustion log message is generated for an IP pool when one of the following events occurs:

NOTE:

If the IP address usage threshold for the IP pool is 100%, the DHCP module generates an alarm message instead of a log message for one of the following events. To set the IP address usage threshold for an IP pool, use the **ip-in-use threshold** command.

- The IP pool does not have assignable IP addresses.
- The IP address usage of the IP pool drops below 90% after IP address exhaustion. The IP address usage is calculated by using the following formula:

  (Total number of IP addresses – number of assignable IP addresses)/total number of IP addresses

For log messages to be sent correctly, configure the information center to set log message filtering and output rules, including output destinations. For more information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

This feature is not affected if DHCP logging is disabled on the DHCP server.

Procedure

1. Enter system view.

   system-view
2. Enter IP pool view.

   dhcp server ip-pool *pool-name*
3. Enable IP exhaustion event logging.

   exhaustion log enable

   By default, IP exhaustion event logging is disabled.

Display and maintenance commands for DHCP server

IMPORTANT:

A restart of the DHCP server or execution of the **reset dhcp server ip-in-use** command deletes all lease information. The DHCP server denies any DHCP request for lease extension, and the client must request an IP address again.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about IP address conflicts.	**display dhcp server conflict** [**ip** *ip-address*] [**vpn-instance** *vpn-instance-name*]
Display information about DHCP binding auto backup.	**display dhcp server database**
Display information about lease-expired IP addresses.	**display dhcp server expired** [[**ip** *ip-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Display information about assignable IP addresses.	**display dhcp server free-ip** [**pool** *pool-name* | **vpn-instance** *vpn-instance-name*]

Task	Command
Display information about assigned IP addresses.	**display dhcp server ip-in-use** [[**ip** *ip-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Display information about DHCP address pools.	**display dhcp server pool** [*pool-name* | **vpn-instance** *vpn-instance-name*]
Display DHCP server statistics.	**display dhcp server statistics** [**pool** *pool-name* | **vpn-instance** *vpn-instance-name*]
Clear information about IP address conflicts.	**reset dhcp server conflict** [**ip** *ip-address*] [**vpn-instance** *vpn-instance-name*]
Clear information about lease-expired IP addresses.	**reset dhcp server expired** [[**ip** *ip-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Clear information about assigned IP addresses.	**reset dhcp server ip-in-use** [[**ip** *ip-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Clear DHCP server statistics.	**reset dhcp server statistics** [**vpn-instance** *vpn-instance-name*]

DHCP server configuration examples

Example: Configuring static IP address assignment

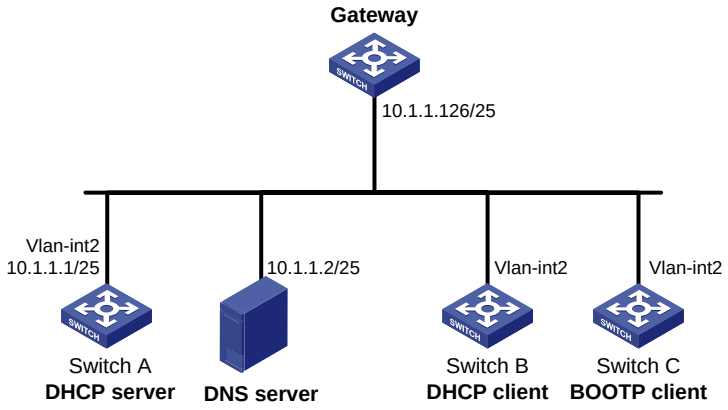
Network configuration

As shown in Figure 10, Switch B (DHCP client) and Switch C (BOOTP client) obtain the IP address, DNS server address, and gateway address from Switch A (DHCP server).

The client ID of VLAN-interface 2 on Switch B is 0030-3030-662e-6532-3030-2e30-3030-322d-4574-6865-726e-6574.

The MAC address of VLAN-interface 2 on Switch C is 000f-e200-01c0.

Figure 10 Network diagram



Procedure

1. Specify an IP address for VLAN-interface 2 on Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 10.1.1.1 25
[SwitchA-Vlan-interface2] quit
```

2. Configure the DHCP server:

Create DHCP address pool 0.

```
[SwitchA] dhcp server ip-pool 0
```

Configure a static binding for Switch B.

```
[SwitchA-dhcp-pool-0] static-bind ip-address 10.1.1.5 25 client-identifier
0030-3030-662e-6532-3030-2e30-3030-322d-4574-6865-726e-6574
```

Configure a static binding for Switch C.

```
[SwitchA-dhcp-pool-0] static-bind ip-address 10.1.1.6 25 hardware-address
000f-e200-01c0
```

Specify the DNS server address and the gateway address.

```
[SwitchA-dhcp-pool-0] dns-list 10.1.1.2
[SwitchA-dhcp-pool-0] gateway-list 10.1.1.126
[SwitchA-dhcp-pool-0] quit
```

Enable DHCP.

```
[SwitchA] dhcp enable
```

Enable the DHCP server on VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] dhcp select server
[SwitchA-Vlan-interface2] quit
```

Verifying the configuration

Verify that Switch B can obtain IP address 10.1.1.5 and all other network parameters from Switch A. (Details not shown.)

Verify that Switch C can obtain IP address 10.1.1.6 and all other network parameters from Switch A. (Details not shown.)

On the DHCP server, display the IP addresses assigned to the clients.

```
[SwitchA] display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration      Type
                 Hardware address
10.1.1.5         0030-3030-662e-6532-  Jan 21 14:27:27 2014  Static(C)
                 3030-2e30-3030-322d-
                 4574-6865-726e-6574
10.1.1.6         000f-e200-01c0        Unlimited             Static(C)
```

Example: Configuring dynamic IP address assignment

Network configuration

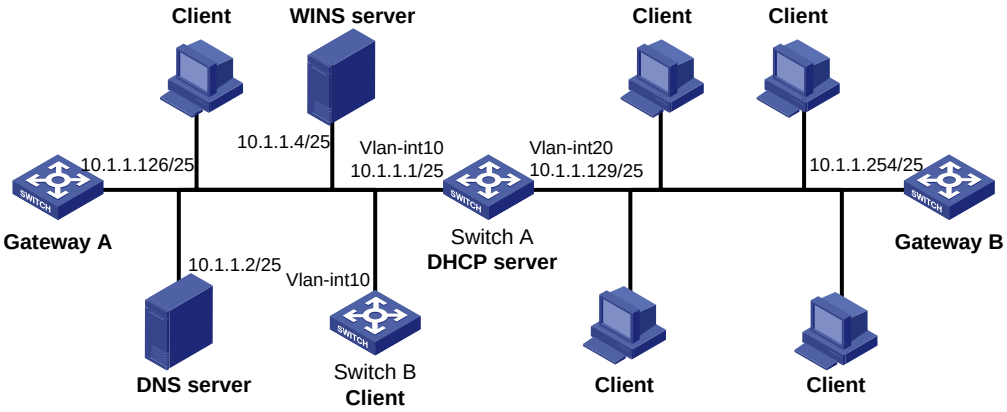
As shown in Figure 11, the DHCP server (Switch A) assigns IP addresses to clients on subnet 10.1.1.0/24, which is subnetted into 10.1.1.0/25 and 10.1.1.128/25.

Configure DHCP server on Switch A to implement the following assignment scheme.

Table 2 Assignment scheme

DHCP clients	IP address	Lease	Other configuration parameters
Clients connected to VLAN-interface 10	IP addresses on subnet 10.1.1.0/25	10 days and 12 hours	• Gateway: 10.1.1.126/25 • DNS server: 10.1.1.2/25 • Domain name: aabbcc.com • WINS server: 10.1.1.4/25
Clients connected to VLAN-interface 20	IP addresses on subnet 10.1.1.128/25	Five days	• Gateway: 10.1.1.254/25 • DNS server: 10.1.1.2/25 • Domain name: aabbcc.com

Figure 11 Network diagram



Procedure

1. Specify IP addresses for the VLAN interfaces. (Details not shown.)
2. Configure the DHCP server:

   # Exclude the DNS server address, WINS server address, and gateway addresses from dynamic allocation.

   ```
   <SwitchA> system-view
   [SwitchA] dhcp server forbidden-ip 10.1.1.2
   [SwitchA] dhcp server forbidden-ip 10.1.1.4
   [SwitchA] dhcp server forbidden-ip 10.1.1.126
   [SwitchA] dhcp server forbidden-ip 10.1.1.254
   ```

   # Configure DHCP address pool 1 to assign IP addresses and other configuration parameters to clients on subnet 10.1.1.0/25.

   ```
   [SwitchA] dhcp server ip-pool 1
   [SwitchA-dhcp-pool-1] network 10.1.1.0 mask 255.255.255.128
   [SwitchA-dhcp-pool-1] expired day 10 hour 12
   [SwitchA-dhcp-pool-1] domain-name aabbcc.com
   [SwitchA-dhcp-pool-1] dns-list 10.1.1.2
   [SwitchA-dhcp-pool-1] gateway-list 10.1.1.126
   [SwitchA-dhcp-pool-1] nbns-list 10.1.1.4
   [SwitchA-dhcp-pool-1] quit
   ```

   # Configure DHCP address pool 2 to assign IP addresses and other configuration parameters to clients on subnet 10.1.1.128/25.

   ```
   [SwitchA] dhcp server ip-pool 2
   [SwitchA-dhcp-pool-2] network 10.1.1.128 mask 255.255.255.128
   ```

```
[SwitchA-dhcp-pool-2] expired day 5
[SwitchA-dhcp-pool-2] domain-name aabbcc.com
[SwitchA-dhcp-pool-2] dns-list 10.1.1.2
[SwitchA-dhcp-pool-2] gateway-list 10.1.1.254
[SwitchA-dhcp-pool-2] quit
```

Enable DHCP.

```
[SwitchA] dhcp enable
```

Enable the DHCP server on VLAN-interface 10 and VLAN-interface 20.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] dhcp select server
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] dhcp select server
[SwitchA-Vlan-interface20] quit
```

Verifying the configuration

Verify that clients on subnets 10.1.1.0/25 and 10.1.1.128/25 can obtain correct IP addresses and all other network parameters from Switch A. (Details not shown.)

On the DHCP server, display the IP addresses assigned to the clients.

```
[SwitchA] display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration      Type
                 Hardware address
10.1.1.3         0031-3865-392e-6262-  Jan 14 22:25:03 2015  Auto(C)
                 3363-2e30-3230-352d-
                 4745-302f-30
10.1.1.5         0031-fe65-4203-7e02-  Jan 14 22:25:03 2015  Auto(C)
                 3063-5b30-3230-4702-
                 620e-712f-5e
10.1.1.130       3030-3030-2e30-3030-  Jan 9 10:45:11 2015   Auto(C)
                 662e-3030-3033-2d45-
                 7568-6572-1e
10.1.1.131       3030-0020-fe02-3020-  Jan 9 10:45:11 2015   Auto(C)
                 7052-0201-2013-1e02
                 0201-9068-23
10.1.1.132       2020-1220-1102-3021-  Jan 9 10:45:11 2015   Auto(C)
                 7e52-0211-2025-3402
                 0201-9068-9a
10.1.1.133       2021-d012-0202-4221-  Jan 9 10:45:11 2015   Auto(C)
                 8852-0203-2022-55e0
                 3921-0104-31
```

Example: Configuring DHCP user class

Network requirement

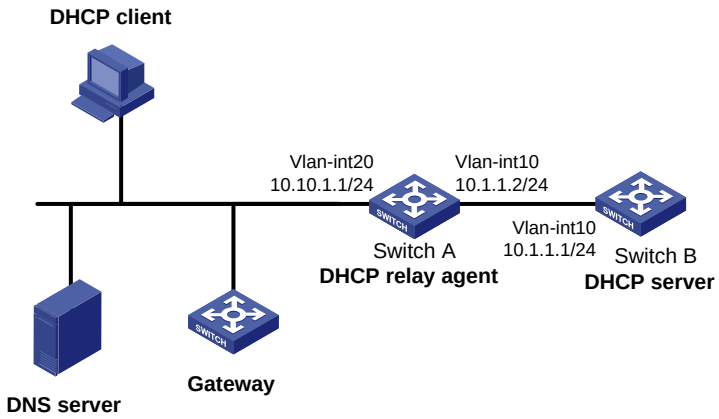
As shown in Figure 12, the DHCP relay agent (Switch A) forwards DHCP packets between DHCP clients and the DHCP server (Switch B). Enable switch A to support Option 82 so that switch A can add Option 82 in the DHCP requests sent by the DHCP clients.

Configure the address allocation scheme as follows:

Assign IP addresses	To clients
10.10.1.2 to 10.10.1.10	The DHCP request contains Option 82.
10.10.1.11 to 10.10.1.26	The hardware address in the request is six bytes long and begins with **aabb-aabb-aab**.

For clients on subnet 10.10.1.0/24, the DNS server address is 10.10.1.20/24 and the gateway address is 10.10.1.254/24.

Figure 12 Network diagram



Procedure

1. Specify IP addresses for interfaces on the DHCP server and the DHCP relay agent. (Details not shown.)
2. Configure DHCP services:

   # Create DHCP user class **tt** and configure a match rule to match client requests with Option 82.

   ```
   <SwitchB> system-view
   [SwitchB] dhcp class tt
   [SwitchB-dhcp-class-tt] if-match rule 1 option 82
   [SwitchB-dhcp-class-tt] quit
   ```

   # Create DHCP user class **ss** and configure a match rule to match DHCP requests in which the hardware address is six bytes long and begins with **aabb-aabb-aab**.

   ```
   [SwitchB] dhcp class ss
   [SwitchB-dhcp-class-ss] if-match rule 1 hardware-address aabb-aabb-aab0 mask
   ffff-ffff-fff0
   [SwitchB-dhcp-class-ss] quit
   ```

   # Create DHCP address pool **aa**.

   ```
   [SwitchB] dhcp server ip-pool aa
   ```

   # Specify the subnet for dynamic allocation.

   ```
   [SwitchB-dhcp-pool-aa] network 10.10.1.0 mask 255.255.255.0
   ```

   # Specify the address range for dynamic allocation.

   ```
   [SwitchB-dhcp-pool-aa] address range 10.10.1.2 10.10.1.100
   ```

   # Specify the address range for user class **tt**.

   ```
   [SwitchB-dhcp-pool-aa] class tt range 10.10.1.2 10.10.1.10
   ```

   # Specify the address range for user class **ss**.

   ```
   [SwitchB-dhcp-pool-aa] class ss range 10.10.1.11 10.10.1.26
   ```

Specify the gateway address and the DNS server address.

```
[SwitchB-dhcp-pool-aa] gateway-list 10.10.1.254
[SwitchB-dhcp-pool-aa] dns-list 10.10.1.20
[SwitchB-dhcp-pool-aa] quit
```

Enable DHCP and configure the DHCP server to handle Option 82.

```
[SwitchB] dhcp enable
[SwitchB] dhcp server relay information enable
```

Enable DHCP server on VLAN-interface 10.

```
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] dhcp select server
[SwitchB-Vlan-interface10] quit
```

Verifying the configuration

Verify that clients matching the user classes can obtain IP addresses in the specified ranges and all other configuration parameters from the DHCP server. (Details not shown.)

Display the IP address assigned by the DHCP server.

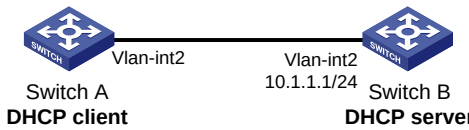
```
[SwitchB] display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration      Type
                 Hardware address
10.10.1.2        0031-3865-392e-6262-  Jan 14 22:25:03 2015  Auto(C)
                 3363-2e30-3230-352d-
                 4745-302f-30
10.10.1.11       aabb-aabb-aab1        Jan 14 22:25:03 2015  Auto(C)
```

Example: Configuring DHCP user class whitelist

Network configuration

As shown in Figure 13, configure the DHCP user class whitelist to allow the DHCP server to assign IP addresses to clients whose hardware addresses are six bytes long and begin with **aabb-aabb**.

Figure 13 Network diagram



Procedure

1. Specify IP addresses for the interfaces on the DHCP server. (Details not shown.)
2. Configure DHCP:

   # Create DHCP user class **ss** and configure a match rule to match DHCP requests in which the hardware address is six bytes long and begins with **aabb-aabb**.

   ```
   <SwitchB> system-view
   [SwitchB] dhcp class ss
   [SwitchB-dhcp-class-ss] if-match rule 1 hardware-address aabb-aabb-0000 mask
   ffff-ffff-0000
   [SwitchB-dhcp-class-ss] quit
   ```

   # Create DHCP address pool **aa**.

   ```
   [SwitchB] dhcp server ip-pool aa
   ```

   # Specify the subnet for dynamic allocation.

```
[SwitchB-dhcp-pool-aa] network 10.1.1.0 mask 255.255.255.0
```

Enable the DHCP user class whitelist.

```
[SwitchB-dhcp-pool-aa] verify class
```

Add DHCP user class **ss** to the DHCP user class whitelist.

```
[SwitchB-dhcp-pool-aa] valid class ss
[SwitchB-dhcp-pool-aa] quit
```

Enable DHCP.

```
[SwitchB] dhcp enable
```

Enable DHCP server on VLAN-interface 2.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] dhcp select server
[SwitchB-Vlan-interface2] quit
```

Verifying the configuration

Verify that clients matching the DHCP user class can obtain IP addresses on subnet 10.1.1.0/24 from the DHCP server. (Details not shown.)

On the DHCP server, display the IP addresses assigned to the clients.

```
[SwitchB] display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration      Type
                 Hardware address
10.1.1.2         aabb-aabb-ab01        Jan 14 22:25:03 2015  Auto(C)
```

Example: Configuring primary and secondary subnets

Network configuration

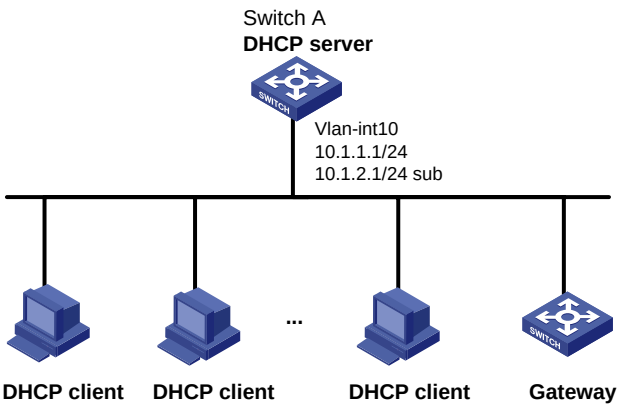
As shown in Figure 14, the DHCP server (Switch A) dynamically assigns IP addresses to clients in the LAN.

Configure two subnets in the address pool on the DHCP server: 10.1.1.0/24 as the primary subnet and 10.1.2.0/24 as the secondary subnet. The DHCP server selects IP addresses from the secondary subnet when the primary subnet has no assignable addresses.

Switch A assigns the following parameters:

- The default gateway 10.1.1.254/24 to clients on subnet 10.1.1.0/24.
- The default gateway 10.1.2.254/24 to clients on subnet 10.1.2.0/24.

Figure 14 Network diagram



Procedure

Create DHCP address pool **aa**.

```
<SwitchA> system-view
[SwitchA] dhcp server ip-pool aa
```

Specify the primary subnet and the gateway address for dynamic allocation.

```
[SwitchA-dhcp-pool-aa] network 10.1.1.0 mask 255.255.255.0
[SwitchA-dhcp-pool-aa] gateway-list 10.1.1.254
```

Specify the secondary subnet and the gateway address for dynamic allocation.

```
[SwitchA-dhcp-pool-aa] network 10.1.2.0 mask 255.255.255.0 secondary
[SwitchA-dhcp-pool-aa-secondary] gateway-list 10.1.2.254
[SwitchA-dhcp-pool-aa-secondary] quit
[SwitchA-dhcp-pool-aa] quit
```

Enable DHCP.

```
[SwitchA] dhcp enable
```

Configure the primary and secondary IP addresses of VLAN-interface 10.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ip address 10.1.1.1 24
[SwitchA-Vlan-interface10] ip address 10.1.2.1 24 sub
```

Enable the DHCP server on VLAN-interface 10.

```
[SwitchA-Vlan-interface10] dhcp select server
[SwitchA-Vlan-interface10] quit
```

Verifying the configuration

Verify that the DHCP server assigns clients IP addresses and gateway address from the secondary subnet when no address is available from the primary subnet. (Details not shown.)

Display the primary and secondary subnet IP addresses the DHCP server has assigned. The following is part of the command output.

```
[SwitchA] display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration      Type
                 Hardware address
10.1.1.2         0031-3865-392e-6262-  Jan 14 22:25:03 2015  Auto(C)
                 3363-2e30-3230-352d-
                 4745-302f-30
10.1.2.2         3030-3030-2e30-3030-  Jan 14 22:25:03 2015  Auto(C)
                 662e-3030-3033-2d45-
                 7568-6572-1e
```

Example: Customizing DHCP option

Network configuration

As shown in Figure 15, DHCP clients obtain IP addresses and PXE server addresses from the DHCP server (Switch A). The subnet for address allocation is 10.1.1.0/24.

Configure the address allocation scheme as follows:

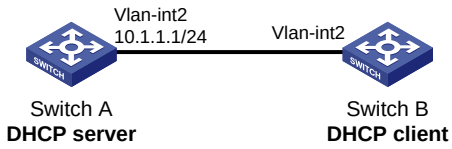
Assign PXE addresses	To clients
2.3.4.5 and 3.3.3.3	The hardware address in the request is six bytes long and begins with **aabb-aabb**.

Assign PXE addresses	To clients
1.2.3.4 and 2.2.2.2.	Other clients.

The DHCP server assigns PXE server addresses to DHCP clients through Option 43, a customized option. The format of Option 43 and that of the PXE server address sub-option are shown in Figure 5 and Figure 7. For example, the value of Option 43 configured in the DHCP address pool is 80 0B 00 00 02 01 02 03 04 02 02 02 02.

- The number 80 is the value of the sub-option type.
- The number 0B is the value of the sub-option length.
- The numbers 00 00 are the value of the PXE server type.
- The number 02 indicates the number of servers.
- The numbers 01 02 03 04 02 02 02 02 indicate that the PXE server addresses are 1.2.3.4 and 2.2.2.2.

Figure 15 Network diagram



Procedure

1. Specify IP addresses for the interfaces. (Details not shown.)
2. Configure the DHCP server:

   # Create DHCP user class **ss** and configure a match rule to match DHCP requests in which the hardware address is six bytes long and begins with **aabb-aabb**.

   ```
   <SwitchA> system-view
   [SwitchA] dhcp class ss
   [SwitchA-dhcp-class-ss] if-match rule 1 hardware-address aabb-aabb-0000 mask
   ffff-ffff-0000
   [SwitchA-dhcp-class-ss] quit
   ```

   # Create DHCP option group 1 and customize Option 43.

   ```
   [SwitchA] dhcp option-group 1
   [SwitchA-dhcp-option-group-1] option 43 hex 800B000002020304050303030303
   ```

   # Enable the DHCP server on VLAN-interface 2.

   ```
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] dhcp select server
   [SwitchA-Vlan-interface2] quit
   ```

   # Create DHCP address pool 0.

   ```
   [SwitchA] dhcp server ip-pool 0
   ```

   # Specify the subnet for dynamic address allocation.

   ```
   [SwitchA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
   ```

   # Customize Option 43.

   ```
   [SwitchA-dhcp-pool-0] option 43 hex 800B000002010203040202020202
   ```

   # Associate DHCP user class **ss** with option group 1.

   ```
   [SwitchA-dhcp-pool-0] class ss option-group 1
   [SwitchA-dhcp-pool-0] quit
   ```

   # Enable DHCP.

```
[SwitchA] dhcp enable
```

Verifying the configuration

Verify that Switch B can obtain an IP address on subnet 10.1.1.0/24 and the corresponding PXE server addresses from Switch A. (Details not shown.)

On the DHCP server, display the IP addresses assigned to the clients.

```
[SwitchA] display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration      Type
                 Hardware address
10.1.1.2         aabb-aabb-ab01        Jan 14 22:25:03 2015  Auto(C)
```

Troubleshooting DHCP server configuration

Failure to obtain a non-conflicting IP address

Symptom

A client's IP address obtained from the DHCP server conflicts with an IP address of another host.

Solution

Another host on the subnet might have the same IP address.

To resolve the problem:

1. Disable the client's network adapter or disconnect the client's network cable. Ping the IP address of the client from another host to check whether there is a host using the same IP address.
2. If a ping response is received, the IP address has been manually configured on a host. Execute the **dhcp server forbidden-ip** command on the DHCP server to exclude the IP address from dynamic allocation.
3. Enable the network adapter or connect the network cable, release the IP address, and obtain another one on the client. For example, to release the IP address and obtain another one on a Windows XP DHCP client:
   a. In Windows environment, execute the **cmd** command to enter the DOS environment.
   b. Enter **ipconfig /release** to relinquish the IP address.
   c. Enter **ipconfig /renew** to obtain another IP address.

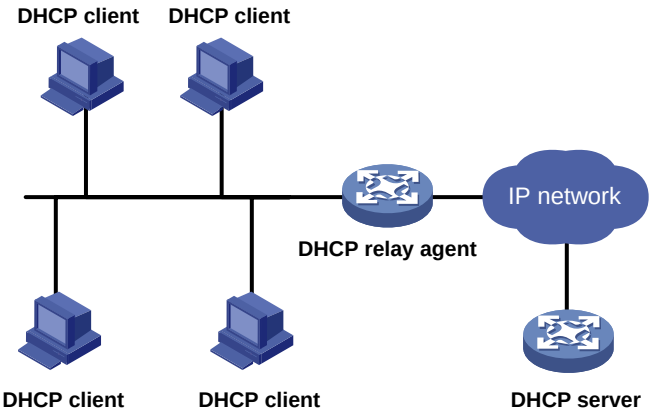
Configuring the DHCP relay agent

About DHCP relay agent

The DHCP relay agent enables clients to get IP addresses and configuration parameters from a DHCP server on another subnet.

Figure 16 shows a typical application of the DHCP relay agent.

Figure 16 DHCP relay agent application

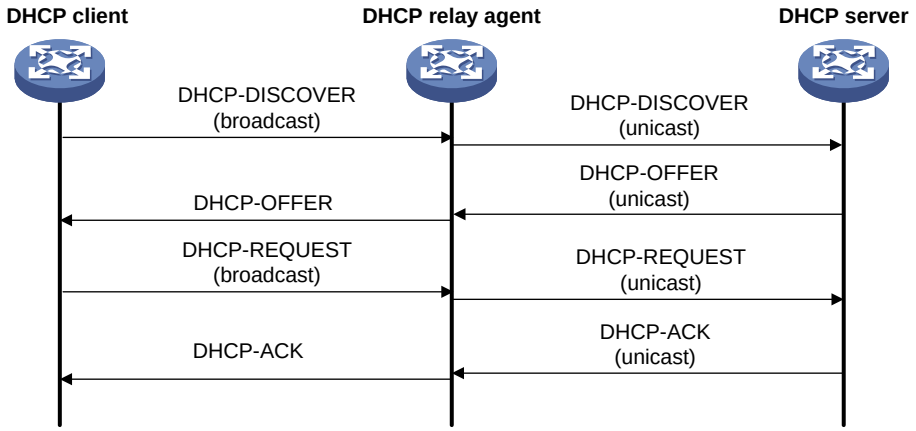


DHCP relay agent operation

The DHCP server and client interact with each other in the same way regardless of whether the relay agent exists. For the interaction details, see "IP address allocation process." The following only describes steps related to the DHCP relay agent:

1. After receiving a DHCP-DISCOVER or DHCP-REQUEST broadcast message from a DHCP client, the DHCP relay agent processes the message as follows:
   a. Fills the **giaddr** field of the message with its IP address.
   b. Unicasts the message to the designated DHCP server.
2. Based on the **giaddr** field, the DHCP server returns an IP address and other configuration parameters in a response.
3. The relay agent conveys the response to the client.

Figure 17 DHCP relay agent operation



DHCP relay agent support for Option 82

Option 82 records the location information about the DHCP client. It enables the administrator to perform the following tasks:

- Locate the DHCP client for security and accounting purposes.
- Assign IP addresses in a specific range to clients.

For more information about Option 82, see "Relay agent option (Option 82)."

If the DHCP relay agent supports Option 82, it handles DHCP requests by following the strategies described in Table 3.

If a response returned by the DHCP server contains Option 82, the DHCP relay agent removes the Option 82 before forwarding the response to the client.

Table 3 Handling strategies of the DHCP relay agent

If a DHCP request has…	Handling strategy	The DHCP relay agent…
Option 82	Drop	Drops the message.
	Keep	Forwards the message without changing Option 82.
	Replace	Forwards the message after replacing the original Option 82 with the Option 82 padded according to the configured padding format, padding content, and code type.
No Option 82	N/A	Forwards the message after adding Option 82 padded according to the configured padding format, padding content, and code type.

DHCP relay agent support for MCE

An MCE device acting as the DHCP relay agent can forward DHCP packets between a DHCP server and clients on either a public network or a private network. For more information about MCE, see *MPLS Configuration Guide*.

DHCP relay agent tasks at a glance

To configure a DHCP relay agent, perform the following tasks:

1. Enabling DHCP
2. Enabling the DHCP relay agent on an interface
3. Specifying DHCP servers
4. (Optional.) Configuring advanced features:
   - Specifying a DHCP relay address pool for DHCP clients
   - Configuring the DHCP relay agent security features
   - Configuring the DHCP relay agent to release an IP address
   - Disabling release notification
   - Configuring DHCP relay agent support for Option 82
   - Enabling Option 60 insertion into DHCP requests
   - Setting the DSCP value for DHCP packets sent by the DHCP relay agent
   - Specifying the DHCP relay agent address for the giaddr field
   - Specifying the source IP address for relayed DHCP requests
   - Configuring support for forwarding DHCP replies based on MAC address table

Enabling DHCP

Restrictions and guidelines

You must enable DHCP to make other DHCP relay agent settings take effect.

Procedure

1. Enter system view.
   `system-view`
2. Enable DHCP.
   `dhcp enable`
   By default, DHCP is disabled.

Enabling the DHCP relay agent on an interface

About this task

With the DHCP relay agent enabled, an interface forwards incoming DHCP requests to a DHCP server.

An IP address pool that contains the IP address of the DHCP relay interface must be configured on the DHCP server. Otherwise, the DHCP clients connected to the relay agent cannot obtain correct IP addresses.

Procedure

1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Enable the DHCP relay agent.
   `dhcp select relay`
   By default, when DHCP is enabled, an interface operates in the DHCP server mode.

Specifying DHCP servers

Specifying DHCP servers on a relay agent

About this task

To improve availability, you can specify several DHCP servers on the DHCP relay agent. When the interface receives request messages from clients, the relay agent forwards them to all DHCP servers.

Restrictions and guidelines

The IP address of any specified DHCP server must not reside on the same subnet as the IP address of the relay interface. Otherwise, the clients might fail to obtain IP addresses.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify a DHCP server address on the relay agent.

   dhcp relay server-address *ip-address* [**class** *class-name*] [**public** | **vpn-instance** *vpn-instance-name*]

   By default, no DHCP server address is specified on the relay agent.

   To specify multiple DHCP server addresses, repeat this step. You can specify a maximum of eight DHCP servers.

Specifying DHCP servers in a DHCP relay address pool

About this task

DHCP address pools created on a DHCP relay agent are called DHCP relay address pools. You can create a relay address pool and specify DHCP servers in this address pool. This feature allows DHCP clients of the same type to obtain IP addresses and other configuration parameters from the DHCP servers specified in the matching DHCP relay address pool.

It applies to scenarios where the DHCP relay agent connects to clients of the same access type but classified into different types by their locations. In this case, the relay interface typically has no IP address configured. You can use the **gateway-list** command to specify gateway addresses for clients matching the same DHCP relay address pool and bind the gateway addresses to the device's MAC address.

Upon receiving a DHCP DISCOVER or REQUEST from a client that matches a DHCP relay address pool, the relay agent processes the packet as follows:

- Fills the **giaddr** field of the packet with a specified gateway address.
- Forwards the packet to all DHCP servers in the matching DHCP relay address pool.

The DHCP servers select a DHCP relay address pool according to the gateway address.

Restrictions and guidelines

If PPPoE users are in the network, follow these restrictions and guidelines when you configure the DHCP relay address pool:

- Enable the DHCP relay agent to record DHCP relay entries by using the **dhcp relay client-information record** command. When a PPPoE user goes offline, the DHCP relay agent can find a matching relay entry and send a DHCP-RELEASE message to the DHCP

server. This mechanism ensures the DHCP server is aware of the releasing of the IP address in a timely manner.

- The **remote-server** command also configures the device as a DHCP relay agent. You do not need to enable the DHCP relay agent by using the **dhcp select relay** command.

Procedure

1. Enter system view.

   system-view

2. Create a DHCP relay address pool and enter its view.

   dhcp server ip-pool *pool-name*

3. Specify gateways in the DHCP relay address pool.

   gateway-list *ip-address*&<1-64> **export-route**

   By default, no gateway address is specified.

4. Specify DHCP servers in the DHCP relay address pool.

   remote-server *ip-address*&<1-8> [**public** | **vpn-instance** *vpn-instance-name*]

   By default, no DHCP server is specified in the DHCP relay address pool.

   You can specify a maximum of eight DHCP servers in one DHCP relay address pool for high availability.

Specifying the DHCP server selecting algorithm

About this task

The DHCP relay agent supports the **polling** and **master-backup** DHCP server selecting algorithms.

By default, the DHCP relay agent uses the **polling** algorithm. It forwards DHCP requests to all DHCP servers. The DHCP clients select the DHCP server from which the first received DHCP reply comes.

If the DHCP relay agent uses the **master-backup** algorithm, it forwards DHCP requests to the master DHCP server first. If the master DHCP server is not available, the relay agent forwards the subsequent DHCP requests to a backup DHCP server. If the backup DHCP server is not available, the relay agent selects the next backup DHCP server, and so on. If no backup DHCP server is available, it repeats the process starting from the master DHCP server.

The master DHCP server is determined in one of the following ways:

- In a common network where multiple DHCP server addresses are specified on the DHCP relay interface, the first specified DHCP server is the master. The other DHCP servers are backup.
- In a network where DHCP relay address pools are configured on the DHCP relay agent, the first specified DHCP server in a DHCP relay address pool is the master. The other DHCP servers in the DHCP relay address pool are backup.

DHCP server selection supports the following functions:

- **DHCP server response timeout time**—The DHCP relay agent determines that a DHCP server is not available if it does not receive any response from the server within the DHCP server response timeout time. The DHCP server response timeout time is configurable and the default is 30 seconds.
- **DHCP server switchback**—If the DHCP relay agent selects a backup DHCP server, it does not switch back to the master DHCP server by default. You can configure the DHCP relay agent to switch back to the master DHCP server after a delay. If the master DHCP server is available, the DHCP relay agent forwards DHCP requests to the master DHCP server. If the master DHCP server is not available, the DHCP relay agent still uses the backup DHCP server.

Specifying the DHCP server selecting algorithm in interface view

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the DHCP server selecting algorithm.
   dhcp relay server-address algorithm { **master-backup** | **polling** }
   By default, the **polling** algorithm is used. The DHCP relay agent forwards DHCP requests to all DHCP servers.
4. (Optional.) Set the DHCP server response timeout time for DHCP server switchover.
   dhcp relay dhcp-server timeout *time*
   By default, the DHCP server response timeout time is 30 seconds.
5. (Optional.) Enable the switchback to the master DHCP server and set the delay time.
   dhcp relay master-server switch-delay *delay-time*
   By default, the DHCP relay agent does not switch back to the master DHCP server.

Specifying the DHCP server selecting algorithm in DHCP relay address pool view

1. Enter system view.
   system-view
2. Enter DHCP relay address pool view.
   dhcp server ip-pool *pool-name*
3. Specify the DHCP server selecting algorithm.
   dhcp relay server-address algorithm { **master-backup** | **polling** }
   By default, the **polling** algorithm is used. The DHCP relay agent forwards DHCP requests to all DHCP servers.
4. (Optional.) Set the DHCP server response timeout time for DHCP server switchover.
   dhcp-server timeout *time*
   By default, the DHCP server response timeout time is 30 seconds.
5. (Optional.) Enable the switchback to the master DHCP server and set the delay time.
   master-server switch-delay *delay-time*
   By default, the DHCP relay agent does not switch back to the master DHCP server.

Specifying a DHCP relay address pool for DHCP clients

About this task

After you configure multiple DHCP relay address pools on a DHCP relay agent, you can specify these pools on an interface. To match DHCP clients based on options, you can define option settings when you specify the relay address pools.

If you specify multiple DHCP relay address pools on an interface, the relay agent selects a DHCP relay address pool for a DHCP client as follows:

1. Compares option values in the DHCP request in descending order against option values in DHCP relay address pools.
   - If a match (other than 60) is found, the matching process stops and the relay agent selects that matching relay address pool.

- If an Option 60 match is found, the relay agent continues to compare the Option 60 content in the request and the Option 60 strings in the relay address pools configured with the *option-text* argument:
  - If the Option 60 content matches the Option 60 string in a relay address pool, the relay address pool is selected.
  - If the Option 60 content does not match the Option 60 string in any relay address pool, the relay agent selects the relay address pool that is not configured with the *option-text* argument.

2. If still no DHCP relay address pool is matched, the relay agent selects the DHCP relay address pool with no options specified.

Restrictions and guidelines

If you specify DHCP servers by configuring both of the following methods on an interface, the DHCP relay address pool setting takes effect.

- Specify DHCP relay address pools by using the **dhcp relay pool** command.
- Specify DHCP servers directly on an interface by using the **dhcp relay server-address** command.

When you specify a DHCP relay address pool on an interface to define the DHCP servers, make sure the **remote-server** command is configured in the DHCP relay address pool. Otherwise, the relay agent drops DHCP requests. The DHCP requests are not forwarded to any DHCP server even if the **dhcp relay server-address** command is configured.

Procedure

1. Enter system view.

   system-view

2. Create a DHCP relay address pool and enter its view.

   dhcp server ip-pool *pool-name*

   By default, no DHCP relay address pools exist.

3. Specify DHCP servers in the DHCP relay address pool.

   remote-server *ip-address*&<1-8> [**public** | **vpn-instance** *vpn-instance-name*]

   By default, no DHCP server is specified in the DHCP relay address pool.

4. Specify gateway addresses for the clients matching the DHCP relay address pool.

   gateway-list *ip-address*&<1-64> **export-route**

   By default, no gateway address is specified.

5. Specify the DHCP server selecting algorithm.

   remote-server algorithm { **master-backup** | **polling** }

   By default, the **polling** algorithm is used. The DHCP relay agent forwards DHCP requests to all DHCP servers at the same time.

6. Return to system view.

   quit

7. Enter interface view.

   interface *interface-type interface-number*

8. Specify a DHCP relay address pool for DHCP clients.

   dhcp relay pool *pool-name* [**option** *code* [*option-text*]]

   By default, no DHCP relay address pool is specified for DHCP clients.

Configuring the DHCP relay agent security features

Restrictions and guidelines for DHCP relay agent security feature configuration

If the relay agent must record and refresh relay entries, enable the DHCP server proxy on the relay agent. When DHCP requests pass through the DHCP server proxy, it records or refreshes the client's IP-to-MAC bindings.

Enabling the DHCP relay agent to record relay entries

About this task

Perform this task to enable the DHCP relay agent to automatically record clients' IP-to-MAC bindings (relay entries) after they obtain IP addresses through DHCP.

Some security features use the relay entries to check incoming packets and block packets that do not match any entry. In this way, illegal hosts are not able to access external networks through the relay agent. Examples of the security features are ARP address check, authorized ARP, and IP source guard.

Restrictions and guidelines

The DHCP relay agent does not record IP-to-MAC bindings for DHCP clients running on synchronous/asynchronous serial interfaces.

Procedure

1. Enter system view.

   `system-view`

2. Enable the relay agent to record relay entries.

   `dhcp relay client-information record`

   By default, the relay agent does not record relay entries.

Enabling periodic refresh of dynamic relay entries

About this task

A DHCP client unicasts a DHCP-RELEASE message to the DHCP server to release its IP address. The DHCP relay agent conveys the message to the DHCP server and does not remove the IP-to-MAC entry of the client.

With this feature, the DHCP relay agent uses the IP address of a relay entry to periodically send a DHCP-REQUEST message to the DHCP server.

The relay agent maintains the relay entries depending on what it receives from the DHCP server:

- If the server returns a DHCP-ACK message or does not return any message within an interval, the DHCP relay agent removes the relay entry. In addition, upon receiving the DHCP-ACK message, the relay agent sends a DHCP-RELEASE message to release the IP address.
- If the server returns a DHCP-NAK message, the relay agent keeps the relay entry.

With this mechanism, the DHCP relay agent might mistakenly remove the relay entry for an online DHCP client when the DHCP server restarts. To resolve this issue, use the **`record-keep-time`** *`time`* option in the **`dhcp relay client-information refresh`** command to prolong the

keep time for DHCP relay entries. When the DHCP server does not reply to a DHCP-REQUEST message three times in a row, the DHCP relay agent does not remove the corresponding relay entry. Instead, it starts the keep time counter and keeps sending the DHCP-REQUEST message to the DHCP server at refresh intervals.

- If the DHCP server returns a DHCP-ACK message or does not return any message before the keep time elapses, the DHCP relay agent removes the relay entry.
- If the DHCP server returns a DHCP-NAK message before the keep time elapses, the DHCP relay agent keeps the relay entry and resets the keep time counter.

  The DHCP relay agent will start the keep time counter again if the DHCP server does not reply to another DHCP-REQUEST message three times in a row.

Procedure

1. Enter system view.

   system-view
2. Enable periodic refresh of dynamic relay entries.

   dhcp relay client-information refresh enable

   By default, periodic refresh of dynamic relay entries is enabled.
3. (Optional.) Set the refresh interval.

   dhcp relay client-information refresh { **auto** | **interval** *interval* } [**record-keep-time** *time*]

   By default, the refresh interval is **auto**, which is calculated based on the number of total relay entries. The keep time for DHCP relay entries is 0 minutes.

Configuring DHCP flood attack protection

About this task

The DHCP flood attack protection enables the DHCP relay agent to detect DHCP flood attacks according to the DHCP packet rate threshold on a per-MAC basis.

When the DHCP relay agent receives a DHCP packet from a client (MAC address), it creates a DHCP flood attack entry in check state. If the number of DHCP packets from the same MAC address reaches the upper limit in the detection duration, the relay agent determines that the client is launching a DHCP flood attack. The DHCP flood attack entry changes to the restrain state, and the DHCP relay agent discards the DHCP packets from that client. When the aging time of the entry is reached, the DHCP relay agent deletes the entry. If a DHCP packet from the MAC address arrives later, the DHCP relay agent will create a flood attack entry and count the number of incoming DHCP packets for that client again.

If you enable this feature in VXLAN VSI view, the DHCP relay agent will perform the following operations:

- Collect statistics about DHCP packets with the same source MAC address received on each VSI interface.
- Create a DHCP flood attack entry in check state for each source MAC address.

Configuring DHCP flood attack protection in a VXLAN network

1. Enter system view.

   system-view
2. (Optional) Set the DHCP packet rate threshold for DHCP flood attack detection.

   dhcp flood-protection threshold *packet-number milliseconds*

   By default, the device allows a maximum of 6 DHCP packets per 5000 milliseconds from each DHCP client.
3. (Optional) Set the DHCP flood attack entry aging time.

dhcp flood-protection aging-time *time*

The default setting is 300 seconds.

4. Enter VSI view.

   vsi *vsi-name*

5. Enable DHCP flood attack protection.

   dhcp flood-protection enable

   By default, DHCP flood attack protection is disabled.

Enabling DHCP starvation attack protection

About this task

A DHCP starvation attack occurs when an attacker constantly sends forged DHCP requests using different MAC addresses in the **chaddr** field to a DHCP server. This exhausts the IP address resources of the DHCP server so legitimate DHCP clients cannot obtain IP addresses. The DHCP server might also fail to work because of exhaustion of system resources. The following methods are available to relieve or prevent such attacks.

- To relieve a DHCP starvation attack that uses DHCP packets encapsulated with different source MAC addresses, you can use one of the following methods:
  - Limit the number of ARP entries that a Layer 3 interface can learn.
  - Set the MAC learning limit for a Layer 2 port, and disable unknown frame forwarding when the MAC learning limit is reached.
- To prevent a DHCP starvation attack that uses DHCP requests encapsulated with the same source MAC address, you can enable MAC address check on the DHCP relay agent. The DHCP relay agent compares the **chaddr** field of a received DHCP request with the source MAC address in the frame header. If they are the same, the DHCP relay agent forwards the request to the DHCP server. If not, the relay agent discards the request.

Enable MAC address check only on the DHCP relay agent directly connected to the DHCP clients. A DHCP relay agent changes the source MAC address of DHCP packets before sending them.

A MAC address check entry has an aging time. When the aging time expires, both of the following occur:

- The entry ages out.
- The DHCP relay agent rechecks the validity of DHCP requests sent from the MAC address in the entry.

Procedure

1. Enter system view.

   system-view

2. Set the aging time for MAC address check entries.

   dhcp relay check mac-address aging-time *time*

   The default aging time is 30 seconds.

   This command takes effect only after you execute the **dhcp relay check mac-address** command.

3. Enter the interface view.

   interface *interface-type interface-number*

4. Enable MAC address check.

   dhcp relay check mac-address

   By default, MAC address check is disabled.

Enabling DHCP server proxy on the DHCP relay agent

About this task

The DHCP server proxy feature isolates DHCP servers from DHCP clients and protects DHCP servers against attacks.

Upon receiving a response from the server, the DHCP server proxy modifies the server's IP address as the relay interface's IP address before sending out the response. The DHCP client takes the DHCP relay agent as the DHCP server.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable DHCP relay agent and DHCP server proxy on the interface.
   dhcp select relay proxy
   By default, the interface operates in DHCP server mode after DHCP is enabled.

Enabling client offline detection on the DHCP relay agent

About this task

The client offline detection on the DHCP relay agent detects the user online status based on the ARP entry aging. When an ARP entry ages out, the DHCP client offline detection feature deletes the relay entry for the IP address and sends a RELEASE message to the DHCP server.

If DHCP relay agent and DHCP snooping are configured on the same device, the DHCP snooping module deletes its DHCP snooping entries after it obtains the RELEASE messages from the relay agent module.

Restrictions and guidelines

The feature does not function if an ARP entry is manually deleted.

Procedure

1. Enter system view.
   system-view
2. Enable the relay agent to record relay entries.
   dhcp relay client-information record
   By default, the relay agent does not record relay entries.
   Without relay entries, client offline detection cannot function correctly.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable the DHCP relay agent.
   dhcp select relay
   By default, when DHCP is enabled, an interface operates in the DHCP server mode.
5. Enable client offline detection.
   dhcp client-detect
   By default, client offline detection is disabled on the DHCP relay agent.

Configuring the DHCP relay agent to release an IP address

About this task

Configure the relay agent to release the IP address for a relay entry. The relay agent sends a DHCP-RELEASE message to the server and meanwhile deletes the relay entry. Upon receiving the DHCP-RELEASE message, the DHCP server releases the IP address.

This command can release only the IP addresses in the recorded relay entries.

Procedure

1. Enter system view.
   system-view
2. Configure the DHCP relay agent to release an IP address.
   dhcp relay release ip *ip-address* [**vpn-instance** *vpn-instance-name*]

Disabling release notification

About this task

Users might roam among different gateways on a wireless network. When a user roams to a new gateway, the ARP entry of the user on the old gateway will age out. With DHCP client offline detection enabled, the DHCP relay agent deletes the relay entry of the user after the user's ARP entry expires. It also sends a DHCP-RELEASE message to the server to release the user address. When the DHCP server assigns that address to another user, an address conflict occurs. To avoid this situation, execute the **undo dhcp relay release-agent** command to disable the relay agent from sending release messages to the DHCP server when it deletes relay entries. This feature also applies when the DHCP relay entries are manually deleted by using the **reset** command.

Restrictions and guidelines

Disabling release notification does not affect relaying DHCP-RELEASE messages that are sent from DHCP clients. The DHCP relay agent forwards these messages to the DHCP server after receiving them from DHCP clients.

Procedure

1. Enter system view.
   system-view
2. Disable release notification.
   undo dhcp relay release-agent
   By default, release notification is enabled.

Configuring DHCP relay agent support for Option 82

For Option 82 to function correctly, you must configure both the DHCP server and relay agent. For server-side Option 82 configuration, see "Enabling handling of Option 82."

When you use the **dhcp relay information link-selection** command, follow these guidelines:

- For this command to take effect, perform the following tasks first:

- Enable the DHCP relay agent to support Option 82 by using the **dhcp relay information enable** command.
- Specify a source IP address for relayed DHCP requests by using the **dhcp relay source-address** command.

- To ensure successful IP address allocation, the IP address specified in sub-option 5 must belong to the same network segment as one of the following:
  - Gateway address in the AAA-authorized address pool or the address pool bound to the relay interface.
  - IP address of the DHCP relay interface.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the relay agent to handle Option 82.

   dhcp relay information enable

   By default, handling of Option 82 is disabled.

4. (Optional.) Configure the strategy for handling DHCP requests that contain Option 82.

   dhcp relay information strategy { **drop** | **keep** | **replace** }

   By default, the handling strategy is **replace**.

   If the handling strategy is **replace**, configure a padding mode and a padding format for Option 82. If the handling strategy is **keep** or **drop**, you do not need to configure a padding mode or padding format for Option 82.

5. (Optional.) Configure the padding mode and padding format for the Circuit ID sub-option.

   dhcp relay information circuit-id { **bas** | **string** *circuit-id* | { **normal** | **verbose** [**node-identifier** { **mac** | **sysname** | **user-defined** *node-identifier* }] [**interface**] } [**format** { **ascii** | **hex** }] }

   By default, the padding mode for Circuit ID sub-option is **normal**, and the padding format is **hex**.

   The device name (**sysname**) must not include spaces if it is configured as the padding content for sub-option 1. Otherwise, the DHCP relay agent will fail to add or replace Option 82.

6. (Optional.) Configure the padding mode and padding format for the Remote ID sub-option.

   dhcp relay information remote-id { **normal** [**format** { **ascii** | **hex** }] | **string** *remote-id* | **sysname** }

   By default, the padding mode for the Remote ID sub-option is **normal**, and the padding format is **hex**.

7. (Optional.) Configure the IP address to be inserted in sub-option 5 of Option 82.

   dhcp relay information link-selection *link-selection-address*

   By default, the relay agent inserts either of the following addresses in sub-option 5 of Option 82:
   - Gateway address in the AAA-authorized address pool or the address pool bound to the relay interface.
   - IP address of the relay interface.

Enabling Option 60 insertion into DHCP requests

About this task

Option 60 records vendor class identifier information of DHCP clients. It allows the clients to obtain IP addresses from different address ranges. After receiving a DHCP request with Option 60 encapsulated, the DHCP server follows the procedure to assign an IP address:

1. Uses Option 60 to determine a user class for the client.
2. Selects an IP address from the address range that matches the user class and assigns the address to the client.

After you enable Option 60 insertion on the DHCP relay agent, the relay agent first examines whether the received DHCP request contains Option 60.

- If the request does not contain Option 60, the relay agent inserts the option string into the request before forwarding the request to the DHCP server.
- If the request contains Option 60, the relay agent forwards the request to the DHCP server without processing this option.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable Option 60 insertion into DHCP requests on the DHCP relay agent.

   dhcp relay insert option60 *option-text*

   By default, the DHCP relay agent does not insert Option 60 into DHCP requests.

Setting the DSCP value for DHCP packets sent by the DHCP relay agent

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet.

Procedure

1. Enter system view.

   system-view
2. Set the DSCP value for DHCP packets sent by the DHCP relay agent.

   dhcp dscp *dscp-value*

   By default, the DSCP value in DHCP packets sent by the DHCP relay agent is 56.

Specifying the DHCP relay agent address for the **giaddr** field

Manually specifying the DHCP relay agent address for the **giaddr** field

About this task

This task allows you to specify the IP addresses to be encapsulated to the **giaddr** field of the DHCP requests. If you do not specify any DHCP relay agent address, the primary IP address of the DHCP relay interface is encapsulated to the **giaddr** field of DHCP requests.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify the DHCP relay agent address to be encapsulated in relayed DHCP requests.

   dhcp relay gateway *ip-address*

   By default, the primary IP address of the DHCP relay interface is encapsulated in the relayed DHCP requests.

Configuring smart relay to specify the DHCP relay agent address for the **giaddr** field

About this task

By default, the relay agent only encapsulates the primary IP address to the **giaddr** field of all requests before relaying them to the DHCP server. The DHCP server then selects an IP address on the same subnet as the address in the **giaddr** filed. If no assignable addresses on the subnet are available, the DHCP server does not assign any IP address. The DHCP smart relay feature is introduced to allow the DHCP relay agent to encapsulate secondary IP addresses when the DHCP server does not send back a DHCP-OFFER message.

The relay agent initially encapsulates its primary IP address to the **giaddr** field before forwarding a request to the DHCP server. If no DHCP-OFFER is received, the relay agent allows the client to send a maximum of two requests to the DHCP server by using the primary IP address. If no DHCP-OFFER is returned after two retries, the relay agent switches to a secondary IP address. If the DHCP server still does not respond, the next secondary IP address is used. After the secondary IP addresses are all tried and the DHCP server does not respond, the relay agent repeats the process by starting from the primary IP address.

Procedure

1. Enter system view.

   system-view

2. Enable the DHCP smart relay feature.

   dhcp smart-relay enable

   By default, the DHCP smart relay feature is disabled.

Specifying the source IP address for relayed DHCP requests

About this task

This task is required if multiple relay interfaces share the same IP address or if a relay interface does not have routes to DHCP servers. You can specify an IP address or the IP address of another interface, typically the loopback interface, on the DHCP relay agent as the source IP address for DHCP requests. The relay interface inserts the source IP address in the source IP address field as well as the **giaddr** field in DHCP requests.

If multiple relay interfaces share the same IP address, you must also configure the relay interface to support Option 82. Upon receiving a DHCP request, the relay interface inserts the subnet information in sub-option 5 in Option 82. The DHCP server assigns an IP address according to sub-option 5. The DHCP relay agent looks up the output interface in the MAC address table to forward the DHCP reply.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify the source IP address for DHCP requests.

   dhcp relay source-address { *ip-address* | **interface** *interface-type interface-number* } [**default-giaddr**]

   By default, the relay agent chooses the default source IP address for relayed requests depending on whether its server-side interface and the DHCP server belong to the same VPN instance:

   - If they belong to the same VPN instance, the relay agent uses the IP address of the output interface for relayed requests as the default source IP address.
   - If they belong to different VPN instances, the relay agent uses the lowest IP address that is in the same VPN instance as the DHCP server as the default source address.

Discarding DHCP requests that are delivered from VXLAN tunnels

About this task

In a VXLAN network, the DHCP relay agent feature can be configured on the VSI or VLAN interface of a VTEP.

When the DHCP relay agent receives a DHCP request on the VLAN interface or from an AC mapped to the VSI interface, the relay agent forwards this request to the DHCP servers and broadcasts this request to other VTEPs. If those VTEPs also function as the DHCP relay agents, each will forward the DHCP request to the DHCP servers they are connecting to. To prevent a DHCP server from receiving the same DHCP request from different VTEPs, you can configure this command on the VSI or VLAN interface of the VTEPs that are not directly connecting to DHCP clients.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

interface *interface-type interface-number*

3. Configure the DHCP relay agent to discard the DHCP requests that are delivered from VXLAN tunnels.

   dhcp relay request-from-tunnel discard

   By default, the DHCP relay agent can forward the DHCP requests that are delivered from VXLAN tunnels.

   If you configure this command on a device that acts as both the distributed and centralized VXLAN IP gateways, make sure the gateways do not use the same VSI or VLAN interface to provide the gateway services.

Configuring support for forwarding DHCP replies based on MAC address table

About this task

When a DHCP client sends a request to its connected EVPN gateway, the gateway records the request forwarding information for the client before relaying the request to the DHCP server. The request forwarding information contains the MAC address of the DHCP client and output interface of the request packet. If the reply for this request is received by another EVPN gateway, the reply is discarded by default because it does not have the matching request forwarding information on the gateway. As a result, the client cannot obtain the IP address.

This feature allows the relay agent to look up the MAC address table for the output interface for the replies that do not have the request forwarding information. This feature ensures that DHCP clients receive DHCP reply packets.

If the **broadcast** keyword is specified, the DHCP relay agent broadcasts the DHCP reply out of the interface in the matching MAC address entry. When the DHCP relay agent connecting to the requesting DHCP client receives the reply, it sends the reply to its CPU before broadcasting the reply. If recording relay entries is enabled on this relay agent, a relay entry is generated for the DHCP client.

If the **broadcast** keyword is not specified, the DHCP relay agent connecting to the requesting client will directly forward the reply based on the MAC address instead of delivering it up to the CPU. In this case, the DHCP relay agent cannot be aware that the packet is a DHCP reply, thus no relay entry is generated for the client even if the recording of relay entries is enabled.

Procedure

1. Enter system view.

   system-view

2. Enable MAC address table lookup for DHCP replies that do not have request forwarding information.

   dhcp relay mac-forward enable [**broadcast**]

   By default, the DHCP relay agent discards a DHCP reply if the relay agent does not have the request forwarding information for the reply.

Display and maintenance commands for DHCP relay agent

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display MAC address check entries on the DHCP relay agent.	**display dhcp relay check mac-address**
Display relay entries on the DHCP relay agent.	**display dhcp relay client-information** [**interface** *interface-type interface-number* | **ip** *ip-address* [**vpn-instance** *vpn-instance-name*]]
Display Option 82 configuration information on the DHCP relay agent.	**display dhcp relay information** [**interface** *interface-type interface-number*]
Display M-LAG status information recorded on the DHCP relay agent.	**display dhcp relay m-lag-status**
Display information about DHCP servers on an interface.	**display dhcp relay server-address** [**interface** *interface-type interface-number*]
Display packet statistics on the DHCP relay agent.	**display dhcp relay statistics** [**interface** *interface-type interface-number*]
Clear relay entries on the DHCP relay agent.	**reset dhcp relay client-information** [**interface** *interface-type interface-number* | **ip** *ip-address* [**vpn-instance** *vpn-instance-name*]]
Clear packet statistics on the DHCP relay agent.	**reset dhcp relay statistics** [**interface** *interface-type interface-number*]

DHCP relay agent configuration examples

Example: Configuring basic DHCP relay agent

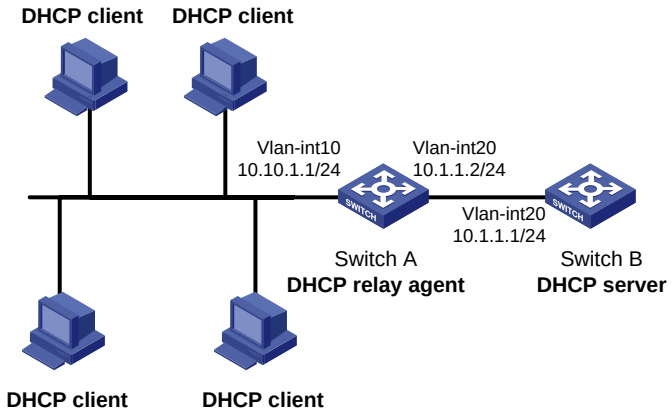
Network configuration

As shown in Figure 18, configure the DHCP relay agent on Switch A. The DHCP relay agent enables DHCP clients to obtain IP addresses and other configuration parameters from the DHCP server on another subnet.

The DHCP relay agent and server are on different subnets. Configure static or dynamic routing to make them reachable to each other.

Perform the configuration on the DHCP server to guarantee the client-server communication. For DHCP server configuration information, see "DHCP server configuration examples."

Figure 18 Network diagram



Procedure

Specify IP addresses for the interfaces. (Details not shown.)

Enable DHCP.

```
<SwitchA> system-view
[SwitchA] dhcp enable
```

Enable the DHCP relay agent on VLAN-interface 10.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] dhcp select relay
```

Specify the IP address of the DHCP server on the relay agent.

```
[SwitchA-Vlan-interface10] dhcp relay server-address 10.1.1.1
```

Verifying the configuration

Verify that DHCP clients can obtain IP addresses and all other network parameters from the DHCP server through the DHCP relay agent. (Details not shown.)

Display the statistics of DHCP packets forwarded by the DHCP relay agent.

```
[SwitchA] display dhcp relay statistics
```

Display relay entries if you have enabled relay entry recording on the DHCP relay agent.

```
[SwitchA] display dhcp relay client-information
```

Example: Configuring Option 82

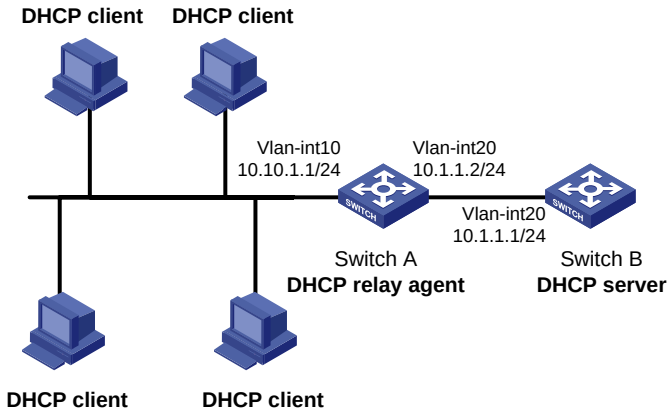
Network configuration

As shown by the network diagram in "Example: Configuring basic DHCP relay agent," the DHCP relay agent (Switch A) replaces Option 82 in DHCP requests before forwarding them to the DHCP server (Switch B).

- The Circuit ID sub-option is **company001.**
- The Remote ID sub-option is **device001**.

To use Option 82, you must also enable the DHCP server to handle Option 82.

Figure 19 Network diagram



Procedure

Specify IP addresses for the interfaces. (Details not shown.)

Enable DHCP.

```
<SwitchA> system-view
[SwitchA] dhcp enable
```

Enable the DHCP relay agent on VLAN-interface 10.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] dhcp select relay
```

Specify the IP address of the DHCP server.

```
[SwitchA-Vlan-interface10] dhcp relay server-address 10.1.1.1
```

Configure the handling strategies and padding content of Option 82.

```
[SwitchA-Vlan-interface10] dhcp relay information enable
[SwitchA-Vlan-interface10] dhcp relay information strategy replace
[SwitchA-Vlan-interface10] dhcp relay information circuit-id string company001
[SwitchA-Vlan-interface10] dhcp relay information remote-id string device001
```

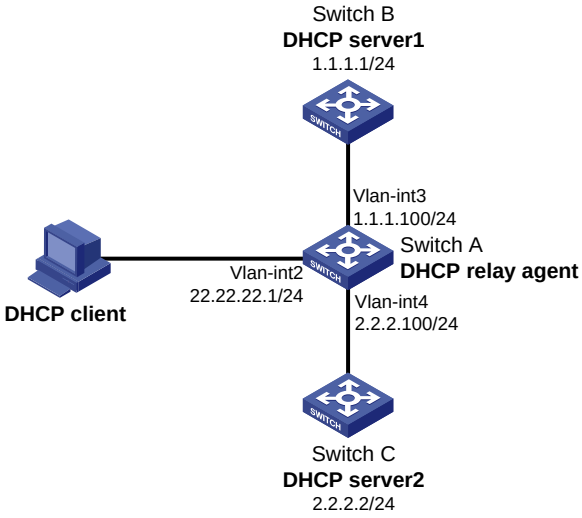
Example: Configuring DHCP server selection

Network configuration

As shown in Figure 20, the DHCP client and the DHCP servers are in different subnets. DHCP server 1 and DHCP server 2 both have a DHCP address pool that contains IP addresses in subnet 22.22.22.0/24, but neither has DHCP enabled.

Configure the DHCP relay agent for the DHCP client to obtain an IP address in subnet 22.22.22.0/24 and other configuration parameters from a DHCP server. The DHCP relay agent is connected to the DHCP client through VLAN-interface 2, to DHCP server 1 through VLAN-interface 3, and to DHCP server 2 through VLAN-interface 4.

Figure 20 Network diagram



Procedure

1. Assign IP addresses to interfaces on the switches. (Details not shown.)
2. Configure Switch B and Switch C as DHCP servers. (Details not shown.)
3. Configure the DHCP relay agent on Switch A:

   # Enable DHCP.

   ```
   <SwitchA> system-view
   [SwitchA] dhcp enable
   ```

   # Enable the DHCP relay agent on VLAN-interface 2.

   ```
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] dhcp select relay
   ```

   # Specify the IP addresses of the DHCP servers.

   ```
   [SwitchA-Vlan-interface2] dhcp relay server-address 1.1.1.1
   [SwitchA-Vlan-interface2] dhcp relay server-address 2.2.2.2
   ```

   # Specify the DHCP server selecting algorithm as **master-backup**.

   ```
   [SwitchA-Vlan-interface2] dhcp relay server-address algorithm master-backup
   ```

   # Configure the DHCP relay agent to switch back to the master DHCP server 3 minutes after it switches to the backup DHCP server.

   ```
   [SwitchA-Vlan-interface2] dhcp relay master-server switch-delay 3
   ```

Verifying the configuration

Verify that the DHCP client cannot obtain an IP address and that the following log is output in about 30 seconds.

```
DHCPR/3/DHCPR_SERVERCHANGE:
 Switched to the server at 2.2.2.2 because the current server did not respond.
```

Enable DHCP on the DHCP server at 1.1.1.1. (Details not shown.)

Verify that the DHCP client cannot obtain an IP address and that the following log is output in about 3 minutes.

```
DHCPR/3/DHCPR_SWITCHMASTER:
 Switched to the master DHCP server at 1.1.1.1.
```

Verify that the DHCP client obtains an IP address. (Details not shown.)

Troubleshooting DHCP relay agent configuration

Failure of DHCP clients to obtain configuration parameters through the DHCP relay agent

Symptom

DHCP clients cannot obtain configuration parameters through the DHCP relay agent.

Solution

Some problems might occur with the DHCP relay agent or server configuration.

To locate the problem, enable debugging and execute the **`display`** command on the DHCP relay agent to view the debugging information and interface state information.

Check that:

- DHCP is enabled on the DHCP server and relay agent.
- The DHCP server has an address pool on the same subnet as the DHCP clients.
- The DHCP server and DHCP relay agent can reach each other.
- The DHCP server address specified on the DHCP relay interface connected to the DHCP clients is correct.

Configuring the DHCP client

About DHCP client

You can configure an interface to act as a DHCP client to obtain configuration parameters such as an IP address from the DHCP server.

Support of DHCP Client for gateway availability detection

The DHCP client performs gateway availability detection after obtaining an IP address, ensuring that it can correctly obtain usable gateway information and access to external networks

The gateway availability detection feature is implemented through a ping operation. The DHCP client sends an ICMP echo request to the gateway address carried in the received DHCP reply, and then waits for an ICMP echo reply from the gateway:

- If the client receives an ICMP echo reply from the gateway within one minute, the gateway is determined reachable. The client can use the obtained IP address to access external networks.
- If the client does not receive an ICMP echo reply from the gateway within one minute, the detection attempt fails and the client pings the gateway again. When the number of consecutive failed detection attempts reaches the upper limit, the gateway availability detection fails and the gateway is determined unreachable. The client will then release the obtained IP address and request a new IP address. The default upper limit is 4, which varies by device.

In the current software version, only VLAN-interface 1 supports the gateway availability detection feature.

Restrictions and guidelines: DHCP client configuration

The DHCP client configuration is supported only on Layer 3 Ethernet interfaces (or subinterfaces), VLAN interfaces, and Layer 3 aggregate interfaces.

DHCP client tasks at a glance

To configure a DHCP client, perform the following tasks:

1. Enabling the DHCP client on an interface
2. Configuring a DHCP client ID for an interface

   Perform this task if the DHCP client uses the client ID to obtain IP addresses.
3. (Optional.) Enabling duplicated address detection
4. (Optional.) Setting the DSCP value for DHCP packets sent by the DHCP client
5. (Optional.) Configuring SNMP notifications for the DHCP client

Enabling the DHCP client on an interface

Restrictions and guidelines

- If the number of IP address request failures reaches the system-defined amount, the DHCP client-enabled interface uses a default IP address.

- An interface can be configured to acquire an IP address in multiple ways. The new configuration overwrites the old.
- Secondary IP addresses cannot be configured on an interface that is enabled with the DHCP client.
- If the interface obtains an IP address on the same segment as another interface on the device, the interface does not use the assigned address. Instead, it requests a new IP address from the DHCP server.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure an interface to use DHCP for IP address acquisition.
   ip address dhcp-alloc
   By default, an interface does not use DHCP for IP address acquisition.

Configuring a DHCP client ID for an interface

About this task

A DHCP client ID is added to the DHCP option 61 to uniquely identify a DHCP client. A DHCP server can assign IP addresses to clients based on their DHCP client IDs.

DHCP client ID includes an ID type and a type value. Each ID type has a fixed type value. You can specify a DHCP client ID by using one of the following methods:

- Use an ASCII string as the client ID. If an ASCII string is used, the type value is 00.
- Use a hexadecimal number as the client ID. If a hexadecimal number is used, the type value is the first two characters in the number.
- Use the MAC address of an interface to generate a client ID. If this method is used, the type value is 01.

To identify the type value of a DHCP client ID, execute the **display dhcp server ip-in-use** or **display dhcp client** command.

Restrictions and guidelines

Make sure the ID for each DHCP client is unique.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure a DHCP client ID for the interface.
   dhcp client identifier { **ascii** *ascii-string* | **hex** *hex-string* | **mac** *interface-type interface-number* }
   By default, an interface generates the DHCP client ID based on its MAC address. If the interface does not have a MAC address, it uses the MAC address of the first Ethernet interface to generate its client ID.

Enabling duplicated address detection

About this task

DHCP client detects IP address conflict through ARP packets. An attacker can act as the IP address owner to send an ARP reply. The spoofing attack makes the client unable to use the IP address assigned by the server. As a best practice, disable duplicate address detection when ARP attacks exist on the network.

Procedure

1. Enter system view.
   system-view
2. Enable duplicate address detection.
   dhcp client dad enable
   By default, the duplicate address detection feature is enabled on an interface.

Setting the DSCP value for DHCP packets sent by the DHCP client

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet.

Procedure

1. Enter system view.
   system-view
2. Set the DSCP value for DHCP packets sent by the DHCP client.
   dhcp client dscp *dscp-value*
   By default, the DSCP value in DHCP packets sent by the DHCP client is 56.

Configuring SNMP notifications for the DHCP client

About this task

This task enables the DHCP client module to report critical events in SNMP notifications to network administrators.

- **IP application failure notifications**—An IP application failure notification is generated when the DHCP client module fails to apply for an IP address.
- **Local address conflict notifications**—A local address conflict notification is generated when the DHCP client module obtains an IP address that is in the same subnet as one of the local interface addresses.

The SNMP notifications are sent to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Enable SNMP notifications for the DHCP client module.

   snmp-agent trap enable dhcp-client [**ip-apply-failed** | **local-conflict**]*

   By default, all SNMP notifications are enabled for the DHCP client module.

Display and maintenance commands for DHCP client

Execute **display** command in any view.

Task	Command
Display DHCP client information.	**display dhcp client** [**verbose**] [**interface** *interface-type interface-number*]

DHCP client configuration examples

Example: Configuring DHCP client

Network configuration

As shown in Figure 22, on a LAN, Switch B contacts the DHCP server through VLAN-interface 2 to obtain an IP address, a DNS server address, and static route information. The DHCP client's IP address resides on subnet 10.1.1.0/24. The DNS server address is 20.1.1.1. The next hop of the static route to subnet 20.1.1.0/24 is 10.1.1.2.

The DHCP server uses Option 121 to assign static route information to DHCP clients. Figure 21 shows the Option 121 format. The destination descriptor field contains the following parts: subnet mask length and destination network address, both in hexadecimal notation. In this example, the destination descriptor is 18 14 01 01 (the subnet mask length is 24 and the network address is 20.1.1.0 in dotted decimal notation). The next hop address is 0A 01 01 02 (10.1.1.2 in dotted decimal notation).

Figure 21 Option 121 format

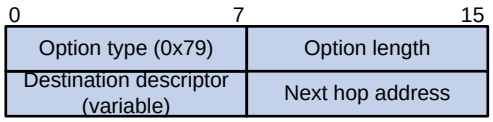
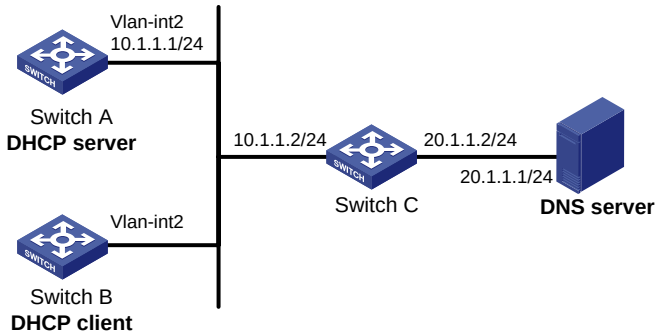


Figure 22 Network diagram



Procedure

1. Configure Switch A:

   # Specify an IP address for VLAN-interface 2.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ip address 10.1.1.1 24
   [SwitchA-Vlan-interface2] quit
   ```

   # Exclude an IP address from dynamic allocation.

   ```
   [SwitchA] dhcp server forbidden-ip 10.1.1.2
   ```

   # Configure DHCP address pool 0. Specify the subnet, lease duration, DNS server address, and a static route to subnet 20.1.1.0/24.

   ```
   [SwitchA] dhcp server ip-pool 0
   [SwitchA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
   [SwitchA-dhcp-pool-0] expired day 10
   [SwitchA-dhcp-pool-0] dns-list 20.1.1.1
   [SwitchA-dhcp-pool-0] option 121 hex 18 14 01 01 0A 01 01 02
   [SwitchA-dhcp-pool-0] quit
   ```

   # Enable DHCP.

   ```
   [SwitchA] dhcp enable
   ```

2. Configure Switch B:

   # Configure VLAN-interface 2 to use DHCP for IP address acquisition.

   ```
   <SwitchB> system-view
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] ip address dhcp-alloc
   [SwitchB-Vlan-interface2] quit
   ```

Verifying the configuration

Display the IP address and other network parameters assigned to Switch B.

```
[SwitchB-Vlan-interface2] display dhcp client verbose
Vlan-interface2 DHCP client information:
 Current state: BOUND
 Allocated IP: 10.1.1.3 255.255.255.0
 Allocated lease: 864000 seconds, T1: 331858 seconds, T2: 756000 seconds
 Lease from May 21 19:00:29 2012   to   May 31 19:00:29 2012
 DHCP server: 10.1.1.1
 Transaction ID: 0xcde72232
```

```
 Classless static routes:
   Destination: 20.1.1.0, Mask: 255.255.255.0, NextHop: 10.1.1.2
 DNS servers: 20.1.1.1
 Client ID type: acsii(type value=00)
 Client ID value: 000c.29d3.8659-Vlan2
 Client ID (with type) hex: 0030-3030-632e-3239-
                            6433-2e38-3635-392d-
                            4574-6830-2f30-2f32
 T1 will timeout in 3 days 19 hours 48 minutes 43 seconds
```

Display the route information on Switch B. The output shows that a static route to subnet 20.1.1.0/24 is added to the routing table.

```
[SwitchB] display ip routing-table
Destinations : 11        Routes : 11
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.1.1.0/24         Direct 0    0            10.1.1.3        Vlan2
10.1.1.3/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Static 70   0            10.1.1.2        Vlan2
10.1.1.255/32       Direct 0    0            10.1.1.3        Vlan2
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

Configuring DHCP snooping

About DHCP snooping

DHCP snooping is a security feature for DHCP.

DHCP snooping works between the DHCP client and server, or between the DHCP client and DHCP relay agent. It guarantees that DHCP clients obtain IP addresses from authorized DHCP servers. Also, it records IP-to-MAC bindings of DHCP clients (called DHCP snooping entries) for security purposes.

DHCP snooping defines trusted and untrusted ports to make sure clients obtain IP addresses only from authorized DHCP servers.

- **Trusted**—A trusted port can forward DHCP messages correctly to make sure the clients get IP addresses from authorized DHCP servers.
- **Untrusted**—An untrusted port discards received DHCP-ACK and DHCP-OFFER messages to prevent unauthorized servers from assigning IP addresses.

DHCP snooping reads DHCP-ACK messages received from trusted ports and DHCP-REQUEST messages to create DHCP snooping entries. A DHCP snooping entry includes the MAC and IP addresses of a client, the port that connects to the DHCP client, and the VLAN.

The following features need to use DHCP snooping entries:

- **ARP fast-reply**—Uses DHCP snooping entries to reduce ARP broadcast traffic. For more information, see "Configuring ARP fast-reply."
- **ARP attack detection**—Uses DHCP snooping entries to filter ARP packets from unauthorized clients. For more information, see *Security Configuration Guide*.
- **MAC-forced forwarding (MFF)**—Auto-mode MFF performs the following tasks:
  - Intercepts ARP requests from clients.
  - Uses DHCP snooping entries to find the gateway address.
  - Returns the gateway MAC address to the clients.

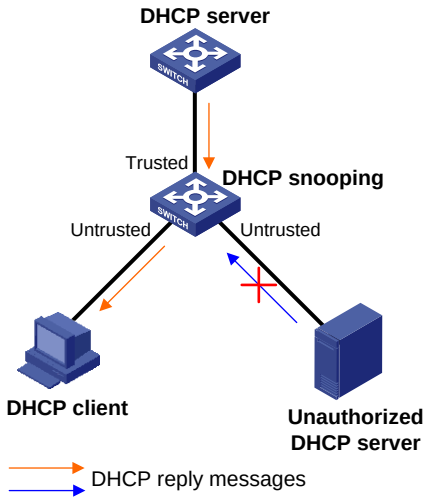
  This feature forces the client to send all traffic to the gateway so that the gateway can monitor client traffic to prevent malicious attacks among clients. For more information, see *Security Configuration Guide*.
- **IP source guard**—Uses DHCP snooping entries to filter illegal packets on a per-port basis. For more information, see *Security Configuration Guide*.
- **VLAN mapping**—Uses DHCP snooping entries to replace service provider VLAN in packets with customer VLAN before sending the packets to clients. For more information, see *Layer 2—LAN Switching Configuration Guide*.

Application of trusted and untrusted ports

Configure ports facing the DHCP server as trusted ports, and configure other ports as untrusted ports.

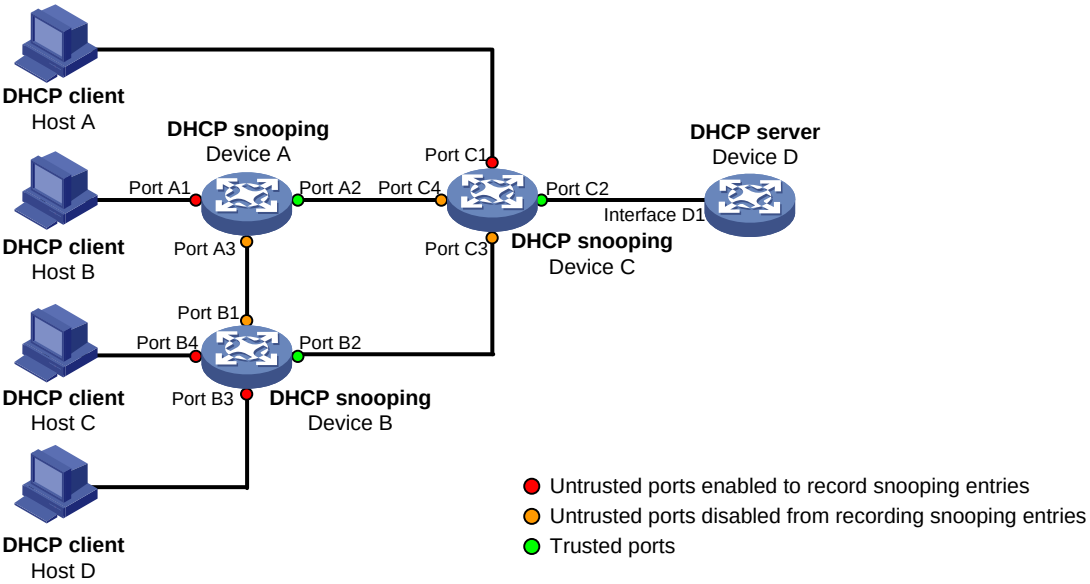
As shown in Figure 23, configure the DHCP snooping device's port that is connected to the DHCP server as a trusted port. The trusted port forwards response messages from the DHCP server to the client. The untrusted port connected to the unauthorized DHCP server discards incoming DHCP response messages.

Figure 23 Trusted and untrusted ports



In a cascaded network as shown in Figure 24, configure the DHCP snooping devices' ports facing the DHCP server as trusted ports. To save system resources, you can enable only the untrusted ports directly connected to the DHCP clients to record DHCP snooping entries.

Figure 24 Trusted and untrusted ports in a cascaded network



DHCP snooping support for Option 82

Option 82 records the location information about the DHCP client so the administrator can locate the DHCP client for security and accounting purposes. For more information about Option 82, see "Relay agent option (Option 82)."

Sub-option 9 (Vendor-Specific) in Option 82 is supported only on DHCP snooping devices. Each DHCP snooping device with the **append** Option 82 handling strategy adds the following information to the sub-option in the received DHCP request:

- Node identifier of the current DHCP snooping device.
- Information about the client-side interface.

- VLAN of the DHCP client.

After the management device receives the DHCP request, it can determine the network topology that the request has travelled and locate the DHCP client.

DHCP snooping uses the same strategies as the DHCP relay agent to handle Option 82 for DHCP request messages, as shown in Table 4. If a response returned by the DHCP server contains Option 82, DHCP snooping removes Option 82 before forwarding the response to the client. If the response contains no Option 82, DHCP snooping forwards it directly.

Table 4 Handling strategies

If a DHCP request has…	Handling strategy	DHCP snooping…
Option 82	Append	• Forwards the message after padding the Vendor-Specific sub-option with the content specified in the **dhcp snooping information vendor-specific** command. • Forwards the message without changing Option 82 if the **dhcp snooping information vendor-specific** command is not configured.
	Drop	Drops the message.
	Keep	Forwards the message without changing Option 82.
	Replace	Forwards the message after replacing the original Option 82 with the Option 82 padded according to the configured padding format, padding content, and code type.
No Option 82	N/A	Forwards the message after adding the Option 82 padded according to the configured padding format, padding content, and code type.

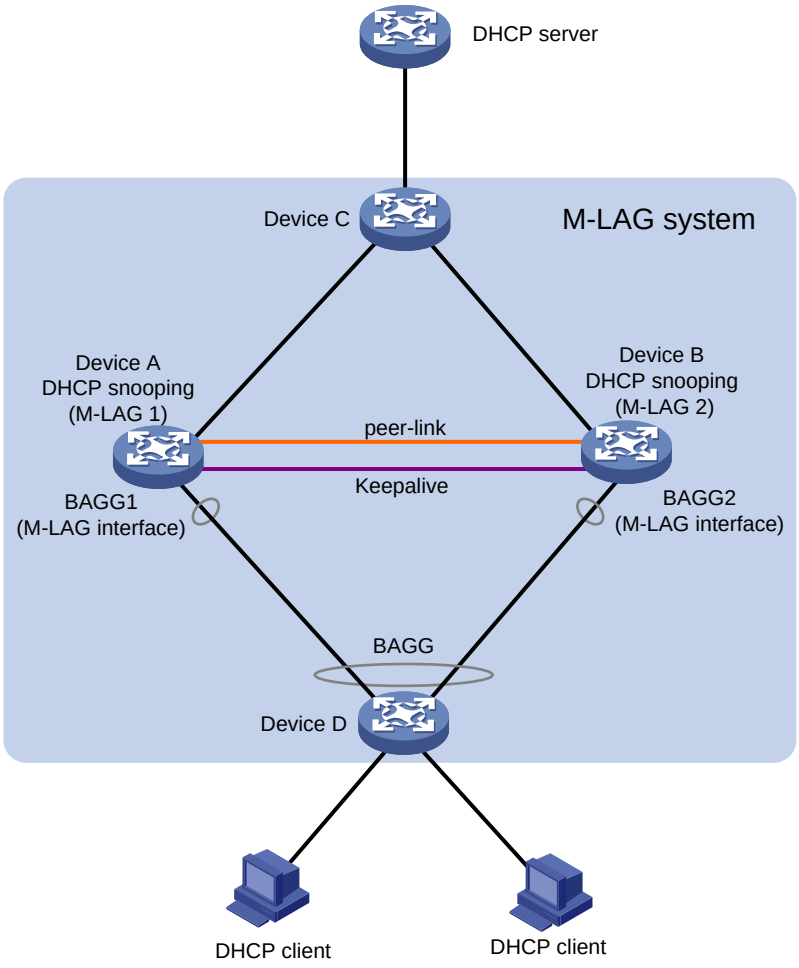
DHCP snooping on an M-LAG network

M-LAG can virtualize two DHCP snooping devices into one system through multichassis link aggregation, improving the availability of DHCP snooping devices and facilitating device-level redundancy and load sharing.

To view M-LAG status information and packet statistics related to DHCP snooping entry synchronization, execute the **display dhcp snooping m-lag-status** and **display dhcp snooping m-lag-statistics** command, respectively. You can determine whether DHCP snooping operates correctly on an M-LAG network based on the command output.

For more information about M-LAG, see M-LAG configuration in *Layer 2—LAN Switching Configuration Guide*.

Figure 25 DHCP snooping on an M-LAG network



Restrictions and guidelines: DHCP snooping configuration

- The DHCP snooping configuration does not take effect on a Layer 2 Ethernet interface that is an aggregation member port. The configuration takes effect when the interface leaves the aggregation group.
- Specify the ports connected to authorized DHCP servers as trusted ports to make sure that DHCP clients can obtain valid IP addresses. The trusted ports and the ports connected to DHCP clients must be in the same VLAN.
- You can specify the following interfaces as trusted ports: Layer 2 Ethernet interfaces, Layer 2 aggregate interfaces, Layer 3 Ethernet interfaces, and Layer 3 aggregate interfaces. For more information about aggregate interfaces, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*.

DHCP snooping tasks at a glance

To configure DHCP snooping, perform the following tasks:

1. Configuring basic DHCP snooping features

2. (Optional.) Configuring DHCP snooping support for Option 82
3. (Optional.) Configuring DHCP snooping entry auto backup
4. (Optional.) Setting the maximum number of DHCP snooping entries
5. (Optional.) Configuring DHCP snooping security features
6. (Optional.) Enabling recording untrusted DHCP servers
7. (Optional.) Enabling the giaddr field check in DHCP requests
8. (Optional.) Enabling client offline detection on the DHCP snooping device
9. (Optional.) Enabling DHCP snooping logging
10. (Optional.) Enabling DHCP snooping alarm notifications
11. (Optional.) Enabling alarm log generation for packet drop events
12. (Optional.) Disabling DHCP snooping on an interface

Configuring basic DHCP snooping features

Configuring basic DHCP snooping features in a common network

About this task

Basic DHCP snooping features refer to the following:

- Enabling DHCP snooping.
- Configuring DHCP snooping trusted ports.
- Enabling recording client information in DHCP snooping entries.

If you enable DHCP snooping globally, DHCP snooping is enabled on all interfaces on the device.

You can also enable DHCP snooping for specific VLANs. After enabling DHCP snooping for a VLAN, you can configure the other basic DHCP snooping features in the VLAN.

Restrictions and guidelines

If the basic DHCP snooping features are configured globally, you can only use the undo form of the global configuration commands to disable the settings globally. The VLAN-specific configuration commands cannot disable the settings.

If the basic DHCP snooping features are configured in a VLAN, you can only use the undo form of the VLAN-specific configuration commands to disable the settings in the VLAN. The global configuration command cannot disable the settings.

Configuring basic DHCP snooping features globally

1. Enter system view.
   `system-view`
2. Enable DHCP snooping globally.
   `dhcp snooping enable`
   By default, DHCP snooping is disabled globally.
3. Enter interface view.
   `interface` *`interface-type interface-number`*
   This interface must connect to the DHCP server.
4. Specify the port as a trusted port.
   `dhcp snooping trust`

By default, all ports are untrusted ports after DHCP snooping is enabled.

5. (Optional.) Enable the recording of DHCP snooping entries.
   a. Return to system view.
      quit
   b. Enter interface view.
      interface *interface-type interface-number*
      This interface must connect to the DHCP client.
   c. Enable the recording of DHCP snooping entries.
      dhcp snooping binding record
      By default, the recording of DHCP snooping entries is disabled.

Configuring basic DHCP snooping features for VLANs

1. Enter system view.
   system-view
2. Enable DHCP snooping for VLANs.
   dhcp snooping enable vlan *vlan-id-list*
   By default, DHCP snooping is disabled for all VLANs.
3. Enter VLAN view
   vlan *vlan-id*
   Make sure DHCP snooping is enabled for the VLAN.
4. Configure an interface in the VLAN as a trusted port.
   dhcp snooping trust interface *interface-type interface-number*
   By default, all interfaces in the VLAN are untrusted ports.
5. (Optional.) Enable recording of client information in DHCP snooping entries.
   dhcp snooping binding record
   By default, recording of client information in DHCP snooping entries is disabled.

Configuring basic DHCP snooping features in a VXLAN network

About this task

In a VXLAN network, you can configure the following interfaces as DHCP snooping trusted interfaces:

- ACs that are mapped to a VSI.
- VXLAN tunnel interfaces that are assigned to a VSI.

When the VTEP with DHCP snooping configured receives a DHCP request, the VTEP forwards this request through the trusted ACs or VXLAN tunnel interfaces.

Restrictions and guidelines

If the DHCP server is in the local site, configure the AC that connects to the DHCP server as trusted. If the DHCP server is in a remote site, configure the VXLAN tunnel interface as trusted.

Configuring basic DHCP snooping features in a VXLAN network (DHCP server at the local site)

1. Enter system view.
   system-view

2. Enable DHCP snooping globally.
   dhcp snooping enable
3. By default, DHCP snooping is disabled globally.
4. Enter interface view.
   interface *interface-type interface-number*
5. Enter Ethernet service instance view.
   service-instance *instance-id*
6. Configure the AC as the DHCP snooping trusted interface.
   dhcp snooping trust

   By default, all ports are untrusted after DHCP snooping is enabled.
7. (Optional.) Enable recording of client information in DHCP snooping entries on the ACs mapped to the VSI and VXLAN tunnel interfaces assigned to the VSI.
   a. Exit to interface view.
      quit
   b. Exit to the system view.
      quit
   c. Enter VSI view.
      vsi *vsi-name*
   d. Enable recording of client information in DHCP snooping entries on the ACs mapped to the VSI and VXLAN tunnel interfaces assigned to the VSI.
      dhcp snooping binding record

      By default, the recording of DHCP snooping entries is disabled.

Configuring basic DHCP snooping features in a VXLAN network (DHCP server at a remote site)

1. Enter system view.
   system-view
2. Enable DHCP snooping globally.
   dhcp snooping enable
3. By default, DHCP snooping is disabled globally.
4. Enter VSI view.
   vsi *vsi-name*
5. Configure the VXLAN tunnel interfaces as the DHCP snooping trusted interface.
   dhcp snooping trust tunnel

   By default, all ports are untrusted after DHCP snooping is enabled.

   This command sets all VXLAN tunnel interfaces in the VSI as DHCP snooping trusted interfaces.
6. (Optional.) Enable recording of client information in DHCP snooping entries on the ACs mapped to the VSI and VXLAN tunnel interfaces assigned to the VSI.
   dhcp snooping binding record

   By default, the recording of DHCP snooping entries is disabled.

Configuring DHCP snooping support for Option 82

Restrictions and guidelines

- The Option 82 configuration on a Layer 2 Ethernet interface that has been added to an aggregation group does not take effect unless the interface leaves the aggregation group.
- To support Option 82, you must configure Option 82 on both the DHCP server and the DHCP snooping device. For information about configuring Option 82 on the DHCP server, see "Enabling handling of Option 82."
- If Option 82 contains the device name, the device name must contain no spaces. Otherwise, DHCP snooping drops the message. You can use the **sysname** command to specify the device name. For more information about this command, see *Fundamentals Command Reference*.
- DHCP snooping uses "outer VLAN tag.inner VLAN tag" to fill the VLAN ID field of sub-option 1 in verbose padding format if either of the following conditions exists:
  - DHCP snooping and QinQ work together.
  - DHCP snooping receives a DHCP packet with two VLAN tags.

  For example, if the outer VLAN tag is 10 and the inner VLAN tag is 20, the VLAN ID field is 000a.0014. The hexadecimal digit **a** represents the outer VLAN tag 10, and the hexadecimal digit **14** represents the inner VLAN tag 20.

Configuring DHCP snooping support for Option 82 in interface view

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable DHCP snooping to support Option 82.

   dhcp snooping information enable

   By default, DHCP snooping does not support Option 82.
4. (Optional.) Configure a handling strategy for DHCP requests that contain Option 82.

   dhcp snooping information strategy { **append** | **drop** | **keep** | **replace** }

   By default, the handling strategy is **replace**.

   If the handling strategy is **append** or **replace**, configure a padding mode and padding format for Option 82. If the handling strategy is **keep** or **drop**, you do not need to configure any padding mode or padding format for Option 82.
5. (Optional.) Configure the padding mode and padding format for the Circuit ID sub-option.

   dhcp snooping information circuit-id { **normal-extended** | [**vlan** *vlan-id*] **string** *circuit-id* | { **normal** | **verbose** [**node-identifier** { **mac** | **sysname** | **user-defined** *node-identifier* }] } [**format** { **ascii** | **hex** }] }

   By default, the padding mode is **normal** and the padding format is **hex** for the Circuit ID sub-option.

   If the device name (**sysname**) is configured as the padding content for sub-option 1, make sure the device name does not include spaces. Otherwise, the DHCP snooping device will fail to add or replace Option 82.
6. (Optional.) Configure the padding mode and padding format for the Remote ID sub-option.

   dhcp snooping information remote-id { **normal** [**format** { **ascii** | **hex** }] | [**vlan** *vlan-id*] { **hex** *remote-id* | **string** *remote-id* | **sysname** } }

   By default, the padding mode is **normal** and the padding format is **hex** for the Remote ID sub-option.

7. (Optional.) Configure the padding mode for the Vendor-Specific sub-option.

   dhcp snooping information vendor-specific [**vlan** *vlan-id*] **bas** [**node-identifier** { **mac** | **sysname** | **user-defined** *string* }]

   By default, the device does not insert any content into the Vendor-Specific sub-option.

Configuring DHCP snooping support for Option 82 in VLAN view

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Enable DHCP snooping to support Option 82.

   dhcp snooping information enable

   By default, DHCP snooping does not support Option 82.

4. (Optional.) Configure a handling strategy for DHCP requests that contain Option 82.

   dhcp snooping information strategy { **append** | **drop** | **keep** | **replace** }

   By default, the handling strategy is **replace**.

   If the handling strategy is **append** or **replace**, configure a padding mode and padding format for Option 82. If the handling strategy is **keep** or **drop**, you do not need to configure any padding mode or padding format for Option 82.

5. (Optional.) Configure the padding mode and padding format for the Circuit ID sub-option.

   dhcp snooping information circuit-id { **normal-extended** | **string** *circuit-id* | { **normal** | **verbose** [**node-identifier** { **mac** | **sysname** | **user-defined** *node-identifier* }] } [**format** { **ascii** | **hex** }] }

   By default, the padding mode is **normal** and the padding format is **hex** for the Circuit ID sub-option.

   If the device name (**sysname**) is configured as the padding content for the Circuit ID sub-option, make sure the device name does not include spaces. Otherwise, the DHCP snooping device will fail to add or replace Option 82.

6. (Optional.) Configure the padding mode and padding format for the Remote ID sub-option.

   dhcp snooping information remote-id { **normal** [**format** { **ascii** | **hex** }] | **hex** *remote-id* | **string** *remote-id* | **sysname** }

   By default, the padding mode is **normal** and the padding format is **hex** for the Remote ID sub-option.

7. (Optional.) Configure the padding mode for the Vendor-Specific sub-option.

   dhcp snooping information vendor-specific bas [**node-identifier** { **mac** | **sysname** | **user-defined** *string* }]

   By default, the device does not insert any content into the Vendor-Specific sub-option.

Configuring DHCP snooping entry auto backup

About this task

The auto backup feature saves DHCP snooping entries to a backup file, and allows the DHCP snooping device to download the entries from the backup file at device reboot. The entries on the DHCP snooping device cannot survive a reboot. The auto backup helps the security features provide services if these features (such as IP source guard) must use DHCP snooping entries for user authentication.

Restrictions and guidelines

If you disable DHCP snooping with the **undo dhcp snooping enable** command, the device deletes all DHCP snooping entries, but entries stored in the backup file still exist. They are deleted next time the device updates the backup file.

Procedure

1. Enter system view.

   system-view

2. Configure the DHCP snooping device to back up DHCP snooping entries to a file.

   dhcp snooping binding database filename { *filename* | **url** *url* [**username** *username* [**password** { **cipher** | **simple** } *string*]] }

   By default, the DHCP snooping device does not back up DHCP snooping entries.

   With this command executed, the DHCP snooping device backs up DHCP snooping entries immediately and runs auto backup.

   This command automatically creates the file if you specify a non-existent file.

3. (Optional.) Manually save DHCP snooping entries to the backup file.

   dhcp snooping binding database update now

4. (Optional.) Set the waiting time after a DHCP snooping entry change for the DHCP snooping device to update the backup file.

   dhcp snooping binding database update interval *interval*

   By default, the DHCP snooping device waits 300 seconds to update the backup file after a DHCP snooping entry change. If no DHCP snooping entry changes, the backup file is not updated.

Setting the maximum number of DHCP snooping entries

About this task

Perform this task to prevent the system resources from being overused.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the maximum number of DHCP snooping entries for the interface to learn.

   dhcp snooping max-learning-num *max-number*

   By default, the number of DHCP snooping entries for an interface to learn is unlimited.

Configuring DHCP snooping security features

Enabling DHCP starvation attack protection

About this task

A DHCP starvation attack occurs when an attacker constantly sends forged DHCP requests that contain identical or different sender MAC addresses in the **chaddr** field to a DHCP server. This

attack exhausts the IP address resources of the DHCP server so legitimate DHCP clients cannot obtain IP addresses. The DHCP server might also fail to work because of exhaustion of system resources. For information about the fields of DHCP packet, see "DHCP message format."

You can prevent DHCP starvation attacks in the following ways:

- If the forged DHCP requests contain different sender MAC addresses, use the **mac-address max-mac-count** command to set the MAC learning limit on a Layer 2 port. For more information about the command, see *Layer 2—LAN Switching Command Reference*.
- If the forged DHCP requests contain the same sender MAC address, perform this task to enable MAC address check for DHCP snooping. This feature compares the **chaddr** field of a received DHCP request with the source MAC address field in the frame header. If they are the same, the request is considered valid and forwarded to the DHCP server. If not, the request is discarded.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable MAC address check.
   dhcp snooping check mac-address
   By default, MAC address check is disabled.

Enabling DHCP-REQUEST attack protection

About this task

DHCP-REQUEST messages include DHCP lease renewal packets, DHCP-DECLINE packets, and DHCP-RELEASE packets. This feature prevents the unauthorized clients that forge the DHCP-REQUEST messages from attacking the DHCP server.

Attackers can forge DHCP lease renewal packets to renew leases for legitimate DHCP clients that no longer need the IP addresses. These forged messages disable the victim DHCP server from releasing the IP addresses.

Attackers can also forge DHCP-DECLINE or DHCP-RELEASE packets to terminate leases for legitimate DHCP clients that still need the IP addresses.

To prevent such attacks, you can enable DHCP-REQUEST check. This feature uses DHCP snooping entries to check incoming DHCP-REQUEST messages.

- If a matching entry is found for a message, this feature compares the entry with the message information.
  - If they are consistent, the message is considered as valid and forwarded to the DHCP server.
  - If they are different, the message is considered as a forged message and is discarded.
- If no matching entry is found, the message is considered valid and forwarded to the DHCP server.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable DHCP-REQUEST check.

dhcp snooping check request-message

By default, DHCP-REQUEST check is disabled.

Configuring a DHCP packet blocking port

About this task

Perform this task to configure a port as a DHCP packet blocking port. This blocking port drops all incoming DHCP requests.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the port to block DHCP requests.

   dhcp snooping deny

   By default, the port does not block DHCP requests.

CAUTION:

To avoid IP address acquisition failure, configure a port to block DHCP packets only if no DHCP clients are attached to it.

Enabling recording untrusted DHCP servers

About this task

Typically, a DHCP snooping device identifies the DHCP servers that are connected to the untrusted ports as untrusted. The snooping device drops incoming DHCP replies through these ports. With feature enabled, the snooping device will record the DHCP servers for dropped DHCP replies, and generate and send log messages to the information center. With the information center, you can set log message filtering and output rules, including output destinations.

Procedure

1. Enter system view.

   system-view
2. Enable recording untrusted DHCP servers.

   dhcp snooping untrusted-server-record enable

   By default, the device does not record untrusted DHCP servers.

Enabling the giaddr field check in DHCP requests

About this task

A DHCP snooping device functions between DHCP clients and a DHCP server, or between DHCP clients and a DHCP relay agent. The **giaddr** field in a DHCP request records the address information of the first relay agent that the request passes by. If the DHCP snooping devices receives a DHCP request where the **giaddr** field value is not 0, it indicates that the DHCP snooping device location is not correct. In this case, the DHCP snooping device cannot work correctly.

This feature enables the DHCP snooping device to examine the **giaddr** field value in received DHCP packets and drop them if the **giaddr** field is not 0. When the number of dropped DHCP requests reaches or exceeds the threshold, the device generates a log for administrators to adjust locations of the DHCP devices.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the **giaddr** field check in DHCP requests.
   dhcp snooping check giaddr
   By default, the device does not check the **giaddr** field in DHCP requests.

Enabling client offline detection on the DHCP snooping device

About this task

When a DHCP client goes offline abnormally, it does not send a message to the DHCP server to release the IP address. As a result, the DHCP server is not aware of the offline event and cannot release the client lease in a timely manner.

With this feature enabled, the DHCP snooping device performs the following operations when the ARP entry of a client ages out:

1. Deletes the DHCP snooping entry for the client.
2. Sends a DHCP-RELEASE message to the DHCP server to inform the server to release the address lease of the client.

Procedure

1. Enter system view.
   system-view
2. Enable client offline detection.
   dhcp snooping client-detect
   By default, client offline detection is disabled.

Enabling DHCP snooping logging

About this task

The DHCP snooping logging feature enables the DHCP snooping device to generate DHCP snooping logs and send them to the information center. The information helps administrators locate and solve problems. For information about the log destination and output rule configuration in the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

As a best practice, disable this feature if the log generation affects the device performance.

Procedure

1. Enter system view.
   system-view

2. Enable DHCP snooping logging.

   dhcp snooping log enable

   By default, DHCP snooping logging is disabled.

Enabling DHCP snooping alarm notifications

About DHCP snooping alarm notifications

The DHCP snooping device can report critical DHCP snooping events to network administrators in the following alarm notifications:

- **Binding-exhaust** alarm notifications: The DHCP snooping device generates this type of alarm notifications when the DHCP snooping entry resource is exhausted or recovers from the exhaustion condition.
- **Binding-mismatch** alarm notifications: The DHCP snooping device generates this type of alarm notifications when the number of DHCP requests dropped due to DHCP snooping entry mismatch has reached the alarm threshold.
- **Binding-threshold** alarm notifications: The DHCP snooping device generates this type of alarm notifications when the usage of DHCP snooping entry resources has reached or dropped below the alarm threshold.
- **Chaddr-mismatch** alarm notifications: The DHCP snooping device generates this type of alarm notifications when the number of DHCP requests dropped due to MAC address mismatch has reached the alarm threshold.
- **Rate-limit** alarm notifications: The DHCP snooping device generates this type of alarm notifications when the number of DHCP requests dropped due to reception rate limit exceeding has reached the alarm threshold.
- **Untrust-reply** alarm notifications: The DHCP snooping device generates this type of alarm notifications when the number of DHCP server replies dropped on untrusted interfaces has reached the alarm threshold.

After you enable DHCP snooping alarm notifications, the DHCP snooping module reports critical DHCP snooping events in alarm notifications to the SNMP module. For DHCP snooping event notifications to be sent correctly, you must configure SNMP on the device. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Enabling the binding-exhaust alarm

1. Enter system view.

   system-view

2. Enable the binding-exhaust alarm.

   snmp-agent trap enable dhcp snooping binding-exhaust

   By default, the binding-exhaust alarm is disabled for DHCP snooping.

3. Enter interface view.

   interface *interface-type interface-number*

4. (Optional.) Enable the binding-exhaust alarm on the interface.

   dhcp snooping exhaustion trap enable

   By default, the binding-exhaust alarm is disabled on the interface.

Enabling the binding-threshold alarm

1. Enter system view.
   system-view
2. Enable the binding-threshold alarm.
   snmp-agent trap enable dhcp snooping binding-threshold
   By default, the binding-threshold alarm is disabled for DHCP snooping.
3. Enter interface view.
   interface *interface-type interface-number*
4. (Optional.) Set the alarm threshold on the usage of DHCP snooping entry resources.
   dhcp snooping learning-num-threshold *threshold-value*
   By default, the alarm threshold on the usage of DHCP snooping entry resources is 100%.

Enabling the binding-mismatch or chaddr-mismatch packet drop alarm

About this task

This chapter introduces how to enable the **binding-mismatch** or **chaddr-mismatch** packet drop alarm.

Procedure

1. Enter system view.
   system-view
2. Enable the **binding-mismatch** or **chaddr-mismatch** alarm for the SNMP module.
   snmp-agent trap enable dhcp snooping [**binding-mismatch** | **chaddr-mismatch**] *
   By default, both the **binding-mismatch** and **chaddr-mismatch** alarms are disabled for the SNMP module.
3. Set the global alarm threshold that triggers packet drop alarm generation.
   dhcp snooping alarm { **mac-address** | **request-message** } **threshold** *threshold*
   By default, the global alarm threshold that triggers packet drop alarm generation is 100.
   For an interface, the **mac-address** keyword has the same effect as the **chaddr-mismatch** keyword in the **dhcp snooping alarm** { **binding-mismatch** | **chaddr-mismatch** } **threshold** command.
   For an interface, the **request-message** keyword has the same effect as the **binding-mismatch** keyword in the **dhcp snooping alarm** { **binding-mismatch** | **chaddr-mismatch** } **threshold** command.
4. Enter interface view.
   interface *interface-type interface-number*
5. (Optional.) Disable alarm log generation for packet drop events on the interface.
   undo dhcp snooping trap { **binding-mismatch** | **chaddr-mismatch** } **enable**
   By default, the packet drop alarm is disabled on the interface.
6. (Optional.) Set the global threshold that triggers alarm log generation for packet drop events on the interface.

dhcp snooping trap { **binding-mismatch** | **chaddr-mismatch** } **threshold** *threshold*

By default, the global threshold configured in system view takes effect on the interface.

Enabling the rate-limit or untrust-reply packet drop alarm

About this task

This chapter introduces how to enable the **rate-limit** or **untrust-reply** packet drop alarm.

After you enable the **rate-limit** or **untrust-reply** packet drop alarm for the SNMP module and the DHCP module, both modules generate alarm notifications when the alarm threshold is reached.

Restrictions and guidelines

- After you enable the **rate-limit** packet drop alarm for the SNMP or DHCP module, the module generates an alarm notification when the alarm threshold is reached. To set the alarm threshold that triggers the **rate-limit** packet drop alarm, use the **dhcp snooping trap threshold rate-limit** command.
- After you enable the **untrust-reply** packet drop alarm for the DHCP module, this feature functions as follows:
  - If you specify the **software** keyword, the software side collects statistics about the DHCP server replies dropped on each interface. When the number of DHCP server replies dropped on an interface reaches the alarm threshold, the DHCP module generates an alarm notification.
  - If you specify the **hardware** keyword, the hardware side collects statistics about the DHCP server replies dropped on untrusted interfaces of the device. When the total number of dropped DHCP server replies reaches the alarm threshold, the DHCP module generates an alarm notification.

Procedure

1. Enter system view.

   system-view
2. Enable the **rate-limit** or **untrust-reply** alarm for the SNMP module.

   snmp-agent trap enable dhcp snooping [**rate-limit** | **untrust-reply**] *

   By default, both the **rate-limit** and **untrust-reply** alarms are disabled for the SNMP module.
3. Enable the **rate-limit** or **untrust-reply** alarm for the DHCP module.

   dhcp snooping trap { **rate-limit** | **untrust-reply** { **software** | **hardware** } } **enable**

   By default, all types of packet drop alarms are disabled.
4. Set the global threshold that triggers packet drop alarm generation.

   dhcp snooping trap { **rate-limit** | **untrust-reply** { **software** | **hardware** } } **threshold** *threshold*

   By default, the global threshold is 100 for all packet drop types.
5. Enter interface view.

   interface *interface-type interface-number*
6. (Optional.) Disable the packet drop alarm on the interface.

   undo dhcp snooping trap { **rate-limit** | **untrust-reply software** } **enable**

   By default, all types of packet drop alarms are disabled on the interface.
7. (Optional.) Set the packet drop alarm threshold on the interface.

dhcp snooping trap { **rate-limit** | **untrust-reply software** } **threshold** *threshold*

By default, the global threshold configured in system view takes effect on the interface.

Enabling alarm log generation for packet drop events

About this task

After you enable alarm log generation for a packet drop type, the device generates an alarm log when the number of packets dropped reaches the alarm threshold. Then, the device clears the current packet drop statistics and counts packet drops again.

The alarm log is sent to the information center. For information about the log destination and output rule configuration in the information center, see *Network Management and Monitoring Configuration Guide*.

To set the alarm threshold, use the **dhcp snooping alarm threshold** command.

Restrictions and guidelines

For this feature to take effect, you must first execute the **dhcp snooping log enable** command to enable DHCP snooping logging.

Procedure

1. Enter system view.

   system-view
2. Enable alarm log generation for packet drop events.

   dhcp snooping alarm { **giaddr** | **mac-address** | **request-message** } **enable**

   By default, alarm log generation is disabled for packet drop events.
3. Set the global threshold that triggers packet drop alarm log generation.

   dhcp snooping alarm { **giaddr** | **mac-address** | **request-message** } **threshold** *threshold*

   By default, the global threshold is 100.

Disabling DHCP snooping on an interface

About this task

This feature allows you to narrow down the interface range where DHCP snooping takes effect. For example, to enable DHCP snooping globally except for a specific interface, you can enable DHCP snooping globally and disable DHCP snooping on the target interface.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Disable DHCP snooping on the interface.

   dhcp snooping disable

   By default:

- If you enable DHCP snooping globally or for a VLAN, DHCP snooping is enabled on all interfaces on the device or on all interfaces in the VLAN.
- If you do not enable DHCP snooping globally or for a VLAN, DHCP snooping is disabled on all interfaces on the device or on all interfaces in the VLAN.

Display and maintenance commands for DHCP snooping

Execute **display** commands in any view, and **reset** commands in user view.

Task	Command
Display the packet drop statistics on an interface enabled with the packet drop alarm.	**display dhcp snooping alarm packet statistics interface** *interface-type interface-number*
Display DHCP snooping entries.	**display dhcp snooping binding** [**ip** *ip-address* [**vlan** *vlan-id*] | **interface** *interface-type interface-number*] [**verbose**]
Display information about the file that stores DHCP snooping entries.	**display dhcp snooping binding database**
Display statistics about the packets exchanged between M-LAG member devices for DHCP snooping entry synchronization.	**display dhcp snooping m-lag-statistics**
Display M-LAG status information.	**display dhcp snooping m-lag-status**
Display Option 82 configuration information on the DHCP snooping device.	**display dhcp snooping information** { **all** | **interface** *interface-type interface-number* }
Display DHCP packet statistics on the DHCP snooping device.	**display dhcp snooping packet statistics** [**slot** *slot-number*]
Display information about trusted ports.	**display dhcp snooping trust**
Clear the packet drop statistics on an interface.	**reset dhcp snooping alarm packet statistics interface** *interface-type interface-number*
Clear DHCP snooping entries.	**reset dhcp snooping binding** { **all** | **ip** *ip-address* [**vlan** *vlan-id*] }
Clear statistics about the packets exchanged between M-LAG member devices for DHCP snooping entry synchronization.	**reset dhcp snooping m-lag-statistics**

Task	Command
Clear DHCP packet statistics on the DHCP snooping device.	**reset dhcp snooping packet statistics** [**slot** *slot-number*]

DHCP snooping configuration examples

Example: Configuring basic DHCP snooping features globally

Network configuration

As shown in Figure 26, Switch B is connected to the authorized DHCP server through GigabitEthernet 1/0/1, to the unauthorized DHCP server through GigabitEthernet 1/0/3, and to the DHCP client through GigabitEthernet 1/0/2.

Configure only the port connected to the authorized DHCP server to forward the responses from the DHCP server. Enable the DHCP snooping device to record clients' IP-to-MAC bindings by reading DHCP-ACK messages received from the trusted port and the DHCP-REQUEST messages.

Figure 26 Network diagram

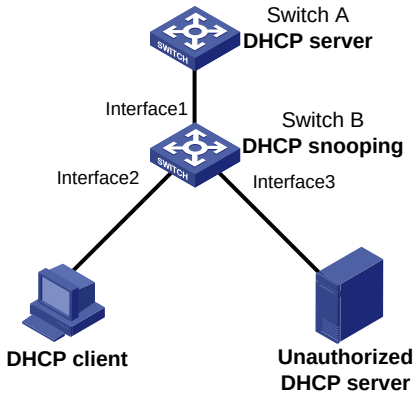


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Enable DHCP snooping globally.

```
<SwitchB> system-view
[SwitchB] dhcp snooping enable
```

Configure GigabitEthernet 1/0/1 as a trusted port.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] dhcp snooping trust
[SwitchB-GigabitEthernet1/0/1] quit
```

Enable recording clients' IP-to-MAC bindings on GigabitEthernet 1/0/2.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] dhcp snooping binding record
[SwitchB-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Verify that the DHCP client can obtain an IP address and other configuration parameters only from the authorized DHCP server. (Details not shown.)

Display the DHCP snooping entry recorded for the client.

```
[SwitchB] display dhcp snooping binding
```

Example: Configuring basic DHCP snooping features for a VLAN

Network configuration

As shown in Figure 27, Switch B is connected to the authorized DHCP server through GigabitEthernet 1/0/1, to the unauthorized DHCP server through GigabitEthernet 1/0/3, and to the DHCP client through GigabitEthernet 1/0/2.

Configure only the port in VLAN 100 connected to the authorized DHCP server to forward the responses from the DHCP server. Enable the port in VLAN 100 to record clients' IP-to-MAC bindings by reading DHCP-ACK messages received from the trusted port and the DHCP-REQUEST messages.

Figure 27 Network diagram

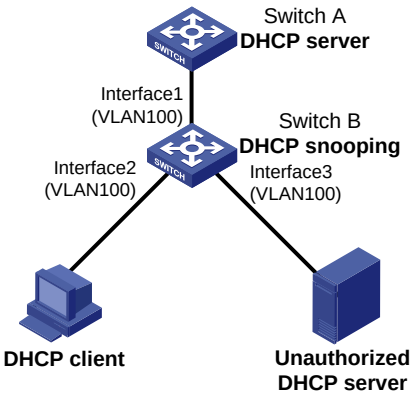


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Assign GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
[SwitchB-vlan100] quit
```

Enable DHCP snooping for VLAN 100.

```
[SwitchB] dhcp snooping enable vlan 100
```

Configure GigabitEthernet 1/0/1 as DHCP snooping trusted port.

```
[SwitchB] vlan 100
[SwitchB-vlan100] dhcp snooping trust interface gigabitethernet 1/0/1
```

Enable recording clients' IP-to-MAC bindings in VLAN 100.

```
[SwitchB-vlan100] dhcp snooping binding record
[SwitchB-vlan100] quit
```

Verifying the configuration

Verify that the DHCP client can obtain an IP address and other configuration parameters only from the authorized DHCP server. (Details not shown.)

Display the DHCP snooping entry recorded for the client.

```
[SwitchB] display dhcp snooping binding
```

Example: Configuring DHCP snooping support for Option 82

Network configuration

As shown in Figure 28, enable DHCP snooping and configure Option 82 on Switch B as follows:

- Configure the handling strategy for DHCP requests that contain Option 82 as **`replace`**.
- On GigabitEthernet 1/0/2, configure the padding content for the Circuit ID sub-option as **`company001`** and for the Remote ID sub-option as **`device001`**.
- On GigabitEthernet 1/0/3, configure the padding mode for the Circuit ID sub-option as **`verbose`**, access node identifier as **`sysname`**, and padding format as **`ascii`**. Configure the padding content for the Remote ID sub-option as **`device001`**.

Figure 28 Network diagram

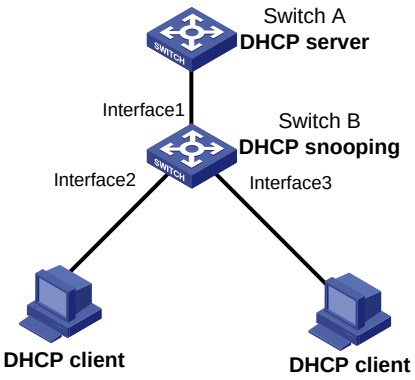


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Enable DHCP snooping.

```
<SwitchB> system-view
[SwitchB] dhcp snooping enable
```

Configure GigabitEthernet 1/0/1 as a trusted port.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] dhcp snooping trust
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure Option 82 on GigabitEthernet 1/0/2.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] dhcp snooping information enable
[SwitchB-GigabitEthernet1/0/2] dhcp snooping information strategy replace
[SwitchB-GigabitEthernet1/0/2] dhcp snooping information circuit-id string company001
[SwitchB-GigabitEthernet1/0/2] dhcp snooping information remote-id string device001
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure Option 82 on GigabitEthernet 1/0/3.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] dhcp snooping information enable
[SwitchB-GigabitEthernet1/0/3] dhcp snooping information strategy replace
[SwitchB-GigabitEthernet1/0/3] dhcp snooping information circuit-id verbose
node-identifier sysname format ascii
[SwitchB-GigabitEthernet1/0/3] dhcp snooping information remote-id string device001
```

Verifying the configuration

Display Option 82 configuration information on GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 on the DHCP snooping device.

```
[SwitchB] display dhcp snooping information
```

Configuring the BOOTP client

About BOOTP client

BOOTP client application

An interface that acts as a BOOTP client can use BOOTP to obtain information (such as IP address) from the BOOTP server.

To use BOOTP, an administrator must configure a BOOTP parameter file for each BOOTP client on the BOOTP server. The parameter file contains information such as MAC address and IP address of a BOOTP client. When a BOOTP client sends a request to the BOOTP server, the BOOTP server searches for the BOOTP parameter file and returns the corresponding configuration information.

BOOTP is usually used in relatively stable environments. In network environments that change frequently, DHCP is more suitable.

Because a DHCP server can interact with a BOOTP client, you can use the DHCP server to assign an IP address to the BOOTP client instead of using a BOOTP server. The DHCP server will assign an IP address to the BOOTP client based on the IP address allocation sequence.

Obtaining an IP address dynamically

A BOOTP client dynamically obtains an IP address from a BOOTP server as follows:

1. The BOOTP client broadcasts a BOOTP request, which contains its own MAC address.
2. Upon receiving the request, the BOOTP server searches the configuration file for the IP address and other information according to the BOOTP client's MAC address.
3. The BOOTP server returns a BOOTP response to the BOOTP client.
4. The BOOTP client obtains the IP address from the received response.

A DHCP server can take the place of the BOOTP server in the following dynamic IP address acquisition.

Protocols and standards

- RFC 951, *Bootstrap Protocol (BOOTP)*
- RFC 2132, *DHCP Options and BOOTP Vendor Extensions*
- RFC 1542, *Clarifications and Extensions for the Bootstrap Protocol*

Configuring an interface to use BOOTP for IP address acquisition

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*

   BOOTP client configuration applies only to Layer 3 Ethernet interfaces (including subinterfaces), Layer 3 aggregate interfaces, and VLAN interfaces.

3. Configure an interface to use BOOTP for IP address acquisition.
   `ip address bootp-alloc`
   By default, an interface does not use BOOTP for IP address acquisition.

Display and maintenance commands for BOOTP client

Execute **`display`** command in any view.

Task	Command
Display BOOTP client information.	**`display bootp client`** [**`interface`** *`interface-type interface-number`*]

BOOTP client configuration examples

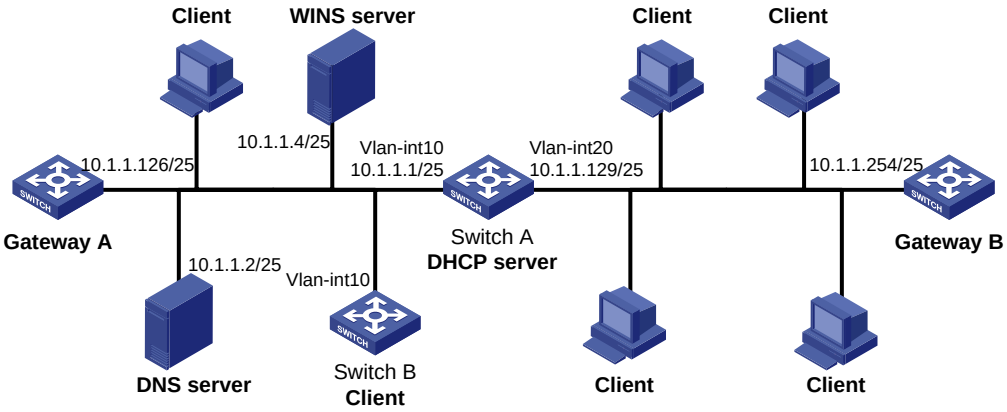
Example: Configuring BOOTP client

Network configuration

As shown in Figure 29, Switch B's port belonging to VLAN 10 is connected to the LAN. VLAN-interface 10 obtains an IP address from the DHCP server by using BOOTP.

To make the BOOTP client obtain an IP address from the DHCP server, you must perform configuration on the DHCP server. For more information, see "DHCP server configuration examples."

Figure 29 Network diagram



Procedure

The following describes the configuration on Switch B, which acts as a client.

Configure VLAN-interface 10 to dynamically obtain an IP address from the DHCP server.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ip address bootp-alloc
```

Verifying the configuration

Display the IP address assigned to the BOOTP client.

```
[SwitchB] display bootp client
```

Contents

Configuring DNS 1
About DNS 1
Types of DNS services 1
Static domain name resolution 1
DNS server-based domain name resolution 1
DNS server group-based domain name resolution 2
DNS proxy 3
DNS spoofing 4
DNS tasks at a glance 5
Configuring the DNS client 5
About this task 5
Configuring static domain name resolution 6
Configuring DNS server-based domain name resolution 6
Configuring DNS server group-based domain name resolution (independent domain name rules) 7
Configuring DNS server group-based domain name resolution (domain name rules associated with domain name groups) 8
Enabling DNS server health check 10
Configuring the DNS proxy 10
Enabling DNS proxy 10
Specifying DNS server addresses 11
Configuring a DNS server group 11
Configuring the DNS transparent proxy 11
Configuring DNS spoofing 12
Configuring DNS fast-reply 13
Specifying the source interface for DNS packets 13
Configuring the DNS trusted interface 14
Setting the DSCP value for outgoing DNS packets 14
Display and maintenance commands for DNS 15
IPv4 DNS configuration examples 15
Example: Configuring static domain name resolution 15
Example: Configuring dynamic domain name resolution 16
Example: Configuring DNS proxy 18
IPv6 DNS configuration examples 20
Example: Configuring static domain name resolution 20
Example: Configuring dynamic domain name resolution 20
Example: Configuring DNS proxy 23
Troubleshooting DNS configuration 24
Failure to resolve IPv4 addresses 24
Failure to resolve IPv6 addresses 25
Configuring DDNS 26
About DDNS 26
Restrictions and guidelines: DDNS configuration 26
DDNS client tasks at a glance 27
Configuring a DDNS policy 27
Applying the DDNS policy to an interface 29
Setting the DSCP value for outgoing DDNS packets 30
Display and maintenance commands for DDNS 30

Configuring DNS

About DNS

Domain Name System (DNS) is a distributed database used by TCP/IP applications to translate domain names into IP addresses. The domain name-to-IP address mapping is called a DNS entry.

Types of DNS services

DNS services can be static or dynamic. After a user specifies a name, the device checks the static name resolution table for an IP address. If no IP address is available, it contacts the DNS server for dynamic name resolution, which takes more time than static name resolution. To improve efficiency, you can put frequently queried name-to-IP address mappings in the local static name resolution table.

Static domain name resolution

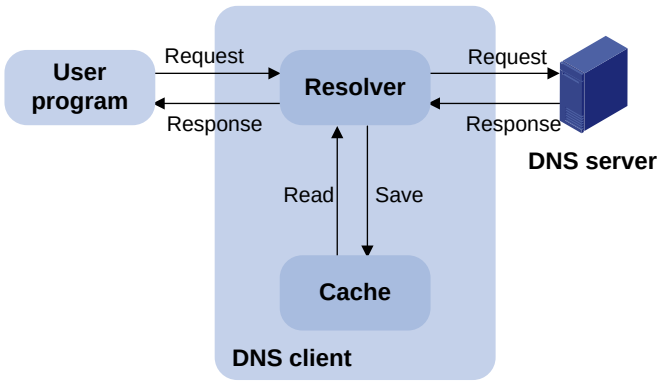
Static domain name resolution means manually creating mappings between domain names and IP addresses. For example, you can create a static DNS mapping for a device so that you can Telnet to the device by using the domain name.

DNS server-based domain name resolution

Architecture

Figure 1 shows the relationship between the user program, DNS client, and DNS server. The DNS client includes the resolver and cache. The user program and DNS client can run on the same device or different devices. The DNS server and the DNS client usually run on different devices.

Figure 1 DNS server-based domain name resolution



The device can function as a DNS client, but not a DNS server.

If an alias is configured for a domain name on the DNS server, the device can resolve the alias into the IP address of the host.

Resolution process

The DNS server-based domain name resolution process is as follows:

1. A user program sends a name query to the resolver of the DNS client.

2. The DNS resolver looks up the local domain name cache for a match. If the resolver finds a match, it sends the corresponding IP address back. If not, it sends a query to the DNS server.
3. The DNS server looks up the corresponding IP address of the domain name in its DNS database. If no match is found, the server sends a query to other DNS servers. This process continues until a result, whether successful or not, is returned.
4. After receiving a response from the DNS server, the DNS client returns the resolution result to the user program.

Caching

DNS server-based domain name resolution allows the DNS client to store latest DNS entries in the DNS cache. The DNS client does not need to send a request to the DNS server for a repeated query within the aging time. To make sure the entries from the DNS server are up to date, a DNS entry is removed when its aging timer expires. The DNS server determines how long a mapping is valid, and the DNS client obtains the aging information from DNS responses.

DNS suffixes

You can configure a domain name suffix list so that the resolver can use the list to supply the missing part of an incomplete name.

For example, you can configure **com** as the suffix for example.com. The user only needs to enter **example** to obtain the IP address of example.com. The resolver adds the suffix and delimiter before passing the name to the DNS server.

The name resolver handles the queries based on the domain names that the user enters:

- If the user enters a domain name without a dot (.) (for example, example), the resolver considers the domain name to be a host name. It adds a DNS suffix to the host name before performing the query operation. If no match is found for any host name and suffix combination, the resolver uses the user-entered domain name (for example, example) for the IP address query.
- If the user enters a domain name with a dot (.) among the letters (for example, www.example), the resolver directly uses this domain name for the query operation. If the query fails, the resolver adds a DNS suffix for another query operation.
- If the user enters a domain name with a dot (.) at the end (for example, example.com.), the resolver considers the domain name an FQDN and returns the successful or failed query result. The dot at the end of the domain name is considered a terminating symbol.

DNS server group-based domain name resolution

This feature allows the device (DNS client or proxy) to search for a matching domain name rule and send queries to servers in the DNS server group bound to the rule.

When the device receives a user query and fails to find a local matching DNS entry, the device determines whether a matching domain name rule exists in the same VPN instance or on the public network as the user. If a matching rule exists, the device forwards the query to DNS servers in the DNS server group bound to the rule. If no matching rule exists, the device does not forward the query to DNS servers in DNS server groups.

A domain name rule can be matched by using one of the following methods:

- **Exact match**—The match succeeds only when the domain name in the query is exactly the same as a domain name in the rule.
- **Fuzzy match**—The match succeeds if the domain name in the query contains a subdomain name.
- **Include the domain name group**—The device matches the domain name in a query against the domain names in the domain name groups. The match succeeds if the domain name in the query matches any domain name in a group.

If the domain name rule includes default domain name group **any**, any domain name can match the rule.

- **Exclude the domain name group**—The device matches the domain name in a query against the domain names in the domain name groups. The match succeeds only when the domain name in the query does not match any domain name in the domain name groups.

  If the domain name rule includes default domain name group **any**, no domain names can match the rule.

The device looks for a matching domain name rule in ascending order of rule IDs and compares the domain name in the query with domain names and subdomain names one by one. The domain name match order is the same as the order displayed in the **`display this`** command for the rule.

- If the domain name in the query matches a domain name or subdomain name in a rule, the device sends the query to DNS servers in the DNS server group bound to the rule.
  - After receiving the reply, the device sends the reply to the user and stores the DNS mapping in its local DNS cache.
  - If no reply is received, the device turns to the next domain name rule.
- If the domain name in the query does not match any domain names or subdomain names in all rules, the domain name resolution fails.

DNS proxy

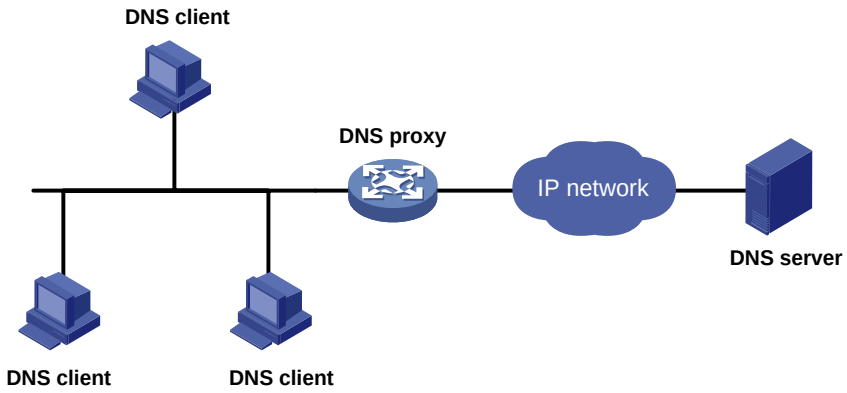
The DNS proxy performs the following functions:

- Forwards the request from the DNS client to the designated DNS server.
- Conveys the reply from the DNS server to the client.

The DNS proxy simplifies network management. When the DNS server address is changed, you can change the configuration only on the DNS proxy instead of on each DNS client.

Figure 2 shows the typical DNS proxy application.

Figure 2 DNS proxy application



A DNS proxy operates as follows:

2. A DNS client considers the DNS proxy as the DNS server, and sends a DNS request to the DNS proxy. The destination address of the request is the IP address of the DNS proxy.
3. The DNS proxy searches the local static domain name resolution table and dynamic domain name resolution cache after receiving the request. If the requested information is found, the DNS proxy returns a DNS reply to the client.
4. If the requested information is not found, the DNS proxy forwards the request as follows:
   a. If a matching DNS server group exists, the proxy forwards the request to DNS servers in the DNS server group bound to the rule.

b. If no matching DNS server group exists but a DNS server is specified, the DNS proxy sends the request to the DNS server for domain name resolution.

5. After receiving a reply from the DNS server, the DNS proxy records the IP address-to-domain name mapping and forwards the reply to the DNS client.

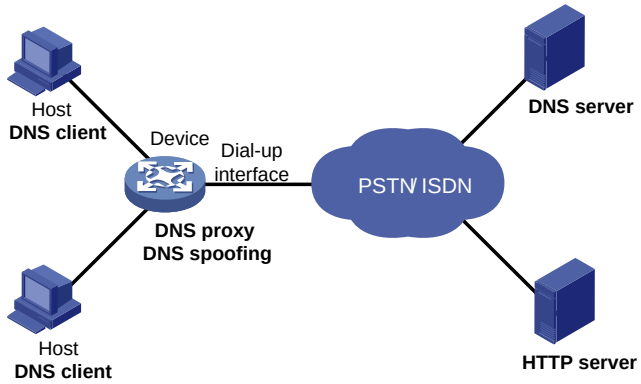
If no DNS server is designated or no route is available to the designated DNS server, the DNS proxy does not forward DNS requests.

DNS spoofing

As shown in Figure 3, DNS spoofing is applied to the dial-up network.

- The device connects to a PSTN/ISDN network through a dial-up interface. The device triggers the establishment of a dial-up connection only when packets are to be forwarded through the dial-up interface.
- The device acts as a DNS proxy and is specified as a DNS server on the hosts. After the dial-up connection is established, the device dynamically obtains the DNS server address through DHCP or another autoconfiguration mechanism.

Figure 3 DNS spoofing application



The DNS proxy does not have the DNS server address or cannot reach the DNS server after startup. A host accesses the HTTP server in the following steps:

2. The host sends a DNS request to the device to resolve the domain name of the HTTP server into an IP address.
3. Upon receiving the request, the device spoofs the host by replying a configured IP address in the following conditions: no dial-up connection is established, or the device does not know the DNS server address or cannot reach any DNS server. The device must have a route to the IP address with the dial-up interface as the output interface.

   The IP address configured for DNS spoofing is not the actual IP address of the requested domain name. Therefore, the TTL field is set to 0 in the DNS reply. When the DNS client receives the reply, it creates a DNS entry and ages it out immediately.
4. Upon receiving the reply, the host sends an HTTP request to the replied IP address.
5. When forwarding the HTTP request through the dial-up interface, the device performs the following operations:
   - Establishes a dial-up connection with the network.
   - Dynamically obtains the DNS server address through DHCP or another autoconfiguration mechanism.
6. Because the DNS entry ages out immediately upon creation, the host sends another DNS request to the device to resolve the HTTP server domain name.
7. The device operates the same as a DNS proxy. For more information, see "DNS proxy."
8. After obtaining the IP address of the HTTP server, the host can access the HTTP server.

Without DNS spoofing, the device forwards the DNS requests from the host to the DNS server if it cannot find a matching local DNS entry. However, the device cannot obtain the DNS server address, because no dial-up connection is established. Therefore, the device cannot forward or answer the requests from the client. DNS resolution fails, and the client cannot access the HTTP server.

DNS tasks at a glance

To configure DNS, perform the following tasks:

1. Configuring the DNS client

   Choose the following tasks as needed:

   - Configuring static domain name resolution
   - Configuring DNS server-based domain name resolution
   - Configuring DNS server group-based domain name resolution (independent domain name rules)
   - Configuring DNS server group-based domain name resolution (domain name rules associated with domain name groups)
   - Enabling DNS server health check
2. (Optional.) Configuring DNS proxy
   - Configuring the DNS proxy
   - Configuring the DNS transparent proxy
3. (Optional.) Configuring DNS spoofing

   This feature is applied to the dial-up network.
4. (Optional.) Configuring DNS fast-reply
5. (Optional.) Specifying the source interface for DNS packets
6. (Optional.) Configuring the DNS trusted interface
7. (Optional.) Setting the DSCP value for outgoing DNS packets

Configuring the DNS client

About this task

You can create domain name-to-address mappings on the DNS client through the following methods:

- **Configure static domain name resolution**—Use this method when you use domain names to access a small number of devices or when the network does not have available DNS servers. The network administrator must configure or maintain the domain name-to-address mappings manually.
- **Configure dynamic domain name resolution**—Use this method when you use domain names to access a large number of devices and the network has an available DNS server.

A DNS client resolves a domain name in the following order:

1. Static domain name resolution.
2. Locally saved DNS mappings that have been resolved.
3. DNS server-based domain name resolution.

The resolution fails if the domain name cannot be resolved after all these methods are used.

Configuring static domain name resolution

Restrictions and guidelines

For the public network or a VPN instance, each host name maps to only one IPv4 address and one IPv6 address.

A maximum of 2048 DNS entries can be configured for the public network or each VPN instance. You can configure DNS entries for both public network and VPN instances.

Procedure

1. Enter system view.

   system-view

2. Configure a host name-to-address mapping. Choose the options to configure as needed:

   IPv4:

   ip host *host-name ip-address* [**vpn-instance** *vpn-instance-name*]

   IPv6:

   ipv6 host *host-name ipv6-address* [**vpn-instance** *vpn-instance-name*]

Configuring DNS server-based domain name resolution

Restrictions and guidelines

- The limit on the number of DNS servers on the device is as follows:
  - In system view, you can specify a maximum of six DNS server IPv4 addresses for the public network or each VPN instance. You can specify DNS server IPv4 addresses for both public network and VPN instances.
  - In system view, you can specify a maximum of six DNS server IPv6 addresses for the public network or each VPN instance. You can specify DNS server IPv6 addresses for both public network and VPN instances.
- A DNS server address is required so that DNS queries can be sent to a correct server for resolution. If you specify both an IPv4 address and an IPv6 address, the device performs the following operations:
  - Sends an IPv4 DNS query first to the DNS server IPv4 addresses. If the query fails, the device turns to the DNS server IPv6 addresses.
  - Sends an IPv6 DNS query first to the DNS server IPv6 addresses. If the query fails, the device turns to the DNS server IPv4 addresses.
- The DNS server address priority is as follows:
  - A DNS server address specified earlier has a higher priority.
  - A DNS server address manually specified takes priority over a DNS server address dynamically obtained, for example, through DHCP.

  The device first sends a DNS query to the DNS server address of the highest priority. If the first query fails, it sends the DNS query to the DNS server address of the second highest priority, and so on.
- You can configure a DNS suffix that the system automatically adds to the incomplete domain name that a user enters.
  - You can configure a maximum of 16 DNS suffixes for the public network or each VPN instance. You can configure DNS suffixes for both public network and VPN instances.
  - A DNS suffix manually configured takes priority over a DNS suffix dynamically obtained, for example, through DHCP. A DNS suffix configured earlier has a higher priority. The device first uses the suffix that has the highest priority. If the query fails, the device uses the suffix that has the second highest priority, and so on.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Configure a DNS suffix.

   dns domain *domain-name* [**vpn-instance** *vpn-instance-name*]

   By default, no DNS suffix is configured and only the domain name that a user enters is resolved.

3. Specify a DNS server address.

   IPv4:

   dns server *ip-address* [**vpn-instance** *vpn-instance-name*] [**priority** *priority-value*]

   IPv6:

   ipv6 dns server *ipv6-address* [*interface-type interface-number*] [**vpn-instance** *vpn-instance-name*] [**priority** *priority-value*]

   By default, no DNS server address is specified.

Configuring DNS server group-based domain name resolution (independent domain name rules)

About this task

An independent domain name rule can be matched by using one of the following methods:

- **Exact match**—The match succeeds only when the domain name in the query is exactly the same as a domain name in the rule.
- **Fuzzy match**—The match succeeds if the domain name in the query contains a subdomain name.

Restrictions and guidelines

You can add both IPv4 and IPv6 DNS server addresses to a DNS server group.

The user query can only match a domain name rule that is in the same VPN instance or on the public network as the user. The device forwards the query to DNS servers of the DNS server group bound to the matching rule as follows:

- If the DNS servers are configured with priorities, the device selects DNS servers in descending order of their priorities.
- If the DNS servers are not configured with priorities or they have the same priority, the device selects DNS servers according to the server order displayed in the **display this** command.

If the query is an IPv4 packet, the device forwards the query first to IPv4 DNS servers. If the query is an IPv6 packet, the device forwards the query first to IPv6 DNS servers.

Procedure

1. Enter system view.

   system-view

2. Create a DNS server group.

   dns server-group *group-id* [**vpn-instance** *vpn-instance-name*]

   By default, no DNS server groups exist.

   The device supports a maximum of 16 DNS server groups. Each DNS server group can contain a maximum of 6 IPv4 DNS server addresses and a maximum of 6 IPv6 DNS server addresses.

3. Add DNS servers to the DNS server group.

   IPv4:

dns server *ip-address* [**priority** *priority-value*]

IPv6:

ipv6 dns server *ipv6-address* [*interface-type interface-number*] [**priority** *priority-value*]

By default, no DNS servers exist in the DNS server group.

4. Return to system view.

   quit

5. Configure a domain name rule.

   dns domain-rule *rule-id* { **domain-name** *domain-name* | **subdomain-name** *subdomain-name* } [**any-vpn-instance** | **vpn-instance** *vpn-instance-name*] **server-group** *group-id*

   By default, no domain name rule is configured.

Configuring DNS server group-based domain name resolution (domain name rules associated with domain name groups)

About this task

When the device uses domain name rules to match DNS queries, you can associate a domain name rule with a domain name group to simplify configuration. When multiple domain names are to be matched, you do not need to configure domain name rules multiple times.

After adding domain names or subdomain names to a domain name group, you can associate it with a domain name rule to enable the device to distribute DNS queries based on the matching results.

The relationship between a domain name rule and its associated domain name groups include the following types:

- **Include the domain name group**—The device matches the domain name in a query against the domain names in the domain name groups. The match succeeds if the domain name in the query matches any domain name in a group.

  If the domain name rule includes default domain name group **any**, any domain name can match the rule.

- **Exclude the domain name group**—The device matches the domain name in a query against the domain names in the domain name groups. The match succeeds only when the domain name in the query does not match any domain name in the domain name groups.

  If the domain name rule includes default domain name group **any**, no domain names can match the rule.

Restrictions and guidelines

You can add both IPv4 and IPv6 DNS server addresses to a DNS server group.

The user query can only match a domain name rule that is in the same VPN instance or on the public network as the user. The device forwards the query to DNS servers of the DNS server group bound to the matching rule as follows:

- If the DNS servers are configured with priorities, the device selects DNS servers in descending order of their priorities.
- If the DNS servers are not configured with priorities or they have the same priority, the device selects DNS servers according to the server order displayed in the **display this** command.

If the query is an IPv4 packet, the device forwards the query first to IPv4 DNS servers. If the query is an IPv6 packet, the device forwards the query first to IPv6 DNS servers.

Procedure

1. Enter system view.

   system-view

2. Create a DNS server group.

   dns server-group *group-id* [**vpn-instance** *vpn-instance-name*]

   By default, no DNS server groups exist.

   The device supports a maximum of 16 DNS server groups. Each DNS server group can contain a maximum of 6 IPv4 DNS server addresses and a maximum of 6 IPv6 DNS server addresses.

3. Add DNS servers to the DNS server group.

   IPv4:

   dns server *ip-address* [**priority** *priority-value*]

   IPv6:

   ipv6 dns server *ipv6-address* [*interface-type interface-number*] [**priority** *priority-value*]

   By default, no DNS servers exist in the DNS server group.

4. Return to system view.

   quit

5. Create a domain name group and enter its view, or enter an existing domain name group view.

   dns domain-name-group *group-name*

   By default, the system has default domain name group **any**.

   The default domain name group cannot be manually created or deleted.

   You can execute this command multiple times to create multiple domain name groups. The maximum number of domain name groups that can be created varies by device.

6. Add a domain name to the domain name group.

   domain-name *domain-name*

   By default, a domain name group does not have domain names.

7. Add a subdomain name to a domain name group.

   subdomain-name *subdomain-name*

   By default, a domain name group does not have subdomain names.

8. Return to system view.

   quit

9. Configure a domain name rule.

   - Associate the domain name group with a domain name rule and configure the domain name rule to match a DNS query when the domain name in the query is in the domain name group.

     dns domain-rule *rule-id* **domain-name-group** *group-name* [**any-vpn-instance** | **vpn-instance** *vpn-instance-name*] **server-group** *group-id*

   - Associate the domain name group with a domain name rule and configure the domain name rule to match a DNS query when the domain name in the query is not in the domain name group.

     dns domain-rule *rule-id* **exclude domain-name-group** *group-name* [**any-vpn-instance** | **vpn-instance** *vpn-instance-name*] **server-group** *group-id*

   By default, no domain name rule is configured.

Enabling DNS server health check

About this task

This feature enables the device to detect the availability of DNS servers in a DNS server group, which prevents the device from forwarding packets to unavailable DNS servers.

When the device receives a DNS request that matches a domain name rule, the device forwards the request to a DNS server. The DNS server is selected as follows:

- If the DNS servers are configured with priorities, the device selects DNS servers in descending order of their priorities.
- If the DNS servers are not configured with priorities or they have the same priority, the device selects DNS servers according to the server order displayed in the **display this** command.

After the device forwards the request to a DNS server, if it does not receive a DNS reply from the DNS server within two seconds, it reselects a DNS server. To prevent the device from selecting an unavailable DNS server, execute the **health-check enable** command to detect the availability of DNS servers. After you execute the command on the device, it works as follows:

1. The device periodically sends a DNS request to each DNS server in the DNS server group.
2. If the device receives a DNS reply from a DNS server, it regards the DNS server available.
3. If the device does not receive any DNS reply from a DNS server after it sends a DNS request to the DNS server for three consecutive times, it regards the DNS server unavailable.

After obtaining the availability of each DNS server in the DNS server group, the device automatically ignores the unavailable DNS servers and selects an available DNS server for packet forwarding according to the above selection mechanism.

Procedure

1. Enter system view.
   system-view
2. Create a DNS server group and enter its view.
   dns server-group *group-id* [**vpn-instance** *vpn-instance-name*]
3. Enable DNS server health check.
   health-check enable
   By default, DNS server health check is disabled.

Configuring the DNS proxy

Enabling DNS proxy

1. Enter system view.
   system-view
2. Enable DNS proxy.
   dns proxy enable
   By default, DNS proxy is disabled.

Specifying DNS server addresses

Restrictions and guidelines

You can specify multiple DNS servers. The DNS proxy forwards a request to the DNS server that has the highest priority. If having not received a reply, it forwards the request to a DNS server that has the second highest priority, and so on.

You can specify both an IPv4 address and an IPv6 address.

- A DNS proxy forwards an IPv4 name query first to IPv4 DNS servers. If no reply is received, it forwards the request to IPv6 DNS servers.
- A DNS proxy forwards an IPv6 name query first to IPv6 DNS servers. If no reply is received, it forwards the request to IPv4 DNS servers.

Procedure

1. Enter system view.

   system-view

2. Specify a DNS server address.

   IPv4:

   dns server *ip-address* [**vpn-instance** *vpn-instance-name*] [**priority** *priority-value*]

   IPv6:

   ipv6 dns server *ipv6-address* [*interface-type interface-number*] [**vpn-instance** *vpn-instance-name*] [**priority** *priority-value*]

   By default, no DNS server address is specified.

Configuring a DNS server group

For more information, see "Configuring DNS server group-based domain name resolution (independent domain name rules)" and "Configuring DNS server group-based domain name resolution (domain name rules associated with domain name groups)."

Configuring the DNS transparent proxy

About this task

DNS transparent proxy modifies the source address in DNS requests so that the DNS client seems to receive a DNS reply directly from the DNS server. This feature is applicable to domain name-based policies, such as security policies and bandwidth policies.

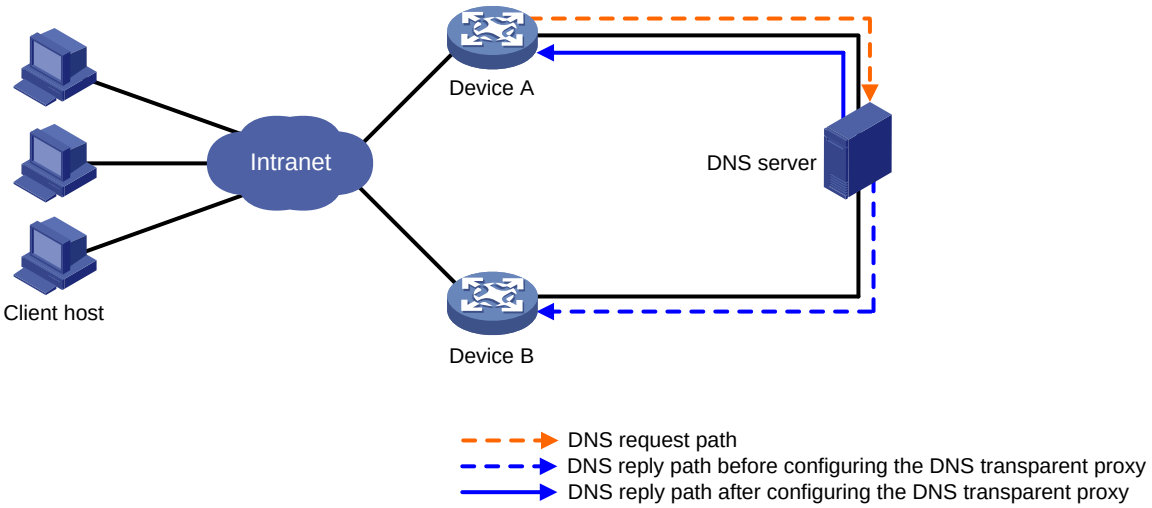
The DNS client does not configure the DNS server address as the DNS transparent proxy address, which simplifies the DNS client configuration. As a best practice, enable DNS transparent proxy in some load balancing scenarios.

The device enabled with DNS transparent proxy monitors received DNS requests and replies and records DNS mappings as follows:

1. The device monitors all received DNS packets. Upon receiving a DNS request, the device specifies a local IP address that can reach the DNS server as the source IP address for the request.
2. Upon receiving the DNS reply, the device records the DNS mapping and forwards the reply to the DNS client.

3. The device searches the local entries after receiving another request. If a match is found, the device returns a DNS reply to the client. If no match is found, the device forwards the query to the DNS server for domain name resolution.

Figure 4 Network diagram



As shown in Figure 4, a DNS transparent proxy operates as follows:

4. Device A is enabled with DNS transparent proxy. Upon receiving a DNS request, Device A changes the source address in the request to its own address and forwards the query to the DNS server.
5. Upon receiving the DNS reply, Device A records the domain name-to-IP address mapping and forwards the reply to the DNS client.

Restrictions and guidelines

Do not configure both the DNS transparent proxy and DNS fast-reply features on the device.

The DNS transparent proxy feature cannot be used across VPNs. Make sure the input and output interfaces of DNS packets on the device belong to the same VPN.

Procedure

1. Enter system view.

   `system-view`

2. Enable DNS transparent proxy.

   `dns transparent-proxy enable [ authorizing ]`

   By default, DNS transparent proxy is disabled.

Configuring DNS spoofing

Restrictions and guidelines

You can configure only one replied IPv4 address and one replied IPv6 address for the public network or a VPN instance. If you execute the command multiple times, the most recent configuration takes effect.

You can configure DNS spoofing for both public network and VPN instances.

After DNS spoofing takes effect, the device spoofs a DNS request even though a matching static DNS entry exists.

Prerequisites

The DNS proxy is enabled on the device.

No DNS server or route to any DNS server is specified on the device.

Procedure

1. Enter system view.
   system-view
2. Enable DNS proxy.
   dns proxy enable
   By default, DNS proxy is disabled.
3. Enable DNS spoofing and specify the IP address used to spoof DNS requests. Choose one option as needed:
   IPv4:
   dns spoofing *ip-address* [**vpn-instance** *vpn-instance-name*]
   IPv6:
   ipv6 dns spoofing *ipv6-address* [**vpn-instance** *vpn-instance-name*]
   By default, DNS spoofing is disabled.

Configuring DNS fast-reply

About this task

With this feature enabled, the device monitors the received DNS requests (only UDP packets are supported in the current software version). Then, it resolves the domain names in the requests, and searches for a match in the local static domain name resolution table.

- If a match is found, the device sends a DNS reply to the DNS client.
- If no match is found, the device forwards the query to the DNS server.

As DNS fast-reply can process a large number of DNS requests per second, use this feature in scenarios that require high DNS packet processing performance.

Restrictions and guidelines

This command does not take effect on labeled DNS packets.

Procedure

1. Enter system view.
   system-view
2. Enable DNS fast-reply.
   dns fast-reply enable
   By default, DNS fast-reply is disabled.

Specifying the source interface for DNS packets

About this task

This task enables the device to always use the primary IP address of the specified source interface as the source IP address of outgoing DNS packets. This feature applies to scenarios in which the DNS server responds only to DNS requests sourced from a specific IP address. If no IP address is configured on the source interface, no DNS packets can be sent out.

Restrictions and guidelines

When sending an IPv6 DNS request, the device follows the method defined in RFC 3484 to select an IPv6 address of the source interface.

You can configure only one source interface on the public network or a VPN instance. You can configure source interfaces for both public network and VPN instances.

Make sure the source interface belongs to the specified VPN instance if you specify the **vpn-instance** *vpn-instance-name* option.

Procedure

1. Enter system view.

   system-view

2. Specify the source interface for DNS packets.

   dns source-interface *interface-type interface-number* [**vpn-instance** *vpn-instance-name*]

   By default, no source interface for DNS packets is specified.

Configuring the DNS trusted interface

About this task

This task enables the device to use only the DNS suffix and domain name server information obtained through the trusted interface. The device can then obtain the correct resolved IP address. This feature protects the device against attackers that act as the DHCP server to assign incorrect DNS suffix and domain name server address.

Restrictions and guidelines

You can configure a maximum of 128 DNS trusted interfaces.

Procedure

1. Enter system view.

   system-view

2. Specify the DNS trusted interface.

   dns trust-interface *interface-type interface-number*

   By default, no DNS trusted interface is specified.

Setting the DSCP value for outgoing DNS packets

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet. A bigger DSCP value represents a higher priority.

Procedure

1. Enter system view.

   system-view

2. Set the DSCP value for DNS packets sent by a DNS client or a DNS proxy.

   IPv4:

   dns dscp *dscp-value*

   By default, the DSCP value is 0 in IPv4 DNS packets sent by a DNS client or a DNS proxy.

   IPv6:

ipv6 dns dscp *dscp-value*

By default, the DSCP value is 0 in IPv6 DNS packets sent by a DNS client or a DNS proxy.

Display and maintenance commands for DNS

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display DNS suffixes.	**display dns domain** [**dynamic**] [**vpn-instance** *vpn-instance-name*]
Display the domain name resolution table.	**display dns host** [**ip** | **ipv6**] [**vpn-instance** *vpn-instance-name*]
Display IPv4 DNS server information.	**display dns server** [**dynamic**] [**vpn-instance** *vpn-instance-name*]
Display health status of DNS servers.	**display dns server health status**
Display dynamic entries for domain name-to-IP address mappings recorded by DNS transparent proxy.	**display dns transparent host** [**ip** | **ipv6**] [**vpn-instance** *vpn-instance-name*]
Display IPv6 DNS server information.	**display ipv6 dns server** [**dynamic**] [**vpn-instance** *vpn-instance-name*]
Clear dynamic DNS entries.	**reset dns host** [**ip** | **ipv6**] [**vpn-instance** *vpn-instance-name*]

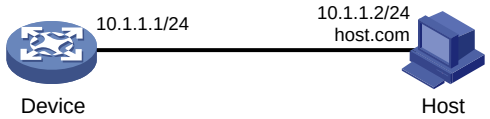
IPv4 DNS configuration examples

Example: Configuring static domain name resolution

Network configuration

As shown in Figure 5, the host at 10.1.1.2 is named **host.com**. Configure static IPv4 DNS on the device so that the device can use the easy-to-remember domain name rather than the IP address to access the host.

Figure 5 Network diagram



Procedure

Configure a mapping between host name **host.com** and IP address 10.1.1.2.

```
<Sysname> system-view
[Sysname] ip host host.com 10.1.1.2
```

Verify that the device can use static domain name resolution to resolve domain name **host.com** into IP address 10.1.1.2.

```
[Sysname] ping host.com
```

```
Ping host.com (10.1.1.2): 56 data bytes, press CTRL+C to break
56 bytes from 10.1.1.2: icmp_seq=0 ttl=255 time=1.000 ms
56 bytes from 10.1.1.2: icmp_seq=1 ttl=255 time=1.000 ms
56 bytes from 10.1.1.2: icmp_seq=2 ttl=255 time=1.000 ms
56 bytes from 10.1.1.2: icmp_seq=3 ttl=255 time=1.000 ms
56 bytes from 10.1.1.2: icmp_seq=4 ttl=255 time=2.000 ms

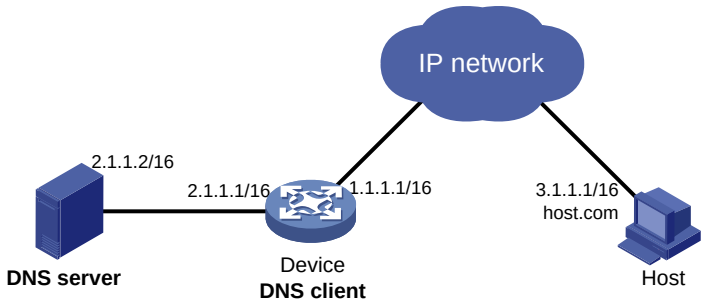
--- Ping statistics for host.com ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/1.200/2.000/0.400 ms
```

Example: Configuring dynamic domain name resolution

Network configuration

As shown in Figure 6, configure the DNS server to store the mapping between the host's domain name **host** and IPv4 address 3.1.1.1/16 in the **com** domain. Configure dynamic IPv4 DNS and DNS suffix **com** on the device so that the device can use domain name **host** to access the host.

Figure 6 Network diagram



Prerequisites

Before performing the following configuration, make sure that:

- The device and the host can reach each other.
- The IP addresses of the interfaces are configured as shown in Figure 6.

Procedure

1. Configure the DNS server:

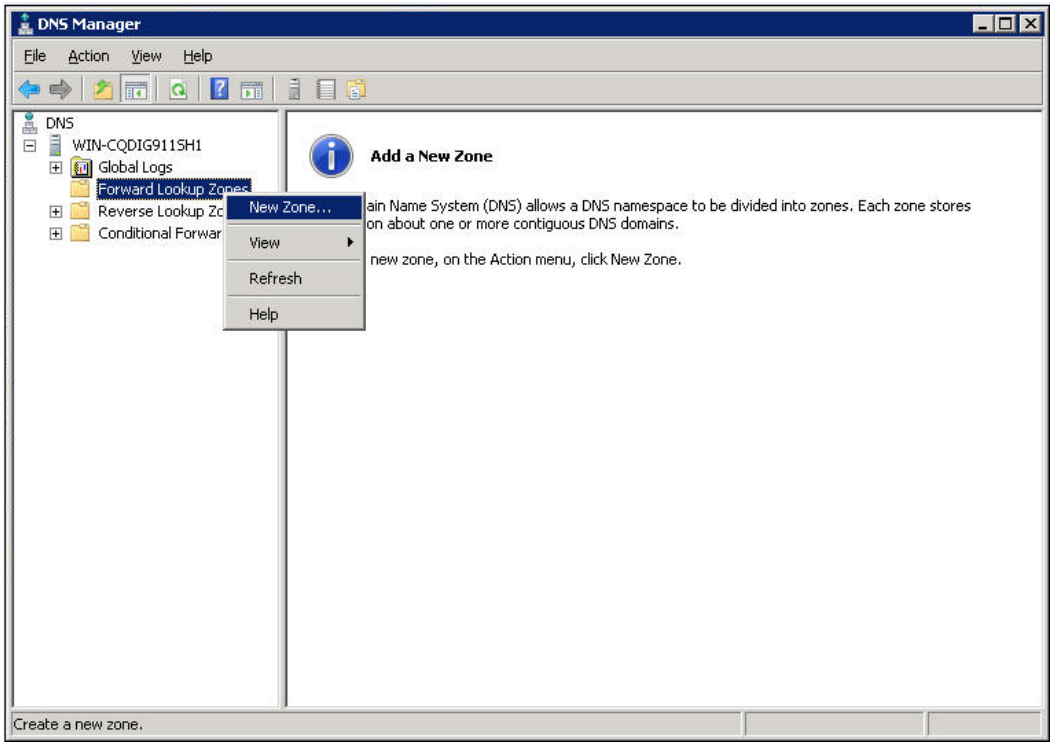
   The configuration might vary by DNS server. The following configuration is performed on a PC running Windows Server 2008 R2.

   a. Select **Start** > **Programs** > **Administrative Tools** > **DNS**.

      The DNS server configuration page appears, as shown in Figure 7.

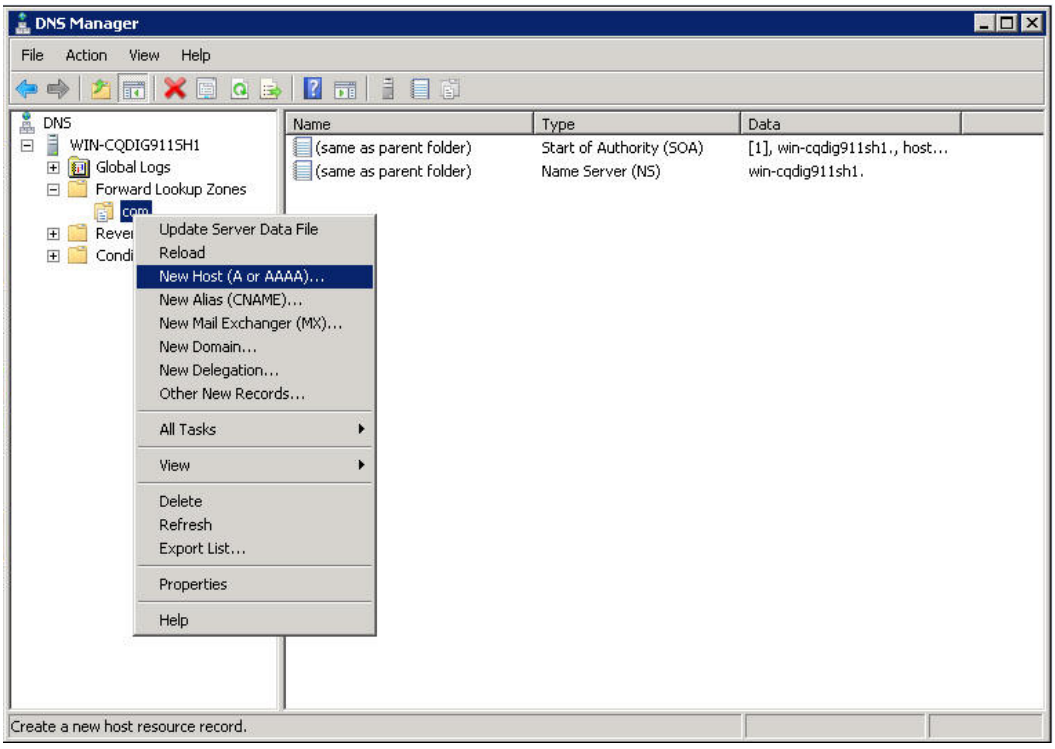
   b. Right-click **Forward Lookup Zones**, select **New Zone**, and then follow the wizard to create a new zone named **com**.

Figure 7 Creating a zone



c. On the DNS server configuration page, right-click zone **com** and select **New Host**.

Figure 8 Adding a host

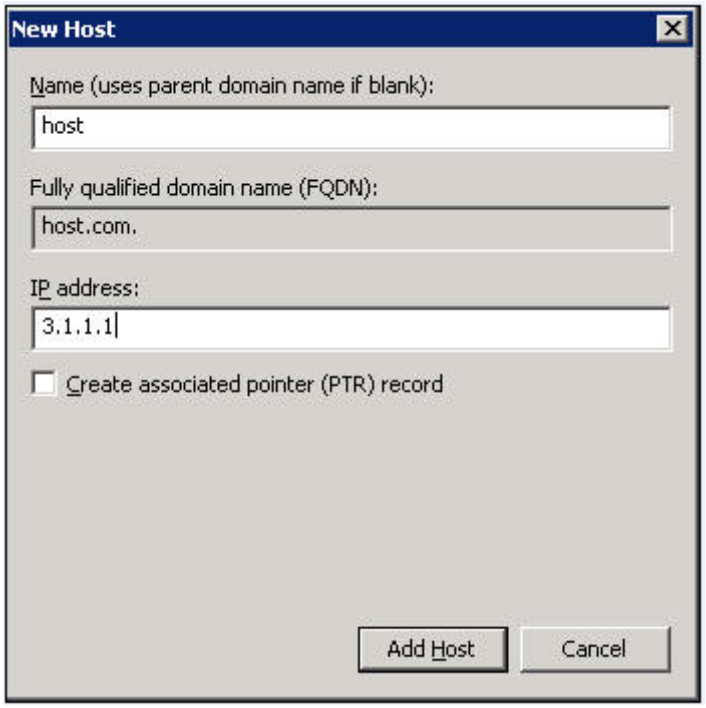


d. On the page that appears, enter host name **host** and IP address **3.1.1.1**.

e. Click **Add Host**.

The mapping between the IP address and host name is created.

Figure 9 Adding a mapping between domain name and IP address



2. Configure the DNS client:

   # Specify the DNS server 2.1.1.2.

   ```
   <Sysname> system-view
   [Sysname] dns server 2.1.1.2
   ```

   # Specify **com** as the name suffix.

   ```
   [Sysname] dns domain com
   ```

Verifying the configuration

Verify that the device can use the dynamic domain name resolution to resolve domain name **host.com** into IP address 3.1.1.1.

```
[Sysname] ping host
Ping host.com (3.1.1.1): 56 data bytes, press CTRL+C to break
56 bytes from 3.1.1.1: icmp_seq=0 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=1 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=2 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=3 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=4 ttl=255 time=2.000 ms

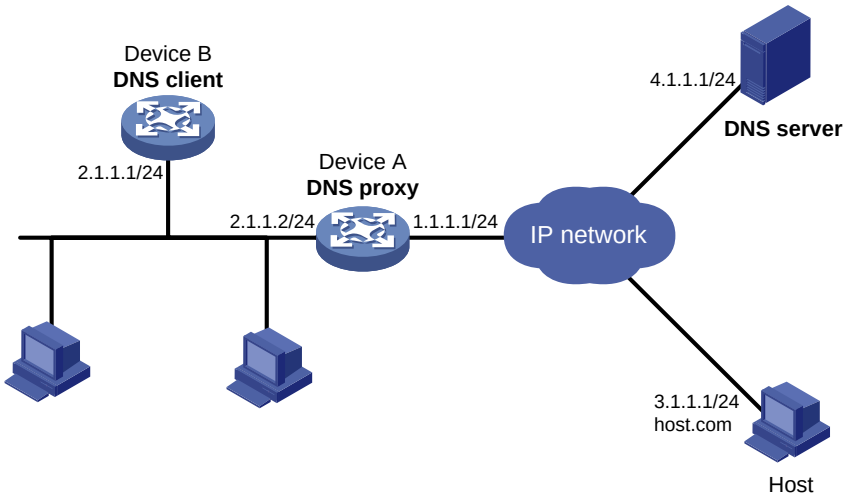
--- Ping statistics for host ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/1.200/2.000/0.400 ms
```

Example: Configuring DNS proxy

Network configuration

As shown in Figure 10, configure Device A as the DNS proxy to forward DNS packets between the DNS client (Device B) and the DNS server at 4.1.1.1.

Figure 10 Network diagram



Prerequisites

Before performing the following configuration, make sure that:

- Device A, the DNS server, and the host can reach each other.
- The IP addresses of the interfaces are configured as shown in Figure 10.

Procedure

1. Configure the DNS server:

   The configuration might vary by DNS server. When a PC running Windows Server 2008 R2 acts as the DNS server, see "Example: Configuring dynamic domain name resolution" for configuration information.

2. Configure the DNS proxy:

   # Specify the DNS server 4.1.1.1.

   ```
   <DeviceA> system-view
   [DeviceA] dns server 4.1.1.1
   ```

   # Enable DNS proxy.

   ```
   [DeviceA] dns proxy enable
   ```

3. Configure the DNS client:

   ```
   <DeviceB> system-view
   ```

   # Specify the DNS server 2.1.1.2.

   ```
   [DeviceB] dns server 2.1.1.2
   ```

Verifying the configuration

Verify that DNS proxy on Device A functions.

```
[DeviceB] ping host.com
Ping host.com (3.1.1.1): 56 data bytes, press CTRL+C to break
56 bytes from 3.1.1.1: icmp_seq=0 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=1 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=2 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=3 ttl=255 time=1.000 ms
56 bytes from 3.1.1.1: icmp_seq=4 ttl=255 time=2.000 ms

--- Ping statistics for host.com ---
```

```
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/1.200/2.000/0.400 ms
```

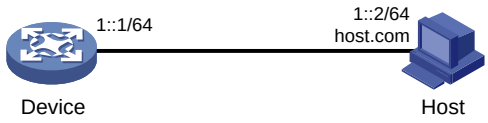
IPv6 DNS configuration examples

Example: Configuring static domain name resolution

Network configuration

As shown in Figure 11, the host at 1::2 is named **host.com**. Configure static IPv6 DNS on the device so that the device can use the easy-to-remember domain name rather than the IPv6 address to access the host.

Figure 11 Network diagram



Procedure

Configure a mapping between host name **host.com** and IPv6 address 1::2.

```
<Device> system-view
[Device] ipv6 host host.com 1::2
```

Verify that the device can use static domain name resolution to resolve domain name **host.com** into IPv6 address 1::2.

```
[Sysname] ping ipv6 host.com
Ping6(56 data bytes) 1::1 --> 1::2, press CTRL+C to break
56 bytes from 1::2, icmp_seq=0 hlim=128 time=1.000 ms
56 bytes from 1::2, icmp_seq=1 hlim=128 time=0.000 ms
56 bytes from 1::2, icmp_seq=2 hlim=128 time=1.000 ms
56 bytes from 1::2, icmp_seq=3 hlim=128 time=1.000 ms
56 bytes from 1::2, icmp_seq=4 hlim=128 time=0.000 ms

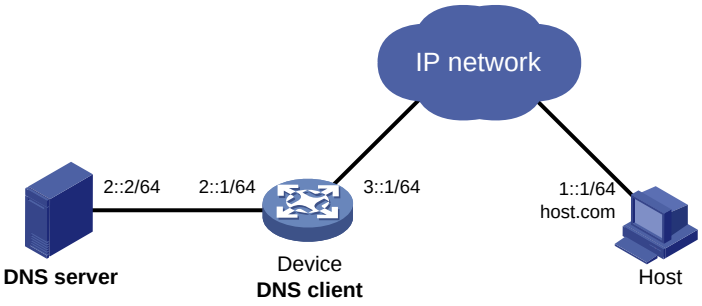
--- Ping6 statistics for host.com ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/0.600/1.000/0.490 ms
```

Example: Configuring dynamic domain name resolution

Network configuration

As shown in Figure 12, configure the DNS server to store the mapping between the host's domain name **host** and IPv6 address 1::1/64 in the **com** domain. Configure dynamic IPv6 DNS and DNS suffix **com** on the device so that the device can use domain name **host** to access the host.

Figure 12 Network diagram



Prerequisites

Before performing the following configuration, make sure that:

- The device and the host can reach each other.
- The IPv6 addresses of the interfaces are configured as shown in Figure 12.

Procedure

1. Configure the DNS server:

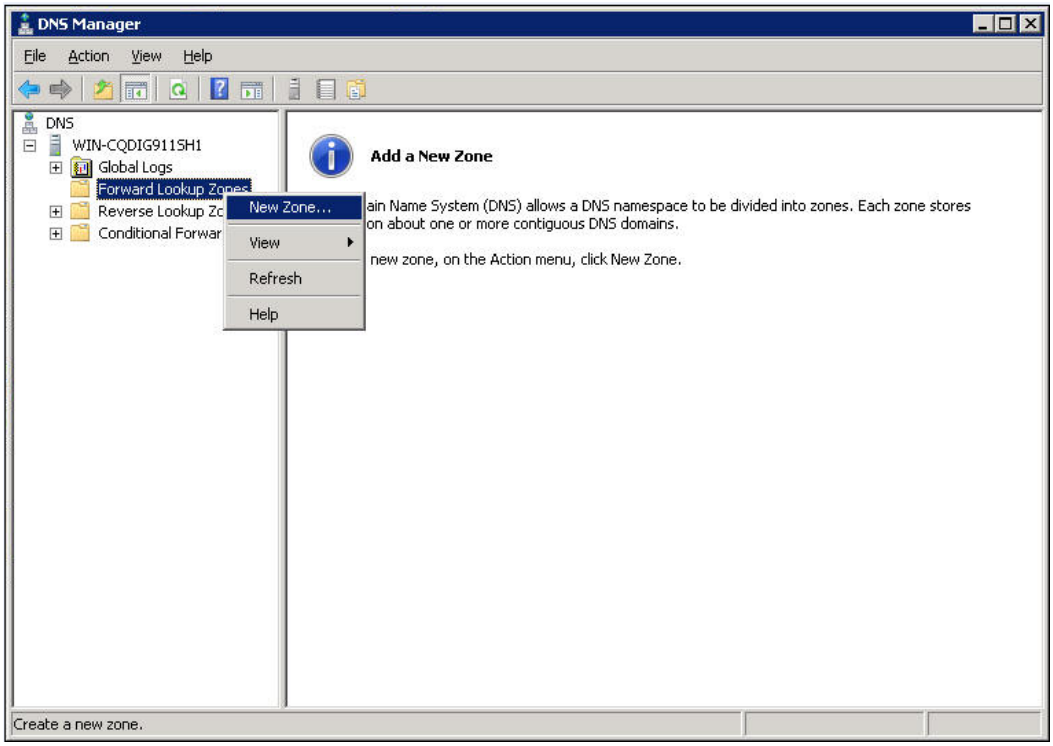
   The configuration might vary by DNS server. The following configuration is performed on a PC running Windows Server 2008 R2. Make sure that the DNS server supports IPv6 DNS so that the server can process IPv6 DNS packets and its interfaces can forward IPv6 packets.

   a. Select **Start** > **Programs** > **Administrative Tools** > **DNS**.

      The DNS server configuration page appears, as shown in Figure 13.

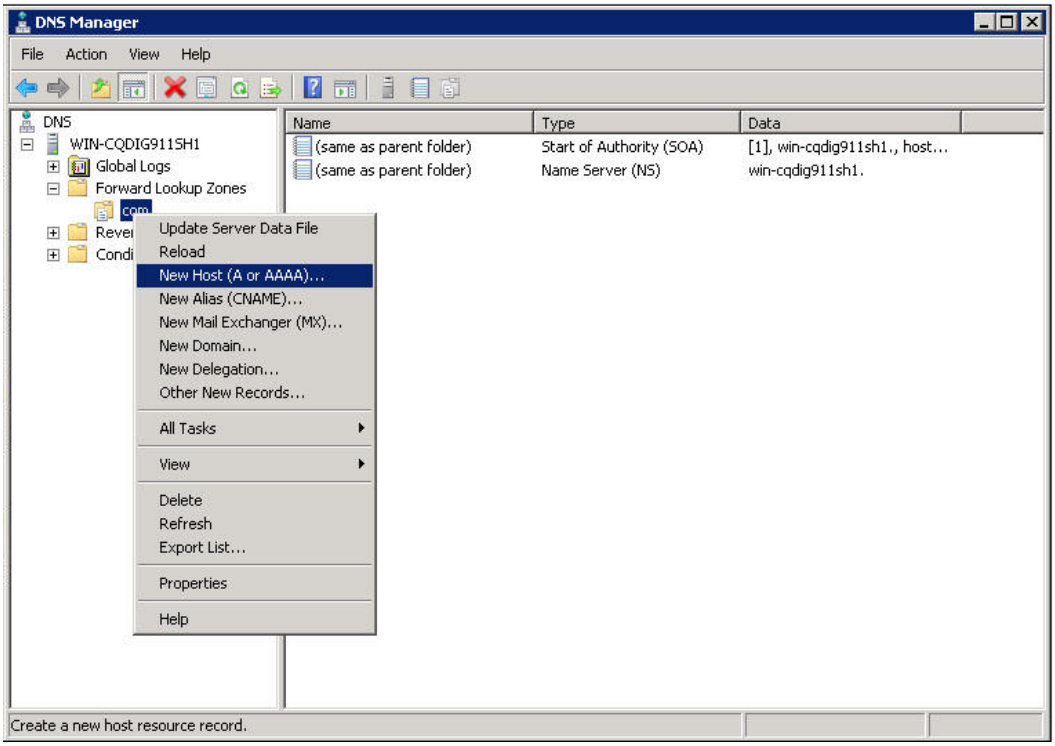
   b. Right-click **Forward Lookup Zones**, select **New Zone**, and then follow the wizard to create a new zone named **com**.

      Figure 13 Creating a zone



   c. On the DNS server configuration page, right-click zone **com** and select **New Host**.

Figure 14 Adding a host

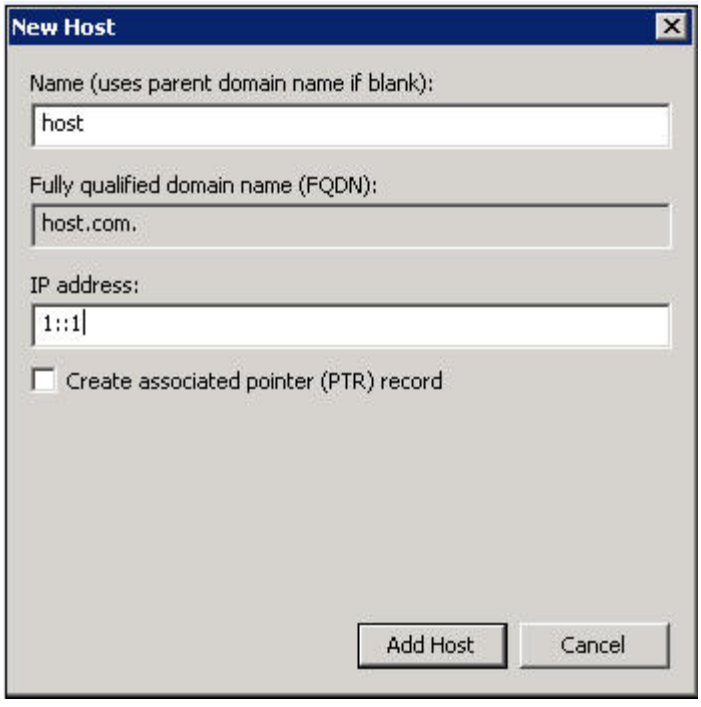


d. On the page that appears, enter host name **host** and IPv6 address **1::1**.

e. Click **Add Host**.

The mapping between the IPv6 address and host name is created.

Figure 15 Adding a mapping between domain name and IPv6 address



2. Configure the DNS client:

Specify the DNS server 2::2.

```
<Device> system-view
[Device] ipv6 dns server 2::2
```

Configure **com** as the DNS suffix.

```
[Device] dns domain com
```

Verifying the configuration

Verify that the device can use the dynamic domain name resolution to resolve the domain name **host.com** into the IP address 1::1.

```
[Device] ping ipv6 host
Ping6(56 data bytes) 3::1 --> 1::1, press CTRL+C to break
56 bytes from 1::1, icmp_seq=0 hlim=128 time=1.000 ms
56 bytes from 1::1, icmp_seq=1 hlim=128 time=0.000 ms
56 bytes from 1::1, icmp_seq=2 hlim=128 time=1.000 ms
56 bytes from 1::1, icmp_seq=3 hlim=128 time=1.000 ms
56 bytes from 1::1, icmp_seq=4 hlim=128 time=0.000 ms

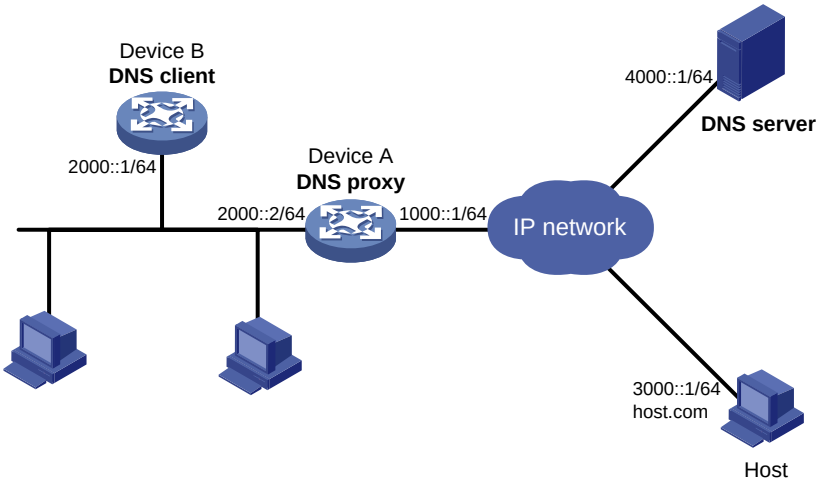
--- Ping6 statistics for host ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/0.600/1.000/0.490 ms
```

Example: Configuring DNS proxy

Network configuration

As shown in Figure 16, configure Device A as the DNS proxy to forward DNS packets between the DNS client (Device B) and the DNS server at 4000::1.

Figure 16 Network diagram



Prerequisites

Before performing the following configuration, make sure that:

- Device A, the DNS server, and the host are reachable to each other.
- The IPv6 addresses of the interfaces are configured as shown in Figure 16.

Procedure

1. Configure the DNS server:

This configuration might vary by DNS server. When a PC running Windows Server 2008 R2 acts as the DNS server, see "Example: Configuring dynamic domain name resolution" for configuration information.

2. Configure the DNS proxy:

   # Specify the DNS server 4000::1.

   ```
   <DeviceA> system-view
   [DeviceA] ipv6 dns server 4000::1
   ```

   # Enable DNS proxy.

   ```
   [DeviceA] dns proxy enable
   ```

3. Configure the DNS client:

   # Specify the DNS server 2000::2.

   ```
   <DeviceB> system-view
   [DeviceB] ipv6 dns server 2000::2
   ```

Verifying the configuration

Verify that DNS proxy on Device A functions.

```
[DeviceB] ping host.com
Ping6(56 data bytes) 2000::1 --> 3000::1, press CTRL+C to break
56 bytes from 3000::1, icmp_seq=0 hlim=128 time=1.000 ms
56 bytes from 3000::1, icmp_seq=1 hlim=128 time=0.000 ms
56 bytes from 3000::1, icmp_seq=2 hlim=128 time=1.000 ms
56 bytes from 3000::1, icmp_seq=3 hlim=128 time=1.000 ms
56 bytes from 3000::1, icmp_seq=4 hlim=128 time=0.000 ms

--- Ping6 statistics for host.com ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/0.600/1.000/0.490 ms
```

Troubleshooting DNS configuration

Failure to resolve IPv4 addresses

Symptom

After enabling dynamic domain name resolution, the user cannot get the correct IP address.

Solution

To resolve the problem:

1. Use the **`display dns host ip`** command to verify that the specified domain name is in the cache.
2. If the specified domain name does not exist, check that the DNS client can communicate with the DNS server.
3. If the specified domain name is in the cache, but the IP address is incorrect, check that the DNS client has the correct IP address of the DNS server.
4. Verify that the mapping between the domain name and IP address is correct on the DNS server.

Failure to resolve IPv6 addresses

Symptom

After enabling dynamic domain name resolution, the user cannot get the correct IPv6 address.

Solution

To resolve the problem:

1. Use the **`display dns host ipv6`** command to verify that the specified domain name is in the cache.
2. If the specified domain name does not exist, check that dynamic domain name resolution is enabled, and that the DNS client can communicate with the DNS server.
3. If the specified domain name is in the cache, but the IPv6 address is incorrect, check that the DNS client has the correct IPv6 address of the DNS server.
4. Verify that the mapping between the domain name and IPv6 address is correct on the DNS server.

Configuring DDNS

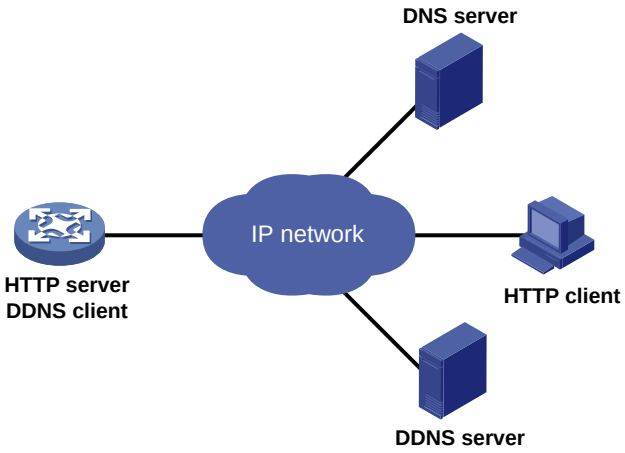
About DDNS

DNS provides only the static mappings between domain names and IP addresses. When the IP address of a node changes, your access to the node fails.

Dynamic Domain Name System (DDNS) can dynamically update the mappings between domain names and IP addresses for DNS servers.

Figure 17 shows the typical DDNS application.

Figure 17 DDNS application



DDNS works on the client-server model.

- **DDNS client**—A device that needs to update the mapping between its domain name and IP address dynamically on the DNS server when its IP address changes. An Internet user typically accesses an application layer server such as an HTTP server or an FTP server by using the server's domain name. When its IP address changes, the application layer server runs as a DDNS client. It sends a request to the DDNS server for updating the mapping between its domain name and its IP address.
- **DDNS server**—Informs the DNS server of latest mappings. When receiving the mapping update request from a DDNS client, the DDNS server tells the DNS server to re-map the domain name and the IP address of the DDNS client. Therefore, the Internet users can use the same domain name to access the DDNS client even if the IP address of the DDNS client has changed.

The device can function as a DDNS client to update the domain name-IP address mappings on the DNS servers through DDNS servers such as www.3322.org and PeanutHull.

NOTE:

The DDNS update process does not have a unified standard but varies by DDNS server that the DDNS client contacts.

Restrictions and guidelines: DDNS configuration

DDNS is supported by only IPv4 DNS and not supported by IPv6 DNS. It is used to update the mappings between domain names and IPv4 addresses.

DDNS client tasks at a glance

To configure a DDNS client, perform the following tasks:

1. Configuring a DDNS policy
2. Applying the DDNS policy to an interface
3. (Optional.) Setting the DSCP value for outgoing DDNS packets

Configuring a DDNS policy

About this task

A DDNS policy contains the DDNS server address, port number, login ID, password, time interval, associated SSL client policy, and update time interval. After creating a DDNS policy, you can apply it to multiple interfaces to simplify DDNS configuration.

Restrictions and guidelines

The URL address for update requests varies by DDNS server.

Table 1 Common URL addresses

DDNS server	URL address for DDNS update requests
www.3322.org	http://members.3322.org/dyndns/update?system=dyndns&hostname=<h>&myip=<a>
DYNDNS	http://members.dyndns.org/nic/update?system=dyndns&hostname=<h>&myip=<a>
DYNS	http://www.dyns.cx/postscript.php?host=<h>&ip=<a>
ZONEEDIT	http://dynamic.zoneedit.com/auth/dynamic.html?host=<h>&dnsto=<a>
TZO	http://cgi.tzo.com/webclient/signedon.html?TZOName=<h>IPAddress=<a>
EASYDNS	http://members.easydns.com/dyn/ez-ipupdate.php?action=edit&myip=<a>&host_id=<h>
HEIPV6TB	http://dyn.dns.he.net/nic/update?hostname=<h>&myip=<a>
CHANGE-IP	http://nic.changeip.com/nic/update?hostname=<h>&offline=1
NO-IP	http://dynupdate.no-ip.com/nic/update?hostname=<h>&myip=<a>
DHS	http://members.dhs.org/nic/hosts?domain=dyn.dhs.org&hostname=<h>&hostscmd=edit&hostscmdstage=2&type=1&ip=<a>
HP	https://server-name/nic/update?group=group-name&myip=<a>
ODS	ods://update.ods.org
GNUDIP	gnudip://*server-name*
PeanutHull	Select the URL according to your network situation: • oray://phddns60.oray.net • oray://phservice2.oray.net

Identify the DDNS server type in your network and follow the following restrictions and guidelines to set an appropriate URL address:

- The URL address for an update request can start with:
  - **http://**—The HTTP-based DDNS server.
  - **https://**—The HTTPS-based DDNS server.

    - **ods://**—The TCP-based ODS server.
    - **gnudip://**—The TCP-based GNUDIP server.
    - **oray://**—The TCP-based PeanutHull DDNS server.
- HP and GNUDIP are common DDNS update protocols. The *server-name* argument is the domain name or IP address of the service provider's server using one of the update protocols.
- The port number in the URL address is optional. If no port is specified, the system uses the default port numbers: port 80 for HTTP, port 443 for HTTPS, and port 6060 for PeanutHull DDNS server.
- The <h> value can be automatically filled with an FQDN if it is specified in the command for applying a DDNS policy to an interface. The <a> value is automatically filled with the primary IP address of the interface to which the DDNS policy is applied. For more information about applying DDNS policies, see "Applying the DDNS policy to an interface."
- You can also manually specify an FQDN and an IP address for the <h> and <a> fields. In this case, the FQDN specified at the CLI does not take effect. As a best practice, do not manually change the <h> and <a> because your configuration might be incorrect.
- No FQDN or IP address can be specified in the URL address for update requests sent to the PeanutHull DDNS server. You can specify the FQDN when applying the DDNS policy to an interface. The IP address is the primary IP address of the interface to which the DDNS policy is applied.

Prerequisites

Visit the website of a DDNS service provider, register an account, and apply for a domain name for the DDNS client. When the DDNS client updates the mapping between the domain name and the IP address through the DDNS server, the DDNS server checks the following:

- Whether the account information is correct.
- Whether the domain name to be updated belongs to the account.

Procedure

1. Enter system view.

   system-view
2. Create a DDNS policy and enter its view.

   ddns policy *policy-name*
3. Specify a URL address for DDNS update requests.

   url *request-url*

   By default, no URL address is specified for DDNS update requests.

   The URL address cannot contain a username or password. To configure them, use the **username** command and the **password** command.
4. Specify the username for logging in to the DDNS server.

   username *username*

   By default, no username is specified.
5. Specify the password for logging in to the DDNS server.

   password { **cipher** | **simple** } *string*

   By default, no password is specified.
6. (Optional.) Specify the parameter transmission method for sending DDNS update requests to HTTP/HTTPS-based DDNS servers.

   method { **http-get** | **http-post** }

   By default, the **http-get** method is used.

   This step is effective for communicating with HTTP/HTTPS-based DDNS servers.

Specify the **http-post** keyword for DDNS update with a DHS server.

7. (Optional.) Associate an SSL client policy with the DDNS policy.

   ssl-client-policy *policy-name*

   By default, no SSL client policy is associated with the DDNS policy.

   This step is only effective and a must for HTTP-based DDNS update requests. For SSL client policy configuration, see *Security Configuration Guide*.

8. (Optional.) Specify the interval for sending update requests.

   interval *days* [*hours* [*minutes*]]

   By default, the time interval is one hour.

Applying the DDNS policy to an interface

About this task

After you apply the DDNS policy to an interface and specify the FQDN for update, the DDNS client can send requests to the DDNS server. The requests are to update the mapping between the domain name and the primary IP address of the interface.

Restrictions and guidelines

- The **fqdn** *domain-name* option is a must for all DDNS servers except the PeanutHull DDNS server.
- The **fqdn** *domain-name* option is optional for PeanutHull DDNS server. If no FQDN is specified, the DDNS server updates all domain names for the DDNS client account. If an FQDN is specified, the DDNS server updates only the mapping between the specified FQDN and the primary IP address.

Prerequisites

Before you apply a DDNS policy to an interface, complete the following tasks:

- Specify the primary IP address of the interface and make sure the DDNS server and the interface can reach each other.
- Configure static or dynamic domain name resolution to translate the domain name of the DDNS server into the IPv4 address. For more information, see "Configuring the DNS client."

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Apply the DDNS policy to the interface to update the mapping between the specified FQDN and the primary IP address of the interface, and enable DDNS update.

   ddns apply policy *policy-name* [**fqdn** *domain-name*]

   By default, no DDNS policy is applied to the interface, no FQDN is specified for update, and DDNS update is disabled.

   An FQDN, including a host name and a domain name, is the only identifier for a network node and can be resolved as an IP address.

Setting the DSCP value for outgoing DDNS packets

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet. A bigger DSCP value represents a higher priority.

Procedure

1. Enter system view.
   system-view
2. Set the DSCP value for outgoing DDNS packets.
   ddns dscp *dscp-value*
   By default, the DSCP value for outgoing DDNS packets is 0.

Display and maintenance commands for DDNS

Execute **display** commands in any view.

Task	Command
Display DDNS policy information.	**display ddns policy** [*policy-name*]

Contents

Configuring mDNS gateway 1

About mDNS gateway 1
mDNS network model 1
mDNS gateway functionalities 2
Protocols and standards 3
mDNS gateway tasks at a glance 3
Enabling mDNS gateway 4
Configuring an mDNS gateway policy 4
Specifying the permitted service types for entry learning 4
Specifying trusted mDNS devices 5
Specifying the source address for mDNS packets 5
About specifying the source address for mDNS packets 5
Restrictions and guidelines 5
Specifying the source address for mDNS packets in a VLAN 5
Specifying the source address for mDNS packets in a VSI 6
Specifying the source address for mDNS packets in system view 6
Enabling mDNS gateway periodic probing 6
About mDNS gateway periodic probing 6
Restrictions and guidelines 7
Enabling mDNS gateway periodic probing in a VLAN 7
Enabling mDNS gateway periodic probing in a VSI 7
Enabling mDNS gateway periodic probing in a mDNS gateway policy 8
Enabling mDNS gateway proxy forwarding 8
Enabling the mDNS gateway to forward mDNS queries on the receiving port in a VLAN 8
Configuring the mDNS gateway to support QinQ 9
Display and maintenance commands for mDNS relay 10
mDNS gateway configuration examples 10
Example: Configuring mDNS gateway 10
Troubleshooting mDNS gateway 12
Failure to receive packets from an mDNS relay 12

Configuring mDNS gateway

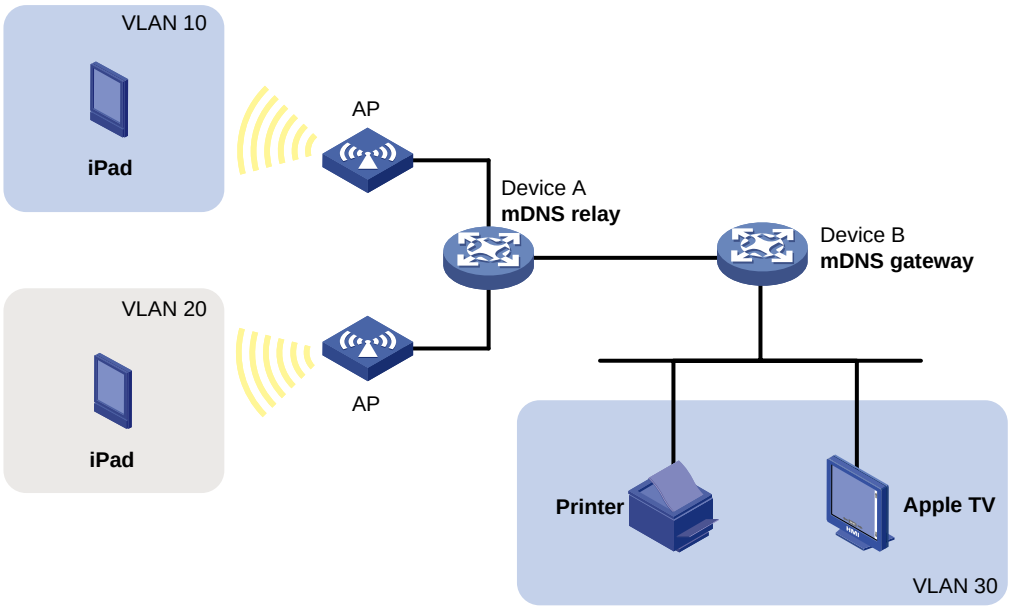
About mDNS gateway

A Multicast Domain Name System (mDNS) gateway can be deployed in a large-scale network to enable inter-VLAN or inter-VSI mDNS service discovery. It records a list of available services in the network, responds to service queries from mDNS clients, and performs active mDNS service discovery as well as access control.

mDNS network model

As shown in Figure 1, an mDNS network contains mDNS endpoint devices, mDNS service providers, the mDNS gateway, and mDNS relays.

Figure 1 mDNS networking model



mDNS endpoint device

An mDNS endpoint device obtains services from mDNS service providers. Examples are iPhones and iPads.

mDNS service provider

An mDNS service provider provides services for mDNS endpoint devices. Examples are mDNS-capable printers and Apple TVs.

mDNS gateway

The mDNS gateway records available services advertised by mDNS service providers and responds to queries from mDNS endpoint devices. With mDNS gateway enabled, Device B in Figure 1 can act as the mDNS gateway to help mDNS endpoint devices discover mDNS service providers in a different VLAN from the endpoint devices.

mDNS relay

An mDNS relay ensures successful mDNS communication when mDNS endpoint devices and mDNS service providers are in different subnets.

mDNS gateway functionalities

The mDNS gateway provides the following functionalities based on mDNS policies:

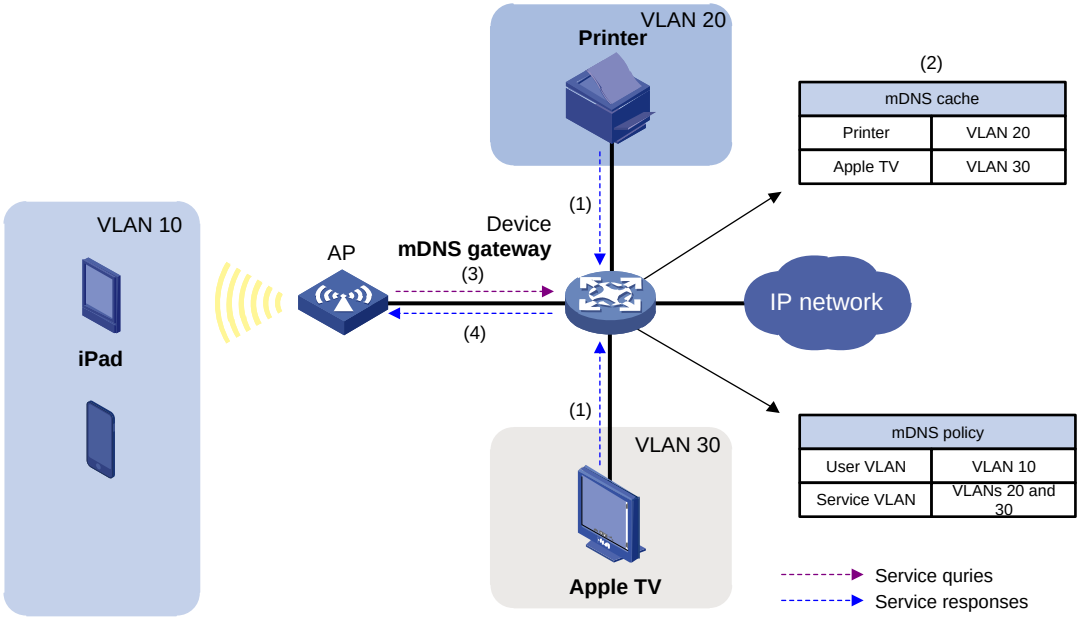
- **Reply mDNS queries.**
- **Discover mDNS services.**

An mDNS gateway policy defines user VLAN or VSI and service VLAN or VSI mappings to ensure that endpoint devices in specific VLANs or VSIs can access only service providers in specific service VLANs or VSIs.

Replying mDNS queries

Upon receiving mDNS queries from mDNS endpoint devices, the gateway checks its service list and sends responses that carry service information to the endpoint devices if matches are found.

Figure 2 Replying mDNS queries



As shown in Figure 2, the mDNS gateway replies mDNS queries in a VLAN-separated network as follows:

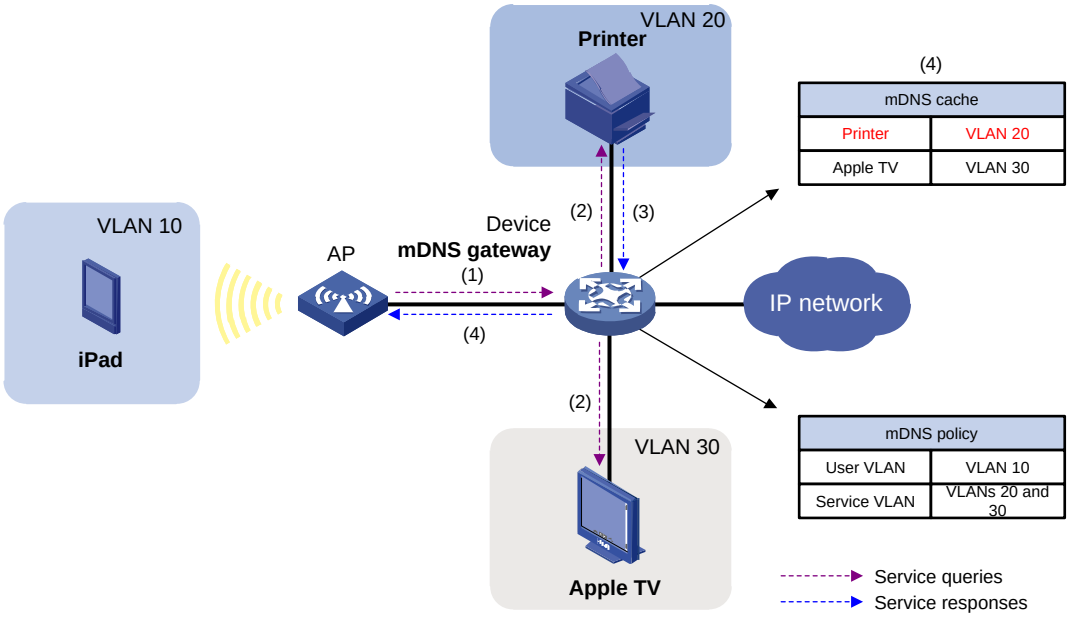
1. mDNS service providers broadcast mDNS service responses to notify services.
2. Upon receiving the responses, the mDNS gateway creates an mDNS service entry that records information such as service instance name and service type.
3. The mDNS endpoint device sends an mDNS query to the mDNS gateway.
4. The mDNS gateway searches its service list for the requested service.
   - If a match is found, the mDNS gateway sends a response to the endpoint device directly.
   - If no match is found, the mDNS gateway starts service discovery in the corresponding service VLANs of the user VLAN in which the endpoint device resides.
5. The mDNS endpoint device obtains information about the service provider from the response.

Discovering mDNS services

The mDNS gateway discovers mDNS services by listening to mDNS service responses sent by mDNS service providers and then records information about the services as entries. Service providers can broadcast service responses actively or reply service responses upon requests.

After receiving a query to a service that is not in the service list, the mDNS gateway forwards the query in the corresponding service VLANs of the user VLAN. Upon receiving the query, service providers that provide the service will send responses.

Figure 3 Discovering mDNS services



As shown in Figure 3, the mDNS gateway discovers services as follows:

1. The mDNS gateway receives a query for the printing service from the mDNS endpoint device (iPad).
2. The mDNS gateway searches its service list for the printing service and no match is found.
3. The mDNS gateway forwards the mDNS query to the corresponding service VLANs.
4. Upon receiving the query, the printer sends an mDNS response.
5. Upon receiving the response, the mDNS gateway records information about the service and sends a response to the mDNS endpoint device.

Protocols and standards

RFC 6762, *Multicast DNS*

mDNS gateway tasks at a glance

To configure the mDNS gateway, perform the following tasks:

1. Enabling mDNS gateway
2. (Optional.) Configuring an mDNS gateway policy
3. (Optional.) Specifying the permitted service types for entry learning
4. (Optional.) Specifying trusted mDNS devices
5. (Optional.) Specifying the source address for mDNS packets
6. (Optional.) Enabling mDNS gateway periodic probing
7. (Optional.) Enabling mDNS gateway proxy forwarding
8. (Optional.) Enabling the mDNS gateway to forward mDNS queries on the receiving port in a VLAN

9. (Optional.) Configuring the mDNS gateway to support QinQ

Enabling mDNS gateway

1. Enter system view.
   system-view
2. Enable mDNS gateway.
   mdns gateway enable
   By default, mDNS gateway is disabled.

Configuring an mDNS gateway policy

Restrictions and guidelines

With mDNS gateway policies configured, the mDNS gateway processes only the queries of mDNS endpoint devices in user VLANs or VSIs configured in the policies.

Procedure

1. Enter system view.
   system-view
2. Create an mDNS gateway policy and enter its view.
   mdns gateway policy *policy-name*
3. Specify service VLANs.
   service vlan *vlan-id-list*
   By default, no service VLAN is specified.
4. Specify service VSIs.
   service vsi *vsi-name*
   By default, no service VSI is specified.
5. Specify user VLANs.
   user vlan *vlan-id-list*
   By default, no user VLAN is specified.
6. Specify user VSIs.
   user vsi *vsi-name*
   By default, no user VSI is specified.

Specifying the permitted service types for entry learning

About this task

By default, the mDNS gateway can learn all mDNS services of a predefined type, which can be memory-consuming if a large number of service providers exist. You can perform this task to specify the permitted service types that can be learned by the mDNS gateway for the gateway to learn only required services.

Procedure

1. Enter system view.

system-view

2. Specify a permitted service type for entry learning.

   mdns gateway permit service-type { *predefined-service-type* | **user-defined** *user-defined-service-type* } [**max-entry-count** *count*]

   By default, the mDNS gateway can learn all mDNS services of a pre-defined type.

Specifying trusted mDNS devices

About this task

To prevent security issues and resource waste caused by attacks performed by rogue mDNS devices, you can specify trustworthy mDNS relays as trusted mDNS devices. With trusted devices specified, the mDNS gateway processes only unicast queries from the trusted mDNS devices.

Procedure

1. Enter system view.

   system-view

2. Specify a trusted mDNS device.

   mdns gateway trust device-ip *ip-address* [**vpn-instance** *vpn-instance-name*]

   By default, no trusted device is specified.

Specifying the source address for mDNS packets

About specifying the source address for mDNS packets

For responses destined for endpoint devices and mDNS queries and probe queries sent within a VLAN or VSI, the default source IP address is the primary IP address of the VLAN interface or VSI interface. If no IP address is configured for the VLAN interface or VSI interface, the mDNS gateway cannot send out the queries.

If the mDNS gateway is connected to a large number of VLANs or VSIs, configuring an IP address to each VLAN interface or VSI interface wastes IP address resources. You can use this command to specify a specific IP address for the mDNS gateway to use.

Restrictions and guidelines

When you configure the same source address for mDNS packets configured in different views, the configuration in the following views is in descending order of priority: VLAN view > system view > VSI view.

Specifying the source address for mDNS packets in a VLAN

1. Enter system view.

   system-view

2. Enable mDNS gateway.

   mdns gateway enable

   By default, mDNS gateway is disabled.

3. Enter VLAN view.

   vlan *vlan-id*

4. Specify the source address for mDNS packets.

   mdns gateway source ip { **interface** *interface-type interface-number* | *ip-address* }

   By default, the source IP address is the primary IP address of the VLAN interface.

   You can specify only one source address for a VLAN. If you execute this command multiple times, the most recent configuration takes effect.

Specifying the source address for mDNS packets in a VSI

1. Enter system view.

   system-view

2. Enable mDNS gateway.

   mdns gateway enable

   By default, mDNS gateway is disabled.

3. Enter VSI view.

   vsi *vsi-id*

4. Specify the source address for mDNS packets.

   mdns gateway source ip { **interface** *interface-type interface-number* | *ip-address* }

   By default, the source IP address is the primary IP address of the VSI interface.

   You can specify only one source address for a VSI. If you execute this command multiple times, the most recent configuration takes effect.

Specifying the source address for mDNS packets in system view

1. Enter system view.

   system-view

2. Enable mDNS gateway.

   mdns gateway enable

   By default, mDNS gateway is disabled.

3. Specify the source address for mDNS packets.

   mdns gateway source ip { **interface** *interface-type interface-number* | *ip-address* }

   By default, the mDNS gateway uses the primary IP address of the VLAN or VSI interface for packet output as the source IP address of mDNS packets.

   In QinQ scenarios, as a best practice to simplify configuration and prevent the mDNS gateway from sending multiple mDNS probe queries, configure this command in system view.

Enabling mDNS gateway periodic probing

About mDNS gateway periodic probing

mDNS service providers broadcast their services at network association. For the mDNS gateway to discover services offered by service providers that accessed the network earlier than the gateway, enable mDNS gateway periodic probing.

This feature enables the mDNS gateway to send mDNS probe queries at intervals. Each mDNS service provider sends a response to the gateway after receiving the query. The mDNS gateway can learn the latest service information of all mDNS service providers in the network in a timely manner.

Restrictions and guidelines

For an mDNS device, you can enable periodic probing on its directly connected mDNS gateway or mDNS relay.

If you execute the **mdns gateway probe interval** command in both mDNS gateway policy view and VLAN/VSI view, the mDNS gateway will send two probe queries. As a best practice, execute the command in only one of the views.

As a best practice to control periodic probing in different VLANs or VSIs, configure multiple mDNS gateway policies and execute the **mdns gateway probe interval** command in the view of each mDNS gateway policy.

In a QinQ scenario, to avoid any impact on the device service networking capabilities due to too many mDNS multicast packets sent within a short period, configure the inner VLANs as accurately as possible.

Enabling mDNS gateway periodic probing in a VLAN

1. Enter system view.
   system-view
2. Enable mDNS gateway.
   mdns gateway enable
   By default, mDNS gateway is disabled.
3. Enter VLAN view.
   vlan *vlan-id*
4. Enable mDNS gateway periodic probing.
   mdns gateway probe interval *interval*
   By default, mDNS gateway periodic probing is disabled.
   If you execute this command multiple times, the most recent configuration takes effect.

Enabling mDNS gateway periodic probing in a VSI

1. Enter system view.
   system-view
2. Enable mDNS gateway.
   mdns gateway enable
   By default, mDNS gateway is disabled.
3. Enter VSI view.
   vsi *vsi-id*
4. Enable mDNS gateway periodic probing.
   mdns gateway probe interval *interval*
   By default, mDNS gateway periodic probing is disabled.
   If you execute this command multiple times, the most recent configuration takes effect.

Enabling mDNS gateway periodic probing in a mDNS gateway policy

1. Enter system view.
   system-view
2. Enable mDNS gateway.
   mdns gateway enable
   By default, mDNS gateway is disabled.
3. Enter mDNS gateway policy view.
   mdns gateway policy *policy-name*
4. Enable mDNS gateway periodic probing.
   mdns gateway probe interval *interval*
   By default, mDNS gateway periodic probing is disabled.
   If you execute this command multiple times, the most recent configuration takes effect.

Enabling mDNS gateway proxy forwarding

About this task

By default, mDNS gateway proxy forwarding is enabled. The mDNS gateway can send probe queries to discover mDNS service providers on the network. If you enable both mDNS gateway proxy forwarding and periodic probing, the mDNS gateway will send two probe queries. In this case, disable mDNS gateway proxy forwarding.

Restrictions and guidelines

As a best practice to control proxy forwarding in different VLANs or VSIs, configure multiple mDNS gateway policies and execute the **mdns gateway probe interval** command in the view of each mDNS gateway policy.

In QinQ scenarios or scenarios with a lot of services VLANs or service VSIs, enable periodic probing on the mDNS gateway and disable mDNS gateway proxy forwarding.

Procedure

1. Enter system view.
   system-view
2. Enable mDNS gateway policy view.
   mdns gateway policy *policy-name*
3. Enable mDNS gateway proxy forwarding.
   mdns gateway proxy-forward enable
   By default, mDNS gateway proxy forwarding **is enabled**.

Enabling the mDNS gateway to forward mDNS queries on the receiving port in a VLAN

About this task

Use this feature in scenarios where the mDNS gateway must receive and forward mDNS queries on the same port.

Procedure

1. Enter system view.
   system-view
2. Enable the mDNS gateway to forward mDNS queries on the receiving port in a VLAN.
   mdns gateway permit source-port-forward enable
   By default, the mDNS gateway does not forward mDNS queries on the receiving port in a VLAN.

Configuring the mDNS gateway to support QinQ

About this task

On a QinQ network, packets sent from both the client and server to the mDNS gateway carry two layers of VLAN tags. The inner VLAN tag (also called CVLAN tag) identifies the VLAN to which the frame belongs when it is transmitted in the customer network. The outer VLAN tag (also called SVLAN tag) identifies the VLAN to which the QinQ frame belongs when it is transmitted in the service provider network. The mDNS gateway identifies the two layers of VLAN tags and implements service discovery as follows:

1. You specify service VLANs and users VLANs for an mDNS gateway policy and specify inner VLANs and outer VLANs.
2. Upon receiving an mDNS query, the mDNS gateway identifies whether the outer and inner VLAN tags of the packet matches the mDNS gateway policy.
   - If a match is found, the mDNS gateway forwards the query to the corresponding service VLANs in the mDNS gateway policy.
   - If no match is found, the mDNS gateway drops the query.

Restrictions and guidelines

In a QinQ scenario, to avoid any impact on the device service networking capabilities due to too many mDNS multicast packets sent within a short period, configure the inner VLANs as accurately as possible.

Procedure

1. Enter system view.
   system-view
2. Enable mDNS gateway.
   mdns gateway enable
   By default, mDNS gateway is disabled.
3. (Optional.) Specify the source address for mDNS packets.
   mdns gateway source ip { **interface** *interface-type interface-number* | *ip-address* }
   By default, the mDNS gateway uses the primary IP address of the VLAN or VSI interface for packet output as the source IP address of mDNS packets.
4. Enter mDNS gateway policy view.
   mdns gateway policy *policy-name*
5. Specify service VLANs for the mDNS gateway policy.
   service vlan *vlan-id-list* **second-vlan** *vlan-id-list*
   By default, no service VLAN is specified for an mDNS gateway policy.
6. Specify user VLANs for the mDNS gateway policy.
   user vlan *vlan-id-list* **second-vlan** *vlan-id-list*

By default, no user VLAN is specified for an mDNS gateway policy.

Display and maintenance commands for mDNS relay

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display mDNS gateway configuration.	**display mdns gateway**
Display mDNS gateway policy information.	**display mdns gateway policy** [*policy-name*]
Display information about services learned by the mDNS gateway.	**display mdns gateway service** { **all** | *predefined-service-type* | **user-defined** *user-defined-service-type* | **vlan** *vlan-id* [**second-vlan** *vlan-id*] | **vsi** *vsi-name* } [**verbose**]
Display mDNS gateway packet statistics.	**display mdns gateway statistics**
Clear information about services learned by the mDNS gateway.	**reset mdns gateway service** { **all** | *predefined-service-type* | **user-defined** *user-defined-service-type* | **vlan** *vlan-id* | **vsi** *vsi-name* }
Clear mDNS gateway packet statistics.	**reset mdns gateway statistics**

mDNS gateway configuration examples

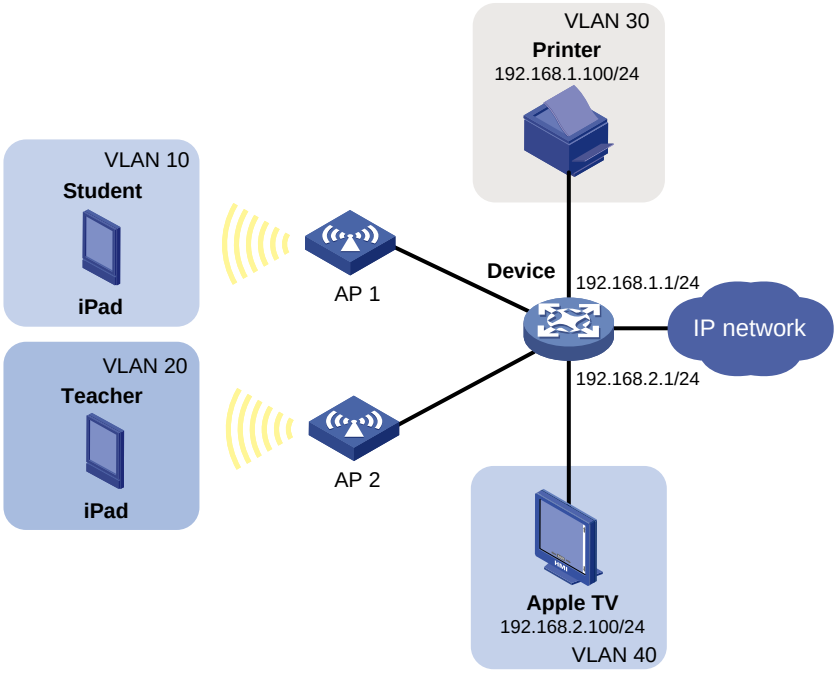
Example: Configuring mDNS gateway

Network configuration

As shown in Figure 4, a campus network uses mDNS to provide services to both teachers and students. Configure mDNS gateway to allow students to access only the printer and allow teachers to access both the Apple TV and the printer.

Enable mDNS gateway periodic probing for the gateway to update mDNS services in time.

Figure 4 Network diagram



Procedure

1. Enable mDNS gateway.
   ```
   <Device> system-view
   [Device] mdns gateway enable
   ```
2. Create an mDNS gateway policy named **student**.
   ```
   [Device] mdns gateway policy student
   ```
3. Specify VLAN 10 as a user VLAN and VLAN 30 as a service VLAN.
   ```
   [Device-mdns-gateway-policy-student] user vlan 10
   [Device-mdns-gateway-policy-student] service vlan 30
   [Device-mdns-gateway-policy-student] quit
   ```
4. Create an mDNS gateway policy named **teacher**.
   ```
   [Device] mdns gateway policy teacher
   ```
5. Specify VLAN 20 as a user VLAN and VLANs 30 and 40 as service VLANs.
   ```
   [Device-mdns-gateway-policy-teacher] user vlan 20
   [Device-mdns-gateway-policy-teacher] service vlan 30 40
   [Device-mdns-gateway-policy-teacher] quit
   ```
6. Permit service types **airplay**, **raop**, and **printer** in entry learning.
   ```
   [Device] mdns gateway permit service-type airplay
   [Device] mdns gateway permit service-type raop
   [Device] mdns gateway permit service-type printer
   ```
7. Specify the source address for mDNS packets in VLAN 10.
   ```
   [Device] vlan 10
   [Device-vlan10] mdns gateway source ip 192.168.1.1
   [Device-vlan10] quit
   ```
8. Specify the source address for mDNS packets in VLAN 20.
   ```
   [Device] vlan 20
   [Device-vlan20] mdns gateway source ip 192.168.2.1
   ```

```
[Device-vlan20] quit
```

9. Enable mDNS gateway periodic probing in VLAN 30 and set the probing interval to 600 seconds.

```
[Device] vlan 30
[Device-vlan30] mdns gateway probe interval 600
[Device-vlan30] quit
```

10. Enable mDNS gateway periodic probing in VLAN 40 and set the probing interval to 600 seconds.

```
[Device] vlan 40
[Device-vlan40] mdns gateway probe interval 600
[Device-vlan40] quit
```

Verifying the configuration

Display mDNS gateway configuration.

```
[Device] display mdns gateway
mDNS gateway status: Enabled
mDNS gateway source IP:
  VLAN 10: 192.168.1.1
  VLAN 20: 192.168.2.1
mDNS gateway probe is enabled in the following VLANs: 30, 40
mDNS gateway probe is enabled in the following VSIs:
mDNS gateway probe is enabled in the following Policys:
```

Display information about services learned by the mDNS gateway.

```
[Device] display mdns gateway service all
Instance name        Service type VLAN/VSI name TTL      Pro IP
Apple TV             airplay      40           120       TCP
192.168.2.100
Apple TV             raop         40           120       TCP 192.168.2.100
Printer              printer      30           120       TCP
192.168.1.100

Total mDNS gateway services: 3
```

Verify that students can access only the printer and teachers can access both the Apple TV and the printer.

Troubleshooting mDNS gateway

Failure to receive packets from an mDNS relay

Symptom

The mDNS gateway cannot receive mDNS packets from an mDNS relay.

Analysis

Possible reasons are:

- The mDNS gateway feature is not enabled.
- Trusted mDNS devices are configured, but the mDNS relay is not a trusted device.

Solution

To resolve this issue:

1. Execute the **display mdns gateway** command to view current configuration. If mDNS gateway is not enabled, execute the **mdns gateway enable** command to enable this feature.
2. Execute the **display current-configuration** command to view trusted mDNS devices. If the mDNS relay is not in the list, execute the **mdns gateway trust device-ip** command to specify the mDNS relay as a trusted device.

Contents

Configuring mDNS relay 1

About mDNS relay 1
mDNS network device types 1
mDNS network model 1
mDNS relay functionalities 2
Relaying mDNS service advertisements 2
Relaying mDNS queries 3
Periodic probing 4
Protocols and standards 4
mDNS relay tasks at a glance 4
Specifying the mDNS gateway address 5
Enabling mDNS relay 5
Restrictions and guidelines for enabling mDNS relay 5
Enabling mDNS relay for a VLAN 5
Enabling mDNS relay for a VSI 5
Specifying the source IP address for mDNS packets 6
About specifying source IP address for mDNS packets 6
Restrictions and guidelines for specifying the source IP address for mDNS packets 6
Specifying the source IP address for mDNS packets in a VLAN 6
Specifying the source IP address for mDNS packets in a VSI 6
Enabling mDNS relay periodic probing 7
About mDNS relay periodic probing 7
Enabling mDNS relay periodic probing in a VLAN 7
Enabling mDNS relay periodic probing in a VSI 7
Display and maintenance commands for mDNS relay 7
mDNS relay configuration examples 8
Example: Configuring mDNS relay 8
Troubleshooting mDNS relay 9
mDNS packet receiving failure 9

Configuring mDNS relay

About mDNS relay

The multicast Domain Name Service (mDNS) protocol is a zero configuration networking protocol for mDNS service discovery. In a LAN network, endpoint devices can automatically discover available services advertised by mDNS service providers without knowing these service providers.

The mDNS relay supports mDNS packet transmission across subnets and expands the use of the mDNS protocol to large-scale networks. Without the mDNS relay, mDNS packets can only be transmitted within a subnet.

mDNS network device types

An mDNS network contains the following device types:

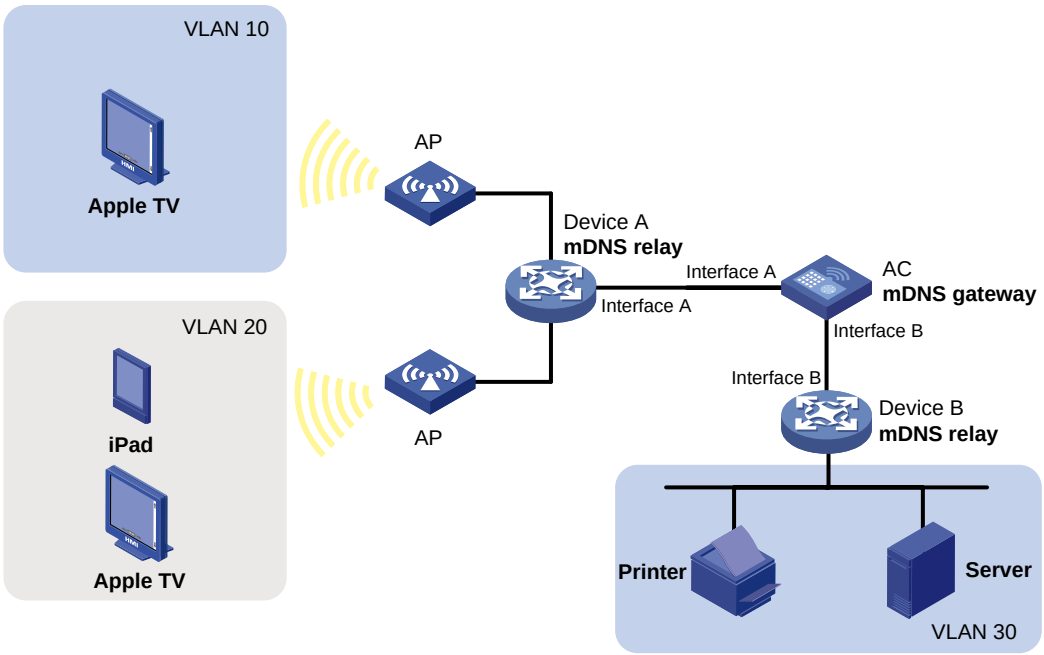
- **mDNS endpoint device**—Obtains services from mDNS service providers. Examples are iPhones and iPads.
- **mDNS service provider**—Provides services for mDNS endpoint devices. Examples are printers and video servers.
- **mDNS gateway**—Records available services advertised by mDNS service providers and responds to queries from mDNS endpoint devices. For more information about the mDNS gateway, see "Configuring mDNS gateway."
- **mDNS relay**—Ensures successful mDNS communication when mDNS endpoint devices and mDNS service providers are in different subnets. The mDNS relay supports periodically probing mDNS service providers in the network.

mDNS network model

As shown in Figure 1, mDNS endpoint devices (Apple TVs and iPad) and mDNS service providers (printer and server) are in different VLANs. APs and the AC require inter-subnet Layer 3 communication. The AC acts as the mDNS gateway and cannot receive multicast mDNS packets sent from the endpoint devices and service providers.

To resolve this problem, enable mDNS relay on Device A and Device B and specify the mDNS gateway address on them. After receiving an mDNS packet from an endpoint device, Device A changes the packet destination IP address to the mDNS gateway address and unicasts it to the gateway. After receiving an mDNS packet from an mDNS service provider, Device B changes the packet destination IP address to the mDNS gateway address and unicasts it to the gateway.

Figure 1 mDNS network model



mDNS relay functionalities

The mDNS relay provides the following functionalities:

- **Relaying mDNS service advertisements**—After receiving advertisements from an mDNS service provider, the relay changes the packet destination IP address to the mDNS gateway address and unicasts them to the gateway.
- **Relaying mDNS queries**—After receiving mDNS queries from endpoint devices, the relay changes the query destination IP address to the mDNS gateway address and unicasts them to the gateway.
- **Periodic probing**—The mDNS relay supports periodically sending mDNS queries to probe mDNS service providers in a subnet and forwarding the responses to the mDNS gateway for service update.

Relaying mDNS service advertisements

An mDNS service provider advertises its service to the mDNS gateway when it is attached to the network. The host name identifies the mDNS service provider and the service name identifies service type and other information provided by the service provider.

As shown in Figure 2, an mDNS service provider probes and advertises the host name and service name in the following workflow:

1. Host name probe.
   a. The mDNS service provider automatically generates a host name after it is attached to the network. To detect the uniqueness of the host name, the service provider sends an mDNS query with multicast destination address 224.0.0.251 and starts the probe timer.
   b. After receiving the query, the mDNS relay replaces the destination IP address with the mDNS gateway address and unicasts it to the mDNS gateway.
   c. The mDNS gateway examines whether the host name conflicts with host names of existing mDNS service entries. If a conflict exists, the mDNS gateway sends the mDNS relay an mDNS response to announce the conflict. The mDNS relay changes the destination IP

address of the response to multicast address 224.0.0.251 and forwards it to the mDNS service provider.

d. After receiving the conflict announcement, the mDNS service provider generates another host name and starts probing again.

2. Host name advertisement.

a. The mDNS service provider advertises its host name if it does not receive any host name conflict announcement within the probe period. The destination of the advertisement packet is 224.0.0.251.

b. After receiving the packet, the mDNS relay changes the packet destination IP address to the mDNS gateway address and unicasts it to the mDNS gateway.

c. The mDNS gateway records the host name and IP address in the received packet.

3. Service name probe.

a. The mDNS service provider sends an mDNS query with destination address 224.0.0.251 to detect the uniqueness of the service name and starts the probe timer.

b. After receiving the query, the mDNS relay replaces the destination IP address with the mDNS gateway address and unicasts it to the mDNS gateway.

c. The mDNS gateway examines whether the service name conflicts with service names of existing mDNS service entries. If a conflict exists, the mDNS gateway sends the mDNS relay an mDNS response to announce the conflict. The mDNS relay changes the destination IP address of the response to multicast address 224.0.0.251 and forwards it to the mDNS service provider.

d. After receiving the conflict announcement, the mDNS service provider generates another service name and starts probing again.

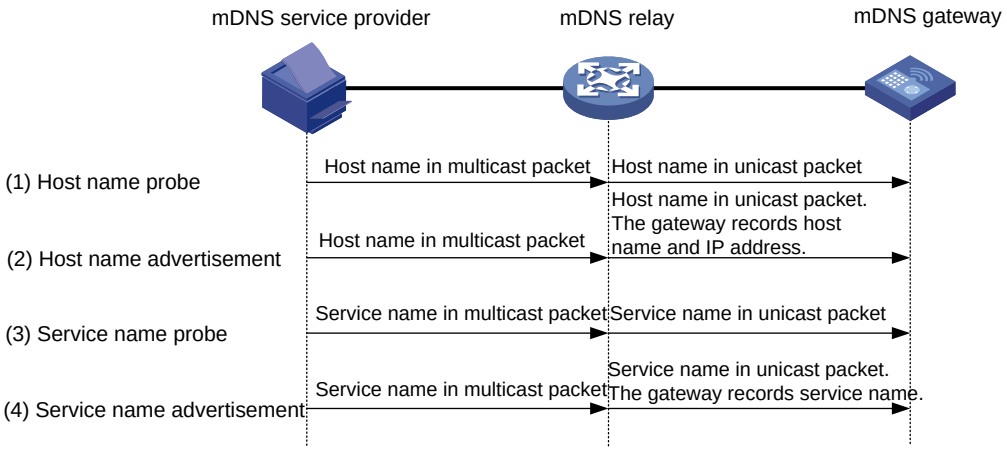
4. Service name advertisement.

a. The mDNS service provider advertises its service name if it does not receive any service name conflict announcement within the probe period. The destination of the advertisement packet is 224.0.0.251.

b. After receiving the packet, the mDNS relay changes the packet destination IP address to the mDNS gateway address and unicasts it to the mDNS gateway.

c. The mDNS gateway records the service name information in the received packet.

Figure 2 Relaying mDNS service advertisements

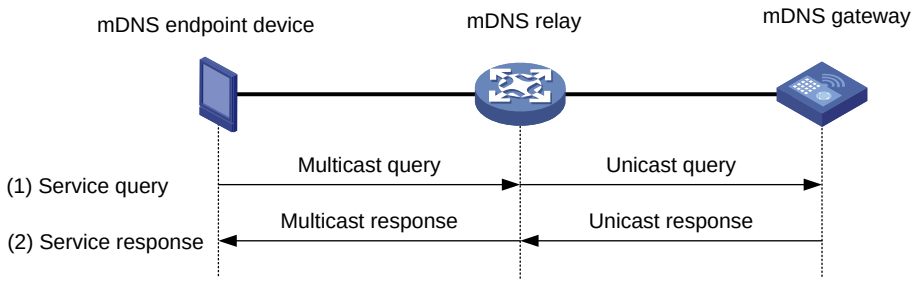


Relaying mDNS queries

As shown in Figure 3, an mDNS query originated from the mDNS endpoint device is relayed as follows:

1. Before accessing a service, the mDNS endpoint device first sends an mDNS query with multicast address 224.0.0.251 to detect whether this service is available.
2. Upon receiving the query, the mDNS relay changes the query destination IP address to the mDNS gateway address and unicasts the query to the gateway.
3. The mDNS gateway looks for a matching service after receiving the query. If the matching service exists, the mDNS gateway encapsulates the host name and IP address of the service provider into the response and unicasts it to the mDNS relay.
4. The mDNS relay changes the response destination IP address to 224.0.0.251 and multicasts the response.
5. After receiving the response, the mDNS endpoint device can access the service provider based on the IP address in the response. Other mDNS endpoints that receive the response will also record information about the service provider for future use.

Figure 3 Relaying mDNS queries



Periodic probing

When an mDNS service provider connects to the network, it unsolicitly sends an advertisement to the mDNS gateway. The mDNS gateway creates a service entry based on the advertisement information. When a new mDNS gateway joins the network where the mDNS service providers already exist, the mDNS service providers do not send advertisements. Therefore, the new mDNS gateway does not know existing mDNS services in the network.

The mDNS gateway creates and maintains service entries for mDNS services. If the mDNS gateway receives an update message for a service before the service entry ages out, the mDNS gateway resets the service entry timer. If the mDNS gateway does not receives any update message for a service before the service entry ages out, the mDNS gateway deletes the entry. In a large-sized network where mDNS service providers exists in many VLANs or VSIs, the mDNS gateway cannot receive update messages timely and might delete service entries mistakenly.

To resolve these issues, enable periodic probing on the mDNS relay and set a probe interval. The mDNS relay sends mDNS queries at intervals to VLANs and VSIs where mDNS relay is enabled. Each mDNS service provider that receives the query sends back an mDNS response. The mDNS relay then unicasts the response to the mDNS gateway, so that the mDNS gateway can quickly learn the latest services provided by all mDNS service providers.

Protocols and standards

RFC 6762, *Multicast DNS*

mDNS relay tasks at a glance

To configure the mDNS relay, perform the following tasks:

1. Specifying the mDNS gateway address

2. Enabling mDNS relay
3. (Optional.) Specifying the source IP address for mDNS packets
4. (Optional.) Enabling mDNS relay periodic probing

Specifying the mDNS gateway address

About this task

After receiving an mDNS query from an mDNS endpoint device, the mDNS relay replaces the multicast destination address with the specified mDNS gateway address and unicasts the query to the mDNS gateway. When receiving the response from the gateway, the mDNS relay multicasts it in the VLAN or VSI where the requesting mDNS endpoint resides. The requesting endpoint device can then receive the response and other mDNS endpoint devices can learn and update the service information.

Procedure

1. Enter system view.
   system-view
2. Specify the IP address of the mDNS gateway.
   mdns relay gateway ip *ip-address* [**vpn-instance** *vpn-instance-name*]
   By default, no mDNS gateway address is specified on the mDNS relay.

Enabling mDNS relay

Restrictions and guidelines for enabling mDNS relay

The mDNS relay feature takes effect only after you specify the mDNS gateway address on the mDNS relay.

Enabling mDNS relay for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable mDNS relay.
   mdns relay enable
   By default, mDNS relay is disabled.

Enabling mDNS relay for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable mDNS relay.
   mdns relay enable

By default, mDNS relay is disabled.

Specifying the source IP address for mDNS packets

About specifying source IP address for mDNS packets

For responses destined for endpoint devices and probe queries sent within a VLAN or VSI, the default source IP address is the IP address of the VLAN interface or VSI interface. If no IP address is configured for the VLAN interface or VSI interface, the mDNS relay cannot send out the queries.

If the mDNS relay is connected to a large number of VLANs or VSIs, configuring an IP address to each VLAN interface or VSI interface wastes IP address resources. You can use this command to specify a specific IP address for the mDNS relay to use.

Restrictions and guidelines for specifying the source IP address for mDNS packets

Make sure the specified source address is on the relay and has a reachable route to the mDNS gateway. If these requirements are not met, the mDNS relay cannot receive responses for the queries or forwarded packets.

As a best practice, specify the interface connecting to the mDNS gateway in the **mdns relay source ip** command.

Specifying the source IP address for mDNS packets in a VLAN

1. Enter system view.

   system-view
2. Enter VLAN view.

   vlan *vlan-id*
3. Specify source IP address for mDNS packets.

   mdns relay source ip { *ip-address* | **interface** *interface-type interface-number* }

   By default, the mDNS relay encapsulates source IP addresses for packets as follows:

   - For responses destined for endpoint devices and VLAN-specific probe queries, the default source IP address is the IP address of the VLAN interface.
   - For multicast packets destined for the gateway, the default source IP address is the IP address of the packet output interface on the mDNS relay.

Specifying the source IP address for mDNS packets in a VSI

1. Enter system view.

   system-view
2. Enter VSI view.

   vsi *vsi-name*
3. Specify source IP address for mDNS packets.

mdns relay source ip { *ip-address* | **interface** *interface-type interface-number* }

By default, the mDNS relay encapsulates source IP addresses for packets as follows:

- For responses destined for endpoint devices and VSI-specific probe queries, the default source IP address is the IP address of the VSI interface.
- For multicast packets destined for the gateway, the default source IP address is the IP address of the packet output interface on the mDNS relay.

Enabling mDNS relay periodic probing

About mDNS relay periodic probing

This feature enables the mDNS relay to send mDNS probe queries within an mDNS relay-enabled VLAN or VSI at intervals. Each mDNS service provider sends a response to the mDNS relay after receiving the query. The mDNS relay then unicasts the responses to the mDNS gateway. The mDNS gateway can learn the latest service information of all mDNS service providers in the network in a timely manner.

Enabling mDNS relay periodic probing in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable mDNS relay periodic probing.
   mdns relay probe interval *interval*
   By default, mDNS relay periodic probing is disabled.

Enabling mDNS relay periodic probing in a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable mDNS relay periodic probing.
   mdns relay probe interval *interval*
   By default, mDNS relay periodic probing is disabled.

Display and maintenance commands for mDNS relay

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display mDNS relay configuration.	**display mdns relay**

Task	Command
Display packet statistics on the mDNS relay.	**display mdns relay statistics**
Clear packet statistics on the mDNS relay.	**reset mdns relay statistics**

mDNS relay configuration examples

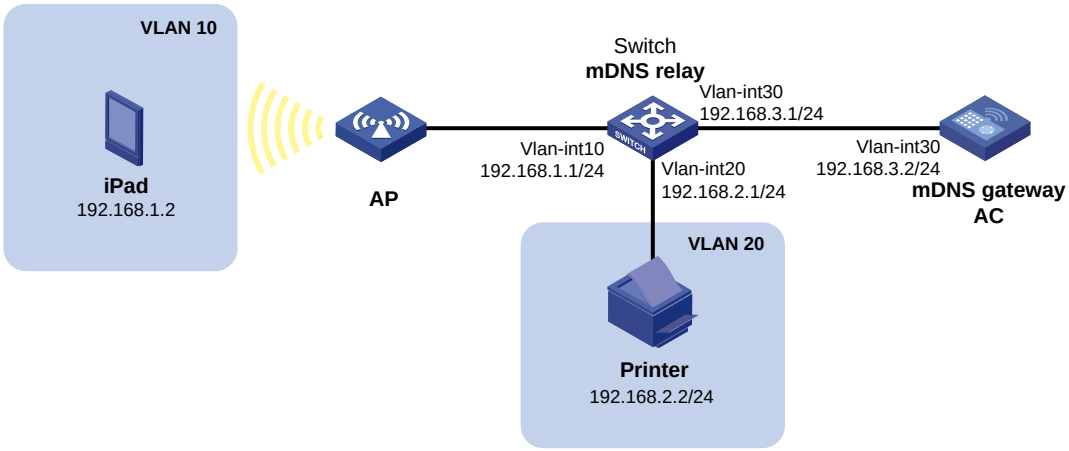
Example: Configuring mDNS relay

Network configuration

As shown in Figure 4, an enterprise network uses mDNS for communication. The working area is in VLAN 10 and uses subnet 192.168.1.0. The service area is in VLAN 20 and uses subnet 192.168.2.0. Subnet 192.168.3.0 runs between the mDNS relay and the mDNS gateway.

Configure mDNS relay on the switch to allow employers in the working area to access the printer in the service area.

Figure 4 Network diagram



Procedure

1. Configure the mDNS gateway and make sure route among the AP, the mDNS relay, and the mDNS gateway is reachable. (Details not shown.)
2. Configure the mDNS relay.

   # Specify 192.168.3.2 as the mDNS gateway address.

   ```
   <Switch> system-view
   [Switch] mdns relay gateway ip 192.168.3.2
   ```

   # Enable mDNS relay in VLAN 10.

   ```
   [Switch] vlan 10
   [Switch-vlan10] mdns relay enable
   ```

   # Specify 192.168.3.1 as the source IP address for mDNS packets.

   ```
   [Switch-vlan10] mdns relay source ip 192.168.3.1
   [Switch-vlan10] quit
   ```

   # Enable mDNS relay in VLAN 20.

   ```
   [Switch] vlan 20
   ```

```
[Switch-vlan20] mdns relay enable
```

Specify 192.168.3.1 as the source IP address for mDNS packets.

```
[Switch-vlan20] mdns relay source ip 192.168.3.1
[Switch-vlan20] quit
```

Verifying the configuration

Display mDNS relay configuration.

```
[Switch] display mdns relay
mDNS relay configuration:
mDNS gateway IP: 192.168.3.2
mDNS relay configuration in VLANs:
VLAN ID  Source IP address           Probe interval (sec)
10       192.168.3.1                --
20       192.168.3.1                --
```

Verify that the user in VLAN 10 can use the iPad to access the printer in VLAN 20. (Details not shown.)

Troubleshooting mDNS relay

mDNS packet receiving failure

Symptom

The mDNS relay cannot receive mDNS packets.

Analysis

Possible reasons are:

- The mDNS relay is not enabled.
- The IP address of the mDNS gateway is not specified on the mDNS relay.

Solution

To resolve this issue:

1. Execute the **`display mdns relay`** command to view current configuration.
2. Perform one of the following tasks:
   - If mDNS relay is not enabled, execute the **`mdns relay enable`** command to enable this feature.
   - If the IP address of the mDNS gateway is not specified on the mDNS relay, execute the **`mdns relay gateway ip`** command to specify the mDNS gateway address.

Contents

Configuring IP forwarding basic settings 2
About FIB table 2
Saving the IP forwarding entries to a file 3
Enabling split horizon forwarding 3
Enabling IPv4 fast traffic migration 3
Forwarding specific packets received from VXLAN tunnels in hardware 4
Enabling IPv4 FIB entry consistency check 4
Enabling SNMP notifications for FIB events 4
Enabling SNMP notifications for IP forwarding events 5
Enabling FIB logging 5
Display and maintenance commands for FIB table 6
Configuring load sharing 6
About load sharing 6
Configuring load sharing mode 6
Enabling local-first load sharing 7
Enabling symmetric load sharing 7
Display and maintenance commands for load sharing 8
Loading sharing configuration examples 8
Example: Configuring load sharing based on source and destination addresses 8

Configuring IP forwarding basic settings

About FIB table

A device uses the FIB table to make packet forwarding decisions.

A device selects optimal routes from the routing table, and puts them into the FIB table. Each FIB entry specifies the next hop IP address and output interface for packets destined for a specific subnet or host.

For more information about the routing table, see *Layer 3—IP Routing Configuration Guide*.

Use the **display fib** command to display the FIB table. The following example displays the entire FIB table.

```
<Sysname> display fib

Destination count: 8 FIB entry count: 8

Flag:
  U:Usable   G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR

Destination/Mask   Nexthop         Flag     OutInterface/Token       Label
0.0.0.0/32         127.0.0.1       UH       InLoop0                  Null
127.0.0.0/8        127.0.0.1       U        InLoop0                  Null
127.0.0.0/32       127.0.0.1       UH       InLoop0                  Null
127.0.0.1/32       127.0.0.1       UH       InLoop0                  Null
127.255.255.255/32 127.0.0.1       UH       InLoop0                  Null
224.0.0.0/4        0.0.0.0         UB       NULL0                    Null
224.0.0.0/24       0.0.0.0         UB       NULL0                    Null
255.255.255.255/32 127.0.0.1       UH       InLoop0                  Null
```

A FIB entry includes the following items:

- **Destination**—Destination IP address.
- **Mask**—Network mask. The mask and the destination address identify the destination network. A logical AND operation between the destination address and the network mask yields the address of the destination network. For example, if the destination address is 192.168.1.40 and the mask 255.255.255.0, the address of the destination network is 192.168.1.0. A network mask includes a certain number of consecutive 1s. It can be expressed in dotted decimal format or by the number of the 1s.
- **Nexthop**—IP address of the next hop.
- **Flag**—Route flag.
- **OutInterface**—Output interface.
- **Token**—MPLS Label Switched Path index number.
- **Label**—Inner label.

Saving the IP forwarding entries to a file

Restrictions and guidelines

The feature automatically creates the file if you specify a nonexistent file. If the file already exists, this feature overwrites the file content.

This feature triggers one-time saving of the IP forwarding entries.

To automatically save the IP forwarding entries periodically, configure a schedule for the device to automatically run the **ip forwarding-table save** command. For information about scheduling a task, see *Fundamentals Configuration Guide*.

Procedure

To save the IP forwarding entries to a file, execute the following command in any view:

ip forwarding-table save filename *filename*

Enabling split horizon forwarding

About this task

This feature prevents IPv4, IPv6, and MPLS packets from being forwarded out of the physical interface on which they were received, avoiding network loops.

Procedure (System view)

1. Enter system view.

   system-view
2. Enable split horizon forwarding.

   forwarding split-horizon

   By default, split horizon forwarding is disabled.

Enabling IPv4 fast traffic migration

About this task

After IPv4 fast traffic migration is enabled on a VLAN interface, the device immediately steers IPv4 traffic forwarded through the VLAN interface to other forwarding paths upon network topology changes or user port migrations. This reduces traffic disruption time to improve overall network response speed and user experience.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable IPv4 fast traffic migration.

   ip forwarding fast-move

   By default, IPv4 fast traffic migration is disabled.

Forwarding specific packets received from VXLAN tunnels in hardware

About this task

By default, the device forwards packets received from VXLAN tunnels to the CPU for processing when acting as a VTEP in a distributed EVPN gateway network. If a large number of packets are received, packet loss might occur because of software rate limit, which might cause service exceptions on downlink devices.

To resolve this issue, you can enable the device to forward specific packets received from VXLAN tunnels in hardware without delivering them to the CPU.

Procedure

1. Enter system view.
   system-view
2. Enable hardware forwarding for specific packets received from VXLAN tunnels.
   forwarding vxlan-packet inner-protocol { **ipv4** | **ipv6** } *
   By default, packets received from VXLAN tunnels are delivered to the CPU for processing.

Enabling IPv4 FIB entry consistency check

About this task

Packet drops or incorrect forwarding might occur when the IPv4 FIB entries in hardware are inconsistent with FIB entries configured in software. To prevent these issues, enable IPv4 FIB entry consistency check.

This feature compares all IPv4 FIB entries in software with the IPv4 FIB entries in hardware regularly. If the device detects an inconsistency, the device performs the following tasks:

- Generates a log.
- Updates the IPv4 FIB entry in hardware with the IPv4 FIB entry in software.

Procedure

1. Enter system view.
   system-view
2. Enable IPv4 FIB entry consistency check.
   fib consistency-check enable
   By default, IPv4 FIB entry consistency check is disabled.

Enabling SNMP notifications for FIB events

About this task

This feature enables the FIB module to generate SNMP notifications for critical FIB events, such as the exceeding of the message queue length threshold. The SNMP notifications are sent to the SNMP module.

For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for FIB events.

   snmp-agent trap enable fib [**deliver-failed** | **ecmp-limit** | **entry-consistency** | **entry-limit**] *

   By default, SNMP notifications for FIB events are enabled.

Enabling SNMP notifications for IP forwarding events

About this task

This feature enables the IP forwarding module to generate SNMP notifications for critical IP forwarding events. The SNMP notifications are sent to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

You can enable SNMP notifications for the following IP forwarding events as needed:

- After you enable SNMP notifications about TTL timeout, the device checks the number of dropped packets at the specified time interval. An SNMP notification is generated and sent to the SNMP module when the number of dropped packets reaches or exceeds the specified threshold.
- After you enable SNMP notifications about MBUF allocation failures, an SNMP notification is generated and sent to the SNMP module when MBUF allocation fails.

If you do not specify any parameters, the command enables all types of SNMP notifications for IP forwarding events.

After you disable the SNMP notifications for IP forwarding events, the device sends only logs to the device's information center module. In this case, you can configure the log output destinations and output rules to view the log information of the basic IPv6 modules. For more information about configuring the information center, see information center configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for IP forwarding events.

   snmp-agent trap enable ip-forwarding [**ttl-expired** | **mbuf-alloc**]

   By default, SNMP notifications for IP forwarding events are enabled.

Enabling FIB logging

About this task

The logs are sent to the information center of the device. For the logs to be output correctly, you must also configure the information center on the device. For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

To avoid memory consumption caused by log recording, you can use the **undo fib log enable** command to disable FIB logging.

Procedure

1. Enter system view.
   system-view
2. Enable FIB logging.
   fib log enable
   By default, FIB logging is disabled.

Display and maintenance commands for FIB table

Execute **display** commands in any view.

Task	Command
Display FIB entries.	**display fib** [**vpn-instance** *vpn-instance-name*] [*ip-address* [*mask* | *mask-length*]]
Display FIB table usage information.	**display fib usage**

Configuring load sharing

About load sharing

If a routing protocol finds multiple equal-cost best routes to the same destination, the device forwards packets over the equal-cost routes to implement load sharing.

Configuring load sharing mode

About this task

Restrictions and guidelines

- When the traffic is not shared equally, you can use the **seed** and **shift** keywords to improve load sharing.
- The load sharing hash seed feature does not take effect on unknown unicast packets.
- The load sharing feature for IP Tunnel packets does not take effect on unknown unicast packets.
- If IPv6 packets are encapsulated by MPLS, they do not support flow label-based per-flow load sharing.

Procedure

1. Enter system view.
   system-view
2. Configure load sharing.
   ip load-sharing mode { **per-flow** [**algorithm** *algorithm-number* [**seed** *seed-number*] [**shift** *shift-number*] \| [**dest-ip** \| **dest-port** \| **flow-label** \| **ingress-port** \| **ip-pro** \| **src-ip** \| **src-port**] * \| **tunnel** { **inner** \| **outer** }] \| **per-packet** } { **global** \| **slot** *slot-number* }

By default, the device performs per-flow load sharing based on the following criteria: source IP address, destination IP address, source port number, destination port number, IP protocol number, and ingress port.

3. Display the load sharing path selected for a flow.

   display ip load-sharing path ingress-port *interface-type interface-number* **packet-format** { **ipv4oe dest-ip** *ip-address* [**src-ip** *ip-address*] | **ipv6oe dest-ipv6** *ipv6-address* [**src-ipv6** *ipv6-address* | **flow-label** *flow-label*] } [**dest-port** *port-id* | **ip-pro** *protocol-id* | **src-port** *port-id* | **vpn-instance** *vpn-instance-name*] *

   The option settings in this command must match both the options displayed in the **display ip load-sharing mode** command and the field values in load shared packets. If the option settings do not meet the requirement, the path displayed by this command might be different from the real path for load sharing.

Enabling local-first load sharing

About this task

Local-first load sharing distributes traffic preferentially across the output interfaces on the receiving IRF member device if output interfaces for multiple equal-cost routes are on different members. This feature enhances packets forwarding efficiency.

Procedure

1. Enter system view.

   system-view

2. Enable local-first load sharing.

   ip load-sharing local-first enable

   By default, local-first load sharing is enabled.

Enabling symmetric load sharing

About this task

Symmetric load sharing ensures that bidirectional traffic specific to a source and destination address pair flow along the same path.

Restrictions and guidelines

Symmetric load sharing takes effect only on known unicast packets.

Procedure

1. Enter system view.

   system-view

2. Enable symmetric load sharing.

   ip load-sharing symmetric enable

   By default, symmetric load sharing is disabled.

Display and maintenance commands for load sharing

Execute **display** commands in any view.

Task	Command
Display the load sharing mode in use.	**display ip load-sharing mode slot** *slot-number*
Display the load sharing path selected for a flow.	**display ip load-sharing path ingress-port** *interface-type interface-number* **packet-format** { **ipv4oe dest-ip** *ip-address* [**src-ip** *ip-address*] | **ipv6oe dest-ipv6** *ipv6-address* [**src-ipv6** *ipv6-address* | **flow-label** *flow-label*] } [**dest-port** *port-id* | **ip-pro** *protocol-id* | **src-port** *port-id* | **vpn-instance** *vpn-instance-name*] *

Loading sharing configuration examples

Example: Configuring load sharing based on source and destination addresses

Network configuration

As shown in Figure 1, Switch A has two equal-cost routes to Switch B. Configure load sharing on Switch A to forward packets through Switch B to the destination IP address 1.2.3.4/24.

Figure 1 Network diagram

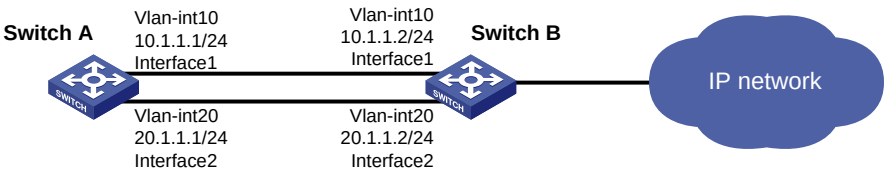


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

On Switch A, assign GigabitEthernet 1/0/1 to VLAN 10, and GigabitEthernet 1/0/2 to VLAN 20.

```
<SwitchA> system-view
[SwitchA] vlan 10
[SwitchA-vlan10] port gigabitethernet 1/0/1
[SwitchA-vlan10] quit
[SwitchA] vlan 20
```

```
[SwitchA-vlan20] port gigabitethernet 1/0/2
[SwitchA-vlan20] quit
```

On Switch A, configure IP addresses for VLAN-interface 10 and VLAN-interface 20.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ip address 10.1.1.1 24
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] ip address 20.1.1.1 24
[SwitchA-Vlan-interface20] quit
```

On Switch B, assign GigabitEthernet 1/0/1 to VLAN 10, and GigabitEthernet 1/0/2 to VLAN 20.

```
<SwitchB> system-view
[SwitchB] vlan 10
[SwitchB-vlan10] port gigabitethernet 1/0/1
[SwitchB-vlan10] quit
[SwitchB] vlan 20
[SwitchB-vlan20] port gigabitethernet 1/0/2
[SwitchB-vlan20] quit
```

On Switch B, configure IP addresses for VLAN-interface 10 and VLAN-interface 20.

```
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ip address 10.1.1.2 24
[SwitchB-Vlan-interface10] quit
[SwitchB] interface vlan-interface 20
[SwitchB-Vlan-interface20] ip address 20.1.1.2 24
[SwitchB-Vlan-interface20] quit
```

On Switch A, configure two static routes to the destination IP address.

```
<SwitchA> system-view
[SwitchA] ip route-static 1.2.3.4 24 10.1.1.2
[SwitchA] ip route-static 1.2.3.4 24 20.1.1.2
[SwitchA] quit
```

On Switch A, display FIB entries matching the destination IP address 1.2.3.4.

```
<SwitchA> display fib 1.2.3.4
Destination count: 1 FIB entry count: 2
Flag:
  U:Usable   G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination/Mask   Nexthop         Flag     OutInterface/Token       Label
1.2.3.0/24         10.1.1.2        USGR     Vlan10                   Null
1.2.3.0/24         20.1.1.2        USGR     Vlan20                   Null
```

On Switch A, configure per-flow load sharing based on the source IP address and destination IP address.

```
<SwitchA> system-view
[SwitchA] ip load-sharing mode per-flow dest-ip src-ip global
```

Verifying the configuration

Verify that Switch A implements load sharing.

```
<SwitchA> display counters outbound interface GigabitEthernet
Interface         Total (pkts)   Broadcast (pkts)   Multicast (pkts)  Err (pkts)
GE1/0/1                 1045                  0                  0           0
```

GE1/0/2	1044	0	0	0

Contents

Configuring fast forwarding 1
About fast forwarding 1
Restrictions and guidelines: Fast forwarding configuration 1
Configuring the aging time for software fast forwarding entries 1
Configuring fast forwarding load sharing for software 1
Display and maintenance commands for fast forwarding 2

Configuring fast forwarding

About fast forwarding

Fast forwarding reduces route lookup time and improves packet forwarding efficiency by using a high-speed cache and data-flow-based technology. It identifies a data flow by using the following fields: source IP address, source port number, destination IP address, destination port number, and protocol number. After a flow's first packet is forwarded through the routing table, fast forwarding creates an entry and uses the entry to forward subsequent packets of the flow.

Restrictions and guidelines: Fast forwarding configuration

Fast forwarding can process fragmented IP packets, but it does not fragment IP packets.

Configuring the aging time for software fast forwarding entries

About this task

The fast forwarding table uses an aging timer for each forwarding entry. If an entry is not updated before the timer expires, the device deletes the entry. If an entry has a hit within the aging time, the aging timer restarts.

Procedure

1. Enter system view.

   system-view
2. Configure the aging time for fast forwarding entries on the software side.

   ip fast-forwarding aging-time *aging-time*

   By default, the aging time is 30 seconds for fast forwarding entries on the software side.

Configuring fast forwarding load sharing for software

About this task

Software fast forwarding load sharing enables the device to identify a data flow by using the packet information.

If fast forwarding load sharing is disabled on the software side, the device identifies a data flow by the packet information and the input interface.

Procedure

1. Enter system view.

   system-view
2. Configure fast forwarding load sharing on the software side. Choose one option as needed:

- Enable fast forwarding load sharing on the software side.
  ip fast-forwarding load-sharing
- Disable fast forwarding load sharing on the software side.
  undo ip fast-forwarding load-sharing

By default, fast forwarding load sharing is enabled on the software side.

Display and maintenance commands for fast forwarding

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the aging time of fast forwarding entries.	**display ip fast-forwarding aging-time**
Display fast forwarding entries.	**display ip fast-forwarding cache** [*ip-address*] [**slot** *slot-number*]
Display fast forwarding entries about fragmented packets.	**display ip fast-forwarding fragcache** [*ip-address*] [**slot** *slot-number*]
Clear the fast forwarding table.	**reset ip fast-forwarding cache** [**slot** *slot-number*]

Contents

Displaying the adjacency table 1
About the adjacency table 1
Display and maintenance commands for adjacency table 1

Displaying the adjacency table

About the adjacency table

The adjacency table stores information about directly connected neighbors for IP forwarding. The neighbor information in this chapter refers to non-Ethernet neighbor information.

This table is not user configurable. The neighbor information is generated, updated, and deleted by link layer protocols through negotiation or through manual configuration. An adjacency entry includes the following information:

- Neighbor network layer address (next hop).
- Output interface.
- Link layer protocol type.
- Link layer address.

When forwarding an IP packet, the device performs the following tasks:

- Searches the FIB to find the output interface and next hop.
- Uses the output interface and next hop address to search the adjacency table for link layer forwarding information.

NOTE:

Ethernet and non-Ethernet neighbor information is stored and managed together.

Display and maintenance commands for adjacency table

NOTE:

The device only supports displaying adjacency entries about tunnel interfaces whose link layer protocol is P2P.

Execute **display** commands in any view.

Task	Command
Display IPv4 adjacency table information.	**display adjacent-table** { **all** | **physical-interface** *interface-type interface-number* | **routing-interface** *interface-type interface-number* | **slot** *slot-number* } [**count** | **verbose**]
Display IPv6 adjacency table information.	**display ipv6 adjacent-table** { **all** | **physical-interface** *interface-type interface-number* | **routing-interface** *interface-type interface-number* | **slot** *slot-number* } [**count** | **verbose**]

Contents

Configuring IRDP 1

About IRDP 1
IRDP operation 1
Protocols and standards 1
IRDP tasks at a glance 1
Restrictions and guidelines: IRDP configuration 2
Enabling IRDP 2
Specifying the preference of advertised IP addresses 2
Setting the lifetime of advertised IP addresses 2
Setting the advertising intervals 3
Specifying the multicast address as the destination address for RAs 3
Specifying a proxy-advertised IP address 4
IRDP configuration examples 4
Example: Configuring IRDP 4

Configuring IRDP

About IRDP

ICMP Router Discovery Protocol (IRDP), an extension of the ICMP, is independent of any routing protocol. It allows hosts to discover the IP addresses of neighboring routers that can act as default gateways to reach devices on other IP networks.

IRDP enables hosts to track dynamic changes in router availability. It does not require you to configure the default gateways on hosts or configure hosts to listen to routing protocols to obtain IP addresses of routers.

IRDP operation

IRDP uses the following types of ICMP messages:

- **Router advertisement (RA)**—Sent by a router to advertise IP addresses (including the primary and secondary IP addresses) and preference.
- **Router solicitation (RS)**—Sent by a host to request the IP addresses of routers on the subnet.

An interface with IRDP enabled periodically broadcasts or multicasts an RA message to advertise its IP addresses and their preferences. A receiving host adds the IP addresses to its routing table, and selects the IP address with the highest preference as the default gateway.

When a host attached to the subnet starts up, the host multicasts an RS message to request immediate advertisements. If the host does not receive any advertisements, it retransmits the RS several times. If the host does not discover the IP addresses of neighboring routers because of network problems, the host can still discover them from periodic RAs.

IRDP allows hosts to discover neighboring routers, but it does not suggest the best route to a destination. If a host sends a packet to a router that is not the best next hop, the host will receive an ICMP redirect message from the router.

Protocols and standards

RFC 1256: *ICMP Router Discovery Messages*

IRDP tasks at a glance

To configure IRDP, perform the following tasks:

1. Enabling IRDP
2. (Optional.) Specifying the preference of advertised IP addresses
3. (Optional.) Setting the lifetime of advertised IP addresses
4. (Optional.) Setting the advertising intervals
5. (Optional.) Specifying the multicast address as the destination address for RAs
6. (Optional.) Specifying a proxy-advertised IP address

Restrictions and guidelines: IRDP configuration

- IRDP takes effect on any Layer 3 interface (except the loopback interface) if the interface has an IP address.
- After IRDP is enabled on an interface, the IRDP configurations take effect, and the device sends RA messages out of the interface.

Enabling IRDP

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IRDP on the interface.
   ip irdp
   By default, IRDP is disabled on an interface.

Specifying the preference of advertised IP addresses

About this task

Every IP address advertised in RAs has a preference value. A larger preference value represents a higher preference. The IP address with the highest preference is selected as the default gateway address.

You can specify the preference for IP addresses to be advertised on a router interface.

An address with the minimum preference value (-2147483648) will not be used as a default gateway address.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the preference of advertised primary and secondary IP addresses on the interface.
   ip irdp preference *preference-value*
   The default preference is 0.
   The preference applies to all advertised primary and secondary IP addresses on the interface.

Setting the lifetime of advertised IP addresses

About this task

An RA contains a lifetime field that specifies the lifetime of advertised IP addresses. If the host does not receive a new RA for an IP address within the address lifetime, the host removes the route entry.

All the IP addresses advertised by an interface, including proxy-advertised IP addresses and the interface IP address, have the same lifetime.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the lifetime of advertised IP addresses.
   ip irdp lifetime *lifetime-value*
   The default lifetime is 1800 seconds.
   The lifetime cannot be shorter than the maximum advertising interval.

Setting the advertising intervals

About this task

A router interface with IRDP enabled sends out RAs randomly between the minimum and maximum advertising intervals. This mechanism prevents the local link from being overloaded by a large number of RAs sent simultaneously from routers.

As a best practice, shorten the advertising interval on a link that suffers high packet loss rates.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the maximum and minimum advertising intervals.
   ip irdp interval *max-interval* [*min-interval*]
   By default, the maximum interval is 600 seconds, and the minimum interval is 3/4 of the maximum interval.

Specifying the multicast address as the destination address for RAs

About this task

An RA uses either of the following destination IP addresses:

- Broadcast address 255.255.255.255.
- Multicast address 224.0.0.1, which identifies all hosts on the local link.

By default, the destination IP address of an RA is the broadcast address. If all listening hosts in a local area network support IP multicast, specify 224.0.0.1 as the destination IP address.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*

3. Specify the multicast address 224.0.0.1 as the destination IP address of RAs.

   ip irdp multicast

   By default, RAs use the broadcast address 255.255.255.255 as the destination IP address.

Specifying a proxy-advertised IP address

About this task

By default, an interface advertises its primary and secondary IP addresses. You can specify IP addresses of other gateways for an interface to proxy-advertise.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify a proxy-advertised IP address and its preference.

   ip irdp address *ip-address preference-value*

   By default, no proxy-advertised IP address is specified.

   You can repeat this command to specify a maximum of four proxy-advertised IP addresses on an interface.

IRDP configuration examples

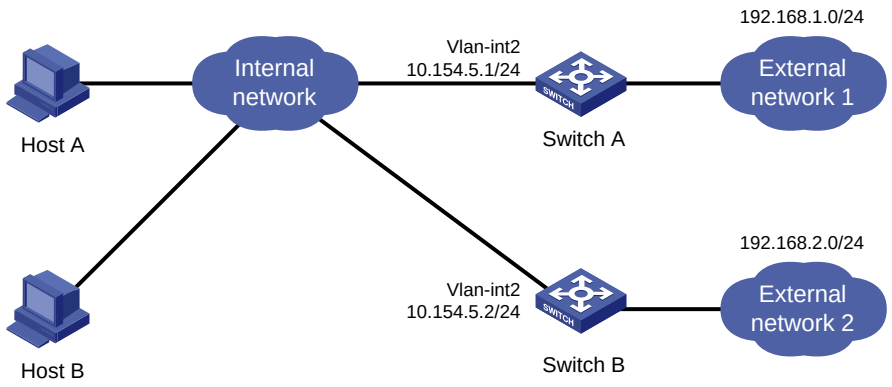
Example: Configuring IRDP

Network configuration

As shown in Figure 1, Host A and Host B that run Linux support IRDP, and they are in the internal network. Switch A and Switch B act as the egress routers and connect to external networks 192.168.1.0/24 and 192.168.2.0/24, respectively.

Configure Switch A as the default gateway for the hosts. Make sure Switch A has routes to reach both External network 1 and External network 2, so that packets to these external networks can be correctly routed.

Figure 1 Network diagram



Procedure

1. Configure Switch A:

Configure a route for Switch A to reach External network 2. (Details not shown.)

Specify an IP address for VLAN-interface 2.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 10.154.5.1 24
```

Enable IRDP on VLAN-interface 2.

```
[SwitchA-Vlan-interface2] ip irdp
```

Specify preference 1000 for advertised IP addresses on VLAN-interface 2.

```
[SwitchA-Vlan-interface2] ip irdp preference 1000
```

Specify the multicast address 224.0.0.1 as the destination IP address for RAs sent by VLAN-interface 2.

```
[SwitchA-Vlan-interface2] ip irdp multicast
```

Specify the IP address 192.168.1.0 and preference 400 for VLAN-interface 2 to proxy-advertise.

```
[SwitchA-Vlan-interface2] ip irdp address 192.168.1.0 400
```

2. Configure Switch B:

Specify an IP address for VLAN-interface 2.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 10.154.5.2 24
```

Enable IRDP on VLAN-interface 2.

```
[SwitchB-Vlan-interface2] ip irdp
```

Specify preference 500 for advertised IP addresses on VLAN-interface 2.

```
[SwitchB-Vlan-interface2] ip irdp preference 500
```

Specify the multicast address 224.0.0.1 as the destination IP address for RAs sent by VLAN-interface 2.

```
[SwitchB-Vlan-interface2] ip irdp multicast
```

Specify the IP address 192.168.2.0 and preference 400 for VLAN-interface 2 to proxy-advertise.

```
[SwitchB-Vlan-interface2] ip irdp address 192.168.2.0 400
```

Verifying the configuration

Display the routing table for Host A.

```
[HostA@localhost ~]$ netstat -rne
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
10.154.5.0      0.0.0.0         255.255.255.0   U     0      0        0 eth1
192.168.1.0     0.0.0.0         255.255.255.0   U     0      0        0 eth1
192.168.2.0     0.0.0.0         255.255.255.0   U     0      0        0 eth1
0.0.0.0         10.154.5.1      0.0.0.0         UG    0      0        0 eth1
```

The output shows that the default route on Host A points to IP address 10.154.5.1, and Host A has routes to 192.168.1.0/24 and 192.168.2.0/24.

Display the routing table for Host B.

```
[HostB@localhost ~]$ netstat -rne
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
```

```
10.154.5.0      0.0.0.0         255.255.255.0   U     0      0        0 eth1
192.168.1.0     0.0.0.0         255.255.255.0   U     0      0        0 eth1
192.168.2.0     0.0.0.0         255.255.255.0   U     0      0        0 eth1
0.0.0.0         10.154.5.1      0.0.0.0         UG    0      0        0 eth1
```

The output shows that the default route on Host B points to IP address 10.154.5.1, and Host B has routes to 192.168.1.0/24 and 192.168.2.0/24.

Contents

Optimizing IP performance 1

IP performance optimization tasks at a glance 1
Enabling an interface to forward directed broadcasts destined for the directly connected network 2
About forwarding broadcasts destined for the directly connected network 2
Procedure 2
Example: Enabling an interface to forward directed broadcasts destined for the directly connected network 2
Setting the interface MTU for IPv4 packets 3
Enabling Layer 3 packet statistics collection 3
Enabling IPv4 local fragment reassembly 4
Enabling processing IP options in IP packets 4
Enabling IP packet isolation between management port and service port 5
Enabling TTL expiry alarm generation 5
Enabling sending ICMP error messages 6
About sending ICMP error messages 6
Enabling sending ICMP redirect messages 6
Enabling sending ICMP time exceeded messages 6
Enable sending ICMP destination unreachable messages 7
Configuring rate limit for ICMP error messages 8
Disabling forwarding ICMP fragments 8
Specifying the source address for ICMP packets 8
Disabling response to broadcast echo requests 9
Disabling sending a specific type of ICMP messages 9
Disabling receiving a specific type of ICMP messages 10
Enable SNMP notifications for attack events on ports 10
Setting TCP MSS for an interface 11
Configuring TCP path MTU discovery 11
Enabling SYN Cookie 12
Setting the TCP buffer size 13
Setting TCP timers 13
Enabling the Timestamps option encapsulation in outgoing TCP packets 13
Enabling SNMP notifications for TCP events 14
Display and maintenance commands for IP performance optimization 14

Optimizing IP performance

IP performance optimization tasks at a glance

All IP performance optimization tasks are optional.

1. Configuring features for IP packets
   - Enabling an interface to forward directed broadcasts destined for the directly connected network
   - Setting the interface MTU for IPv4 packets
   - Enabling Layer 3 packet statistics collection
   - Enabling IPv4 local fragment reassembly
   - This feature is applicable in IRF networks.
   - Enabling processing IP options in IP packets
   - Enabling IP packet isolation between management port and service port
   - Enabling TTL expiry alarm generation
2. Configuring features for ICMP messages
   - Enabling sending ICMP error messages
   - Configuring rate limit for ICMP error messages
   - Disabling forwarding ICMP fragments
   - Specifying the source address for ICMP packets
   - Disabling response to broadcast echo requests
   - Disabling sending a specific type of ICMP messages
   - Disabling receiving a specific type of ICMP messages
   - Enable SNMP notifications for attack events on ports
3. Configuring features for TCP packets
   - Setting TCP MSS for an interface
   - Configuring TCP path MTU discovery
   - Enabling SYN Cookie
   - Setting the TCP buffer size
   - Setting TCP timers
   - Enabling the Timestamps option encapsulation in outgoing TCP packets
   - Enabling SNMP notifications for TCP events

Enabling an interface to forward directed broadcasts destined for the directly connected network

About forwarding broadcasts destined for the directly connected network

A directed broadcast packet is destined for all hosts on a specific network. In the destination IP address of the directed broadcast, the network ID identifies the target network, and the host ID is made up of all ones.

If an interface is allowed to forward directed broadcasts destined for the directly connected network, hackers can exploit this vulnerability to attack the target network. In some scenarios, however, an interface must send such directed broadcast packets to support the following features:

- **UDP helper**—Converts the directed broadcasts to unicasts and forwards them to a specific server.
- **Wake on LAN**—Sends the directed broadcasts to wake up the hosts on the target network.

You can configure this function to enable the interface to forward directed broadcast packets that are destined for directly connected network.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the interface to forward directed broadcasts destined for the directly connected network.
   ip forward-broadcast [**acl** *acl-number*]
   By default, an interface cannot forward directed broadcasts destined for the directly connected network.

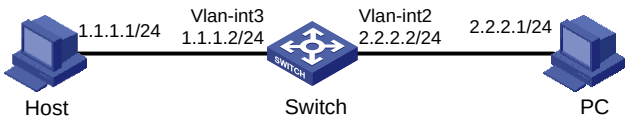
Example: Enabling an interface to forward directed broadcasts destined for the directly connected network

Network configuration

As shown in Figure 1, the default gateway of the host is 1.1.1.2/24, the IP address of VLAN-interface 3 of the switch.

The PC can receive directed broadcasts from the host to IP address 2.2.2.255.

Figure 1 Network diagram



Procedure

1. Specify IP addresses for VLAN-interface 3 and VLAN-interface 2.

```
<Switch> system-view
[Switch] interface vlan-interface 3
[Switch-Vlan-interface3] ip address 1.1.1.2 24
[Switch-Vlan-interface3] quit
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 2.2.2.2 24
```

2. Enable VLAN-interface 2 to forward broadcasts directed for the directly connected network.

```
[Switch-Vlan-interface2] ip forward-broadcast
```

Verifying the configuration

After the configurations are completed, if you ping the subnet-directed broadcast address 2.2.2.255 on the host, the PC can receive the ping packets. If you remove the **ip forward-broadcast** configuration on the switch, the PC cannot receive the ping packets.

Setting the interface MTU for IPv4 packets

About this task

The interface MTU for IPv4 packets defines the largest size of an IPv4 packet that an interface can transmit without fragmentation. When a packet exceeds the MTU of the sending interface, the device processes the packet in one of the following ways:

- If the packet disallows fragmentation, the device discards it.
- If the packet allows fragmentation, the device fragments it and forwards the fragments.

Fragmentation and reassembling consume system resources, so set the MTU based on the network environment to avoid fragmentation.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the interface MTU for IPv4 packets.

   ip mtu *mtu-size*

   By default, the interface MTU is not set.

Enabling Layer 3 packet statistics collection

About this task

With this feature enabled on an interface, the device counts incoming and outgoing IP packets on the interface. To display the collected statistics, execute the **display interface** command.

Restrictions and guidelines

When the interface is processing a large number of packets, Layer 3 packet statistics collection will cause high CPU usage and degrade forwarding performance. If the statistics are not necessary, disable this feature to ensure device performance.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable Layer 3 packet statistics collection.
   statistics l3-packet enable { **inbound** | **outbound** }
   By default, Layer 3 packet statistics collection is disabled.

Enabling IPv4 local fragment reassembly

About this task

By default, a subordinate in a multichassis IRF fabric delivers received packet fragments to the master for centralized fragment reassembly. To improve fragment reassembly efficiency, use this feature. When a subordinate receives fragments of an IPv4 packet, it reassembles them directly.

Use this feature with caution, because fragment reassembly fails on a subordinate when the fragments of an IPv4 packet are received on different subordinates.

Procedure

1. Enter system view.
   system-view
2. Enable IPv4 local fragment reassembly.
   ip reassemble local enable
   By default, IPv4 local fragment reassembly is disabled.

Enabling processing IP options in IP packets

About this task

IP options are typically used for network path diagnosis or temporary transmission of specific services.

By default, when a packet with IP options arrives at an intermediate device, the device sends the packet to the CPU to process IP options before forwarding it out. In a network with excessive packet exchanges, processing IP options will prevent the intermediate device from processing packets in a timely manner and cause packet loss. To avoid this situation, you can perform either of the following operations:

- Execute the **ip option forward** command. This command disables the device from processing IP options in packets that are not destined for the device. Then, those packets will be forwarded through hardware.
- Execute the **ip option drop** command. This command enables the device to drop IP packets that carry IP options and are not destined for the device.

CAUTION:
As a best practice, execute this command only when IP options are not used in the network.

When IP options are required and the device does not support hardware forwarding, you can execute the **undo ip option** command. The device will then deliver IP packets carrying IP options to the CPU for processing.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable the device to process IP options in IP packets.
   ```
   ip option { drop | forward }
   ```
   By default, the device delivers IP packets carrying IP options to the CPU for processing.

Enabling IP packet isolation between management port and service port

About this task

With this feature enabled, the device discards IP packets received and sent on different types of ports (management or service), for example, packets received on the management port but sent out of the service port. This isolates management IP traffic and service IP traffic from each other, preventing attacks against the management network through the service network.

In scenarios where the management port forwards IP service traffic, do not enable this feature as a best practice.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable IP packet isolation between management port and service port.
   ```
   management-port isolate enable
   ```
   By default, IP packet isolation between management port and service port is disabled.

Enabling TTL expiry alarm generation

About this task

NOTE:

A TTL expiry packet is an IP packet with a TTL value less than or equal to one.

When the device has a loop or is under TTL expiry attacks, it receives a large number of TTL expiry packets in a short period. The number of TTL expiry packets might grow exponentially. Processing those packets will occupy the resources used for processing service packets.

To avoid the above issue, use this feature to check for loops or TTL expiry attacks. The device periodically obtains the number of packet bytes with TTL expiry errors at the specified intervals. If the number of bytes in TTL expiry packets received during an interval equals or exceeds that during the previous interval multiplied by the specified alarm threshold, the device will generate a TTL expiry alarm. To have the device notify TTL expiry alarm clearance, specify the **no-drop-notify** keyword.

Procedure

1. Enter system view.
   ```
   system-view
   ```

2. Enable the device to generate TTL expiry alarms.

   ip ttl-expired drop-rate alarm enable [**interval** *interval-value*] [**threshold** *threshold-value*] [**no-drop-notify** [**cycles** *cycle-value*]]

   By default, the device supports generating TTL expiry alarms.

Enabling sending ICMP error messages

About sending ICMP error messages

ICMP messages are used by network layer and transport layer protocols to communicate updates and errors with other devices, facilitating network management.

Sending excessive ICMP messages increases network traffic. The device performance degrades if it receives a lot of malicious ICMP messages that cause it to respond with ICMP error messages. To prevent such problems, the sending of ICMP error messages is disabled by default. You can enable sending ICMP error messages of different types as needed.

ICMP error messages include redirect messages, time exceeded messages, and destination unreachable messages.

Enabling sending ICMP redirect messages

About this task

A host that has only one default route sends all packets to the default gateway. The default gateway sends an ICMP redirect message to inform the host of a correct next hop when the following conditions are met:

- The receiving and sending interfaces are the same.
- The packet source IP address and the IP address of the packet receiving interface are on the same segment.
- There is no source route option in the received packet.

ICMP redirect messages simplify host management and enable hosts to gradually optimize their routing table.

Procedure

1. Enter system view.

   system-view

2. Enable sending ICMP redirect messages.
3. **ip redirects enable**

   By default, the device does not send ICMP redirect messages.

Enabling sending ICMP time exceeded messages

About this task

A device sends ICMP time exceeded messages by following these rules:

- The device sends the source an ICMP TTL exceeded in transit message when the following conditions are met:
  - The received packet is not destined for the device.
  - The TTL field of the packet is 1.

- When the device receives the first fragment of an IP datagram destined for it, it starts a timer. If the timer expires before all the fragments of the datagram are received, the device sends an ICMP fragment reassembly time exceeded message to the source.

Restrictions and guidelines

If the ICMP time exceeded message sending is disabled, the device does not send ICMP TTL exceeded in transit messages. However, it can still send ICMP fragment reassembly time exceeded messages.

Procedure

1. Enter system view.

   `system-view`
2. Enable sending ICMP time exceeded messages.
3. **`ip ttl-expires enable`**

   By default, the device does not send ICMP time exceeded messages.

Enable sending ICMP destination unreachable messages

About this task

A device sends ICMP destination unreachable messages by following these rules:

- The device sends the source an ICMP network unreachable message when the following conditions are met:
  - The packet does not match any route.
  - No default route exists in the routing table.
- The device sends the source an ICMP protocol unreachable message when the following conditions are met:
  - The packet is destined for the device.
  - The transport layer protocol of the packet is not supported by the device.
- The device sends the source an ICMP port unreachable message when the following conditions are met:
  - The UDP packet is destined for the device.
  - The packet's port number does not match the corresponding process.
- The device sends the source an ICMP source route failed message when the following conditions are met:
  - The source uses Strict Source Routing to send packets.
  - The intermediate device finds that the next hop specified by the source is not directly connected.
- The device sends the source an ICMP fragmentation needed and DF set message when the following conditions are met:
  - The MTU of the sending interface is smaller than the packet.
  - The packet has DF set.

Restrictions and guidelines

If a DHCP-enabled device receives an ICMP echo reply without sending any ICMP echo requests, the device does not send any ICMP protocol unreachable messages to the source. To enable DHCP, use the **`dhcp enable`** command. For more information about this command, see DHCP commands in *Layer 3—IP Services Command Reference*.

Procedure

1. Enter system view.

system-view

2. Enable sending ICMP destination unreachable messages.
3. **ip unreachables enable**

   By default, the device does not send ICMP destination unreachable messages.

Configuring rate limit for ICMP error messages

About this task

To avoid sending excessive ICMP error messages within a short period that might cause network congestion, you can limit the rate at which ICMP error messages are sent. A token bucket algorithm is used with one token representing one ICMP error message.

A token is placed in the bucket at intervals until the maximum number of tokens that the bucket can hold is reached.

A token is removed from the bucket when an ICMP error message is sent. When the bucket is empty, ICMP error messages are not sent until a new token is placed in the bucket.

Procedure

1. Enter system view.

   system-view
2. Set the interval for tokens to arrive in the bucket and the bucket size for ICMP error messages.

   ip icmp error-interval *interval* [*bucketsize*]

   By default, a token is placed in the bucket at intervals of 100 milliseconds and the bucket allows a maximum of 10 tokens.

   To disable the ICMP rate limit, set the interval to 0 milliseconds.

Disabling forwarding ICMP fragments

Restrictions and guidelines

Disabling forwarding ICMP fragments can protect your device from ICMP fragment attacks.

Procedure

1. Enter system view.

   system-view
2. Disable forwarding ICMP fragments.

   ip icmp fragment discarding

   By default, forwarding ICMP fragments is enabled.

Specifying the source address for ICMP packets

About this task

Specifying the source IP address for outgoing ping echo requests and ICMP error messages helps users to locate the sending device easily. As a best practice, specify the IP address of the loopback interface as the source IP address.

Restrictions and guidelines

If you specify an IP address in the **ping** command, ping echo requests use the specified address as the source IP address rather than the IP address specified by the **ip icmp source** command.

Procedure

1. Enter system view.
   system-view
2. Specify the source address for outgoing ICMP packets.
   ip icmp source [**vpn-instance** *vpn-instance-name*] *ip-address*
   By default, no source address is specified for outgoing ICMP packets. The default source IP addresses for different types of ICMP packets vary as follows:
   - For an ICMP error message, the source IP address is the IP address of the receiving interface of the packet that triggers the ICMP error message. ICMP error messages include Time Exceeded, Port Unreachable, and Parameter Problem messages.
   - For an ICMP echo request, the source IP address is the IP address of the sending interface.
   - For an ICMP echo reply, the source IP address is the destination IP address of the ICMP echo request specific to this reply.

Disabling response to broadcast echo requests

About this task

When you use the **ping** command to ping a broadcast address, all devices that have received the ICMP echo request must respond to the request. Attackers might exploit this mechanism to launch an ICMP flood attack. The victims will respond to ICMP echo requests frequently, which degrades their forwarding performance. To resolve this issue, perform this task to disable the device from responding to broadcast echo requests.

You can also disable response to broadcast echo requests in the following scenarios:

- The device has received too many broadcast packets.
- The CPU usage of the device is too high.

Procedure

1. Enter system view.
   system-view
2. Disable response to broadcast echo requests.
   undo ip icmp broadcast-echo-reply enable
   By default, the device responds to broadcast echo requests.

Disabling sending a specific type of ICMP messages

About this task

By default, the device sends all types of ICMP messages except Destination Unreachable, Time Exceeded, and Redirect messages. Attackers might obtain information from specific types of ICMP messages, causing security issues.

For security purposes, you can perform this task disable sending ICMP messages of specific types.

Restrictions and guidelines

Disabling sending ICMP messages of a specific type might affect network operation. Please use this feature with caution.

To enable sending Destination Unreachable, Time Exceeded, or Redirect messages, you can perform one of the following tasks:

- Execute the **ip icmp send enable** command.
- Execute one of the following commands as needed:
  - **ip unreachables enable**
  - **ip ttl-expires enable**
  - **ip redirects enable**

Procedure

1. Enter system view.

   system-view
2. Disable the device from sending a specific type of ICMP messages.

   undo ip icmp { **name** *icmp-name* | **type** *icmp-type* **code** *icmp-code* } **send enable**

   By default, the device sends all types of ICMP messages except Destination Unreachable, Time Exceeded, and Redirect messages.

Disabling receiving a specific type of ICMP messages

About this task

By default, the device receives all types of ICMP messages. Such a setting might affect device performance if a large number of ICMP responses are received within a short time. To resolve this issue, you can perform this task to disable the device from receiving a specific type of ICMP messages.

Restrictions and guidelines

Disabling receiving ICMP messages of a specific type might affect network operation. Please use this feature with caution.

Procedure

1. Enter system view.

   system-view
2. Disable the device from receiving a specific type of ICMP messages.

   undo ip icmp { **name** *icmp-name* | **type** *icmp-type* **code** *icmp-code* } **receive enable**

   By default, the device receives all types of ICMP messages.

Enable SNMP notifications for attack events on ports

About this task

This feature enables the port attack module to generate SNMP notifications for critical events, such as ICMP packet reception overspeeding on a port. The SNMP notifications are sent to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable SNMP notifications for attack events on ports.
   ```
   snmp-agent trap enable port-attack
   ```
   By default, SNMP notifications are disabled for attack events on ports.

Setting TCP MSS for an interface

About this task

The maximum segment size (MSS) option informs the receiver of the largest segment that the sender can accept. Each end announces its MSS during TCP connection establishment. If the size of a TCP segment is smaller than the MSS of the receiver, TCP sends the TCP segment without fragmentation. If not, it fragments the segment according to the receiver's MSS.

Restrictions and guidelines

- If you set the TCP MSS on an interface, the size of each TCP segment received or sent on the interface cannot exceed the MSS value.
- This configuration takes effect only for TCP connections established after the configuration rather than the TCP connections that already exist.
- This configuration is effective only for IP packets.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the TCP MSS for the interface.
   tcp mss *value*
   By default, the TCP MSS is not set.

Configuring TCP path MTU discovery

About this task

TCP path MTU discovery (in RFC 1191) discovers the path MTU between the source and destination ends of a TCP connection. The device uses the path MTU to calculate the MSS to avoid IP fragmentation. The path MTU uses an aging mechanism to ensure that the source device can increase the path MTU when the minimum link MTU on the path increases.

TCP path MTU discovery works as follows:

1. A TCP source device sends a packet with the Don't Fragment (DF) bit set.
2. A router discards the packet that exceeds the MTU of the outgoing interface and returns an ICMP error message. The error message contains the MTU of the outgoing interface.
3. Upon receiving the ICMP message, the TCP source device calculates the current path MTU of the TCP connection.
4. The TCP source device sends subsequent TCP segments that are smaller than the MSS (MSS = path MTU – IP header length – TCP header length).

If the TCP source device still receives ICMP error messages when the MSS is smaller than 32 bytes, the TCP source device will fragment packets.

An ICMP error message received from a router that does not support RFC 1191 has the MTU of the outgoing interface set to 0. Upon receiving the ICMP message, the TCP source device selects the path MTU smaller than the current path MTU from the MTU table as described in RFC 1191. Based on the selected path MTU, the TCP source device calculates the TCP MSS. The MTU table contains MTUs of 68, 296, 508, 1006, 1280, 1492, 2002, 4352, 8166, 17914, 32000, and 65535 bytes. Because the minimum TCP MSS specified by the system is 32 bytes, the actual minimum MTU is 72 bytes.

The aging mechanism of the path MTU is as follows:

- When the TCP source device receives an ICMP error message, it reduces the path MTU and starts an aging timer for the path MTU.
- After the aging timer expires, the source device uses a larger MSS in the MTU table, as described in RFC 1191.
- If no ICMP error message is received within two minutes, the source device increases the MSS again until the MSS negotiated during TCP three-way handshake is reached.

Prerequisites

Make sure all devices on a TCP connection are enabled to send ICMP error messages by using the **ip unreachables enable** command.

Procedure

1. Enter system view.

   system-view

2. Enable TCP path MTU discovery.

   tcp path-mtu-discovery [**aging** *age-time* | **no-aging**]

   By default, TCP path MTU discovery is disabled.

Enabling SYN Cookie

About this task

A TCP connection is established through a three-way handshake. An attacker can exploit this mechanism to mount SYN Flood attacks. The attacker sends a large number of SYN packets, but does not respond to the SYN ACK packets from the server. As a result, the server establishes a large number of TCP semi-connections and can no longer handle normal services.

SYN Cookie can protect the server from SYN Flood attacks. When the server receives a SYN packet, it responds with a SYN ACK packet without establishing a TCP semi-connection. The server establishes a TCP connection and enters ESTABLISHED state only when it receives an ACK packet from the client.

Procedure

1. Enter system view.

   system-view

2. Enable SYN Cookie.

   tcp syn-cookie enable

   By default, SYN Cookie is disabled.

Setting the TCP buffer size

1. Enter system view.
   system-view
2. Set the size of TCP receive/send buffer.
   tcp window *window-size*
   The default buffer size is 63 KB.

Setting TCP timers

About this task

You can set the following TCP timers:

- **SYN wait timer**—TCP starts the SYN wait timer after sending a SYN packet. Within the SYN wait timer if no response is received or the upper limit on TCP connection tries is reached, TCP fails to establish the connection.
- **FIN wait timer**—TCP starts the FIN wait timer when TCP changes the connection state to FIN_WAIT_2. If no FIN packet is received within the timer interval, TCP terminates the connection. If a FIN packet is received, TCP changes the connection state to TIME_WAIT. If a non-FIN packet is received, TCP restarts the timer, and tears down the connection when the timer expires.

Procedure

1. Enter system view.
   system-view
2. Set the TCP SYN wait timer.
3. **tcp timer syn-timeout** *time-value*
   - By default, the TCP SYN wait timer is 75 seconds.
4. Set the TCP FIN wait timer.
5. **tcp timer fin-timeout** *time-value*
   By default, the TCP FIN wait timer is 675 seconds.

Enabling the Timestamps option encapsulation in outgoing TCP packets

About this task

Devices at each end of the TCP connection can calculate the RTT value by using the TCP Timestamps option carried in TCP packets. For security purpose in some networks, you can disable this feature at one end of the TCP connection to prevent intermediate devices from obtaining the Timestamps option information.

Procedure

1. Enter system view.
   system-view
2. Enable the device to encapsulate the TCP Timestamps option in outgoing TCP packets.
   tcp timestamps enable
   By default, the TCP timestamps option is encapsulated in outgoing TCP packets.

Enabling SNMP notifications for TCP events

About this task

This feature enables the TCP module to generate SNMP notifications for critical TCP events, such as MD5 authentication failure for TCP connection. The SNMP notifications are sent to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Enable SNMP notifications for TCP events.
   `snmp-agent trap enable tcp`
   By default, SNMP notifications for TCP events are enabled.

Display and maintenance commands for IP performance optimization

Execute **`display`** commands in any view and **`reset`** commands in user view.

Task	Command
Display ICMP statistics.	**`display icmp statistics`** [**`slot`** *`slot-number`*]
Display IP packet statistics.	**`display ip statistics`** [**`slot`** *`slot-number`*]
Display brief information about RawIP connections.	**`display rawip`** [**`slot`** *`slot-number`*]
Display detailed information about RawIP connections.	**`display rawip verbose`** [**`slot`** *`slot-number`* [**`pcb`** *`pcb-index`*]]
Display brief information about TCP connections.	**`display tcp`** [**`slot`** *`slot-number`*]
Display TCP traffic statistics.	**`display tcp statistics`** [**`slot`** *`slot-number`*]
Display detailed information about TCP connections.	**`display tcp verbose`** [**`slot`** *`slot-number`* [**`pcb`** *`pcb-index`*]]
Display brief information about UDP connections.	**`display udp`** [**`slot`** *`slot-number`*]
Display UDP traffic statistics.	**`display udp statistics`** [**`slot`** *`slot-number`*]
Display detailed information about UDP connections.	**`display udp verbose`** [**`slot`** *`slot-number`* [**`pcb`** *`pcb-index`*]]
Clear IP packet statistics.	**`reset ip statistics`** [**`slot`** *`slot-number`*]
Clear TCP traffic statistics.	**`reset tcp statistics`**
Clear UDP traffic statistics.	**`reset udp statistics`**

Contents

Configuring UDP helper 1
About UDP helper 1
Configuring UDP helper to convert broadcast to unicast 1
Configuring UDP helper to convert broadcast to multicast 2
Display and maintenance commands for UDP helper 2
UDP helper configuration examples 3
Example: Configuring UDP helper to convert broadcast to unicast 3
Example: Configuring UDP helper to convert broadcast to multicast 3

Configuring UDP helper

About UDP helper

UDP helper can provide the following packet conversion services for packets with specific UDP destination port numbers:

- Convert broadcast to unicast, and forward the unicast packets to specific destinations.
- Convert broadcast to multicast, and forward the multicast packets.

Configuring UDP helper to convert broadcast to unicast

About this task

You can configure UDP helper to convert broadcast packets with specific UDP port numbers to unicast packets.

Upon receiving a UDP broadcast packet, UDP helper uses the configured UDP ports to match the UDP destination port number of the packet.

- If a match is found, UDP helper duplicates the packet and modifies the destination IP address of the copy to the configured unicast address. Then UDP helper forwards the unicast packet to the unicast address.
- If no match is found, UDP helper does not process the packet.

Procedure

1. Enter system view.

   system-view

2. Enable UDP helper.

   udp-helper enable

   By default, UDP helper is disabled.

3. Specify a UDP port number for UDP helper.

   udp-helper port { *port-number* | **dns** | **netbios-ds** | **netbios-ns** | **tacacs** | **tftp** | **time** }

   By default, no UDP port numbers are specified for UDP helper.

   Do not set UDP ports 67 and 68 for UDP helper, because UDP helper cannot forward DHCP broadcast packets.

   You can specify a maximum of 256 UDP ports for UDP helper.

4. Enter interface view.

   interface *interface-type interface-number*

5. Specify a destination server for UDP helper to convert broadcast to unicast.

   udp-helper server *ip-address* [**global** | **vpn-instance** *vpn-instance-name*]

   By default, no destination servers are specified.

   Use this command on the interface that receives broadcast packets.

Configuring UDP helper to convert broadcast to multicast

About this task

You can configure UDP helper to convert broadcast packets with specific UDP port numbers to multicast packets.

Upon receiving a UDP broadcast packet, UDP helper uses the configured UDP ports to match the UDP destination port number of the packet.

- If a match is found, UDP helper duplicates the packet and modifies the destination IP address of the copy to the configured multicast address. Then UDP helper forwards the packet to the multicast group.
- If no match is found, UDP helper does not process the packet.

Procedure

1. Enter system view.

   system-view
2. Enable UDP helper.

   udp-helper enable

   By default, UDP helper is disabled.
3. Specify a UDP port number for UDP helper.

   udp-helper port { *port-number* | **dns** | **netbios-ds** | **netbios-ns** | **tacacs** | **tftp** | **time** }

   By default, no UDP port numbers are specified for UDP helper.

   Do not set UDP ports 67 and 68 for UDP helper, because UDP helper cannot forward DHCP broadcast packets.

   You can specify a maximum of 256 UDP ports for UDP helper.
4. Enter interface view.

   interface *interface-type interface-number*
5. Specify a destination multicast address for UDP helper to convert broadcast to multicast.

   udp-helper broadcast-map *multicast-address* [**acl** *acl-number*]

   By default, no destination multicast addresses are specified for UDP helper.

   Use this command on the interface that receives broadcast packets.

Display and maintenance commands for UDP helper

Execute **display** command in any view and **reset** commands in user view.

Task	Command
Display information about broadcast to unicast conversion by UDP helper on an interface.	**display udp-helper interface** *interface-type interface-number*
Clear packet statistics for UDP helper.	**reset udp-helper statistics**

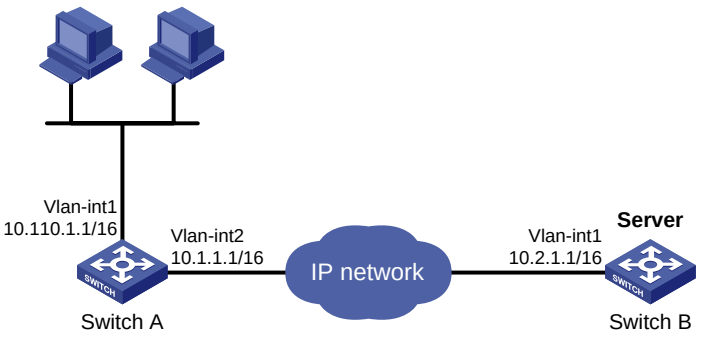
UDP helper configuration examples

Example: Configuring UDP helper to convert broadcast to unicast

Network configuration

As shown in Figure 1, configure UDP helper to convert broadcast to unicast on VLAN-interface 1 of Switch A. This feature enables Switch A to forward broadcast packets with UDP destination port number 55 to the destination server 10.2.1.1/16.

Figure 1 Network diagram



Procedure

Make sure Switch A can reach the subnet 10.2.0.0/16.

Enable UDP helper.

```
[SwitchA] System-view
[SwitchA] udp-helper enable
```

Enable the UDP port 55 for UDP helper.

```
[SwitchA] udp-helper port 55
```

Specify the destination server 10.2.1.1 for UDP helper to convert broadcast to unicast on VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.110.1.1 16
[SwitchA-Vlan-interface1] udp-helper server 10.2.1.1
```

Verifying the configuration

Display information about broadcast to unicast conversion by UDP helper on VLAN-interface 1.

```
[SwitchA-Vlan-interface1] display udp-helper interface vlan-interface 1
Interface                Server VPN instance          Server address Packets sent
Vlan-interface1          N/A                          10.2.1.1       5
```

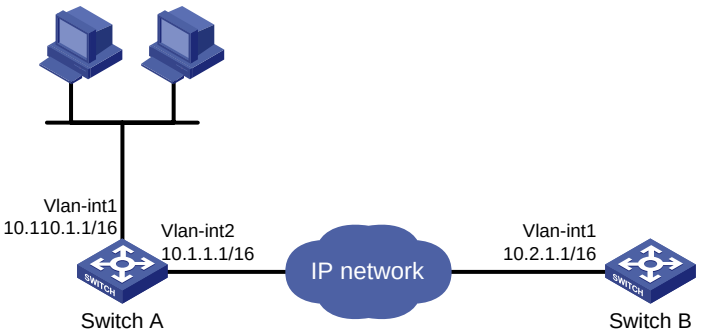
Example: Configuring UDP helper to convert broadcast to multicast

Network configuration

As shown in Figure 2, VLAN-interface 1 of Switch B can receive multicast packets destined to 225.1.1.1.

Configure UDP helper to convert broadcast to multicast on VLAN-interface 1 of Switch A. This feature enables Switch A to forward broadcast packets with UDP destination port number 55 to the multicast group 225.1.1.1.

Figure 2 Network diagram



Procedure

Make sure Switch A can reach the subnet 10.2.0.0/16.

1. Configure Switch A:

Enable UDP helper.

```
<SwitchA> system-view
[SwitchA] udp-helper enable
```

Enable the UDP port 55 for UDP helper.

```
[SwitchA] udp-helper port 55
```

Configure UDP helper to convert broadcast packets to multicast packets destined for 225.1.1.1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.110.1.1 16
[SwitchA-Vlan-interface1] udp-helper broadcast-map 225.1.1.1
[SwitchA-Vlan-interface1] quit
```

Enable IP multicast routing globally.

```
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Enable PIM-DM on VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] pim dm
[SwitchA-Vlan-interface1] quit
```

Enable PIM-DM and IGMP on VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] pim dm
[SwitchA-Vlan-interface2] igmp enable
```

Configure VLAN-interface 2 as a static member of multicast group 225.1.1.1.

```
[SwitchA-Vlan-interface2] igmp static-group 225.1.1.1
```

2. Configure Switch B:

Enable IP multicast routing globally.

```
<SwitchB> system-view
[SwitchB] multicast routing
[SwitchB-mrib] quit
```

Enable PIM-DM and IGMP on VLAN-interface 1.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] pim dm
[SwitchB-Vlan-interface1] igmp enable
```

Configure VLAN-interface 1 as a static member of multicast group 225.1.1.1.

```
[SwitchB-Vlan-interface1] igmp static-group 225.1.1.1
```

Verifying the configuration

Verify that you can capture multicast packets from Switch A on Switch B.

Contents

Configuring basic IPv6 settings 1

About IPv6 1
IPv6 features 1
IPv6 addresses 2
IPv6 path MTU discovery 4
IPv6 transition technologies 5
Protocols and standards 5
IPv6 basics tasks at a glance 6
Configuring an IPv6 global unicast address 6
About IPv6 global unicast address 6
Restrictions and guidelines 7
Generating an EUI-64 IPv6 address 7
Manually assigning an IPv6 global unicast address 7
Stateless address autoconfiguration 7
Configuring prefix-specific address autoconfiguration 9
Configuring an IPv6 link-local address 9
About IPv6 link-local address 9
Restrictions and guidelines 9
Configuring automatic generation of an IPv6 link-local address for an interface 10
Manually assigning an IPv6 link-local address to an interface 10
Configuring an IPv6 anycast address 10
Configuring path MTU discovery 11
Setting the interface MTU for IPv6 packets 11
Setting a static path MTU for an IPv6 address 11
Setting the aging time for dynamic path MTUs 12
Setting the IPv6 TCP MSS for an interface 12
Controlling sending ICMPv6 messages 13
Disabling receiving a specific type of ICMPv6 messages 13
Disabling sending a specific type of ICMPv6 messages 13
Configuring the rate limit for ICMPv6 error messages 14
Enabling replying to multicast echo requests 14
Enabling sending ICMPv6 destination unreachable messages 14
Enabling sending ICMPv6 time exceeded messages 15
Enabling sending ICMPv6 redirect messages 15
Specifying the source address for unsolicited ICMPv6 packets 16
Specifying the source IPv6 address for outgoing ICMPv6 replies 16
Enabling Layer 3 packet statistics collection 16
Enabling IPv6 local fragment reassembly 17
Enabling discarding IPv6 packets that contain extension headers 17
Enabling IPv6 fast traffic migration 18
Enabling IPv6 FIB entry consistency check 18
Enabling source MAC unknown IPv6 packets to trigger ND snooping 18
Enabling HopLimit expiry alarm generation 19
Enabling SNMP notifications for the IPv6 basics module 20
Enabling SNMP notifications for the IPv6 forwarding module 20
Display and maintenance commands for IPv6 basics 21
Basic IPv6 settings configuration examples 22
Example: Configuring basic IPv6 settings 22

Configuring IPv6 neighbor discovery 28

About IPv6 neighbor discovery 28
ICMPv6 messages used by IPv6 neighbor discovery 28
Address resolution 28
Neighbor reachability detection 29
Duplicate address detection 29
Router/prefix discovery and stateless address autoconfiguration 30
Redirection 30

Protocols and standards ····30
IPv6 neighbor discovery tasks at a glance ····30
Configuring a static neighbor entry ····31
Setting the dynamic neighbor learning limit ····32
Enabling unsolicited NA learning ····33
Setting the aging timer for ND entries in stale state ····34
Setting ND entry probing parameters ····34
Minimizing link-local ND entries ····35
Setting the hop limit ····35
Configuring RA message sending and parameters ····35
About RA message parameters ····35
Restrictions and guidelines ····36
Enabling the sending of RA messages ····37
Configuring parameters for RA messages ····37
Specifying DNS server information in RA messages ····38
Specifying DNS suffix information in RA messages ····39
Suppressing advertising DNS information in RA messages ····40
Setting the maximum number of attempts to send an NS message for DAD ····40
Enabling duplicate address detection ····41
Configuring ND snooping in a VLAN ····41
About ND snooping in a VLAN ····41
Enabling ND snooping ····43
Specifying an ACL for ND snooping entry filtering ····43
Setting the ND snooping entry learning limit ····44
Configuring ND snooping entry related parameters ····44
Configuring ND snooping in a VXLAN ····45
About ND snooping in a VXLAN ····45
Procedure ····45
Enabling ND proxy ····45
About ND proxy ····45
Enabling common ND proxy ····47
Enabling local ND proxy ····47
Enabling DAD message forwarding for a VSI ····47
Configuring IPv6 ND direct route advertisement ····48
About IPv6 ND direct route advertisement ····48
Restrictions and guidelines ····48
Application in Layer 3 access networks ····48
Procedure ····48
Enabling ND direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state ····49
Adding prefix information advertised through RA messages to IPSG binding entries ····50
Enabling recording user IPv6 address conflicts ····50
Enabling recording user port migrations ····50
Enabling ND logging for user online and offline events ····51
Logging ND entry deployment events ····51
Enabling ND entry consistency check ····52
Disabling the device from aging or deleting ND entries in response to network topology changes ····52
Enabling recording ND entry learning events ····53
Disabling sending NS messages when data packets trigger ND resolution ····53
Performing ND-ping ····54
Using ND-ping to verify the availability of an IPv6 address ····54
Using ND-Ping to obtain the IPv6 address for a MAC address ····54
Display and maintenance commands for IPv6 ND ····55
IPv6 ND configuration examples ····56
Example: Configuring ND snooping ····56

Configuring basic IPv6 settings

About IPv6

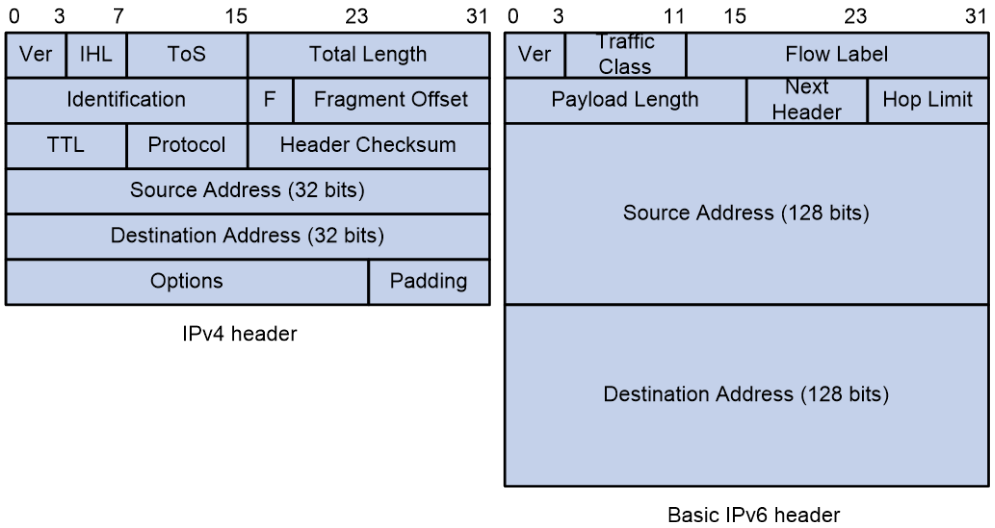
IPv6, also called IP next generation (IPng), was designed by the IETF as the successor to IPv4. One significant difference between IPv6 and IPv4 is that IPv6 increases the IP address size from 32 bits to 128 bits.

IPv6 features

Simplified header format

IPv6 removes several IPv4 header fields or moves them to the IPv6 extension headers to reduce the length of the basic IPv6 packet header. The basic IPv6 packet header has a fixed length of 40 bytes to simplify IPv6 packet handling and improve forwarding efficiency. Although the IPv6 address size is four times the IPv4 address size, the basic IPv6 packet header size is only twice the size of the option-less IPv4 packet header.

Figure 1 IPv4 packet header format and basic IPv6 packet header format



Larger address space

IPv6 can provide 3.4 x 10^{38} addresses to meet the requirements of hierarchical address assignment for both public and private networks.

Hierarchical address structure

IPv6 uses a hierarchical address structure to speed up route lookup and reduce the IPv6 routing table size through route aggregation.

Address autoconfiguration

To simplify host configuration, IPv6 supports stateful and stateless address autoconfiguration.

- Stateful address autoconfiguration enables a host to acquire an IPv6 address and other configuration information from a server (for example, a DHCPv6 server). For more information about DHCPv6 server, see "Configuring the DHCPv6 server."

- Stateless address autoconfiguration enables a host to automatically generate an IPv6 address and other configuration information by using its link-layer address and the prefix information advertised by a router.

To communicate with other hosts on the same link, a host automatically generates a link-local address based on its link-layer address and the link-local address prefix (FE80::/10).

Built-in security

IPv6 defines extension headers to support IPsec. IPsec provides end-to-end security and enhances interoperability among different IPv6 applications.

QoS support

The Flow Label field in the IPv6 header allows the device to label the packets of a specific flow for special handling.

Enhanced neighbor discovery mechanism

The IPv6 neighbor discovery protocol uses a group of ICMPv6 messages to manage information exchange among neighboring nodes on the same link. The group of ICMPv6 messages replaces ARP messages, ICMPv4 router discovery messages, and ICMPv4 redirect messages and provides a series of other functions.

Flexible extension headers

IPv6 eliminates the Options field in the header and introduces optional extension headers to provide scalability and improve efficiency. The Options field in the IPv4 packet header contains a maximum of 40 bytes, whereas the IPv6 extension headers are restricted to the maximum size of IPv6 packets.

IPv6 addresses

IPv6 address format

An IPv6 address is represented as a set of 16-bit hexadecimals separated by colons (:). An IPv6 address is divided into eight groups, and each 16-bit group is represented by four hexadecimal numbers, for example, 2001:0000:130F:0000:0000:09C0:876A:130B.

To simplify the representation of IPv6 addresses, you can handle zeros in IPv6 addresses by using the following methods:

- The leading zeros in each group can be removed. For example, the above address can be represented in a shorter format as 2001:0:130F:0:0:9C0:876A:130B.
- If an IPv6 address contains one or more consecutive groups of zeros, they can be replaced by a double colon (::). For example, the above address can be represented in the shortest format as 2001:0:130F::9C0:876A:130B.

IMPORTANT:

A double colon can appear once or not at all in an IPv6 address. This limit allows the device to determine how many zeros the double colon represents and correctly convert it to zeros to restore a 128-bit IPv6 address.

An IPv6 address consists of an address prefix and an interface ID, which are equivalent to the network ID and the host ID of an IPv4 address.

An IPv6 address prefix is written in IPv6-address/prefix-length notation. The prefix-length is a decimal number indicating how many leftmost bits of the IPv6 address are in the address prefix.

IPv6 address types

IPv6 addresses include the following types:

- **Unicast address**—An identifier for a single interface, similar to an IPv4 unicast address. A packet sent to a unicast address is delivered to the interface identified by that address.
- **Multicast address**—An identifier for a set of interfaces (typically belonging to different nodes), similar to an IPv4 multicast address. A packet sent to a multicast address is delivered to all interfaces identified by that address.

  Broadcast addresses are replaced by multicast addresses in IPv6.
- **Anycast address**—An identifier for a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to the nearest interface among the interfaces identified by that address. The nearest interface is chosen according to the routing protocol's measure of distance.

The type of an IPv6 address is designated by the first several bits, called the format prefix.

Table 1 Mappings between address types and format prefixes

Type		Format prefix (binary)	IPv6 prefix ID
Unicast address	Unspecified address	00...0 (128 bits)	::/128
	Loopback address	00...1 (128 bits)	::1/128
	Link-local address	1111111010	FE80::/10
	Global unicast address	Other forms	N/A
Multicast address		11111111	FF00::/8
Anycast address		Anycast addresses use the unicast address space and have the identical structure of unicast addresses.	

Unicast addresses

Unicast addresses include global unicast addresses, link-local unicast addresses, the loopback address, and the unspecified address.

- **Global unicast addresses**—Equivalent to public IPv4 addresses, global unicast addresses are provided for Internet service providers. This type of address allows for prefix aggregation to restrict the number of global routing entries.
- **Link-local addresses**—Used for communication among link-local nodes for neighbor discovery and stateless autoconfiguration. Packets with link-local source or destination addresses are not forwarded to other links.
- **A loopback address**—0:0:0:0:0:0:0:1 (or ::1). It has the same function as the loopback address in IPv4. It cannot be assigned to any physical interface. A node uses this address to send an IPv6 packet to itself.
- **An unspecified address**—0:0:0:0:0:0:0:0 (or ::). It cannot be assigned to any node. Before acquiring a valid IPv6 address, a node fills this address in the source address field of IPv6 packets. The unspecified address cannot be used as a destination IPv6 address.

Multicast addresses

IPv6 multicast addresses listed in Table 2 are reserved for special purposes.

Table 2 Reserved IPv6 multicast addresses

Address	Application
FF01::1	Node-local scope all-nodes multicast address.
FF02::1	Link-local scope all-nodes multicast address.
FF01::2	Node-local scope all-routers multicast address.
FF02::2	Link-local scope all-routers multicast address.

Multicast addresses also include solicited-node addresses. A node uses a solicited-node multicast address to acquire the link-layer address of a neighboring node on the same link and to detect duplicate addresses. Each IPv6 unicast or anycast address has a corresponding solicited-node address. The format of a solicited-node multicast address is FF02:0:0:0:0:1:FFXX:XXXX. FF02:0:0:0:0:1:FF is fixed and consists of 104 bits, and XX:XXXX is the last 24 bits of an IPv6 unicast address or anycast address.

EUI-64 address-based interface identifiers

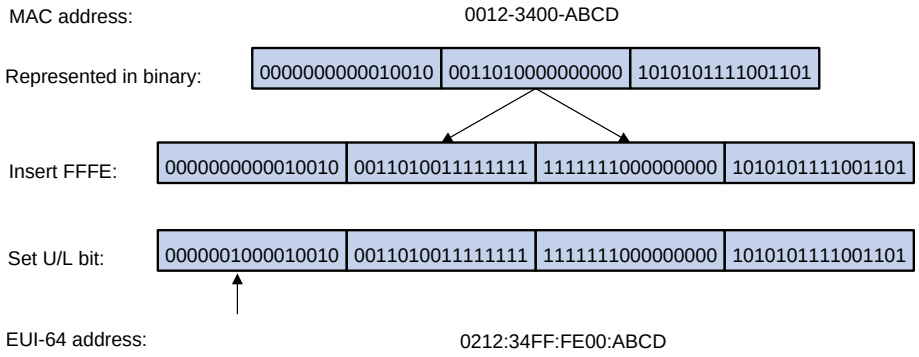
An interface identifier is 64 bits long and uniquely identifies an interface on a link.

On an IEEE 802 interface (such as a VLAN interface), the interface identifier is derived from the link-layer address (typically a MAC address) of the interface. The MAC address is 48 bits long.

To obtain an EUI-64 address-based interface identifier, follow these steps:

1. Insert the 16-bit binary number 1111111111111110 (hexadecimal value of FFFE) behind the 24th high-order bit of the MAC address.
2. Invert the universal/local (U/L) bit (the seventh high-order bit). This operation makes the interface identifier have the same local or global significance as the MAC address.

Figure 2 Converting a MAC address into an EUI-64 address-based interface identifier



On a tunnel interface, the lower 32 bits of the EUI-64 address-based interface identifier are the source IPv4 address of the tunnel interface. The higher 32 bits of the EUI-64 address-based interface identifier of an ISATAP tunnel interface are 0000:5EFE, whereas those of other tunnel interfaces are all zeros. For more information about tunnels, see "Configuring tunneling."

On an interface of another type (such as a serial interface), the EUI-64 address-based interface identifier is generated randomly by the device.

IPv6 path MTU discovery

The links that a packet passes from a source to a destination can have different MTUs, among which the minimum MTU is the path MTU. If a packet exceeds the path MTU, the source end fragments the packet to reduce the processing pressure on intermediate devices and to use network resources effectively.

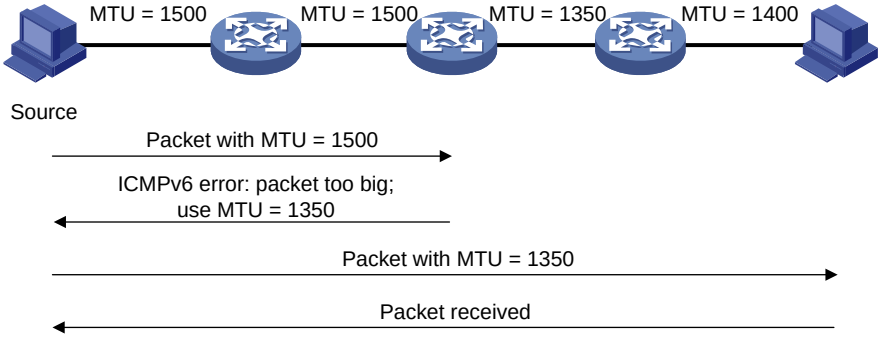
A source end uses path MTU discovery to find the path MTU to a destination, as shown in Figure 3.

1. The source host sends a packet no larger than its MTU to the destination host.
2. If the MTU of an intermediate device's output interface is smaller than the packet, the device performs the following operations:
   - Discards the packet.
   - Returns an ICMPv6 error message containing the interface MTU to the source host.
3. Upon receiving the ICMPv6 error message, the source host performs the following operations:
   - Uses the returned MTU to limit the packet size.

- Performs fragmentation.
- Sends the fragments to the destination host.

4. Step 2 and step 3 are repeated until the destination host receives the packet. In this way, the source host finds the minimum MTU of all links in the path to the destination host.

Figure 3 Path MTU discovery process



IPv6 transition technologies

IPv6 transition technologies enable communication between IPv4 and IPv6 networks.

Dual stack

Dual stack is the most direct transition approach. A network node that supports both IPv4 and IPv6 is a dual-stack node. A dual-stack node configured with an IPv4 address and an IPv6 address can forward both IPv4 and IPv6 packets. An application that supports both IPv4 and IPv6 prefers IPv6 at the network layer.

Dual stack is suitable for communication between IPv4 nodes or between IPv6 nodes. It is the basis of all transition technologies. However, it does not solve the IPv4 address depletion issue because each dual-stack node must have a globally unique IPv4 address.

Tunneling

Tunneling uses one network protocol to encapsulate the packets of another network protocol and transfers them over the network. For more information about tunneling, see "Configuring tunneling."

Protocols and standards

- RFC 1881, *IPv6 Address Allocation Management*
- RFC 1887, *An Architecture for IPv6 Unicast Address Allocation*
- RFC 1981, *Path MTU Discovery for IP version 6*
- RFC 2375, *IPv6 Multicast Address Assignments*
- RFC 2460, *Internet Protocol, Version 6 (IPv6) Specification*
- RFC 2464, *Transmission of IPv6 Packets over Ethernet Networks*
- RFC 2526, *Reserved IPv6 Subnet Anycast Addresses*
- RFC 3307, *Allocation Guidelines for IPv6 Multicast Addresses*
- RFC 4191, *Default Router Preferences and More-Specific Routes*
- RFC 4291, *IP Version 6 Addressing Architecture*
- RFC 4443, *Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification*
- RFC 4862, *IPv6 Stateless Address Autoconfiguration*

IPv6 basics tasks at a glance

To configure basic IPv6 settings, perform the following tasks:

1. Configuring an IPv6 address

   Choose the following tasks as needed:

   - Configuring an IPv6 global unicast address
   - Configuring an IPv6 link-local address
   - Configuring an IPv6 anycast address
2. (Optional.)Configuring path MTU discovery
   - Setting the interface MTU for IPv6 packets
   - Setting a static path MTU for an IPv6 address
   - Setting the aging time for dynamic path MTUs
3. (Optional.) Setting the IPv6 TCP MSS for an interface
4. (Optional.) Controlling sending ICMPv6 messages
   - Disabling receiving a specific type of ICMPv6 messages
   - Disabling sending a specific type of ICMPv6 messages
   - Configuring the rate limit for ICMPv6 error messages
   - Enabling replying to multicast echo requests
   - Enabling sending ICMPv6 destination unreachable messages
   - Enabling sending ICMPv6 time exceeded messages
   - Enabling sending ICMPv6 redirect messages
   - Specifying the source address for unsolicited ICMPv6 packets
   - Specifying the source IPv6 address for outgoing ICMPv6 replies
5. (Optional.) Enabling Layer 3 packet statistics collection
6. (Optional.) Enabling IPv6 local fragment reassembly
7. (Optional.) Enabling discarding IPv6 packets that contain extension headers
8. (Optional.) Enabling IPv6 fast traffic migration
9. (Optional.) Enabling IPv6 FIB entry consistency check
10. (Optional.) Enabling source MAC unknown IPv6 packets to trigger ND snooping
11. (Optional.) Enabling HopLimit expiry alarm generation
12. (Optional.) Enabling SNMP notifications for the IPv6 basics module
13. (Optional.) Enabling SNMP notifications for the IPv6 forwarding module

Configuring an IPv6 global unicast address

About IPv6 global unicast address

Use one of the following methods to configure an IPv6 global unicast address for an interface:

- **EUI-64 IPv6 address**—The IPv6 address prefix of the interface is manually configured, and the interface ID is generated automatically by the interface.
- **Manual configuration**—The IPv6 global unicast address is manually configured.
- **Stateless address autoconfiguration**—The IPv6 global unicast address is generated automatically based on the address prefix information contained in the RA message.

- **Prefix-specific address autoconfiguration**—The IPv6 global unicast address is generated automatically based on the prefix specified by its ID. The prefix can be manually configured or obtained through DHCPv6.

You can configure multiple IPv6 global unicast addresses on an interface.

Manually configured global unicast addresses (including EUI-64 IPv6 addresses) take precedence over automatically generated ones. If you manually configure a global unicast address with the same address prefix as an existing global unicast address on an interface, the manually configured one takes effect. However, it does not overwrite the automatically generated address. If you delete the manually configured global unicast address, the device uses the automatically generated one.

Restrictions and guidelines

An interface ID is used in EUI-64 IPv6 address configuration and stateless address autoconfiguration. The interface ID is derived from the MAC address of the interface. If the MAC address changes, the interface ID, global unicast address, and link-local address will also change. This will cause the entry table to rebuild for some protocols, such as MLD, IPv6 PIM, and OSPFv3. If you do not want this situation to occur, use other IPv6 address configuration methods that do not use MAC addresses, for example, manually specify an IPv6 address.

Generating an EUI-64 IPv6 address

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure an EUI-64 IPv6 address on the interface.
   ipv6 address { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* } **eui-64**
   By default, no EUI-64 IPv6 address is configured on an interface.

Manually assigning an IPv6 global unicast address

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Assign an IPv6 global unicast address to the interface.
   ipv6 address { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }
   By default, no IPv6 global unicast address is configured on an interface.

Stateless address autoconfiguration

About this task

Stateless address autoconfiguration enables an interface to automatically generate an IPv6 global unicast address by using the address prefix in the received RA message and the interface ID. On an IEEE 802 interface (such as an Ethernet interface or a VLAN interface), the interface ID is generated based on the interface's MAC address and is globally unique. An attacker can exploit this rule to identify the sending device easily.

To fix the vulnerability, you can configure the temporary address feature. With this feature, an IEEE 802 interface generates the following addresses:

- **Public IPv6 address**—Includes the address prefix in the RA message and a fixed interface ID generated based on the MAC address of the interface.
- **Temporary IPv6 address**—Includes the address prefix in the RA message and a random interface ID generated through MD5.

You can also configure the interface to preferentially use the temporary IPv6 address as the source address of sent packets. When the valid lifetime of the temporary IPv6 address expires, the interface deletes the address and generates a new one. This feature enables the system to send packets with different source addresses through the same interface. If the temporary IPv6 address cannot be used because of a DAD conflict, the public IPv6 address is used.

The preferred lifetime and valid lifetime for a temporary IPv6 address are determined as follows:

- The preferred lifetime of a temporary IPv6 address takes the smaller of the following values:
  - The preferred lifetime of the address prefix in the RA message.
  - The preferred lifetime configured for temporary IPv6 addresses minus DESYNC_FACTOR (a random number in the range of 0 to 600 seconds).
- The valid lifetime of a temporary IPv6 address takes the smaller of the following values:
  - The valid lifetime of the address prefix.
  - The valid lifetime configured for temporary IPv6 addresses.

Restrictions and guidelines

If the IPv6 prefix in the RA message is not 64 bits long, stateless address autoconfiguration fails to generate an IPv6 global unicast address.

To generate a temporary address, an interface must be enabled with stateless address autoconfiguration. Temporary IPv6 addresses do not overwrite public IPv6 addresses, so an interface can have multiple IPv6 addresses with the same address prefix but different interface IDs.

If an interface fails to generate a public IPv6 address because of a prefix conflict or other reasons, it does not generate any temporary IPv6 address.

Executing the **undo ipv6 address auto** command on an interface deletes all IPv6 global unicast addresses and link-local addresses that are automatically generated on the interface.

Enabling stateless address autoconfiguration

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable stateless address autoconfiguration on an interface, so that the interface can automatically generate a global unicast address.

   ipv6 address auto

   By default, the stateless address autoconfiguration feature is disabled on an interface.

Configuring the temporary address feature and preferentially using the temporary IPv6 address as the source address of outgoing packets

1. Enter system view.

   system-view
2. Enable the temporary IPv6 address feature.

   ipv6 temporary-address [*valid-lifetime preferred-lifetime*]

   By default, the temporary IPv6 address feature is disabled.

3. Enable the system to preferentially use the temporary IPv6 address as the source address of the outgoing packets.

   ipv6 prefer temporary-address

   By default, the system does not preferentially use the temporary IPv6 address as the source address of the outgoing packets.

Configuring prefix-specific address autoconfiguration

1. Enter system view.

   system-view

2. Configure an IPv6 prefix.

   Choose one option as needed:

   - Configure a static IPv6 prefix.

     ipv6 prefix *prefix-number ipv6-prefix/prefix-length*

     By default, no static IPv6 prefixes exist.

   - Use DHCPv6 to obtain a dynamic IPv6 prefix.

     For more information about IPv6 prefix acquisition, see "Configuring the DHCPv6 client."

3. Enter interface view.

   interface *interface-type interface-number*

4. Specify an IPv6 prefix for an interface to automatically generate an IPv6 global unicast address and advertise the prefix.

   ipv6 address *prefix-number sub-prefix/prefix-length*

   By default, no IPv6 prefix is specified for the interface to automatically generate an IPv6 global unicast address.

Configuring an IPv6 link-local address

About IPv6 link-local address

Configure IPv6 link-local addresses using one of the following methods:

- **Automatic generation**—The device automatically generates a link-local address for an interface according to the link-local address prefix (FE80::/10) and the link-layer address of the interface.
- **Manual assignment**—Manually configure an IPv6 link-local address for an interface.

Restrictions and guidelines

After you configure an IPv6 global unicast address for an interface, the interface automatically generates a link-local address. This link-local address is the same as the one generated by using the **ipv6 address auto link-local** command. If a link-local address is manually assigned to an interface, this manual assigned link-local address takes effect. If the manually assigned link-local address is deleted, the automatically generated link-local address takes effect.

Using the **undo ipv6 address auto link-local** command on an interface deletes only the link-local address generated by the **ipv6 address auto link-local** command. If the interface has an IPv6 global unicast address, it still has a link-local address. If the interface has no IPv6 global unicast address, it has no link-local address.

An interface can have only one link-local address. As a best practice, use the automatic generation method to avoid link-local address conflicts. If both the automatic generation and manual assignment methods are used, the manual assignment takes precedence.

- If you first use automatic generation and then manual assignment, the manually assigned link-local address overwrites the automatically generated one.
- If you first use manual assignment and then automatic generation, both of the following occur:
  - The link-local address is still the manually assigned one.
  - The automatically generated link-local address does not take effect. If you delete the manually assigned address, the automatically generated link-local address takes effect.

An interface ID is used for generating the link-local address for an interface. On an IEEE 802 interface (such as an Ethernet interface or a VLAN interface), the interface ID is derived from the MAC address of the interface. If the MAC address changes, the interface ID, global unicast address, and link-local address will also change. This will cause the entry table to rebuild for some protocols, such as MLD, IPv6 PIM, and OSPFv3. If you do not want this situation to occur, use other IPv6 address configuration methods that do not use MAC addresses, for example, manually specify a link-local address.

Configuring automatic generation of an IPv6 link-local address for an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the interface to automatically generate an IPv6 link-local address.
   ipv6 address auto link-local

   By default, no link-local address is configured on an interface.

   After an IPv6 global unicast address is configured on the interface, a link-local address is generated automatically.

Manually assigning an IPv6 link-local address to an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Manually assign an IPv6 link-local address to the interface.
   ipv6 address { *ipv6-address* [*prefix-length*] | *ipv6-address/prefix-length* } **link-local**

   By default, no link-local address is configured on an interface.

Configuring an IPv6 anycast address

4. Enter system view.
   system-view
5. Enter interface view.
   interface *interface-type interface-number*

6. Configure an IPv6 anycast address.

   ipv6 address { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* } **anycast**

   By default, no IPv6 anycast address is configured on an interface.

Configuring path MTU discovery

Setting the interface MTU for IPv6 packets

About this task

If the size of a packet exceeds the MTU of the sending interface, the device discards the packet. If the device is an intermediate device, it also sends the source host an ICMPv6 Packet Too Big message with the MTU of the sending interface. The source host fragments the packets according to the MTU. To avoid this situation, set a proper interface MTU.

Restrictions and guidelines

When you configure the MTU of an interface, the system prompts configuration failure but the MTU configuration actually takes effect on the interface. If the size of a packet exceeds the specified MTU, the device fragments the packet.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the interface MTU for IPv6 packets.

   ipv6 mtu *size*

   By default, no interface MTU is set.

Setting a static path MTU for an IPv6 address

About this task

You can set a static path MTU for an IPv6 address. Before sending a packet to the IPv6 address, the device compares the output interface MTU with the static path MTU. If the packet size exceeds the smaller one of the two values, the device fragments the packet according to the smaller value. After sending the fragmented packets, the device dynamically finds the path MTU to a destination host (see "IPv6 path MTU discovery").

Procedure

1. Enter system view.

   system-view

2. Set a static path MTU for an IPv6 address.

   ipv6 pathmtu [**vpn-instance** *vpn-instance-name*] *ipv6-address value*

   By default, no path MTU is set for any IPv6 address.

Setting the aging time for dynamic path MTUs

About this task

After the device dynamically discovers the path MTU to a destination host (see "IPv6 path MTU discovery"), it performs the following operations:

- Sends packets to the destination host based on this path MTU.
- Starts the aging timer for this path MTU.

When the aging timer expires, the device removes the dynamic path MTU and discovers the path MTU again.

Restrictions and guidelines

The aging time is invalid for a static path MTU.

Procedure

1. Enter system view.
   system-view
2. Set the aging time for dynamic path MTUs.
   ipv6 pathmtu age *age-time*
   The default setting is 10 minutes.

Setting the IPv6 TCP MSS for an interface

About this task

The maximum segment size (MSS) option informs the receiver of the largest segment that the sender can accept. Each end announces its MSS during TCP connection establishment. If the size of a TCP segment does not exceed the MSS value of the receiver, TCP sends the TCP segment without fragmentation. If not, it fragments the segment according to the receiver's MSS.

Restrictions and guidelines

- If you set the TCP MSS on an interface, the size of each TCP segment received or sent on the interface cannot exceed the MSS value.
- This configuration takes effect only for TCP connections established after the configuration rather than the TCP connections that already exist.
- This configuration is effective only for IPv6 packets. If MPLS is enabled on the interface, do not set the TCP MSS on the interface.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the IPv6 TCP MSS for the interface.
   ipv6 tcp mss *value*
   By default, the IPv6 TCP MSS is not set.

Controlling sending ICMPv6 messages

Disabling receiving a specific type of ICMPv6 messages

About this task

By default, the device receives all types of ICMPv6 messages. Such a setting might affect device performance if a large number of ICMPv6 responses are received within a short time. To resolve this issue, you can perform this task to disable the device from receiving a specific type of ICMPv6 messages.

Restrictions and guidelines

Disabling receiving ICMPv6 messages of a specific type might affect network operation. Please use this feature with caution.

Procedure

1. Enter system view.

   system-view

2. Disable the device from receiving a specific type of ICMPv6 messages.

   undo ipv6 icmpv6 { **name** *name* | **type** *type* **code** *code* } **receive enable**

   By default, the device receives all types of ICMPv6 messages.

Disabling sending a specific type of ICMPv6 messages

About this task

By default, the device sends all types of ICMPv6 messages except Destination Unreachable and Redirect messages. Attackers might obtain device information from specific types of ICMPv6 messages, causing security issues.

For security purposes, you can perform this task to disable the device from sending specific types of ICMPv6 messages.

Restrictions and guidelines

Disabling sending ICMPv6 messages of a specific type might affect network operation. Please use this feature with caution.

To enable sending Destination Unreachable, Time Exceeded, or Redirect messages, you can perform one of the following tasks:

- Execute the **ipv6 icmpv6 send enable** command.
- Execute one of the following commands as needed:
  - **ipv6 unreachables enable**
  - **ipv6 hoplimit-expires enable**
  - **ipv6 redirects enable**

Procedure

1. Enter system view.

   system-view

2. Disable the device from sending a specific type of ICMPv6 messages.

   undo ipv6 icmpv6 { **name** *name* | **type** *type* **code** *code* } **send enable**

   By default, the device can send all types of ICMPv6 messages except Destination Unreachable and Redirect messages.

Configuring the rate limit for ICMPv6 error messages

About this task

To avoid sending excessive ICMPv6 error messages within a short period that might cause network congestion, you can limit the rate at which ICMPv6 error messages are sent. A token bucket algorithm is used with one token representing one ICMPv6 error message.

A token is placed in the bucket at intervals until the maximum number of tokens that the bucket can hold is reached.

A token is removed from the bucket when an ICMPv6 error message is sent. When the bucket is empty, ICMPv6 error messages are not sent until a new token is placed in the bucket.

Procedure

1. Enter system view.

   `system-view`

2. Set the bucket size and the interval for tokens to arrive in the bucket for ICMPv6 error messages.

   `ipv6 icmpv6 error-interval` *`interval`* [*`bucketsize`*]

   By default, the bucket allows a maximum of 10 tokens. A token is placed in the bucket at an interval of 100 milliseconds.

   To disable the ICMPv6 rate limit, set the interval to 0 milliseconds.

Enabling replying to multicast echo requests

1. Enter system view.

   `system-view`

2. Enable replying to multicast echo requests.

   `ipv6 icmpv6 multicast-echo-reply enable`

   By default, this feature is disabled.

Enabling sending ICMPv6 destination unreachable messages

About this task

The device sends the source the following ICMPv6 destination unreachable messages:

- **ICMPv6 No Route to Destination message**—A packet to be forwarded does not match any route.
- **ICMPv6 Communication with Destination Administratively Prohibited message**—An administrative prohibition is preventing successful communication with the destination. This is typically caused by a firewall or an ACL on the device.
- **ICMPv6 Beyond Scope of Source Address message**—The destination is beyond the scope of the source IPv6 address. For example, a packet's source IPv6 address is a link-local address, and its destination IPv6 address is a global unicast address.
- **ICMPv6 Address Unreachable message**—The device fails to resolve the link layer address for the destination IPv6 address of a packet.
- **ICMPv6 Port Unreachable message**—No port process on the destination device exists for a received UDP packet.

Restrictions and guidelines

An ICMPv6 destination unreachable message indicates that the destination is not reachable from the source device. Attackers can launch malicious attacks to make the device generate incorrect ICMPv6 destination unreachable messages, which will affect the function of the network. To protect the network from malicious attacks and decrease unnecessary network traffic, you can disable the sending of ICMPv6 destination unreachable messages.

Procedure

1. Enter system view.

   `system-view`

2. Enable sending ICMPv6 destination unreachable messages.

   `ipv6 unreachables enable`

   By default, the device does not send ICMPv6 destination unreachable messages.

Enabling sending ICMPv6 time exceeded messages

About this task

The device sends the source ICMPv6 time exceeded messages as follows:

- If a received packet is not destined for the device and its hop limit is 1, the device sends an ICMPv6 hop limit exceeded in transit message to the source.
- Upon receiving the first fragment of an IPv6 datagram destined for the device, the device starts a timer. If the timer expires before all fragments arrive, the device sends an ICMPv6 fragment reassembly time exceeded message to the source.

Restrictions and guidelines

If the device receives large numbers of malicious packets, its performance degrades greatly because it must send back ICMP time exceeded messages. To prevent such attacks, disable sending ICMPv6 time exceeded messages.

Procedure

1. Enter system view.

   `system-view`

2. Enable sending ICMPv6 time exceeded messages.

   `ipv6 hoplimit-expires enable`

   By default, sending ICMPv6 time exceeded messages is enabled.

Enabling sending ICMPv6 redirect messages

About this task

Upon receiving a packet from a host, the device sends an ICMPv6 redirect message to inform the host of a better next hop when the following conditions are met:

- The interface receiving the packet is the interface forwarding the packet.
- The selected route is not created or modified by any ICMPv6 redirect messages.
- The selected route is not a default route.
- The forwarded packet does not contain the routing extension header.

The ICMPv6 redirect feature simplifies host management by enabling hosts that hold few routes to optimize their routing table gradually. However, to avoid adding too many routes on hosts, this feature is disabled by default.

Procedure

1. Enter system view.
   system-view
2. Enable sending ICMPv6 redirect messages.
   ipv6 redirects enable
   By default, the device does not send ICMPv6 redirect messages.

Specifying the source address for unsolicited ICMPv6 packets

About this task

Perform this task to specify the source IPv6 address for outgoing ping echo requests and ICMPv6 error messages. It is a good practice to specify the IPv6 address of a loopback interface as the source IPv6 address. This feature helps users easily locate the sending device.

Restrictions and guidelines

For ICMPv6 echo requests, the source IPv6 address specified in the **ping ipv6** command has higher priority than the source IPv6 address specified in the **ipv6 icmpv6 source** command.

Procedure

1. Enter system view.
   system-view
2. Specify a source IPv6 address for unsolicited ICMPv6 packets.
   ipv6 icmpv6 source [**vpn-instance** *vpn-instance-name*] *ipv6-address*
   By default, no source address is specified for unsolicited ICMPv6 packets.

Specifying the source IPv6 address for outgoing ICMPv6 replies

About this task

To help receiver hosts identify sources of ICMPv6 replies, you can specify the IPv6 address of a loopback interface or a local interface as the source IPv6 address for ICMPv6 replies.

Procedure

1. Enter system view.
   system-view
2. Specify a source IPv6 address for ICMPv6 replies.
   ipv6 icmpv6 reply source [**vpn-instance** *vpn-instance-name*] { **address** *ipv6-address* | **interface** *interface-type* [*interface-number*] }
   By default, no source IP address is specified for outgoing ICMPv6 replies.

Enabling Layer 3 packet statistics collection

About this task

With this feature enabled on an interface, the device counts incoming and outgoing IPv6 packets on the interface. To display the collected statistics, execute the **display interface** command.

Restrictions and guidelines

When the interface is processing a large number of packets, Layer 3 packet statistics collection will cause high CPU usage and degrade the forwarding performance. If the statistics are not necessary, disable this feature to ensure the device performance.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable Layer 3 packet statistics collection.
   statistics l3-packet enable { **inbound** | **outbound** }
   By default, Layer 3 packet statistics collection is disabled.

Enabling IPv6 local fragment reassembly

About this task

By default, a subordinate in a multichassis IRF fabric delivers received packet fragments to the master for centralized fragment reassembly. To improve fragment reassembly efficiency, use this feature. When a subordinate receives fragments of an IPv6 packet, it reassembles them directly.

Restrictions and guidelines

Use this feature with caution, because fragment reassembly fails on a subordinate when the fragments of an IPv6 packet are received on different subordinates.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 local fragment reassembly.
   ipv6 reassemble local enable
   By default, IPv6 local fragment reassembly is disabled.

Enabling discarding IPv6 packets that contain extension headers

About this task

This feature enables a device to discard a received IPv6 packet in which the extension headers cannot be processed by the device.

Procedure

1. Enter system view
   system-view
2. Enable the device to discard IPv6 packets that contain extension headers.
   ipv6 extension-header drop enable
   By default, the device does not discard IPv6 packets that contain extension headers.

Enabling IPv6 fast traffic migration

About this task

After IPv6 fast traffic migration is enabled on a VLAN interface, the device immediately steers IPv6 traffic forwarded through the VLAN interface to other forwarding paths upon network topology changes or user port migrations. This reduces traffic disruption time to improve overall network response speed and user experience.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IPv6 fast traffic migration.
   ipv6 forwarding fast-move
   By default, IPv6 fast traffic migration is disabled.

Enabling IPv6 FIB entry consistency check

About this task

Packet drops or incorrect forwarding might occur when the IPv6 FIB entries in hardware are inconsistent with FIB entries configured in software. To prevent these issues, enable IPv6 FIB entry consistency check.

This feature compares all IPv6 FIB entries in software with the IPv6 FIB entries in hardware regularly. If the device detects an inconsistency, the device performs the following tasks:

- Generates a log.
- Updates the IPv6 FIB entry in hardware with the IPv6 FIB entry in software.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 FIB entry consistency check.
   ipv6 fib consistency-check enable
   By default, IPv6 FIB entry consistency check is disabled.

Enabling source MAC unknown IPv6 packets to trigger ND snooping

About this task

WLANs allow mobile clients to roam from one AP to another with their IPv6 addresses unchanged. The device might receive source MAC unknown packets from a client that has roamed to one of the APs attached to the device.

By default, the device learns the MAC address of an unknown packet, and then forwards the packet. However, it does not generate an ND entry. Packets returned to the client will be unable to reach the device.

To solve this issue, you can enable this feature to send an NS message out the interface on which a source MAC unknown IPv6 packet is received.

The NS message contains the following address information:

- The target IPv6 address is the source IPv6 address of the data packet.
- The destination IPv6 address is the IPv6 address of the multicast group to which the target IPv6 address belongs.
- The destination MAC address is the multicast MAC address for the multicast group.

When the device receives an NA message in response to the NS message, the AP generates an ND entry.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable source MAC unknown IPv6 packets to trigger ND snooping.

   ipv6 nd unknown-source-mac-probing enable

   By default, source MAC unknown IPv6 packets cannot trigger ND snooping.

Enabling HopLimit expiry alarm generation

About this task

NOTE:

A HopLimit expiry packet is an IP packet with a HopLimit value less than or equal to one.

When the device has a loop or is under HopLimit expiry attacks, it receives a large number of HopLimit expiry packets in a short period. The number of HopLimit expiry packets might grow exponentially. Processing those packets will occupy the resources used for processing service packets.

To avoid the above issue, use this feature to check for loops or HopLimit expiry attacks. The device periodically obtains the number of packet bytes with HopLimit expiry errors at the specified intervals. If the number of bytes in HopLimit expiry packets received during an interval equals or exceeds that during the previous interval multiplied by the specified alarm threshold, the device will generate a HopLimit expiry alarm. To have the device notify HopLimit expiry alarm clearance, specify the **no-drop-notify** keyword.

Procedure

1. Enter system view.

   system-view
2. Enable the device to generate HopLimit expiry alarms.

   ipv6 hoplimit-expired drop-rate alarm enable [**interval** *interval-value*] [**threshold** *threshold-value*] [**no-drop-notify** [**cycles** *resume-value*]]

   By default, the device supports generating HopLimit expiry alarms.

Enabling SNMP notifications for the IPv6 basics module

About this task

This feature enables the IPv6 basics module to generate SNMP notifications for critical events. The SNMP notifications will be sent to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

- If you specify the **interface-state-change** keyword, the IPv6 basics module will send SNMP notifications for IPv6 interface state changes. When an IPv6 interface has a state change, the device sends information about the interface and status of the interface in SNMP notifications to the SNMP module.
- If you specify the **protocol-enable-failure** keyword, the IPv6 basics module will send SNMP notifications for interface-level IPv6 enabling failures. When an IPv6 interface has an IPv6 enabling failure, the device sends the name, chassis number, and slot number of the interface in SNMP notifications to the SNMP module.

Restrictions and guidelines

To enable or disable all SNMP notifications for the IPv6 basics module, do not specify any parameters when you use the **snmp-agent trap enable ipv6 address** command.

After you disable SNMP notifications for the IPv6 basics module, the IPv6 basics module will not send any SNMP notifications to the SNMP module. The device can report the critical events of the IPv6 basics module only by sending log messages to the information center. For log messages to be sent correctly, configure the information center to set log message filtering and output rules, including output destinations. For information about the log destination and output rule configuration in the information center, see information center configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for the IPv6 basics module.

   snmp-agent trap enable ipv6 address [**interface-state-change** | **protocol-enable-failure**] *

   By default, SNMP notifications are enabled for the IPv6 basics module.

   To enable all SNMP notifications for the IPv6 basics module, do not specify any parameters.

Enabling SNMP notifications for the IPv6 forwarding module

About this task

This feature enables the IPv6 forwarding module to generate SNMP notifications for critical events. The SNMP notifications are delivered to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Enable SNMP notifications for the IPv6 forwarding module.

   snmp-agent trap enable ipv6-forwarding

   By default, SNMP notifications are enabled for the IPv6 forwarding module.

Display and maintenance commands for IPv6 basics

Execute **display** commands in any view and **reset** commands in user view.

For information about the **display tcp statistics**, **display udp statistics**, **reset tcp statistics**, and **reset udp statistics** command, see the IP performance commands in *Layer 3—IP Services Command Reference*.

Task	Command
Display IPv6 FIB entries.	**display ipv6 fib** [**vpn-instance** *vpn-instance-name*] [*ipv6-address* [*prefix-length*]]
Display the IPv6 FIB table usage.	**display ipv6 fib usage**
Display ICMPv6 traffic statistics.	**display ipv6 icmp statistics** [**slot** *slot-number*]
Display IPv6 information about the interface.	**display ipv6 interface** [*interface-type* [*interface-number*]] [**brief** [**description**]]
Display IPv6 prefix information about the interface.	**display ipv6 interface** *interface-type interface-number* **prefix**
Display the IPv6 path MTU information.	**display ipv6 pathmtu** [**vpn-instance** *vpn-instance-name*] { *ipv6-address* | { **all** | **dynamic** | **static** } [**count**] }
Display the IPv6 prefix information.	**display ipv6 prefix** [*prefix-number*]
Display brief information about IPv6 RawIP connections.	**display ipv6 rawip** [**slot** *slot-number*]
Display detailed information about IPv6 RawIP connections.	**display ipv6 rawip verbose** [**slot** *slot-number* [**pcb** *pcb-index*]]
Display IPv6 and ICMPv6 packet statistics.	**display ipv6 statistics** [**slot** *slot-number*]
Display brief information about IPv6 TCP connections.	**display ipv6 tcp** [**slot** *slot-number*]
Display detailed information about IPv6 TCP connections.	**display ipv6 tcp verbose** [**slot** *slot-number* [**pcb** *pcb-index*]]
Display brief information about IPv6 UDP connections.	**display ipv6 udp** [**slot** *slot-number*]
Display detailed information about IPv6 UDP connections.	**display ipv6 udp verbose** [**slot** *slot-number* [**pcb** *pcb-index*]]
Display IPv6 TCP traffic statistics.	**display tcp statistics** [**slot** *slot-number*]

Task	Command
Display IPv6 UDP traffic statistics.	**display udp statistics** [**slot** *slot-number*]
Clear path MTUs.	**reset ipv6 pathmtu** { **all** | **dynamic** | **static** }
Clear IPv6 and ICMPv6 packet statistics.	**reset ipv6 statistics** [**slot** *slot-number*]
Clear IPv6 TCP traffic statistics.	**reset tcp statistics**
Clear IPv6 UDP traffic statistics.	**reset udp statistics**

Basic IPv6 settings configuration examples

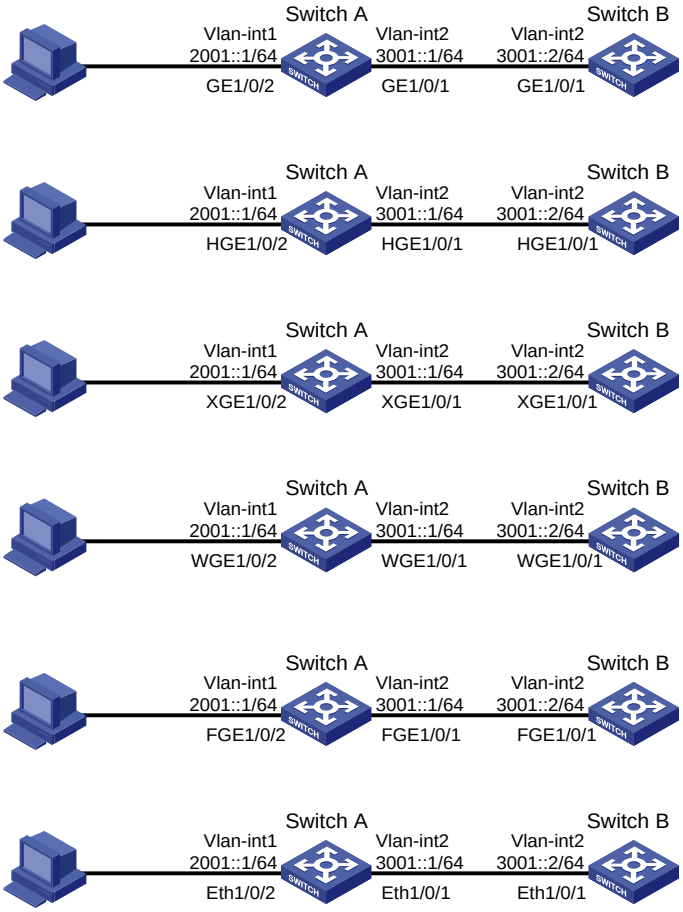
Example: Configuring basic IPv6 settings

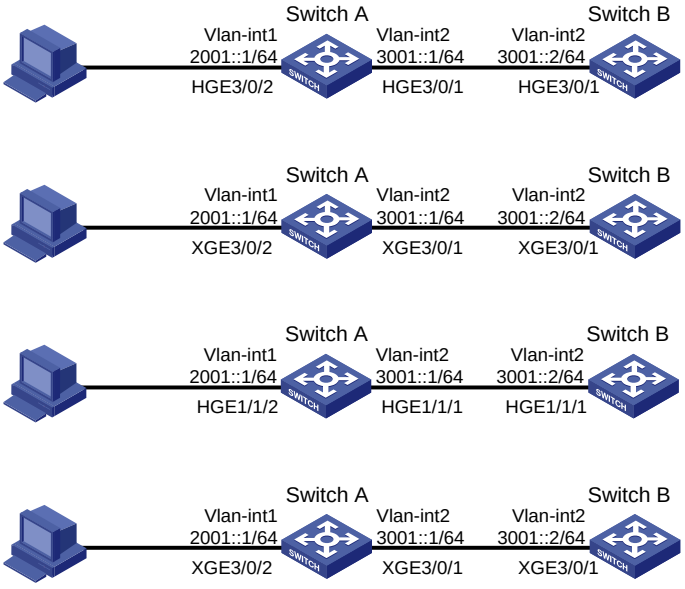
Network configuration

As shown in Figure 4, a host, Switch A, and Switch B are connected through Ethernet ports. Add the Ethernet ports to corresponding VLANs. Configure IPv6 addresses for the VLAN interfaces and verify that they are connected. Switch B can reach the host.

Enable IPv6 on the host to automatically obtain an IPv6 address through IPv6 ND.

Figure 4 Network diagram





Procedure

This example assumes that the VLAN interfaces have been created on the switches.

1. Configure Switch A:

Specify a global unicast address for VLAN-interface 2.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address 3001::1/64
[SwitchA-Vlan-interface2] quit
```

Specify a global unicast address for VLAN-interface 1, and allow it to advertise RA messages (no interface advertises RA messages by default).

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ipv6 address 2001::1/64
[SwitchA-Vlan-interface1] undo ipv6 nd ra halt
[SwitchA-Vlan-interface1] quit
```

2. Configure Switch B:

Configure a global unicast address for VLAN-interface 2.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address 3001::2/64
[SwitchB-Vlan-interface2] quit
```

Configure an IPv6 static route with destination IPv6 address 2001::/64 and next hop address 3001::1.

```
[SwitchB] ipv6 route-static 2001:: 64 3001::1
```

3. Configure the host:

Enable IPv6 for the host to automatically obtain an IPv6 address through IPv6 ND.

Display neighbor information for GigabitEthernet 1/0/2 on Switch A.

```
[SwitchA] display ipv6 neighbors interface gigabitethernet 1/0/2
Type: S-Static    D-Dynamic    O-Openflow     R-Rule    IS-Invalid static
IPv6 address              MAC address    VLAN/VSI   Interface     State T  Aging
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          GE1/0/2       STALE D  1238
```

```
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          GE1/0/2       STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          HGE1/0/2      STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          HGE1/0/2      STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          XGE1/0/2      STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          XGE1/0/2      STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          WGE1/0/2      STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          WGE1/0/2      STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          FGE1/0/2      STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          FGE1/0/2      STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          Eth1/0/2      STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          Eth1/0/2      STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          GE1/0/2       STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          GE1/0/2       STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          XGE3/0/2      STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          XGE3/0/2      STALE D  1248
FE80::215:E9FF:FEA6:7D14  0015-e9a6-7d14 1          XGE3/0/2      STALE D  1238
2001::15B:E0EA:3524:E791  0015-e9a6-7d14 1          XGE3/0/2      STALE D  1248
```

The output shows that the IPv6 global unicast address that Host obtained is 2001::15B:E0EA:3524:E791.

Verifying the configuration

Display the IPv6 interface settings on Switch A. All IPv6 global unicast addresses configured on the interface are displayed.

```
[SwitchA] display ipv6 interface vlan-interface 2
Vlan-interface2 current state: UP
Line protocol current state: UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:2
  Global unicast address(es):
    3001::1, subnet is 3001::/64
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::1:FF00:1
    FF02::1:FF00:2
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 1200000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   25829
  InTooShorts:                  0
  InTruncatedPkts:              0
  InHopLimitExceeds:            0
  InBadHeaders:                 0
  InBadOptions:                 0
  ReasmReqds:                   0
  ReasmOKs:                     0
  InFragDrops:                  0
```

```
  InFragTimeouts:               0
  OutFragFails:                 0
  InUnknownProtos:              0
  InDelivers:                   47
  OutRequests:                  89
  OutForwDatagrams:             48
  InNoRoutes:                   0
  InTooBigErrors:               0
  OutFragOKs:                   0
  OutFragCreates:               0
  InMcastPkts:                  6
  InMcastNotMembers:            25747
  OutMcastPkts:                 48
  InAddrErrors:                 0
  InDiscards:                   0
  OutDiscards:                  0
[SwitchA] display ipv6 interface vlan-interface 1
Vlan-interface1 current state: UP
Line protocol current state: UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:1C0
  Global unicast address(es):
    2001::1, subnet is 2001::/64
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::1:FF00:1
    FF02::1:FF00:1C0
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 1200000 milliseconds
  ND retransmit interval is 1000 milliseconds
  ND advertised reachable time is 0 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 600 seconds
  ND router advertisements live for 1800 seconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   272
  InTooShorts:                  0
  InTruncatedPkts:              0
  InHopLimitExceeds:            0
  InBadHeaders:                 0
  InBadOptions:                 0
  ReasmReqds:                   0
  ReasmOKs:                     0
  InFragDrops:                  0
  InFragTimeouts:               0
  OutFragFails:                 0
```

```
  InUnknownProtos:              0
  InDelivers:                   159
  OutRequests:                  1012
  OutForwDatagrams:             35
  InNoRoutes:                   0
  InTooBigErrors:               0
  OutFragOKs:                   0
  OutFragCreates:               0
  InMcastPkts:                  79
  InMcastNotMembers:            65
  OutMcastPkts:                 938
  InAddrErrors:                 0
  InDiscards:                   0
  OutDiscards:                  0
```

Display the IPv6 interface settings on Switch B. All IPv6 global unicast addresses configured on the interface are displayed.

```
[SwitchB] display ipv6 interface vlan-interface 2
Vlan-interface2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:1234
  Global unicast address(es):
    3001::2, subnet is 3001::/64
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::1:FF00:2
    FF02::1:FF00:1234
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 1200000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   117
  InTooShorts:                  0
  InTruncatedPkts:              0
  InHopLimitExceeds:            0
  InBadHeaders:                 0
  InBadOptions:                 0
  ReasmReqds:                   0
  ReasmOKs:                     0
  InFragDrops:                  0
  InFragTimeouts:               0
  OutFragFails:                 0
  InUnknownProtos:              0
  InDelivers:                   117
  OutRequests:                  83
  OutForwDatagrams:             0
```

```
  InNoRoutes:                  0
  InTooBigErrors:              0
  OutFragOKs:                  0
  OutFragCreates:              0
  InMcastPkts:                 28
  InMcastNotMembers:           0
  OutMcastPkts:                7
  InAddrErrors:                0
  InDiscards:                  0
  OutDiscards:                 0
```

Ping Switch A and Switch B on the host, and ping Switch A and the host on Switch B to verify that they are connected.

NOTE:

When you ping a link-local address, use the **-i** parameter to specify an interface for the link-local address.

```
[SwitchB] ping ipv6 -c 1 3001::1
Ping6(56 data bytes) 3001::2 --> 3001::1, press CTRL+C to break
56 bytes from 3001::1, icmp_seq=0 hlim=64 time=4.404 ms

--- Ping6 statistics for 3001::1 ---
1 packet(s) transmitted, 1 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 4.404/4.404/4.404/0.000 ms
[SwitchB] ping ipv6 -c 1 2001::15B:E0EA:3524:E791
Ping6(56 data bytes) 3001::2 --> 2001::15B:E0EA:3524:E791, press CTRL+C to break
56 bytes from 2001::15B:E0EA:3524:E791, icmp_seq=0 hlim=64 time=5.404 ms

--- Ping6 statistics for 2001::15B:E0EA:3524:E791 ---
1 packet(s) transmitted, 1 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 5.404/5.404/5.404/0.000 ms
```

The output shows that Switch B can ping Switch A and the host. The host can also ping Switch B and Switch A.

Configuring IPv6 neighbor discovery

About IPv6 neighbor discovery

ICMPv6 messages used by IPv6 neighbor discovery

The IPv6 neighbor discovery (ND) process uses ICMP messages for address resolution, neighbor reachability verification, and neighboring device tracking.

Table 3 describes the ICMPv6 messages used by the IPv6 ND protocol.

Table 3 ICMPv6 messages used by ND

ICMPv6 message	Type	Function
Neighbor Solicitation (NS)	135	Acquires the link-layer address of a neighbor on the local link.
		Verifies the reachability of a neighbor.
		Detects duplicate addresses.
Neighbor Advertisement (NA)	136	Responds to an NS message.
		Notifies the neighboring nodes of link layer changes.
Router Solicitation (RS)	133	Requests an address prefix and other configuration information for autoconfiguration after startup.
Router Advertisement (RA)	134	Responds to an RS message.
		Advertises information, such as the Prefix Information options and flag bits.
Redirect	137	Informs the source host of a better next hop on the path to a particular destination when certain conditions are met.

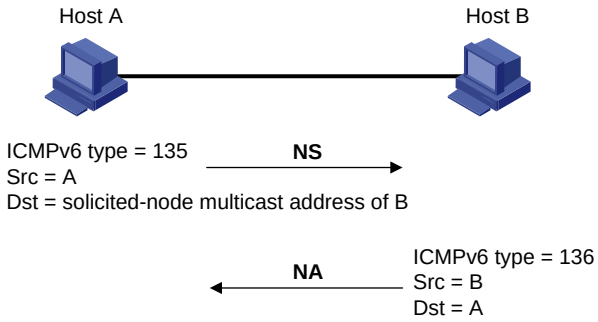
Address resolution

This function is similar to ARP in IPv4. An IPv6 node acquires the link-layer addresses of neighboring nodes on the same link through NS and NA messages.

Figure 5 shows how Host A acquires the link-layer address of Host B on the same link. The address resolution procedure is as follows:

1. Host A multicasts an NS message. The source address of the NS message is the IPv6 address of the sending interface of Host A. The destination address is the solicited-node multicast address of Host B. The NS message body contains the link-layer address of Host A and the target IPv6 address.
2. After receiving the NS message, Host B determines whether the target address of the packet is its IPv6 address. If it is, Host B learns the link-layer address of Host A, and then unicasts an NA message containing its link-layer address.
3. Host A acquires the link-layer address of Host B from the NA message.

Figure 5 Address resolution



Neighbor reachability detection

After Host A acquires the link-layer address of its neighbor Host B, Host A can use NS and NA messages to test the reachability of Host B as follows:

1. Host A sends an NS message whose destination address is the IPv6 address of Host B.
2. If Host A receives an NA message from Host B, Host A decides that Host B is reachable. Otherwise, Host B is unreachable.

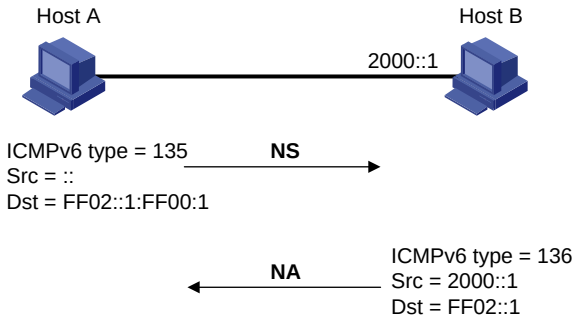
Duplicate address detection

After Host A acquires an IPv6 address, it performs Duplicate Address Detection (DAD) to check whether the address is being used by any other node. This is similar to gratuitous ARP in IPv4. DAD is accomplished through NS and NA messages.

The DAD procedure is as follows:

1. Host A sends an NS message. The source address is the unspecified address and the destination address is the corresponding solicited-node multicast address of the IPv6 address to be detected. The NS message body contains the detected IPv6 address.
2. If Host B uses this IPv6 address, Host B returns an NA message that contains its IPv6 address.
3. Host A knows that the IPv6 address is being used by Host B after receiving the NA message from Host B. If receiving no NA message, Host A decides that the IPv6 address is not in use and uses this address.

Figure 6 Duplicate address detection



Router/prefix discovery and stateless address autoconfiguration

Router/prefix discovery allows an IPv6 node to find the neighboring routers and learn the prefix and network configuration parameters of the network from receiving RA messages.

Stateless address autoconfiguration allows an IPv6 node to automatically generate an IPv6 address based on the information learned through router/prefix discovery.

A node performs router/prefix discovery and stateless address autoconfiguration as follows:

1. At startup, a node sends an RS message to request configuration information from a router.
2. The router returns an RA message containing the Prefix Information option and other configuration information. (The router also periodically sends an RA message.)
3. The node automatically generates an IPv6 address and other configuration parameters according to the configuration information in the RA message.

The Prefix Information option contains an address prefix and the preferred lifetime and valid lifetime of the address prefix. A node updates the preferred lifetime and valid lifetime upon receiving a periodic RA message.

The generated IPv6 address is valid within the valid lifetime and becomes invalid when the valid lifetime expires.

After the preferred lifetime expires, the node cannot use the generated IPv6 address to establish new connections, but can receive packets destined for the IPv6 address. The preferred lifetime cannot be greater than the valid lifetime.

Redirection

Upon receiving a packet from a host, the gateway sends an ICMPv6 redirect message to inform the host of a better next hop when the following conditions are met:

- The interface receiving the packet is the same as the interface forwarding the packet.
- The selected route is not created or modified by an ICMPv6 redirect message.
- The selected route is not a default route on the device.
- The forwarded IPv6 packet does not contain the routing extension header.

Protocols and standards

- RFC 4861, *Neighbor Discovery for IP Version 6 (IPv6)*
- RFC 8106, *IPv6 Router Advertisement Options for DNS Configuration*

IPv6 neighbor discovery tasks at a glance

All IPv6 neighbor discovery tasks are optional.

- Configuring ND entry related features
  - Configuring a static neighbor entry
  - Setting the dynamic neighbor learning limit
  - Enabling unsolicited NA learning
  - Setting the aging timer for ND entries in stale state
  - Setting ND entry probing parameters
  - Minimizing link-local ND entries

- Setting the hop limit
- Configuring RA message sending and parameters
- Setting the maximum number of attempts to send an NS message for DAD
- Enabling duplicate address detection
- Configuring ND snooping
  - Configuring ND snooping in a VLAN
  - Configuring ND snooping in a VXLAN
- Enabling ND proxy
- Enabling DAD message forwarding for a VSI
- Configuring IPv6 ND direct route advertisement
- Enabling ND direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state
- Adding prefix information advertised through RA messages to IPSG binding entries
- Configuring user information recording
  - Enabling recording user IPv6 address conflicts
  - Enabling recording user port migrations
  - Enabling ND logging for user online and offline events
- Logging ND entry deployment events
- Enabling ND entry consistency check
- Disabling the device from aging or deleting ND entries in response to network topology changes
- Enabling recording ND entry learning events
- Disabling sending NS messages when data packets trigger ND resolution
- Performing ND-ping

Configuring a static neighbor entry

About this task

A neighbor entry stores information about a link-local node. The entry can be created dynamically through NS and NA messages, or configured statically.

The device uniquely identifies a static neighbor entry by using the neighbor's IPv6 address and the number of the Layer 3 interface that connects to the neighbor. You can configure a static neighbor entry by using one of the following methods:

- **Method 1**—Associate a neighbor's IPv6 address and link-layer address with the local Layer 3 interface.
- **Method 2**—Associate a neighbor's IPv6 address and link-layer address with a Layer 2 port in a VLAN.
- **Method 3**—Specify a neighbor IPv6 address, MAC address, input interface (VSI interface), output interface (tunnel interface), and VSI name.
- **Method 4**—Specify a neighbor IPv6 address, MAC address, input interface (VSI interface), output interface (determined by a Layer 2 interface and Ethernet service instance), and VSI name.

Restrictions and guidelines

To configure a static neighbor entry for a VLAN interface, use Method 1 or Method 2.

- If you use Method 1, the device is required to resolve the Layer 2 port in the related VLAN.

- If you use Method 2, make sure the Layer 2 port belongs to the specified VLAN and the corresponding VLAN interface already exists. After the configuration, the device associates the VLAN interface with the neighbor IPv6 address to identify the static neighbor entry.

If the device and its neighbor are connected through a VSI interface, use Method 3 or Method 4 to configure the neighbor entry.

- If Method 3 is used, the neighbor entry is in REACH state. This method is applicable to the network where VXLAN gateways are connected through tunnel interfaces. In the network, a VXLAN gateway is identified by both the VSI and VSI interface. A VSI interface is associated with multiple tunnel interfaces. To create a neighbor entry, you must specify the VSI interface, VSI, and tunnel interface.
- If Method 4 is used, the neighbor entry is in REACH state. This method is applicable to the network where VXLAN gateways are associated with local sites. A VXLAN gateway is identified by both the VSI and VSI interface. One VXLAN gateway might have multiple local sites. Local sites access the VXLAN network through Layer 2 interfaces where Ethernet service instance and VSI mappings are configured. To create a neighbor entry, you must specify the VSI interface, Layer 2 interface connected to the local site, Ethernet service instance, and VSI.

For more information about VSI, VSI interfaces, and Ethernet service instances, see VXLAN overview in *VXLAN Configuration Guide*.

For more information about tunnel interfaces, see "Configuring tunneling."

To delete a static neighbor entry for a VSI interface, specify only the VSI interface.

To delete a static neighbor entry for a VLAN interface, specify only the VLAN interface.

You can use the **undo ipv6 neighbor** command to delete both static and dynamic neighbor entries.

Procedure

1. Enter system view.

   system-view

2. Configure a static neighbor entry.

   ipv6 neighbor *ipv6-address mac-address* { *vlan-id port-type port-number* | **interface** *interface-type interface-number* | **vsi-interface** *vsi-interface-id* **tunnel** *number* **vsi** *vsi-name* | **vsi-interface** *vsi-interface-id interface-type interface-number* **service-instance** *instance-id* **vsi** *vsi-name* } [**vpn-instance** *vpn-instance-name*]

   By default, no static neighbor entries exist.

Setting the dynamic neighbor learning limit

About this task

The device can dynamically acquire the link-layer address of a neighboring node through NS and NA messages and add it into the neighbor table. A large neighbor table degrades the forwarding performance. To avoid excessive resource consumption by neighbor entries, use one of the following methods:

- Set the maximum number of dynamic neighbor entries that an interface can learn. When the number of dynamic neighbor entries reaches the limit, the interface stops learning neighbor information.
- Set the maximum number of dynamic neighbor entries that the device can learn. When the number of dynamic neighbor entries reaches the limit, the device stops learning neighbor information.

The number of dynamic neighbor entries learnt by the interfaces on a device cannot exceed the maximum number of dynamic neighbor entries that the device can learn.

Setting the maximum number of dynamic neighbor entries on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the dynamic neighbor learning limit on the interface.
   ipv6 neighbors max-learning-num *max-number*
   By default, an interface can learn a maximum of 12288 dynamic neighbor entries.

Setting the maximum number of dynamic neighbor entries on the device

1. Enter system view.
   system-view
2. Set the dynamic neighbor learning limit on the device.
   ipv6 neighbors max-learning-number *max-number* **slot** *slot-number* [**cpu** *cpu-number*]
   By default, the device can learn a maximum of 12288 dynamic neighbor entries.
   To disable the device from learning dynamic neighbor entries, set the value for the *max-number* argument to 0.

Enabling unsolicited NA learning

About this task

On some networks, a server multicasts NA messages to two peer devices for link backup. The peer devices cannot learn ND entry for the server from these NA messages by default. If no ND learning is triggered by data exchange between the server and peer devices, the peer devices learn the entry for the server only when the server unicasts messages to them.

This feature enables an interface to learn ND entries from unsolicited NA messages. The ND entries generated by using this method are in stale state.

Restrictions and guidelines

To ensure that the device learns ND entries from trusted NA messages, enable this feature only on a secure network.

This feature might cause the device to learn excessive ND entries that consume too many system resources. As a best practice, execute the **ipv6 neighbor stale-aging** command to set a smaller aging timer before you enable this feature. The smaller aging timer accelerates the aging of ND entries in stale state.

This feature is available only on Layer 3 interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Enable unsolicited NA learning.

ipv6 nd unsolicited-na-learning enable

By default, unsolicited NA learning is disabled.

Setting the aging timer for ND entries in stale state

About this task

ND entries in stale state have an aging timer. If an ND entry in stale state is not refreshed before the timer expires, the ND entry changes to the delay state. If it is still not refreshed in 5 seconds, the ND entry changes to the probe state, and the device sends an NS message three times. If no response is received, the device deletes the ND entry.

Restrictions and guidelines

You can set the aging timer for ND entries in stale state in system view and interface view. For ND entries in stale state on an interface, the aging timer in interface view has higher priority than the aging timer in system view.

Procedure

1. Enter system view.

   system-view

2. Set the aging timer for ND entries in stale state.
   - Set the aging timer for ND entries in stale state in system view.

     ipv6 neighbor stale-aging { *aging-minutes* | **second** *aging-seconds* }

     The default setting is 240 minutes.
   - Execute the following commands in sequence to set the aging timer in interface view.

     interface *interface-type interface-number*

     ipv6 neighbor timer stale-aging { *aging-minutes* | **second** *aging-seconds* }

     By default, the aging timer of ND entries in stale state is not configured on an interface. The aging timer is determined by the configuration of the **ipv6 neighbor stale-aging** command in system view.

Setting ND entry probing parameters

About this task

After an ND entry changes to PROBE state, the device sends an NS message to test the reachability of the neighbor in the ND entry. A maximum of three attempts is allowed by default. The device handles the ND entry as follows:

- If the device does not receive an NA message from the neighbor, it deletes the ND entry.
- If the device receives an NA message from the neighbor, it sets the ND entry state to REACHABLE.

Perform this task to set the maximum number of probes and the probe interval for ND entries.

Restrictions and guidelines

When you modify the probe count, the modification does not take effect on ND entries in PROBE state. Neighbor entries in DELAY state will adopt this setting when they enter into the PROBE state.

When you modify the probe interval, the modification takes effect immediately.

Procedure

1. Enter system view.

system-view

system-view

2. Set the maximum number of probes to test the reachability of neighbors in ND entries.

   ipv6 neighbor aging probe-count *count*

   By default, the device performs a maximum of three probes to test the reachability of neighbors in ND entries.

3. Set the interval for testing the reachability of neighbors in ND entries.

   ipv6 neighbor aging probe-interval *interval*

   By default, the interval for testing the reachability of neighbors in ND entries is the same as the interval for retransmitting an NS message.

Minimizing link-local ND entries

About this task

Perform this task to minimize link-local ND entries assigned to the hardware. Link-local ND entries refer to ND entries that contain link-local addresses.

By default, the device assigns all ND entries to the hardware. With this feature enabled, the newly learned link-local ND entries are not assigned to the hardware if the link-local addresses of the entries are not the next hops of any routes. This feature saves hardware resources.

This feature takes effect only on newly learned link-local ND entries.

Procedure

1. Enter system view.

   system-view

2. Minimize link-local ND entries.

   ipv6 neighbor link-local minimize

   By default, the device assigns all ND entries to the hardware.

Setting the hop limit

About this task

You can set the hop limit value to fill in the Hop Limit field for IPv6 packets to be sent.

Procedure

1. Enter system view.

   system-view

2. Set the value for the Hop Limit field in the IP header.

   ipv6 hop-limit *value*

   The default setting is 64.

Configuring RA message sending and parameters

About RA message parameters

You can enable an interface to send RA messages, and configure the interval for sending RA messages and parameters in RA messages. After receiving an RA message, a host can use these

parameters to perform corresponding operations. Table 4 describes the configurable parameters in an RA message.

Table 4 Parameters in an RA message and their descriptions

Parameter	Description
Hop Limit	Maximum number of hops in RA messages. A host receiving the RA message fills the value in the Hop Limit field of sent IPv6 packets.
Prefix information	After receiving the prefix information, the hosts on the same link can perform stateless autoconfiguration.
MTU	Guarantees that all nodes on the link use the same MTU.
Boot file URL	Specifies the URL address for downloading the boot file in RA messages. The device can use the ND protocol to obtain both the IPv6 address and the boot file URL for automatic configuration instead of using DHCPv6.
M flag	Determines whether a host uses stateful autoconfiguration to obtain an IPv6 address. If the M flag is set to 1, the host uses stateful autoconfiguration (for example, from a DHCPv6 server) to obtain an IPv6 address. Otherwise, the host uses stateless autoconfiguration to generate an IPv6 address according to its link-layer address and the prefix information in the RA message.
O flag	Determines whether a host uses stateful autoconfiguration to obtain configuration information other than the IPv6 address. If the O flag is set to 1, the host uses stateful autoconfiguration (for example, from a DHCPv6 server) to obtain configuration information other than the IPv6 address. Otherwise, the host uses stateless autoconfiguration.
Router Lifetime	Tells the receiving hosts how long the advertising router can live. If the lifetime of a router is 0, the router cannot be used as the default gateway.
Retrans Timer	If the device does not receive a response message within the specified time after sending an NS message, it retransmits the NS message.
Reachable Time	If the neighbor reachability detection shows that a neighbor is reachable, the device considers the neighbor reachable within the specified reachable time. If the device needs to send a packet to the neighbor after the specified reachable time expires, the device reconfirms whether the neighbor is reachable.
Router Preference	Specifies the router preference in an RA message. A host selects a router as the default gateway according to the router preference. If router preferences are the same, the host selects the router from which the first RA message is received.
DNS server option	DNS server information for IPv6 hosts. Hosts can obtain DNS server information from received RA messages instead of using DHCPv6.
DNS suffix information in DNS Search List (DNSSL) option	DNS suffix information for IPv6 hosts. Hosts can obtain DNS suffix information from received RA messages instead of using DHCPv6.

Restrictions and guidelines

The maximum interval for sending RA messages should be less than (or equal to) the router lifetime in RA messages. In this way, the router can be updated by an RA message before expiration.

The values of the NS retransmission timer and the reachable time configured for an interface are sent in RA messages to hosts. This interface sends NS messages at the interval of the NS retransmission timer and considers a neighbor reachable within the reachable time.

Enabling the sending of RA messages

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the sending of RA messages.

   undo ipv6 nd ra halt

   The default setting is disabled.

4. Set the maximum and minimum intervals for sending RA messages.

   ipv6 nd ra interval *max-interval min-interval*

   By default, the maximum interval for sending RA messages is 600 seconds, and the minimum interval is 200 seconds.

   The device sends RA messages at random intervals between the maximum interval and the minimum interval.

   The minimum interval should be less than or equal to 0.75 times the maximum interval.

Configuring parameters for RA messages

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the prefix information in RA messages.

   ipv6 nd ra prefix { *ipv6-prefix prefix-length* | *ipv6-prefix*/*prefix-length* } [*valid-lifetime preferred-lifetime* [**no-autoconfig** | **off-link** | **prefix-preference** *level*] * | **no-advertise**]

   By default, no prefix information is configured for RA messages, and the IPv6 address of the interface sending RA messages is used as the prefix information. If the IPv6 address is manually configured, the prefix uses a fixed valid lifetime of 2592000 seconds (30 days) and a preferred lifetime of 604800 seconds (7 days). If the IPv6 address is automatically obtained, the prefix uses the valid lifetime and preferred lifetime configured for the IPv6 address.

4. Configure the default settings for prefixes advertised in RA messages.

   ipv6 nd ra prefix default [*valid-lifetime preferred-lifetime* [**no-autoconfig** | **off-link**] * | **no-advertise**]

   By default, no default settings are configured for prefixes advertised in RA messages.

5. Turn off the MTU option in RA messages.

   ipv6 nd ra no-advlinkmtu

   By default, RA messages contain the MTU option.

6. Specify unlimited hops in RA messages.

   ipv6 nd ra hop-limit unspecified

   By default, the maximum number of hops in RA messages is 64.

7. Specify the URL of the boot file in RA messages.

   ipv6 nd ra boot-file-url *url-string*

   By default, RA messages do not contain the URL of the boot file.

8. Set the M flag bit to 1.

   ipv6 nd autoconfig managed-address-flag

   By default, the M flag bit is set to 0 in RA advertisements. Hosts receiving the advertisements will obtain IPv6 addresses through stateless autoconfiguration.

9. Set the O flag bit to 1.

   ipv6 nd autoconfig other-flag

   By default, the O flag bit is set to 0 in RA advertisements. Hosts receiving the advertisements will acquire other configuration information through stateless autoconfiguration.

10. Set the router lifetime in RA messages.

    ipv6 nd ra router-lifetime *time*

    By default, the router lifetime is three times as long as the maximum interval for advertising RA messages.

11. Set the NS retransmission timer.

    ipv6 nd ns retrans-timer *value*

    By default, an interface sends NS messages every 1000 milliseconds, and the value of the Retrans Timer field in RA messages is 0.

12. Set the router preference in RA messages.

    ipv6 nd router-preference { **high** | **low** | **medium** }

    By default, the router preference is medium.

13. Set the reachable time.

    ipv6 nd nud reachable-time *time*

    By default, the neighbor reachable time is 1200000 milliseconds, and the value of the Reachable Time field in sent RA messages is 0.

Specifying DNS server information in RA messages

About this task

The DNS server options in RA messages provide DNS server information for IPv6 hosts. The RA messages allow hosts to obtain their IPv6 addresses and the DNS server through stateless autoconfiguration. This method is useful in a network where DHCPv6 infrastructure is not provided.

One DNS server option contains one DNS server. All DNS server options are sorted in ascending order of the DNS server sequence number.

After you execute the **ipv6 nd ra dns server** command, the device immediately sends an RA message with the existing and newly specified DNS server information.

After you execute the **undo ipv6 nd ra dns server** command, the device immediately sends two RA messages.

- The first RA message contains information about all DNS servers, including the DNS servers specified in the **undo** command with their lifetime set to 0 seconds.
- The second RA message contains information about remaining DNS servers.

Each time the device sends an RA message from an interface, it immediately refreshes the RA message advertisement interval for that interface.

Restrictions and guidelines

You can configure a maximum of eight DNS servers on an interface.

The default lifetime of a DNS server is three times the maximum interval for advertising RA messages. To set the maximum interval, use the **ipv6 nd ra interval** command.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify DNS server information to be advertised in RA messages.

   ipv6 nd ra dns server *ipv6-address* [*seconds* | **infinite**] **sequence** *seqno*

   By default, no DNS server information is specified and RA messages do not contain DNS server options.

Specifying DNS suffix information in RA messages

About this task

The DNSSL option in RA messages provides suffix information for IPv6 hosts. The RA messages allow hosts to obtain their IPv6 addresses and the DNS suffix through stateless autoconfiguration. This method is useful in a network where DHCPv6 infrastructure is not provided.

One DNSSL option contains one DNS suffix. All DNSSL options are sorted in ascending order of the sequence number of the DNS suffix.

After you execute the **ipv6 nd ra dns search-list** command, the device immediately sends an RA message with the existing and newly specified DNS suffix information.

After you execute the **undo ipv6 nd ra dns search-list** command, the device immediately sends two RA messages.

- The first RA message contains information about all DNS suffixes, including DNS suffixes specified in the **undo** command with their lifetime set to 0 seconds.
- The second RA message contains information about remaining DNS suffixes.

Each time the device sends an RA message from an interface, it immediately refreshes the RA message advertisement interval for that interface.

Restrictions and guidelines

You can configure a maximum of eight DNS suffixes on an interface.

The default lifetime of a DNS suffix is three times the maximum interval for advertising RA messages. To set the maximum interval, use the **ipv6 nd ra interval** command.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify DNS suffix information to be advertised in RA messages.

   ipv6 nd ra dns search-list *domain-name* [*seconds* | **infinite**] **sequence** *seqno*

   By default, no DNS suffix information is specified and RA messages do not contain DNS suffix options.

Suppressing advertising DNS information in RA messages

About this task

Perform this task to suppress the device from advertising information about DNS server addresses and DNS suffixes in RA messages.

Whether enabling this feature on an interface will trigger sending RA message immediately depends on the interface configuration:

- If the interface has DNS server information configured or has obtained an AAA-authorized DNS server address, the device immediately sends two RA messages. In the first message, the lifetime for DNS server addresses is 0 seconds. The second RA message does not contain any DNS server options.
- If the interface has no DNS server information specified or no AAA-authorized DNS server address assigned, no RA messages are triggered.

If you specify a new DNS server or remove a DNS server on the interface after enabling DNS server suppression, the device immediately sends an RA message without any DNS server options.

Whether disabling this feature on an interface will trigger sending RA message immediately depends on the interface configuration:

- If the interface has DNS server information configured or has obtained an AAA-authorized DNS server address, the device immediately sends an RA message containing the DNS server information.
- If the interface has no DNS server information specified or no AAA-authorized DNS server address assigned, no RA messages are triggered.

Each time the device sends an RA message from an interface, it immediately refreshes the RA message advertisement interval for that interface.

The same suppression mechanism applies when you enable or disable DNS suffix suppression in RA messages.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable DNS server suppression in RA messages.

   ipv6 nd ra dns server suppress

   By default, DNS server suppression in RA messages is disabled.
4. Enable DNS suffix suppression in RA messages.

   ipv6 nd ra dns search-list suppress

   By default, DNS suffix suppression in RA messages is disabled.

Setting the maximum number of attempts to send an NS message for DAD

About this task

An interface sends an NS message for DAD for an obtained IPv6 address. The interface resends the NS message if it does not receive a response within the time specified by the **ipv6 nd ns**

retrans-timer command. If the interface receives no response after making the maximum attempts specified by the **ipv6 nd dad attempts** command, the interface uses the IPv6 address.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the number of attempts to send an NS message for DAD.
   ipv6 nd dad attempts *times*
   The default setting is 1. When the *times* argument is set to 0, DAD is disabled.

Enabling duplicate address detection

About this task

If the device detects that an IPv6 address on an interface has been used on the network, the device marks that IPv6 address as duplicate. The interface cannot use the address for communication.

By default, an interface does not perform duplicate address detection. Once an IPv6 address is marked as duplicate on an interface, it will be unusable even after it becomes unique on the link later.

To resolve this issue, enable duplicate address detection. This feature sends NS messages to the duplicate address at random intervals until it does not receive an NA response message from that address or until duplicate detection is disabled for duplicate addresses. For more information about the detection process, see "Duplicate address detection."

You can set the maximum duplicate detection interval for duplicate addresses. After the device marks a detected address as duplicate, it waits for a random amount of time between 1 and the maximum detection interval. Then, the device resends an NS message to the solicited-node multicast address of the duplicate address. This mechanism helps reduce the risk of congestion that results from the NS messages sent for duplicate detection.

Procedure

1. Enter system view.
   system-view
2. Enable duplicate address detection.
   ipv6 address duplicate-detect enable
   By default, duplicate address detection is disabled.
3. (Optional.) Set the maximum duplicate detection interval for duplicate addresses.
   ipv6 address duplicate-detect interval *interval*
   The default maximum interval is 5 seconds.

Configuring ND snooping in a VLAN

About ND snooping in a VLAN

The ND snooping feature is used in Layer 2 switching networks. It learns the source MAC addresses, source IPv6 addresses, input interfaces, and VLANs of arriving ND messages and data packets to build the ND snooping table.

ND snooping entries can be used by ND detection to prevent spoofing attacks.

ND detection processes the ND messages received on ND trusted and untrusted interfaces as follows:

- ND detection forwards all ND messages received on an ND trusted interface.
- ND detection compares all ND messages received on an ND untrusted interface with the ND snooping entries except for RA and redirect messages.

You can use the **ipv6 nd detection trust** command to specify a Layer 2 Ethernet or aggregate port as an ND trusted interface. For more information about the **ipv6 nd detection trust** command, see *Security Command Reference*.

ND snooping entries can be used by IPv6 source guard to prevent spoofing attacks. For more information about IPv6 source guard, see *Security Configuration Guide*.

ND snooping provides device liveness tracking so that the ND snooping table can be updated in a timely manner. After ND snooping is enabled for a VLAN, the device uses the following mechanisms to create, update, and delete ND snooping entries. The following example uses ND messages for illustration.

Creation of ND snooping entries

Upon receiving an ND message or data packet from an unknown source, the device creates an ND snooping entry in INVALID status and performs DAD for the source IPv6 address. The device sends NS messages out of the ND trusted interfaces in the receiving VLAN twice. The sending interval is set by the **ipv6 nd snooping dad retrans-timer** command.

- If the device does not receive an NA message within the invalid entry lifetime (set by the **ipv6 nd snooping lifetime invalid** command), the entry becomes valid.
- If the device receives an NA message within the invalid entry lifetime, it deletes this entry.

Updating of ND snooping entries

When the ND untrusted interface that receives an ND message is different from that in the entry for an IPv6 address, the device performs DAD for the entry. It sends NS messages twice. The sending interval is set by the **ipv6 nd snooping dad retrans-timer** command.

- If the device does not receive an NA message within the invalid entry lifetime, it updates the entry with the new receiving interface.
- If the device receives an NA message within the invalid entry lifetime, the ND snooping entry remains unchanged.

Deletion of ND snooping entries

- When an ND trusted interface in the VLAN receives an ND message from the IPv6 address in a learned ND snooping entry, it performs DAD for the entry. The device sends NS messages twice. The sending interval is set by the **ipv6 nd snooping dad retrans-timer** command.
  - If the device does not receive an NA message within the invalid entry lifetime, it deletes the entry.
  - If the device receives an NA message within the invalid entry lifetime, the ND snooping entry remains unchanged.
- If an ND snooping entry has no matching ND messages within the valid entry lifetime (set by the **ipv6 nd snooping lifetime valid** command), the entry becomes invalid. The device then performs DAD for the entry by sending NS messages out of the interface in the entry. The device sends NS messages twice by default. You can use the **ipv6 nd snooping aging** command to set the sending interval and probe times.
  - If the device does not receive an NA message within the invalid entry lifetime, it deletes the entry.
  - If the device receives an NA message within the invalid entry lifetime, the ND snooping entry remains unchanged and becomes valid.

Enabling ND snooping

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ND snooping for IPv6 addresses. Choose the options to configure as needed:
   - Enable ND snooping for global unicast addresses.
     ipv6 nd snooping enable global
   - Enable ND snooping for link-local addresses.
     ipv6 nd snooping enable link-local

   By default, ND snooping is disabled for IPv6 global unicast addresses and link-local addresses.
4. (Optional.) Enable ND snooping for data packets from unknown sources.
   ipv6 nd snooping glean source

   By default, ND snooping is disabled for data packets from unknown sources.

   After executing this command for a VLAN, you must configure IPv6 source guard on all untrusted interfaces in the VLAN to ensure correct forwarding of the data packets received on these interfaces.

Specifying an ACL for ND snooping entry filtering

About this task

ND snooping is used in networks deployed with a Layer 2 switch. The device creates ND snooping entries by listening to ND or IPv6 messages. To control network traffic and ensure the effective use of network resources, you can specify an ACL to filter ND snooping entries.

When the device creates an ND snooping entry, it matches the specified ACL against the source IPv6 address in the related ND or IPv6 message.

- If the source IPv6 address hits a permit rule in the ACL, the device continues to create that ND snooping entry.
- If the source IPv6 address cannot hit a permit rule in the ACL, the device does not create that ND snooping entry.

Restrictions and guidelines

After you specify a new ACL for ND snooping entry filtering in a VLAN, the device will match the new ACL against all existing ND snooping entries in that VLAN:

- If the source IPv6 address in an ND snooping entry hits a permit rule in the ACL, the device will retain the ND snooping entry.
- If the source IPv6 address in an ND snooping entry cannot hit a permit rule in the ACL, the device will mark the ND snooping entry stale and delete it immediately.
- If the ACL specified in a VLAN does not exist or does not have rules, the device will delete all ND snooping entries from that VLAN.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Specify an ACL for ND snooping entry filtering.

ipv6 nd snooping match acl { *ipv6-acl-number* | **name** *ipv6-acl-name* }

By default, no ACL is specified for ND snooping entry filtering.

Setting the ND snooping entry learning limit

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet or Layer 2 aggregate interface view.

   interface *interface-type interface-number*

3. (Optional.) Set the ND snooping entry learning limit on the interface.

   ipv6 nd snooping max-learning-num *max-number*

   By default, an interface can learn a maximum of 6144 ND snooping entries.

Configuring ND snooping entry related parameters

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet or Layer 2 aggregate interface view.

   interface *interface-type interface-number*

3. (Optional.) Configure the port as an ND snooping uplink port. The ND snooping uplink port cannot learn ND snooping entries.

   ipv6 nd snooping uplink

   By default, the port is not an ND snooping uplink port. After ND snooping is enabled, the port can learn ND snooping entries.

4. Return to system view.

   quit

5. (Optional.) Set timeout timers for ND snooping entries.

   ipv6 nd snooping lifetime { **invalid** *invalid-lifetime* | **valid** *valid-lifetime* }

   The default settings are as follows:

   - The timeout timer for ND snooping entries in INVALID status (TENTATIVE, TESTING_TPLT, or TESTING_VP) is 500 milliseconds.
   - The timeout timer for ND snooping entries in VALID status is 300 seconds.

6. (Optional.) Set the DAD NS message retransmission interval for ND snooping entry creation or update.

   ipv6 nd snooping dad retrans-timer *interval*

   The default value is 250 milliseconds.

7. (Optional.) Set the probe times and interval for the device to send NS messages for aging ND snooping entries.

   ipv6 nd snooping aging { *probe-count count* | *probe-interval interval* }

   By default, the device sends two NS messages at an interval of 250 milliseconds for aging ND snooping entries.

Configuring ND snooping in a VXLAN

About ND snooping in a VXLAN

The ND snooping feature in a VXLAN learns the source MAC addresses, source IPv6 addresses, VSI name, and link IDs to build the ND snooping table. For more information about VXLAN, see VXLAN overview in *VXLAN Configuration Guide*.

ND snooping entries can be used by ND detection to prevent spoofing attacks. For more information about the ND attack detection, see *Security Configuration Guide*.

ND snooping entries can be used by IPv6 source guard to prevent spoofing attacks. For more information about IPv6 source guard, see *Security Configuration Guide*.

After ND snooping is enabled in a VSI, the device uses the following mechanisms to create and maintain ND snooping entries.

Creation of ND snooping entries

With this feature enabled in a VXLAN, all receiving ND messages are delivered to the CPU. The CPU obtains the source IPv6 addresses, source MAC addresses, VSI name, and link IDs to create ND snooping entries.

Aging of ND snooping entries

The aging time of an ND snooping entry is 5 minutes. If no matching packet is received within 30 seconds before the end of the entry aging time, the entry is set to TENTATIVE status. At the same time, the device sends a NS message for the entry. If the source IP and source MAC in the received ND message matches the ND snooping entry, the device sets the entry to VALID and resets the aging time.

When the aging time of an ND snooping entry is reached, the entry is deleted.

Procedure

1. Enter system view.
   system-view
2. Create a VSI and enter its view.
   vsi *vsi-name*
3. Enable ND snooping for IPv6 addresses. Choose the options to configure as needed:
   - Enable ND snooping for global unicast addresses.
     ipv6 nd snooping enable global
   - Enable ND snooping for link-local addresses.
     ipv6 nd snooping enable link-local

   By default, ND snooping is disabled for IPv6 global unicast addresses and link-local addresses.

Enabling ND proxy

About ND proxy

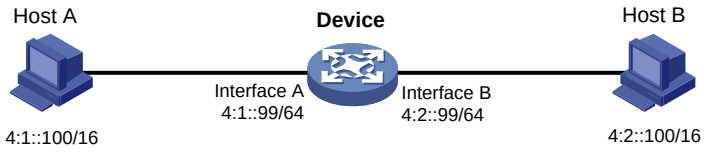
ND proxy enables a device to answer an NS message requesting the hardware address of a host on another network. With ND proxy, hosts in different broadcast domains can communicate with each other as they would on the same network.

ND proxy includes common ND proxy and local ND proxy.

Common ND proxy

As shown in Figure 7, Interface A with IPv6 address 4:1::99/64 and Interface B with IPv6 address 4:2::99/64 belong to different subnets. Host A and Host B reside on the same network but in different broadcast domains.

Figure 7 Application environment of ND proxy



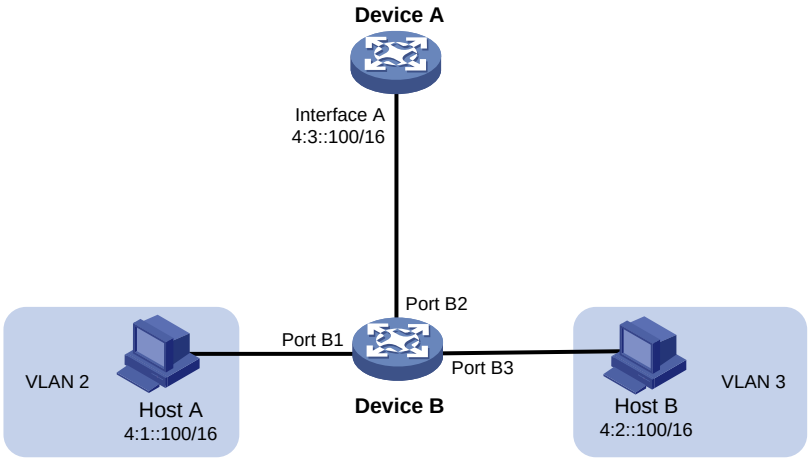
Because Host A's IPv6 address is on the same subnet as Host B's, Host A directly sends an NS message to obtain Host B's MAC address. However, Host B cannot receive the NS message because they belong to different broadcast domains.

To solve this problem, enable common ND proxy on Interface A and Interface B of the device. The device replies to the NS message from Host A, and forwards packets from other hosts to Host B.

Local ND proxy

As shown in Figure 8, Host A belongs to VLAN 2 and Host B belongs to VLAN 3. Host A and Host B connect to Port B1 and Port B3, respectively.

Figure 8 Application environment of local ND proxy



Because Host A's IPv6 address is on the same subnet as Host B's, Host A directly sends an NS message to obtain Host B's MAC address. However, Host B cannot receive the NS message because they belong to different VLANs.

To solve this problem, enable local ND proxy on Interface A of Device A so that Device A can forward messages between Host A and Host B.

Local ND proxy implements Layer 3 communication for four hosts in the following cases:

Local ND proxy implements Layer 3 communication for two hosts in the following cases:

Local ND proxy implements Layer 3 communication for two hosts in the following cases:

- The two hosts connect to ports of the same device and the ports must be in different VLANs.
- The two hosts connect to isolated Layer 2 ports in the same isolation group of a VLAN.

Enabling common ND proxy

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable common ND proxy.
   proxy-nd enable
   By default, common ND proxy is disabled.

Enabling local ND proxy

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable local ND proxy.
   local-proxy-nd enable
   By default, local ND proxy is disabled.

Enabling DAD message forwarding for a VSI

About this task

When an access user in an authentication domain initiates an authentication request, the device looks for an available authentication server. If all authentication servers in that authentication domain are detected unavailable, the device changes to the critical state (also known as the fail-permit state). In this state, the device assigns the access user to the critical domain (also known as the fail-permit domain) so that the user can come online in the critical domain without being authenticated. If the local ND proxy feature is enabled by using the **local-proxy-nd enable** command, the device discards the DAD NS and NA messages received from the VSI of the critical domain instead of forwarding those messages.

To resolve this issue, enable DAD message forwarding for the VSI of the critical domain. The device will then forward the DAD NS and NA messages received from the VSI of the critical domain.

Restrictions and guidelines

Enable this feature only for the VSIs associated with critical domains.

If you disable this feature for a VSI, the device directly discards the DAD NS and NA messages received from that VSI.

Procedure

1. Enter system view.
   system-view
2. Create a VSI and enter its view.
   vsi *vsi-name*
3. Enable DAD message forwarding for the VSI.
   ipv6 nd local-proxy dad forward enable
   By default, DAD message forwarding is enabled for a VSI.

Configuring IPv6 ND direct route advertisement

About IPv6 ND direct route advertisement

This feature generates 128-bit host routes based on ND entries for packet forwarding and route advertisement. The route preference value determines the match order of a route. Dynamic routing protocols use the tag value as the route identifier when redistributing the route.

After you enable this feature, the routing table might be populated with excessive host routes. To reduce the routing table size, execute the **ipv6 nd route-direct prefix convert-length** command for the device to generate network routes for identified ND entries instead of host routes.

After you enable ND direct route advertisement by using the **ipv6 nd route-direct advertise** command, the device uses ND entries to generate direct routes and adjacency entries. If the direct routes are generated earlier than the adjacency entries, packet loss will occur due to lack of Layer 2 information during packet encapsulation. To avoid such an issue, use the **ipv6 nd route-direct advertise delay** command to set the delay time for ND direct route generation.

After you enable ND direct route advertisement and set the delay time for ND direct route generation, the device starts a timer after it has learned an ND entry. Then, it generates an ND direct route when the timer value reaches the delay time. If you edit the direct route advertisement configuration within the delay time, the device immediately advertises the direct route with the new configuration.

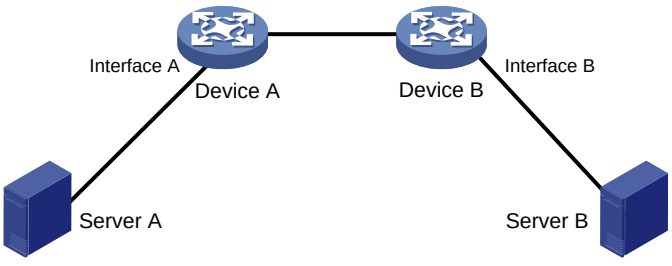
Restrictions and guidelines

The **ipv6 nd route-direct advertise** command is mutually exclusive with the **ipv6 nd route-direct advertise mad-down-single-homed** command. If you execute both of them on the same interface, the most recent command takes effect.

Application in Layer 3 access networks

As shown in Figure 9, ND direct route advertisement is enabled on Interface A and Interface B. This feature generates a host route to Server A and a host route to Server B for the routing protocols to advertise. Therefore, each device forwards only the traffic destined to the server within the network, which saves bandwidth.

Figure 9 Application in a Layer 3 access network



Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*

3. (Optional.) Set the delay time for ND direct route generation.

   ipv6 nd route-direct advertise delay *delay-time*

   By default, the device does not delay ND direct route generation.

4. Enable ND direct route advertisement.

   ipv6 nd route-direct advertise [**preference** *preference-value* | **tag** *tag-value*] *

   By default, ND direct route advertisement is disabled.

5. (Optional.) Specify a prefix length for generating a network route for identified ND entries.

   ipv6 nd route-direct prefix *ipv6-prefix prefix-length* **convert-length** *convert-length*

   By default, no prefix length is specified for generating a network route for identified ND entries, and the device generates 128-bit host routes based on ND entries.

   Make sure the IPv6 prefix is consistent with the prefix of the IPv6 address configured for the interface.

Enabling ND direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state

About this task

If the peer link of an M-LAG system fails, M-LAG Multi-Active Detection (MAD) will shut down all interfaces on the secondary member device upon M-LAG system split, except for the interfaces excluded from the shutdown action by IRF MAD or M-LAG MAD. An interface in M-LAG MAD DOWN state cannot forward traffic. As a result, traffic cannot be forwarded for the devices single-homed to the secondary member device.

To ensure traffic forwarding for single-homed devices, enable ND direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state. When the peer link fails, the ND module on the secondary member device will send ND entries to the routing management module, and the routing management module will generate direct routes based on the ND entries. The direct routes are used to direct traffic forwarding or are advertised by routing protocols to the single-homed devices.

Restrictions and guidelines

The **ipv6 nd route-direct advertise mad-down-single-homed** and **ipv6 nd route-direct advertise** commands are mutually exclusive. If you execute both of them on the same interface, the most recent command takes effect.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN interface view.

   interface vlan-interface *interface-number*

3. Enable ND direct route advertisement for single-homing interfaces in M-LAG MAD DOWN state.

   ipv6 nd route-direct advertise mad-down-single-homed

   By default, ND direct route advertisement is disabled for single-homing interfaces in M-LAG MAD DOWN state.

Adding prefix information advertised through RA messages to IPSG binding entries

About this task

This feature enables a VLAN interface to add prefix information advertised through RA messages to IP Source Guard (IPSG) binding entries. IPSG can then filter incoming IPv6 packets by IPv6 prefix information in the related IPSG binding entries.

This feature also supports collaboration with the ND detection feature to validate access users.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN interface view.

   interface vlan-interface *interface-number*

3. Add prefix information advertised through RA messages to IP Source Guard (IPSG) binding entries.

   ipv6 nd ra prefix source binding

   By default, the VLAN interface does not add prefix information advertised through RA messages to IPSG binding entries.

Enabling recording user IPv6 address conflicts

About this task

This feature detects and records user IPv6 address conflicts. A conflict occurs if an incoming NA packet has the same source IP address as an existing ND entry but a different source MAC address. The device generates a user IPv6 address conflict record, logs the conflict, and sends the log to the information center. For information about the log destination and output rule configuration in the information center, see the information center in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable recording user IPv6 address conflicts.

   ipv6 nd user-ip-conflict record enable

   By default, recording user IPv6 address conflicts is disabled.

Enabling recording user port migrations

About this task

This feature enables the device to detect and record user port migrations. A user port migrates if an incoming NA packet has the same source IPv6 address and source MAC address as an existing ND entry but a different ingress port. The device generates a user port migration record, logs the migration event, and sends the log to the information center. For information about the log destination and output rule configuration in the information center, see the information center in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable recording user port migrations.
   ipv6 nd user-move record enable
   By default, recording user port migrations is disabled.

Enabling ND logging for user online and offline events

About this task

This feature enables the device to generate user online or offline logs upon such events and send these logs to the information center. For information about the log destination and output rule configuration in the information center, see the information center in *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

A higher log output rate consumes more CPU resources. Adjust the log output rate based the CPU performance and usage.

Procedure

1. Enter system view.
   system-view
2. Enable ND logging for user online and offline events.
   ipv6 nd online-offline-log enable [**rate** *rate*]
   By default, ND logging for user online and offline events is disabled.

Logging ND entry deployment events

About this task

After learning an ND entry, the device deploys that ND entry to hardware for packet forwarding. To record the events that occur during ND entry deployment to hardware, you can enable this feature. This feature enables the device to generate log messages for ND deployment events and deliver such log messages to the information center.

For log messages to be sent correctly, configure the information center to set log message filtering and output rules, including output destinations. For information about information center configuration, see information center configuration in *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Log generation consumes memory resources. To save device memory, enable this feature only when you need to troubleshoot traffic forwarding issues.

If the number of log messages generated by the device per second reaches the limit, the device will not generate any log message for a new ND deployment event.

Procedure

1. Enter system view.
   system-view

2. Enable logging ND entry deployment events.
   ipv6 nd hardware log enable [**count-limit** *count-limit-value*]
   By default, the device does not log ND entry deployment events.

Enabling ND entry consistency check

About this task

Packet drops or incorrect forwarding might occur when the ND entries in hardware are inconsistent with ND entries configured in software. To prevent these issues, enable ND entry consistency check.

This feature compares all ND entries in software with the ND entries in hardware regularly. If the device detects an inconsistency, the device performs the following tasks:

- Generates a log.
- Updates the ND entry in hardware with the ND entry in software.

Procedure

1. Enter system view.
   system-view
2. Enable ND entry consistency check.
   ipv6 nd consistency-check enable
   By default, ND entry consistency check is disabled.

Disabling the device from aging or deleting ND entries in response to network topology changes

About this task

After you configure STP for the device in a tree-shaped network topology, the system notifies the ND module to age or delete the learned ND entries when the loop protocol detects a network topology change. Then, the device must learn ND entries again to obtain the latest ND entry information. If the network topology changes frequently, a large number of ND messages might occur when the device relearns ND entries, which occupies too many system resources and affects the normal operation of other services. To avoid such an issue, you can execute the **undo ipv6 nd topology-change enable** command. The device will not age or delete ND entries even if the network topology changes.

Restrictions and guidelines

After you execute the **undo ipv6 nd topology-change enable** command, the ND entries saved by the device might not be the latest ones, which causes user traffic interruption. As a best practice, use this command only when necessary.

Procedure

1. Enter system view.
   system-view
2. Disable the device from aging or deleting ND entries in response to network topology changes.
   undo ipv6 nd topology-change enable
   By default, the device ages or deletes ND entries in response to network topology changes.

Enabling recording ND entry learning events

About this task

An ND entry learning event occurs when the number of ND entries that an IRF member device or an interface has learnt exceeds the threshold or drops below the threshold.

After you enable this feature, the ND module logs ND entry learning events and sends them to the information center. For log messages to be sent correctly, configure the information center to set log message filtering and output rules, including output destinations. For information about the log destination and output rule configuration in the information center, see information center configuration in *Network Management and Monitoring Configuration Guide*.

If you enable ND entry limit notifications for the ND module, the device sends a notification to the SNMP module when the number of ND entries exceeds the alarm threshold. For ND entry limit notifications to be sent correctly, you must also configure SNMP on the device.

For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*. For more information about ND entry limit notifications, see ND attack defense configuration in *Security Configuration Guide*.

Procedure

1. Enter system view.

   `system-view`

2. (Optional.) Enable SNMP notifications for the ND module.

   `snmp-agent trap enable nd` [**`entry-limit`**]

   By default, SNMP notifications are disabled for the ND module.

   If you do not specify a keyword, this command enables all types of SNMP notifications for the ND module. If you specify the **`entry-limit`** keyword, this command enables only ND entry limit notifications for the ND module.

3. Enable recording ND entry learning events.

   `ipv6 nd entry-limit record enable`

   By default, the ND module does not record ND entry learning events.

Disabling sending NS messages when data packets trigger ND resolution

About this task

By default, when the device receives a data packet not destined for it and cannot find a match for the next hop in the ND table, it performs the following tasks:

1. Multicasts an NS message to obtain the MAC address of the next hop.
2. Generates an ND entry based on the obtained MAC address.

A large number of NS messages consume too many network resources, affecting normal service operation. To resolve the issue, use this command to disable the device from sending NS messages for ND entry learning when IPv6 data packets trigger ND resolution. This suppresses ND flooding by reducing ND packets on the network.

Restrictions and guidelines

With this feature configured on an interface, the device does not multicast an NS message for ND entry learning when the interface attempts to send a data packet that triggers ND resolution. As a best practice, configure this feature only when the network is under ND flood attacks.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Disable the device from sending NS messages for ND entry learning when IPv6 data packets trigger ND resolution.

   ipv6 nd fib-miss drop

   By default, the device sends NS messages for ND entry learning when IPv6 data packets trigger ND resolution.

Performing ND-ping

Using ND-ping to verify the availability of an IPv6 address

About this task

This feature allows users to verify if an IPv6 address is being used by another device in the LAN.

After you perform this task, the device sends an NS packet to the specified IPv6 address. If no NA packet is received within the timeout time, the device retransmits the NS packet. If no NA packet is received after the maximum number of transmission attempts is reached, the device considers that the IPv6 address is not being used.

You can also use the **ping ipv6** command to verify the IPv6 address availability. However, the test result might be inaccurate, because the peer device cannot respond if a firewall is configured to forbid the device from responding to ICMPv6 packets. Compared with ping operations, ND-ping uses Layer 2 packets (ND packets), which are not blocked by firewalls in most cases, and NS packets are shorter than ICMPv6 packets and require less network resources.

Prerequisites

To test the address availability by specifying a host name, configure DNS for the device to translate the host name to an IPv6 address. For more information about DNS, see "Configuring DNS."

Restrictions and guidelines

If multiple devices exist in the LAN, executing the **ping nd ipv6** command might take a long time. To stop the command execution, press **Ctrl+C**.

Procedure

Execute the command in any view to verify the availability of an IPv6 address in the LAN.

ping nd ipv6 *host* [**interface** *interface-type interface-number* [**vlan** *vlan-id*]] [**timeout** *timeout*] [**count** *count*]

Using ND-Ping to obtain the IPv6 address for a MAC address

About this task

Perform this task to obtain the IPv6 address of the device that uses the specified MAC address in a specific subnet.

After you perform this task, the device broadcasts a Layer 3 ICMPv6 packet. If no ICMPv6 response is received within the timeout time, the device resends the broadcast packet. If no ICMPv6 response is received after the number of maximum transmission attempts is received, the device considers that the MAC address does not exist in the subnet.

Restrictions and guidelines

If multiple devices exist in the subnet, executing the **ping nd mac** command might take a long time. To stop the command execution, press **Ctrl+C**.

Procedure

Execute the command in any view to obtain the IPv6 address of the device that uses the specified MAC address in a specific subnet.

ping nd mac *mac-address* { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*] } [**timeout** *timeout*] [**count** *count*]

Display and maintenance commands for IPv6 ND

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display statistics for ND proxy reply packets.	**display ipv6 nd proxy statistics**
Display information about direct routes generated advertised by the ND module.	**display ipv6 nd route-direct advertise interface** *interface-type interface-number*
Display the number of IPv6 ND snooping entries in VLANs.	**display ipv6 nd snooping count vlan** [**interface** *interface-type interface-number*]
Display the number of IPv6 ND snooping entries in a VSI.	**display ipv6 nd snooping count vsi** [*vsi-name*]
Display IPv6 ND snooping entries in VLANs.	**display ipv6 nd snooping vlan** [[*vlan-id* | **interface** *interface-type interface-number*] [**global** | **link-local**] | *ipv6-address*] [**verbose**]
Display IPv6 ND snooping entries in a VSI.	**display ipv6 nd snooping vsi** [*vsi-name*] [**slot** *slot-number*
Display user IPv6 address conflict records.	**display ipv6 nd user-ip-conflict record** [**slot** *slot-number*]
Display user port migration records.	**display ipv6 nd user-move record** [**slot** *slot-number*]
Display the total number of neighbor entries.	**display ipv6 neighbors** { { **all** | **dynamic** | **static** } [**slot** *slot-number*] | **interface** *interface-type interface-number* | **vlan** *vlan-id* } **count**
Display neighbor information.	**display ipv6 neighbors** { { *ipv6-address* | **all** | **dynamic** | **static** } [**slot** *slot-number*] | **interface** *interface-type interface-number* | **vlan** *vlan-id* } [**verbose**]
Display the maximum number of ND entries that a device supports.	**display ipv6 neighbors entry-limit**
Display ND entry statistics.	**display ipv6 neighbors statistics** { [**by-slot**] **all** | **interface** { *interface-name* | *interface-type interface-number* } | **slot**

	slot-number }
Display the ND table usage.	**display ipv6 neighbors usage**
Display neighbor information for a VPN.	**display ipv6 neighbors vpn-instance** *vpn-instance-name* [**count**]
Clear IPv6 ND snooping entries in a VLAN.	**reset ipv6 nd snooping vlan** { [*vlan-id*] [**global** | **link-local**] | *vlan-id ipv6-address* }
Clear IPv6 ND snooping entries in a VSI.	**reset ipv6 nd snooping vsi** [*vsi-name*]
Clear IPv6 neighbor information.	**reset ipv6 neighbors** { **all** | **dynamic** | **interface** *interface-type interface-number* | **slot** *slot-number* | **static** }

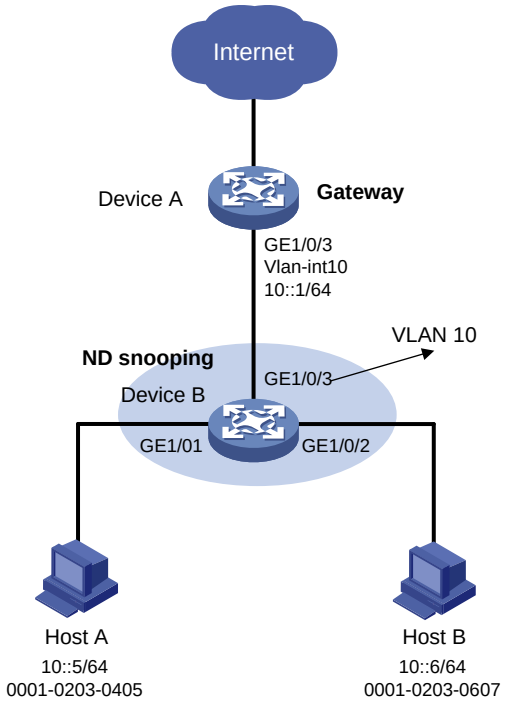
IPv6 ND configuration examples

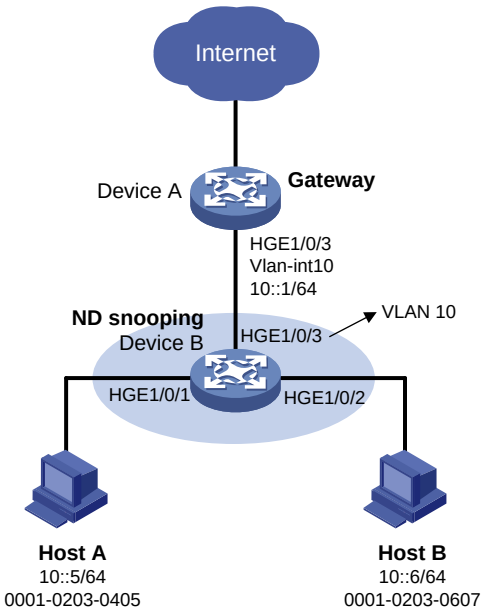
Example: Configuring ND snooping

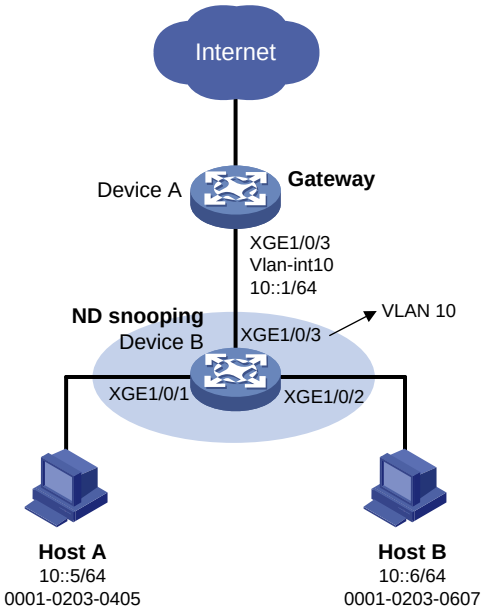
Network configuration

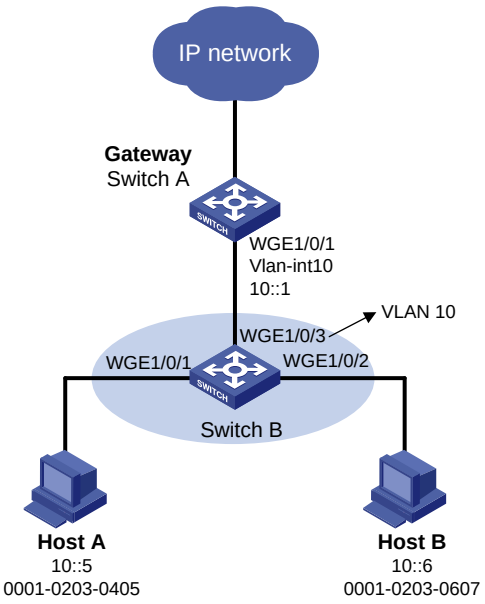
As shown in Figure 10, Host A and Host B are connected to the gateway through Device B. Enable ND snooping on Device B to learn ND snooping entries about Host A and Host B.

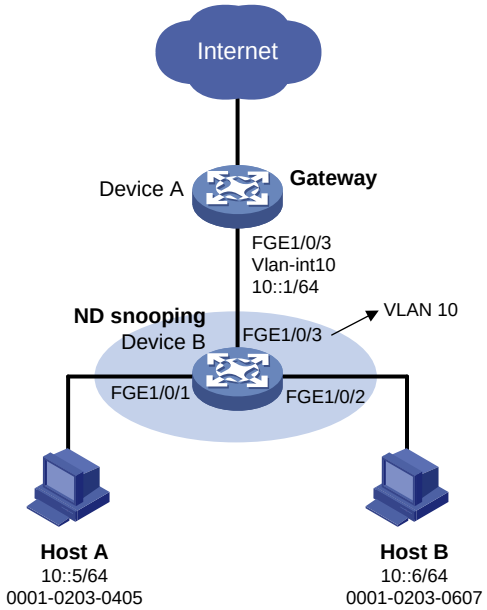
Figure 10 Network diagram

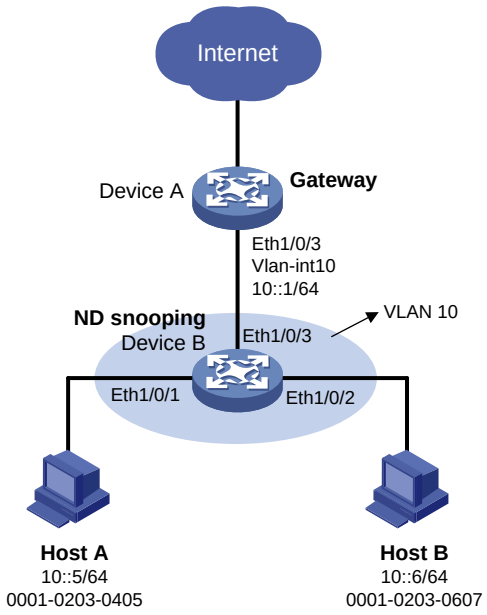


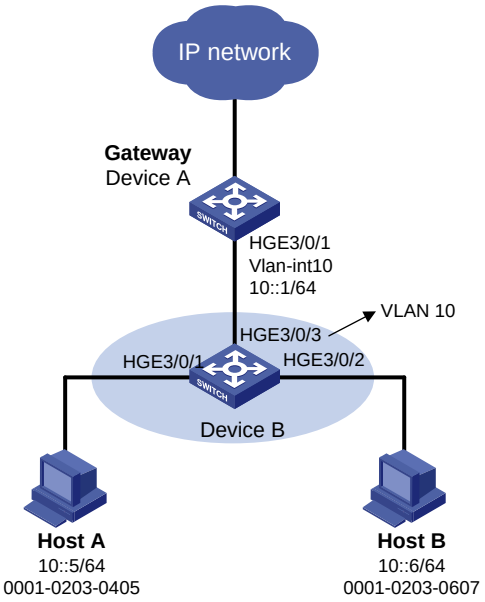

Internet
Device A
Gateway
HGE1/0/3
Vlan-int10
10::1/64
ND snooping
Device B
HGE1/0/3
VLAN 10
HGE1/0/1
HGE1/0/2
Host A
10::5/64
0001-0203-0405
Host B
10::6/64
0001-0203-0607


Internet
Device A
Gateway
XGE1/0/3
Vlan-int10
10::1/64
ND snooping
Device B
XGE1/0/3
VLAN 10
XGE1/0/1
XGE1/0/2
Host A
10::5/64
0001-0203-0405
Host B
10::6/64
0001-0203-0607


IP network
Gateway
Switch A
WGE1/0/1
Vlan-int10
10::1
VLAN 10
WGE1/0/3
WGE1/0/1
WGE1/0/2
Switch B
Host A
10::5
0001-0203-0405
Host B
10::6
0001-0203-0607


Internet
Device A
Gateway
FGE1/0/3
Vlan-int10
10::1/64
ND snooping
Device B
VLAN 10
FGE1/0/3
FGE1/0/1
FGE1/0/2
Host A
10::5/64
0001-0203-0405
Host B
10::6/64
0001-0203-0607





Procedure

1. Configure Device A:

   # Create VLAN 10.

   ```
   <DeviceA> system-view
   [DeviceA] vlan 10
   [DeviceA-vlan10] quit
   ```

   # Configure GigabitEthernet 1/0/3 to trunk VLAN 10.

   ```
   [DeviceA] interface gigabitethernet 1/0/3
   [DeviceA-GigabitEthernet1/0/3] port link-type trunk
   [DeviceA-GigabitEthernet1/0/3] port trunk permit vlan 10
   [DeviceA-GigabitEthernet1/0/3] quit
   ```

   # Assign IPv6 address 10::1/64 to VLAN-interface 10.

   ```
   [DeviceA] interface vlan-interface 10
   ```

```
[DeviceA-Vlan-interface10] ipv6 address 10::1/64
[DeviceA-Vlan-interface10] quit
```

2. Configure Device B:

Create VLAN 10.

```
[DeviceB] vlan 10
[DeviceB-vlan10] quit
```

Configure GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 to trunk VLAN 10.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type access
[DeviceB-GigabitEthernet1/0/1] port access vlan 10
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type access
[DeviceB-GigabitEthernet1/0/2] port access vlan 10
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan 10
[DeviceB-GigabitEthernet1/0/3] quit
```

Enable ND snooping for global unicast addresses and link-local addresses in VLAN 10.

```
[DeviceB] vlan 10
[DeviceB-vlan10] ipv6 nd snooping enable global
[DeviceB-vlan10] ipv6 nd snooping enable link-local
```

Enable ND snooping for data packets from unknown sources in VLAN 10.

```
[DeviceB-vlan10] ipv6 nd snooping glean source
[DeviceB-vlan10] quit
```

Configure GigabitEthernet 1/0/3 as ND trusted interface.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] ipv6 nd detection trust
[DeviceB-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet 1/0/1 to learn a maximum number of 200 ND snooping entries.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] ipv6 nd snooping max-learning-num 200
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 to learn a maximum number of 200 ND snooping entries.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ipv6 nd snooping max-learning-num 200
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] commit
```

Verifying the configuration

Verify that Device B has learned ND snooping entries for Host A and Host B.

```
[DeviceB] display ipv6 nd snooping vlan 10
IPv6 address              MAC address     VID  Interface     Status       Age
10::5                     0001-0203-0405  10   GE1/0/1       VALID        157
10::6                     0001-0203-0607  10   GE1/0/2       VALID        105
10::5                     0001-0203-0405  10   HGE1/0/1     VALID        157
```

```
10::6                    0001-0203-0607  10   HGE1/0/2     VALID        105
10::5                    0001-0203-0405  10   XGE1/0/1      VALID        157
10::6                    0001-0203-0607  10   XGE1/0/2      VALID        105
10::5                    0001-0203-0405  10   WGE1/0/1      VALID        157
10::6                    0001-0203-0607  10   WGE1/0/2      VALID        105
10::5                    0001-0203-0405  10   FGE1/0/1      VALID        157
10::6                    0001-0203-0607  10   FGE1/0/2      VALID        105
10::5                    0001-0203-0405  10   Eth1/0/1       VALID        157
10::6                    0001-0203-0607  10   Eth1/0/2       VALID        105
10::5                    0001-0203-0405  10   HGE3/0/1      VALID        157
10::6                    0001-0203-0607  10   HGE3/0/2      VALID        105
```

Contents

DHCPv6 overview 1
DHCPv6 address/prefix assignment 1
Rapid assignment involving two messages 1
Assignment involving four messages 1
Address/prefix lease renewal 2
Stateless DHCPv6 3
DHCPv6 options 3
Option 18 3
Option 37 4
Option 79 5
Option 65253 5
Protocols and standards 5
Configuring the DHCPv6 server 6
About DHCPv6 server 6
IPv6 address assignment 6
IPv6 prefix assignment 6
Concepts 7
DHCPv6 address pool 7
IPv6 address/prefix allocation sequence 8
DHCPv6 server tasks at a glance 9
Configuring IPv6 prefix assignment 9
Configuring IPv6 address assignment 11
Configuring network parameters assignment 12
About network parameters assignment 12
Configuring network parameters in a DHCPv6 address pool 13
Configuring network parameters in a DHCPv6 option group 13
Configuring the DHCPv6 server on an interface 14
Configuring a DHCPv6 policy for IPv6 address and prefix assignment 15
Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 server 16
Configuring DHCPv6 binding auto backup 16
Advertising subnets assigned to clients 17
Enabling the DHCPv6 server to advertise IPv6 prefixes 18
Configuring DHCPv6 vendor-specific options 18
Applying a DHCPv6 address pool to a VPN instance 19
Configuring DHCPv6 flood attack protection 19
Enabling IPv6 address binding conversion for IP source guard 20
Configuring SNMP notifications for the DHCPv6 server 21
About SNMP DHCPv6 server notifications 21
Enabling IPv6 resource exhaustion notifications 21
Enabling IPv6 address usage alarm notifications 21
Enabling IPv6 prefix usage alarm notifications 21
Enabling DHCPv6 logging on the DHCPv6 server 22
Enabling IPv6 resource exhaustion logging 22
Display and maintenance commands for DHCPv6 server 23
DHCPv6 server configuration examples 24
Example: Configuring dynamic IPv6 prefix assignment 24
Example: Configuring dynamic IPv6 address assignment 27
Configuring the DHCPv6 relay agent 29
About DHCPv6 relay agent 29
Typical application 29
DHCPv6 relay agent operating process 29
DHCPv6 relay agent tasks at a glance 30
Enabling the DHCPv6 relay agent on an interface 30
Specifying DHCPv6 servers on the relay agent 31
Specifying DHCPv6 server IP addresses 31

Specifying DHCPv6 servers for a DHCPv6 address pool on the DHCPv6 relay agent ····31
Specifying the DHCPv6 server selection algorithm ····32
Specifying a gateway address for DHCPv6 clients ····33
Setting the DUID entry aging timer for smart relay ····33
Configuring DHCPv6 smart relay for address acquisition from different IPv6 subnets for a client ····34
Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 relay agent ····35
Specifying a padding mode for the Interface-ID option ····35
Configuring the DHCPv6 relay agent to add Option 37 in Relay-forward messages ····35
Enabling the DHCPv6 relay agent to support Option 79 ····36
Enabling the DHCPv6 relay agent to advertise IPv6 prefixes ····36
Enabling advertising host routes for assigned IPv6 addresses ····37
Specifying the source IPv6 address for relayed DHCPv6 requests ····37
Configuring DHCPv6 relay security features ····38
Enabling the DHCPv6 relay agent to record relay entries ····38
Enabling IPv6 release notification ····38
Enabling client offline detection ····39
Configuring DHCPv6 flood attack protection ····39
Enabling checking the Reply messages in response to Confirm messages ····40
Discarding DHCPv6 requests that are delivered from VXLAN tunnels ····41
Configuring support for forwarding DHCPv6 replies based on MAC address table ····42
Display and maintenance commands for DHCPv6 relay agent ····42
DHCPv6 relay agent configuration examples ····43
Example: Configuring DHCPv6 relay agent ····43
Configuring the DHCPv6 client ····46
About the DHCPv6 client ····46
Restrictions and guidelines: DHCPv6 client configuration ····46
DHCPv6 client tasks at a glance ····46
Configuring the DHCPv6 client DUID ····46
Configuring IPv6 address acquisition ····47
Configuring IPv6 prefix acquisition ····47
Configuring IPv6 address and prefix acquisition ····47
Configuring acquisition of configuration parameters except IP addresses and prefixes ····48
Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 client ····48
Display and maintenance commands for DHCPv6 client ····49
DHCPv6 client configuration examples ····49
Example: Configuring IPv6 address acquisition ····49
Example: Configuring IPv6 prefix acquisition ····51
Example: Configuring IPv6 address and prefix acquisition ····53
Example: Configuring stateless DHCPv6 ····55
Configuring DHCPv6 snooping ····57
About DHCPv6 snooping ····57
Application of trusted and untrusted ports ····57
DHCPv6 snooping on an M-LAG network ····58
Restrictions and guidelines: DHCPv6 snooping configuration ····59
DHCPv6 snooping tasks at a glance ····60
Configuring basic DHCPv6 snooping features ····60
Configuring basic DHCPv6 snooping features in a common network ····60
Configuring basic DHCPv6 snooping features in a VXLAN network ····61
Configuring DHCP snooping support for Option 18 ····63
Configuring DHCP snooping support for Option 37 ····63
Configuring DHCPv6 snooping entry auto backup ····63
Setting the maximum number of DHCPv6 snooping entries ····64
Configuring DHCPv6 packet rate limit ····65
Configuring DHCPv6 snooping security features ····65
Enabling DHCPv6-REQUEST check ····65
Configuring a DHCPv6 packet blocking port ····66
Enabling LDRA on the DHCPv6 snooping device ····66
Enabling Relay-Forward packet check ····67
Enabling client offline detection on the DHCPv6 snooping device ····67
Enabling DHCPv6 snooping logging and alarm ····68

Enabling DHCPv6 snooping logging ····· 68
Configuring packet drop alarm ····· 68
Disabling DHCPv6 snooping on an interface ····· 69
Display and maintenance commands for DHCPv6 snooping ····· 69
DHCPv6 snooping configuration examples ····· 70
Example: Configuring DHCPv6 snooping globally ····· 70
Example: Configuring DHCPv6 snooping for a VLAN ····· 73
Configuring DHCPv6 guard ····· 78
About DHCPv6 guard ····· 78
DHCPv6 guard operating mechanism ····· 78
Restrictions and guidelines: DHCPv6 guard configuration ····· 79
DHCPv6 guard tasks at a glance ····· 79
Configuring a DHCPv6 guard policy ····· 79
Applying a DHCPv6 guard policy to an interface ····· 80
Applying a DHCPv6 guard policy to a VLAN ····· 80
Display and maintenance commands for DHCPv6 guard ····· 81
DHCPv6 guard configuration examples ····· 81
Example: Configuring DHCPv6 guard ····· 81

DHCPv6 overview

DHCPv6 provides a framework to assign IPv6 prefixes, IPv6 addresses, and other configuration parameters to hosts.

DHCPv6 address/prefix assignment

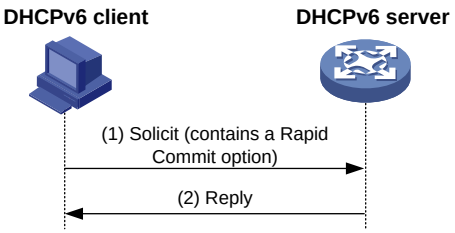
An address/prefix assignment process involves two or four messages.

Rapid assignment involving two messages

As shown in Figure 1, rapid assignment operates in the following steps:

1. The DHCPv6 client sends to the DHCPv6 server a Solicit message that contains a Rapid Commit option to prefer rapid assignment.
2. If the DHCPv6 server supports rapid assignment, it responds with a Reply message containing the assigned IPv6 address/prefix and other configuration parameters. If the DHCPv6 server does not support rapid assignment, Assignment involving four messages is performed.

Figure 1 Rapid assignment involving two messages

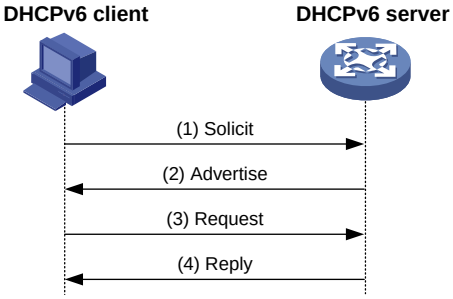


Assignment involving four messages

As shown in Figure 2, four-message assignment operates using the following steps:

1. The DHCPv6 client sends a Solicit message to request an IPv6 address/prefix and other configuration parameters.
2. The DHCPv6 server responds with an Advertise message that contains the assignable address/prefix and other configuration parameters if either of the following conditions exists:
   - The Solicit message does not contain a Rapid Commit option.
   - The DHCPv6 server does not support rapid assignment even though the Solicit message contains a Rapid Commit option.
3. The DHCPv6 client might receive multiple Advertise messages offered by different DHCPv6 servers. It selects an offer according to the receiving sequence and server priority, and sends a Request message to the selected server for confirmation.
4. The DHCPv6 server sends a Reply message to the client, confirming that the address/prefix and other configuration parameters are assigned to the client.

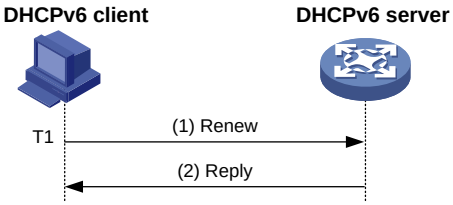
Figure 2 Assignment involving four messages



Address/prefix lease renewal

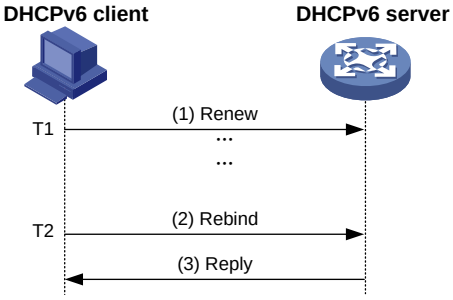
An IPv6 address/prefix assigned by a DHCPv6 server has a valid lifetime. After the valid lifetime expires, the DHCPv6 client cannot use the IPv6 address/prefix. To use the IPv6 address/prefix, the DHCPv6 client must renew the lease time.

Figure 3 Using the Renew message for address/prefix lease renewal



As shown in Figure 3, at T1, the DHCPv6 client sends a Renew message to the DHCPv6 server. The recommended value of T1 is half the preferred lifetime. The DHCPv6 server responds with a Reply message, informing the client whether the lease is renewed.

Figure 4 Using the Rebind message for address/prefix lease renewal



As shown in Figure 4:

- If the DHCPv6 client does not receive a response from the DHCPv6 server after sending a Renew message at T1, it multicasts a Rebind message to all DHCPv6 servers at T2. Typically, the value of T2 is 0.8 times the preferred lifetime.
- The DHCPv6 server responds with a Reply message, informing the client whether the lease is renewed.
- If the DHCPv6 client does not receive a response from any DHCPv6 server before the valid lifetime expires, the client stops using the address/prefix.

For more information about the valid lifetime and the preferred lifetime, see "Configuring basic IPv6 settings."

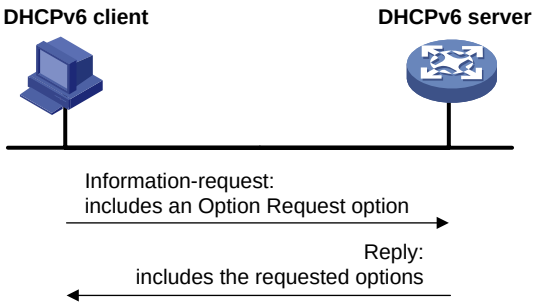
Stateless DHCPv6

Stateless DHCPv6 enables a device that has obtained an IPv6 address/prefix to get other configuration parameters from a DHCPv6 server.

The device performs stateless DHCPv6 if an RA message with the following flags is received from the router during stateless address autoconfiguration:

- The managed address configuration flag (M flag) is set to 0.
- The other stateful configuration flag (O flag) is set to 1.

Figure 5 Stateless DHCPv6 operation



As shown in Figure 5, stateless DHCPv6 operates in the following steps:

2. The DHCPv6 client sends an Information-request message to the multicast address of all DHCPv6 servers and DHCPv6 relay agents. The Information-request message contains an Option Request option that specifies the requested configuration parameters.
3. The DHCPv6 server returns to the client a Reply message containing the requested configuration parameters.
4. The client checks the Reply message. If the obtained configuration parameters match those requested in the Information-request message, the client uses these parameters to complete configuration. If not, the client ignores the configuration parameters. If the client receives multiple replies with configuration parameters matching those requested in the Information-request message, it uses the first received reply.

DHCPv6 options

Option 18

Option 18, also called the interface-ID option, is used by the DHCPv6 relay agent to determine the interface to use to forward RELAY-REPLY message.

The DHCPv6 snooping device adds Option 18 to the received DHCPv6 request message before forwarding it to the DHCPv6 server. The server then assigns IP address to the client based on the client information in Option 18.

Figure 6 Option 18 format

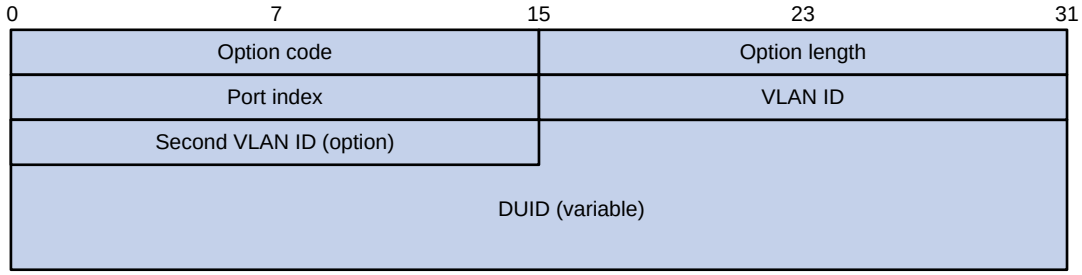


Figure 6 shows the Option 18 format, which includes the following fields:

- **Option code**—Option code. The value is 18.
- **Option length**—Size of the option data.
- **Port index**—Port that receives the DHCPv6 request from the client.
- **VLAN ID**—ID of the outer VLAN.
- **Second VLAN ID**—ID of the inner VLAN. This field is optional. If the received DHCPv6 request does not contain a second VLAN, Option 18 also does not contain it.
- **DUID**—DUID of the DHCPv6 client.

Option 37

Option 37, also called the remote-ID option, is used to identify the client.

The DHCPv6 snooping device adds Option 37 to the received DHCPv6 request message before forwarding it to the DHCPv6 server. This option provides client information about address allocation.

Figure 7 Option 37 format

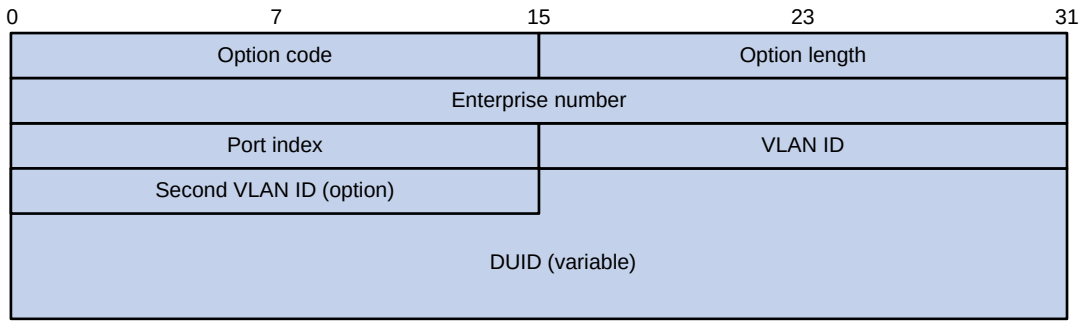


Figure 7 shows the Option 37 format, which includes the following fields:

- **Option code**—Option code. The value is 37.
- **Option length**—Size of the option data.
- **Enterprise number**—Enterprise number.
- **Port index**—Port to which the client is attached.
- **VLAN ID**—ID of the outer VLAN.
- **Second VLAN ID**—ID of the inner VLAN. This field is optional. Option 37 does not contain this information if the received DHCPv6 request does not contain a second VLAN ID.
- **DUID**—DUID information to be sent to DHCP servers. This information is configurable with the **`ipv6 dhcp relay remote-id duid`** command.

Option 79

Option 79, also called the client link-layer address option, is used to record the MAC address of the DHCPv6 client. The first relay agent that a DHCPv6 request passes learns the MAC address of the client and encapsulates this address into Option 79 in the Relay-Forward message for the request. The DHCPv6 server verifies the client or assigns IPv6 address/prefix to the client based on the MAC address of the client.

Figure 8 Option 79 format

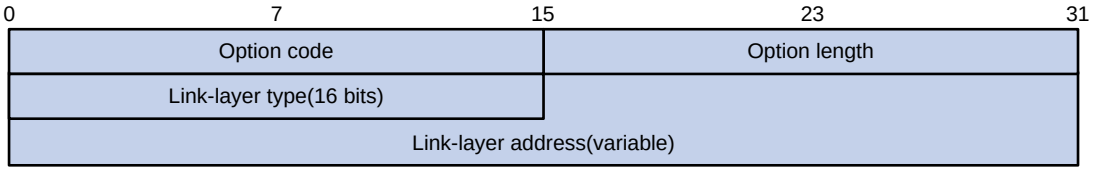


Figure 8 shows the Option 79 format, which includes the following fields:

- **Option code**—Option code. The value is 79.
- **Option length**—Size of the option data.
- **Link-layer type**—Link-layer address type of the client.
- **Link-layer address**—Link-layer address of the client.

Option 65253

As a vendor-specific DHCP option, Option 65253 contains the IPv6 domain name of a cloud server. After a DHCP client receives a DHCPv6 server reply that contains Option 65253, it automatically obtains the cloud server's domain name by parsing Option 65253. Then, the client can remotely log in to the cloud server for management purposes. For more information about cloud servers, see cloud connection configuration in *Network Management and Monitoring Configuration Guide*.

To have the device correctly obtain the IPv6 domain name in Option 65253, make sure the value in Option 65253 uses one of the following formats:

- Domain name with a port number specified:
  - **Domain name format**: *domain name:port number*, for example, example.com:17443.
  - **IPv4 address format**: *X.X.X.X:port number*, for example, 192.168.1.1:17443.
  - **IPv6 address format**: *[X:X::X:X]:port number*, for example, [1::1]:64.
- Domain name without a port number specified (the default port number is used):
  - **Domain name format**: *domain name*, for example, example.com.
  - **IPv4 address format**: *X.X.X.X*, for example, 192.168.1.1.
  - **IPv6 address format**: *X:X::X:X* or *[X:X::X:X]*, for example, 1::1 or [1::1].

Protocols and standards

- RFC 3736, *Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6*
- RFC 3315, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*
- RFC 2462, *IPv6 Stateless Address Autoconfiguration*
- RFC 3633, *IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6*
- RFC 6939, *Client Link-Layer Address Option in DHCPv6*

Configuring the DHCPv6 server

About DHCPv6 server

A DHCPv6 server can assign IPv6 addresses, IPv6 prefixes, and other configuration parameters to DHCPv6 clients.

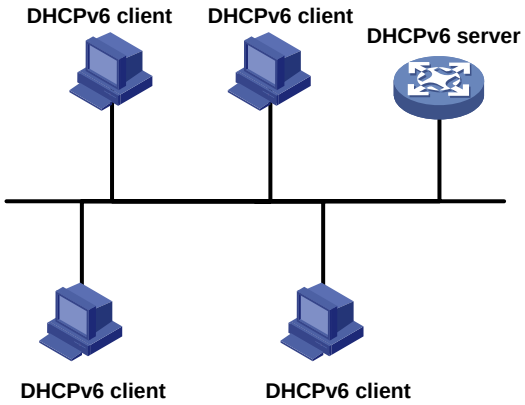
IPv6 address assignment

As shown in Figure 9, the DHCPv6 server assigns IPv6 addresses, domain name suffixes, DNS server addresses, and other configuration parameters to DHCPv6 clients.

The IPv6 addresses assigned to the clients include the following types:

- **Temporary IPv6 addresses**—Frequently changed without lease renewal.
- **Non-temporary IPv6 addresses**—Correctly used by DHCPv6 clients, with lease renewal.

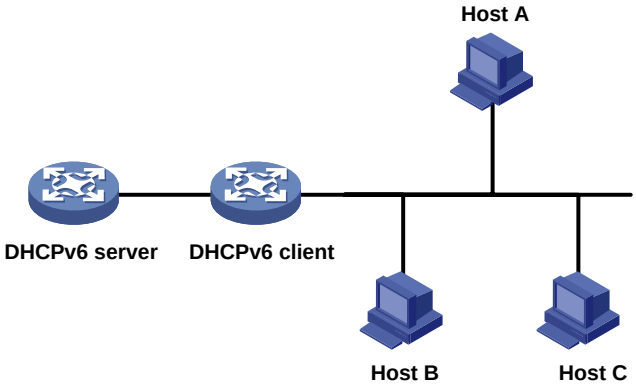
Figure 9 IPv6 address assignment



IPv6 prefix assignment

As shown in Figure 10, the DHCPv6 server assigns an IPv6 prefix to the DHCPv6 client. The client advertises the prefix information in a multicast RA message so that hosts on the subnet can automatically configure their IPv6 addresses by using the prefix.

Figure 10 IPv6 prefix assignment



Concepts

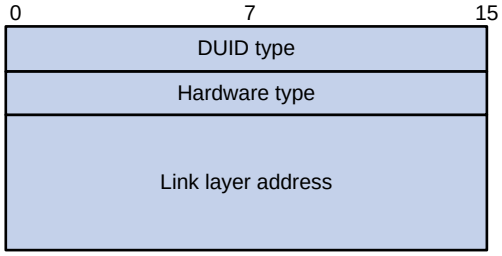
Multicast addresses used by DHCPv6

DHCPv6 uses the multicast address FF05::1:3 to identify all site-local DHCPv6 servers. It uses the multicast address FF02::1:2 to identify all link-local DHCPv6 servers and relay agents.

DUID

A DHCP unique identifier (DUID) uniquely identifies a DHCPv6 device (DHCPv6 client, server, or relay agent). A DHCPv6 device adds its DUID in a sent packet.

Figure 11 DUID-LL format



The device supports the DUID format based on link-layer address (DUID-LL) defined in RFC 3315. Figure 11 shows the DUID-LL format, which includes the following fields:

- **DUID type**—The device supports the DUID type of DUID-LL with the value of 0x0003.
- **Hardware type**—The device supports the hardware type of Ethernet with the value of 0x0001.
- **Link layer address**—Takes the value of the bridge MAC address of the device.

IA

Identified by an IAID, an identity association (IA) provides a construct through which a client manages the obtained addresses, prefixes, and other configuration parameters. A client can have multiple IAs, for example, one for each of its interfaces.

IAID

An IAID uniquely identifies an IA. It is chosen by the client and must be unique on the client.

PD

The DHCPv6 server creates a prefix delegation (PD) for each assigned prefix to record the following details:

- IPv6 prefix.
- Client DUID.
- IAID.
- Valid lifetime.
- Preferred lifetime.
- Lease expiration time.
- IPv6 address of the requesting client.

DHCPv6 address pool

The DHCP server selects IPv6 addresses, IPv6 prefixes, and other parameters from an address pool, and assigns them to the DHCP clients.

Address allocation mechanisms

DHCPv6 supports the following address allocation mechanisms:

- **Static address allocation**—To implement static address allocation for a client, create a DHCPv6 address pool, and manually bind the DUID and IAID of the client to an IPv6 address in the DHCPv6 address pool. When the client requests an IPv6 address, the DHCPv6 server assigns the IPv6 address in the static binding to the client.
- **Dynamic address allocation**—To implement dynamic address allocation for clients, create a DHCPv6 address pool, specify a subnet for the pool, and divide the subnet into temporary and non-temporary IPv6 address ranges. Upon receiving a DHCP request, the DHCPv6 server selects an IPv6 address from the temporary or non-temporary IPv6 address range based on the address type in the client request.

Prefix allocation mechanisms

DHCPv6 supports the following prefix allocation mechanisms:

- **Static prefix allocation**—To implement static prefix allocation for a client, create a DHCPv6 address pool, and manually bind the DUID and IAID of the client to an IPv6 prefix in the DHCPv6 address pool. When the client requests an IPv6 prefix, the DHCPv6 server assigns the IPv6 prefix in the static binding to the client.
- **Dynamic prefix allocation**—To implement dynamic prefix allocation for clients, create a DHCPv6 address pool and a prefix pool, specify a subnet for the address pool, and apply the prefix pool to the address pool. Upon receiving a DHCP request, the DHCPv6 server dynamically selects an IPv6 prefix from the prefix pool in the address pool.

Address pool selection

The DHCPv6 server observes the following principles when selecting an IPv6 address or prefix for a client:

1. If there is an address pool where an IPv6 address is statically bound to the DUID or IAID of the client, the DHCPv6 server selects this address pool. It assigns the statically bound IPv6 address or prefix and other configuration parameters to the client.
2. If the receiving interface has a DHCP policy and the DHCP client matches a user class, the DHCP server selects the address pool that is bound to the matching user class. If no matching user class is found, the server assigns an IP address and other parameters from the default DHCP address pool. If no default address pool is specified or the default address pool does not have assignable IP addresses, the address assignment fails.
3. If the receiving interface has an address pool applied, the DHCP server selects an IPv6 address or prefix and other configuration parameters from this address pool.
4. If the above conditions are not met, the DHCPv6 server selects an address pool depending on the client location.
    - **Client on the same subnet as the server**—The DHCPv6 server compares the IPv6 address of the receiving interface with the subnets of all address pools. It selects the address pool with the longest-matching subnet.
    - **Client on a different subnet than the server**—The DHCPv6 server compares the IPv6 address of the DHCPv6 relay agent interface closest to the client with the subnets of all address pools. It also selects the address pool with the longest-matching subnet.

    To make sure IPv6 address allocation functions correctly, keep the subnet used for dynamic assignment consistent with the subnet where the interface of the DHCPv6 server or DHCPv6 relay agent resides.

IPv6 address/prefix allocation sequence

The DHCPv6 server selects an IPv6 address/prefix for a client in the following sequence:

1. IPv6 address/prefix statically bound to the client's DUID and IAID and expected by the client.

2. IPv6 address/prefix statically bound to the client's DUID and IAID.
3. IPv6 address/prefix statically bound to the client's DUID and expected by the client.
4. IPv6 address/prefix statically bound to the client's DUID.
5. EUI-64 IPv6 address generated based on the client MAC address if EUI-64 address allocation is enabled.
6. Assignable IPv6 address/prefix in the address pool/prefix pool expected by the client.
7. IPv6 address/prefix that was ever assigned to the client.
8. Assignable IPv6 address/prefix in the address pool/prefix pool.
9. IPv6 address/prefix that was a conflict or passed its lease duration. If no IPv6 address/prefix is assignable, the server does not respond.

If a client moves to another subnet, the DHCPv6 server selects an IPv6 address/prefix from the address pool that matches the new subnet.

Conflicted IPv6 addresses can be assigned to other DHCPv6 clients only after the addresses are in conflict for one hour.

DHCPv6 server tasks at a glance

To configure the DHCPv6 server, perform the following tasks:

1. Configuring the DHCPv6 server to assign IPv6 prefixes, IPv6 addresses, and other network parameters

   Choose the following tasks as needed:
   - Configuring IPv6 prefix assignment
   - Configuring IPv6 address assignment
   - Configuring network parameters assignment
2. Modifying the address pool selection method on the DHCPv6 server

   Choose the following tasks as needed:
   - Configuring the DHCPv6 server on an interface
   - Configuring a DHCPv6 policy for IPv6 address and prefix assignment
3. (Optional.) Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 server
4. (Optional.) Configuring DHCPv6 binding auto backup
5. (Optional.) Advertising subnets assigned to clients
6. (Optional.) Enabling the DHCPv6 server to advertise IPv6 prefixes
7. (Optional.) Configuring DHCPv6 vendor-specific options
8. (Optional.) Applying a DHCPv6 address pool to a VPN instance
9. (Optional.) Configuring DHCPv6 flood attack protection
10. (Optional.) Enabling IPv6 address binding conversion for IP source guard
11. (Optional.) Configuring SNMP notifications for the DHCPv6 server
12. (Optional.) Enabling DHCPv6 logging on the DHCPv6 server
13. (Optional.) Enabling IPv6 resource exhaustion logging

Configuring IPv6 prefix assignment

About this task

Use the following methods to configure IPv6 prefix assignment:

- **Configure a static IPv6 prefix binding in an address pool**—If you bind a DUID and an IAID to an IPv6 prefix, the DUID and IAID in a request must match those in the binding before the DHCPv6 server can assign the IPv6 prefix to the DHCPv6 client. If you only bind a DUID to an IPv6 prefix, the DUID in the request must match the DUID in the binding before the DHCPv6 server can assign the IPv6 prefix to the DHCPv6 client.
- **Apply a prefix pool to an address pool**—The DHCPv6 server dynamically assigns an IPv6 prefix from the prefix pool in the address pool to a DHCPv6 client.

Restrictions and guidelines

When you configure IPv6 prefix assignment, follow these restrictions and guidelines:

- An IPv6 prefix can be bound to only one DHCPv6 client. You cannot modify bindings that have been created. To change the binding for a DHCPv6 client, you must delete the existing binding first.
- One address pool can have only one prefix pool applied. You cannot modify prefix pools that have been applied. To change the prefix pool for an address pool, you must remove the prefix pool application first.
- You can apply a prefix pool that has not been created to an address pool. The setting takes effect after the prefix pool is created.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Specify the IPv6 prefixes excluded from dynamic assignment.

   ipv6 dhcp server forbidden-prefix *start-prefix*/*prefix-len* [*end-prefix*/*prefix-len*] [**vpn-instance** *vpn-instance-name*]

   By default, no IPv6 prefixes in the prefix pool are excluded from dynamic assignment.

   If the excluded IPv6 prefix is in a static binding, the prefix still can be assigned to the client.

3. Create a prefix pool.

   ipv6 dhcp prefix-pool *prefix-pool-number* **prefix** { *prefix-number* | *prefix*/*prefix-len* } **assign-len** *assign-len* [**vpn-instance** *vpn-instance-name*]

   This step is required for dynamic prefix assignment.

   If you specify an IPv6 prefix by its ID, make sure the IPv6 prefix is in effect. Otherwise, the configuration does not take effect.

4. Enter DHCP address pool view.

   ipv6 dhcp pool *pool-name*

5. Specify an IPv6 subnet for dynamic assignment.

   network { *prefix/prefix-length* | **prefix** *prefix-number* [*sub-prefix/sub-prefix-length*] } [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

   By default, no IPv6 subnet is specified for dynamic assignment.

   The IPv6 subnets cannot be the same in different address pools.

   If you specify an IPv6 prefix by its ID, make sure the IPv6 prefix is in effect. Otherwise, the configuration does not take effect.

6. Configure the prefix assignment. Choose the options to configure as needed:

   - Configure a static prefix binding:

     static-bind prefix *prefix*/*prefix-len* **duid** *duid* [**iaid** *iaid*] [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

By default, no static prefix binding is configured.

To add multiple static IPv6 prefix bindings, repeat this step.

- Apply the prefix pool to the address pool:

  prefix-pool *prefix-pool-number* [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

  By default, static or dynamic prefix assignment is not configured for an address pool.

Configuring IPv6 address assignment

About this task

Use one of the following methods to configure IPv6 address assignment:

- Configure a static IPv6 address binding in an address pool.

  If you bind a DUID and an IAID to an IPv6 address, the DUID and IAID in a request must match those in the binding before the DHCPv6 server can assign the IPv6 address to the requesting client. If you only bind a DUID to an IPv6 address, the DUID in a request must match the DUID in the binding before the DHCPv6 server can assign the IPv6 address to the requesting client.

- Specify a subnet and address ranges in an address pool.
  - **Non-temporary address assignment**—The server selects addresses from the non-temporary address range specified by the **address range** command. If no non-temporary address range is specified, the server selects addresses on the subnet specified by the **network** command.
  - **Temporary address assignment**—The server selects addresses from the temporary address range specified by the **temporary address range** command. If no temporary address range is specified in the address pool, the DHCPv6 server cannot assign temporary addresses to clients.

Restrictions and guidelines

- You can specify only one non-temporary address range and one temporary address range in an address pool.
- The address ranges specified by the **address range** and **temporary address range** commands must be on the subnet specified by the **network** command. Otherwise, the addresses are unassignable.
- An IPv6 address can be bound to only one DHCPv6 client. You cannot modify bindings that have been created. To change the binding for a DHCPv6 client, you must delete the existing binding first.
- Only one subnet can be specified in an address pool. If you use the **network** command multiple times in a DHCPv6 address pool, the most recent configuration takes effect. If you use this command to specify only new lifetimes, the settings do not affect existing leases. The IPv6 addresses assigned after the modification will use the new lifetimes.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Specify the IPv6 addresses excluded from dynamic assignment.

   ipv6 dhcp server forbidden-address *start-ipv6-address* [*end-ipv6-address*] [**vpn-instance** *vpn-instance-name*]

   By default, all IPv6 addresses except for the DHCPv6 server's IP address in a DHCPv6 address pool are assignable.

   If the excluded IPv6 address is in a static binding, the address still can be assigned to the client.

3. Enter DHCPv6 address pool view.

   ipv6 dhcp pool *pool-name*

4. Specify an IPv6 subnet for dynamic assignment.

   network { *prefix/prefix-length* | **prefix** *prefix-number* [*sub-prefix/sub-prefix-length*] } [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

   By default, no IPv6 address subnet is specified.

   The IPv6 subnets cannot be the same in different address pools.

   If you specify an IPv6 prefix by its ID, make sure the IPv6 prefix is in effect. Otherwise, the configuration does not take effect.

5. (Optional.) Specify a non-temporary IPv6 address range.

   address range *start-ipv6-address end-ipv6-address* [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

   By default, no non-temporary IPv6 address range is specified, and all unicast addresses on the subnet are assignable.

6. (Optional.) Specify a temporary IPv6 address range.

   temporary address range *start-ipv6-address end-ipv6-address* [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

   By default, no temporary IPv6 address range is specified, and the DHCPv6 server cannot assign temporary IPv6 addresses.

7. (Optional.) Enable EUI-64 address allocation mode.

   address-alloc-mode eui-64

   By default, EUI-64 address allocation mode is disabled.

   This feature enables the DHCPv6 server to obtain the client MAC address in the DHCP request and generates an EUI-64 IPv6 address to assign to the client.

8. (Optional.) Create a static binding.

   static-bind address *ipv6-address*/*addr-prefix-length* **duid** *duid* [**iaid** *iaid*] [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

   By default, no static binding is configured.

   To add more static bindings, repeat this step.

Configuring network parameters assignment

About network parameters assignment

In addition to IPv6 prefixes and IPv6 addresses, you can configure the following network parameters in an address pool:

- A maximum of eight DNS server addresses.
- One domain name.
- A maximum of eight SIP server addresses.
- A maximum of eight SIP server domain names.

You can configure network parameters on a DHCPv6 server by using one of the following methods:

- Configure network parameters in a DHCPv6 address pool.

- Configure network parameters in a DHCPv6 option group, and specify the option group for a DHCPv6 address pool.

Network parameters configured in a DHCPv6 address pool take precedence over those configured in a DHCPv6 option group.

Configuring network parameters in a DHCPv6 address pool

1. Enter system view.

   system-view

2. Enter DHCPv6 address pool view.

   ipv6 dhcp pool *pool-name*

3. Specify an IPv6 subnet for dynamic assignment.

   network { *prefix/prefix-length* | **prefix** *prefix-number* [*sub-prefix/sub-prefix-length*] } [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*]

   By default, no IPv6 subnet is specified.

   The IPv6 subnets cannot be the same in different address pools.

   If you specify an IPv6 prefix by its ID, make sure the IPv6 prefix is in effect. Otherwise, the configuration does not take effect.

4. Specify a DNS server address.

   dns-server *ipv6-address*

   By default, no DNS server address is specified.

5. Specify a domain name.

   domain-name *domain-name*

   By default, no domain name is specified.

6. Specify a SIP server address or domain name.

   sip-server { **address** *ipv6-address* | **domain-name** *domain-name* }

   By default, no SIP server address or domain name is specified.

7. Configure a self-defined DHCPv6 option.

   option *code* **hex** *hex-string*

   By default, no self-defined DHCPv6 option is configured.

Configuring network parameters in a DHCPv6 option group

About this task

A DHCPv6 option group can be created by using the following methods:

- Create a static DHCPv6 option group by using the **ipv6 dhcp option-group** command. The static DHCPv6 option group takes precedence over the dynamic DHCPv6 option group.
- When the device acts as a DHCPv6 client, it automatically creates a dynamic DHCPv6 option group for saving the obtained parameters. For more information about creating a dynamic DHCPv6 option group, see "Configuring the DHCPv6 client."

Procedure

1. Enter system view.

   system-view

2. Create a static DHCPv6 option group and enter its view.

   ipv6 dhcp option-group *option-group-number*

3. Specify a DNS server address.
   dns-server *ipv6-address*
   By default, no DNS server address is specified.
4. Specify a domain name suffix.
   domain-name *domain-name*
   By default, no domain name suffix is specified.
5. Specify a SIP server address or domain name.
   sip-server { **address** *ipv6-address* | **domain-name** *domain-name* }
   By default, no SIP server address or domain name is specified.
6. Configure a self-defined DHCPv6 option.
   option *code* **hex** *hex-string*
   By default, no self-defined DHCPv6 option is configured.
7. Return to system view.
   quit
8. Enter DHCPv6 address pool view.
   ipv6 dhcp pool *pool-name*
9. Specify a DHCPv6 option group.
   option-group *option-group-number*
   By default, no DHCPv6 option group is specified.

Configuring the DHCPv6 server on an interface

About this task

Enable the DHCP server and configure one of the following address/prefix assignment methods on an interface:

- **Apply an address pool on the interface**—The DHCPv6 server selects an IPv6 address/prefix from the applied address pool for a requesting client. If there is no assignable IPv6 address/prefix in the address pool, the DHCPv6 server cannot assign an IPv6 address/prefix to a client.
- **Configure global address assignment on the interface**—The DHCPv6 server selects an IPv6 address/prefix in the global DHCPv6 address pool that matches the server interface address or the DHCPv6 relay agent address for a requesting client.

If you configure both methods on an interface, the DHCPv6 server uses the specified address pool for address assignment without performing global address assignment.

Restrictions and guidelines

- An interface cannot act as a DHCPv6 server and DHCPv6 relay agent at the same time.
- Do not enable DHCPv6 server and DHCPv6 client on the same interface.
- You can apply an address pool that has not been created to an interface. The setting takes effect after the address pool is created.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the DHCPv6 server on the interface.

ipv6 dhcp select server

By default, the interface does not act as a DHCP server or a DHCP relay agent, and discards DHCPv6 packets from DHCPv6 clients.

4. Configure an assignment method.
   - Configure global address assignment.

     ipv6 dhcp server { **allow-hint** | **preference** *preference-value* | **rapid-commit** } *

     By default, the DHCP server supports global address assignment, but it does not support desired address/prefix assignment or rapid assignment. The default preference is 0.
   - Apply a DHCPv6 address pool to the interface.

     ipv6 dhcp server apply pool *pool-name* [**allow-hint** | **preference** *preference-value* | **rapid-commit**] *

Configuring a DHCPv6 policy for IPv6 address and prefix assignment

About this task

In a DHCPv6 policy, each DHCPv6 user class has a bound DHCPv6 address pool. Clients matching different user classes obtain IPv6 addresses, IPv6 prefixes, and other parameters from different address pools. When receiving a DHCPv6 request, the DHCPv6 server compares the packet against the user classes in the order that they are configured.

If a match is found and the bound address pool has assignable IPv6 addresses or prefixes, the server uses the address pool for assignment. If the bound address pool does not have assignable IPv6 addresses or prefixes, the assignment fails.

If no match is found, the server uses the default DHCPv6 address pool for assignment. If no default address pool is specified or the default address pool does not have assignable IPv6 addresses or prefixes, the assignment fails.

For successful assignment, make sure the applied DHCPv6 policy and the bound address pools exist.

A match rule cannot match an option added by the DHCPv6 device, for example, Option 18 or Option 37.

Procedure

1. Enter system view.

   system-view
2. Create a DHCPv6 user class and enter DHCPv6 user class view.

   ipv6 dhcp class *class-name*
3. Configure a match rule for the DHCPv6 user class.

   if-match rule *rule-number* { **option** *option-code* [**ascii** *ascii-string* [**offset** *offset* | **partial**] | **hex** *hex-string* [**mask** *mask* | **offset** *offset* **length** *length* | **partial**]] | **relay-agent** *gateway-ipv6-address* }

   By default, no match rule is configured for a DHCPv6 user class.
4. Return to system view.

   quit
5. Create a DHCPv6 policy and enter DHCPv6 policy view.

   ipv6 dhcp policy *policy-name*

The DHCPv6 policy takes effect only after it is applied to the interface that acts as the DHCPv6 server.

6. Specify a DHCPv6 address pool for a DHCPv6 user class.
   class *class-name* **pool** *pool-name*
   By default, no address pool is specified for a user class.
7. (Optional.) Specify the default DHCPv6 address pool.
   default pool *pool-name*
   By default, the default address pool is not specified.
8. Return to system view.
   quit
9. Enter interface view.
   interface *interface-type interface-number*
10. Apply the DHCPv6 policy to the interface.
    ipv6 dhcp apply-policy *policy-name*
    By default, no DHCPv6 policy is applied to an interface.

Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 server

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet.

Procedure

1. Enter system view.
   system-view
2. Set the DSCP value for DHCPv6 packets sent by the DHCPv6 server.
   ipv6 dhcp dscp *dscp-value*
   By default, the DSCP value in DHCPv6 packets sent by the DHCPv6 server is 56.

Configuring DHCPv6 binding auto backup

About this task

The auto backup feature saves DHCPv6 bindings to a backup file, and allows the DHCPv6 server to download the bindings from the backup file at the server reboot. The bindings include the lease bindings and conflicted IPv6 addresses. They cannot survive a reboot on the DHCPv6 server.

The DHCPv6 server does not provide services during the download process. If a connection error occurs during the process and cannot be repaired in a short amount of time, you can terminate the download operation. Manual interruption allows the DHCPv6 server to provide services without waiting for the connection to be repaired.

Procedure

1. Enter system view.
   system-view
2. Configure the DHCPv6 server to back up the bindings to a file.

ipv6 dhcp server database filename { *filename* | **url** *url* [**username** *username* [**password** { **cipher** | **simple** } *string*]] }

By default, the DHCPv6 server does not back up the DHCPv6 bindings.

With this command executed, the DHCPv6 server backs up its bindings immediately and runs auto backup.

3. (Optional.) Manually save the DHCPv6 bindings to the backup file.

   ipv6 dhcp server database update now

4. (Optional.) Set the waiting time after a DHCPv6 binding change for the DHCPv6 server to update the backup file.

   ipv6 dhcp server database update interval *interval*

   By default, the DHCP server waits 300 seconds to update the backup file after a DHCP binding change. If no DHCP binding changes, the backup file is not updated.

5. (Optional.) Terminate the download of DHCPv6 bindings from the backup file.

   ipv6 dhcp server database update stop

   This command only triggers one termination.

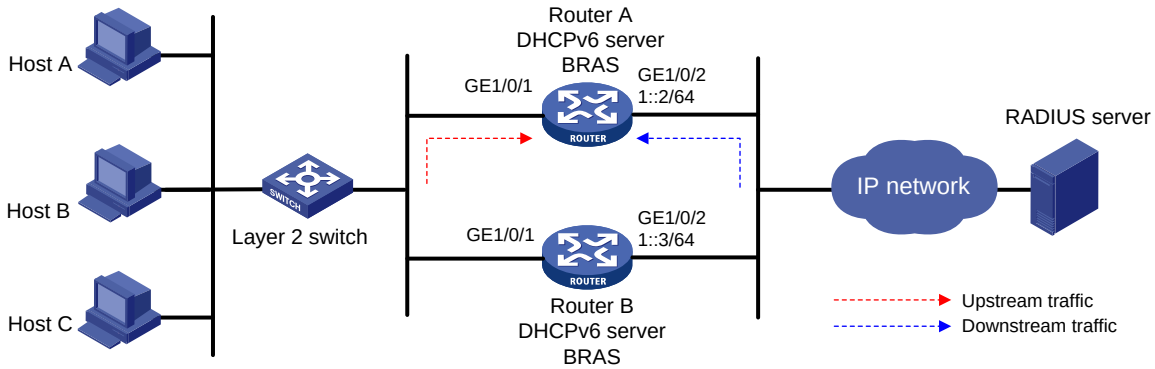
Advertising subnets assigned to clients

About this task

This feature enables the route management module to advertise subnets assigned to DHCPv6 clients. Use this feature in a node redundancy scenario to make sure the bidirectional traffic of each DHCPv6 client traverses the same node for purposes such as traffic accounting.

Figure 12 shows a typical scenario in which Router A and Router B act as both DHCPv6 servers and BRAS devices. The BRAS devices send accounting packets to the RADIUS server for each client. Upon receipt of the accounting packets, the RADIUS server overwrites the previous traffic statistics for that client with the new statistics in the received packets. If the upstream and downstream traffic of a client traverses different BRAS devices, the RADIUS server will be unable to collect complete traffic statistics for that client. To make sure the RADIUS server collect complete traffic statistics for each client from the BRAS devices, configure the DHCPv6 servers to advertise subnets assigned to clients. This configuration makes sure the upstream and downstream traffic of a client traverses the same BRAS device.

Figure 12 Network diagram



Restrictions and guidelines

If the address pool is bound to a VPN instance, you must make sure the VPN instance already exists for the subnet advertisement feature to take effect.

Procedure

1. Enter system view.
   system-view
2. Enter DHCP address pool view.
   ipv6 dhcp pool *pool-name*
3. Advertise the subnet that is assigned to DHCPv6 clients.
   network { *prefix/prefix-length* | **prefix** *prefix-number* [*sub-prefix/sub-prefix-length*] } [**preferred-lifetime** *preferred-lifetime* **valid-lifetime** *valid-lifetime*] **export-route**

   By default, the subnet assigned to DHCPv6 clients is not advertised.

Enabling the DHCPv6 server to advertise IPv6 prefixes

About this task

A DHCPv6 client can obtain an IPv6 prefix through DHCPv6 and use this IPv6 prefix to assign IPv6 addresses for clients in a downstream network. If the IPv6 prefix is in a different subnet than the IPv6 address of the DHCPv6 client's upstream interface, the clients in the downstream network cannot access the external network. If the DHCPv6 server is on the same link as the DHCPv6 client, enable the DHCPv6 server to advertise the IPv6 prefix.

Procedure

1. Enter system view.
   system-view
2. Enable the DHCPv6 server to advertise IPv6 prefixes.
   ipv6 dhcp advertise pd-route

   By default, the DHCPv6 server does not advertise IPv6 prefixes.

Configuring DHCPv6 vendor-specific options

About this task

In addition to the options defined by RFC, the DHCPv6 server also supports a group of options that require manual configuration. These options are known as vendor-specific options.

The DHCPv6 server can use vendor-specific options to assign additional network settings to a DHCPv6 client. These network settings include TFTP server information and the client's configuration file name. After receiving a DHCPv6 reply that carries vendor-specific options, the DHCPv6 client can parse these options to obtain related information.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 address pool view.
   ipv6 dhcp pool *pool-name*
3. Configure a vendor-specific option and enter its view.
   vendor-specific *vendor-id*

   By default, no vendor-specific options are configured.

4. Configure a DHCPv6 vendor-specific suboption.

   suboption *suboption-code* { **address** *ipv6-address*&<1-4> | **ascii** *ascii-string* | **hex** *hex-string* }

   By default, no vendor-specific suboptions are configured.

Applying a DHCPv6 address pool to a VPN instance

About this task

If a DHCPv6 address pool is applied to a VPN instance, the DHCPv6 server assigns IPv6 addresses in this address pool to clients in the VPN instance. Addresses in this address pool will not be assigned to clients on the public network.

The DHCPv6 server can obtain the VPN instance to which a DHCPv6 client belongs from the following information:

- The client's VPN information stored in authentication modules.
- The VPN information of the DHCPv6 server's interface that receives DHCPv6 packets from the client.

The VPN information from authentication modules takes priority over the VPN information of the receiving interface.

An MCE acting as the DHCP server can assign IP addresses not only to clients on public networks, but also to clients on private networks. The IP address ranges of public and private networks or those of private networks on the DHCP server cannot overlap. For more information about MCE, see *MCE Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter DHCP address pool view.

   ipv6 dhcp pool *pool-name*

3. Apply the address pool to a VPN instance.

   vpn-instance *vpn-instance-name*

   By default, the address pool is not applied to any VPN instance.

Configuring DHCPv6 flood attack protection

About this task

The DHCPv6 flood attack protection enables the DHCPv6 server to detect DHCPv6 flood attacks according to the DHCPv6 packet rate threshold on a per-MAC basis.

When the DHCPv6 server receives a DHCPv6 packet from a client (MAC address), it creates a DHCPv6 flood attack entry in check state. If the number of DHCPv6 packets from the same MAC address reaches the upper limit in the detection duration, the server determines that the client is launching a DHCPv6 flood attack. The DHCPv6 flood attack entry changes to the restrain state, and the DHCPv6 server discards the DHCPv6 packets from that client. When the aging time of the entry is reached, the DHCPv6 server deletes the entry. If a DHCPv6 packet from the MAC address arrives later, the DHCPv6 server will create a flood attack entry and count the number of incoming DHCP packets for that client again.

This feature is available only in VXLAN networks. For more information about VXLANs, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Set the DHCPv6 packet rate threshold for DHCPv6 flood attack detection.
   ipv6 dhcp flood-protection threshold *packet-number milliseconds*
   By default, the device allows a maximum of 6 DHCPv6 packets per 5000 milliseconds from each DHCPv6 client.
3. (Optional.) Set the DHCP flood attack entry aging time.
   ipv6 dhcp flood-protection aging-time *time*
   The default setting is 300 seconds.
4. Enter VSI view.
   vsi *vsi-name*
5. Enable DHCPv6 flood attack protection.
   ipv6 dhcp flood-protection enable
   By default, DHCPv6 flood attack protection is disabled.

Enabling IPv6 address binding conversion for IP source guard

About this task

In some networks, the gateway acts as the DHCPv6 server to assign IPv6 addresses to users and reports user information to the controller. The user information report supports only IP source guard entries based on which the controller manages and monitors users. For the DHCPv6 server to report information about all online DHCP users, use this feature to convert user IPv6 address bindings to dynamic IP source guard bindings.

For more information about IP source guard, see *Security Configuration Guide*.

Restrictions and guidelines

If the device does not have enough storage space, execute the **undo ipv6 dhcp server entry-convert enable** command to disable this feature. The system does not delete IP source guard bindings that have been converted before you disable this feature.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 address binding conversion for IP source guard.
   ipv6 dhcp server entry-convert enable
   By default, IPv6 address binding conversion for IP source guard is disabled.

Configuring SNMP notifications for the DHCPv6 server

About SNMP DHCPv6 server notifications

Perform this task to configure the DHCPv6 module to send SNMP notifications to report DHCPv6 server events, including:

- Resource exhaustion events or recoveries from resource exhaustion conditions in an IPv6 address pool.
- Resource usage threshold violations or recoveries from threshold violation conditions in an IPv6 pool.

The SNMP notifications are sent to the SNMP module. For the SNMP notifications to be sent correctly, you must also configure SNMP. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Enabling IPv6 resource exhaustion notifications

1. Enter system view.
   system-view
2. Enable IPv6 resource exhaustion notifications.
   snmp-agent trap enable ipv6 dhcp server address-exhaust pd-exhaust
   By default, IPv6 resource exhaustion notifications are enabled.
3. Enter IPv6 address pool view.
   ipv6 dhcp pool *pool-name*
4. (Optional.) Enable IPv6 resource exhaustion alarming for an IPv6 pool.
   exhaustion trap enable
   By default, IPv6 resource exhaustion alarming is enabled for an IPv6 pool.

Enabling IPv6 address usage alarm notifications

1. Enter system view.
   system-view
2. Enable IPv6 address usage alarm notifications.
   snmp-agent trap enable ipv6 dhcp server ip-in-use
   By default, IPv6 address usage alarm notifications are enabled.
3. Enter IPv6 address pool view.
   ipv6 dhcp pool *pool-name*
4. (Optional.) Set the IPv6 address usage threshold.
   ip-in-use threshold *threshold-value*
   The default threshold is 100%.

Enabling IPv6 prefix usage alarm notifications

1. Enter system view.
   system-view

2. Enable IPv6 prefix usage alarm notifications.

   snmp-agent trap enable ipv6 dhcp server pd-in-use

   By default, IPv6 prefix usage alarm notifications are enabled.
3. Enter IPv6 address pool view.

   ipv6 dhcp pool *pool-name*
4. (Optional.) Set the IPv6 prefix usage threshold.

   pd-in-use threshold *threshold-value*

   The default threshold is 100%.

Enabling DHCPv6 logging on the DHCPv6 server

About this task

The DHCPv6 logging feature enables the DHCPv6 server to generate DHCPv6 logs and send them to the information center. The information helps administrators locate and solve problems. For information about the log destination and output rule configuration in the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

As a best practice, disable this feature if the log generation affects the device performance or reduces the address and prefix allocation efficiency. For example, this situation might occur when a large number of clients frequently come online or go offline.

Procedure

1. Enter system view.

   system-view
2. Enable DHCPv6 logging.

   ipv6 dhcp log enable

   By default, DHCPv6 logging is disabled.

Enabling IPv6 resource exhaustion logging

About this task

This feature enables the DHCPv6 module to send IPv6 resource exhaustion log messages to the information center. For log messages to be sent correctly, configure the information center to set log message filtering and output rules, including output destinations. For information about the information center configuration, see *Network Management and Monitoring Configuration Guide*.

The DHCPv6 module generates an IPv6 resource exhaustion log message when the IPv6 address pool encounters one of the following events:

NOTE:

If the IPv6 address usage or prefix usage threshold for the pool is 100%, the DHCPv6 module generates an alarm message instead of a log message for one of the following events.

To set the IPv6 address usage threshold for an IPv6 address pool, use the **ip-in-use threshold** command. To set the IPv6 prefix usage threshold for an IPv6 address pool, use the **pd-in-use threshold** command.

- The pool does not have assignable address resources or prefix resources.
- The IPv6 address usage or prefix usage of the pool drops to or below 90% after exhaustion.

You can use the following formulas to calculate the IPv6 resource usage of an IPv6 address pool:

- *IPv6 address usage = (total number of IPv6 addresses – number of assignable IPv6 addresses)/total number of IPv6 addresses*
- *IPv6 prefix usage = (total number of IPv6 prefixes – number of assignable IPv6 prefixes)/total number of IPv6 prefixes*

Restrictions and guidelines

This feature is not affected if DHCPv6 logging is disabled.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 address pool view.
   ipv6 dhcp pool *pool-name*
3. Enable IPv6 resource exhaustion logging.
   exhaustion log enable
   By default, IPv6 resource exhaustion logging is disabled.

Display and maintenance commands for DHCPv6 server

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the DUID of the local device.	**display ipv6 dhcp duid**
Display information about a DHCPv6 option group.	**display ipv6 dhcp option-group** [*option-group-number*]
Display DHCPv6 address pool information.	**display ipv6 dhcp pool** [*pool-name* | **vpn-instance** *vpn-instance-name*]
Display prefix pool information.	**display ipv6 dhcp prefix-pool** [*prefix-pool-number*] [**vpn-instance** *vpn-instance-name*]
Display DHCPv6 server information on an interface.	**display ipv6 dhcp server** [**interface** *interface-type interface-number*]
Display information about IPv6 address conflicts.	**display ipv6 dhcp server conflict** [**address** *ipv6-address*] [**vpn-instance** *vpn-instance-name*]
Display information about DHCPv6 binding auto backup	**display ipv6 dhcp server database**
Display information about expired IPv6 addresses.	**display ipv6 dhcp server expired** [[**address** *ipv6-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Display information about IPv6 address bindings.	**display ipv6 dhcp server ip-in-use** [[**address** *ipv6-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Display information about IPv6 prefix	**display ipv6 dhcp server pd-in-use** [**pool**

Task	Command
bindings.	*pool-name* | [**prefix** *prefix/prefix-len*] [**vpn-instance** *vpn-instance-name*]]
Display packet statistics on the DHCPv6 server.	**display ipv6 dhcp server statistics** [**pool** *pool-name* | **vpn-instance** *vpn-instance-name*]
Clear information about IPv6 address conflicts.	**reset ipv6 dhcp server conflict** [**address** *ipv6-address*] [**vpn-instance** *vpn-instance-name*]
Clear information about expired IPv6 address bindings.	**reset ipv6 dhcp server expired** [[**address** *ipv6-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Clear information about IPv6 address bindings.	**reset ipv6 dhcp server ip-in-use** [[**address** *ipv6-address*] [**vpn-instance** *vpn-instance-name*] | **pool** *pool-name*]
Clear information about IPv6 prefix bindings.	**reset ipv6 dhcp server pd-in-use** [**pool** *pool-name* | [**prefix** *prefix/prefix-len*] [**vpn-instance** *vpn-instance-name*]]
Clear packets statistics on the DHCPv6 server.	**reset ipv6 dhcp server statistics** [**vpn-instance** *vpn-instance-name*]

DHCPv6 server configuration examples

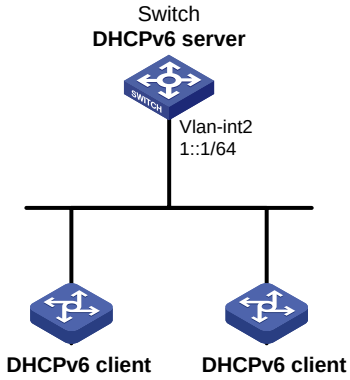
Example: Configuring dynamic IPv6 prefix assignment

Network configuration

As shown in Figure 13, the switch acts as a DHCPv6 server to assign an IPv6 prefix, a DNS server address, a domain name, a SIP server address, and a SIP server name to each DHCPv6 client.

The switch assigns prefix 2001:0410:0201::/48 to the client whose DUID is 00030001CA0006A40000, and assigns prefixes in the range of 2001:0410::/48 to 2001:0410:FFFF::/48 (excluding 2001:0410:0201::/48) to other clients. The DNS server address is 2::2:3. The DHCPv6 clients reside in the domain **aaa.example.com**. The SIP server address is 2:2::4, and the SIP server name is **bbb.example.com**.

Figure 13 Network diagram



Procedure

Specify an IPv6 address for VLAN-interface 2.

```
<Switch> system-view
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ipv6 address 1::1/64
```

Disable RA message suppression on VLAN-interface 2.

```
[Switch-Vlan-interface2] undo ipv6 nd ra halt
```

Set the M flag to 1 in RA advertisements to be sent on VLAN-interface 2. Hosts that receive the RA advertisements will obtain IPv6 addresses through DHCPv6.

```
[Switch-Vlan-interface2] ipv6 nd autoconfig managed-address-flag
```

Set the O flag to 1 in RA advertisements to be sent on VLAN-interface 2. Hosts that receive the RA advertisements will obtain information other than IPv6 address through DHCPv6.

```
[Switch-Vlan-interface2] ipv6 nd autoconfig other-flag
[Switch-Vlan-interface2] quit
```

Create prefix pool 1, and specify the prefix 2001:0410::/32 with the assigned prefix length 48.

```
[Switch] ipv6 dhcp prefix-pool 1 prefix 2001:0410::/32 assign-len 48
```

Create address pool 1.

```
[Switch] ipv6 dhcp pool 1
```

In address pool 1, configure subnet 1::/64 where VLAN interface-2 resides.

```
[Switch-dhcp6-pool-1] network 1::/64
```

Apply prefix pool 1 to address pool 1, and set the preferred lifetime to one day, and the valid lifetime to three days.

```
[Switch-dhcp6-pool-1] prefix-pool 1 preferred-lifetime 86400 valid-lifetime 259200
```

In address pool 1, bind prefix 2001:0410:0201::/48 to the client DUID 00030001CA0006A40000, and set the preferred lifetime to one day, and the valid lifetime to three days.

```
[Switch-dhcp6-pool-1] static-bind prefix 2001:0410:0201::/48 duid 00030001CA0006A40000
preferred-lifetime 86400 valid-lifetime 259200
```

Configure the DNS server address 2:2::3.

```
[Switch-dhcp6-pool-1] dns-server 2:2::3
```

Configure the domain name as **aaa.example.com**.

```
[Switch-dhcp6-pool-1] domain-name aaa.example.com
```

Configure the SIP server address as 2:2::4, and the SIP server name as **bbb.example.com**.

```
[Switch-dhcp6-pool-1] sip-server address 2:2::4
[Switch-dhcp6-pool-1] sip-server domain-name bbb.example.com
[Switch-dhcp6-pool-1] quit
```

Enable the DHCPv6 server on VLAN-interface 2, enable desired prefix assignment and rapid prefix assignment, and set the preference to the highest.

```
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ipv6 dhcp select server
[Switch-Vlan-interface2] ipv6 dhcp server allow-hint preference 255 rapid-commit
```

Verifying the configuration

Display DHCPv6 server configuration on VLAN-interface 2.

```
[Switch-Vlan-interface2] display ipv6 dhcp server interface vlan-interface 2
Using pool: global
Preference value: 255
Allow-hint: Enabled
```

```
Rapid-commit: Enabled
```

Display information about address pool 1.

```
[Switch-Vlan-interface2] display ipv6 dhcp pool 1
DHCPv6 pool: 1
  Network: 1::/64
    Preferred lifetime 604800 seconds, valid lifetime 2592000 seconds
  Prefix pool: 1
    Preferred lifetime 86400 seconds, valid lifetime 259200 seconds
  Static bindings:
    DUID: 00030001ca0006a40000
    IAID: Not configured
    Prefix: 2001:410:201::/48
      Preferred lifetime 86400 seconds, valid lifetime 259200 seconds
  DNS server addresses:
    2:2::3
  Domain name:
    aaa.example.com
  SIP server addresses:
    2:2::4
  SIP server domain names:
    bbb.example.com
  undo exhaustion log enable
  IP-in-use threshold: 100
  PD-in-use threshold: 100
  exhaustion trap enable
```

Display information about prefix pool 1.

```
[Switch-Vlan-interface2] display ipv6 dhcp prefix-pool 1
Prefix: 2001:410::/32
Assigned length: 48
Total prefix number: 65536
Available: 65535
In-use: 0
Static: 1
```

After the client with the DUID 00030001CA0006A40000 obtains an IPv6 prefix, display the binding information on the DHCPv6 server.

```
[Switch-Vlan-interface2] display ipv6 dhcp server pd-in-use
Pool: 1
 IPv6 prefix                              Type      Lease expiration
 2001:410:201::/48                        Static(C) Jul 10 19:45:01 2019
```

After the other client obtains an IPv6 prefix, display binding information on the DHCPv6 server.

```
[Switch-Vlan-interface2] display ipv6 dhcp server pd-in-use
Pool: 1
 IPv6 prefix                              Type      Lease expiration
 2001:410:201::/48                        Static(C) Jul 10 19:45:01 2019
 2001:410::/48                            Auto(C)   Jul 10 20:44:05 2019
```

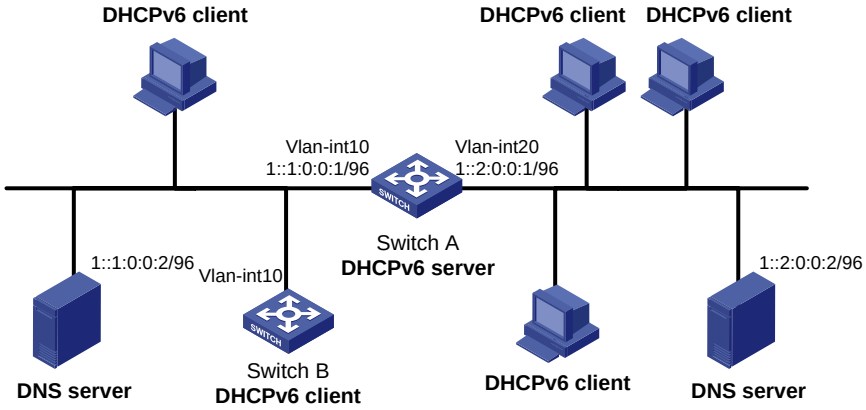
Example: Configuring dynamic IPv6 address assignment

Network configuration

As shown in Figure 14, Switch A acts as a DHCPv6 server to assign IPv6 addresses to the clients on subnets 1::1:0:0:0/96 and 1::2:0:0:0/96.

On Switch A, configure the IPv6 address 1::1:0:0:1/96 for VLAN-interface 10 and 1::2:0:0:1/96 for VLAN-interface 20. The lease duration of the addresses on subnet 1::1:0:0:0/96 is 172800 seconds (two days), the valid time is 345600 seconds (four days), the domain name suffix is aabbcc.com, and the DNS server address is 1::1:0:0:2/96. The lease duration of the addresses on subnet 1::2:0:0:0/96 is 432000 seconds (five days), the valid time is 864000 seconds (ten days), the domain name is aabbcc.com, and the DNS server address is 1::2:0:0:2/96.

Figure 14 Network diagram



Procedure

1. Configure the interfaces on the DHCPv6 server:

   # Specify an IPv6 address for VLAN-interface 10.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 10
   [SwitchA-Vlan-interface10] ipv6 address 1::1:0:0:1/96
   ```

   # Disable RA message suppression on VLAN-interface 10.

   ```
   [SwitchA-Vlan-interface10] undo ipv6 nd ra halt
   ```

   # Set the M flag to 1 in RA advertisements to be sent on VLAN-interface 10. Hosts that receive the RA advertisements will obtain IPv6 addresses through DHCPv6.

   ```
   [SwitchA-Vlan-interface10] ipv6 nd autoconfig managed-address-flag
   ```

   # Set the O flag to 1 in RA advertisements to be sent on VLAN-interface 10. Hosts that receive the RA advertisements will obtain information other than IPv6 address through DHCPv6.

   ```
   [SwitchA-Vlan-interface10] ipv6 nd autoconfig other-flag
   [SwitchA-Vlan-interface10] quit
   ```

   # Specify an IPv6 address for VLAN-interface 20.

   ```
   [SwitchA] interface vlan-interface 20
   [SwitchA-Vlan-interface20] ipv6 address 1::2:0:0:1/96
   ```

   # Disable RA message suppression on VLAN-interface 20.

   ```
   [SwitchA-Vlan-interface20] undo ipv6 nd ra halt
   ```

   # Set the M flag to 1 in RA advertisements to be sent on VLAN-interface 20. Hosts that receive the RA advertisements will obtain IPv6 addresses through DHCPv6.

   ```
   [SwitchA-Vlan-interface20] ipv6 nd autoconfig managed-address-flag
   ```

Set the O flag to 1 in RA advertisements to be sent on VLAN-interface 20. Hosts that receive the RA advertisements will obtain information other than IPv6 address through DHCPv6.

```
[SwitchA-Vlan-interface20] ipv6 nd autoconfig other-flag
[SwitchA-Vlan-interface20] quit
```

2. Enable DHCPv6:

Enable DHCPv6 server on VLAN-interface 10 and VLAN-interface 20.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ipv6 dhcp select server
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] ipv6 dhcp select server
[SwitchA-Vlan-interface20] quit
```

Exclude the DNS server addresses from dynamic assignment.

```
[SwitchA] ipv6 dhcp server forbidden-address 1::1:0:0:2
[SwitchA] ipv6 dhcp server forbidden-address 1::2:0:0:2
```

Configure the DHCPv6 address pool 1 to assign IPv6 addresses and other configuration parameters to clients on subnet 1::1:0:0:0/96.

```
[SwitchA] ipv6 dhcp pool 1
[SwitchA-dhcp6-pool-1] network 1::1:0:0:0/96 preferred-lifetime 172800
valid-lifetime 345600
[SwitchA-dhcp6-pool-1] domain-name aabbcc.com
[SwitchA-dhcp6-pool-1] dns-server 1::1:0:0:2
[SwitchA-dhcp6-pool-1] quit
```

Configure the DHCPv6 address pool 2 to assign IPv6 addresses and other configuration parameters to clients on subnet 1::2:0:0:0/96.

```
[SwitchA] ipv6 dhcp pool 2
[SwitchA-dhcp6-pool-2] network 1::2:0:0:0/96 preferred-lifetime 432000
valid-lifetime 864000
[SwitchA-dhcp6-pool-2] domain-name aabbcc.com
[SwitchA-dhcp6-pool-2] dns-server 1::2:0:0:2
[SwitchA-dhcp6-pool-2] quit
```

Verifying the configuration

Verify that the clients on subnets 1::1:0:0:0/96 and 1::2:0:0:0/96 can obtain IPv6 addresses and all other configuration parameters from the DHCPv6 server (Switch A). (Details not shown.)

On the DHCPv6 server, display IPv6 addresses assigned to the DHCPv6 clients.

```
[SwitchA] display ipv6 dhcp server ip-in-use
```

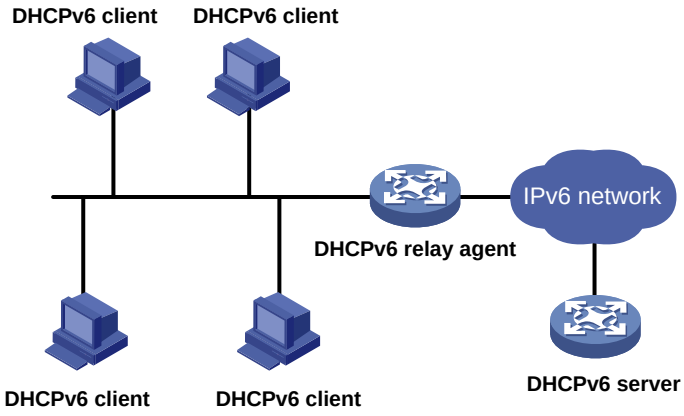
Configuring the DHCPv6 relay agent

About DHCPv6 relay agent

Typical application

A DHCPv6 client usually uses a multicast address to contact the DHCPv6 server on the local link to obtain an IPv6 address and other configuration parameters. As shown in Figure 15, if the DHCPv6 server resides on another subnet, the DHCPv6 clients need a DHCPv6 relay agent to contact the server. The relay agent feature avoids deploying a DHCPv6 server on each subnet.

Figure 15 Typical DHCPv6 relay agent application

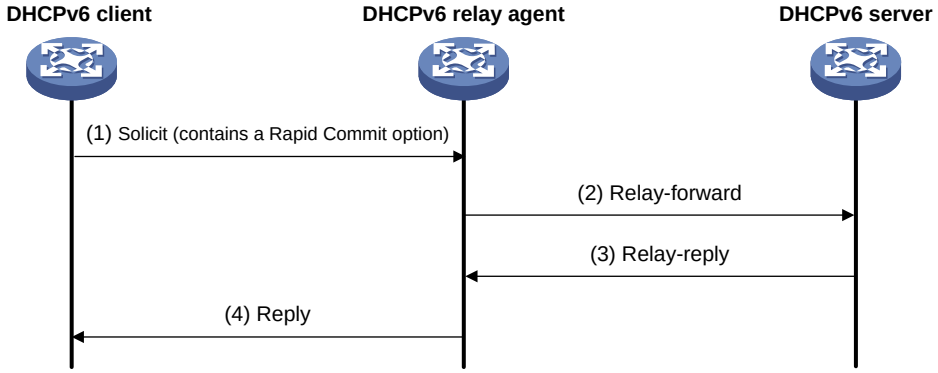


DHCPv6 relay agent operating process

As shown in Figure 16, a DHCPv6 client obtains an IPv6 address and other network configuration parameters from a DHCPv6 server through a DHCPv6 relay agent. The following example uses rapid assignment to describe the process:

- The DHCPv6 client sends a Solicit message containing the Rapid Commit option to the multicast address FF02::1:2 of all the DHCPv6 servers and relay agents.
- After receiving the Solicit message, the DHCPv6 relay agent encapsulates the message into the Relay Message option of a Relay-forward message, and sends the message to the DHCPv6 server.
- After obtaining the Solicit message from the Relay-forward message, the DHCPv6 server performs the following tasks:
  - Selects an IPv6 address and other required parameters.
  - Adds them to a reply that is encapsulated within the Relay Message option of a Relay-reply message.
  - Sends the Relay-reply message to the DHCPv6 relay agent.
- The DHCPv6 relay agent obtains the reply from the Relay-reply message and sends the reply to the DHCPv6 client.
- The DHCPv6 client uses the IPv6 address and other network parameters assigned by the DHCPv6 server to complete network configuration.

Figure 16 Operating process of a DHCPv6 relay agent



DHCPv6 relay agent tasks at a glance

To configure a DHCPv6 relay agent, perform the following tasks:

1. Enabling the DHCPv6 relay agent on an interface
2. Specifying DHCPv6 servers on the relay agent
3. (Optional.) Specifying a gateway address for DHCPv6 clients
4. (Optional.) Setting the DUID entry aging timer for smart relay
5. (Optional.) Configuring DHCPv6 smart relay for address acquisition from different IPv6 subnets for a client
6. (Optional.) Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 relay agent
7. (Optional.) Specifying a padding mode for the Interface-ID option
8. (Optional.) Configuring the DHCPv6 relay agent to add Option 37 in Relay-forward messages
9. (Optional.) Enabling the DHCPv6 relay agent to support Option 79
10. (Optional.) Enabling the DHCPv6 relay agent to advertise IPv6 prefixes
11. (Optional.) Enabling advertising host routes for assigned IPv6 addresses
12. (Optional.) Specifying the source IPv6 address for relayed DHCPv6 requests
13. (Optional.) Configuring DHCPv6 relay security features
14. (Optional.) Enabling checking the Reply messages in response to Confirm messages
15. (Optional.) Discarding DHCPv6 requests that are delivered from VXLAN tunnels
16. (Optional.) Configuring support for forwarding DHCPv6 replies based on MAC address table

Enabling the DHCPv6 relay agent on an interface

Restrictions and guidelines

As a best practice, do not enable DHCPv6 relay agent and DHCPv6 client on the same interface.

Procedure

1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Enable DHCPv6 relay agent on the interface.

ipv6 dhcp select relay

By default, the DHCPv6 relay agent is disabled on the interface.

Specifying DHCPv6 servers on the relay agent

Specifying DHCPv6 server IP addresses

Restrictions and guidelines

- You can use the **ipv6 dhcp relay server-address** command to specify a maximum of eight DHCPv6 servers on the DHCPv6 relay agent interface. The DHCPv6 relay agent forwards DHCP requests to all the specified DHCPv6 servers.
- If a DHCPv6 server address is a link-local address or multicast address, you must specify an outgoing interface by using the **interface** keyword in this command. Otherwise, DHCPv6 packets might fail to reach the DHCPv6 server.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify a DHCPv6 server.

   ipv6 dhcp relay server-address *ipv6-address* [**interface** *interface-type interface-number*] [**public** | **vpn-instance** *vpn-instance-name*]

   By default, no DHCPv6 server is specified.

Specifying DHCPv6 servers for a DHCPv6 address pool on the DHCPv6 relay agent

About this task

This feature allows DHCPv6 clients of the same type to obtain IPv6 addresses, IPv6 prefixes, and other configuration parameters from the DHCPv6 servers in the matching DHCPv6 address pool.

It applies to scenarios where the DHCPv6 relay agent connects to clients of the same access type but classified into different types by their locations. In this case, the relay interface typically has no IPv6 address configured. You can use the **gateway-list** command to specify the gateway addresses for clients matching the same DHCPv6 address pool.

Upon receiving a DHCPv6 Solicit or Request from a client that matches a DHCPv6 address pool, the relay agent processes the packet as follows:

- Fills the **link-address** field of the packet with a specified gateway address.
- Forwards the packet to all DHCPv6 servers in the matching DHCPv6 address pool.

The DHCPv6 servers select a DHCPv6 address pool according to the gateway address.

Restrictions and guidelines

- You can specify a maximum of eight DHCPv6 servers for one DHCPv6 address pool for high availability. The relay agent forwards DHCPv6 Solicit and Request packets to all DHCPv6 servers in the DHCPv6 address pool.

Procedure

1. Enter system view.

system-view

2. Create a DHCPv6 address pool and enter its view.

   ipv6 dhcp pool *pool-name*

3. Specify gateway addresses for the clients matching the DHCPv6 address pool.

   gateway-list *ipv6-address*&<1-8>

   By default, no gateway address is specified.

4. Specify DHCPv6 servers for the DHCPv6 address pool.

   remote-server *ipv6-address* [**interface** *interface-type interface-number*]

   By default, no DHCPv6 server is specified for the DHCPv6 address pool.

Specifying the DHCPv6 server selection algorithm

About this task

The DHCPv6 relay agent supports the **polling** and **master-backup algorithms for** DHCPv6 server selection.

By default, the DHCPv6 relay agent uses the **polling** algorithm. It forwards DHCPv6 requests to all DHCPv6 servers. The DHCPv6 clients select the DHCPv6 server from which the first received DHCP reply comes.

If the DHCPv6 relay agent uses the **master-backup** algorithm, it forwards DHCP requests to the master DHCPv6 server first. If the master DHCPv6 server is not available, the relay agent forwards the subsequent DHCPv6 requests to a backup DHCPv6 server. If the backup DHCPv6 server is not available, the relay agent selects the next backup DHCPv6 server, and so on. If none of the backup DHCPv6 servers are available, it repeats the process, starting from the master DHCPv6 server.

DHCPv6 server selection also supports the following mechanisms:

- **DHCPv6 server response timeout time**—The DHCPv6 relay agent determines that a DHCPv6 server is not available if it does not receive any response from the server within the DHCPv6 server response timeout time. The DHCPv6 server response timeout time is configurable and the default is 30 seconds.
- **DHCPv6 server switchback**—If the DHCPv6 relay agent selects a backup DHCPv6 server, it does not switch back to the master DHCPv6 server by default. You can configure the DHCPv6 relay agent to switch back to the master DHCPv6 server after a delay. If the master DHCPv6 server is available, the DHCPv6 relay agent forwards DHCPv6 requests to the master DHCPv6 server. If the master DHCPv6 server is not available or does not have idle addresses for assignment, the DHCPv6 relay agent still uses the backup DHCP server.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify the DHCPv6 server selection algorithm.

   ipv6 dhcp relay server-address algorithm { **master-backup** | **polling** }

   By default, the polling algorithm is used. The DHCPv6 relay agent forwards DHCPv6 requests to all DHCPv6 servers.

4. (Optional.) Set the DHCPv6 server response timeout time for DHCP server switchover.

   ipv6 dhcp relay dhcp-server timeout *time*

   By default, the DHCPv6 server response timeout time is 30 seconds.

5. (Optional.) Enable the switchback to the master DHCPv6 server and set the delay.

   ipv6 dhcp relay master-server switch-delay *delay-time*

   By default, the DHCPv6 relay agent does not switch back to the master DHCPv6 server.

Specifying a gateway address for DHCPv6 clients

About this task

By default, the DHCPv6 relay agent fills the **link-address** field of DHCPv6 Solicit and Request packets with the first IPv6 address of the relay interface. You can specify a gateway address on the relay agent for DHCPv6 clients. The DHCPv6 relay agent uses the specified gateway address to fill the **link-address** field of DHCPv6 Solicit and Request packets.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify a gateway address for DHCPv6 clients.

   ipv6 dhcp relay gateway *ipv6-address*

   By default, the DHCPv6 relay agent uses the first IPv6 address of the relay interface as the clients' gateway address.

Setting the DUID entry aging timer for smart relay

About this task

This task sets the aging timer for entries in the DUID table for smart relay.

When receiving a Solicit message from a client, DHCPv6 smart relay uses the DUID table to determine the link address set in the Relay-forward message sent to forward the Solicit message.

The DHCPv6 relay DUID entry for a client contains the following information:

- MAC address of the DHCPv6 client.
- Index of the relay interface connected to the client.
- The remaining lifetime of the entry.
- Link address (one of the global unicast addresses on the relay interface).
- Number of Relay-forward messages sent with this link address to forward a Solicit message from the client.
- The amount of time that has elapsed since the first Relay-forward message was sent with that link address to forward a Solicit message from the client.

To avoid smart relay failure, make sure the aging timer for DUID entries is greater than the time-based link address change trigger set by using the **ipv6 dhcp smart-relay** command. For more information about DHCPv6 smart relay, see "Configuring DHCPv6 smart relay for address acquisition from different IPv6 subnets for a client."

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the aging timer for DUID entries on the DHCPv6 relay interface.

   ipv6 dhcp relay duid aging-time *seconds*

   By default, the aging timer for DUID entries is 60 seconds on the DHCPv6 relay agent.

Configuring DHCPv6 smart relay for address acquisition from different IPv6 subnets for a client

About this task

DHCPv6 smart relay enables a client to obtain an IPv6 address or prefix from one of the subnets to which the relay interface is attached in the following situations:

- The DHCPv6 server does not have free IPv6 addresses or prefixes in one subnet.
- The DHCPv6 server does not respond.

When receiving a Solicit message from a client on a relay interface, DHCPv6 smart relay forwards that message in a Relay-forward message to DHCPv6 servers. In the Relay-forward message, DHCPv6 smart relay sets the lowest global unicast address on the relay interface in the **link-address** field.

When a DHCPv6 server receives the Relay-forward message, it searches for free IPv6 address or prefix resources on the same subnet as the received link address.

- If free resources are available, the DHCPv6 server replies with a free address or prefix.
- If no free resources are available, the DHCPv6 server returns a NOADDRS-AVAIL or NOPREIFIX-AVAIL message.

Upon receipt of the NOADDRS-AVAIL or NOPREIFIX-AVAIL message, DHCPv6 smart relay changes the link address of the relay interface to forward the subsequent Solicit message. DHCPv6 smart relay selects one of the global unicast addresses on the relay interface in a round-robin manner to forward Solicit messages from the client until the client obtains an address.

When forwarding the Solicit message from a client, DHCPv6 smart relay agent also changes the link address if the following conditions are met:

- The number of Relay-forward messages sent with that link address for the client has reached the limit.
- The amount of time that has elapsed since the first Relay-forward message was sent with that link address has reached the limit.

Prerequisites

Before you enable DHCPv6 smart relay, you must first enable DHCPv6 relay entry recording by executing the **ipv6 dhcp relay client-information record** command. For more information about DHCPv6 relay entry recording, see "Enabling the DHCPv6 relay agent to record relay entries".

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. (Optional.) Configure smart relay.

   ipv6 dhcp smart-relay { **count** *count* | **time** *seconds* } *

   By default, DHCPv6 smart relay changes the link address if the following conditions are met:

   - That link address has been used in a minimum of three Relay-forward messages sent to forward the Solicit message from a client.

- A minimum of 24 seconds has elapsed since the first Relay-forward message was sent with that link address.

4. Enable DHCPv6 smart relay.

   ipv6 dhcp smart-relay enable

   By default, DHCPv6 smart relay is disabled.

Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 relay agent

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet.

Procedure

1. Enter system view.

   system-view

2. Set the DSCP value for DHCPv6 packets sent by the DHCPv6 relay agent.

   ipv6 dhcp dscp *dscp-value*

   The default DSCP value is 56.

Specifying a padding mode for the Interface-ID option

About this task

This feature enables the relay agent to fill the Interface-ID option in the specified mode. When receiving a DHCPv6 packet from a client, the relay agent fills the Interface-ID option in the mode and then forwards the packet to the DHCPv6 server.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify a padding mode for the Interface-ID option.

   ipv6 dhcp relay interface-id { **bas** | **interface** | **user-mac** }

   By default, the relay agent fills the Interface-ID option with the interface index of the interface.

Configuring the DHCPv6 relay agent to add Option 37 in Relay-forward messages

About this task

This feature enables the DHCPv6 relay agent to add Option 37 in the Relay-forward message sent to DHCPv6 server for a client. Option 37 provides client location information for DHCPv6 servers to decide the IPv6 address and other settings to be assigned to the requesting client.

For more information about Option 37, see "Option 37".

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the DHCPv6 relay agent to add Option 37 in Relay-forward messages.
   ipv6 dhcp relay remote-id enable
   By default, the DHCPv6 relay agent does not add Option 37 in Relay-forward messages.
4. (Optional.) Configure the DUID to be set in Option 37.
   ipv6 dhcp relay remote-id duid { **ascii** *ascii-string* | **hex** *hex-string* }
   By default, the DHCPv6 relay agent uses its own DUID in Option 37.

Enabling the DHCPv6 relay agent to support Option 79

About this task

If DHCPv6 relay agents exist in the network, the DHCPv6 server needs the MAC address of the DHCPv6 client for authentication or for IPv6 address or prefix assignment. To meet the requirement, enable the DHCPv6 relay agent that the client first passes to support Option 79. This feature allows the DHCPv6 relay agent to learn the MAC address in the client request. When the relay agent generates a Relay-Forward packet for the request, it fills the MAC address of the client in Option 79. The Relay-Forward packet is then forwarded to the DHCPv6 server.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the DHCPv6 relay agent to support Option 79.
   ipv6 dhcp relay client-link-address enable
   By default, the DHCPv6 relay agent does not support Option 79.

Enabling the DHCPv6 relay agent to advertise IPv6 prefixes

About this task

A DHCPv6 client can obtain an IPv6 prefix through DHCPv6 and use this IPv6 prefix to assign IPv6 address to clients in a downstream network. If the IPv6 prefix is in a different subnet than the IPv6 address of the DHCPv6 client's upstream interface, the clients in the downstream network cannot access the external network. You can enable the DHCPv6 relay agent that is on the same link as the DHCPv6 client to advertise the IPv6 prefix.

Procedure

1. Enter system view.

```
system-view
```

2. Enable the DHCPv6 relay agent to advertise IPv6 prefixes.

   ```
   ipv6 dhcp advertise pd-route
   ```

   By default, the DHCPv6 relay agent does not advertise IPv6 prefixes.

   Before using this command, make sure the DHCPv6 relay agent is enabled to record DHCPv6 relay entries.

Enabling advertising host routes for assigned IPv6 addresses

About this task

In a network where ND cannot resolve global unicast addresses, network devices cannot generate ND entries for all global unicast addresses. If a DHCPv6 client obtains a global unicast address, the neighboring devices do not have the ND entries for this global unicast address, thus cannot forward the packets destined for the client. To resolve this problem, enable the DHCPv6 relay agent to advertise host routes for assigned IPv6 addresses in DHCPv6 replies. The advertised route information is as follows:

- The destination IP address is the assigned IPv6 address.
- The next hop is the link-local address of the DHCPv6 client.
- The output interface is the interface that forwards the reply.

After the relay agent receives a packet destined for the assigned IPv6 address, the relay agent looks up the routing table for the next hop. ND resolution can succeed because the next hop is the link-local address of the client. The relay agent searches the ND table for the MAC address of the client based on the next hop and then forwards the packet.

Restrictions and guidelines

Before using this feature on the DHCPv6 relay agent, enable the DHCPv6 relay agent to record DHCPv6 relay entries first.

Procedure

1. Enter system view.

   ```
   system-view
   ```

2. Enable the DHCPv6 relay agent to advertise host routes for IPv6 addresses assigned to DHCPv6 clients.

   ```
   ipv6 dhcp advertise address-route
   ```

   By default, the DHCPv6 relay agent does not advertise host routes for IPv6 addresses assigned to DHCPv6 clients.

Specifying the source IPv6 address for relayed DHCPv6 requests

About this task

This task is required if a relay interface does not have routes to DHCPv6 servers. You can specify a global unicast address or the IPv6 address of another interface (typically the loopback interface) as the source IPv6 address for DHCPv6 requests. The relay interface inserts the source IPv6 address in the source IPv6 address field of DHCPv6 requests.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify the source IPv6 address for relayed DHCPv6 requests.

   ipv6 dhcp relay source-address { *ipv6-address* | **interface** *interface-type interface-number* }

   By default, the DHCPv6 relay agent uses the IPv6 global unicast address of the interface that connects to the DHCPv6 server as the source IPv6 address for relayed DHCPv6 requests.

   If the specified interface does not have an IPv6 global unicast address, the IPv6 address of the output interface is used as the source address for relayed DHCPv6 requests.

Configuring DHCPv6 relay security features

Enabling the DHCPv6 relay agent to record relay entries

About this task

This feature enables the DHCPv6 relay agent to automatically record DHCPv6 relay entries after DHCPv6 clients obtain IPv6 addresses or prefixes through DHCPv6. A DHCPv6 relay entry contains the binding between a client's hardware address and IPv6 address or prefix.

Some security features, such as IP source guard, use DHCPv6 relay entries to check incoming packets and block packets that do not match any entry. Hosts using manually configured IPv6 addresses are denied to access external networks through the relay agent. For more information about IP source guard, see *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the recording of DHCPv6 relay entries.

   ipv6 dhcp relay client-information record

   By default, the DHCPv6 relay agent does not record relay entries.

Enabling IPv6 release notification

About this task

This feature enables the DHCPv6 relay agent to send a Release message to the DHCPv6 server after it deletes a DHCPv6 relay entry. After the DHCPv6 server receives the message, it reclaims the IPv6 address or prefix and marks the lease as expired.

If you do not enable this feature, the DHCPv6 relay agent will not send a Release message after it deletes a relay entry.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IPv6 release notification.
   ipv6 dhcp relay release-agent
   By default, IPv6 release notification is disabled.

Enabling client offline detection

About this task

This feature enables the DHCPv6 relay agent to detect the status of ND entries. After an ND entry ages out, the DHCPv6 relay agent considers the client offline and deletes the relay entry for the client. For more information about ND, see "Configuring basic IPv6 settings."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable client offline detection.
   ipv6 dhcp client-detect
   By default, client offline detection is disabled.

Configuring DHCPv6 flood attack protection

About this task

The DHCPv6 flood attack protection enables the DHCPv6 relay agent to detect DHCPv6 flood attacks according to the DHCPv6 packet rate threshold on a per-MAC basis.

When the DHCPv6 relay agent receives a DHCPv6 packet from a client (MAC address), it creates a DHCPv6 flood attack entry in check state. If the number of DHCPv6 packets from the same MAC address reaches the upper limit in the detection duration, the relay agent determines that the client is launching a DHCPv6 flood attack. The DHCPv6 flood attack entry changes to the restrain state, and the DHCPv6 relay agent discards the DHCPv6 packets from that client. When the aging time of the entry is reached, the DHCPv6 relay agent deletes the entry. If a DHCPv6 packet from the MAC address arrives later, the DHCPv6 relay agent will create a flood attack entry and count the number of incoming DHCP packets for that client again.

This feature is available only in VXLAN networks. For more information about VXLANs, see *VXLAN Configuration Guide*.

Restrictions and guidelines

Configure this feature on the downlink interface of the DHCPv6 relay agent if the relay agent uses its MAC address as the source MAC address in the packets relayed to the DHCP server.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Set the DHCPv6 packet rate threshold for DHCPv6 flood attack detection.
   ipv6 dhcp flood-protection threshold *packet-number milliseconds*
   By default, the device allows a maximum of 6 DHCPv6 packets per 5000 milliseconds from each DHCPv6 client.

3. (Optional.) Set the DHCP flood attack entry aging time.

   ipv6 dhcp flood-protection aging-time *time*

   The default setting is 300 seconds.
4. Enter VSI view.

   vsi *vsi-name*
5. Enable DHCPv6 flood attack protection.

   ipv6 dhcp flood-protection enable

   By default, DHCPv6 flood attack protection is disabled.

Enabling checking the Reply messages in response to Confirm messages

About this task

Multiple DHCPv6 servers might be deployed on a relay network, for example, DHCPv6 server A and DHCPv6 server B. When DHCPv6 server A receives a Request message from a DHCPv6 client, it assigns an IPv6 address to that client. In this example, the assigned IPv6 address does not belong to any address pool on DHCPv6 server B. Upon receiving the IPv6 address, the client multicasts a Confirm message to all DHCPv6 servers on the network to verify that the IPv6 address is usable. On receipt of the Confirm message, DHCPv6 server A and DHCPv6 server B act differently as follows:

- DHCPv6 server A returns a Reply message to notify the client that the IPv6 address is usable.
- DHCPv6 server B returns a Reply message to notify the client that it fails to assign the IPv6 address.

The DHCPv6 relay agent then forwards the two Reply messages to the client. As a result, the client might fail to obtain the IPv6 address.

To avoid the issue above, enable the DHCPv6 relay agent to check the Reply messages in response to Confirm messages. When the DHCPv6 relay agent receives the first Reply message in response to a client's Confirm message, it performs the following task:

1. Starts a check timer that is fixed at 0.5 seconds.
2. Identifies whether the Reply messages received before the check timer expires confirm successful address allocation.
3. Forwards the Reply messages received after the check timer expires, using the original forwarding mechanism without checking them.

The operating mechanism of this feature is as follows:

- If the first received Reply message confirms successful address allocation to the client, the relay agent forwards that Reply message to the client. The relay agent processes the Reply messages subsequently received for the client within the check timer as follows:
  - Discards the Reply messages that confirm unsuccessful address allocation.
  - Forwards the Reply messages that confirm successful address allocation.
- If the first received Reply message confirms unsuccessful address allocation to the client, the relay agent caches that Reply message and waits for a Reply message of successful address allocation:
  - If the relay agent does not receive such a Reply message for the client within the check timer, it performs the following operations after the check timer expires:
    - Forwards the cached Reply message to the client.
    - Discards the other Reply messages received for the client before the check timer expires.

    - If the relay agent receives such a Reply message for the client within the check timer, it forwards that Reply message to the client and discards the cached Reply message. The relay agent processes the Reply messages received for the client within the check timer as follows:
        - Discards the Reply messages that confirm unsuccessful address allocation.
        - Forwards the Reply messages that confirm successful address allocation.
- If the relay agent receives new DHCP requests (including Confirm messages) from the client, it deletes the cached Reply message and disables the check timer.
- If the relay agent receives a new DHCP request from the client, it determines that the client has initiated a new request. In this situation, the relay agent will directly forward the Reply messages in response to the Confirm message of the client, without checking them.

Procedure

1. Enter system view.

   system-view
2. Enable the DHCPv6 relay agent to check the Reply messages in response to Confirm messages.

   ipv6 dhcp relay confirm-reply-check

   By default, the DHCPv6 relay agent does not check the Reply messages in response to Confirm messages.

Discarding DHCPv6 requests that are delivered from VXLAN tunnels

About this task

In a VXLAN network, the DHCPv6 relay agent feature can be configured on the VSI interface of a VTEP.

When the DHCPv6 relay agent receives a DHCPv6 request from an AC mapped to the VSI interface, the relay agent forwards this request to the DHCPv6 servers and broadcasts this request to other VTEPs. If those VTEPs also function as the DHCPv6 relay agents, each will forward the DHCPv6 request to the DHCPv6 servers they are connecting to. To prevent a DHCPv6 server from receiving the same DHCPv6 request from different VTEPs, you can configure this feature on the VSI interfaces of the VTEPs that are not directly connecting to DHCPv6 clients.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *vsi-interface vsi-number*
3. Configure the DHCP relay agent to discard the DHCP requests that are delivered from VXLAN tunnels.

   ipv6 dhcp relay request-from-tunnel discard

   By default, the DHCPv6 relay agent can forward the DHCPv6 requests that are delivered from VXLAN tunnels.

   If you configure this command on a device that acts as both the distributed and centralized VXLAN IP gateways, make sure the gateways do not use the same VSI interface to provide the gateway services.

Configuring support for forwarding DHCPv6 replies based on MAC address table

About this task

When a DHCPv6 client sends a request to its gateway enabled with DHCPv6 relay agent, the gateway records the request forwarding information for the client before relaying the request to the DHCPv6 server. The request forwarding information contains the MAC address of the DHCPv6 client and output interface of the request packet. If another gateway receives the reply for this request, the gateway discards the reply by default because it does not have the matching forwarding information for the reply. As a result, the client cannot obtain an IPv6 address or prefix.

To solve this problem, configure the support for forwarding replies based on MAC address table on the relay agent that is connected to the DHCPv6 server. This feature allows the relay agent to look up the MAC address table for the output interface for the replies that do not have the request forwarding information. This feature ensures that DHCPv6 clients receive DHCPv6 reply packets.

If the **broadcast** keyword is specified, the DHCPv6 relay agent broadcasts the DHCPv6 reply out of the interface found in the matching MAC address entry. When the DHCPv6 relay agent connecting to the requesting DHCPv6 client receives the reply, it sends the reply to its CPU before broadcasting the reply. If recording relay entries is enabled on this relay agent, a relay entry is generated for the DHCPv6 client.

If the **broadcast** keyword is not specified, the DHCPv6 relay agent connecting to the requesting client will directly forward the reply based on the MAC address instead of delivering it up to the CPU. In this case, the DHCPv6 relay agent cannot be aware that the packet is a DHCPv6 reply, thus no relay entry is generated for the client even if the recording of relay entries is enabled.

Restrictions and guidelines

This task takes effect only when the replies from the DHCPv6 server to the relay agent carry client MAC address information. To meet this requirement, enable the DHCPv6 relay agent to insert client MAC addresses into Option 18 for DHCPv6 requests by using the **ipv6 dhcp relay interface-id user mac** command.

Procedure

1. Enter system view.

   system-view

2. Enable MAC address table lookup for DHCPv6 replies that do not have request forwarding information.

   ipv6 dhcp relay mac-forward enable [**broadcast**]

   By default, the DHCPv6 relay agent discards a DHCPv6 reply if the relay agent does not have the request forwarding information for the reply.

Display and maintenance commands for DHCPv6 relay agent

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the DUID of the local device.	**display ipv6 dhcp duid**
Display DHCPv6 relay entries that record clients' IPv6 address information.	**display ipv6 dhcp relay client-information address** [**interface**

Task	Command
	interface-type interface-number | **ipv6** *ipv6-address*] [**vpn-instance** *vpn-instance-name*]
Display DHCPv6 relay entries that record clients' IPv6 prefix information.	**display ipv6 dhcp relay client-information pd** [**interface** *interface-type interface-number* | **prefix** *prefix/prefix-len*] [**vpn-instance** *vpn-instance-name*]
Display M-LAG status information recorded on the DHCPv6 relay agent.	**display ipv6 dhcp relay m-lag-status**
Display DHCPv6 server addresses specified on a DHCPv6 relay interface.	**display ipv6 dhcp relay server-address** [**interface** *interface-type interface-number*]
Display packet statistics on the DHCPv6 relay agent.	**display ipv6 dhcp relay statistics** [**interface** *interface-type interface-number*]
Clear DHCPv6 relay entries that record clients' IPv6 address information.	**reset ipv6 dhcp relay client-information address** [**interface** *interface-type interface-number* | **ipv6** *ipv6-address*] [**vpn-instance** *vpn-instance-name*]
Clear DHCPv6 relay entries that record clients' IPv6 prefix information.	**reset ipv6 dhcp relay client-information pd** [**interface** *interface-type interface-number* | **prefix** *prefix/prefix-len*] [**vpn-instance** *vpn-instance-name*]
Clear packets statistics on the DHCPv6 relay agent.	**reset ipv6 dhcp relay statistics** [**interface** *interface-type interface-number*]

DHCPv6 relay agent configuration examples

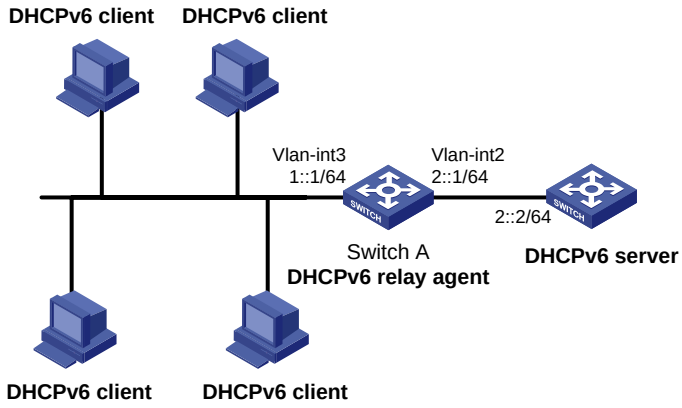
Example: Configuring DHCPv6 relay agent

Network configuration

As shown in Figure 17, configure the DHCPv6 relay agent on Switch A to relay DHCPv6 packets between DHCPv6 clients and the DHCPv6 server.

Switch A acts as the gateway of network 1::/64. It sends RA messages to notify the hosts to obtain IPv6 addresses and other configuration parameters through DHCPv6. For more information about RA messages, see "Configuring basic IPv6 settings."

Figure 17 Network diagram



Procedure

Specify IPv6 addresses for VLAN-interface 2 and VLAN-interface 3.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address 2::1 64
[SwitchA-Vlan-interface2] quit
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ipv6 address 1::1 64
```

Disable RA message suppression on VLAN-interface 3.

```
[SwitchA-Vlan-interface3] undo ipv6 nd ra halt
```

Set the M flag to 1 in RA advertisements to be sent on VLAN-interface 3. Hosts that receive the RA advertisements will obtain IPv6 addresses through DHCPv6.

```
[SwitchA-Vlan-interface3] ipv6 nd autoconfig managed-address-flag
```

Set the O flag to 1 in RA advertisements to be sent on VLAN-interface 3. Hosts that receive the RA advertisements will obtain information other than IPv6 address through DHCPv6.

```
[SwitchA-Vlan-interface3] ipv6 nd autoconfig other-flag
```

Enable the DHCPv6 relay agent on VLAN-interface 3 and specify the DHCPv6 server on the relay agent.

```
[SwitchA-Vlan-interface3] ipv6 dhcp select relay
[SwitchA-Vlan-interface3] ipv6 dhcp relay server-address 2::2
```

Verifying the configuration

Display DHCPv6 server address information on Switch A.

```
[SwitchA-Vlan-interface3] display ipv6 dhcp relay server-address
Interface: Vlan-interface3
 Server address      Outgoing Interface         Public/VRF name
 2::2                                           --/--
```

Display packet statistics on the DHCPv6 relay agent.

```
[SwitchA-Vlan-interface3] display ipv6 dhcp relay statistics
Packets dropped               :  0
Packets received              :  14
    Solicit                   :  0
    Request                   :  0
    Confirm                   :  0
```

```
    Renew                   :  0
    Rebind                  :  0
    Release                 :  0
    Decline                 :  0
    Information-request     :  7
    Relay-forward           :  0
    Relay-reply             :  7
Packets sent                :  14
    Advertise               :  0
    Reconfigure             :  0
    Reply                   :  7
    Relay-forward           :  7
    Relay-reply             :  0
```

Configuring the DHCPv6 client

About the DHCPv6 client

With DHCPv6 client configured, an interface can obtain configuration parameters from the DHCPv6 server.

A DHCPv6 client can use DHCPv6 to complete the following functions:

- Obtain an IPv6 address, an IPv6 prefix, or both, and obtain other configuration parameters. If DHCPv6 server is enabled on the device, the client can automatically save the obtained parameters to a DHCPv6 option group. With the obtained IPv6 prefix, the client can generate its global unicast address.
- Support stateless DHCPv6 to obtain configuration parameters except IPv6 address and IPv6 prefix. The client obtains an IPv6 address through stateless IPv6 address autoconfiguration. If the client receives an RA message with the M flag set to 0 and the O flag set to 1 during address acquisition, stateless DHCPv6 starts.

Restrictions and guidelines: DHCPv6 client configuration

Do not configure the DHCPv6 client on the same interface as the DHCPv6 server or the DHCPv6 relay agent.

DHCPv6 client tasks at a glance

To configure a DHCPv6 client, perform the following tasks:

1. (Optional.) Configuring the DHCPv6 client DUID
2. Configuring the DHCPv6 client to obtain IPv6 addresses, IPv6 prefixes and other network parameters

   Choose the following tasks as needed:

   - Configuring IPv6 address acquisition
   - Configuring IPv6 prefix acquisition
   - Configuring IPv6 address and prefix acquisition
   - Configuring acquisition of configuration parameters except IP addresses and prefixes
3. (Optional.) Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 client

Configuring the DHCPv6 client DUID

About this task

The DUID of a DHCPv6 client is the globally unique identifier of the client. The client pads its DUID into Option 1 of the DHCPv6 packet that it sends to the DHCPv6 server. The DHCPv6 server can assign specific IPv6 addresses or prefixes to DHCPv6 clients with specific DUIDs.

Restrictions and guidelines

Make sure the DUID that you configure is unique.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the DHCPv6 client DUID.
   ipv6 dhcp client duid { **ascii** *ascii-string* | **hex** *hex-string* | **mac** *interface-type interface-number* }
   By default, the interface uses the device bridge MAC address to generate its DHCPv6 client DUID.

Configuring IPv6 address acquisition

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the interface to use DHCPv6 to obtain an IPv6 address and other configuration parameters.
   ipv6 address dhcp-alloc [**option-group** *group-number* | **rapid-commit**] *
   - If the switch starts up with the initial configuration, the default setting of this command takes effect. An interface does not use DHCPv6 to obtain an IPv6 address and other network settings.
   - If the switch starts up with the factory defaults, the factory default setting of this command takes effect. Only VLAN-interface 1 uses DHCPv6 to obtain an IPv6 address and other network settings.

     For more information about initial configuration and factory defaults, see configuration file management configuration in *Fundamentals Configuration Guide*.

Configuring IPv6 prefix acquisition

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the interface to use DHCPv6 to obtain an IPv6 prefix and other configuration parameters.
   ipv6 dhcp client pd *prefix-number* [**option-group** *group-number* | **rapid-commit**] *
   By default, the interface does not use DHCPv6 for IPv6 prefix acquisition.

Configuring IPv6 address and prefix acquisition

1. Enter system view.
   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the interface to use DHCPv6 to obtain an IPv6 address, an IPv6 prefix, and other configuration parameters.

   ipv6 dhcp client stateful prefix *prefix-number* [**option-group** *option-group-number* | **rapid-commit**] *

   By default, the interface does not use DHCPv6 for IPv6 address and prefix acquisition.

Configuring acquisition of configuration parameters except IP addresses and prefixes

About this task

When a DHCPv6 client has obtained an IPv6 address and prefix, you can configure the following methods for the client to obtain other network configuration parameters:

- Execute the **ipv6 address auto** command to enable an interface to automatically generate an IPv6 global unicast address and a link-local address. Then stateless DHCPv6 will be triggered when the M flag is set to 0 and the O flag is set to 1 in a received RA message. For more information about the commands, see *Layer 3—IP services Command Reference*.
- Executing the **ipv6 dhcp client stateless enable** command on an interface to enable the interface to act as a DHCPv6 client to obtain configuration parameters from a DHCPv6 server.

If you execute both the **ip address auto** and **ipv6 dhcp client stateless enable** commands, the interface acts as follows:

- Generate a global unicast address and a link-local address.
- Obtain other configuration parameters from a DHCPv6 server.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the interface to support stateless DHCPv6. Choose the options to configure as needed:
   - Enable stateless IPv6 address autoconfiguration:

     ipv6 address auto
   - Configure the client to obtain network parameters from DHCPv6 servers:

     ipv6 dhcp client stateless enable

   By default, the interface does not support stateless DHCPv6.

Setting the DSCP value for DHCPv6 packets sent by the DHCPv6 client

About this task

The DSCP value of a packet specifies the priority level of the packet and affects the transmission priority of the packet.

Procedure

1. Enter system view.
   system-view
2. Set the DSCP value for DHCPv6 packets sent by the DHCPv6 client.
   ipv6 dhcp client dscp *dscp-value*
   By default, the DSCP value in DHCPv6 packets sent by the DHCPv6 client is 56.

Display and maintenance commands for DHCPv6 client

Execute the **display** commands in any view, and execute the **reset** command in user view.

Task	Command
Display the DHCPv6 client information.	**display ipv6 dhcp client** [**interface** *interface-type interface-number*]
Display the DHCPv6 client statistics.	**display ipv6 dhcp client statistics** [**interface** *interface-type interface-number*]
Clear the DHCPv6 client statistics.	**reset ipv6 dhcp client statistics** [**interface** *interface-type interface-number*]

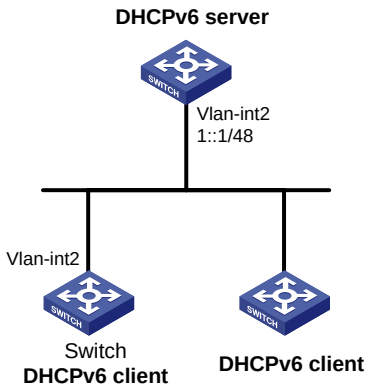
DHCPv6 client configuration examples

Example: Configuring IPv6 address acquisition

Network configuration

As shown in Figure 18, configure the switch to use DHCPv6 to obtain configuration parameters from the DHCPv6 server. The parameters include IPv6 address, DNS server address, domain name suffix, SIP server address, and SIP server domain name.

Figure 18 Network diagram



Procedure

You must configure the DHCPv6 server first before configuring the DHCPv6 client. For information about configuring DHCPv6 server, see "Configuring the DHCPv6 server."

Configure VLAN-interface 2 as a DHCPv6 client for IPv6 address acquisition. Configure the DHCPv6 client to support DHCPv6 rapid address assignment. Configure the DHCPv6 client to create a dynamic DHCPv6 option group for saving configuration parameters.

```
<Switch> system-view
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ipv6 address dhcp-alloc rapid-commit option-group 1
[Switch-Vlan-interface2] quit
```

Verifying the configuration

Verify that the client has obtained an IPv6 address and other configuration parameters from the server.

```
[Switch] display ipv6 dhcp client
Vlan-interface2:
  Type: Stateful client requesting address
    State: OPEN
    Client DUID: 0003000100e002000000
    Preferred server:
      Reachable via address: FE80::2E0:1FF:FE00:18
      Server DUID: 0003000100e001000000
    IA_NA: IAID 0x00000642, T1 50 sec, T2 80 sec
      Address: 1:1::2/128
       Preferred lifetime 100 sec, valid lifetime 200 sec
       Will expire on Mar 27 2014 at 08:06:57 (198 seconds left)
    DNS server addresses:
      2000::FF
    Domain name:
      example.com
    SIP server addresses:
      2:2::4
    SIP server domain names:
      bbb.example.com
```

After DHCPv6 server is enabled on the device, verify that configuration parameters are saved in a dynamic DHCPv6 option group.

```
[Switch] display ipv6 dhcp option-group 1
DHCPv6 option group: 1
  DNS server addresses:
    Type: Dynamic (DHCPv6 address allocation)
    Interface: Vlan-interface2
    2000::FF
  Domain name:
    Type: Dynamic (DHCPv6 address allocation)
    Interface: Vlan-interface2
    example.com
  SIP server addresses:
    Type: Dynamic (DHCPv6 address allocation)
    Interface: Vlan-interface2
```

```
    2:2::4
  SIP server domain names:
    Type: Dynamic (DHCPv6 address allocation)
    Interface: Vlan-interface2
    bbb.example.com
```

Verify that the DHCPv6 client has obtained an IPv6 address..

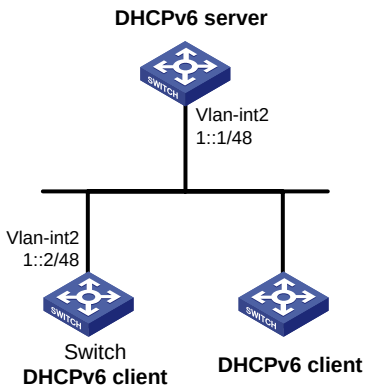
```
[Switch] display ipv6 interface brief
*down: administratively down
(s): spoofing
Interface                              Physical   Protocol   IPv6 Address
Vlan-interface2                        up         up         1:1::2
```

Example: Configuring IPv6 prefix acquisition

Network configuration

As shown in Figure 19, configure the switch to use DHCPv6 to obtain configuration parameters from the DHCPv6 server. The parameters include IPv6 prefix, DNS server address, domain name suffix, SIP server address, and SIP server domain name.

Figure 19 Network diagram



Procedure

You must configure the DHCPv6 server first before configuring the DHCPv6 client. For information about configuring DHCPv6 server, see "Configuring the DHCPv6 server."

Configure an IPv6 address for VLAN-interface 2 that is connected to the DHCPv6 server.

```
<Switch> system-view
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ipv6 address 1::2/48
```

Configure VLAN-interface 2 as a DHCPv6 client for IPv6 prefix acquisition. Configure the DHCPv6 client to support DHCPv6 rapid prefix assignment. Configure the DHCPv6 client to assign an ID to the obtained IPv6 prefix and create a dynamic DHCPv6 option group for saving configuration parameters.

```
[Switch-Vlan-interface2] ipv6 dhcp client pd 1 rapid-commit option-group 1
[Switch-Vlan-interface2] quit
```

Verifying the configuration

Verify that the DHCPv6 client has obtained an IPv6 prefix and other configuration parameters from the DHCPv6 server.

```
[Switch] display ipv6 dhcp client
Vlan-interface2:
  Type: Stateful client requesting prefix
    State: OPEN
    Client DUID: 0003000100e002000000
    Preferred server:
      Reachable via address: FE80::2E0:1FF:FE00:18
      Server DUID: 0003000100e001000000
    IA_PD: IAID 0x00000642, T1 50 sec, T2 80 sec
      Prefix: 12:34::/48
        Preferred lifetime 100 sec, valid lifetime 200 sec
        Will expire on Feb 4 2014 at 15:37:20(80 seconds left)
    DNS server addresses:
      2000::FF
    Domain name:
      example.com
    SIP server addresses:
      2:2::4
    SIP server domain names:
      bbb.example.com
```

Verify that the client has obtained an IPv6 prefix.

```
[Switch] display ipv6 prefix 1
Number: 1
Type  : Dynamic
Prefix: 12:34::/48
Preferred lifetime 100 sec, valid lifetime 200 sec
```

After DHCPv6 server is enabled on the device, verify that configuration parameters are saved in a dynamic DHCPv6 option group.

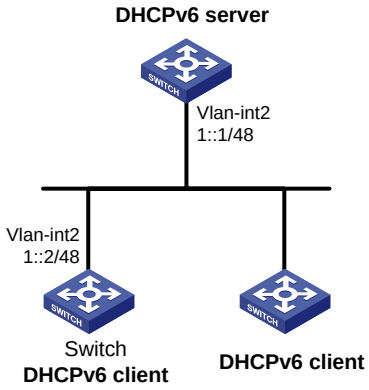
```
[Switch] display ipv6 dhcp option-group 1
DHCPv6 option group: 1
  DNS server addresses:
    Type: Dynamic (DHCPv6 prefix allocation)
    Interface: Vlan-interface2
    2000::FF
  Domain name:
    Type: Dynamic (DHCPv6 prefix allocation)
    Interface: Vlan-interface2
    example.com
  SIP server addresses:
    Type: Dynamic (DHCPv6 prefix allocation)
    Interface: Vlan-interface2
    2:2::4
  SIP server domain names:
    Type: Dynamic (DHCPv6 prefix allocation)
    Interface: Vlan-interface2
    bbb.example.com
```

Example: Configuring IPv6 address and prefix acquisition

Network configuration

As shown in Figure 20, configure the switch to use DHCPv6 to obtain configuration parameters from the DHCPv6 server. The parameters include IPv6 address, IPv6 prefix, DNS server address, domain name suffix, SIP server address, and SIP server domain name.

Figure 20 Network diagram



Procedure

You must configure the DHCPv6 server before configuring the DHCPv6 client. For information about configuring the DHCPv6 server, see "Configuring the DHCPv6 server."

Configure an IPv6 address for VLAN-interface 2 that is connected to the DHCPv6 server.

```
<Switch> system-view
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ipv6 address 1::2/48
```

Configure VLAN-interface 2 as a DHCPv6 client for IPv6 address and prefix acquisition. Specify IDs for the dynamic IPv6 prefix and dynamic DHCPv6 option group, and configure the client to support rapid address and prefix assignment.

```
[Switch-Vlan-interface2] ipv6 dhcp client stateful prefix 1 rapid-commit option-group 1
[Switch-Vlan-interface2] quit
```

Verifying the configuration

Verify that the DHCPv6 client has obtained an IPv6 address, an IPv6 prefix, and other configuration parameters from the DHCPv6 server.

```
[Switch] display ipv6 dhcp client
Vlan-interface2:
  Type: Stateful client requesting address and prefix
    State: OPEN
    Client DUID: 0003000100e002000000
    Preferred server:
      Reachable via address: FE80::2E0:1FF:FE00:18
      Server DUID: 0003000100e001000000
    IA_NA: IAID 0x00000642, T1 50 sec, T2 80 sec
      Address: 1:1::2/128
        Preferred lifetime 100 sec, valid lifetime 200 sec
        Will expire on Mar 27 2014 at 08:02:00 (199 seconds left)
    IA_PD: IAID 0x00000642, T1 50 sec, T2 80 sec
```

```
      Prefix: 12:34::/48
        Preferred lifetime 100 sec, valid lifetime 200 sec
        Will expire on Mar 27 2014 at 08:02:00 (199 seconds left)
    DNS server addresses:
      2000::FF
    Domain name:
      example.com
    SIP server addresses:
      2:2::4
    SIP server domain names:
      bbb.example.com
```

Verify that the DHCPv6 client has obtained an IPv6 address.

```
[Switch] display ipv6 interface brief
*down: administratively down
(s): spoofing
Interface                              Physical   Protocol   IPv6 Address
Vlan-interface2                        up         up         1:1::2
```

Verify that the client has obtained an IPv6 prefix.

```
[Switch] display ipv6 prefix 1
Number: 1
Type  : Dynamic
Prefix: 12:34::/48
Preferred lifetime 100 sec, valid lifetime 200 sec
```

After DHCPv6 server is enabled on the device, verify that configuration parameters are saved in a dynamic DHCPv6 option group.

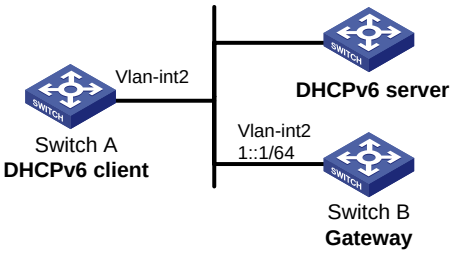
```
[Switch] display ipv6 dhcp option-group 1
DNS server addresses:
    Type: Dynamic (DHCPv6 address and prefix allocation)
    Interface: Vlan-interface2
    2000::FF
  Domain name:
    Type: Dynamic (DHCPv6 address and prefix allocation)
    Interface: Vlan-interface2
    example.com
  SIP server addresses:
    Type: Dynamic (DHCPv6 address and prefix allocation)
    Interface: Vlan-interface2
    2:2::4
  SIP server domain names:
    Type: Dynamic (DHCPv6 address and prefix allocation)
    Interface: Vlan-interface2
    bbb.example.com
```

Example: Configuring stateless DHCPv6

Network configuration

As shown in Figure 21, configure Switch A to use stateless DHCPv6 to obtain configuration parameters except IPv6 address and IPv6 prefix. Switch B acts as the gateway and advertises RA messages periodically.

Figure 21 Network diagram



Procedure

You must configure the DHCPv6 server first before configuring the DHCPv6 client. For information about configuring DHCPv6 server, see "Configuring the DHCPv6 server."

1. Configure the gateway Switch B.

   # Configure an IPv6 address for VLAN-interface 2.

   ```
   <SwitchB> system-view
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] ipv6 address 1::1 64
   ```

   # Set the O flag to 1 in RA advertisements to be sent on VLAN-interface 2. Hosts that receive the RA advertisements will obtain information other than IPv6 address through DHCPv6.

   ```
   [SwitchB-Vlan-interface2] ipv6 nd autoconfig other-flag
   ```

   # Disable RA message suppression on VLAN-interface 2.

   ```
   [SwitchB-Vlan-interface2] undo ipv6 nd ra halt
   ```

2. Configure the DHCPv6 client Switch A.

   # Enable stateless IPv6 address autoconfiguration on VLAN-interface 2.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ipv6 address auto
   ```

   With stateless IPv6 address autoconfiguration enabled, but no IPv6 address configured for VLAN-interface 2, Switch A automatically generates a link-local address. It sends an RS message to Switch B to request configuration information for IPv6 address generation. Upon receiving the RS message, Switch B sends back an RA message. After receiving an RA message with the M flag set to 0 and the O flag set to 1, Switch A performs stateless DHCPv6 to get other configuration parameters.

Verifying the configuration

Display the DHCPv6 client information.

```
[SwitchA-Vlan-interface2] display ipv6 dhcp client interface vlan-interface 2
Vlan-interface2:
 Type: Stateless client
    State: OPEN
    Client DUID: 00030001000fe2ff0000
    Preferred server:
```

```
      Reachable via address: FE80::213:7FFF:FEF6:C818
      Server DUID: 0003000100137ff6c818
    DNS server addresses:
      1:2:4::5
      1:2:4::7
    Domain name:
      abc.example.com
```

Display the DHCPv6 client statistics.

```
[SwitchA-Vlan-interface2] display ipv6 dhcp client statistics
Interface                    :  Vlan-interface2
Packets received             :  1
        Reply                :  1
        Advertise            :  0
        Reconfigure          :  0
        Invalid              :  0
Packets sent                 :  5
        Solicit              :  0
        Request              :  0
        Renew                :  0
        Rebind               :  0
        Information-request  :  5
        Release              :  0
        Decline              :  0
```

Configuring DHCPv6 snooping

About DHCPv6 snooping

It guarantees that DHCPv6 clients obtain IPv6 addresses or prefixes from authorized DHCPv6 servers. Also, it records IP-to-MAC bindings of DHCPv6 clients (called DHCPv6 snooping address entries) and prefix-to-port bindings of DHCPv6 clients (called DHCPv6 snooping prefix entries) for security purposes.

DHCPv6 snooping defines trusted and untrusted ports to make sure that clients obtain IPv6 addresses only from authorized DHCPv6 servers.

- **Trusted**—A trusted port can forward DHCPv6 messages correctly to make sure the clients get IPv6 addresses from authorized DHCPv6 servers.
- **Untrusted**—An untrusted port discards received messages sent by DHCPv6 servers to prevent unauthorized servers from assigning IPv6 addresses.

DHCPv6 snooping reads DHCP-ACK messages received from trusted ports and DHCP-REQUEST messages to create DHCPv6 snooping entries. A DHCPv6 snooping entry can be an address entry or a prefix entry.

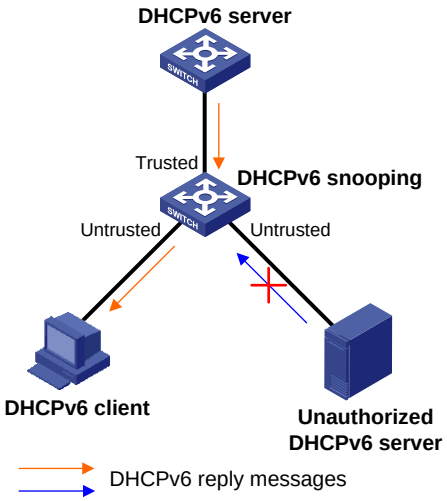
- A DHCPv6 address entry includes the MAC and IP addresses of a client, the port that connects to the DHCPv6 client, and the VLAN. You can use the **`display ipv6 dhcp snooping binding`** command to display the IP addresses of users for management.
- A DHCPv6 prefix entry includes the prefix and lease information assigned to the client, the port that connects to the DHCPv6 client, and the VLAN. You can use the **`display ipv6 dhcp snooping pd binding`** command to display the prefixes of the users for management.

Application of trusted and untrusted ports

Configure ports facing the DHCPv6 server as trusted ports, and configure other ports as untrusted ports.

As shown in Figure 22, configure the DHCPv6 snooping device's port that is connected to the DHCPv6 server as a trusted port. The trusted port forwards response messages from the DHCPv6 server to the client. The untrusted port connected to the unauthorized DHCPv6 server discards incoming DHCPv6 response messages.

Figure 22 Trusted and untrusted ports



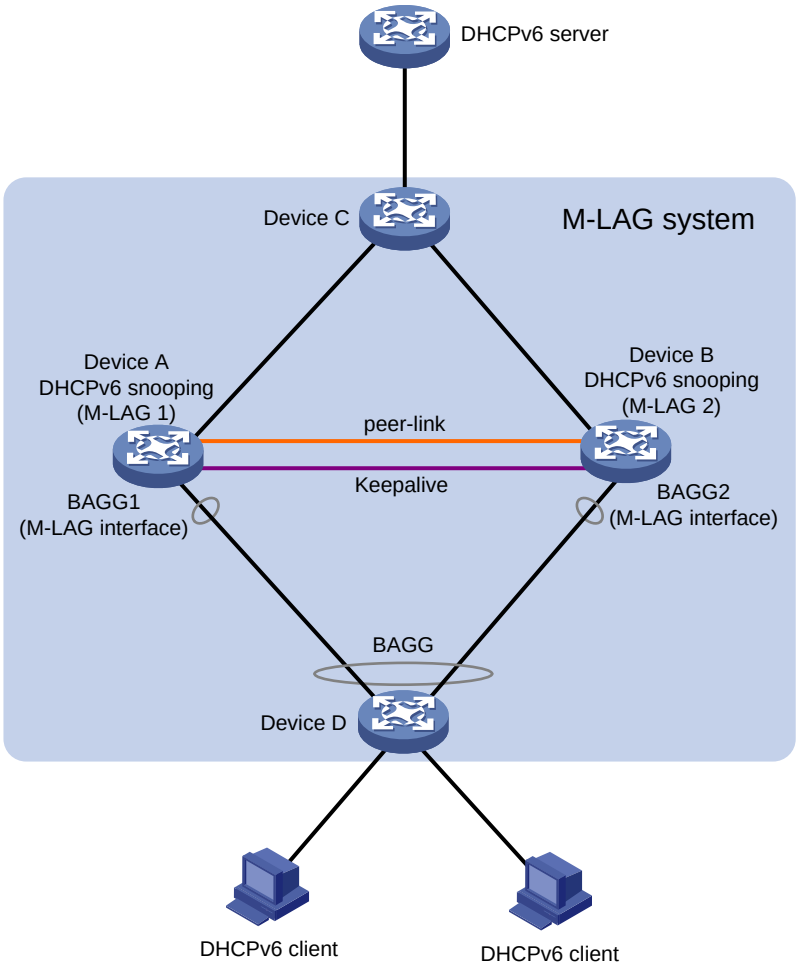
DHCPv6 snooping on an M-LAG network

M-LAG can virtualize two DHCPv6 snooping devices into one system through multichassis link aggregation, improving the availability of DHCPv6 snooping devices and facilitating device-level redundancy and load sharing.

To view the M-LAG status information and packet statistics related to DHCPv6 snooping entry synchronization, execute the **display ipv6 dhcp snooping m-lag-status** and **display ipv6 dhcp snooping m-lag-statistics** command, respectively. You can determine whether DHCPv6 snooping operates correctly on an M-LAG network based on the command output.

For more information about M-LAG, see M-LAG configuration in *Layer 2—LAN Switching Configuration Guide*.

Figure 23 DHCPv6 snooping on an M-LAG network



Restrictions and guidelines: DHCPv6 snooping configuration

DHCPv6 snooping works between the DHCPv6 client and server, or between the DHCPv6 client and DHCPv6 relay agent.

DHCPv6 snooping does not work between the DHCPv6 server and DHCPv6 relay agent.

To make sure DHCPv6 clients can obtain valid IPv6 addresses, specify the ports connected to authorized DHCPv6 servers as trusted ports. The trusted ports and the ports connected to DHCPv6 clients must be in the same VLAN.

If you configure DHCPv6 snooping settings on a Layer 2 Ethernet interface that is a member port of a Layer 2 aggregate interface, the settings do not take effect unless the interface is removed from the aggregation group.

After you enable DHCPv6 snooping and configure VLAN mapping on an interface, the interface will generate DHCPv6 snooping entries. The VLAN information in those DHCPv6 snooping entries are original, not processed by VLAN mapping. The device can communicate with a destination IP address in a DHCPv6 snooping entry only if the MAC address associated with the destination IP address matches the MAC address in the DHCPv6 snooping entry. If the MAC entry for the MAC address expires in the MAC address table, the communication will be terminated.

DHCPv6 snooping tasks at a glance

To configure DHCPv6 snooping, perform the following tasks:

1. Configuring basic DHCPv6 snooping features
2. (Optional.) Configuring DHCP snooping support for Option 18
3. (Optional.) Configuring DHCP snooping support for Option 37
4. (Optional.) Configuring DHCPv6 snooping entry auto backup
5. (Optional.) Setting the maximum number of DHCPv6 snooping entries
6. (Optional.) Configuring DHCPv6 packet rate limit
7. (Optional.) Configuring DHCPv6 snooping security features
8. (Optional.) Enabling LDRA on the DHCPv6 snooping device
9. (Optional.) Enabling Relay-Forward packet check
10. (Optional.) Enabling client offline detection on the DHCPv6 snooping device
11. (Optional.) Enabling DHCPv6 snooping logging and alarm
12. (Optional.) Disabling DHCPv6 snooping on an interface

Configuring basic DHCPv6 snooping features

Configuring basic DHCPv6 snooping features in a common network

About this task

Basic DHCPv6 snooping features include enabling DHCPv6 snooping, configuring trusted ports, and enabling recording DHCPv6 snooping entries.

When you enable DHCPv6 snooping globally on a device, DHCPv6 snooping is also enabled in all VLANs on the device. Enable snooping in specific VLANs if you do not need to enable DHCPv6 snooping globally in some networks. You can also other basic DHCP snooping features in these VLANs.

Restrictions and guidelines

If the basic DHCPv6 snooping features are configured globally, you can only use the undo form of the global configuration commands to disable the settings globally. The VLAN-specific configuration commands cannot disable the settings.

If the basic DHCPv6 snooping features are configured in a VLAN, you can only use the undo form of the VLAN-specific configuration commands to disable the settings in the VLAN. The global configuration command cannot disable the settings.

Configuring basic DHCPv6 snooping features globally

1. Enter system view.
   `system-view`
2. Enable DHCPv6 snooping globally.
   `ipv6 dhcp snooping enable`
   By default, DHCPv6 snooping is disabled globally.
3. Enter interface view.
   `interface` *`interface-type interface-number`*
   This interface must connect to the DHCPv6 server.

4. Specify the port as a trusted port.
   ipv6 dhcp snooping trust
   By default, all ports are untrusted ports after DHCPv6 snooping is enabled.
5. Enable recording DHCPv6 snooping entries.
   a. Return to system view.
      quit
   b. Enter interface view.
      interface *interface-type interface-number*
      This interface must connect to the DHCPv6 client.
   c. Enable recording DHCPv6 snooping entries. Choose the following tasks as needed:
      - Enable recording DHCPv6 snooping address entries.
        ipv6 dhcp snooping binding record
        By default, recording of DHCPv6 snooping address entries is disabled.
      - Enable recording DHCPv6 snooping prefix entries.
        ipv6 dhcp snooping pd binding record
        By default, recording of DHCPv6 snooping prefix entries is disabled.

Configuring basic DHCPv6 snooping features for VLANs

1. Enter system view.
   system-view
2. Enable DHCPv6 snooping for VLANs.
   ipv6 dhcp snooping enable vlan *vlan-id-list*
   By default, DHCPv6 snooping is disabled in all VLANs.
3. Enter VLAN view.
   vlan *vlan-id*
   Make sure DHCP snooping is enabled for the VLAN.
4. Specify a port as a trusted port.
   ipv6 dhcp snooping trust interface *interface-type interface-number*
   By default, all ports are untrusted ports after DHCPv6 snooping is enabled.
5. (Optional.) Enable recording DHCPv6 snooping entries in the VLAN. Choose the following tasks as needed:
   - Enable recording DHCPv6 snooping address entries.
     ipv6 dhcp snooping binding record
     By default, recording of DHCPv6 snooping address entries is disabled in a VLAN.
   - Enable recording DHCPv6 snooping prefix entries.
     ipv6 dhcp snooping pd binding record
     By default, recording of DHCPv6 snooping prefix entries is disabled in a VLAN.

Configuring basic DHCPv6 snooping features in a VXLAN network

About this task

In a VXLAN network, you can configure the following interfaces as DHCPv6 snooping trusted interfaces:

- ACs that are mapped to a VSI.
- VXLAN tunnel interfaces that are assigned to a VSI.

When the VTEP with DHCPv6 snooping configured receives a DHCPv6 request, the VTEP forwards this request through the trusted ACs or VXLAN tunnel interfaces.

Restrictions and guidelines

If the DHCPv6 server is in the local site, configure the AC that connects to the DHCPv6 server as trusted. If the DHCPv6 server is in a remote site, configure the VXLAN tunnel interface as trusted.

Configuring basic DHCPv6 snooping features in a VXLAN network (DHCPv6 server at the local site)

1. Enter system view.
   system-view
2. Enable DHCPv6 snooping globally.
   ipv6 dhcp snooping enable
   By default, DHCPv6 snooping is disabled globally.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enter Ethernet service instance view.
   service-instance *instance-id*
5. Configure the AC as the DHCPv6 snooping trusted interface.
   ipv6 dhcp snooping trust
   By default, all ports are untrusted after DHCPv6 snooping is enabled.
6. (Optional.) Enable recording of client information in DHCPv6 snooping entries on the ACs mapped to the VSI and VXLAN tunnel interfaces assigned to the VSI.
   a. Exit to interface view.
      quit
   b. Exit to the system view.
      quit
   c. Enter VSI view.
      vsi *vsi-name*
   d. Enable recording of client information in DHCP snooping entries on the ACs mapped to the VSI and VXLAN tunnel interfaces assigned to the VSI.
      ipv6 dhcp snooping binding record
      By default, the recording of DHCP snooping entries is disabled.

Configuring basic DHCPv6 snooping features in a VXLAN network (DHCPv6 server at a remote site)

1. Enter system view.
   system-view
2. Enable DHCPv6 snooping globally.
   ipv6 dhcp snooping enable
   By default, DHCP snooping is disabled globally.
3. Enter VSI view.
   vsi *vsi-name*
4. Configure the VXLAN tunnel interfaces as the DHCP snooping trusted interface.
   ipv6 dhcp snooping trust tunnel

By default, all ports are untrusted after DHCPv6 snooping is enabled.

This command sets all VXLAN tunnel interfaces in the VSI as DHCPv6 snooping trusted interfaces.

5. (Optional.) Enable recording of client information in DHCP snooping entries on the ACs mapped to the VSI and VXLAN tunnel interfaces assigned to the VSI.

   ipv6 dhcp snooping binding record

   By default, the recording of DHCP snooping entries is disabled.

Configuring DHCP snooping support for Option 18

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable DHCP snooping support for Option 18.

   ipv6 dhcp snooping option interface-id enable

   By default, DHCP snooping support for Option 18 is disabled.

4. (Optional.) Specify the content as the interface ID for Option 18.

   ipv6 dhcp snooping option interface-id { **bas** | [**vlan** *vlan-id*] **string** *interface-id* }

   By default, the DHCPv6 snooping device uses its DUID as the content for Option 18.

Configuring DHCP snooping support for Option 37

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable DHCP snooping support for Option 37.

   ipv6 dhcp snooping option remote-id enable

   By default, DHCP snooping support for Option 37 is disabled.

4. (Optional.) Specify the content as the remote ID.

   ipv6 dhcp snooping option remote-id [**vlan** *vlan-id*] **string** *remote-id*

   By default, the DHCPv6 snooping device uses its DUID as the content for Option 37.

Configuring DHCPv6 snooping entry auto backup

About this task

The auto backup feature saves DHCPv6 snooping entries to a backup file, and allows the DHCPv6 snooping device to download the entries from the backup file at reboot. The entries on the DHCPv6 snooping device cannot survive a reboot. The auto backup helps the security features provide services if these features (such as IP source guard) must use DHCPv6 snooping entries for user authentication.

Restrictions and guidelines

- If you disable DHCPv6 snooping with the **undo ipv6 dhcp snooping enable** command, the device deletes all DHCPv6 snooping entries, including those stored in the backup file.
- If you execute the **ipv6 dhcp snooping binding database filename** command, the DHCPv6 snooping device backs up DHCPv6 snooping entries immediately and runs auto backup. This command automatically creates the file if you specify a non-existent file.
- The waiting period starts when a DHCPv6 snooping entry is learned, updated, or removed. The DHCPv6 snooping device updates the backup file when the specified waiting period is reached. All changed entries during the period will be saved to the backup file. If no DHCPv6 snooping entry changes, the backup file is not updated.

Procedure

1. Enter system view.

   system-view
2. Configure the DHCPv6 snooping device to back up DHCPv6 snooping entries to a file.

   ipv6 dhcp snooping binding database filename { *filename* | **url** *url* [**username** *username* [**password** { **cipher** | **simple** } *string*]] }

   By default, the DHCPv6 snooping device does not back up the DHCPv6 snooping entries.
3. (Optional.) Manually save DHCPv6 snooping entries to the backup file.

   ipv6 dhcp snooping binding database update now
4. (Optional.) Set the waiting time after a DHCPv6 snooping entry change for the DHCPv6 snooping device to update the backup file.

   ipv6 dhcp snooping binding database update interval *interval*

   By default, the DHCP snooping device waits 300 seconds to update the backup file after a DHCP snooping entry change. If no DHCP snooping entry changes, the backup file is not updated.

Setting the maximum number of DHCPv6 snooping entries

About this task

Perform this task to prevent the system resources from being overused.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set the maximum number of DHCPv6 snooping entries for the interface to learn.

   ipv6 dhcp snooping max-learning-num *max-number*

   By default, the number of DHCPv6 snooping entries for an interface to learn is not limited.

Configuring DHCPv6 packet rate limit

About this task

This DHCPv6 packet rate limit feature discards exceeding DHCPv6 packets to prevent attacks that send large numbers of DHCPv6 packets.

Restrictions and guidelines

The rate set on the Layer 2 aggregate interface applies to all members of the aggregate interface. If a member interface leaves the aggregation group, it uses the rate set in its Ethernet interface view.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the maximum rate at which an interface can receive DHCPv6 packets.
   ipv6 dhcp snooping rate-limit *rate*
   By default, incoming DHCPv6 packets on an interface are not rate limited.

Configuring DHCPv6 snooping security features

Enabling DHCPv6-REQUEST check

About this task

Perform this task to use the DHCPv6-REQUEST check feature to protect the DHCPv6 server against DHCPv6 client spoofing attacks. Attackers can forge DHCPv6-RENEW messages to renew leases for legitimate DHCPv6 clients that no longer need the IP addresses. The forged messages disable the victim DHCPv6 server from releasing the IP addresses. Attackers can also forge DHCPv6-DECLINE or DHCPv6-RELEASE messages to terminate leases for legitimate DHCPv6 clients that still need the IP addresses.

The DHCPv6-REQUEST check feature enables the DHCPv6 snooping device to check every received DHCPv6-RENEW, DHCPv6-DECLINE, or DHCPv6-RELEASE message against DHCPv6 snooping entries.

- If any criterion in an entry is matched, the device compares the entry with the message information.
  - If they are consistent, the device considers the message valid and forwards it to the DHCPv6 server.
  - If they are different, the device considers the message forged and discards it.
- If no matching entry is found, the device forwards the message to the DHCPv6 server.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable DHCPv6-REQUEST check.
   ipv6 dhcp snooping check request-message

By default, DHCPv6-REQUEST check is disabled.

Configuring a DHCPv6 packet blocking port

About this task

Perform this task to configure a port as a DHCPv6 packet blocking port. The DHCPv6 packet blocking port drops all incoming DHCP requests.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the port to block DHCPv6 requests.

   ipv6 dhcp snooping deny

   By default, the port does not block DHCPv6 requests.

CAUTION:

To avoid IPv6 address and prefix acquisition failure, make sure no DHCPv6 clients are connected to the port.

Enabling LDRA on the DHCPv6 snooping device

About this task

Some DHCPv6 servers assign IPv6 addresses or prefixes to DHCPv6 clients only based on the Interface ID option in Relay-Forward packets generated by DHCPv6 relay agents. If no DHCPv6 relay agents exist in the network, these DHCPv6 servers cannot assign IPv6 addresses or prefixes based on the Interface ID option.

To solve this problem, you can enable the lightweight DHCPv6 relay agent (LDRA) on the receiving interface of DHCPv6 requests on the DHCPv6 snooping device. The feature allows the interface to generate a Relay-Forward packet for a received DHCPv6 request and to insert the Interface ID option in the packet. The DHCPv6 server then can assign an IPv6 address or prefix based on the Interface ID option.

Restrictions and guidelines

A network might have multiple cascaded lightweight DHCPv6 relay agents. As a best practice, do not specify the **trust** keyword if illegal Relay-Forward packets exist in the network.

For this feature to take effect, you must first execute the **ipv6 dhcp snooping enable** command to enable DHCPv6 snooping.

If you execute both the **ipv6 dhcp snooping relay-agent enable** and **ipv6 dhcp snooping option interface-id enable** commands, the **ipv6 dhcp snooping relay-agent enable** command does not take effect.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*

3. Enable LDRA on the interface.

   ipv6 dhcp snooping relay-agent enable [**trust**]

   By default, LDRA is disabled on an interface.

Enabling Relay-Forward packet check

About this task

A DHCPv6 snooping device functions between DHCPv6 clients and a DHCPv6 server, or between DHCPv6 clients and a DHCPv6 relay agent. When a DHCPv6 relay agent receives a DHCPv6 request, it generates a Relay-Forward packet, adds client information to Option 79, and then forwards the packet to the DHCPv6 server. If the DHCPv6 snooping device receives a Relay-Forward packet, it indicates that the DHCPv6 snooping device location is not correct. In this case, the DHCPv6 snooping device cannot work correctly.

This feature enables the DHCPv6 snooping device to drop Relay-Forward packets. When the number of dropped Relay-Forward packets reaches or exceeds the threshold, the device generates a log for administrators to adjust locations of the DHCPv6 devices.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the Relay-Forward packet check.

   ipv6 dhcp snooping check relay-forward

   By default, the Relay-Forward packet check is disabled.

Enabling client offline detection on the DHCPv6 snooping device

About this task

When a DHCPv6 client goes offline abnormally, it does not send a message to the DHCPv6 server to release the IPv6 address or prefix. As a result, the DHCPv6 server is not aware of the offline event and cannot release the address or prefix lease of the client in a timely manner.

With this feature enabled, the DHCPv6 snooping device performs the following operations when the ND entry of a client ages out:

1. Deletes the DHCPv6 snooping entry for the client.
2. Sends a release message to the DHCPv6 server to inform the server to release the address or prefix lease of the client.

Procedure

1. Enter system view.

   system-view

2. Enable client offline detection.

   ipv6 dhcp snooping client-detect

   By default, client offline detection is disabled.

Enabling DHCPv6 snooping logging and alarm

Enabling DHCPv6 snooping logging

About this task

The DHCPv6 snooping logging feature enables the DHCPv6 snooping device to generate DHCPv6 snooping logs and send them to the information center. The information helps administrators locate and solve problems. For information about the log destination and output rule configuration in the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

As a best practice, disable this feature if the log generation affects the device performance.

Procedure

1. Enter system view.
   system-view
2. Enable DHCPv6 snooping logging.
   ipv6 dhcp snooping log enable
   By default, DHCPv6 snooping logging is disabled.

Configuring packet drop alarm

About this task

After you enable the packet drop alarm for a feature, the device generates an alarm log when the number of packets dropped by this feature reaches or exceeds the threshold. The alarm log is sent to the information center. For information about the log destination and output rule configuration in the information center, see *Network Management and Monitoring Configuration Guide*.

To set the alarm threshold, use the **ipv6 dhcp snooping alarm threshold** command.

Restrictions and guidelines

For this feature to take effect, you must first execute the **ipv6 dhcp snooping log enable** command to enable DHCPv6 snooping logging.

Procedure

1. Enter system view.
   system-view
2. Enable the packet drop alarm.
   ipv6 dhcp snooping alarm { **relay-forward** | **request-message** } **enable**
   By default, the packet drop alarm is disabled.
3. Set a packet drop alarm threshold.
   ipv6 dhcp snooping alarm { **relay-forward** | **request-message** } **threshold** *threshold*
   By default, the packet drop alarm threshold is 100.

Disabling DHCPv6 snooping on an interface

About this task

This feature allows you to narrow down the interface range where DHCPv6 snooping takes effect. For example, to enable DHCP snooping globally except for a specific interface, you can enable DHCPv6 snooping globally and disable DHCPv6 snooping on the target interface.

Procedure

1. Enter system view.
   system-view
2. Enter interface view
   interface *interface-type interface-number*
3. Disable DHCPv6 snooping on the interface.
   ipv6 dhcp snooping disable
   By default:
   - If you enable DHCPv6 snooping globally or for a VLAN, DHCPv6 snooping is enabled on all interfaces on the device or on all interfaces in the VLAN.
   - If you do not enable DHCPv6 snooping globally or for a VLAN, DHCPv6 snooping is disabled on all interfaces on the device or on all interfaces in the VLAN.

Display and maintenance commands for DHCPv6 snooping

Execute **display** commands in any view, and **reset** commands in user view.

Task	Command
Display DHCPv6 snooping address entries.	**display ipv6 dhcp snooping binding** [**address** *ipv6-address* [**vlan** *vlan-id*] | **interface** *interface-type interface-number*]
Display information about the file that stores DHCPv6 snooping entries.	**display ipv6 dhcp snooping binding database**
Display statistics about the packets exchanged between M-LAG member devices for DHCPv6 snooping entry synchronization.	**display ipv6 dhcp snooping m-lag-statistics**
Display M-LAG status information.	**display ipv6 dhcp snooping m-lag-status**
Display DHCPv6 packet statistics for DHCPv6 snooping.	**display ipv6 dhcp snooping packet statistics** [**slot** *slot-number*]
Display DHCPv6 snooping prefix entries.	**display ipv6 dhcp snooping pd binding** [**interface** *interface-type interface-number* | **prefix** *prefix/prefix-length* [**vlan** *vlan-id*]]
Display information about trusted ports.	**display ipv6 dhcp snooping trust**

Task	Command
Clear DHCPv6 snooping address entries.	**reset ipv6 dhcp snooping binding** { **all** | **address** *ipv6-address* [**vlan** *vlan-id*] }
Clear statistics about the packets exchanged between M-LAG member devices for DHCPv6 snooping entry synchronization.	**reset ipv6 dhcp snooping m-lag-statistics**
Clear DHCPv6 packet statistics for DHCPv6 snooping.	**reset ipv6 dhcp snooping packet statistics** [**slot** *slot-number*]
Clear DHCPv6 snooping prefix entries.	**reset ipv6 dhcp snooping pd binding** { **all** | **prefix** *prefix/prefix-length* [**vlan** *vlan-id*] }

DHCPv6 snooping configuration examples

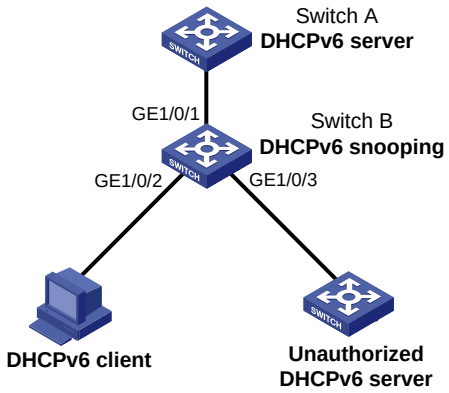
Example: Configuring DHCPv6 snooping globally

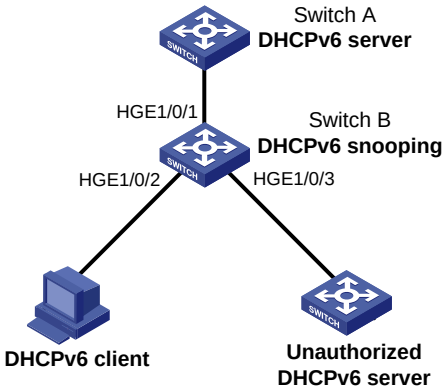
Network configuration

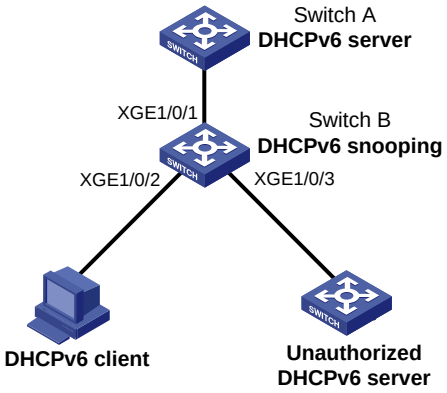
As shown in Figure 24, Switch B is connected to the authorized DHCPv6 server through GigabitEthernet 1/0/1, to the unauthorized DHCPv6 server through GigabitEthernet 1/0/3, and to the DHCPv6 client through GigabitEthernet 1/0/2.

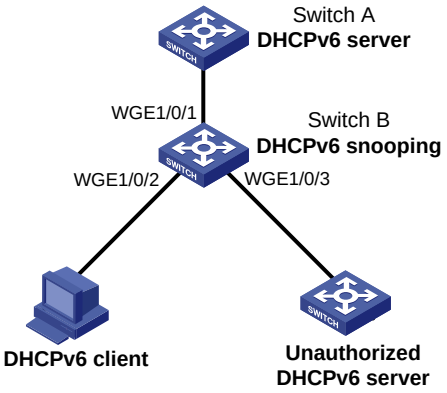
Configure only the port connected to the authorized DHCPv6 server to forward the responses from the DHCPv6 server. Enable the DHCPv6 snooping device to record DHCPv6 snooping address entries.

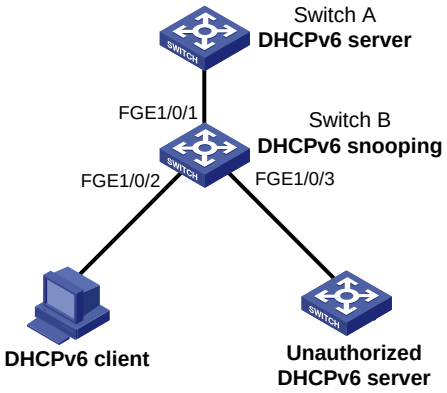
Figure 24 Network diagram

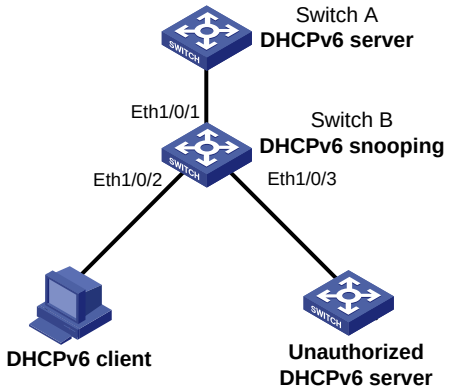


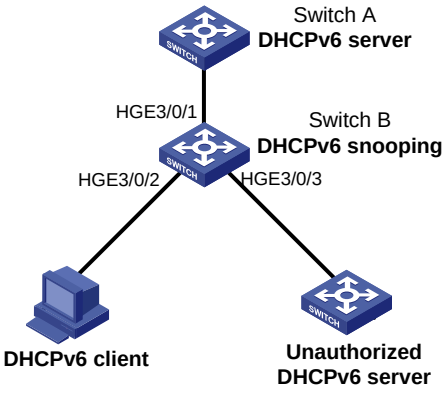

Switch A
DHCPv6 server
HGE1/0/1
Switch B
DHCPv6 snooping
HGE1/0/2
HGE1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server

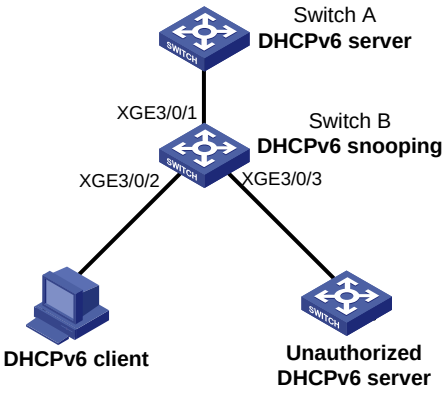

Switch A
DHCPv6 server
XGE1/0/1
Switch B
DHCPv6 snooping
XGE1/0/2
XGE1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server

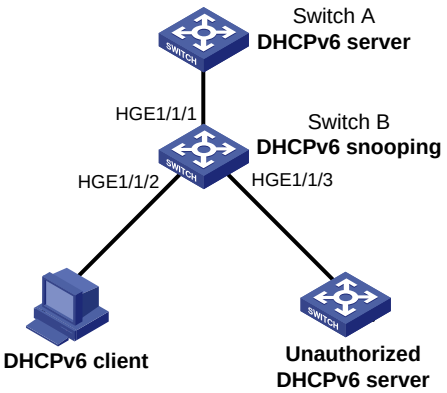

Switch A
DHCPv6 server
WGE1/0/1
Switch B
DHCPv6 snooping
WGE1/0/2
WGE1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server

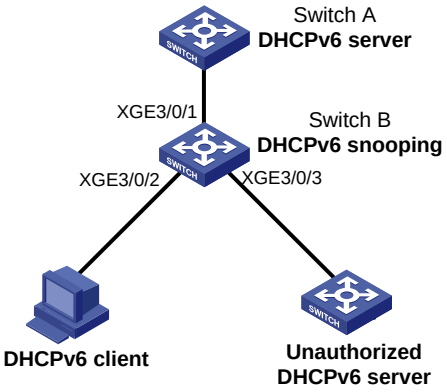

Switch A
DHCPv6 server
FGE1/0/1
Switch B
DHCPv6 snooping
FGE1/0/2
FGE1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server


Switch A
DHCPv6 server
Eth1/0/1
Switch B
DHCPv6 snooping
Eth1/0/2
Eth1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server


Switch A
DHCPv6 server
HGE3/0/1
Switch B
DHCPv6 snooping
HGE3/0/2
HGE3/0/3
DHCPv6 client
Unauthorized
DHCPv6 server


Switch A
DHCPv6 server
XGE3/0/1
Switch B
DHCPv6 snooping
XGE3/0/2
XGE3/0/3
DHCPv6 client
Unauthorized
DHCPv6 server


Switch A
DHCPv6 server
HGE1/1/1
Switch B
DHCPv6 snooping
HGE1/1/2
HGE1/1/3
DHCPv6 client
Unauthorized
DHCPv6 server



Procedure

Enable DHCPv6 snooping.

```
<SwitchB> system-view
[SwitchB] ipv6 dhcp snooping enable
```

Specify GigabitEthernet 1/0/1 as a trusted port.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] ipv6 dhcp snooping trust
[SwitchB-GigabitEthernet1/0/1] quit
```

Enable recording DHCPv6 snooping address entries on GigabitEthernet 1/0/2.

```
[SwitchB]interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] ipv6 dhcp snooping binding record
[SwitchB-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Verify that the DHCPv6 client obtains an IPv6 address and all other configuration parameters only from the authorized DHCPv6 server. (Details not shown.)

Display DHCPv6 snooping address entries on the DHCPv6 snooping device.

```
[SwitchB] display ipv6 dhcp snooping binding
```

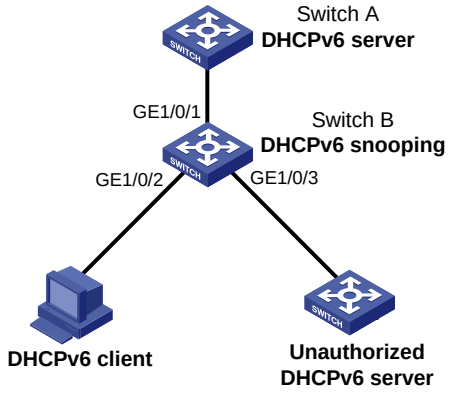
Example: Configuring DHCPv6 snooping for a VLAN

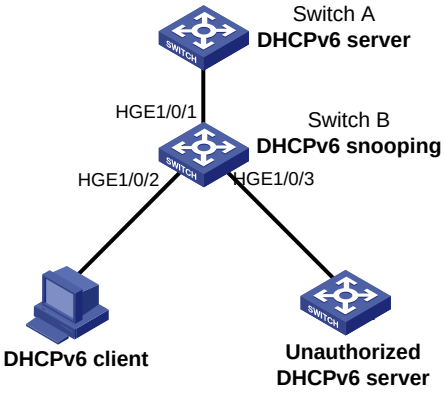
Network configuration

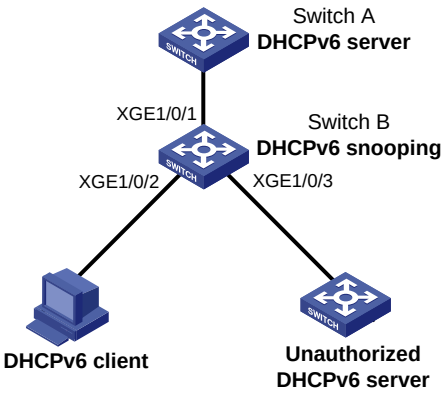
As shown in Figure 25, Switch B is connected to the authorized DHCPv6 server through GigabitEthernet 1/0/1, to the unauthorized DHCPv6 server through GigabitEthernet 1/0/3, and to the DHCPv6 client through GigabitEthernet 1/0/2.

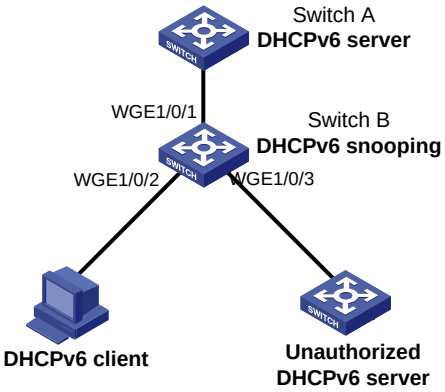
In VLAN 100, configure only the port connected to the authorized DHCPv6 server to forward the responses from the DHCPv6 server. Enable the DHCPv6 snooping device to record DHCPv6 snooping address entries.

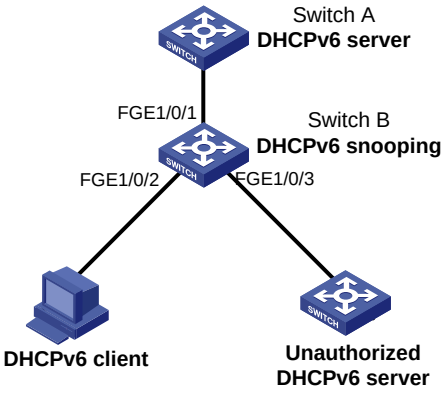
Figure 25 Network diagram

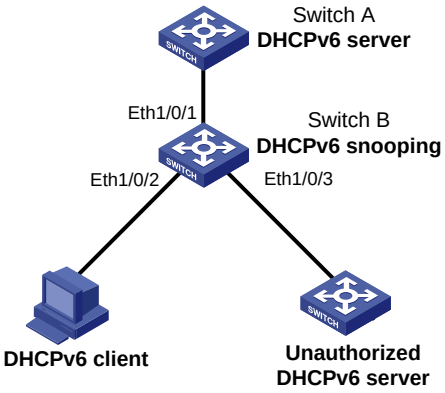


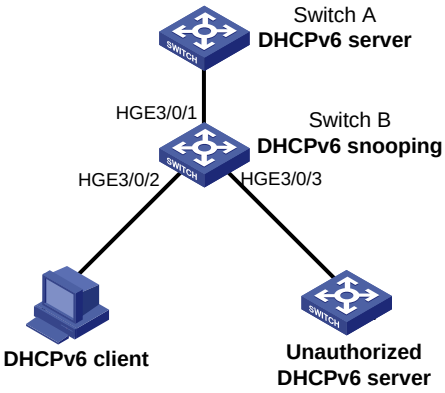


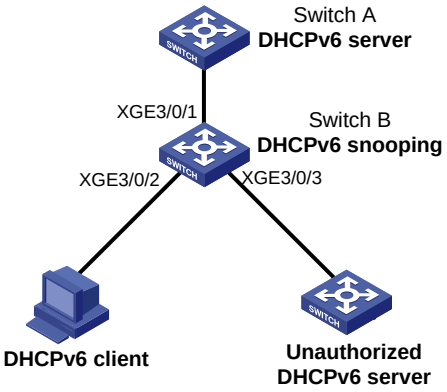


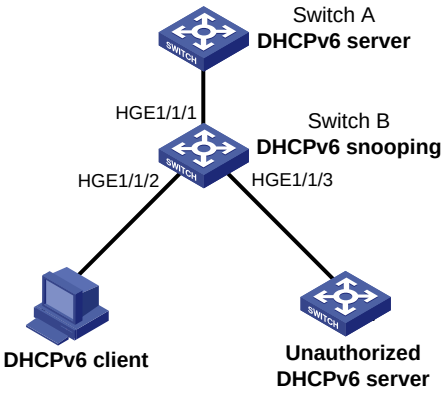

Switch A
DHCPv6 server
WGE1/0/1
Switch B
DHCPv6 snooping
WGE1/0/2
WGE1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server

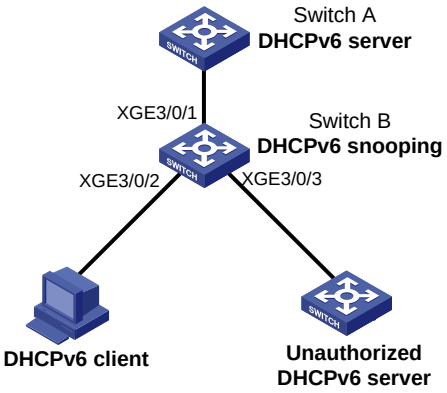

Switch A
DHCPv6 server
FGE1/0/1
Switch B
DHCPv6 snooping
FGE1/0/2
FGE1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server


Switch A
DHCPv6 server
Eth1/0/1
Switch B
DHCPv6 snooping
Eth1/0/2
Eth1/0/3
DHCPv6 client
Unauthorized
DHCPv6 server


Switch A
DHCPv6 server
HGE3/0/1
Switch B
DHCPv6 snooping
HGE3/0/2
HGE3/0/3
DHCPv6 client
Unauthorized
DHCPv6 server







Procedure

Assign access ports GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
[SwitchB-vlan100] quit
```

Enable DHCPv6 snooping for VLAN 100.

```
[SwitchB] ipv6 dhcp snooping enable vlan 100
```

Configure GigabitEthernet 1/0/1 as a trusted port in VLAN 100.

```
[SwitchB] vlan 100
[SwitchB-vlan100] ipv6 dhcp snooping trust interface gigabitethernet 1/0/1
```

Enable recording DHCPv6 snooping entries in VLAN 100.

```
[SwitchB-vlan100] ipv6 dhcp snooping binding record
[SwitchB-vlan100] quit
```

Verifying the configuration

Verify that the DHCPv6 client obtains an IPv6 address and all other configuration parameters only from the authorized DHCPv6 server. (Details not shown.)

Display DHCPv6 snooping address entries on the DHCPv6 snooping device.

```
[SwitchB] display ipv6 dhcp snooping binding
```

Configuring DHCPv6 guard

About DHCPv6 guard

The DHCPv6 guard feature filters DHCPv6 Advertise and Reply messages by using DHCPv6 guard policies to make sure DHCPv6 clients obtain addresses/prefixes from authorized DHCPv6 servers. To provide finer level of filtering granularity, you can specify the following parameters for a DHCPv6 guard policy:

- Device role of the device that attached to the target interface or VLAN. The interface or VLAN to which the DHCPv6 guard policy is applied is called the target interface or VLAN.
- DHCPv6 server match criterion.
- Match criterion for IPv6 addresses/prefixes assigned by DHCPv6 servers.
- Allowed DHCPv6 server preference range.

To meet requirements of DHCPv6 clients in different locations, apply DHCPv6 guard policies to different interfaces or VLANs on the same device.

DHCPv6 guard operating mechanism

Upon receiving a DHCPv6 Solicit or Request message, the DHCPv6 guard device forwards the message without performing the DHCPv6 guard policy check.

When receiving a DHCPv6 reply, the DHCPv6 guard device performs the DHCPv6 guard policy check in the following order:

1. Examines whether the receiving port is a trusted port. The device forwards the message if the message is from the a trusted port.

   Configure trusted ports in a DHCPv6 guard policy only in one of the following conditions:

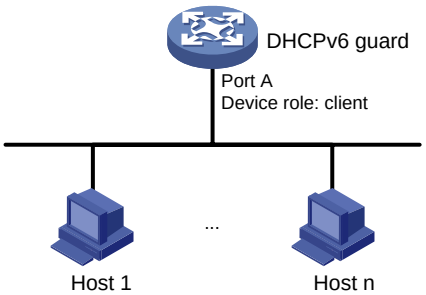
   - The port to which the DHCPv6 guard policy applies is connected to an authorized server.
   - All ports in the VLAN to which the DHCPv6 guard policy applies are connected to authorized servers.

2. Examines the message based on the device role:

   - If the message is received from the device with the DHCPv6 client device role, the device drops the message.

     If the interface to which the DHCPv6 guard policy applies is not connected to any authorized DHCPv6 servers, set the device role to **client** for the policy, as shown in Figure 26.

     Figure 26 Setting the device role to client

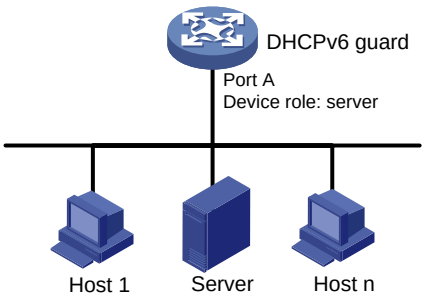
     

   - If the message is received from the device with the DHCPv6 server device role, the device examines the message as follows:

- For an Advertise message, the message passes the policy check if the source IP address in the message is permitted by the ACL and the server preference is in the match range.
- For a Reply message, the message passes the policy check if the assigned IPv6 addresses/prefixes in the message are permitted by the ACL.

If the interface to which the DHCPv6 guard policy applies is connected to an authorized DHCPv6 server, set the device role to **server** for the policy, as shown in Figure 27.

Figure 27 Setting the device role to server



The device forwards the reply after the message passes the DHCPv6 guard policy check.

Restrictions and guidelines: DHCPv6 guard configuration

The DHCPv6 guard feature operates correctly only when the device is located between the DHCPv6 client and the DHCPv6 server or between the DHCPv6 client and the DHCPv6 relay agent. If the device is located between the DHCPv6 server and the DHCPv6 relay agent, the DHCPv6 guard feature cannot operate correctly.

When the DHCPv6 guard feature is configured on a DHCPv6 snooping device, both features can take effect. The device forwards DHCPv6 reply packets received on a DHCPv6 snooping trusted port only if they pass the DHCPv6 guard check. These packets are dropped if they fail the DHCPv6 guard check.

DHCPv6 guard tasks at a glance

To configure DHCPv6 guard, perform the following tasks:

1. Configuring a DHCPv6 guard policy
2. Applying the DHCPv6 guard policy

   Choose the following tasks as needed:

   - Applying a DHCPv6 guard policy to an interface
   - Applying a DHCPv6 guard policy to a VLAN

   If DHCPv6 guard polices are applied to both an interface and the VLAN of the interface, the interface-specific policy is used on the interface.

Configuring a DHCPv6 guard policy

1. Enter system view.

   ```
   system-view
   ```

2. Create a DHCPv6 guard policy and enter its view.
   ipv6 dhcp guard policy *policy-name*
3. Specify the role of the device attached to the target interface or VLAN.
   device-role { **client** | **server** }
   By default, the device role is DHCPv6 client for the device attached to the target interface or VLAN.
4. Configure a DHCPv6 guard policy.
   - Configure a DHCPv6 server match criterion.
     if-match server acl { *acl-number* | **name** *acl-name* }
     By default, no DHCPv6 server match criterion is configured, and all DHCPv6 servers are authorized.
   - Configure a match criterion for the assigned IPv6 addresses/prefixes.
     if-match reply acl { *acl-number* | **name** *acl-name* }
     By default, no match criterion is configured for the assigned IPv6 addresses/prefixes, and all assigned IPv6 addresses/prefixes can pass the address/prefix check.
   - Configure an allowed DHCPv6 server preference range.
     preference { **max** *max-value* | **min** *min-value* } *
     By default, no DHCPv6 server preference range is configured, and DHCPv6 servers with preferences 1 to 255 can pass the preference check.
   - Configure the port to which the policy applies as a trusted port for the policy.
     trust port
     By default, no trusted port is configured for a DHCPv6 guard policy.

Applying a DHCPv6 guard policy to an interface

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   interface *interface-type interface-number*
3. Apply a DHCPv6 guard policy to the interface.
   ipv6 dhcp guard apply policy *policy-name*
   By default, no DHCPv6 guard policy is applied to the interface.

Applying a DHCPv6 guard policy to a VLAN

1. Enter system view.
   system-view
2. Create a VLAN and enter its view.
   vlan *vlan-number*
3. Apply a DHCPv6 guard policy to the VLAN.
   ipv6 dhcp guard apply policy *policy-name*
   By default, no DHCPv6 guard policy is applied to the VLAN.

Display and maintenance commands for DHCPv6 guard

Execute **display** commands in any view.

Task	Command
Display information about DHCPv6 guard policies.	**display ipv6 dhcp guard policy** [*policy-name*]

DHCPv6 guard configuration examples

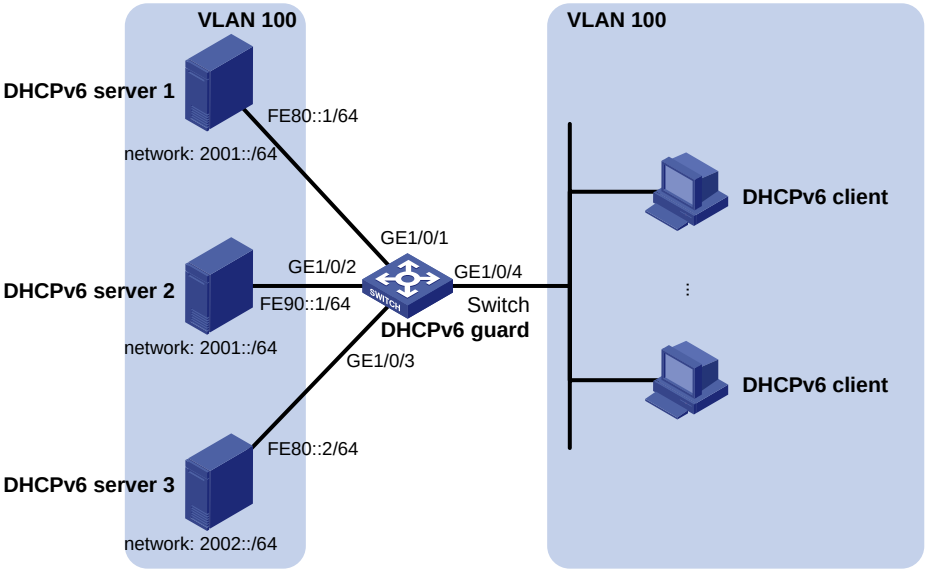
Example: Configuring DHCPv6 guard

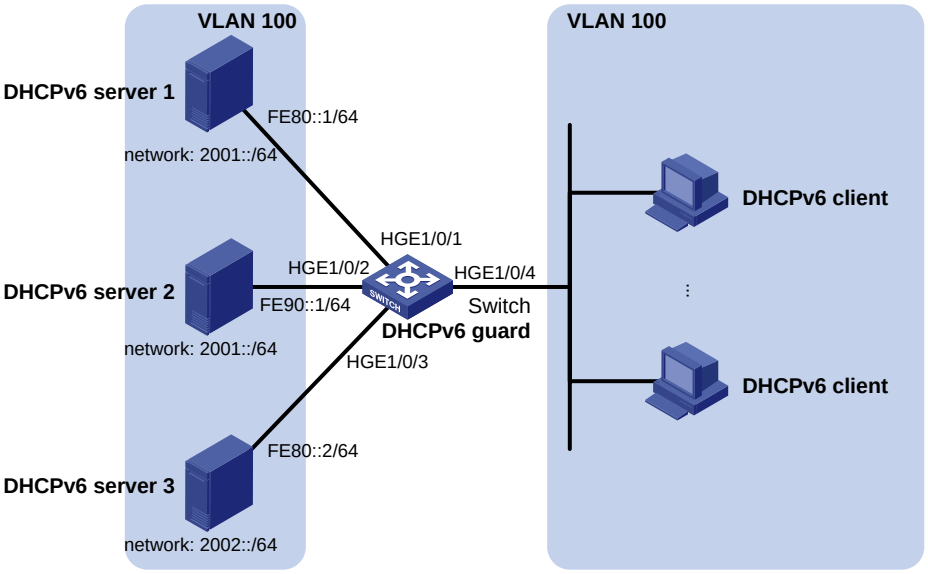
Network configuration

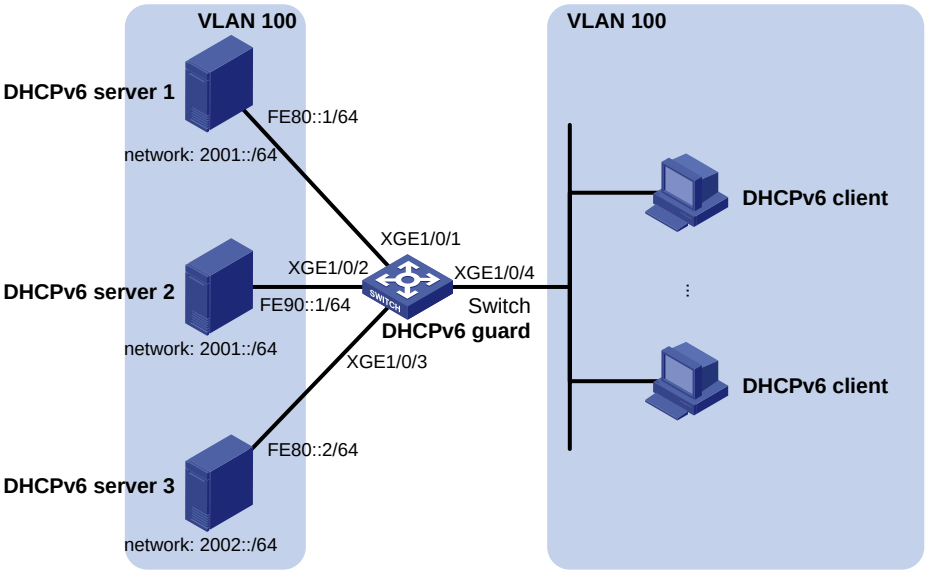
As shown in Figure 28, all DHCPv6 servers and clients are in VLAN 100. The assignable IPv6 address ranges on the DHCPv6 server 1, server 2, and server 3 are 2001::/64, 2001::/64, and 2002::/64, respectively.

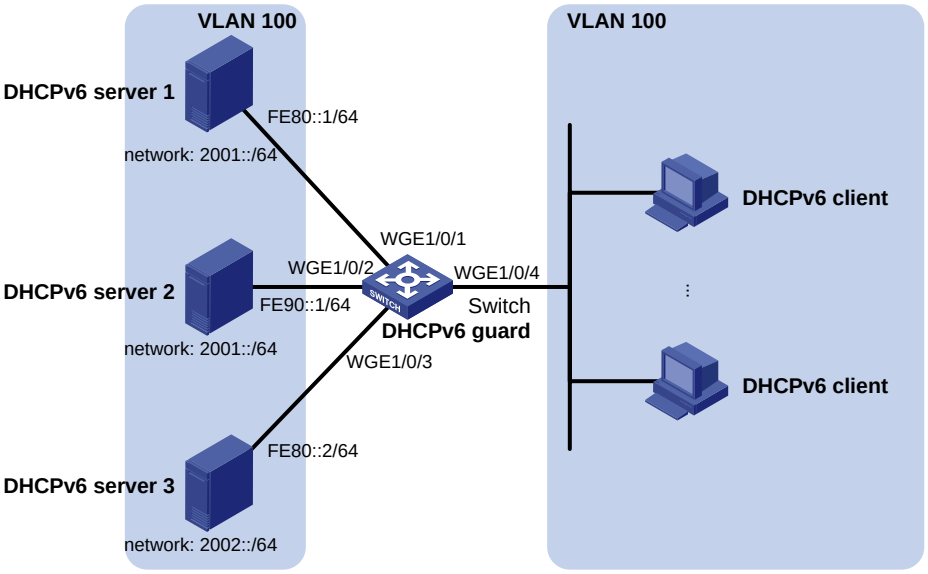
Configure DHCPv6 guard on the switch, so that the switch forwards only DHCPv6 replies with the source IPv6 address in the range of FE80::/12 and assigned prefixes in the range of 2001::/16.

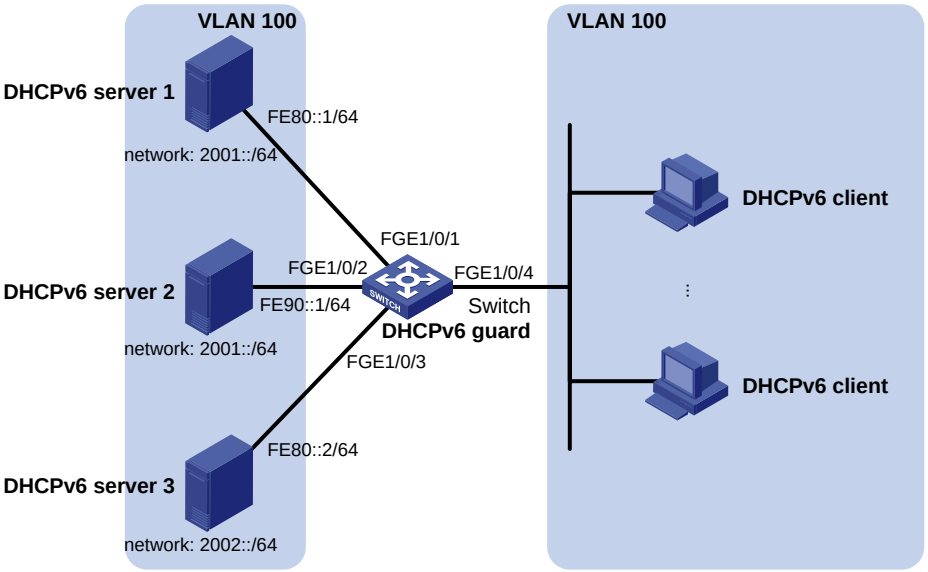
Figure 28 Network diagram

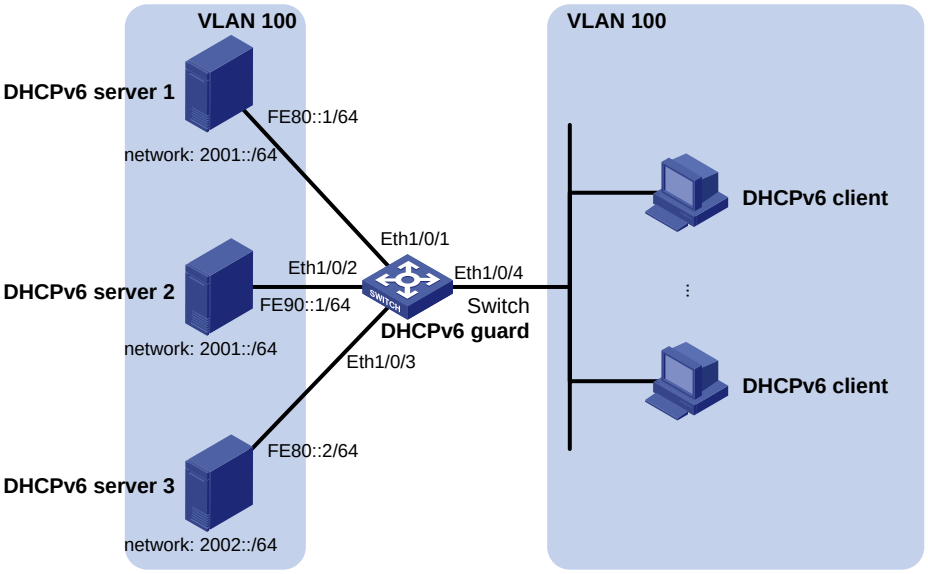


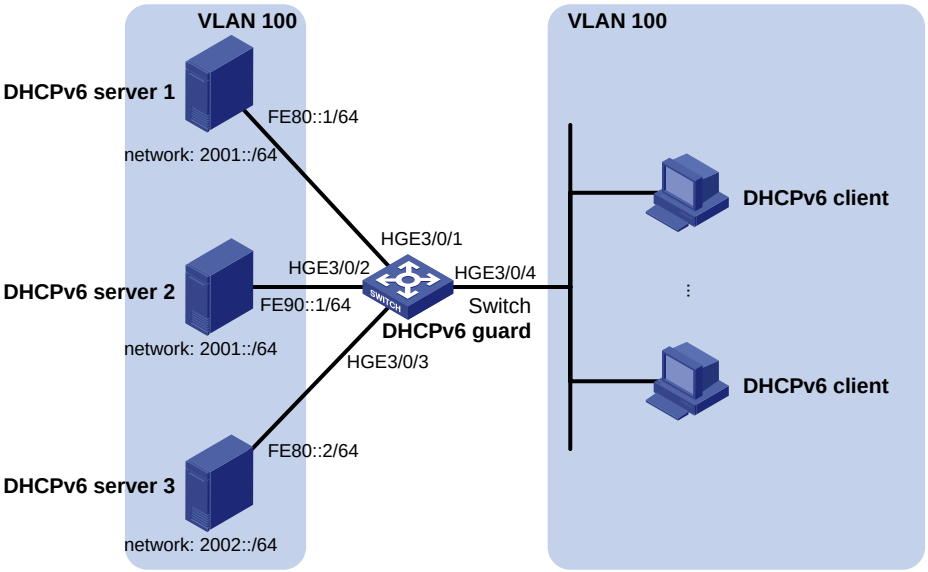

VLAN 100
VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
HGE1/0/1
HGE1/0/2
HGE1/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
HGE1/0/3
DHCPv6 client
DHCPv6 client
FE80::2/64
DHCPv6 server 3
network: 2002::/64

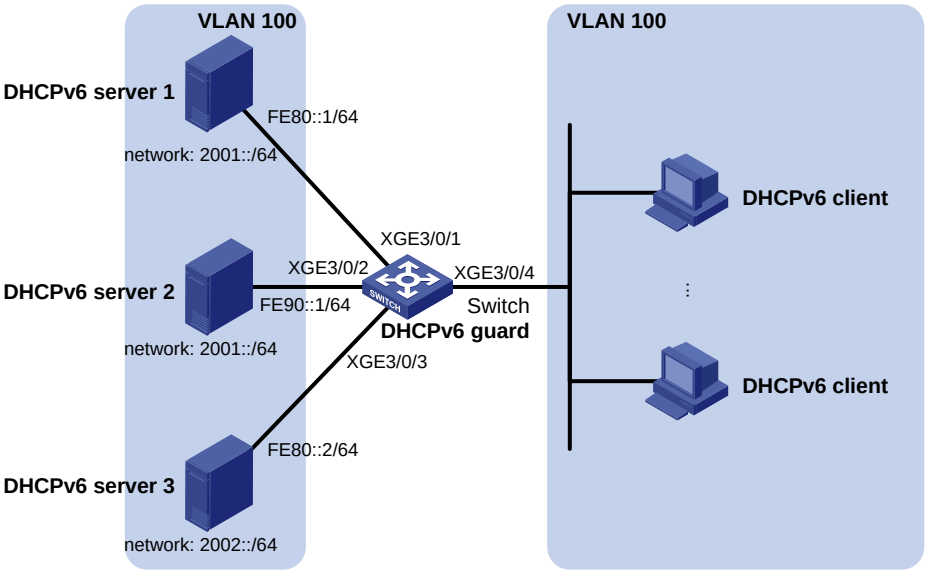

VLAN 100
VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
XGE1/0/1
XGE1/0/2
XGE1/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
XGE1/0/3
DHCPv6 client
DHCPv6 client
FE80::2/64
DHCPv6 server 3
network: 2002::/64

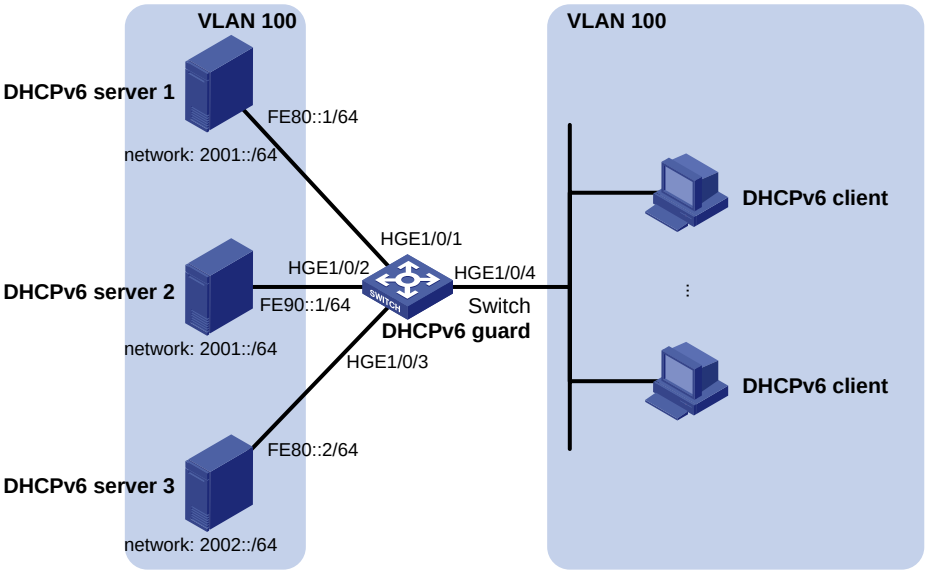

VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
WGE1/0/1
WGE1/0/2
WGE1/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
WGE1/0/3
FE80::2/64
DHCPv6 server 3
network: 2002::/64
VLAN 100
DHCPv6 client
DHCPv6 client

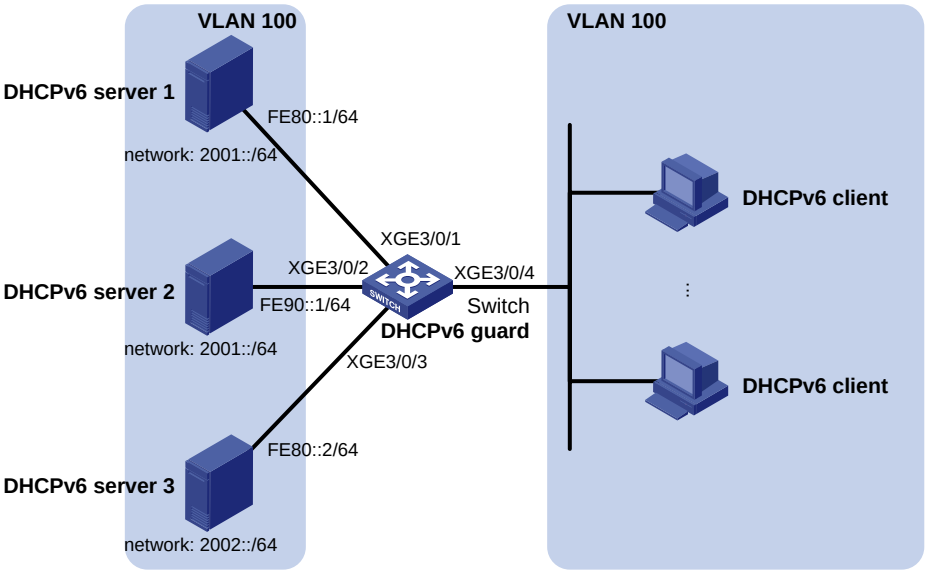

VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
FGE1/0/1
FGE1/0/2
FGE1/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
FGE1/0/3
FE80::2/64
DHCPv6 server 3
network: 2002::/64
VLAN 100
DHCPv6 client
DHCPv6 client


VLAN 100
VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
Eth1/0/1
Eth1/0/2
Eth1/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
Eth1/0/3
DHCPv6 client
DHCPv6 client
FE80::2/64
DHCPv6 server 3
network: 2002::/64


VLAN 100
VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
HGE3/0/1
HGE3/0/2
HGE3/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
HGE3/0/3
DHCPv6 client
DHCPv6 client
FE80::2/64
DHCPv6 server 3
network: 2002::/64


VLAN 100
VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
XGE3/0/1
XGE3/0/2
XGE3/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
XGE3/0/3
DHCPv6 client
DHCPv6 client
FE80::2/64
DHCPv6 server 3
network: 2002::/64


VLAN 100
VLAN 100
DHCPv6 server 1
FE80::1/64
network: 2001::/64
HGE1/0/1
HGE1/0/2
HGE1/0/4
DHCPv6 server 2
FE90::1/64
Switch
DHCPv6 guard
network: 2001::/64
HGE1/0/3
DHCPv6 client
DHCPv6 client
FE80::2/64
DHCPv6 server 3
network: 2002::/64



Procedure

Before you configure DHCPv6 guard, complete the configuration on DHCPv6 servers.

Create VLAN 100, and assign GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, GigabitEthernet 1/0/3, and GigabitEthernet 1/0/4 to VLAN 100.

```
<Switch> system-view
[Switch] vlan 100
[Switch-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[Switch-vlan100] quit
```

Create an IPv6 basic ACL numbered 2001.

```
[Switch] acl ipv6 number 2001
```

Create rule **1** to permit only packets with source IPv6 addresses in the range of FE80::/12.

```
[Switch-acl-ipv6-basic-2001] rule 1 permit source fe80:: 12
[Switch-acl-ipv6-basic-2001] quit
```

Create an IPv6 basic ACL numbered 2002.

```
[Switch] acl ipv6 number 2002
```

Create rule **1** to permit only packets with source IPv6 addresses in the range of 2001::/16.

```
[Switch-acl-ipv6-basic-2002] rule 1 permit source 2001:: 16
[Switch-acl-ipv6-basic-2002] quit
```

Create DHCPv6 guard policy named **p1**.

```
[Switch] ipv6 dhcp guard policy p1
```

Set the device role to the DHCPv6 server for the device attached to the target VLAN.

```
[Switch-dhcp6-guard-policy-p1] device-role server
```

Specify ACL 2001 to match DHCPv6 servers.

```
[Switch-dhcp6-guard-policy-p1] if-match server acl 2001
```

Specify ACL 2002 to match IPv6 addresses/prefixes assigned by DHCPv6 servers.

```
[Switch-dhcp6-guard-policy-p1] if-match reply acl 2002
[Switch-dhcp6-guard-policy-p1] quit
```

Create DHCPv6 guard policy named **p2**.

```
[Switch] ipv6 dhcp guard policy p2
```

Set the device role to the DHCPv6 client for the device attached to the target interface.

```
[Switch-dhcp6-guard-policy-p2] device-role client
[Switch-dhcp6-guard-policy-p2] quit
```

Apply DHCPv6 guard policy **p1** to VLAN 100.

```
[Switch] vlan 100
[Switch-vlan100] ipv6 dhcp guard apply policy p1
[Switch-vlan100] quit
```

Apply DHCPv6 guard policy **p2** to GigabitEthernet 1/0/4.

```
[Switch]interface gigabitethernet 1/0/4
[Switch-GigabitEthernet1/0/4] ipv6 dhcp guard apply policy p2
[Switch-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Verify that the switch forwards DHCPv6 replies with the source IPv6 address in the range of FE80::/12 and the assigned IPv6 prefixes in the range of 2001::/16. The switch forwards DHCPv6 replies from the DHCPv6 server 1 and drops replies from DHCPv6 server 2 and server 3.

Contents

Configuring IPv6 fast forwarding 1
About IPv6 fast forwarding 1
Configuring the aging time for IPv6 software fast forwarding entries 1
Configuring IPv6 fast forwarding load sharing for software 1
Display and maintenance commands for IPv6 fast forwarding 2

Configuring IPv6 fast forwarding

About IPv6 fast forwarding

Fast forwarding reduces route lookup time and improves packet forwarding efficiency by using a high-speed cache and data-flow-based technology. It identifies a data flow by using the following fields:

- Source IPv6 address.
- Destination IPv6 address.
- Source port number.
- Destination port number.
- Protocol number.

After a flow's first packet is forwarded through the routing table, fast forwarding creates an entry and uses the entry to forward subsequent packets of the flow.

Configuring the aging time for IPv6 software fast forwarding entries

About this task

The IPv6 fast forwarding table uses an aging timer for each forwarding entry. If an entry is not updated before the timer expires, the device deletes the entry. If an entry has a hit within the aging time, the aging timer restarts.

Procedure

1. Enter system view.
   `system-view`
2. Set the aging time for IPv6 fast forwarding entries on the software side.
   `ipv6 fast-forwarding aging-time` *`aging-time`*
   By default, the aging time is 30 seconds for IPv6 fast forwarding entries on the software side.

Configuring IPv6 fast forwarding load sharing for software

About this task

IPv6 software fast forwarding load sharing enables the device to identify a data flow by using the packet information.

If IPv6 fast forwarding load sharing is disabled on the software side, the device identifies a data flow by the packet information and the input interface.

Procedure

1. Enter system view.
   `system-view`
2. Configure IPv6 fast forwarding load sharing on the software side. Choose one option as needed:

- Enable IPv6 fast forwarding load sharing on the software side.
  Ipv6 fast-forwarding load-sharing
- Disable IPv6 fast forwarding load sharing on the software side.
  undo ipv6 fast-forwarding load-sharing

By default, IPv6 fast forwarding load sharing is enabled on the software side.

Display and maintenance commands for IPv6 fast forwarding

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the aging time of the IPv6 fast forwarding entries.	**display ipv6 fast-forwarding aging-time**
Display IPv6 fast forwarding entries.	**display ipv6 fast-forwarding cache** [*ipv6-address*] [**slot** *slot-number*]
Clear the IPv6 fast forwarding table.	**reset ipv6 fast-forwarding cache** [**slot** *slot-number*]

Contents

Configuring tunneling 1
About tunneling 1
Supported tunneling technologies 1
Restrictions and guidelines: Tunnel interface configuration 1
Configuring a tunnel interface 1
About tunnel interface configuration 1
Tunnel interface configuration tasks at a glance 1
Creating a tunnel interface 1
Configuring parameters for tunneled packets 3
Restoring the default settings of the tunnel interface 3
Enabling SNMP notifications for the tunneling module 3
Display and maintenance commands for tunnel interface configuration 4
Troubleshooting tunnel interface configuration 5
Tunnel interface not up 5
IPv6 over IPv4 tunneling 6
About IPv6 over IPv4 tunneling 6
Implementation 6
Tunnel modes 6
IPv6 over IPv4 tunneling tasks at a glance 8
Configuring an IPv6 over IPv4 manual tunnel 8
Restrictions and guidelines 8
Procedure 9
Example: Configuring an IPv6 over IPv4 manual tunnel 9
Configuring a 6to4 tunnel 12
Restrictions and guidelines 12
Procedure 12
Example: Configuring a 6to4 tunnel 12
Example: Configuring 6to4 relay 15
Configuring an ISATAP tunnel 17
Restrictions and guidelines 17
Procedure 17
Example: Configuring an ISATAP tunnel 18
IPv4 over IPv4 tunneling 22
About IPv4 over IPv4 tunneling 22
Restrictions and guidelines: IPv4 over IPv4 tunnel configuration 22
Configuring an IPv4 over IPv4 tunnel 23
IPv4 over IPv4 tunnel configuration examples 24
Example: Configuring an IPv4 over IPv4 tunnel 24
IPv4 over IPv6 tunneling 27
About IPv4 over IPv6 tunneling 27
Implementation 27
Configuring an IPv4 over IPv6 tunnel 27
Restrictions and guidelines 27
Procedure 28
Example: Configuring an IPv4 over IPv6 tunnel 28
IPv6 over IPv6 tunneling 32
About IPv6 over IPv6 tunneling 32
Restrictions and guidelines: IPv6 over IPv6 tunnel configuration 32
IPv6 over IPv6 tunnel configuration tasks at a glance 33
Configuring an IPv6 over IPv6 tunnel 33
IPv6 over IPv6 tunnel configuration examples 34
Example: Configuring an IPv6 over IPv6 tunnel 34

Configuring tunneling

This chapter describes tunnel interface configuration. For information about tunnel modes, see the subsequent chapters.

About tunneling

Tunneling encapsulates the packets of a network protocol within the packets of a second network protocol and transfers them over a virtual point-to-point connection. The virtual connection is called a tunnel. Packets are encapsulated at the tunnel source and de-encapsulated at the tunnel destination.

Supported tunneling technologies

Tunneling supports the following technologies:

- GRE tunneling. For more information, see "Configuring GRE."
- VXLAN tunneling. For more information, see *VXLAN Configuration Guide*.
- IPv6 over IPv4 tunneling, IPv4 over IPv4 tunneling, and IPv4/IPv6 over IPv6 tunneling.

Restrictions and guidelines: Tunnel interface configuration

Do not specify the same tunnel source and destination addresses for the tunnel interfaces on the same device.

On IPv4 and IPv6 VXLAN tunnel interfaces, IP addresses (assigned by **ip address** and **ipv6 address** commands) are meaningless. As a best practice, do not assign IP addresses to those types of tunnel interfaces.

Configuring a tunnel interface

About tunnel interface configuration

Configure a tunnel interface (Layer 3 virtual interface) at both ends of a tunnel. The devices use the tunnel interface to identify, process, and send packets for the tunnel.

Tunnel interface configuration tasks at a glance

To configure a tunnel interface, perform the following tasks:

1. Creating a tunnel interface
2. (Optional.) Configuring parameters for tunneled packets
3. (Optional.) Restoring the default settings of the tunnel interface

Creating a tunnel interface

1. Enter system view.

system-view

2. Create a tunnel interface, specify the tunnel mode, and enter tunnel interface view.

   interface tunnel *number* **[mode** { **gre** [**ipv6**] | **ipv4-ipv4** | **ipv4-ipv6** | **ipv6-ipv4** [**6to4** | **isatap**] | **ipv6-ipv6** | { **vxlan** [**ipv6**] }]

   The IPv4 VXLAN tunnel mode and IPv6 VXLAN tunnel mode are supported only in expert operating mode.

   For packet tunneling to succeed, the two ends of a tunnel must use the same tunnel mode.

3. Configure a source address or source interface for the tunnel interface.

   source { *ipv4-address* | *ipv6-address* | *interface-type interface-number* }

   By default, no source address or source interface is configured for the tunnel interface.

   If you specify a source address, it is used as the source address of tunneled packets.

   If you specify a source interface, the primary IP address of this interface is used as the source IP address of tunneled packets.

4. Configure a destination address for the tunnel interface.

   destination { *ipv4-address* | *ipv6-address* }

   By default, no destination address is configured for the tunnel interface.

   The tunnel destination address must be the IP address of the receiving interface on the tunnel peer. It is used as the destination IP address of tunneled packets.

5. (Optional.) Configure a description for the interface.

   description *text*

   By default, the description for a tunnel interface is **Tunnel** *number* **Interface**.

6. (Optional.) Set the MTU of the tunnel interface.

   mtu *size*

   The default settings are as follows:

   - If the tunnel interface has never been up, the MTU is 64000 bytes.
   - If the tunnel interface is up, its MTU is identical to the outgoing interface's MTU minus the length of the tunnel headers. The outgoing interface is automatically obtained through routing table lookup based on the tunnel destination address.

   You can use the **ip mtu**, **ipv6 mtu**, or **mtu** command to set the MTU of a tunnel interface.

   - The **mtu** command applies to both IPv4 and IPv6 tunnel interfaces.
   - The **ip mtu** and **ipv6 mtu** commands have higher priority than the **mtu** command on a tunnel interface.
   - If all of the commands are used on a tunnel interface, the **ip mtu** and **ipv6 mtu** commands take effect. After you restore the default settings for the **ip mtu** and **ipv6 mtu** commands, the **mtu** command takes effect.
   - If none of the commands is used on a tunnel interface, the MTU of the tunnel interface is identical to the outgoing interface's MTU minus the length of the tunnel headers. The outgoing interface is automatically obtained through routing table lookup based on the tunnel destination address.

7. (Optional.) Set the expected bandwidth for the tunnel interface.

   bandwidth *bandwidth-value*

   The default expected bandwidth (in kbps) is the interface maximum rate divided by 1000.

   The expected bandwidth is an informational parameter used only by higher-layer protocols for calculation. You cannot adjust the actual bandwidth of an interface by using this command.

8. Bring up the tunnel interface.

   undo shutdown

By default, a tunnel interface is not administratively down.

Configuring parameters for tunneled packets

1. Enter system view.
   system-view
2. Enter tunnel interface view.
   interface tunnel *number*
3. Set the ToS for tunneled packets.
   tunnel tos *tos-value*

   The default setting is the same as the ToS of the original packets.

   This command is not supported on IPv4 VXLAN tunnel interfaces or IPv6 VXLAN tunnel interfaces.
4. Set the TTL for tunneled packets.
   tunnel ttl *ttl-value*

   The default TTL for tunneled packets is 255.

Restoring the default settings of the tunnel interface

Restrictions and guidelines

CAUTION:
This operation might interrupt ongoing network services. Make sure you are fully aware of the impact of this operation when you perform it on a live network.

This operation might fail to restore the default settings for some commands for reasons such as command dependencies or system restrictions. Use the **display this** command in interface view to identify these commands. Use their **undo** forms or follow the command reference to restore their default settings. If your restoration attempt still fails, follow the error message instructions to resolve the problem.

Procedure

1. Enter system view.
   system-view
2. Enter tunnel interface view.
   interface tunnel *number*
3. Restore the default settings of the tunnel interface.
   default

Enabling SNMP notifications for the tunneling module

About this task

After you enable the **max-threshold-reached** notification, when the number of tunnel interfaces exceeds the warning threshold (80% of the maximum), the device generates SNMP notifications that record the chassis number and slot number of the exceeding tunnel interfaces.

After you enable the **vxlan-config-failure** notification, when a VXLAN tunnel failed to be deployed, the device generates an SNMP notification that records the interface number of the tunnel.

After you enable the **vxlan-decap-failure** notification, when a VXLAN packet failed to be decapsulated, the device generates an SNMP notification that records the source IP type, source IP address, destination IP type, and destination IP address of the VXLAN packet.

After you enable the **vxlan-tunnel-status** notification, the device generates SNMP notifications for VXLAN over IPv4 tunnels when the tunnel status changes.

After you enable the **vxlan-ipv6-tunnel-status** notification, the device generates SNMP notifications for VXLAN over IPv6 tunnels when the tunnel status changes.

If you do not specify the notification types to be enabled, the device enables all SNMP notifications for the tunneling module.

To output tunnel notifications correctly, you must also configure parameters for sending SNMP notifications. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for the tunneling module.

   snmp-agent trap enable tunnel [**max-threshold-reached** | **vxlan-config-failure** | **vxlan-decap-failure** | **vxlan-tunnel-status** | **vxlan-ipv6-tunnel-status**] *

   By default, SNMP notifications for the tunneling module are disabled.

Display and maintenance commands for tunnel interface configuration

Execute **display** commands in any view and **reset** commands in user view.

Task	Command	Remarks
Display information about tunnel interfaces.	**display interface** [**tunnel** [*number*]] [**brief** [**description** | **down**]]	N/A
Display IPv6 information on tunnel interfaces.	**display ipv6 interface** [**tunnel** [*number*]] [**brief**]	For more information about this command, see IPv6 basics in *Layer 3—IP Services Command Reference*.
Clear statistics on tunnel interfaces.	**reset counters interface** [**tunnel** [*number*]]	N/A
Clear IPv6 statistics on tunnel interfaces.	**reset ipv6 statistics** [**slot** *slot-number*]	For more information about this command, see IPv6 basics in *Layer 3—IP Services Command Reference*.

Troubleshooting tunnel interface configuration

Tunnel interface not up

Symptom

A tunnel interface configured with related parameters such as tunnel source address, tunnel destination address, and tunnel mode cannot come up.

Analysis

The physical interface of the tunnel does not come up, or the tunnel destination is unreachable.

Solution

1. To resolve the problem:
   - Use the **`display interface`** or **`display ipv6 interface`** command to verify that the physical interface of the tunnel is up. If the physical interface is down, check the network connection.
   - Use the **`display ipv6 routing-table`** or **`display ip routing-table`** command to verify that the tunnel destination is reachable. If the route is not available, configure a route to reach the tunnel destination.
2. If the problem persists, contact H3C Support.

IPv6 over IPv4 tunneling

About IPv6 over IPv4 tunneling

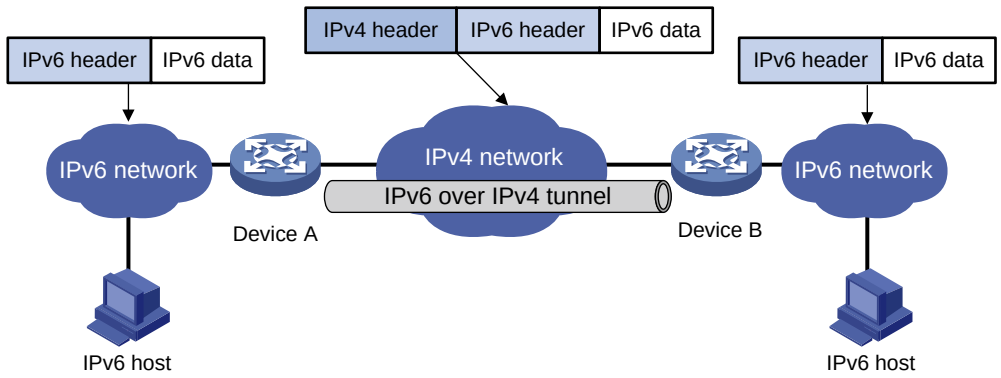
Implementation

IPv6 over IPv4 tunneling enables isolated IPv6 networks to communicate, as shown in Figure 1.

NOTE:

The devices at both ends of an IPv6 over IPv4 tunnel must support the IPv4/IPv6 dual stack.

Figure 1 IPv6 over IPv4 tunnel



The IPv6 over IPv4 tunnel processes packets by using the following steps:

1. A host in the IPv6 network sends an IPv6 packet to Device A at the tunnel source.
2. After Device A receives the IPv6 packet, it processes the packet as follows:
   a. Searches the routing table to identify the outgoing interface for the IPv6 packet.

      The outgoing interface is the tunnel interface, so Device A knows that the packet needs to be forwarded through the tunnel.

   b. Adds an IPv4 header to the IPv6 packet and forwards the packet through the physical interface of the tunnel.

      In the IPv4 header, the source IPv4 address is the IPv4 address of the tunnel source, and the destination IPv4 address is the IPv4 address of the tunnel destination.
3. Upon receiving the packet, Device B de-encapsulates the packet.
4. If the destination address of the IPv6 packet is itself, Device B forwards it to the upper-layer protocol. If it is not, Device B forwards it according to the routing table.

Tunnel modes

IPv6 over IPv4 tunnels include manually configured tunnels and automatic tunnels, depending on how the IPv4 address of the tunnel destination is obtained.

- **Manually configured tunnel**—The destination IPv4 address of the tunnel cannot be automatically obtained from the destination IPv6 address of an IPv6 packet at the tunnel source. It must be manually configured.

- **Automatic tunnel**—The destination IPv4 address of the tunnel can be automatically obtained from the destination IPv6 address (with an IPv4 address embedded) of an IPv6 packet at the tunnel source.

The source IPv4 addresses for all IPv6 over IPv4 tunnels are manually configured.

According to the way an IPv6 packet is encapsulated, IPv6 over IPv4 tunnels are divided into the modes shown in the following sections.

IPv6 over IPv4 manual tunneling

An IPv6 over IPv4 manual tunnel is a point-to-point link. To establish a manual tunnel, you must manually configure the source and destination addresses of the tunnel at both ends of the tunnel.

Manual tunneling provides the following solutions:

- Connects isolated IPv6 networks over an IPv4 network.
- Connects an IPv6 network and an IPv4/IPv6 dual-stack host over an IPv4 network.

6to4 tunneling

- Ordinary 6to4 tunneling

  A 6to4 tunnel is a point-to-multipoint automatic tunnel. It is used to connect multiple isolated IPv6 networks over an IPv4 network.

  The ends of a 6to4 tunnel are 6to4 addresses. The address format is 2002:abcd:efgh:subnet number::interface ID/48.

  - 2002 is the fixed IPv6 address prefix.
  - abcd:efgh represents a 32-bit globally unique IPv4 address in hexadecimal notation.

    For example, 1.1.1.1 can be represented by 0101:0101. The IPv4 address identifies a 6to4 network (an IPv6 network where all hosts use 6to4 addresses). The border router of a 6to4 network must have the IPv4 address abcd:efgh configured on the interface connected to the IPv4 network.
  - The subnet number identifies a subnet in the 6to4 network.
  - The subnet number::interface ID uniquely identifies a host in the 6to4 network.

  The destination IPv4 address of a 6to4 tunnel is embedded in the destination 6to4 address. This mechanism enables the device to automatically obtain the tunnel destination address.

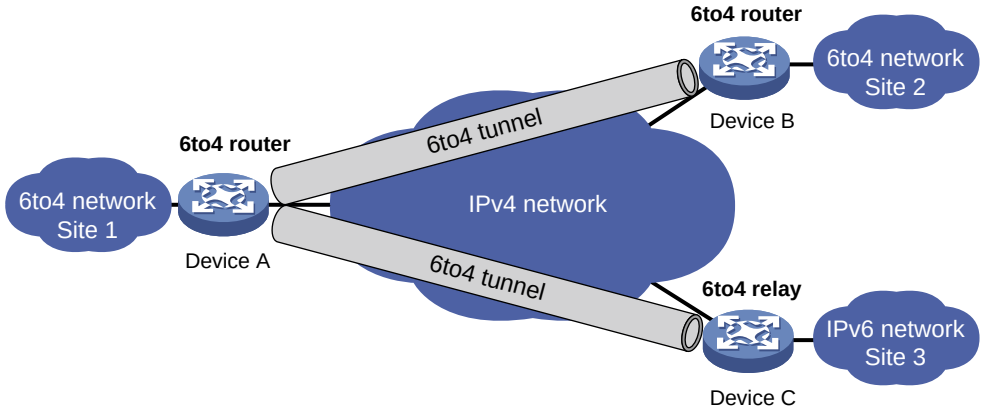
  6to4 tunneling uses an IPv4 address to identify a 6to4 network. This method overcomes the limitations of automatic IPv4-compatible IPv6 tunneling.

- 6to4 relay

  6to4 relay connects a 6to4 network and an IPv6 network that uses an IP prefix other than 2002::/16. A 6to4 relay router is a gateway that forwards packets from a 6to4 network to an IPv6 network.

  As shown in Figure 2, 6to4 network Site 1 communicates with IPv6 network Site 3 over a 6to4 tunnel. Configure a static route on the border router (Device A) in the 6to4 network. The next hop address must be the 6to4 address of the 6to4 relay router (Device C). Device A forwards all packets destined for the IPv6 network over the 6to4 tunnel, and Device C then forwards them to the IPv6 network.

Figure 2 Principle of 6to4 tunneling and 6to4 relay



ISATAP tunneling

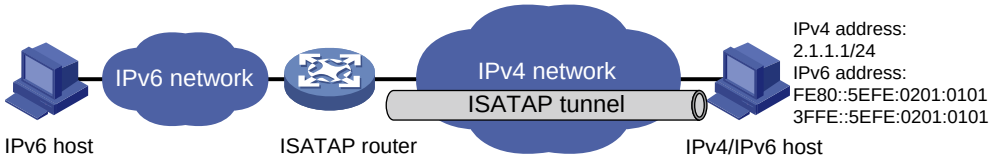
An ISATAP tunnel is a point-to-multipoint automatic tunnel. It provides a solution to connect an IPv6 host and an IPv6 network over an IPv4 network.

The destination address of an ISATAP tunnel is an ISATAP address. The address format is prefix:0:5EFE:abcd:efgh/64.

- The 64-bit prefix is a valid IPv6 unicast address prefix.
- The abcd:efgh/64 segments represent a 32-bit IPv4 address in hexadecimal notation, which identifies the tunnel destination but does not require global uniqueness.

ISATAP tunnels are mainly used for communication between IPv6 routers or between an IPv6 host and an IPv6 router over an IPv4 network.

Figure 3 Principle of ISATAP tunneling



IPv6 over IPv4 tunneling tasks at a glance

To configure IPv6 over IPv4 tunneling, perform the following tasks:

1. Configuring an IPv6 over IPv4 tunnel

   Choose one of the following tasks:

   - Configuring an IPv6 over IPv4 manual tunnel
   - Configuring a 6to4 tunnel
   - Configuring an ISATAP tunnel

Configuring an IPv6 over IPv4 manual tunnel

Restrictions and guidelines

When you perform tasks in this section, follow these restrictions and guidelines:

- The tunnel destination address specified on the local device must be identical with the tunnel source address specified on the tunnel peer device.
- Do not specify the same tunnel source and destination addresses for the tunnel interfaces in the same mode on a device.
- To ensure correct packet forwarding, identify whether the destination IPv6 network and the IPv6 address of the local tunnel interface are on the same subnet. If they are not, configure a route reaching the destination IPv6 network through the tunnel interface. You can configure the route by using one of the following methods:
  - Configure a static route, and specify the local tunnel interface as the egress interface or specify the IPv6 address of the peer tunnel interface as the next hop.
  - Enable a dynamic routing protocol on the tunnel interface.

  The route configuration is required on both ends of the tunnel. For more information about route configuration, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter IPv6 over IPv4 manual tunnel interface view.

   interface tunnel *number* [**mode ipv6-ipv4**]
3. Specify an IPv6 address for the tunnel interface.

   See "Configuring basic IPv6 settings."
4. Configure a source address or source interface for the tunnel interface.

   source { *ipv4-address* | *interface-type interface-number* }

   By default, no source address or source interface is configured for the tunnel interface.

   If you specify a source address, it is used as the source IP address of tunneled packets.

   If you specify a source interface, the primary IP address of this interface is used as the source IP address of tunneled packets.
5. Configure a destination address for the tunnel interface.

   destination *ipv4-address*

   By default, no destination address is configured for the tunnel interface.

   The tunnel destination address must be the IP address of the receiving interface on the tunnel peer. It is used as the destination IP address of tunneled packets.
6. (Optional.) Set the DF bit for tunneled packets.

   tunnel dfbit enable

   By default, the DF bit is not set for tunneled packets.

Example: Configuring an IPv6 over IPv4 manual tunnel

Network configuration

As shown in Figure 4, configure an IPv6 over IPv4 tunnel between Switch A and Switch B so the two IPv6 networks can reach each other over the IPv4 network. Because the tunnel destination IPv4 address cannot be automatically obtained from the destination IPv6 addresses, configure an IPv6 over IPv4 manual tunnel.

Figure 4 Network diagram

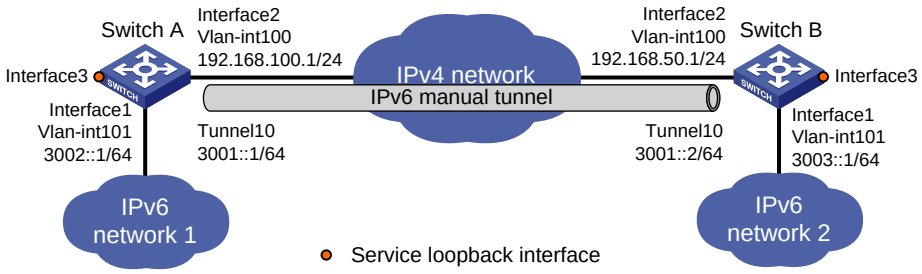


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Make sure Switch A and Switch B have the corresponding VLAN interfaces created and can reach each other through IPv4.

1. Configure Switch A:

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/2
[SwitchA-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.168.100.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/1 to VLAN 101.

```
[SwitchA] vlan 101
[SwitchA-vlan101] port gigabitethernet 1/0/1
[SwitchA-vlan101] quit
```

Specify an IPv6 address for VLAN-interface 101.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 3002::1 64
[SwitchA-Vlan-interface101] quit
```

Create IPv6 over IPv4 manual tunnel interface Tunnel10.

```
[SwitchA] interface tunnel 10 mode ipv6-ipv4
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchA-Tunnel10] ipv6 address 3001::1/64
```

Specify VLAN-interface 100 as the source interface of the tunnel interface.

```
[SwitchA-Tunnel10] source vlan-interface 100
```

Specify the destination address for the tunnel interface as the IP address of the VLAN-interface 100 on Switch B.

```
[SwitchA-Tunnel10] destination 192.168.50.1
```

```
[SwitchA-Tunnel10] quit
```

Configure a static route destined for IPv6 network 2 through Tunnel10.

```
[SwitchA] ipv6 route-static 3003:: 64 tunnel 10
```

2. Configure Switch B:

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/2
[SwitchB-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 192.168.50.1 255.255.255.0
[SwitchB-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/1 to VLAN 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port gigabitethernet 1/0/1
[SwitchB-vlan101] quit
```

Specify an IPv6 address for VLAN-interface 101.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 3003::1 64
[SwitchB-Vlan-interface101] quit
```

Create IPv6 over IPv4 manual tunnel interface Tunnel10.

```
[SwitchB] interface tunnel 10 mode ipv6-ipv4
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchB-Tunnel10] ipv6 address 3001::2/64
```

Specify VLAN-interface 100 as the source interface of the tunnel interface.

```
[SwitchB-Tunnel10] source vlan-interface 100
```

Specify the destination address for the tunnel interface as the IP address of VLAN-interface 100 of Switch A.

```
[SwitchB-Tunnel10] destination 192.168.100.1
[SwitchB-Tunnel10] quit
```

Configure a static route destined for IPv6 network 1 through Tunnel10.

```
[SwitchB] ipv6 route-static 3002:: 64 tunnel 10
```

Verifying the configuration

Use the **display ipv6 interface** command to display tunnel interface status on Switch A and Switch B. Verify that the interface Tunnel10 is up. (Details not shown.)

Verify that Switch B and Switch A can ping the IPv6 address of VLAN-interface 101 of each other. This example uses Switch A.

```
[SwitchA] ping ipv6 3003::1
Ping6(56 data bytes) 3001::1 --> 3003::1, press CTRL+C to break
56 bytes from 3003::1, icmp_seq=0 hlim=64 time=45.000 ms
56 bytes from 3003::1, icmp_seq=1 hlim=64 time=10.000 ms
56 bytes from 3003::1, icmp_seq=2 hlim=64 time=4.000 ms
56 bytes from 3003::1, icmp_seq=3 hlim=64 time=10.000 ms
56 bytes from 3003::1, icmp_seq=4 hlim=64 time=11.000 ms

--- Ping6 statistics for 3003::1 ---
```

```
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 4.000/16.000/45.000/14.711 ms
```

Configuring a 6to4 tunnel

Restrictions and guidelines

Follow these guidelines when you configure a 6to4 tunnel:

- You do not need to configure a destination address for a 6to4 tunnel, because the destination IPv4 address is embedded in the 6to4 IPv6 address.
- Do not specify the same source addresses for local tunnel interfaces in the same tunnel mode.
- Automatic tunnels do not support dynamic routing. You must configure a static route destined for the destination IPv6 network if the destination IPv6 network is not in the same subnet as the IPv6 address of the tunnel interface. You can specify the local tunnel interface as the egress interface of the route or specify the IPv6 address of the peer tunnel interface as the next hop of the route. The route configuration is required on both ends of the tunnel. For more information about route configuration, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Enter 6to4 tunnel interface view.
   `interface tunnel` *`number`* [**`mode ipv6-ipv4 6to4`**]
3. Specify an IPv6 address for the tunnel interface.
   See "Configuring basic IPv6 settings."
4. Configure a source address or source interface for the tunnel interface.
   `source` { *`ipv4-address`* | *`interface-type interface-number`* }
   By default, no source address or source interface is configured for the tunnel interface.
   If you specify a source address, it is used as the source IP address of tunneled packets.
   If you specify a source interface, the primary IP address of this interface is used as the source IP address of tunneled packets.
5. (Optional.) Set the DF bit for tunneled packets.
   `tunnel dfbit enable`
   By default, the DF bit is not set for tunneled packets.

Example: Configuring a 6to4 tunnel

Network configuration

As shown in Figure 5, configure a 6to4 tunnel between 6to4 switches Switch A and Switch B so the two hosts can reach each other over the IPv4 network.

Figure 5 Network diagram

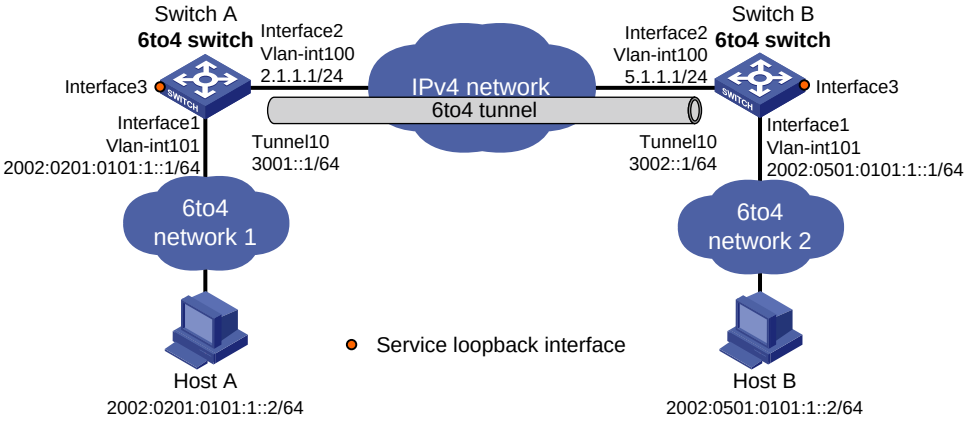


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Requirements analysis

To enable communication between 6to4 networks, configure 6to4 addresses for 6to4 switches and hosts in the 6to4 networks.

- The IPv4 address of VLAN-interface 100 on Switch A is 2.1.1.1/24, and the corresponding 6to4 prefix is 2002:0201:0101::/48. Host A must use this prefix.
- The IPv4 address of VLAN-interface 100 on Switch B is 5.1.1.1/24, and the corresponding 6to4 prefix is 2002:0501:0101::/48. Host B must use this prefix.

Procedure

Make sure Switch A and Switch B have the corresponding VLAN interfaces created and can reach each other through IPv4.

1. Configure Switch A:

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/2
[SwitchA-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 2.1.1.1 24
[SwitchA-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/1 to VLAN 101.

```
[SwitchA] vlan 101
[SwitchA-vlan101] port gigabitethernet 1/0/1
[SwitchA-vlan101] quit
```

Specify a 6to4 address for VLAN-interface 101.

```
[SwitchA] interface vlan-interface 101
```

```
[SwitchA-Vlan-interface101] ipv6 address 2002:0201:0101:1::1/64
[SwitchA-Vlan-interface101] quit
```

Create 6to4 tunnel interface Tunnel10.

```
[SwitchA] interface tunnel 10 mode ipv6-ipv4 6to4
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchA-Tunnel10] ipv6 address 3001::1/64
```

Specify the source interface as VLAN-interface 100 for the tunnel interface.

```
[SwitchA-Tunnel10] source vlan-interface 100
[SwitchA-Tunnel10] quit
```

Configure a static route destined for 2002::/16 through the tunnel interface.

```
[SwitchA] ipv6 route-static 2002:: 16 tunnel 10
```

2. Configure Switch B:

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/2
[SwitchB-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 5.1.1.1 24
[SwitchB-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/1 to VLAN 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port gigabitethernet 1/0/1
[SwitchB-vlan101] quit
```

Specify a 6to4 address for VLAN-interface 101.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 2002:0501:0101:1::1/64
[SwitchB-Vlan-interface101] quit
```

Create 6to4 tunnel interface Tunnel10.

```
[SwitchB] interface tunnel 10 mode ipv6-ipv4 6to4
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchB-Tunnel10] ipv6 address 3002::1/64
```

Specify the source interface as VLAN-interface 100 for the tunnel interface.

```
[SwitchB-Tunnel10] source vlan-interface 100
[SwitchB-Tunnel10] quit
```

Configure a static route destined for 2002::/16 through the tunnel interface.

```
[SwitchB] ipv6 route-static 2002:: 16 tunnel 10
```

Verifying the configuration

Verify that Host A and Host B can ping each other.

```
D:\>ping6 -s 2002:201:101:1::2 2002:501:101:1::2

Pinging 2002:501:101:1::2
from 2002:201:101:1::2 with 32 bytes of data:

Reply from 2002:501:101:1::2: bytes=32 time=13ms
```

```
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time<1ms

Ping statistics for 2002:501:101:1::2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 13ms, Average = 3ms
```

Example: Configuring 6to4 relay

Network configuration

As shown in Figure 6, Switch A is a 6to4 switch, and 6to4 addresses are used on the connected IPv6 network. Switch B acts as a 6to4 relay switch and is connected to the IPv6 network (2001::/64). Configure a 6to4 tunnel between Switch A and Switch B to make Host A and Host B reachable to each other.

The configuration on a 6to4 relay switch is similar to that on a 6to4 switch. To enable communication between the 6to4 network and the IPv6 network, configure a route destined for the IPv6 network on the 6to4 switch. The IPv4 address of VLAN-interface 100 on the relay switch is 6.1.1.1/24 and its corresponding 6to4 prefix is 2002:0601:0101::/48. The next hop of the static route must be an address using this prefix.

Figure 6 Network diagram

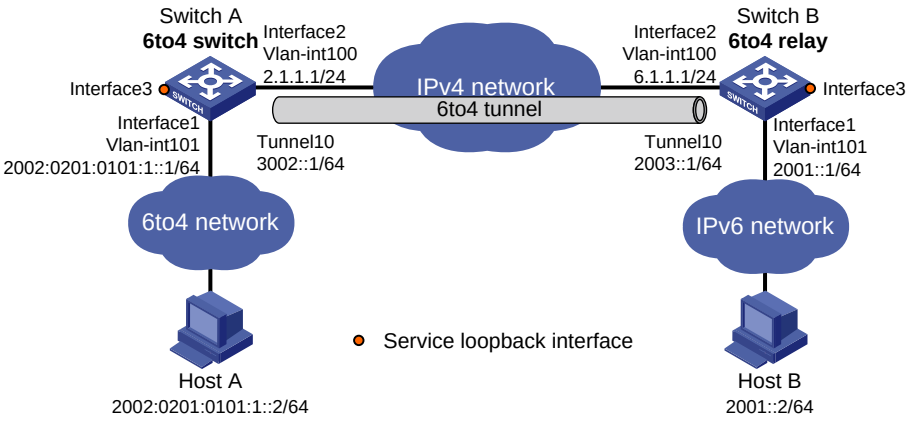


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Make sure Switch A and Switch B have the corresponding VLAN interfaces created and can reach each other through IPv4.

1. Configure Switch A:

   # Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 100.

   ```
   <SwitchA> system-view
   ```

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/2
[SwitchA-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 2.1.1.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/1 to VLAN 101.

```
[SwitchA] vlan 101
[SwitchA-vlan101] port gigabitethernet 1/0/1
[SwitchA-vlan101] quit
```

Specify a 6to4 address for VLAN-interface 101.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 2002:0201:0101:1::1/64
[SwitchA-Vlan-interface101] quit
```

Create 6to4 tunnel interface Tunnel10.

```
[SwitchA] interface tunnel 10 mode ipv6-ipv4 6to4
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchA-Tunnel10] ipv6 address 3002::1/64
```

Specify the source interface as VLAN-interface 100 for the tunnel interface.

```
[SwitchA-Tunnel10] source vlan-interface 100
[SwitchA-Tunnel10] quit
```

Configure a static route destined for the 6to4 relay switch.

```
[SwitchA] ipv6 route-static 2002:0601:0101:: 64 tunnel 10
```

Configure a default route to reach the IPv6 network, which specifies the 6to4 address of the 6to4 relay switch as the next hop.

```
[SwitchA] ipv6 route-static :: 0 2002:0601:0101::1
```

2. Configure Switch B:

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/2
[SwitchB-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 6.1.1.1 255.255.255.0
[SwitchB-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/1 to VLAN 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port gigabitethernet 1/0/1
[SwitchB-vlan101] quit
```

Specify an IPv6 address for VLAN-interface 101.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 2001::1/64
[SwitchB-Vlan-interface101] quit
```

Create 6to4 tunnel interface Tunnel10.

```
[SwitchB] interface tunnel 10 mode ipv6-ipv4 6to4
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchB-Tunnel10] ipv6 address 2003::1/64
```

Specify VLAN-interface 100 as the source interface of the tunnel interface.

```
[SwitchB-Tunnel10] source vlan-interface 100
[SwitchB-Tunnel10] quit
```

Configure a static route destined for 2002::/16 through the tunnel interface.

```
[SwitchB] ipv6 route-static 2002:: 16 tunnel 10
```

Verifying the configuration

Verify that Host A and Host B can ping each other.

```
D:\>ping6 -s 2002:201:101:1::2 2001::2

Pinging 2001::2
from 2002:201:101:1::2 with 32 bytes of data:

Reply from 2001::2: bytes=32 time=13ms
Reply from 2001::2: bytes=32 time=1ms
Reply from 2001::2: bytes=32 time=1ms
Reply from 2001::2: bytes=32 time<1ms

Ping statistics for 2001::2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 13ms, Average = 3ms
```

Configuring an ISATAP tunnel

Restrictions and guidelines

Follow these guidelines when you configure an ISATAP tunnel:

- You do not need to configure a destination address for an ISATAP tunnel, because the destination IPv4 address is embedded in the ISATAP address.
- Do not specify the same source addresses for local tunnel interfaces in the same tunnel mode.
- Automatic tunnels do not support dynamic routing. You must configure a static route destined for the destination IPv6 network if the destination IPv6 network is not in the same subnet as the IPv6 address of the tunnel interface. You can specify the local tunnel interface as the egress interface of the route or specify the IPv6 address of the peer tunnel interface as the next hop of the route. The route configuration is required on both ends of the tunnel. For more information about route configuration, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter ISATAP tunnel interface view.
   interface tunnel *number* [**mode ipv6-ipv4 isatap**]
3. Specify an IPv6 address for the tunnel interface.
   See "Configuring basic IPv6 settings."

4. Configure a source address or source interface for the tunnel interface.

   source { *ipv4-address* | *interface-type interface-number* }

   By default, no source address or source interface is configured for the tunnel interface.

   If you specify a source address, it is used as the source IP address of tunneled packets.

   If you specify a source interface, the primary IP address of this interface is used as the source IP address of tunneled packets.

5. (Optional.) Set the DF bit for tunneled packets.

   tunnel dfbit enable

   By default, the DF bit is not set for tunneled packets.

Example: Configuring an ISATAP tunnel

Network configuration

As shown in Figure 7, configure an ISATAP tunnel between the switch and the ISATAP host so the ISATAP host in the IPv4 network can access the IPv6 network.

Figure 7 Network diagram

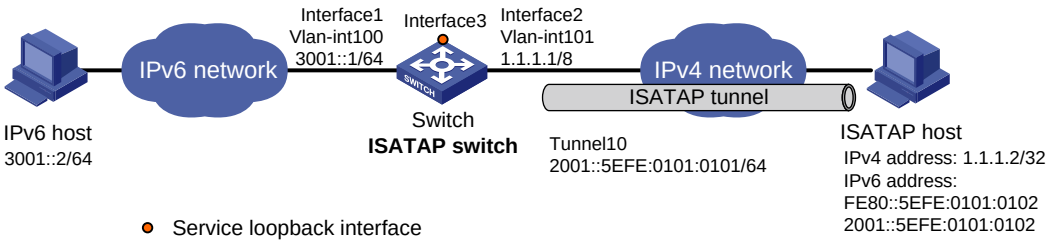


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure the switch:

   # Add GigabitEthernet 1/0/1 to VLAN 100.

   ```
   <Sysname> system-view
   [Sysname] sysname Switch
   [Switch] vlan 100
   [Switch-vlan100] port gigabitethernet 1/0/1
   [Switch-vlan100] quit
   ```

   # Specify an IPv6 address for VLAN-interface 100.

   ```
   [Switch] interface vlan-interface 100
   [Switch-Vlan-interface100] ipv6 address 3001::1/64
   [Switch-Vlan-interface100] quit
   ```

   # Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 101.

   ```
   [Switch] vlan 101
   [Switch-vlan101] port gigabitethernet 1/0/2
   ```

```
[Switch-vlan101] quit
```

Specify an IPv4 address for VLAN-interface 101.

```
[Switch] interface vlan-interface 101
[Switch-Vlan-interface101] ip address 1.1.1.1 255.0.0.0
[Switch-Vlan-interface101] quit
```

Create ISATAP tunnel interface Tunnel10.

```
[Switch] interface tunnel 10 mode ipv6-ipv4 isatap
```

Specify an EUI-64 IPv6 address for the tunnel interface.

```
[Switch-Tunnel10] ipv6 address 2001:: 64 eui-64
```

Specify VLAN-interface 101 as the source interface of the tunnel interface.

```
[Switch-Tunnel10] source vlan-interface 101
```

Disable RA suppression so that the ISATAP host can acquire information such as the address prefix from the RA message advertised by the ISATAP switch.

```
[Switch-Tunnel10] undo ipv6 nd ra halt
[Switch-Tunnel10] quit
```

2. Configure the ISATAP host:

Configurations on the ISATAP host vary by operating system. The following configuration is performed on Windows 10.

On a Windows 10 host, specify the IPv4 address of the ISATAP switch, for example, 1.1.1.1.

```
C:\>netsh
netsh>interface ipv6 isatap
netsh interface ipv6 isatap>set router 1.1.1.1
netsh interface ipv6 isatap>set state enable
netsh interface ipv6 isatap>exit
```

Display information about the ISATAP interface. In this example, the ISATAP host obtained prefix 2001::/64 and automatically generated global IPv6 unicast address 2001::5efe:1.1.1.2.

```
C:\>ipconfig

Windows IP Configuration



Ethernet adapter Ethernet 3:

   Connection-specific DNS Suffix  . :
   Link-local IPv6 Address . . . . . : fe80::3d6d:7f04:bed1:c4ab%20
   IPv4 Address. . . . . . . . . . . : 1.1.1.2
   Subnet Mask . . . . . . . . . . . : 255.0.0.0
   Default Gateway . . . . . . . . . :

Tunnel adapter isatap.{FD960BD5-5546-4B3C-A8E8-4B9EC352AA05}:

   Connection-specific DNS Suffix  . :
   IPv6 Address. . . . . . . . . . . : 2001::200:5efe:1.1.1.2
   Link-local IPv6 Address . . . . . : fe80::200:5efe:1.1.1.2%52
   Default Gateway . . . . . . . . . : fe80::5efe:1.1.1.1%52
```

Display information about IPv6 routes on the host.

```
C:\>route print -6
===========================================================================
```

```
Interface List
20...08 00 27 00 98 dd ......VirtualBox Host-Only Ethernet Adapter
52...00 00 00 00 00 00 00 e0 Microsoft ISATAP Adapter
  1...........................Software Loopback Interface 1
===========================================================================

IPv6 Route Table
===========================================================================
Active Routes:
If Metric Network Destination     Gateway
52    301 ::/0                     fe80::5efe:1.1.1.1
  1    331 ::1/128                  On-link
52    301 2001::/64                On-link
52    301 2001::200:5efe:1.1.1.2/128
                                    On-link
20    291 fe80::/64                On-link
52    301 fe80::200:5efe:1.1.1.2/128
20    291 fe80::3d6d:7f04:bed1:c4ab/128
                                    On-link
  1    331 ff00::/8                 On-link
20    291 ff00::/8                 On-link
===========================================================================
Persistent Routes:
  None
```

3. On the IPv6 host, configure a route to the ISATAP switch.

```
C:\>netsh interface ipv6 set route 2001::/64 5 3001::1
```

Verifying the configuration

Verify that the ISATAP host can ping the IPv6 host.

```
C:\>ping 3001::2

Pinging 3001::2 with 32 bytes of data:
Reply from 3001::2: time<1ms
Reply from 3001::2: time<1ms
Reply from 3001::2: time<1ms
Reply from 3001::2: time<1ms

Ping statistics for 3001::2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 0ms, Average = 0ms
```

Configuration files

Switch:

```
#
 sysname Switch
#
vlan 100 to 101
#
```

```
interface Vlan-interface100
 ipv6 address 3001::1/64
#
interface Vlan-interface101
 ip address 1.1.1.1 255.0.0.0
#
interface GigabitEthernet1/0/1
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port access vlan 101
#
interface Tunnel10 mode ipv6-ipv4 isatap
 source Vlan-interface101
 ipv6 address 2001::/64 eui-64
 undo ipv6 nd ra halt
#
return
```

IPv4 over IPv4 tunneling

About IPv4 over IPv4 tunneling

IPv4 over IPv4 tunneling (RFC 1853) enables isolated IPv4 networks to communicate. For example, an IPv4 over IPv4 tunnel can connect isolated private IPv4 networks over a public IPv4 network.

Figure 8 IPv4 over IPv4 tunnel

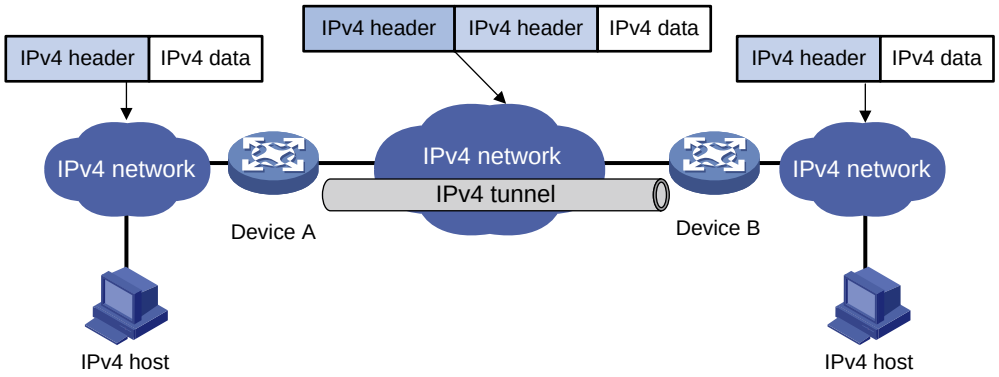


Figure 8 shows the encapsulation and de-encapsulation processes.

- Encapsulation:
  - **a.** Device A receives an IP packet from an IPv4 host and submits it to the IP protocol stack.
  - **b.** The IPv4 protocol stack determines how to forward the packet according to the destination address in the IP header. If the packet is destined for the IPv4 host connected to Device B, Device A delivers the packet to the tunnel interface.
  - **c.** The tunnel interface adds a new IPv4 header to the IPv4 packet and submits it to the IP protocol stack.

    In the new header, the source IP address specifies the tunnel source, and the destination IP address specifies the tunnel destination.
  - **d.** The IP protocol stack uses the destination IP address of the new IP header to look up the routing table, and then sends the packet out.
- De-encapsulation:
  - **a.** After receiving the packet, Device B delivers it to the IP protocol stack.
  - **b.** If the protocol number is 4 (indicating an IPv4 packet is encapsulated within the packet), the IP protocol stack delivers the packet to the tunnel module for de-encapsulation.
  - **c.** The tunnel module de-encapsulates the IP packet and sends it back to the IP protocol stack.
  - **d.** The protocol stack forwards the de-encapsulated packet.

Restrictions and guidelines: IPv4 over IPv4 tunnel configuration

Follow these guidelines when you configure an IPv4 over IPv4 tunnel:

- The tunnel destination address specified on the local device must be identical with the tunnel source address specified on the tunnel peer device.

- Do not specify the same source and destination addresses for local tunnel interfaces in the same tunnel mode.
- The IPv4 address of the local tunnel interface cannot be on the same subnet as the destination address configured on the tunnel interface.
- To ensure correct packet forwarding, identify whether the destination IPv4 network and the IPv4 address of the local tunnel interface are on the same subnet. If they are not, configure a route reaching the destination IPv4 network through the tunnel interface. You can configure the route by using one of the following methods:
  - Configure a static route, and specify the local tunnel interface as the egress interface or specify the IPv4 address of the peer tunnel interface as the next hop.
  - Enable a dynamic routing protocol on the tunnel interface.

  The route configuration is required on both ends of the tunnel. For more information about route configuration, see *Layer 3—IP Routing Configuration Guide*.
- The destination address of the route passing the tunnel interface cannot be on the same subnet as the destination address configured on the tunnel interface.

Configuring an IPv4 over IPv4 tunnel

1. Enter system view.

   system-view
2. Enter IPv4 over IPv4 tunnel interface view.

   interface tunnel *number* [**mode ipv4-ipv4**]
3. Configure an IPv4 address for the tunnel interface.

   ip address *ip-address* { *mask* | *mask-length* } [**sub**]
4. Configure a source address or source interface for the tunnel interface.

   source { *ipv4-address* | *interface-type interface-number* }

   By default, no source address or source interface is configured for the tunnel interface.

   If you specify a source address, it is used as the source IP address of tunneled packets.

   If you specify a source interface, the primary IP address of this interface is used as the source IP address of tunneled packets.
5. Configure a destination address for the tunnel interface.

   destination *ipv4-address*

   By default, no destination address is configured for the tunnel interface.

   The tunnel destination address must be the IP address of the receiving interface on the tunnel peer. It is used as the destination IP address of tunneled packets.
6. (Optional.) Set the DF bit for tunneled packets.

   tunnel dfbit enable

   By default, the DF bit is not set for tunneled packets.

IPv4 over IPv4 tunnel configuration examples

Example: Configuring an IPv4 over IPv4 tunnel

Network configuration

As shown in Figure 9, the two subnets IPv4 group 1 and IPv4 group 2 use private IPv4 addresses. Configure an IPv4 over IPv4 tunnel between Switch A and Switch B to make the two subnets reachable to each other.

Figure 9 Network diagram

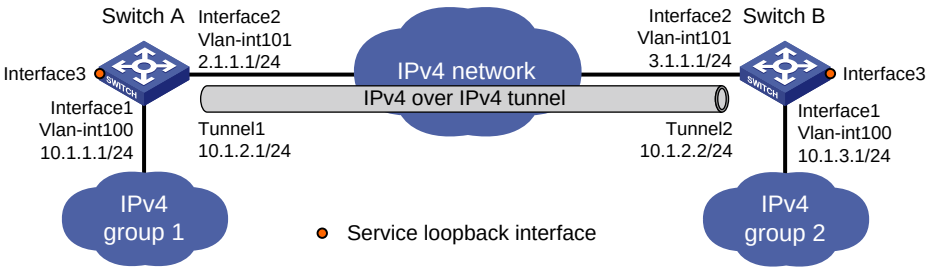


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Make sure Switch A and Switch B have the corresponding VLAN interfaces created and can reach each other through IPv4.

1. Configure Switch A:

   # Add GigabitEthernet 1/0/1 to VLAN 100.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 100
   [SwitchA-vlan100] port gigabitethernet 1/0/1
   [SwitchA-vlan100] quit
   ```

   # Specify an IPv4 address for VLAN-interface 100.

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] ip address 10.1.1.1 255.255.255.0
   [SwitchA-Vlan-interface100] quit
   ```

   # Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 101.

   ```
   [SwitchA] vlan 101
   [SwitchA-vlan101] port gigabitethernet 1/0/2
   [SwitchA-vlan101] quit
   ```

   # Specify an IPv4 address for VLAN-interface 101.

   ```
   [SwitchA] interface vlan-interface 101
   [SwitchA-Vlan-interface101] ip address 2.1.1.1 255.255.255.0
   [SwitchA-Vlan-interface101] quit
   ```

Create IPv4 over IPv4 tunnel interface Tunnel 1.

```
[SwitchA] interface tunnel 1 mode ipv4-ipv4
```

Specify an IPv4 address for the tunnel interface.

```
[SwitchA-Tunnel1] ip address 10.1.2.1 255.255.255.0
```

Specify the IP address of VLAN-interface 101 as the source address for the tunnel interface.

```
[SwitchA-Tunnel1] source 2.1.1.1
```

Specify the IP address of VLAN-interface 101 on Switch B as the destination address for the tunnel interface.

```
[SwitchA-Tunnel1] destination 3.1.1.1
[SwitchA-Tunnel1] quit
```

Configure a static route destined for IPv4 group 2 through the tunnel interface.

```
[SwitchA] ip route-static 10.1.3.0 255.255.255.0 tunnel 1
```

2. Configure Switch B:

Add GigabitEthernet 1/0/1 to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1
[SwitchB-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 10.1.3.1 255.255.255.0
[SwitchB-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port gigabitethernet 1/0/2
[SwitchB-vlan101] quit
```

Specify an IPv4 address for VLAN-interface 101.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ip address 3.1.1.1 255.255.255.0
[SwitchB-Vlan-interface101] quit
```

Create IPv4 over IPv4 tunnel interface Tunnel 2.

```
[SwitchB] interface tunnel 2 mode ipv4-ipv4
```

Specify an IPv4 address for the tunnel interface.

```
[SwitchB-Tunnel2] ip address 10.1.2.2 255.255.255.0
```

Specify the IP address of VLAN-interface 101 as the source address for the tunnel interface.

```
[SwitchB-Tunnel2] source 3.1.1.1
```

Specify the IP address of VLAN-interface 101 on Switch A as the destination address for the tunnel interface.

```
[SwitchB-Tunnel2] destination 2.1.1.1
[SwitchB-Tunnel2] quit
```

Configure a static route destined for IPv4 group 1 through the tunnel interface.

```
[SwitchB] ip route-static 10.1.1.0 255.255.255.0 tunnel 2
```

Verifying the configuration

Use the **`display interface tunnel`** command to display the status of the tunnel interfaces on Switch A and Switch B. Verify that the tunnel interfaces are up. (Details not shown.)

Verify that Switch A and Switch B can ping the IPv4 address of the peer interface VLAN-interface 100. This example uses Switch A.

```
[SwitchA] ping -a 10.1.1.1 10.1.3.1
Ping 10.1.3.1 (10.1.3.1) from 10.1.1.1: 56 data bytes, press CTRL+C to break
56 bytes from 10.1.3.1: icmp_seq=0 ttl=255 time=2.000 ms
56 bytes from 10.1.3.1: icmp_seq=1 ttl=255 time=1.000 ms
56 bytes from 10.1.3.1: icmp_seq=2 ttl=255 time=0.000 ms
56 bytes from 10.1.3.1: icmp_seq=3 ttl=255 time=1.000 ms
56 bytes from 10.1.3.1: icmp_seq=4 ttl=255 time=1.000 ms

--- Ping statistics for 10.1.3.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/1.000/2.000/0.632 ms
```

IPv4 over IPv6 tunneling

About IPv4 over IPv6 tunneling

Implementation

IPv4 over IPv6 tunneling adds an IPv6 header to IPv4 packets so that the IPv4 packets can pass an IPv6 network through a tunnel to realize interworking between isolated IPv4 networks.

Figure 10 IPv4 over IPv6 tunnel

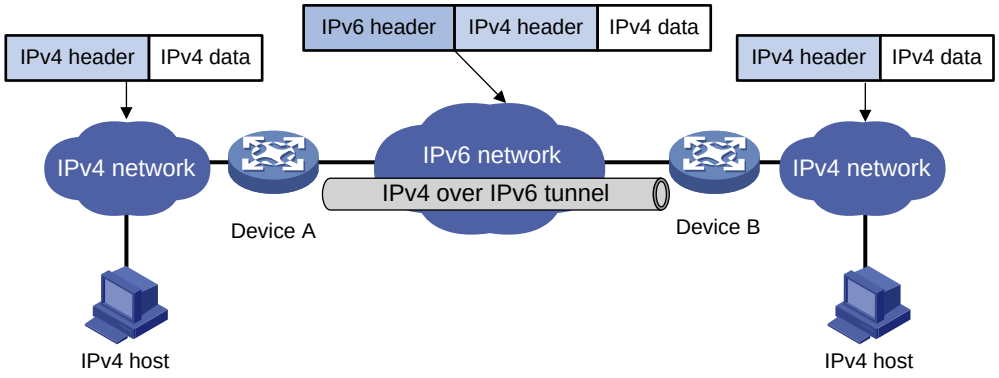


Figure 10 shows the encapsulation and de-encapsulation processes.

- Encapsulation:
  a. Upon receiving an IPv4 packet, Device A delivers it to the IPv4 protocol stack.
  b. The IPv4 protocol stack uses the destination address of the packet to determine the egress interface. If the egress interface is the tunnel interface, the IPv4 protocol stack delivers the packet to the tunnel interface.
  c. The tunnel interface adds an IPv6 header to the original IPv4 packet and delivers the packet to the IPv6 protocol stack.
  d. The IPv6 protocol stack uses the destination IPv6 address of the packet to look up the routing table, and then sends it out.
- De-encapsulation:
  a. Upon receiving the IPv6 packet from the attached IPv6 network, Device B delivers the packet to the IPv6 protocol stack to examine the protocol type encapsulated in the data portion of the packet.
  b. If the protocol type is IPv4, the IPv6 protocol stack delivers the packet to the tunneling module.
  c. The tunneling module removes the IPv6 header and delivers the remaining IPv4 packet to the IPv4 protocol stack.
  d. The IPv4 protocol stack forwards the IPv4 packet.

Configuring an IPv4 over IPv6 tunnel

Restrictions and guidelines

When you perform the tasks in this section, follow these restrictions and guidelines:

- The tunnel destination address specified on the local device must be identical with the tunnel source address specified on the tunnel peer device.
- Do not specify the same source and destination addresses for local tunnel interfaces in the same tunnel mode.
- To ensure correct packet forwarding, identify whether the destination IPv4 network and the IPv4 address of the local tunnel interface are on the same subnet. If they are not, configure a route reaching the destination IPv4 network through the tunnel interface. You can configure the route by using one of the following methods:
  - Configure a static route, and specify the local tunnel interface as the egress interface or specify the IPv6 address of the peer tunnel interface as the next hop.
  - Enable a dynamic routing protocol on the tunnel interface.

  The route configuration is required on both ends of the tunnel. For more information about route configuration, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter IPv4 over IPv6 tunnel interface view.

   interface tunnel *number* [**mode ipv4-ipv6**]
3. Configure an IPv4 address for the tunnel interface.

   ip address *ip-address* { *mask* | *mask-length* } [**sub**]
4. Configure the source address or interface for the tunnel interface.

   source { *ipv6-address* | *interface-type interface-number* }

   By default, no source address or interface is configured for the tunnel.

   If you specify a source address, it is used as the source IPv6 address of tunneled packets.

   If you specify a source interface, the lowest IPv6 address of this interface is used as the source IPv6 address of tunneled packets.
5. Configure the destination address for the tunnel interface.

   destination *ipv6-address*

   By default, no destination address is configured for the tunnel.

   The tunnel destination address must be the IPv6 address of the receiving interface on the tunnel peer. It is used as the destination IPv6 address of tunneled packets.

Example: Configuring an IPv4 over IPv6 tunnel

Network configuration

As shown in Figure 11, configure an IPv4 over IPv6 manual tunnel between Switch A and Switch B so the two networks can reach each other over the IPv6 network.

Figure 11 Network diagram

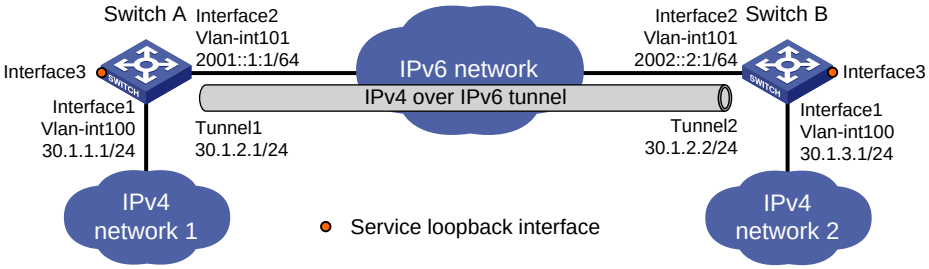


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Make sure Switch A and Switch B have the corresponding VLAN interfaces created and can reach each other through IPv6.

1. Configure Switch A:

Add GigabitEthernet 1/0/1 to VLAN 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1
[SwitchA-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 30.1.1.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 101.

```
[SwitchA] vlan 101
[SwitchA-vlan101] port gigabitethernet 1/0/2
[SwitchA-vlan101] quit
```

Specify an IPv6 address for VLAN-interface 101.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 2001::1:1 64
[SwitchA-Vlan-interface101] quit
```

Create IPv4 over IPv6 tunnel interface Tunnel 1.

```
[SwitchA] interface tunnel 1 mode ipv4-ipv6
```

Specify an IPv4 address for the tunnel interface.

```
[SwitchA-Tunnel1] ip address 30.1.2.1 255.255.255.0
```

Specify the IP address of VLAN-interface 101 as the source address for the tunnel interface.

```
[SwitchA-Tunnel1] source 2001::1:1
```

Specify the IP address of VLAN-interface 101 on Switch B as the destination address for the tunnel interface.

```
[SwitchA-Tunnel1] destination 2003::2:1
```

```
[SwitchA-Tunnel1] quit
```

Configure a static route destined for IPv4 network 2 through the tunnel interface.

```
[SwitchA] ip route-static 30.1.3.0 255.255.255.0 tunnel 1
```

2. Configure Switch B:

Add GigabitEthernet 1/0/1 to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1
[SwitchB-vlan100] quit
```

Specify an IPv4 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 30.1.3.1 255.255.255.0
[SwitchB-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port gigabitethernet 1/0/2
[SwitchB-vlan101] quit
```

Specify an IPv6 address for VLAN-interface 101.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 2003::2:1 64
[SwitchB-Vlan-interface101] quit
```

Create IPv4 over IPv6 tunnel interface Tunnel 2.

```
[SwitchB] interface tunnel 2 mode ipv4-ipv6
```

Specify an IPv4 address for the tunnel interface.

```
[SwitchB-Tunnel2] ip address 30.1.2.2 255.255.255.0
```

Specify the IP address of VLAN-interface 101 as the source address for the tunnel interface (IP address of VLAN-interface 101).

```
[SwitchB-Tunnel2] source 2003::2:1
```

Specify the IP address of VLAN-interface 101 on Switch A as the destination address for the tunnel interface.

```
[SwitchB-Tunnel2] destination 2001::1:1
[SwitchB-Tunnel2] quit
```

Configure a static route destined for IPv4 network 1 through the tunnel interface.

```
[SwitchB] ip route-static 30.1.1.0 255.255.255.0 tunnel 2
```

Verifying the configuration

Use the **`display interface tunnel`** command to display the status of the tunnel interfaces on Switch A and Switch B. Verify that the tunnel interfaces are up. (Details not shown.)

Verify that Switch A and Switch B can ping the IPv4 address of the peer interface. This example uses Switch A.

```
[SwitchA] ping -a 30.1.1.1 30.1.3.1
Ping 30.1.3.1 (30.1.3.1) from 30.1.1.1: 56 data bytes, press CTRL+C to break
56 bytes from 30.1.3.1: icmp_seq=0 ttl=255 time=3.000 ms
56 bytes from 30.1.3.1: icmp_seq=1 ttl=255 time=1.000 ms
56 bytes from 30.1.3.1: icmp_seq=2 ttl=255 time=0.000 ms
56 bytes from 30.1.3.1: icmp_seq=3 ttl=255 time=1.000 ms
56 bytes from 30.1.3.1: icmp_seq=4 ttl=255 time=1.000 ms
```

```
--- Ping statistics for 30.1.3.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/1.200/3.000/0.980 ms
```

IPv6 over IPv6 tunneling

About IPv6 over IPv6 tunneling

IPv6 over IPv6 tunneling (RFC 2473) enables isolated IPv6 networks to communicate with each other over another IPv6 network. For example, two isolated IPv6 networks that do not want to show their addresses to the Internet can use an IPv6 over IPv6 tunnel to communicate with each other.

Figure 12 Principle of IPv6 over IPv6 tunneling

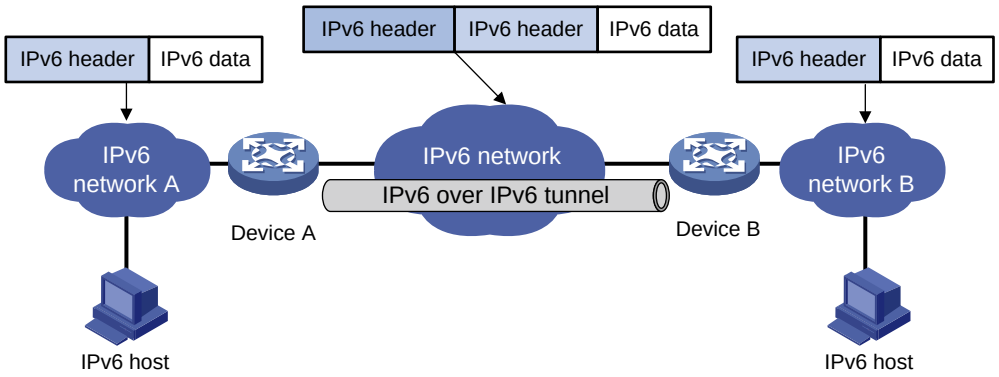


Figure 12 shows the encapsulation and de-encapsulation processes.

- Encapsulation:
  - **a.** After receiving an IPv6 packet, Device A submits it to the IPv6 protocol stack.
  - **b.** The IPv6 protocol stack uses the destination IPv6 address of the packet to find the egress interface. If the egress interface is the tunnel interface, the stack delivers it to the tunnel interface.
  - **c.** After receiving the packet, the tunnel interface adds an IPv6 header to it and submits it to the IPv6 protocol stack.
  - **d.** The IPv6 protocol stack forwards the packet according to its destination IPv6 address.
- De-encapsulation:
  - **a.** Upon receiving the IPv6 packet, Device B delivers it to the IPv6 protocol stack.
  - **b.** The IPv6 protocol stack checks the protocol type of the data portion encapsulated in the IPv6 packet. If the encapsulation protocol is IPv6, the stack delivers the packet to the tunneling module.
  - **c.** The tunneling module de-encapsulates the packet and sends it back to the IPv6 protocol stack.
  - **d.** The IPv6 protocol stack forwards the IPv6 packet.

Restrictions and guidelines: IPv6 over IPv6 tunnel configuration

Follow these guidelines when you configure an IPv6 over IPv6 tunnel:

- The tunnel destination address specified on the local device must be identical with the tunnel source address specified on the tunnel peer device.
- Do not specify the same source and destination addresses for local tunnel interfaces in the same tunnel mode.

- The IPv6 address of the tunnel interface cannot be on the same subnet as the destination address configured for the tunnel interface.
- To ensure correct packet forwarding, identify whether the destination IPv6 network and the IPv6 address of the local tunnel interface are on the same subnet. If they are not, configure a route reaching the destination IPv6 network through the tunnel interface. You can configure the route by using one of the following methods:
  - Configure a static route, and specify the local tunnel interface as the egress interface or specify the IPv6 address of the peer tunnel interface as the next hop.
  - Enable a dynamic routing protocol on the tunnel interface.

  The route configuration is required on both ends of the tunnel. For more information about route configuration, see *Layer 3—IP Routing Configuration Guide*.
- The destination address of the route passing the tunnel interface cannot be on the same subnet as the destination address configured for the tunnel interface.

IPv6 over IPv6 tunnel configuration tasks at a glance

To configure an IPv6 over IPv6 tunnel, perform the following tasks:

1. Configuring an IPv6 over IPv6 tunnel

Configuring an IPv6 over IPv6 tunnel

1. Enter system view.

   system-view
2. Enter IPv6 over IPv6 tunnel interface view.

   interface tunnel *number* [**mode ipv6-ipv6**]
3. Configure an IPv6 address for the tunnel interface.

   See "Configuring basic IPv6 settings."
4. Configure the source address or source interface for the tunnel interface.

   source { *ipv6-address* | *interface-type interface-number* }

   By default, no source address or interface is configured for the tunnel.

   If you specify a source address, it is used as the source IPv6 address of tunneled packets.

   If you specify a source interface, the lowest IPv6 address of this interface is used as the source IPv6 address of tunneled packets.
5. Configure the destination address for the tunnel interface.

   destination *ipv6-address*

   By default, no destination address is configured for the tunnel.

   The tunnel destination address must be the IPv6 address of the receiving interface on the tunnel peer. It is used as the destination IPv6 address of tunneled packets.

IPv6 over IPv6 tunnel configuration examples

Example: Configuring an IPv6 over IPv6 tunnel

Network configuration

As shown in Figure 13, configure an IPv6 over IPv6 tunnel between Switch A and Switch B so the two networks can reach each other without disclosing their IPv6 addresses.

Figure 13 Network diagram

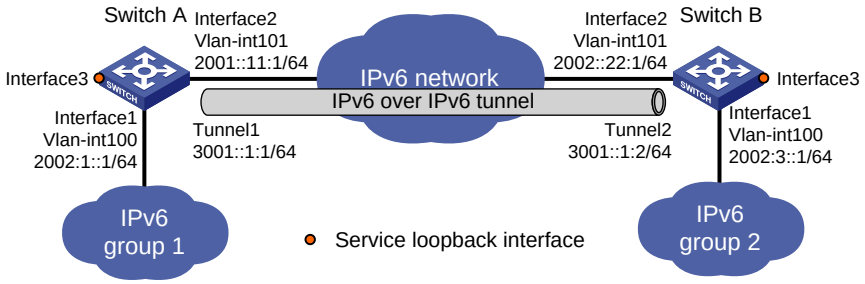


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Make sure Switch A and Switch B have the corresponding VLAN interfaces created and can reach each other through IPv6.

1. Configure Switch A:

   # Add GigabitEthernet 1/0/1 to VLAN 100.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 100
   [SwitchA-vlan100] port gigabitethernet 1/0/1
   [SwitchA-vlan100] quit
   ```

   # Specify an IPv6 address for VLAN-interface 100.

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] ipv6 address 2003:1::1 64
   [SwitchA-Vlan-interface100] quit
   ```

   # Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 101.

   ```
   [SwitchA] vlan 101
   [SwitchA-vlan101] port gigabitethernet 1/0/2
   [SwitchA-vlan101] quit
   ```

   # Specify an IPv6 address for VLAN-interface 101.

   ```
   [SwitchA] interface vlan-interface 101
   [SwitchA-Vlan-interface101] ipv6 address 2001::11:1 64
   [SwitchA-Vlan-interface101] quit
   ```

   # Create IPv6 over IPv6 tunnel interface Tunnel 1.

```
[SwitchA] interface tunnel 1 mode ipv6-ipv6
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchA-Tunnel1] ipv6 address 3001::1:1 64
```

Specify the IP address of VLAN-interface 101 as the source address for the tunnel interface.

```
[SwitchA-Tunnel1] source 2001::11:1
```

Specify the IP address of VLAN-interface 101 on Switch B as the destination address for the tunnel interface.

```
[SwitchA-Tunnel1] destination 2003::22:1
[SwitchA-Tunnel1] quit
```

Configure a static route destined for the IPv6 network group 2 through the tunnel interface.

```
[SwitchA] ipv6 route-static 2003:3:: 64 tunnel 1
```

2. Configure Switch B:

Add GigabitEthernet 1/0/1 to VLAN 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1
[SwitchB-vlan100] quit
```

Specify an IPv6 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ipv6 address 2003:3::1 64
[SwitchB-Vlan-interface100] quit
```

Add GigabitEthernet 1/0/2 (the physical interface of the tunnel) to VLAN 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port gigabitethernet 1/0/2
[SwitchB-vlan101] quit
```

Specify an IPv6 address for VLAN-interface 101.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 2003::22:1 64
[SwitchB-Vlan-interface101] quit
```

Create IPv6 over IPv6 tunnel interface Tunnel 2.

```
[SwitchB] interface tunnel 2 mode ipv6-ipv6
```

Specify an IPv6 address for the tunnel interface.

```
[SwitchB-Tunnel2] ipv6 address 3001::1:2 64
```

Specify the IP address of VLAN-interface 101 as the source address for the tunnel interface.

```
[SwitchB-Tunnel2] source 2003::22:1
```

Specify the IP address of VLAN-interface 101 on Switch A as the destination address for the tunnel interface.

```
[SwitchB-Tunnel2] destination 2001::11:1
[SwitchB-Tunnel2] quit
```

Configure a static route destined for the IPv6 network group 1 through the tunnel interface.

```
[SwitchB] ipv6 route-static 2003:1:: 64 tunnel 2
```

Verifying the configuration

Use the **display ipv6 interface** command to display the status of the tunnel interfaces on Switch A and Switch B. Verify that the tunnel interfaces are up. (Details not shown.)

Verify that Switch A and Switch B can ping the IPv4 address of the peer interface. This example uses Switch A.

```
[SwitchA] ping ipv6 -a 2003:1::1 2003:3::1
```

```
Ping6(56 data bytes) 2003:1::1 --> 2003:3::1, press CTRL+C to break
56 bytes from 2003:3::1, icmp_seq=0 hlim=64 time=9.000 ms
56 bytes from 2003:3::1, icmp_seq=1 hlim=64 time=1.000 ms
56 bytes from 2003:3::1, icmp_seq=2 hlim=64 time=0.000 ms
56 bytes from 2003:3::1, icmp_seq=3 hlim=64 time=0.000 ms
56 bytes from 2003:3::1, icmp_seq=4 hlim=64 time=0.000 ms

--- Ping6 statistics for 2003:3::1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/2.000/9.000/3.521 ms
```

Contents

Configuring GRE 1

About GRE 1
GRE encapsulation format 1
GRE tunnel operating principle 1
GRE security mechanisms 2
GRE application scenarios 2
Protocols and standards 3
Restrictions and guidelines: GRE configuration 3
Configuring a GRE/IPv4 tunnel 4
Configuring a GRE/IPv6 tunnel 5
Display and maintenance commands for GRE 6
GRE configuration examples 6
Example: Configuring an IPv4 over IPv4 GRE tunnel 6
Example: Configuring an IPv4 over IPv6 GRE tunnel 9
Troubleshooting GRE 11
Hosts at both ends of a GRE tunnel cannot ping each other 12

Configuring GRE

About GRE

Generic Routing Encapsulation (GRE) is a tunneling protocol that can encapsulate a protocol (such as IP or Ethernet) into a virtual point-to-point tunnel over a network (such as an IP network). Packets are encapsulated at one tunnel end and de-encapsulated at the other tunnel end. The network layer protocol of the packets before encapsulation and after encapsulation can be the same or different.

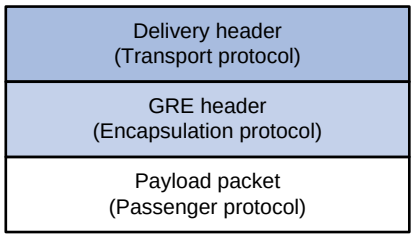
GRE encapsulation format

As shown in Figure 1, a GRE-tunneled packet includes the following parts:

- **Payload packet**—Original packet. The protocol type of the payload packet is called the passenger protocol. The passenger protocol can be any network layer protocol.
- **GRE header**—Header that is added to the payload packet to change the payload packet to a GRE packet. A GRE header includes the number of encapsulations, version, passenger protocol type, checksum, and key. GRE is called the encapsulation protocol.
- **Delivery header**—Header that is added to the GRE packet to deliver it to the tunnel end. The transport protocol (or delivery protocol) is the network layer protocol that transfers GRE packets.

The device supports GRE tunnels with IPv4 and IPv6 as the transport protocols. When the transport protocol is IPv4, the GRE tunnel mode is GRE over IPv4 (GRE/IPv4). When the transport protocol is IPv6, the GRE tunnel mode is GRE over IPv6 (GRE/IPv6).

Figure 1 GRE encapsulation format



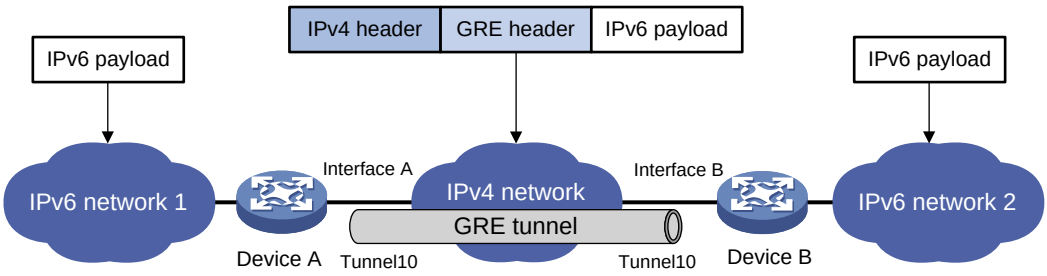
GRE tunnel operating principle

As shown in Figure 2, an IPv6 protocol packet traverses an IPv4 network through a GRE tunnel as follows:

1. After receiving an IPv6 packet from the interface connected to IPv6 network 1, Device A processes the packet as follows:
   a. Looks up the routing table to identify the outgoing interface for the IPv6 packet.
   b. Submits the IPv6 packet to the outgoing interface—the GRE tunnel interface Tunnel 10.
2. Upon receiving the packet, the tunnel interface encapsulates the packet with GRE and then with IPv4. In the IPv4 header:
   - The source address is the tunnel's source address (the IP address of Interface A of Device A).
   - The destination address is the tunnel's destination address (the IP address of Interface B of Device B).

3. Device A looks up the routing table according to the destination address in the IPv4 header, and forwards the IPv4 packet out of the physical interface (Interface A) of the GRE tunnel.
4. When the IPv4 arrives at the GRE tunnel destination Device B, Device B checks the destination address. Because the destination is Device B itself and the protocol number in the IP header is 47 (the protocol number for GRE), Device B submits the packet to GRE for de-encapsulation.
5. GRE first removes the IPv4 header, and then checks the GRE key and packet sequence number. After GRE finishes the checking, it removes the GRE header, and submits the payload to the IPv6 protocol for forwarding.

Figure 2 IPv6 networks interconnected through a GRE tunnel



GRE security mechanisms

GRE supports only the GRE key security mechanism.

GRE key

GRE keys ensure packet validity. The sender adds a GRE key into a packet. The receiver compares the GRE key with its own GRE key. If the two keys are the same, the receiver accepts the packet. If the two keys are different, the receiver drops the packet.

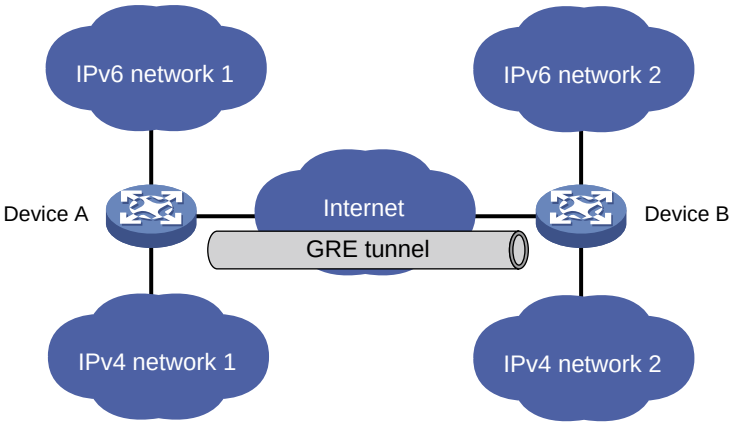
GRE application scenarios

The following shows typical GRE application scenarios:

Connecting networks running different protocols over a single backbone

As shown in Figure 3, IPv6 network 1 and IPv6 network 2 are IPv6 networks, and IPv4 network 1 and IPv4 network 2 are IPv4 networks. Through the GRE tunnel between Device A and Device B, IPv6 network 1 can communicate with IPv6 network 2 and IPv4 network 1 can communicate with IPv4 network 2, without affecting each other.

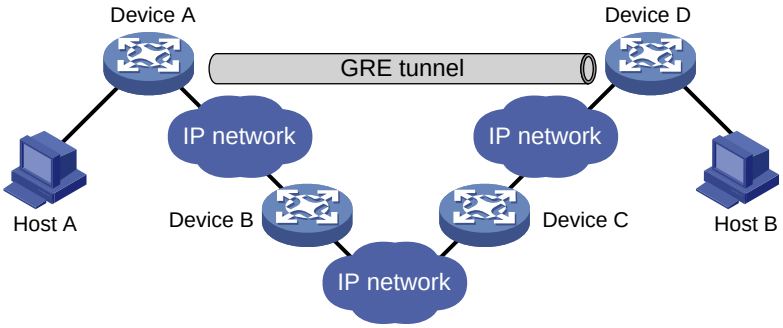
Figure 3 Network diagram



Enlarging network scope

In an IP network, the maximum TTL value of a packet is 255. If two devices have more than 255 hops in between, they cannot communicate with each other. By using a GRE tunnel, you can hide some hops to enlarge the network scope. As shown in Figure 4, only the tunnel-end devices (Device A and Device D) of the GRE tunnel are counted in hop count calculation. Therefore, there are only three hops between Host A and Host B.

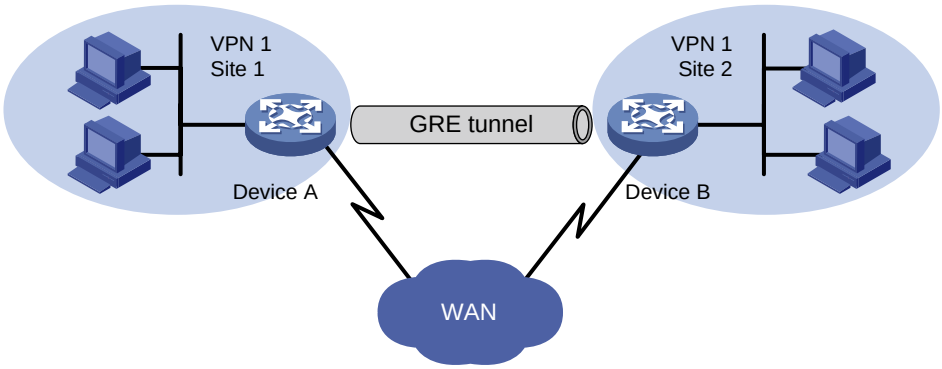
Figure 4 Network diagram



Constructing VPN

As shown in Figure 5, Site 1 and Site 2 both belong to VPN 1 and are located in different cities. Using a GRE tunnel can connect the two VPN sites across the WAN.

Figure 5 Network diagram



Protocols and standards

- RFC 1701, *Generic Routing Encapsulation (GRE)*
- RFC 1702, *Generic Routing Encapsulation over IPv4 networks*
- RFC 2784, *Generic Routing Encapsulation (GRE)*
- RFC 2890, *Key and Sequence Number Extensions to GRE*

Restrictions and guidelines: GRE configuration

When you configure a GRE tunnel, follow these restrictions and guidelines:

- You must configure the tunnel source address and destination address at both ends of a tunnel. The tunnel source or destination address at one end must be the tunnel destination or source address at the other end.

- As a best practice, do not configure the same tunnel source and destination addresses for local tunnel interfaces that use the same tunnel mode.
- To ensure correct packet forwarding, identify whether the destination network of packets and the IP address of the local tunnel interface are on the same subnet. If they are not, configure a route reaching the destination network through the tunnel interface. You can configure the route by using one of the following methods:
  - Configure a static route, using the local tunnel interface as the outgoing interface of the route.
  - Enable a dynamic routing protocol on both the tunnel interface and the interface connecting the private network. This allows the dynamic routing protocol to establish a routing entry with the tunnel interface as the outgoing interface.
- GRE encapsulation and de-encapsulation can decrease the forwarding efficiency of tunnel-end devices.

Configuring a GRE/IPv4 tunnel

Restrictions and guidelines

This task describes only GRE/IPv4 tunnel required tunnel interface commands (the **interface tunnel**, **source**, and **destination** commands). For more tunnel interface commands, see "Configuring tunneling."

Procedure

1. Enter system view.

   system-view

2. Create a GRE tunnel interface, and specify the tunnel mode as GRE/IPv4.

   interface tunnel *number* **mode gre**

   You must configure the same tunnel mode on both ends of a tunnel. Otherwise, packet delivery might fail.

3. Configure an IP address for the tunnel interface based on the passenger protocol.

   IPv4:

   For information about how to assign an IPv4 address to an interface, see "Configuring IP addressing."

   IPv6:

   For information about how to assign an IPv6 address to an interface, see "Configuring basic IPv6 settings."

   By default, no IP address is configured for a tunnel interface.

4. Configure a source address or source interface for the tunnel interface.

   source { *ip-address* | *interface-type interface-number* }

   By default, no source address or interface is configured for a tunnel interface.

   If you configure a source address for a tunnel interface, the tunnel interface uses the source address as the source address of the encapsulated packets.

   If you configure a source interface for a tunnel interface, the tunnel interface uses the primary IP address of the source interface as the source address of the encapsulated packets.

5. Configure a destination address for the tunnel interface.

   destination *ip-address*

   By default, no destination address is configured for a tunnel interface.

   The destination address is the address of the physical interface that the tunnel remote end uses to receive packets from the GRE tunnel.

The tunnel local end uses this address as the destination address of the encapsulated packets.

The tunnel destination address and the IP address of the tunnel interface must be in different subnets.

6. (Optional.) Enable GRE keepalive, and set the keepalive interval and keepalive number.

   keepalive [*interval* [*times*]]

   By default, GRE keepalive is disabled.

7. (Optional.) Configure GRE security mechanisms.
   - Configure a GRE key for the GRE tunnel interface.

     gre key *key*

     By default, no GRE key is configured for a GRE tunnel interface.

     The two ends of a GRE tunnel must have the same key or both have no key.

8. (Optional.) Set the DF bit for encapsulated packets.

   tunnel dfbit enable

   By default, the DF bit is not set, allowing encapsulated packets to be fragmented.

Configuring a GRE/IPv6 tunnel

Restrictions and guidelines

This task describes only GRE/IPv6 tunnel required tunnel interface commands (the **interface tunnel**, **source**, and **destination** commands). For more tunnel interface commands, see "Configuring tunneling."

Procedure

1. Enter system view.

   system-view

2. Create a GRE tunnel interface, and specify the tunnel mode as GRE/IPv6.

   interface tunnel *number* **mode gre ipv6**

   You must configure the same tunnel mode on both ends of a tunnel. Otherwise, packet delivery might fail.

3. Configure an IP address for the tunnel interface based on the passenger protocol.

   IPv4:

   For information about how to assign an IPv4 address to an interface, see "Configuring IP addressing."

   IPv6:

   For information about how to assign an IPv6 address to an interface, see "Configuring basic IPv6 settings."

   By default, no IP address is configured for a tunnel interface.

4. Configure a source IPv6 address or source interface for the tunnel interface.

   source { *ipv6-address* | *interface-type interface-number* }

   By default, no source IPv6 address or interface is configured for a tunnel interface.

   If you configure a source IPv6 address for a tunnel interface, the tunnel interface uses the source IPv6 address as the source IPv6 address of the encapsulated packets.

   If you configure a source interface for a tunnel interface, the tunnel interface uses the IPv6 address of the source interface as the source IPv6 address of the encapsulated packets.

5. Configure a destination IPv6 address for the tunnel interface.

   destination *ipv6-address*

By default, no destination IPv6 address is configured for a tunnel interface.

The destination IPv6 address is the IPv6 address of the physical interface that the tunnel remote end uses to receive packets from the GRE tunnel.

The tunnel local end uses this address as the destination IPv6 address of the encapsulated packets.

The tunnel destination address and the IP address of the tunnel interface must be in different subnets.

6. (Optional.) Configure GRE security mechanisms.
   - (Optional.) Configure a GRE key for the tunnel interface.

     gre key *key*

     By default, no GRE key is configured for a GRE tunnel interface.

     The two ends of a GRE tunnel must have the same key or both have no key.

Display and maintenance commands for GRE

Execute **display** commands in any view and **reset** commands in user view.

Task	Command	Remarks
Display information about tunnel interfaces.	**display interface** [**tunnel** [*number*]] [**brief** [**description** | **down**]]	For more information about the commands, see *Layer 3—IP Services Command Reference*.
Display IPv6 information about tunnel interface.	**display ipv6 interface** [**tunnel** [*number*]] [**brief**]	For more information about this command, see *Layer 3—IP Services Command Reference*.
Clear tunnel interface statistics.	**reset counters interface** [**tunnel** [*number*]]	For more information about this command, see *Layer 3—IP Services Command Reference*.
Clear IPv6 statistics on tunnel interfaces.	**reset ipv6 statistics** [**slot** *slot-number*]	For more information about this command, see IPv6 basics in *Layer 3—IP Services Command Reference*.

GRE configuration examples

Example: Configuring an IPv4 over IPv4 GRE tunnel

Network configuration

As shown in Figure 6, Group 1 and Group 2 are two private IPv4 networks. The two networks both use private network addresses on the network segment of 10.0.0.0/8 and belong to the same VPN. Establish a GRE tunnel between Switch A and Switch B to interconnect the two private IPv4 networks Group 1 and Group 2.

Figure 6 Network diagram

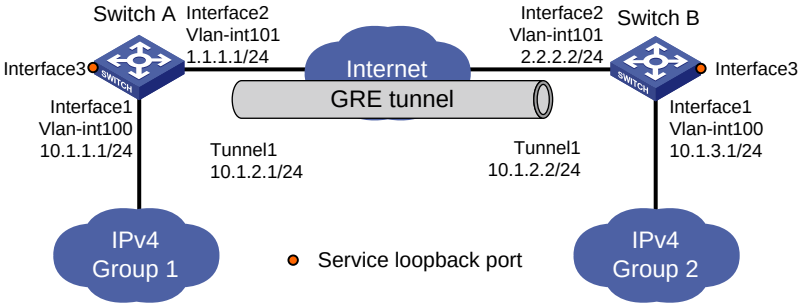


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Before performing the following configuration, configure an IP address for each interface, and make sure Switch A and Switch B can reach each other.

1. Configure Switch A:

Enter system view.

```
<SwitchA> system-view
```

Create a tunnel interface Tunnel 1, and specify the tunnel mode as GRE/IPv4.

```
[SwitchA] interface tunnel 1 mode gre
```

Configure an IP address for the tunnel interface.

```
[SwitchA-Tunnel1] ip address 10.1.2.1 255.255.255.0
```

Configure the source address of tunnel interface as the IP address of VLAN-interface 101 on Switch A.

```
[SwitchA-Tunnel1] source 1.1.1.1
```

Configure the destination address of the tunnel interface as the IP address of VLAN-interface 101 on Switch B.

```
[SwitchA-Tunnel1] destination 2.2.2.2
[SwitchA-Tunnel1] quit
```

Configure a static route from Switch A through the tunnel interface to Group 2.

```
[SwitchA] ip route-static 10.1.3.0 255.255.255.0 tunnel 1
```

2. Configure Switch B:

Enter system view.

```
<SwitchB> system-view
```

Create a tunnel interface Tunnel 1, and specify the tunnel mode as GRE/IPv4.

```
[SwitchB] interface tunnel 1 mode gre
```

Configure an IP address for the tunnel interface.

```
[SwitchB-Tunnel1] ip address 10.1.2.2 255.255.255.0
```

Configure the source address of tunnel interface as the IP address of VLAN-interface 101 on Switch B.

```
[SwitchB-Tunnel1] source 2.2.2.2
```

Configure the destination address of the tunnel interface as the IP address of VLAN-interface 101 on Switch A.

```
[SwitchB-Tunnel1] destination 1.1.1.1
[SwitchB-Tunnel1] quit
```

Configure a static route from Switch B through the tunnel interface to Group 1.

```
[SwitchB] ip route-static 10.1.1.0 255.255.255.0 Tunnel 1
```

Verifying the configuration

Display tunnel interface information on Switch A.

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64kbps
Maximum transmission unit: 1476
Internet address: 10.1.2.1/24 (primary)
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel keepalive disabled
Tunnel TTL 255
Tunnel protocol/transport GRE/IP
    GRE key disabled
    Checksumming of GRE packets disabled
Last clearing of counters: Never
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Display tunnel interface information on Switch B.

```
[SwitchB] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64kbps
Maximum transmission unit: 1476
Internet address: 10.1.2.2/24 (primary)
Tunnel source 2.2.2.2, destination 1.1.1.1
Tunnel keepalive disabled
Tunnel TTL 255
Tunnel protocol/transport GRE/IP
    GRE key disabled
    Checksumming of GRE packets disabled
Last clearing of counters: Never
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

From Switch B, ping the IP address of VLAN-interface 100 on Switch A.

```
[SwitchB] ping -a 10.1.3.1 10.1.1.1
Ping 10.1.1.1 (10.1.1.1) from 10.1.3.1: 56 data bytes, press CTRL+C to break
56 bytes from 10.1.1.1: icmp_seq=0 ttl=255 time=11.000 ms
56 bytes from 10.1.1.1: icmp_seq=1 ttl=255 time=1.000 ms
56 bytes from 10.1.1.1: icmp_seq=2 ttl=255 time=0.000 ms
56 bytes from 10.1.1.1: icmp_seq=3 ttl=255 time=0.000 ms
56 bytes from 10.1.1.1: icmp_seq=4 ttl=255 time=0.000 ms

--- Ping statistics for 10.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/2.400/11.000/4.317 ms
```

The output shows that Switch B can successfully ping Switch A.

Example: Configuring an IPv4 over IPv6 GRE tunnel

Network configuration

As shown in Figure 7, two IPv4 subnets Group 1 and Group 2 are connected to an IPv6 network. Create a GRE/IPv6 tunnel between Switch A and Switch B, so the two IPv4 subnets can communicate with each other through the GRE tunnel over the IPv6 network.

Figure 7 Network diagram

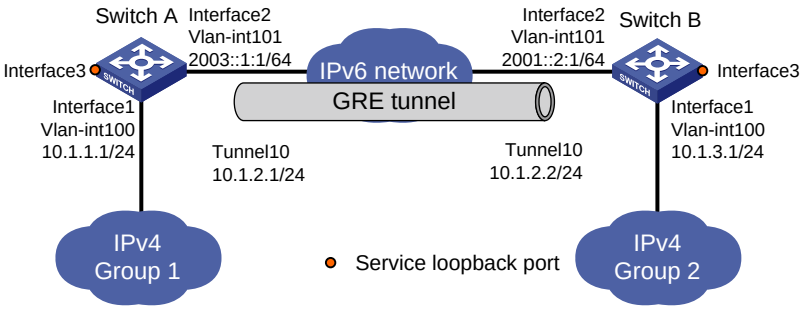


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Before performing the following configuration, configure an IP address for each interface, and make sure Switch A and Switch B can reach each other.

1. Configure Switch A:

   # Enter system view.

   ```
   <SwitchA> system-view
   ```

   # Create a tunnel interface Tunnel 10, and specify the tunnel mode as GRE/ IPv6.

   ```
   [SwitchA] interface tunnel 10 mode gre ipv6
   ```

   # Configure an IP address for the tunnel interface.

```
[SwitchA-Tunnel10] ip address 10.1.2.1 255.255.255.0
```

Configure the source address of the tunnel interface as the IPv6 address of VLAN-interface 101 on Switch A.

```
[SwitchA-Tunnel10] source 2003::1:1
```

Configure the destination address of the tunnel interface as the IPv6 address of VLAN-interface 101 on Switch B.

```
[SwitchA-Tunnel10] destination 2001::2:1
[SwitchA-Tunnel10] quit
```

Configure a static route from Switch A through the tunnel interface to Group 2.

```
[SwitchA] ip route-static 10.1.3.0 255.255.255.0 tunnel 10
```

2. Configure Switch B:

Enter system view.

```
<SwitchB> system-view
```

Create a tunnel interface Tunnel 10, and specify the tunnel mode as GRE/IPv6.

```
[SwitchB] interface tunnel 10 mode gre ipv6
```

Configure an IP address for the tunnel interface.

```
[SwitchB-Tunnel10] ip address 10.1.2.2 255.255.255.0
```

Configure the source address of tunnel interface as the IPv6 address of VLAN-interface 101 on Switch B.

```
[SwitchB-Tunnel10] source 2001::2:1
```

Configure the destination address of the tunnel interface as the IPv6 address of VLAN-interface 101 on Switch A.

```
[SwitchB-Tunnel10] destination 2003::1:1
[SwitchB-Tunnel10] quit
```

Configure a static route from Switch B through the tunnel interface to Group 1.

```
[SwitchB] ip route-static 10.1.1.0 255.255.255.0 tunnel 10
```

Verifying the configuration

Display tunnel interface information on Switch A.

```
[SwitchA] display interface tunnel 10
Tunnel10
Current state: UP
Line protocol state: UP
Description: Tunnel10 Interface
Bandwidth: 64kbps
Maximum transmission unit: 1456
Internet address: 10.1.2.1/24 (primary)
Tunnel source 2003::1:1, destination 2001::2:1
Tunnel TTL 255
Tunnel protocol/transport GRE/IPv6
    GRE key disabled
    Checksumming of GRE packets disabled
Last clearing of counters: Never
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Display tunnel interface information on Switch B.

```
[SwitchB] display interface tunnel 10
Tunnel10
Current state: UP
Line protocol state: UP
Description: Tunnel10 Interface
Bandwidth: 64kbps
Maximum transmission unit: 1456
Internet address: 10.1.2.2/24 (primary)
Tunnel source 2001::2:1, destination 2003::1:1
Tunnel TTL 255
Tunnel protocol/transport GRE/IPv6
    GRE key disabled
    Checksumming of GRE packets disabled
Last clearing of counters: Never
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

From Switch B, ping the IP address of VLAN-interface 100 on Switch A.

```
[SwitchB] ping -a 10.1.3.1 10.1.1.1
Ping 10.1.1.1 (10.1.1.1) from 10.1.3.1: 56 data bytes, press CTRL+C to break
56 bytes from 10.1.1.1: icmp_seq=0 ttl=255 time=2.000 ms
56 bytes from 10.1.1.1: icmp_seq=1 ttl=255 time=1.000 ms
56 bytes from 10.1.1.1: icmp_seq=2 ttl=255 time=1.000 ms
56 bytes from 10.1.1.1: icmp_seq=3 ttl=255 time=0.000 ms
56 bytes from 10.1.1.1: icmp_seq=4 ttl=255 time=1.000 ms

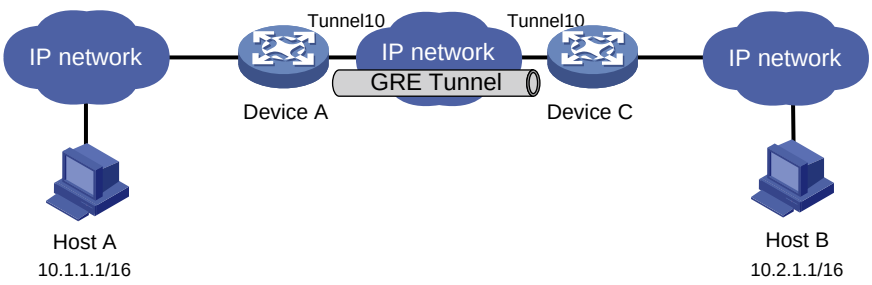
--- Ping statistics for 10.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/1.000/2.000/0.632 ms
```

The output shows that Switch B can successfully ping Switch A.

Troubleshooting GRE

The key to configuring GRE is to keep the configuration consistent. Most faults can be located by using the **debugging gre** or **debugging tunnel** command. This section analyzes one type of fault for illustration, with the scenario shown in Figure 8.

Figure 8 Network diagram



Hosts at both ends of a GRE tunnel cannot ping each other

Symptom

The interfaces at both ends of the tunnel are configured correctly and can ping each other, but Host A and Host B cannot ping each other.

Solution

To resolve the issue:

1. Execute the **`display ip routing-table`** command on Device A and Device C to view whether Device A has a route over tunnel 0 to 10.2.0.0/16 and whether Device C has a route over tunnel 0 to 10.1.0.0/16.
2. If such a route does not exist, execute the **`ip route-static`** command in system view to add the route. Take Device A as an example:

   ```
   [DeviceA] ip route-static 10.2.0.0 255.255.0.0 tunnel 0
   ```
3. If the issue persists, contact Technical Support.

Contents

Configuring HTTP redirect 1
About HTTP redirect 1
HTTP redirect tasks at a glance 1
Specifying the HTTPS redirect listening port number 1
Associating an SSL server policy with the HTTPS redirect service 2
Enabling the endpoint denylist feature for HTTP redirect 2
Display and maintenance commands for HTTP redirect 3

Configuring HTTP redirect

About HTTP redirect

HTTP redirect is a method to redirect users' HTTP or HTTPS requests to a specific URL.

It is used in the following features:

- Redirect URL assignment in 802.1X authentication, MAC authentication, and port security.
- EAD assistant URL redirection in 802.1X authentication.
- URL redirection services in portal.

HTTP redirect tasks at a glance

No configuration is required to redirect HTTP requests.

To redirect HTTPS requests, perform the following tasks:

1. Specifying the HTTPS redirect listening port number
2. (Optional.) Associating an SSL server policy with the HTTPS redirect service
3. (Optional.) Enabling the endpoint denylist feature for HTTP redirect

Specifying the HTTPS redirect listening port number

About this task

The device can redirect HTTPS requests only after you specify the TCP port number on which the HTTPS redirect service listens for HTTPS requests.

Restrictions and guidelines

To avoid service unavailability caused by port conflict, do not specify a TCP port number used by a well-known protocol or used by any other TCP-based service. To display TCP port numbers that have been used by services, use the **display tcp** command. For more information about this command, see IP performance optimization commands in *Layer 3—IP Services Command Reference*.

If you perform this task multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Specify the HTTPS redirect listening port number.
   http-redirect https-port *port-number*
   By default, the HTTPS redirect listening port number is 6654.

Associating an SSL server policy with the HTTPS redirect service

About this task

An SSL server policy is a set of SSL parameters used by the device when the device acts as the SSL server. You can configure parameters such as supported cipher suites and whether to perform digital certificate-based authentication on SSL clients for the SSL server policy.

You can use one of the following local certificates for HTTPS redirect service according to the security requirements and the configuration complexity:

- **Self-signed certificate**—Using this type of certificate is simple in configuration but has low security. You do not need to associate an SSL server policy with the HTTPS redirect service and the default SSL parameters are used. However, a self-signed certificate is not trusted by the browser. When the device redirects HTTPS requests to the specified URL, a certificate security warning prompt might appear on the browser. If you accept the security risks stated in the prompt, you can ignore the prompt to browse the page.
- **CA-signed certificate**—Using this type of certificate is complex in configuration but has high security. You must obtain a CA certificate, request a local certificate from the CA, create an SSL server policy, and associate the SSL server policy with the HTTPS redirect service.

For more information about digital certificates, see PKI in *Security Configuration Guide*. For more information about the SSL server policy configuration, see SSL in *Security Configuration Guide*.

Restrictions and guidelines

HTTPS redirect is unavailable if the associated SSL server policy does not exist. You can first associate a nonexistent SSL server policy with the HTTPS redirect service and then configure the SSL server policy.

If you change the SSL server policy associated with the HTTPS redirect service, the new policy takes effect immediately.

If you perform this task multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.

   `system-view`

2. Associate an SSL server policy with the HTTPS redirect service.

   `http-redirect ssl-server-policy` *`policy-name`*

   By default, no SSL server policy is associated with the HTTPS redirect service. The HTTPS redirect service uses the self-assigned certificate and the default SSL parameters.

Enabling the endpoint denylist feature for HTTP redirect

About this task

In portal authentication scenarios, the device redirects HTTP/HTTPS requests of users to the portal authentication page. When users frequently initiate HTTP/HTTPS requests and trigger redirection to the authentication page, the device processes many HTTP/HTTPS packets. This causes high CPU usage, which affects normal services. To avoid this issue, you can enable the endpoint denylist feature for HTTP redirect.

This feature enables the device to use the configured denylist parameters to uniquely identify endpoint entries by IP or MAC address and collect statistics on the received users' HTTP redirect

packets. When the device detects that the number of HTTP redirect packets sent by a user reaches the threshold for adding an endpoint denylsit entry during the statistics period, the device will add the user's IP or MAC address to the endpoint denylist entry.

After a user's IP or MAC address is added to an endpoint denylist entry, the device stops redirecting HTTP packets from that user until the entry expires or the administrator removes the entry.

Restrictions and guidelines

After you enable this feature, if you edit the endpoint denylist parameters again, the device clears the existing endpoint denylist entries and restarts collecting statistics on users' HTTP redirect packets.

Procedures

1. Enter system view.

   system-view

2. Enable the endpoint denylist feature for HTTP redirect.

   http-redirect endpoint-denylist enable [**packet** *packet-count*] [**period** *period*] [**aging-time** *aging-time*] [**ipbase** | **macbase**]

   By default, the endpoint denylist feature is disabled for HTTP redirect.

Display and maintenance commands for HTTP redirect

Execute **display** commands in any view and execute **reset** commands in user view.

Task	Command
Display endpoint denylist entries for HTTP redirect.	**display http-redirect endpoint-denylist** [**interface** *interface-type interface-number*] [**ip** *ipv4-address* | **ipv6** *ipv6-address* | **mac** *mac-address*] [**slot** *slot-number*]
Delete endpoint denylist entries for HTTP redirect.	**reset http-redirect endpoint-denylist** [**ip** *ipv4-address* | **ipv6** *ipv6-address* | **mac** *mac-address*]

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Layer 3—IP Routing Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the routing fundamentals and configuration procedures. It covers the mainstream routing protocols for IPv4 and IPv6 networks, and describes how to use policies to filter routes and affect routing decisions.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** >

	Folder.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring basic IP routing 1

About IP routing 1
Routing table 1
Route categories 1
Dynamic routing protocols 1
Route preference 2
Load sharing 2
Route backup 3
Route recursion 3
Route redistribution 3
Extension attribute redistribution 3
Setting the maximum lifetime for routes and labels in the RIB 3
Setting the maximum lifetime for routes in the FIB 4
Enabling the RIB to flush route attribute information to the FIB 5
Setting the maximum number of ECMP routes 5
Configuring the ECMP mode 5
Configuring RIB NSR 6
Configuring inter-protocol FRR 6
Configuring BFD for primary route next hop availability detection 7
About configuring BFD for primary route next hop detection 7
Restrictions and guidelines for configuring BFD for primary route next hop availability detection 8
Configuring BFD control packet mode 8
Configuring BFD echo packet mode 9
Enabling route fast switchover 10
Configuring routing policy-based recursive lookup 10
Setting the maximum number of active routes supported by the device 11
Configuring RIB SNMP notifications 12
Display and maintenance commands for basic IP routing 12

Configuring basic IP routing

This chapter focuses on unicast routing protocols. For more information about multicast routing protocols, see *IP Multicast Configuration Guide*.

About IP routing

IP routing directs IP packet forwarding on routers. Based on the destination IP address in the packet, a router looks up a route for the packet in a routing table and forwards the packet to the next hop. Routes are path information used to direct IP packets.

Routing table

A RIB contains the global routing information and related information, including route recursion, route redistribution, and route extension information. The router selects optimal routes from the routing table and puts them into the FIB table. It uses the FIB table to forward packets. For more information about the FIB table, see *Layer 3—IP Services Configuration Guide*.

Route categories

Table 1 categorizes routes by different criteria.

Table 1 Route categories

Criterion	Categories
Origin	• **Direct route**—A direct route is discovered by the data link protocol on an interface, and is also called an interface route. • **Static route**—A static route is manually configured by an administrator. • **Dynamic route**—A dynamic route is dynamically discovered by a routing protocol.
Destination	• **Network route**—The destination is a network. The subnet mask is less than 32 bits. • **Host route**—The destination is a host. The subnet mask is 32 bits.
Whether the destination is directly connected	• **Direct route**—The destination is directly connected. • **Indirect route**—The destination is indirectly connected.

Dynamic routing protocols

Static routes work well in small, stable networks. They are easy to configure and require fewer system resources. However, in networks where topology changes occur frequently, a typical practice is to configure a dynamic routing protocol. Compared with static routing, a dynamic routing protocol is complicated to configure, requires more router resources, and consumes more network resources.

Dynamic routing protocols dynamically collect and report reachability information to adapt to topology changes. They are suitable for large networks.

Dynamic routing protocols can be classified by different criteria, as shown in Table 2.

Table 2 Categories of dynamic routing protocols

Criterion	Categories
Operation scope	• **IGPs**—Work within an AS. Examples include RIP, OSPF, and IS-IS. • **EGPs**—Work between ASs. The most popular EGP is BGP.
Routing algorithm	• **Distance-vector protocols**—Examples include RIP and BGP. BGP is also considered a path-vector protocol. • **Link-state protocols**—Examples include OSPF and IS-IS.
Destination address type	• **Unicast routing protocols**—Examples include RIP, OSPF, BGP, and IS-IS. • **Multicast routing protocols**—Examples include PIM-SM and PIM-DM.
IP version	• **IPv4 routing protocols**—Examples include RIP, OSPF, BGP, and IS-IS. • **IPv6 routing protocols**—Examples include RIPng, OSPFv3, IPv6 BGP, and IPv6 IS-IS.

An AS refers to a group of routers that use the same routing policy and work under the same administration.

Route preference

Routing protocols, including static and direct routing, each by default have a preference. If they find multiple routes to the same destination, the router selects the route with the highest preference as the optimal route.

The preference of a direct route is always 0 and cannot be changed. You can configure a preference for each static route and each dynamic routing protocol. The following table lists the route types and default preferences. The smaller the value, the higher the preference.

Table 3 Route types and default route preferences

Route type	Preference
Direct route	0
Multicast static route	1
OSPF	10
IS-IS	15
Unicast static route	60
RIP	100
OSPF ASE	150
OSPF NSSA	150
IBGP	255
EBGP	255
Unknown (route from an untrusted source)	256

Load sharing

A routing protocol might find multiple optimal equal-cost routes to the same destination. You can use these routes to implement equal-cost multi-path (ECMP) load sharing.

Static routing, IPv6 static routing, RIP, RIPng, OSPF, OSPFv3, BGP, IPv6 BGP, IS-IS, and IPv6 IS-IS support ECMP load sharing.

Route backup

Route backup can improve network availability. Among multiple routes to the same destination, the route with the highest priority is the primary route and others are secondary routes.

The router forwards matching packets through the primary route. When the primary route fails, the route with the highest preference among the secondary routes is selected to forward packets. When the primary route recovers, the router uses it to forward packets.

Route recursion

To use a static, BGP, or RIP route that has an indirectly connected next hop, a router must perform route recursion to find the output interface to reach the next hop.

Link-state routing protocols, such as OSPF and IS-IS, do not need route recursion, because they obtain directly connected next hops through route calculation.

The RIB records and saves route recursion information, including brief information about related routes, recursive paths, and recursion depth.

Route redistribution

Route redistribution enables routing protocols to learn routing information from each other. A dynamic routing protocol can redistribute routes from other routing protocols, including direct and static routing. For more information, see the respective chapters on those routing protocols in this configuration guide.

The RIB records redistribution relationships of routing protocols.

Extension attribute redistribution

Extension attribute redistribution enables routing protocols to learn route extension attributes from each other, including BGP extended community attributes, OSPF area IDs, route types, and router IDs.

The RIB records extended attributes of each routing protocol and redistribution relationships of different routing protocol extended attributes.

Setting the maximum lifetime for routes and labels in the RIB

About this task

Perform this task to prevent routes and labels of a certain protocol from being aged out due to slow protocol convergence resulting from a large number of route entries or long GR period.

Restrictions and guidelines

The configuration takes effect at the next protocol or RIB process switchover.

Procedure (IPv4)

1. Enter system view.

   ```
   system-view
   ```

2. Enter RIB view.
   rib
3. Create the RIB IPv4 address family and enter its view.
   address-family ipv4
4. Set the maximum lifetime for IPv4 routes and labels in the RIB.
   protocol *protocol* [**instance** *instance-name*] **lifetime** *seconds*
   By default, the maximum lifetime for routes and labels in the RIB is 900 seconds.

Procedure (IPv6)

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv6 address family and enter its view.
   address-family ipv6
4. Set the maximum lifetime for IPv6 routes and labels in the RIB.
   protocol *protocol* [**instance** *instance-name*] **lifetime** *seconds*
   By default, the maximum lifetime for routes and labels in the RIB is 900 seconds.

Setting the maximum lifetime for routes in the FIB

About this task

When GR or NSR is disabled, FIB entries must be retained for some time after a protocol process switchover or RIB process switchover. When GR or NSR is enabled, FIB entries must be removed immediately after a protocol or RIB process switchover to avoid routing issues. Perform this task to meet such requirements.

Procedure (IPv4)

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv4 address family and enter its view.
   address-family ipv4
4. Set the maximum lifetime for IPv4 routes in the FIB.
   fib lifetime *seconds*
   By default, the maximum lifetime for routes in the FIB is 600 seconds.

Procedure (IPv6)

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv6 address family and enter its view.
   address-family ipv6
4. Set the maximum lifetime for IPv6 routes in the FIB.

fib lifetime *seconds*

By default, the maximum lifetime for routes in the FIB is 600 seconds.

Enabling the RIB to flush route attribute information to the FIB

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv4 address family and enter its view.
   address-family ipv4
4. Enable the RIB to flush route attribute information to the FIB.
   flush route-attribute *protocol*
   By default, the RIB does not flush route attribute information to the FIB.

Setting the maximum number of ECMP routes

Restrictions and guidelines

IMPORTANT:

This task is used only for version rollback in an IRF network environment and must be performed under the guidance of H3C technical support.

This configuration takes effect at reboot. Make sure the reboot does not impact your network.

Procedure

1. Enter system view.
   system-view
2. Set the maximum number of ECMP routes.
   max-ecmp-num *number*
   By default, the maximum number of ECMP routes is 64.

Configuring the ECMP mode

About this task

An ECMP route group contains ECMP routes to the same destination. The following ECMP modes are supported:

- **Normal mode**—When one or multiple ECMP routes in an ECMP group fail, this mode enables the device to reallocate all traffic to the remaining routes in the ECMP group.
- **Enhanced mode**—When one or multiple ECMP routes in an ECMP group fail, this mode enables the device to reallocate only the traffic of the failed routes to the remaining routes.

Restrictions and guidelines

The task applies to both IPv4 and IPv6 ECMP routes.

Procedure

1. Enter system view.
   system-view
2. Configure the ECMP mode.
   ecmp mode enhanced
   By default, the normal ECMP mode is used.

Configuring RIB NSR

About this task

When an active/standby switchover occurs, nonstop routing (NSR) backs up routing information from the active process to the standby process to avoid routing flapping and ensure forwarding continuity.

RIB NSR provides faster route convergence than protocol NSR during an active/standby switchover.

Restrictions and guidelines

Use this feature with protocol GR or NSR to avoid route timeouts and traffic interruption.

Procedure (IPv4)

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv4 address family and enter its view.
   address-family ipv4
4. Enable IPv4 RIB NSR.
   non-stop-routing
   By default, RIB NSR is disabled.

Procedure (IPv6)

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv6 address family and enter its view.
   address-family ipv6
4. Enable IPv6 RIB NSR.
   non-stop-routing
   By default, RIB NSR is disabled.

Configuring inter-protocol FRR

About this task

This feature is applicable to services that are sensitive to packet loss and latency. Inter-protocol fast reroute (FRR) enables fast rerouting between routes of different protocols. A backup next hop is

automatically selected to reduce the service interruption time caused by unreachable next hops. When the next hop of the primary link fails, the traffic is redirected to the backup next hop.

Among the routes to the same destination in the RIB, a router adds the route with the highest preference to the FIB table. After you configure this feature, the device will select a route of another routing protocol as the backup route. When the next hop of the primary route is unreachable, the device forwards packets through the backup route.

Restrictions and guidelines

This feature uses the next hop of a route from a different protocol as the backup next hop, which might cause loops.

If you configure both inter-protocol FRR and FRR for the specified routing protocol, FRR for the specified routing protocol takes effect.

Procedure (IPv4)

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv4 address family and enter its view.
   address-family ipv4
4. Enable IPv4 RIB inter-protocol FRR.
   inter-protocol fast-reroute [**vpn-instance** *vpn-instance-name*]
   By default, inter-protocol FRR is disabled.
   If you do not specify a VPN instance, inter-protocol FRR is enabled for the public network.

Procedure (IPv6)

1. Enter system view.
   system-view
2. Enter RIB view.
   rib
3. Create the RIB IPv6 address family and enter its view.
   address-family ipv6
4. Enable IPv6 RIB inter-protocol FRR.
   inter-protocol fast-reroute [**vpn-instance** *vpn-instance-name*]
   By default, inter-protocol FRR is disabled.
   If you do not specify a VPN instance, inter-protocol FRR is enabled for the public network.

Configuring BFD for primary route next hop availability detection

About configuring BFD for primary route next hop detection

This feature uses BFD to detect the availability of the primary route next hop for inter-protocol FRR or ECMP routes. When the primary route next hop unreachable, this feature enables the upper layer routing protocol to reallocate service traffic to the backup next hops or to the remaining ECMP next hops.

For inter-protocol FRR, the primary route is the route with the highest preference among the routes to the same destination. If BFD detects that the next hop of the primary route is unreachable, it notifies the FIB of the issue. The FIB will use the backup next hops to guide traffic forwarding.

For ECMP routes of a routing protocol, the primary route is each ECMP route. If BFD detects that the next hop of an ECMP route is unreachable, it notifies the routing protocol of the issue. The routing protocol will reallocate the service traffic to the remaining ECMP next hops. When BFD detects that the next hops of all ECMP routes are invalid, BFD notifies the routing protocol to withdraw the routes. In addition, the device reselects optimal routes for the service traffic.

For more information about BFD, see *High Availability Configuration Guide*.

Restrictions and guidelines for configuring BFD for primary route next hop availability detection

If you specify the **ecmp-shared** keyword when configuring FRR for a routing protocol, the device will use the LFA algorithm to calculate backup next hops for each ECMP route. The routes destined for the backup next hops are also added to the routing table as ECMP routes and the state of these routes is backup. BFD does not detect these backup ECMP routes.

The source IPv6 address of packets for an echo-packet-mode BFD session can only be a global unicast address. If the IPv6 next hop address of the primary link is not a global unicast address, as a best practice, use a control-packet-mode BFD session to detect next hop reachability for the primary link.

Configuring BFD control packet mode

About this task

This mode uses BFD control packets to detect the status of a link bidirectionally at a millisecond level.

Restrictions and guidelines for BFD control packet mode

A control-packet-mode BFD session can be successfully established through negotiation between the two ends. To use a control-packet-mode BFD session to detect the primary route next hop availability, you must create a static BFD session that meets the following requirements on the next hop device:

- Create a single-hop BFD session in control packet mode.
- The source IP address of the BFD session is the destination IP address of the BFD session automatically created by the local end.
- The destination IP address of the BFD session is the source IP address of the BFD session automatically created by the local end.
- The remote discriminator of the BFD session is the local discriminator of the BFD session automatically created by the local end.

The local device can use a control-packet-mode BFD session to detect the primary route next hop only after you create the static BFD session on the primary route next hop device.

Procedure

1. Enter system view.

   system-view
2. Enter RIB view.

   rib
3. Create RIB IPv4 or IPv6 address family view and enter the view.

   IPv4:

address-family ipv4

IPv6:

address-family ipv6

4. Enable BFD for primary link availability detection.

   primary-path-detect bfd ctrl [**inter-protocol-frr** | **protocol-ecmp** *protocol*]

   By default, BFD is disabled for primary link availability detection.

Configuring BFD echo packet mode

About this task

With BFD echo packet mode enabled, the output interface sends BFD echo packets to the destination device, and the device sends the packets back to test the link reachability.

Restrictions and guidelines

You do not need to configure BFD echo packet mode at the peer end.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Configure the source IP address of echo packets.

   IPv4:

   bfd echo-source-ip *ip-address*

   By default, the source IPv4 address of echo packets is not configured.

   As a best practice to avoid network congestion caused by excessive ICMP redirect packets from the peer, use this command. Make sure the source IPv4 address is not on the subnet of any interfaces on the device.

   For more information about this command, see BFD in *High Availability Command Reference*.

   IPv6:

   bfd echo-source-ipv6 *ipv6-address*

   By default, the source IPv6 address of echo packets is not configured.

   As a best practice to avoid network congestion caused by excessive ICMPv6 redirect packets from the peer, use this command. Make sure the source IPv6 address is not on the subnet of any interfaces on the device.

   You must specify a global unicast address as the source IPv6 address of BFD echo packets.

   For more information about this command, see BFD in *High Availability Command Reference*.

3. Enter RIB view.

   rib

4. Create RIB IPv4 or IPv6 address family view and enter the view.

   IPv4:

   address-family ipv4

   IPv6:

   address-family ipv6

5. Enable BFD for primary link availability detection.

   primary-path-detect bfd echo [**inter-protocol-frr** | **protocol-ecmp** *protocol*]

   By default, BFD is disabled for primary link availability detection.

Enabling route fast switchover

About this task

This feature applies to a device that provides the same physical output interface for large numbers of routes, including ECMP routes and primary/secondary routes. When a link failure occurs on the interface, the device must perform the following tasks before switching the traffic to another route:

1. Deletes all ARP or ND entries for the link.
2. Instructs the FIB to delete the associated FIB entries.

This procedure is time consuming and interrupts traffic for a long time. To resolve this problem, you can enable route fast switchover. This feature allows the device to instruct the FIB to delete the invalid FIB entries for route switchover first.

Procedure (IPv4)

1. Enter system view.
   `system-view`
2. Enable IPv4 route fast switchover.
   `ip route fast-switchover enable`
   By default, IPv4 route fast switchover is disabled.

Procedure (IPv6)

1. Enter system view.
   `system-view`
2. Enable IPv6 route fast switchover.
   `ipv6 route fast-switchover enable`
   By default, IPv6 route fast switchover is disabled.

Configuring routing policy-based recursive lookup

About this task

Configure routing policy-based recursive lookup to control route recursion results. For example, when a route changes, the routing protocol has to perform a route recursion if the next hop is indirectly connected. The routing protocol might select an incorrect path, which can cause traffic loss. To prevent this problem, you can use a routing policy to filter out incorrect routes. The routes that pass the filtering of the routing policy will be used for route recursion.

Restrictions and guidelines

The **`apply`** clauses in the specified routing policy cannot take effect.

Make sure a minimum of one related route can match the routing policy for correct traffic forwarding.

Procedure (IPv4)

1. Enter system view.
   `system-view`
2. Enter RIB view.
   `rib`
3. Create the RIB IPv4 address family and enter its view.
   `address-family ipv4`
4. Configure routing policy-based recursive lookup.

protocol *protocol* **nexthop recursive-lookup route-policy** *route-policy-name*

By default, routing policy-based recursive lookup is not configured.

Procedure (IPv6)

1. Enter system view.

   system-view
2. Enter RIB view.

   rib
3. Create the RIB IPv6 address family and enter its view.

   address-family ipv6
4. Configure routing policy-based recursive lookup.

   protocol *protocol* **nexthop recursive-lookup route-policy** *route-policy-name*

   By default, routing policy-based recursive lookup is not configured.

Setting the maximum number of active routes supported by the device

About this task

To perform this task:

- You can use the **routing-table limit** *number warn-threshold* command to specify an alarm threshold. When the percentage of active routes exceeds the alarm threshold, the system generates a system log message but still accepts active routes. You can take relevant actions based on the message to save system resources. If the number of active routes reaches the maximum number, no more routes can be added and new routes are discarded.
- You can use the **routing-table limit** *number* **simply-alert** command to specify an alarm threshold. When the maximum number of active IPv4/IPv6 routes is exceeded, the device still accepts new active routes but generates a system log message. You can take relevant actions based on the message to save system resources.

Procedure (IPv4)

1. Enter system view.

   system-view
2. Enter RIB view.

   rib
3. Create the RIB IPv4 address family and enter its view.

   address-family ipv4
4. Set the maximum number of active IPv4 routes supported by the device.

   routing-table limit *number* { *warn-threshold* | **simply-alert** }

   By default, the maximum number of active IPv4 routes is not set for the device.

   Configuration in RIB IPv4 address family view limits the number of active IPv4 routes for the public network and all VPN instances.

Procedure (IPv6)

1. Enter system view.

   system-view

2. Enter RIB view.

   rib

3. Create the RIB IPv6 address family and enter its view.

   address-family ipv6

4. Set the maximum number of active IPv6 routes supported by the device.

   routing-table limit *number* { *warn-threshold* | **simply-alert** }

   By default, the maximum number of active IPv6 routes is not set for the device.

   Configuration in RIB IPv6 address family view limits the number of active IPv6 routes for the public network and all VPN instances.

Configuring RIB SNMP notifications

About this task

To report critical RIB events to an NMS, enable SNMP notifications for RIB. For SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure (IPv4)

1. Enter system view.

   system-view

2. Enable IPv4 RIB SNMP notifications.

   snmp-agent trap enable rib [**prefix-exceed** | **prefix-exceed-clear** | **prefix-threshold-exceed** | **prefix-thresholdexceed-clear** | **public-prefix-exceed** | **public-prefixexceed-clear** | **public-prefixthreshold-exceed** | **public-prefixthresholdexceed-clear** | **vrf-prefix-exceed** | **vrf-prefixexceed-clear** | **vrf-prefixthreshold-exceed** | **vrf-prefixthresholdexceed-clear**] *

   By default, IPv4 RIB SNMP notifications are enabled.

Procedure (IPv6)

1. Enter system view.

   system-view

2. Enable IPv6 RIB SNMP notifications.

   snmp-agent trap enable ipv6 rib [**prefix-exceed** | **prefix-exceed-clear** | **prefix-threshold-exceed** | **prefix-thresholdexceed-clear** | **public-prefix-exceed** | **public-prefixexceed-clear** | **public-prefixthreshold-exceed** | **public-prefixthresholdexceed-clear** | **vrf-prefix-exceed** | **vrf-prefixexceed-clear** | **vrf-prefixthreshold-exceed** | **vrf-prefixthresholdexceed-clear**] *

   By default, IPv6 RIB SNMP notifications are enabled.

Display and maintenance commands for basic IP routing

Execute **display** commands in any view and **reset** commands in user view.

IMPORTANT:

The **display max-ecmp-num** command is used only for version rollback in an IRF network environment and must be used under the guidance of H3C technical support.

Task	Command
Display the ECMP mode.	**display ecmp mode**
Display routing table information.	**display ip routing-table** [**all-vpn-instance** | **vpn-instance** *vpn-instance-name*] [**verbose**] **display ip routing-table all-routes**
Display information about routes permitted by an IPv4 basic ACL.	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] **acl** *ipv4-acl-number* [**verbose**]
Display information about routes to a specific destination address.	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] *ip-address* [*mask-length* | *mask*] [**longer-match**] [**verbose**]
Display information about routes to a range of destination addresses.	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] *ip-address1* **to** *ip-address2* [**verbose**]
Display information about routes permitted by an IP prefix list.	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] **prefix-list** *prefix-list-name* [**verbose**]
Display information about routes installed by a protocol.	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] **protocol** *protocol* [**inactive** | **verbose**]
Display IPv4 route statistics.	**display ip routing-table** [**all-routes** | **all-vpn-instance** | **vpn-instance** *vpn-instance-name*] **statistics**
Display brief IPv4 routing table information.	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] **summary**
Display route attribute information in the IPv6 RIB.	**display ipv6 rib attribute** [*attribute-id*]
Display IPv6 RIB GR state information.	**display ipv6 rib graceful-restart**
Display next hop information in the IPv6 RIB.	**display ipv6 rib nib** [**self-originated**] [*nib-id*] [**verbose**] **display ipv6 rib nib protocol** *protocol* [**verbose**]
Display next hop information for IPv6 direct routes.	**display ipv6 route-direct nib** [*nib-id*] [**verbose**]
Display IPv6 routing table information.	**display ipv6 routing-table** [**all-vpn-instance** | **vpn-instance** *vpn-instance-name*] [**verbose**] **display ipv6 routing-table all-routes**
Display information about routes permitted by an IPv6 basic ACL.	**display ipv6 routing-table** [**vpn-instance** *vpn-instance-name*] **acl** *ipv6-acl-number* [**verbose**]

Task	Command
Display information about routes to an IPv6 destination address.	**display ipv6 routing-table** [**vpn-instance** *vpn-instance-name*] *ipv6-address* [*prefix-length*] [**longer-match**] [**verbose**]
Display information about routes to a range of IPv6 destination addresses.	**display ipv6 routing-table** [**vpn-instance** *vpn-instance-name*] *ipv6-address1* **to** *ipv6-address2* [**verbose**]
Display information about routes permitted by an IPv6 prefix list.	**display ipv6 routing-table** [**vpn-instance** *vpn-instance-name*] **prefix-list** *prefix-list-name* [**verbose**]
Display information about routes installed by an IPv6 protocol.	**display ipv6 routing-table** [**vpn-instance** *vpn-instance-name*] **protocol** *protocol* [**inactive** | **verbose**]
Display IPv6 route statistics.	**display ipv6 routing-table** [**all-routes** | **all-vpn-instance** | **vpn-instance** *vpn-instance-name*] **statistics**
Display brief IPv6 routing table information.	**display ipv6 routing-table** [**vpn-instance** *vpn-instance-name*] **summary**
Display the maximum number of ECMP routes.	**display max-ecmp-num**
Display route attribute information in the RIB.	**display rib attribute** [*attribute-id*]
Display RIB GR state information.	**display rib graceful-restart**
Display next hop information in the RIB.	**display rib nib** [**self-originated**] [*nib-id*] [**verbose**] **display rib nib protocol** *protocol* [**verbose**]
Display next hop information for direct routes.	**display route-direct nib** [*nib-id*] [**verbose**]
Clear IPv4 route statistics.	**reset ip routing-table statistics protocol** [**vpn-instance** *vpn-instance-name*] { *protocol* | **all** } **reset ip routing-table** [**all-routes** | **all-vpn-instance**] **statistics protocol** { *protocol* | **all** }
Clear IPv6 route statistics.	**reset ipv6 routing-table statistics protocol** [**vpn-instance** *vpn-instance-name*] { *protocol* | **all** } **reset ipv6 routing-table** [**all-routes** | **all-vpn-instance**] **statistics protocol** { *protocol* | **all** }

Contents

Configuring static routing 1
About static routes 1
Configuring a static route 1
Configuring a static route group 2
Configuring floating static routes 3
Deleting static routes 3
Configuring BFD for static routes 4
About BFD 4
Configuring BFD control packet mode 4
Configuring BFD echo packet mode 5
Configuring static route FRR 6
About static route FRR 6
Restrictions and guidelines for static route FRR 6
Configuring static route FRR by specifying a backup next hop 6
Configuring static route FRR to automatically select a backup next hop 7
Enabling BFD echo packet mode for static route FRR 7
Display and maintenance commands for static routing 7
Static route configuration examples 8
Example: Configuring basic static routes 8
Example: Configuring BFD for static routes (direct next hop) 10
Example: Configuring BFD for static routes (indirect next hop) 12
Example: Configuring static route FRR 14
Configuring a default route 18

Configuring static routing

About static routes

Static routes are manually configured. If a network's topology is simple, you only need to configure static routes for the network to work correctly.

Static routes cannot adapt to network topology changes. If a fault or a topological change occurs in the network, the network administrator must modify the static routes manually.

Configuring a static route

1. Enter system view.

   system-view

2. Configure a static route.

   Public network:

   ip route-static *dest-address* { *mask-length* | *mask* } *interface-type interface-number* [*next-hop-address* [**nexthop-index** *index-string*]] [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ip route-static *dest-address* { *mask-length* | *mask* } { *next-hop-address* [**nexthop-index** *index-string*] [**recursive-lookup** { **host-route** | **longest-match** }] | **vpn-instance** *d-vpn-instance-name next-hop-address* [**nexthop-index** *index-string*] [**recursive-lookup** { **host-route** | **longest-match** }] } [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ip route-static *dest-address* { *mask-length* | *mask* } **vpn-instance** *d-vpn-instance-name* [**track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   ip route-static *dest-address* { *mask-length* | *mask* } *interface-type interface-number next-hop-address* [**nexthop-index** *index-string*] **track-arp** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, no static route is configured.

   You can associate Track with a static route to monitor the reachability of the next hops. For more information about Track, see *High Availability Configuration Guide*.

   The **recursive-lookup host-route**, **recursive-lookup longest-match**, and **recursive-lookup** keywords are mutually exclusive. You cannot specify the keywords at the same time.

3. (Optional.) Enable periodic sending of ARP requests to the next hops of static routes.

   ip route-static arp-request interval *interval*

   By default, the device does not send ARP requests to the next hops of static routes.

4. (Optional.) Configure the default preference for static routes.

   ip route-static default-preference *default-preference*

   The default setting is 60.

Configuring a static route group

About this task

This task allows you to batch create static routes with different prefixes but the same output interface and next hop.

You can create a static route group, and specify the static group in the **ip route-static** command. All prefixes in the static route group will be assigned the next hop and output interface specified in the **ip route-static** command.

Procedure

1. Enter system view.

   system-view

2. Create a static route group and enter its view.

   ip route-static-group *group-name*

   By default, no static route group is configured.

3. Add a static route prefix to the static route group.

   prefix *dest-address* { *mask-length* | *mask* }

   By default, no static route prefix is added to the static route group.

4. Return to system view.

   quit

5. Configure a static route.

   Public network:

   ip route-static group *group-name interface-type interface-number* [*next-hop-address*] [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ip route-static group *group-name* { *next-hop-address* [**recursive-lookup** { **host-route** | **longest-match** }] | **vpn-instance** *d-vpn-instance-name next-hop-address* [**recursive-lookup** { **host-route** | **longest-match** }] } [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ip route-static group *group-name* **vpn-instance** *d-vpn-instance-name* [**track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   ip route-static group *group-name interface-type interface-number next-hop-address* **track-arp** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   The **recursive-lookup host-route**, **recursive-lookup longest-match**, and **recursive-lookup** keywords are mutually exclusive. You cannot specify the keywords at the same time.

   VPN:

   ip route-static vpn-instance *s-vpn-instance-name* **group** *group-name interface-type interface-number* [*next-hop-address*] [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ip route-static vpn-instance *s-vpn-instance-name* **group** *group-name* { *next-hop-address* [**recursive-lookup** { **host-route** | **longest-match** }] [**public**] | **vpn-instance** *d-vpn-instance-name next-hop-address*

[**recursive-lookup** { **host-route** | **longest-match** }] } [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

ip route-static vpn-instance *s-vpn-instance-name* **group** *group-name* { **public** | **vpn-instance** *d-vpn-instance-name* } [**track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

ip route-static vpn-instance *s-vpn-instance-name* **group** *group-name interface-type interface-number next-hop-address* **track-arp** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

The **recursive-lookup host-route**, **recursive-lookup longest-match**, and **recursive-lookup** keywords are mutually exclusive. You cannot specify the keywords at the same time.

If you specify a destination VPN instance without specifying a next hop address, the destination VPN instance cannot be the same as the source VPN instance.

By default, no static route is configured.

Configuring floating static routes

Perform this task to configure route backup to improve network reliability.

When a static or dynamic route to a destination address already exists on the device, you configure another static route with a lower priority as the backup route to improve the network reliability. This backup static route is called a floating static route and is activated only when the primary route fails. After the primary route recovers from failure, the floating static route becomes inactive and data forwarding switches back to the primary route.

A floating static route can be configured in either of the following ways

- Configure different priorities for multiple static routes to the same destination address. The route with the lower priority automatically becomes the floating static route.
- When a route to a destination address already exists on the device, configure a static route with a lower priority to the same destination address.

When you configure a floating static route, the priority value of the route must be larger than then priority value of the primary route. For more information, see "Configuring a static route."

Deleting static routes

About this task

To delete a static route, use the **undo ip route-static** command. To delete all static routes including the default route, use the **delete static-routes all** command.

Procedure

1. Enter system view.

   system-view
2. Delete all static routes.

   Public network:

   delete static-routes all

   VPN:

   delete vpn-instance *vpn-instance-name* **static-routes all**

CAUTION:

Deleting all static routes with caution. This operation might cause network connectivity failure and packet forwarding failure.

Configuring BFD for static routes

IMPORTANT:

Enabling BFD for a flapping route could worsen the situation.

About BFD

BFD provides a general-purpose, standard, medium-, and protocol-independent fast failure detection mechanism. It can uniformly and quickly detect the failures of the bidirectional forwarding paths between two routers for protocols, such as routing protocols.

For more information about BFD, see *High Availability Configuration Guide*.

Configuring BFD control packet mode

About this task

This mode uses BFD control packets to detect the status of a link bidirectionally at a millisecond level.

BFD control packet mode can be applied to static routes with a direct next hop or with an indirect next hop.

Restrictions and guidelines for BFD control packet mode

If you use BFD control packet mode at the local end, you must use this mode also at the peer end.

Configuring BFD control packet mode for a static route (direct next hop)

1. Enter system view.

   system-view

2. Configure BFD control packet mode for a static route.

   Public network:

   ip route-static *dest-address* { *mask-length* | *mask* } *interface-type interface-number next-hop-address* **bfd control-packet** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   VPN:

   ip route-static vpn-instance *s-vpn-instance-name dest-address* { *mask-length* | *mask* } *interface-type interface-number next-hop-address* **bfd control-packet** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, BFD control packet mode for a static route is not configured.

Configuring BFD control packet mode for a static route (indirect next hop)

1. Enter system view.

   system-view

2. Configure BFD control packet mode for a static route.

   Public network:

   ip route-static *dest-address* { *mask-length* | *mask* } { *next-hop-address* **bfd control-packet bfd-source** *ip-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* **bfd control-packet bfd-source** *ip-address* } [**preference** *preference*] [**tag tag-value**] [**description** *text*]

   VPN:

   ip route-static vpn-instance *s-vpn-instance-name dest-address* { *mask-length* | *mask* } { *next-hop-address* **bfd control-packet bfd-source** *ip-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* **bfd control-packet bfd-source** *ip-address* } [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, BFD control packet mode for a static route is not configured.

Configuring BFD echo packet mode

About this task

With BFD echo packet mode enabled for a static route, the output interface sends BFD echo packets to the destination device, which loops the packets back to test the link reachability.

Restrictions and guidelines

You do not need to configure BFD echo packet mode at the peer end.

Do not use BFD for a static route with the output interface in spoofing state.

Procedure

1. Enter system view.

   system-view

2. Configure the source address of echo packets.

   bfd echo-source-ip *ip-address*

   By default, the source address of echo packets is not configured.

   As a best practice to avoid network congestion caused by excessive ICMP redirect packets from the peer, use this command. Make sure the source IPv4 address is not on the subnet of any interfaces on the device.

   For more information about this command, see *High Availability Command Reference*.

3. Configure BFD echo packet mode for a static route.

   Public network:

   ip route-static *dest-address* { *mask-length* | *mask* } *interface-type interface-number next-hop-address* **bfd echo-packet** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   VPN:

   ip route-static vpn-instance *s-vpn-instance-name dest-address* { *mask-length* | *mask* } *interface-type interface-number next-hop-address* **bfd echo-packet** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

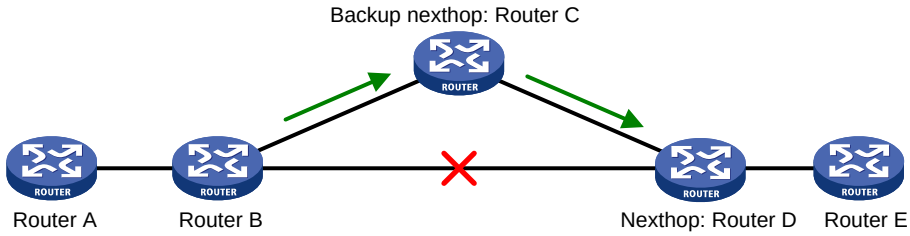
   By default, BFD echo packet mode for a static route is not configured.

Configuring static route FRR

About static route FRR

A link or router failure on a path can cause packet loss. Static route fast reroute (FRR) enables fast rerouting to minimize the impact of link or node failures.

Figure 1 Network diagram



As shown in Figure 1, upon a link failure, packets are directed to the backup next hop to avoid traffic interruption. You can either specify a backup next hop for FRR or enable FRR to automatically select a backup next hop (which must be configured in advance).

Restrictions and guidelines for static route FRR

- Do not use static route FRR and BFD (for a static route) at the same time.
- Equal-cost routes do not support static route FRR.
- Besides the configured static route for FRR, the device must have another route to reach the destination.

  When the state of the primary link (with Layer 3 interfaces staying up) changes from bidirectional to unidirectional or down, static route FRR quickly redirects traffic to the backup next hop. When the Layer 3 interfaces of the primary link are down, static route FRR temporarily redirects traffic to the backup next hop. In addition, the device searches for another route to reach the destination and redirects traffic to the new path if a route is found. If no route is found, traffic interruption occurs.

Configuring static route FRR by specifying a backup next hop

Restrictions and guidelines

A static route does not take effect when the backup output interface is unavailable.

To change the backup output interface or next hop, you must first remove the current setting. The backup output interface and next hop must be different from the primary output interface and next hop.

Procedure

1. Enter system view.

   system-view

2. Configure static route FRR.

   Public network:

   ip route-static *dest-address* { *mask-length* | *mask* } *interface-type interface-number* [*next-hop-address* [**backup-interface** *interface-type interface-number* [**backup-nexthop** *backup-nexthop-address*]]]

[**permanent**] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

VPN:

ip route-static vpn-instance *s-vpn-instance-name dest-address* { *mask-length* | *mask* } *interface-type interface-number* [*next-hop-address* [**backup-interface** *interface-type interface-number* [**backup-nexthop** *backup-nexthop-address*]]] [**permanent**] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

By default, static route FRR is disabled.

Configuring static route FRR to automatically select a backup next hop

1. Enter system view.

   system-view

2. Configure static route FRR to automatically select a backup next hop.

   ip route-static fast-reroute auto

   By default, static route FRR is disabled from automatically selecting a backup next hop.

Enabling BFD echo packet mode for static route FRR

About this task

By default, static route FRR uses ARP to detect primary link failures. Perform this task to enable static route FRR to use BFD echo packet mode for fast failure detection on the primary link.

Procedure

1. Enter system view.

   system-view

2. Configure the source IP address of BFD echo packets.

   bfd echo-source-ip *ip-address*

   By default, the source IP address of BFD echo packets is not configured.

   As a best practice to avoid network congestion caused by excessive ICMP redirect packets from the peer, use this command. Make sure the source IPv4 address is not on the subnet of any interfaces on the device.

   For more information about this command, see *High Availability Command Reference*.

3. Enable BFD echo packet mode for static route FRR.

   ip route-static primary-path-detect bfd echo

   By default, BFD echo packet mode for static route FRR is disabled.

Display and maintenance commands for static routing

Execute **display** commands in any view.

Task	Command
Display static route information.	**display ip routing-table protocol static**

Task	Command
	[**inactive** | **verbose**]
Display static route next hop information.	**display route-static nib** [*nib-id*] [**verbose**]
Display static routing table information.	**display route-static routing-table** [**vpn-instance** *vpn-instance-name*] [*ip-address* { *mask-length* | *mask* }]

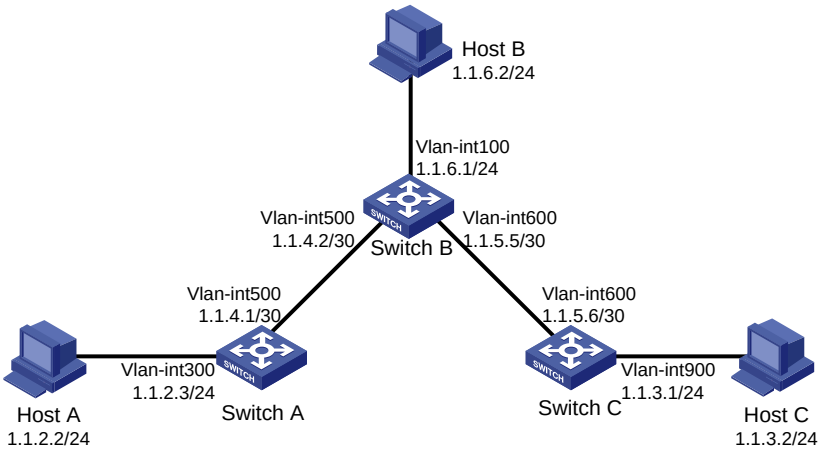
Static route configuration examples

Example: Configuring basic static routes

Network configuration

As shown in Figure 2, configure static routes on the switches for interconnections between any two hosts.

Figure 2 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure static routes:

   # Configure a default route on Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] ip route-static 0.0.0.0 0.0.0.0 1.1.4.2
   ```

   # Configure two static routes on Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] ip route-static 1.1.2.0 255.255.255.0 1.1.4.1
   [SwitchB] ip route-static 1.1.3.0 255.255.255.0 1.1.5.6
   ```

   # Configure a default route on Switch C.

   ```
   <SwitchC> system-view
   [SwitchC] ip route-static 0.0.0.0 0.0.0.0 1.1.5.5
   ```

3. Configure the default gateways of Host A, Host B, and Host C as 1.1.2.3, 1.1.6.1, and 1.1.3.1. (Details not shown.)

Verifying the configuration

Display static routes on Switch A.

```
[SwitchA] display ip routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/0           Static 60   0            1.1.4.2         Vlan500

Static Routing table status : <Inactive>
Summary count : 0
```

Display static routes on Switch B.

```
[SwitchB] display ip routing-table protocol static

Summary count : 2

Static Routing table status : <Active>
Summary count : 2

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
1.1.2.0/24          Static 60   0            1.1.4.1         Vlan500

Static Routing table status : <Inactive>
Summary count : 0
```

Use the **`ping`** command on Host B to test the reachability of Host A (Windows XP runs on the two hosts).

```
C:\Documents and Settings\Administrator>ping 1.1.2.2

Pinging 1.1.2.2 with 32 bytes of data:

Reply from 1.1.2.2: bytes=32 time=1ms TTL=126
Reply from 1.1.2.2: bytes=32 time=1ms TTL=126
Reply from 1.1.2.2: bytes=32 time=1ms TTL=126
Reply from 1.1.2.2: bytes=32 time=1ms TTL=126

Ping statistics for 1.1.2.2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 1ms, Maximum = 1ms, Average = 1ms
```

Use the **`tracert`** command on Host B to test the reachability of Host A.

```
C:\Documents and Settings\Administrator>tracert 1.1.2.2

Tracing route to 1.1.2.2 over a maximum of 30 hops
```

```
  1    <1 ms    <1 ms    <1 ms  1.1.6.1
  2    <1 ms    <1 ms    <1 ms  1.1.4.1
  3     1 ms    <1 ms    <1 ms  1.1.2.2

Trace complete.
```

Example: Configuring BFD for static routes (direct next hop)

Network configuration

Configure the following, as shown in Figure 3:

- Configure a static route to subnet 120.1.1.0/24 on Switch A.
- Configure a static route to subnet 121.1.1.0/24 on Switch B.
- Enable BFD for both routes.
- Configure a static route to subnet 120.1.1.0/24 and a static route to subnet 121.1.1.0/24 on Switch C.

When the link between Switch A and Switch B through the Layer 2 switch fails, BFD can detect the failure immediately. Switch A then communicates with Switch B through Switch C.

Figure 3 Network diagram

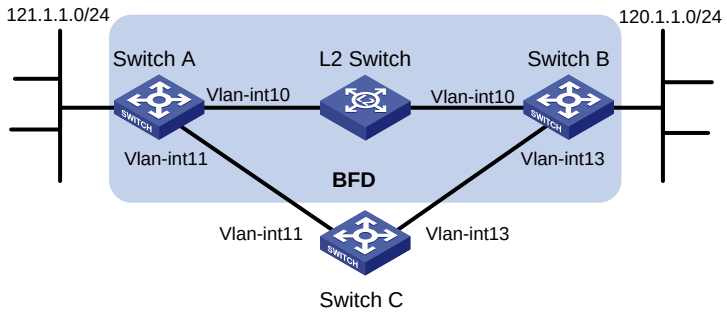


Table 1 Interface and IP address assignment

Device	Interface	IP address
Switch A	VLAN-interface 10	12.1.1.1/24
Switch A	VLAN-interface 11	10.1.1.102/24
Switch B	VLAN-interface 10	12.1.1.2/24
Switch B	VLAN-interface 13	13.1.1.1/24
Switch C	VLAN-interface 11	10.1.1.100/24
Switch C	VLAN-interface 13	13.1.1.2/24

Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Configure static routes and BFD:

   # Configure static routes on Switch A and enable BFD control packet mode for the static route that traverses the Layer 2 switch.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 10
   ```

```
[SwitchA-vlan-interface10] bfd min-control-interval 500
[SwitchA-vlan-interface10] bfd detect-multiplier 9
[SwitchA-vlan-interface10] quit
[SwitchA] ip route-static 120.1.1.0 24 vlan-interface 10 12.1.1.2 bfd control-packet
[SwitchA] ip route-static 120.1.1.0 24 vlan-interface 11 10.1.1.100 preference 65
[SwitchA] quit
```

Configure static routes on Switch B and enable BFD control packet mode for the static route that traverses the Layer 2 switch.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 10
[SwitchB-vlan-interface10] bfd min-control-interval 500
[SwitchB-vlan-interface10] bfd detect-multiplier 9
[SwitchB-vlan-interface10] quit
[SwitchB] ip route-static 121.1.1.0 24 vlan-interface 10 12.1.1.1 bfd control-packet
[SwitchB] ip route-static 121.1.1.0 24 vlan-interface 13 13.1.1.2 preference 65
[SwitchB] quit
```

Configure static routes on Switch C.

```
<SwitchC> system-view
[SwitchC] ip route-static 120.1.1.0 24 13.1.1.1
[SwitchC] ip route-static 121.1.1.0 24 10.1.1.102
```

Verifying the configuration

Display BFD sessions on Switch A.

```
<SwitchA> display bfd session
 Total sessions: 1        Up sessions: 1        Init mode: Active

 IPv4 session working in control packet mode:

 LD/RD            SourceAddr      DestAddr        State  Holdtime    Interface
 4/7              12.1.1.1        12.1.1.2        Up     2000ms      Vlan10
```

The output shows that the BFD session has been created.

Display the static routes on Switch A.

```
<SwitchA> display ip routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
120.1.1.0/24        Static 60   0            12.1.1.2        Vlan10

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 10. Then the link over VLAN-interface 10 fails.

Display static routes on Switch A.

```
<SwitchA> display ip routing-table protocol static
```

```
Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
120.1.1.0/24        Static 65   0            10.1.1.100      Vlan11

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 11.

Example: Configuring BFD for static routes (indirect next hop)

Network configuration

Figure 4 shows the network topology as follows:

- Switch A has a route to interface Loopback 1 (2.2.2.9/32) on Switch B, with the output interface VLAN-interface 10.
- Switch B has a route to interface Loopback 1 (1.1.1.9/32) on Switch A, with the output interface VLAN-interface 12.
- Switch D has a route to 1.1.1.9/32, with the output interface VLAN-interface 10, and a route to 2.2.2.9/32, with the output interface VLAN-interface 12.

Configure the following:

- Configure a static route to subnet 120.1.1.0/24 on Switch A.
- Configure a static route to subnet 121.1.1.0/24 on Switch B.
- Enable BFD for both routes.
- Configure a static route to subnet 120.1.1.0/24 and a static route to subnet 121.1.1.0/24 on both Switch C and Switch D.

When the link between Switch A and Switch B through Switch D fails, BFD can detect the failure immediately. Switch A then communicates with Switch B through Switch C.

Figure 4 Network diagram

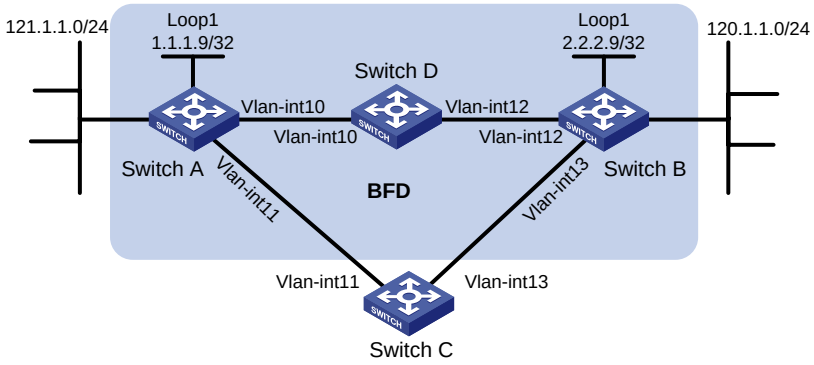


Table 2 Interface and IP address assignment

Device	Interface	IP address
Switch A	VLAN-interface 10	12.1.1.1/24
Switch A	VLAN-interface 11	10.1.1.102/24
Switch A	Loopback 1	1.1.1.9/32
Switch B	VLAN-interface 12	11.1.1.1/24
Switch B	VLAN-interface 13	13.1.1.1/24
Switch B	Loopback 1	2.2.2.9/32
Switch C	VLAN-interface 11	10.1.1.100/24
Switch C	VLAN-interface 13	13.1.1.2/24
Switch D	VLAN-interface 10	12.1.1.2/24
Switch D	VLAN-interface 12	11.1.1.2/24

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure static routes and BFD:

   # Configure static routes on Switch A and enable BFD control packet mode for the static route that traverses Switch D.

```
<SwitchA> system-view
[SwitchA] bfd multi-hop min-control-interval 500
[SwitchA] bfd multi-hop detect-multiplier 9
[SwitchA] ip route-static 2.2.2.9 32 12.1.1.2
[SwitchA] ip route-static 120.1.1.0 24 2.2.2.9 bfd control-packet bfd-source 1.1.1.9
[SwitchA] ip route-static 120.1.1.0 24 vlan-interface 11 10.1.1.100 preference 65
[SwitchA] quit
```

   # Configure static routes on Switch B and enable BFD control packet mode for the static route that traverses Switch D.

```
<SwitchB> system-view
[SwitchB] bfd multi-hop min-control-interval 500
[SwitchB] bfd multi-hop detect-multiplier 9
[SwitchB] ip route-static 1.1.1.9 32 11.1.1.2
[SwitchB] ip route-static 121.1.1.0 24 1.1.1.9 bfd control-packet bfd-source 2.2.2.9
[SwitchB] ip route-static 121.1.1.0 24 vlan-interface 13 13.1.1.2 preference 65
[SwitchB] quit
```

   # Configure static routes on Switch C.

```
<SwitchC> system-view
[SwitchC] ip route-static 120.1.1.0 24 13.1.1.1
[SwitchC] ip route-static 121.1.1.0 24 10.1.1.102
```

   # Configure static routes on Switch D.

```
<SwitchD> system-view
[SwitchD] ip route-static 120.1.1.0 24 11.1.1.1
[SwitchD] ip route-static 121.1.1.0 24 12.1.1.1
```

Verifying the configuration

Display BFD sessions on Switch A.

```
<SwitchA> display bfd session
 Total sessions: 1        Up sessions: 1        Init mode: Active

 IPv4 session working in control packet mode:

 LD/RD            SourceAddr      DestAddr        State  Holdtime    Interface
 4/7              1.1.1.9         2.2.2.9         Up     2000ms      N/A
```

The output shows that the BFD session has been created.

Display the static routes on Switch A.

```
<SwitchA> display ip routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
120.1.1.0/24        Static 60   0            12.1.1.2        Vlan10

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 10. Then the link over VLAN-interface 10 fails.

Display static routes on Switch A.

```
<SwitchA> display ip routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
120.1.1.0/24        Static 65   0            10.1.1.100      Vlan11

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 11.

Example: Configuring static route FRR

Network configuration

As shown in Figure 5, configure static routes on Switch A, Switch B, and Switch C, and configure static route FRR. When Link A becomes unidirectional, traffic can be switched to Link B immediately.

Figure 5 Network diagram

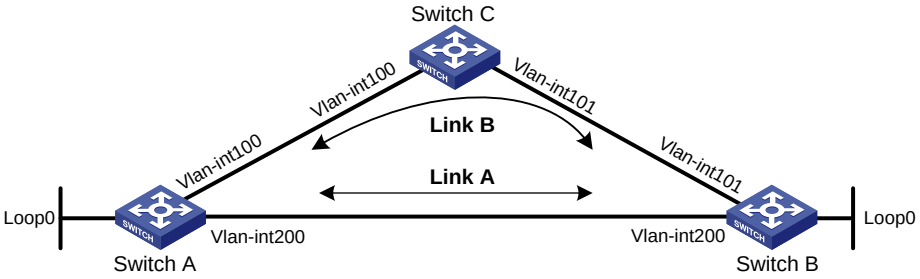


Table 3 Interface and IP address assignment

Device	Interface	IP address
Switch A	VLAN-interface 100	12.12.12.1/24
Switch A	VLAN-interface 200	13.13.13.1/24
Switch A	Loopback 0	1.1.1.1/32
Switch B	VLAN-interface 101	24.24.24.4/24
Switch B	VLAN-interface 200	13.13.13.2/24
Switch B	Loopback 0	4.4.4.4/32
Switch C	VLAN-interface 100	12.12.12.2/24
Switch C	VLAN-interface 101	24.24.24.2/24

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure static route FRR on link A:

   Use one of the following methods:

   - (Method 1.) Specify a backup next hop for static route FRR:

     # Configure a static route on Switch A, and specify VLAN-interface 100 as the backup output interface and 12.12.12.2 as the backup next hop.

     ```
     <SwitchA> system-view
     [SwitchA] ip route-static 4.4.4.4 32 vlan-interface 200 13.13.13.2
     backup-interface vlan-interface 100 backup-nexthop 12.12.12.2
     ```

     # Configure a static route on Switch B, and specify VLAN-interface 101 as the backup output interface and 24.24.24.2 as the backup next hop.

     ```
     <SwitchB> system-view
     [SwitchB] ip route-static 1.1.1.1 32 vlan-interface 200 13.13.13.1
     backup-interface vlan-interface 101 backup-nexthop 24.24.24.2
     ```

   - (Method 2.) Configure static route FRR to automatically select a backup next hop:

     # Configure static routes on Switch A, and enable static route FRR.

     ```
     <SwitchA> system-view
     [SwitchA] ip route-static 4.4.4.4 32 vlan-interface 200 13.13.13.2
     [SwitchA] ip route-static 4.4.4.4 32 vlan-interface 100 12.12.12.2 preference 70
     [SwitchA] ip route-static fast-reroute auto
     ```

     # Configure static routes on Switch B, and enable static route FRR.

     ```
     <SwitchB> system-view
     [SwitchB] ip route-static 1.1.1.1 32 vlan-interface 200 13.13.13.1
     ```

```
[SwitchB] ip route-static 1.1.1.1 32 vlan-interface 101 24.24.24.2 preference 70
[SwitchB] ip route-static fast-reroute auto
```

3. Configure static routes on Switch C.

```
<SwitchC> system-view
[SwitchC] ip route-static 4.4.4.4 32 vlan-interface 101 24.24.24.4
[SwitchC] ip route-static 1.1.1.1 32 vlan-interface 100 12.12.12.1
```

Verifying the configuration

Display route 4.4.4.4/32 on Switch A to view the backup next hop information.

```
[SwitchA] display ip routing-table 4.4.4.4 verbose

Summary count : 1

Destination: 4.4.4.4/32
   Protocol: Static
 Process ID: 0
  SubProtID: 0x0                  Age: 04h20m37s
       Cost: 0              Preference: 60
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                 State: Active Adv
  OrigTblID: 0x0              OrigVrf: default-vrf
    TableID: 0x2               OrigAs: 0
      NibID: 0x26000002         LastAs: 0
     AttrID: 0xffffffff       Neighbor: 0.0.0.0
      Flags: 0x1008c       OrigNextHop: 13.13.13.2
      Label: NULL          RealNextHop: 13.13.13.2
    BkLabel: NULL            BkNextHop: 12.12.12.2
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface200
BkTunnel ID: Invalid       BkInterface: Vlan-interface100
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Display route 1.1.1.1/32 on Switch B to view the backup next hop information.

```
[SwitchB] display ip routing-table 1.1.1.1 verbose

Summary count : 1

Destination: 1.1.1.1/32
   Protocol: Static
 Process ID: 0
  SubProtID: 0x0                  Age: 04h20m37s
       Cost: 0              Preference: 60
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                 State: Active Adv
  OrigTblID: 0x0              OrigVrf: default-vrf
    TableID: 0x2               OrigAs: 0
```

```
      NibID: 0x26000002          LastAs: 0
     AttrID: 0xffffffff        Neighbor: 0.0.0.0
      Flags: 0x1008c        OrigNextHop: 13.13.13.1
      Label: NULL           RealNextHop: 13.13.13.1
    BkLabel: NULL             BkNextHop: 24.24.24.2
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface200
BkTunnel ID: Invalid        BkInterface: Vlan-interface101
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Configuring a default route

A default route is used to forward packets that do not match any specific routing entry in the routing table. Without a default route, packets that do not match any routing entries are discarded and an ICMP destination-unreachable packet is sent to the source.

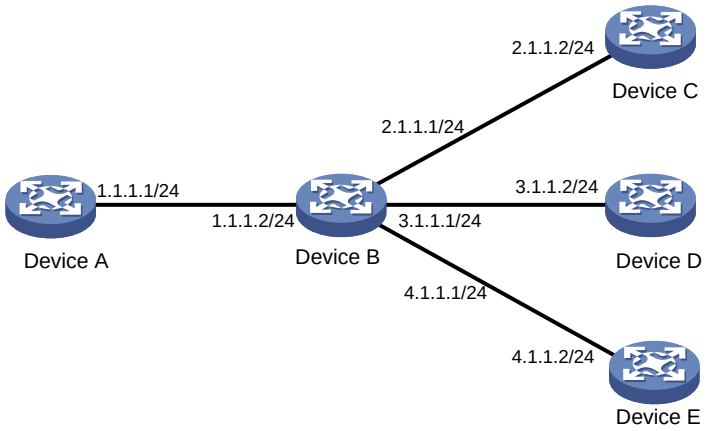
A default route can be configured in either of the following ways:

- The network administrator can configure a default route with both destination and mask being 0.0.0.0. For more information, see "Configuring static routing."
- Some dynamic routing protocols (such as OSPF, IS-IS, and RIP) can generate a default route. For example, an upstream router running OSPF can generate a default route and advertise it to other routers. These routers install the default route with the next hop being the upstream router. For more information, see the respective chapters on these routing protocols in this configuration guide.

As shown in Figure 6, Device B is the next hop for packets from Device A to Device C, Device D, and Device E. You can configure a default route on Device A to replace the three static routes from Device A to Device C, Device D, and Device E, respectively. Then, packets that do not match any other routes will be forwarded based on the default route.

The next hop address, destination address, and subnet mask of the default route configured on Device A are 1.1.1.2, 0.0.0.0, and 0.0.0.0, respectively.

Figure 6 Configuring a default route



Contents

Configuring RIP 1

About RIP 1
RIP routing metrics 1
RIP route entries 1
RIP operation 1
Routing loop prevention 1
RIP versions 2
Protocols and standards 2
RIP tasks at a glance 2
Configuring basic RIP 3
Restrictions and guidelines for configuring basic RIP 3
Enabling RIP 3
Controlling RIP reception and advertisement on interfaces 4
Configuring a RIP version 5
Specifying a RIP neighbor 6
Configuring RIP route control 6
Configuring an additional routing metric 6
Configuring RIPv2 route summarization 7
Disabling host route reception 8
Advertising a default route 8
Configuring received/redistributed route filtering 9
Setting a preference for RIP 10
Configuring RIP route redistribution 10
Tuning and optimizing RIP networks 11
Setting RIP timers 11
Enabling split horizon and poison reverse 11
Setting the maximum number of RIP ECMP routes 12
Setting the RIP triggered update interval 12
Configuring the RIP packet sending rate 13
Setting the maximum length of RIP packets 13
Setting the DSCP value for outgoing RIP packets 14
Configuring RIP network management 14
Configuring RIP GR 14
Enabling RIP NSR 15
Configuring BFD for RIP 16
About BFD for RIP 16
Restrictions and guidelines 16
Configuring single-hop echo detection (for a directly connected RIP neighbor) 16
Configuring single-hop echo detection (for a specific destination) 17
Configuring bidirectional control detection 17
Configuring RIP FRR 18
About RIP FRR 18
Restrictions and guidelines for RIP FRR 18
Enabling RIP FRR 18
Enabling BFD for RIP FRR 18
Enhancing RIP security 19
Enabling zero field check for incoming RIPv1 messages 19
Enabling source IP address check for incoming RIP updates 19
Configuring RIPv2 message authentication 20
Display and maintenance commands for RIP 20
RIP configuration examples 21
Example: Configuring basic RIP 21
Example: Configuring RIP route redistribution 24
Example: Configuring an additional metric for a RIP interface 26
Example: Configuring RIP to advertise a summary route 28
Example: Configuring RIP GR 30
Example: Configuring RIP NSR 31

Example: Configuring BFD for RIP (single-hop echo detection for a directly connected neighbor)·····33
Example: Configuring BFD for RIP (single hop echo detection for a specific destination)·····35
Example: Configuring BFD for RIP (bidirectional detection in BFD control packet mode)·····37
Example: Configuring RIP FRR ·····40

Configuring RIP

About RIP

Routing Information Protocol (RIP) is a distance-vector IGP suited to small-sized networks. It employs UDP to exchange route information through port 520.

RIP routing metrics

RIP uses a hop count to measure the distance to a destination. The hop count from a router to a directly connected network is 0. The hop count from a router to a directly connected router is 1. To limit convergence time, RIP restricts the value range of the metric from 0 to 15. A destination with a metric value of 16 (or greater) is considered unreachable. For this reason, RIP is not suitable for large-sized networks.

RIP route entries

RIP stores routing entries in a database. Each routing entry contains the following elements:

- **Destination address**—IP address of a destination host or a network.
- **Next hop**—IP address of the next hop.
- **Egress interface**—Egress interface of the route.
- **Metric**—Cost from the local router to the destination.
- **Route time**—Time elapsed since the last update. The time is reset to 0 when the routing entry is updated.
- **Route tag**—Used for route control. For more information, see "Configuring routing policies."

RIP operation

RIP works as follows:

1. RIP sends request messages to neighboring routers. Neighboring routers return response messages that contain their routing tables.
2. RIP uses the received responses to update the local routing table and sends triggered update messages to its neighbors. All RIP routers on the network do this to learn latest routing information.
3. RIP periodically sends the local routing table to its neighbors. After a RIP neighbor receives the message, it updates its routing table, selects optimal routes, and sends an update to other neighbors. RIP ages routes to keep only valid routes.

Routing loop prevention

RIP uses the following mechanisms to prevent routing loops:

- **Counting to infinity**—A destination with a metric value of 16 is considered unreachable. When a routing loop occurs, the metric value of a route will increment to 16 to avoid endless looping.
- **Triggered updates**—RIP immediately advertises triggered updates for topology changes to reduce the possibility of routing loops and to speed up convergence.
- **Split horizon**—Disables RIP from sending routes through the interface where the routes were learned to prevent routing loops and save bandwidth.

- **Poison reverse**—Enables RIP to set the metric of routes received from a neighbor to 16 and sends these routes back to the neighbor. The neighbor can delete such information from its routing table to prevent routing loops.

RIP versions

There are two RIP versions, RIPv1 and RIPv2.

RIPv1 is a classful routing protocol. It advertises messages only through broadcast. RIPv1 messages do not carry mask information, so RIPv1 can only recognize natural networks such as Class A, B, and C. For this reason, RIPv1 does not support discontiguous subnets.

RIPv2 is a classless routing protocol. It has the following advantages over RIPv1:

- Supports route tags to implement flexible route control through routing policies.
- Supports masks, route summarization, and CIDR.
- Supports designated next hops to select the best ones on broadcast networks.
- Supports multicasting route updates so only RIPv2 routers can receive these updates to reduce resource consumption.
- Supports plain text authentication and MD5 authentication to enhance security.

RIPv2 supports two transmission modes: broadcast and multicast. Multicast is the default mode using 224.0.0.9 as the multicast address. An interface operating in RIPv2 broadcast mode can also receive RIPv1 messages.

Protocols and standards

- RFC 1058, *Routing Information Protocol*
- RFC 1723, *RIP Version 2 - Carrying Additional Information*
- RFC 1721, *RIP Version 2 Protocol Analysis*
- RFC 1722, *RIP Version 2 Protocol Applicability Statement*
- RFC 1724, *RIP Version 2 MIB Extension*
- RFC 2082, *RIPv2 MD5 Authentication*
- RFC 2091, *Triggered Extensions to RIP to Support Demand Circuits*
- RFC 2453, *RIP Version 2*

RIP tasks at a glance

To configure RIP, perform the following tasks:

1. Configuring basic RIP
   a. Enabling RIP
   b. (Optional.) Controlling RIP reception and advertisement on interfaces
   c. (Optional.) Configuring a RIP version
   d. Specifying a RIP neighbor

      To enable RIP on a link that does not support broadcast or multicast, you must manually specify a RIP neighbor.
2. (Optional.) Configuring RIP route control
   - Configuring an additional routing metric
   - Configuring RIPv2 route summarization
   - Disabling host route reception

- Advertising a default route
- Configuring received/redistributed route filtering
- Setting a preference for RIP
- Configuring RIP route redistribution

3. (Optional.) Tuning and optimizing RIP networks
   - Setting RIP timers
   - Enabling split horizon and poison reverse
   - Setting the maximum number of RIP ECMP routes
   - Setting the RIP triggered update interval
   - Configuring the RIP packet sending rate
   - Setting the maximum length of RIP packets
   - Setting the DSCP value for outgoing RIP packets
4. (Optional.) Configuring RIP network management
5. Enhancing RIP availability
   - Configuring RIP GR
   - Enabling RIP NSR
   - Configuring BFD for RIP
   - Configuring RIP FRR
6. (Optional.) Enhancing RIP security
   - Enabling zero field check for incoming RIPv1 messages
   - Enabling source IP address check for incoming RIP updates
   - Configuring RIPv2 message authentication

Configuring basic RIP

Restrictions and guidelines for configuring basic RIP

To enable multiple RIP processes on a router, you must specify an ID for each process. A RIP process ID has only local significance. Two RIP routers having different process IDs can also exchange RIP packets.

Enabling RIP

About this task

You can enable RIP on a network and specify a wildcard mask for the network. After that, only the interface attached to the network runs RIP.

Restrictions and guidelines

If you configure RIP settings in interface view before enabling RIP, the settings do not take effect until RIP is enabled.

If a physical interface is attached to multiple networks, you cannot advertise these networks in different RIP processes.

You cannot enable multiple RIP processes on a physical interface.

The **`rip enable`** command takes precedence over the **`network`** command.

Enabling RIP on a network

1. Enter system view.
   system-view
2. Enable RIP and enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
   By default, RIP is disabled.
3. Enable RIP on a network.
   network *network-address* [*wildcard-mask*]
   By default, RIP is disabled on a network.
   The **network** 0.0.0.0 command can enable RIP on all interfaces in a single process, but does not apply to multiple RIP processes.

Enabling RIP on an interface

1. Enter system view.
   system-view
2. Enable RIP and enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
   By default, RIP is disabled.
3. Return to system view.
   quit
4. Enter interface view.
   interface *interface-type interface-number*
5. Enable RIP on the interface.
   rip *process-id* **enable** [**exclude-subip**]
   By default, RIP is disabled on an interface.

Controlling RIP reception and advertisement on interfaces

About this task

You can perform this task to configure the following features:

- Suppressing an interface. The suppressed interface can receive RIP messages but cannot send RIP messages.
- Disabling an interface from sending RIP messages.
- Disabling an interface from receiving RIP messages.

Restrictions and guidelines for RIP reception and advertisement control on interfaces

An interface suppressed by using the **silent-interface** command can only receive RIP messages. It cannot send RIP messages. You can use the **silent-interface all** command to suppress all interfaces. The **silent-interface** command takes precedence over the **rip input** and **rip output** commands.

Suppressing an interface

1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Suppress an interface.

   silent-interface { *interface-type interface-number* | **all** }

   By default, all RIP-enabled interfaces can send RIP messages.

   The suppressed interface can still receive RIP messages and respond to unicast requests containing unknown ports.

Disabling an interface from receiving RIP messages

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Disable an interface from receiving RIP messages.

   undo rip input

   By default, a RIP-enabled interface can receive RIP messages.

Disabling an interface from sending RIP messages

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Disable an interface from sending RIP messages.

   undo rip output

   By default, a RIP-enabled interface can send RIP messages.

Configuring a RIP version

About this task

You can configure a global RIP version in RIP view or an interface-specific RIP version in interface view.

An interface preferentially uses the interface-specific RIP version. If no interface-specific version is specified, the interface uses the global RIP version. If neither a global nor interface-specific RIP version is configured, the interface sends RIPv1 broadcasts and can receive the following:

- RIPv1 broadcasts and unicasts.
- RIPv2 broadcasts, multicasts, and unicasts.

Procedure

1. Enter system view.

   system-view

2. Specify a RIP version.

   - Execute the following commands in sequence to specify a global RIP version:

     rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

     version { **1** | **2** }

     By default, no global version is specified. An interface sends RIPv1 broadcasts, and can receive RIPv1 broadcasts and unicasts, and RIPv2 broadcasts, multicasts, and unicasts.

   - Execute the following commands in sequence to specify a RIP version on an interface:

     interface *interface-type interface-number*

     rip version { **1** | **2** [**broadcast** | **multicast**] }

By default, no interface-specific RIP version is specified. The interface sends RIPv1 broadcasts, and can receive RIPv1 broadcasts and unicasts, and RIPv2 broadcasts, multicasts, and unicasts.

Specifying a RIP neighbor

About this task

Typically RIP messages are sent in broadcast or multicast. To enable RIP on a link that does not support broadcast or multicast, you must manually specify a RIP neighbor.

Restrictions and guidelines

As a best practice, do not use the **peer** *ip-address* command to specify a directly connected neighbor. The neighbor might receive a route update in both unicast and multicast (or broadcast) messages from the device.

Procedure

1. Enter system view.

   system-view
2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Specify a RIP neighbor.

   peer *ip-address*

   By default, RIP does not unicast updates to any peer.
4. Disable source IP address check on inbound RIP updates.

   undo validate-source-address

   By default, source IP address check is enabled on inbound RIP updates.

   If the specified neighbor is not directly connected, disable source address check on incoming updates.

Configuring RIP route control

Configuring an additional routing metric

About this task

An additional routing metric (hop count) can be added to the metric of an inbound or outbound RIP route.

- An outbound additional metric is added to the metric of a sent route, and it does not change the route's metric in the routing table.
- An inbound additional metric is added to the metric of a received route before the route is added into the routing table, and the route's metric is changed. If the sum of the additional metric and the original metric is greater than 16, the metric of the route is 16.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify an inbound additional routing metric.

rip metricin [**route-policy** *route-policy-name*] *value*

By default, the additional metric of an inbound route is 0.

4. Specify an outbound additional routing metric.

rip metricout [**route-policy** *route-policy-name*] *value*

By default, the additional metric of an outbound route is 1.

Configuring RIPv2 route summarization

About this task

Perform this task to summarize contiguous subnets into a summary network and sends the network to neighbors. The smallest metric among all summarized routes is used as the metric of the summary route.

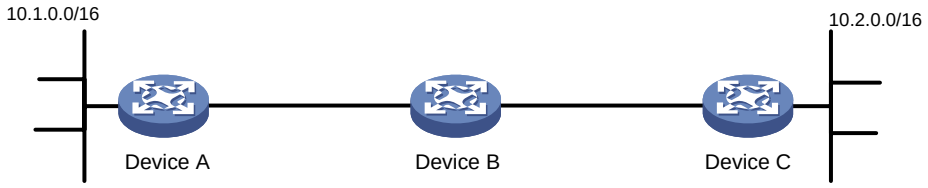
You can use the following methods to summarize routes in RIPv2:

- **Automatic summarization**—Configure RIPv2 to generate a natural network for contiguous subnets. For example, suppose there are three subnet routes 10.1.1.0/24, 10.1.2.0/24, and 10.1.3.0/24. Automatic summarization automatically creates and advertises a summary route 10.0.0.0/8 instead of the more specific routes.
- **Manual summarization**—Manually configure a summary route. RIPv2 advertises the summary route rather than more specific routes. For example, suppose contiguous subnets routes 10.1.1.0/24, 10.1.2.0/24, and 10.1.3.0/24 exist in the routing table. You can create a summary route 10.1.0.0/16 on GigabitEthernet 1/0/1 to advertise the summary route instead of the more specific routes. By default, natural masks are used to advertise summary routes. To manually configure a summary route on an interface, you must first disable RIPv2 automatic route summarization.

In certain cases, route summarization might result in incorrect route calculation because it disables subnet route advertisement. If the routing table contains non-contiguous subnets, disable automatic route summarization so that RIPv2 can advertise subnet routes and host routes for correct route calculation.

As shown in Figure 1, RIPv2 is enabled on all interfaces on Device A, Device B, and Device C. If automatic route summarization is enabled (default setting), both Device A and Device C will send 10.1.0.0/8 to Device B. To ensure correct route calculation on Device B based on subnets 10.1.0.0/16 and 10.2.0.0/16, disable RIPv2 automatic route summarization.

Figure 1 Disabling RIPv2 automatic route summarization



Restrictions and guidelines

To prevent loops caused by route summarization, create a black hole route by specifying interface NULL 0 as the output interface of the summary route. Packets that match the black hole route are dropped.

Enabling RIPv2 automatic route summarization

1. Enter system view.

system-view

2. Enter RIP view.

rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enable RIPv2 automatic route summarization.

   summary

   By default, RIPv2 automatic route summarization is enabled.

   If subnets in the routing table are not contiguous, disable automatic route summarization to advertise more specific routes.

Advertising a summary route

1. Enter system view.

   system-view

2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Disable RIPv2 automatic route summarization.

   undo summary

   By default, RIPv2 automatic route summarization is enabled.

4. Return to system view.

   quit

5. Enter interface view.

   interface *interface-type interface-number*

6. Configure a summary route.

   rip summary-address *ip-address* { *mask-length* | *mask* }

   By default, no summary route is configured.

Disabling host route reception

About this task

This task disables RIPv2 from receiving host routes from the same network to save network resources. This feature does not apply to RIPv1.

Procedure

1. Enter system view.

   system-view

2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Disable RIP from receiving host routes.

   undo host-route

   By default, RIP receives host routes.

Advertising a default route

About this task

You can advertise a default route on all RIP interfaces in RIP view or on a specific RIP interface in interface view. The interface view setting takes precedence over the RIP view settings.

To disable an interface from advertising a default route, use the **rip default-route no-originate** command on the interface.

The router enabled to advertise a default route does not accept default routes from RIP neighbors.

Procedure

1. Enter system view.

   system-view

2. Advertise a default route.
   - Execute the following commands in sequence to configure RIP to advertise a default route:

     rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

     default-route { **only** | **originate** } [**cost** *cost-value* | **route-policy** *route-policy-name*] *

     By default, RIP does not advertise a default route.
   - Execute the following commands in sequence to configure a RIP interface to advertise a default route:

     interface *interface-type interface-number*

     rip default-route { { **only** | **originate** } [**cost** *cost-value* | **route-policy** *route-policy-name*] * | **no-originate** }

     By default, a RIP interface can advertise a default route if the RIP process is enabled to advertise a default route.

Configuring received/redistributed route filtering

About this task

This task allows you to create a policy to filter received or redistributed routes that match specific criteria such as an ACL, IP prefix list, or routing policy.

Procedure

1. Enter system view.

   system-view

2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Configure the filtering of received routes.

   filter-policy { *ipv4-acl-number* | **gateway** *prefix-list-name* | **prefix-list** *prefix-list-name* [**gateway** *prefix-list-name*] | **route-policy** *route-policy-name* } **import** [*interface-type interface-number*]

   By default, the filtering of received routes is not configured.

   This command filters received routes. Filtered routes are not installed into the routing table or advertised to neighbors.

4. Configure the filtering of redistributed routes.

   filter-policy { *ipv4-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } **export** [*protocol* [*process-id*] | *interface-type interface-number*]

   By default, the filtering of redistributed routes is not configured.

   This command filters redistributed routes, including routes redistributed with the **import-route** command.

Setting a preference for RIP

About this task

If multiple IGPs find routes to the same destination, the route found by the IGP that has the highest priority is selected as the optimal route. Perform this task to assign a preference to RIP. The smaller the preference value, the higher the priority.

Procedure

1. Enter system view.

   system-view
2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set a preference for RIP.

   preference { *preference* | **route-policy** *route-policy-name* } *

   The default preference for RIP is 100.

Configuring RIP route redistribution

About this task

Perform this task to configure RIP to redistribute routes from other routing protocols, including OSPF, IS-IS, BGP, static, and direct.

Procedure

1. Enter system view.

   system-view
2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Configure route redistribution.
   - Redistribute routes from BGP.

     import-route bgp [*as-number*] [**allow-ibgp**] [**cost** *cost-value* | **route-policy** *route-policy-name* | **tag** *tag*] *
   - Redistribute direct or static routes.

     import-route { **direct** | **static** } [**cost** *cost-value* | **route-policy** *route-policy-name* | **tag** *tag*] *
   - Redistribute routes from OSPF, IS-IS, or other RIP processes.

     import-route { **isis** | **ospf** | **rip** } [*process-id* | **all-processes**] [**allow-direct** | **cost** *cost-value* | **route-policy** *route-policy-name* | **tag** *tag*] *

   By default, RIP does not redistribute routes.

   This command can redistribute only active routes. To view active routes, use the **display ip routing-table protocol** command.
4. (Optional.) Set a default cost for redistributed routes.

   default cost *cost-value*

   The default cost for redistributed routes is 0.

Tuning and optimizing RIP networks

Setting RIP timers

About this task

You can change the RIP network convergence speed by adjusting the following RIP timers:

- **Update timer**—Specifies the interval between route updates.
- **Timeout timer**—Specifies the route aging time. If no update for a route is received within the aging time, the metric of the route is set to 16.
- **Suppress timer**—Specifies how long a RIP route stays in suppressed state. When the metric of a route is 16, the route enters the suppressed state. A suppressed route can be replaced by an updated route that is received from the same neighbor before the suppress timer expires and has a metric less than 16.
- **Garbage-collect timer**—Specifies the interval from when the metric of a route becomes 16 to when it is deleted from the routing table. RIP advertises the route with a metric of 16. If no update is announced for that route before the garbage-collect timer expires, the route is deleted from the routing table.

Restrictions and guidelines

To avoid unnecessary traffic or route flapping, configure identical RIP timer settings on RIP routers.

Procedure

1. Enter system view.

   system-view

2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Set RIP timers.

   timers { **garbage-collect** *garbage-collect-value* | **suppress** *suppress-value* | **timeout** *timeout-value* | **update** *update-value* } *

   The default settings are as follows:

   - The garbage-collect timer is 120 seconds.
   - The suppress timer is 120 seconds.
   - The timeout timer is 180 seconds.
   - The update timer is 30 seconds.

Enabling split horizon and poison reverse

About this task

The split horizon and poison reverse features can prevent routing loops.

- Split horizon disables RIP from sending routes through the interface where the routes were learned to prevent routing loops between adjacent routers.
- Poison reverse allows RIP to send routes through the interface where the routes were learned. The metric of these routes is always set to 16 (unreachable) to avoid routing loops between neighbors.

Restrictions and guidelines

If both split horizon and poison reverse are configured, only the poison reverse feature takes effect.

Enabling split horizon

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable split horizon.
   rip split-horizon
   By default, split horizon is enabled.

Enabling poison reverse

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable poison reverse.
   rip poison-reverse
   By default, poison reverse is disabled.

Setting the maximum number of RIP ECMP routes

About this task

Perform this task to use ECMP routes to share the load of the RIP network.

Procedure

1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the maximum number of RIP ECMP routes.
   maximum load-balancing *number*
   By default, the maximum number of RIP ECMP routes equals the maximum number of ECMP routes, which is configurable by using the **max-ecmp-num** command.

Setting the RIP triggered update interval

About this task

Perform this task to avoid network overhead and reduce system resource consumption caused by frequent RIP triggered updates.

You can use the **timer triggered** command to set the maximum interval, minimum interval, and incremental interval for sending RIP triggered updates.

- For a stable network, the *minimum-interval* is used.
- If network changes become frequent, the incremental interval *incremental-interval* is used to extend the triggered update sending interval until the *maximum-interval* is reached.

Procedure

1. Enter system view.

system-view

2. Enter RIP view.

   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Set the RIP triggered update interval.

   timer triggered *maximum-interval* [*minimum-interval* [*incremental-interval*]]

   The default settings are as follows:

   - The maximum interval is 5 seconds.
   - The minimum interval is 50 milliseconds.
   - The incremental interval is 200 milliseconds.

Configuring the RIP packet sending rate

About this task

Perform this task to set the interval for sending RIP packets and the maximum number of RIP packets that can be sent at each interval. This feature can avoid excessive RIP packets from affecting system performance and consuming too much bandwidth.

Procedure

1. Enter system view.

   system-view

2. Configure the RIP packet sending rate.

   - Execute the following commands in sequence to configure the RIP packet sending rate for all interfaces:

     rip [*process-id*] [**vpn-instance** *vpn-instance-name*]

     output-delay *time* **count** *count*

     By default, an interface sends up to three RIP packets every 20 milliseconds.

   - Execute the following commands in sequence to configure the RIP packet sending rate for an interface:

     interface *interface-type interface-number*

     rip output-delay *time* **count** *count*

     By default, the interface uses the RIP packet sending rate configured for the RIP process that the interface runs.

Setting the maximum length of RIP packets

About this task

The packet length of RIP packets determines how many routes can be carried in a RIP packet. Set the maximum length of RIP packets to make good use of link bandwidth.

When authentication is enabled, follow these guidelines to ensure packet forwarding:

- For simple authentication, the maximum length of RIP packets must be no less than 52 bytes.
- For MD5 authentication (with packet format defined in RFC 2453), the maximum length of RIP packets must be no less than 56 bytes.
- For MD5 authentication (with packet format defined in RFC 2082), the maximum length of RIP packets must be no less than 72 bytes.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the maximum length of RIP packets.
   rip max-packet-length *value*
   By default, the maximum length of RIP packets is 512 bytes.

Setting the DSCP value for outgoing RIP packets

About this task

The DSCP value specifies the precedence of outgoing packets.

Procedure

1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the DSCP value for outgoing RIP packets.
   dscp *dscp-value*
   By default, the DSCP value for outgoing RIP packets is 48.

Configuring RIP network management

About this task

You can use network management software to manage the RIP process to which MIB is bound.

Procedure

1. Enter system view.
   system-view
2. Bind MIB to a RIP process.
   rip mib-binding *process-id*
   By default, MIB is bound to the RIP process with the smallest process ID.

Configuring RIP GR

About this task

GR ensures forwarding continuity when a routing protocol restarts or an active/standby switchover occurs.

Two routers are required to complete a GR process. The following are router roles in a GR process:

- **GR restarter**—Graceful restarting router. It must have GR capability.
- **GR helper**—A neighbor of the GR restarter. It helps the GR restarter to complete the GR process.

After RIP restarts on a router, the router must learn RIP routes again and update its FIB table, which causes network disconnections and route reconvergence.

With the GR feature, the restarting router (known as the GR restarter) can notify the event to its GR capable neighbors. GR capable neighbors (known as GR helpers) maintain their adjacencies with the router within a GR interval. During this process, the FIB table of the router does not change. After the restart, the router contacts its neighbors to retrieve its FIB.

By default, a RIP-enabled device acts as the GR helper. Perform this task on the GR restarter.

Procedure

1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable GR for RIP.
   graceful-restart
   By default, RIP GR is disabled.
4. (Optional.) Set the GR interval.
   graceful-restart interval *interval*
   By default, the GR interval is 60 seconds.

Enabling RIP NSR

About this task

Nonstop Routing (NSR) allows the device to back up the routing information from the active RIP process to the standby RIP process. After an active/standby switchover, NSR can complete route regeneration without tearing down adjacencies or impacting forwarding services.

To use RIP NSR, the system must have a minimum of two IRF member devices. NSR ensures nonstop services when RIP has redundant processes on multiple IRF member devices.

In contrast to GR, NSR does not require a neighbor device to recover routing information.

Restrictions and guidelines

When both GR and NSR are configured for RIP, the following rules apply:

- NSR has a higher priority than GR. When the active RIP process goes down, the device does not act as the GR restarter and performs RIP NSR to ensure nonstop forwarding.
- The device can participate in a GR process only when it acts as the GR helper. If it performs an active/standby RIP process switchover before the GR process finishes, the GR process might fail even if RIP NSR is enabled on the device.

Procedure

1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable RIP NSR.
   non-stop-routing
   By default, RIP NSR is disabled.
   RIP NSR enabled for a RIP process takes effect only on that process. As a best practice, enable RIP NSR for each process if multiple RIP processes exist.

Configuring BFD for RIP

About BFD for RIP

RIP detects route failures by periodically sending requests. If it receives no response for a route within a certain time, RIP considers the route unreachable. To speed up convergence, perform this task to enable BFD for RIP. For more information about BFD, see *High Availability Configuration Guide*.

RIP supports the following BFD detection modes:

- **Single-hop echo detection**—Detection mode for a directly connected neighbor. In this mode, a BFD session is established only when the directly connected neighbor has route information to send.
- **Single-hop echo detection for a specific destination**—Detection mode for a directly connected neighbor. In this mode, a BFD session is established to the specified RIP neighbor when RIP is enabled on the local interface. When BFD detects a unidirectional link, the local device will not receive or send any RIP packets through the interface to improve convergence speed. When the link recovers, the interface can send RIP packets again.
- **Bidirectional control detection**—Detection mode for both directly and indirectly connected neighbors. In this mode, a BFD session is established only when both ends have routes to send and BFD is enabled on the receiving interface.

Restrictions and guidelines

The **rip bfd enable** and **rip bfd enable destination** commands are mutually exclusive.

Configuring single-hop echo detection (for a directly connected RIP neighbor)

1. Enter system view.

   system-view

2. Configure the source IP address of BFD echo packets.

   bfd echo-source-ip *ip-address*

   By default, the source IP address of BFD echo packets is not configured.

   As a best practice, configure this command and do not specify a source IP address on the same network segment as any local interfaces. This can avoid network congestion caused by a large number of ICMP redirect packets sent from the peer.

   For more information about this command, see *High Availability Command Reference*.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable BFD for RIP.

   rip bfd enable

   By default, BFD for RIP is disabled.

Configuring single-hop echo detection (for a specific destination)

Restrictions and guidelines

This feature applies only to RIP neighbors that are directly connected.

Procedure

1. Enter system view.
   system-view
2. Configure the source IP address of BFD echo packets.
   bfd echo-source-ip *ip-address*
   By default, no source IP address is configured for BFD echo packets.
   As a best practice, configure this command and do not specify a source IP address on the same network segment as any local interfaces. This can avoid network congestion caused by a large number of ICMP redirect packets sent from the peer.
   For more information about this command, see *High Availability Command Reference*.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable BFD for RIP.
   rip bfd enable destination *ip-address*
   By default, BFD for RIP is disabled.

Configuring bidirectional control detection

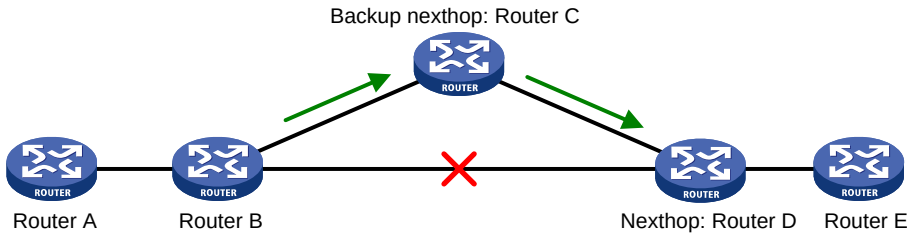
1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Specify a RIP neighbor.
   peer *ip-address*
   By default, RIP does not unicast updates to any peer.
   Because the **undo peer** command does not remove the neighbor relationship immediately, executing the command cannot bring down the BFD session immediately.
4. Enter interface view.
   interface *interface-type interface-number*
5. Enable BFD for RIP.
   rip bfd enable [**ctrl**]
   By default, BFD for RIP is disabled.

Configuring RIP FRR

About RIP FRR

A link or router failure on a path can cause packet loss and even routing loop until RIP completes routing convergence based on the new network topology. FRR enables fast rerouting to minimize the impact of link or node failures.

Figure 2 Network diagram for RIP FRR



As shown in Figure 2, configure FRR on Router B by using a routing policy to specify a backup next hop. When the primary link fails, RIP directs packets to the backup next hop. At the same time, RIP calculates the shortest path based on the new network topology, and forwards packets over that path after network convergence.

Restrictions and guidelines for RIP FRR

RIP FRR takes effect only for RIP routes learned from directly connected neighbors.

RIP FRR is available only when the state of primary link (with Layer 3 interfaces staying up) changes from bidirectional to unidirectional or down.

Equal-cost routes do not support RIP FRR.

Enabling RIP FRR

1. Enter system view.
   system-view
2. Configure a routing policy for FRR.
   You must specify a next hop by using the **apply fast-reroute backup-interface** command in the routing policy.
   For more information about routing policy configuration, see "Configuring routing policies."
3. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
4. Enable RIP FRR.
   fast-reroute route-policy *route-policy-name*
   By default, RIP FRR is disabled.

Enabling BFD for RIP FRR

About this task

By default, RIP FRR does not use BFD to detect primary link failures. For quicker RIP FRR, use BFD single-hop echo detection on the primary link of redundant links to detect link failure.

Procedure

1. Enter system view.
   system-view
2. Configure the source IP address of BFD echo packets.
   bfd echo-source-ip *ip-address*
   By default, the source IP address of BFD echo packets is not configured.
   As a best practice, configure this command and do not specify a source IP address on the same network segment as any local interfaces. This can avoid network congestion caused by a large number of ICMP redirect packets sent from the peer.
   Execute this command only when configuring echo-packet-mode BFD sessions.
   For more information about this command, see *High Availability Command Reference*.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable BFD for RIP FRR.
   rip primary-path-detect bfd { **ctrl** | **echo** }
   By default, BFD for RIP FRR is disabled.

Enhancing RIP security

Enabling zero field check for incoming RIPv1 messages

About this task

Some fields in the RIPv1 message must be set to zero. These fields are called "zero fields." You can enable zero field check for incoming RIPv1 messages. If a zero field of a message contains a non-zero value, RIP does not process the message. If you are certain that all messages are trustworthy, disable zero field check to save CPU resources.

This feature does not apply to RIPv2 packets, because they have no zero fields.

Procedure

1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable zero field check for incoming RIPv1 messages.
   checkzero
   By default, zero field check is disabled for incoming RIPv1 messages.

Enabling source IP address check for incoming RIP updates

About this task

Perform this task to enable source IP address check for incoming RIP updates.

Upon receiving a message on an Ethernet interface, RIP compares the source IP address of the message with the IP address of the interface. If they are not in the same network segment, RIP discards the message.

Upon receiving a message on a PPP interface, RIP checks whether the source address of the message is the IP address of the peer interface. If not, RIP discards the message.

Procedure

1. Enter system view.
   system-view
2. Enter RIP view.
   rip [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable source IP address check for incoming RIP messages.
   validate-source-address
   By default, source IP address check is disabled for incoming RIP updates.

Configuring RIPv2 message authentication

About this task

Perform this task to enable authentication on RIPv2 messages.

RIPv2 supports simple authentication, MD5 authentication, and keychain authentication.

- **Simple authentication**—As a best practice, do not configure the simple authentication mode in network requirements with high security requirements, because the passwords are sent with authentication packets.
- **MD5 authentication**—Uses MD5 passwords to perform packet encryption and decryption. This authentication method is securer than the simple authentication mode.
- **Keychain authentication**—Uses keychains to perform dynamic authentication. For more information about keychains, see keychain configuration in *Security Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure RIPv2 authentication.
   rip authentication-mode { **keychain** *keychain-name* { **rfc2453** | **rfc4822** } | **md5** { **rfc2082** { **cipher** | **plain** } *string key-id* | **rfc2453** { **cipher** | **plain** } *string* } | **simple** { **cipher** | **plain** } *string* }
   By default, RIPv2 authentication is not configured.
   RIPv1 does not support authentication. Although you can specify an authentication mode for RIPv1 in interface view, the configuration does not take effect.
   For more information about keychains, see keychain configuration in *Security Configuration Guide*.

Display and maintenance commands for RIP

Execute **display** commands in any view and execute **reset** commands in user view.

Task	Command
Display RIP current status and configuration information.	**display rip** [*process-id*]
Display RIP GR information.	**display rip** [*process-id*] **graceful-restart**

Task	Command
Display RIP NSR information.	**display rip** [*process-id*] **non-stop-routing**
Display active routes in the RIP database.	**display rip** *process-id* **database** [*ip-address* { *mask-length* | *mask* }]
Display RIP interface information.	**display rip** *process-id* **interface** [*interface-type interface-number*]
Display neighbor information for a RIP process.	**display rip** *process-id* **neighbor** [*interface-type interface-number*]
Display routing information for a RIP process.	**display rip** *process-id* **route** [*ip-address* { *mask-length* | *mask* } [**verbose**] | **peer** *ip-address* | **statistics**]
Reset a RIP process.	**reset rip** *process-id* **process**
Clear the statistics for a RIP process.	**reset rip** *process-id* **statistics**

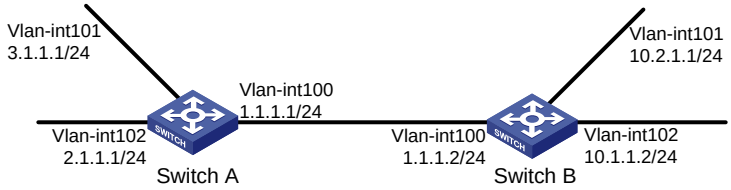
RIP configuration examples

Example: Configuring basic RIP

Network configuration

As shown in Figure 3, enable RIPv2 on all interfaces on Switch A and Switch B. Configure Switch B to not advertise route 10.2.1.0/24 to Switch A, and to accept only route 2.1.1.0/24 from Switch A.

Figure 3 Network diagram



Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Enable RIP.

   # Enable RIP on the specified networks on Switch A.

```
<SwitchA> system-view
[SwitchA] rip
[SwitchA-rip-1] network 1.0.0.0
[SwitchA-rip-1] network 2.0.0.0
[SwitchA-rip-1] network 3.0.0.0
[SwitchA-rip-1] quit
```

   # Enable RIP on the specified interfaces on Switch B.

```
<SwitchB> system-view
[SwitchB] rip
```

```
[SwitchB-rip-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] rip 1 enable
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] rip 1 enable
[SwitchB-Vlan-interface101] quit
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] rip 1 enable
[SwitchB-Vlan-interface102] quit
```

Display the RIP routing table of Switch A.

```
[SwitchA] display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
              D - Direct, O - Optimal, F - Flush to RIB
------------------------------------------------------------------------------
 Peer 1.1.1.2 on Vlan-interface100
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      10.0.0.0/8              1.1.1.2          1       0       RAOF    11
 Local route
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      1.1.1.0/24              0.0.0.0          0       0       RDOF    -
      2.1.1.0/24              0.0.0.0          0       0       RDOF    -
      3.1.1.0/24              0.0.0.0          0       0       RDOF    -
```

The output shows that RIPv1 uses a natural mask.

3. Configure a RIP version:

Configure RIPv2 on Switch A.

```
[SwitchA] rip
[SwitchA-rip-1] version 2
[SwitchA-rip-1] undo summary
[SwitchA-rip-1] quit
```

Configure RIPv2 on Switch B.

```
[SwitchB] rip
[SwitchB-rip-1] version 2
[SwitchB-rip-1] undo summary
[SwitchB-rip-1] quit
```

Display the RIP routing table on Switch A.

```
[SwitchA] display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
              D - Direct, O - Optimal, F - Flush to RIB
------------------------------------------------------------------------------

 Peer 1.1.1.2 on Vlan-interface100
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      10.0.0.0/8              1.1.1.2          1       0       RAOF    50
      10.2.1.0/24             1.1.1.2          1       0       RAOF    16
      10.1.1.0/24             1.1.1.2          1       0       RAOF    16
```

```
 Local route
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      1.1.1.0/24              0.0.0.0          0       0       RDOF    -
      2.1.1.0/24              0.0.0.0          0       0       RDOF    -
      3.1.1.0/24              0.0.0.0          0       0       RDOF    -
```

The output shows that RIPv2 uses classless subnet masks.

NOTE:

After RIPv2 is configured, RIPv1 routes might still exist in the routing table until they are aged out.

Display the RIP routing table on Switch B.

```
[SwitchB] display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
              D - Direct, O - Optimal, F - Flush to RIB
------------------------------------------------------------------------------
 Peer 1.1.1.1 on Vlan-interface100
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      2.1.1.0/24              1.1.1.1          1       0       RAOF    19
      3.1.1.0/24              1.1.1.1          1       0       RAOF    19
 Local route
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      1.1.1.0/24              0.0.0.0          0       0       RDOF    -
      10.1.1.0/24             0.0.0.0          0       0       RDOF    -
      10.2.1.0/24             0.0.0.0          0       0       RDOF    -
```

4. Configure route filtering:

Reference IP prefix lists on Switch B to filter received and redistributed routes.

```
[SwitchB] ip prefix-list aaa index 10 permit 2.1.1.0 24
[SwitchB] ip prefix-list bbb index 10 deny 10.2.1.0 24
[SwitchB] ip prefix-list bbb index 11 permit 0.0.0.0 0 less-equal 32
[SwitchB] rip 1
[SwitchB-rip-1] filter-policy prefix-list aaa import
[SwitchB-rip-1] filter-policy prefix-list bbb export
[SwitchB-rip-1] quit
```

Display the RIP routing table on Switch A.

```
[SwitchA] display rip 100 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
------------------------------------------------------------------------------
 Peer 1.1.1.2 on Vlan-interface100
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      10.1.1.0/24             1.1.1.2          1       0       RAOF    19
 Local route
      Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
      1.1.1.0/24              0.0.0.0          0       0       RDOF    -
      2.1.1.0/24              0.0.0.0          0       0       RDOF    -
```

```
     3.1.1.0/24              0.0.0.0          0       0       RDOF    -
```

Display the RIP routing table on Switch B.

```
[SwitchB] display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
              D - Direct, O - Optimal, F - Flush to RIB
------------------------------------------------------------------------------
 Peer 1.1.1.1 on Vlan-interface100
     Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
     2.1.1.0/24              1.1.1.1          1       0       RAOF    19
 Local route
     Destination/Mask        Nexthop          Cost    Tag     Flags   Sec
     1.1.1.0/24              0.0.0.0          0       0       RDOF    -
     10.1.1.0/24             0.0.0.0          0       0       RDOF    -
     10.2.1.0/24             0.0.0.0          0       0       RDOF    -
```

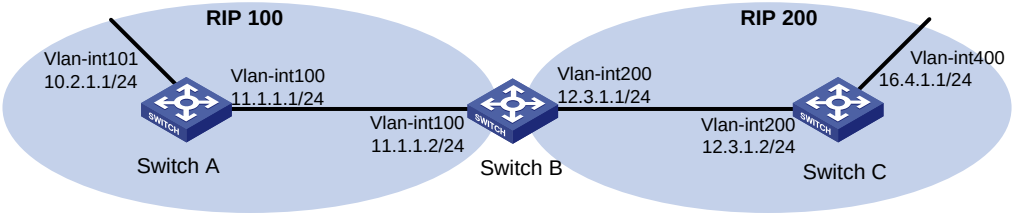
Example: Configuring RIP route redistribution

Network configuration

As shown in Figure 4, Switch B communicates with Switch A through RIP 100 and with Switch C through RIP 200.

Configure RIP 200 to redistribute direct routes and routes from RIP 100 on Switch B so Switch C can learn routes destined for 10.2.1.0/24 and 11.1.1.0/24. Switch A cannot learn routes destined for 12.3.1.0/24 and 16.4.1.0/24.

Figure 4 Network diagram



Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Configure basic RIP settings:

   # Enable RIP 100, and configure RIPv2 on Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] rip 100
   [SwitchA-rip-100] network 10.0.0.0
   [SwitchA-rip-100] network 11.0.0.0
   [SwitchA-rip-100] version 2
   [SwitchA-rip-100] undo summary
   [SwitchA-rip-100] quit
   ```

   # Enable RIP 100 and RIP 200, and configure RIPv2 on Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] rip 100
   [SwitchB-rip-100] network 11.0.0.0
   ```

```
[SwitchB-rip-100] version 2
[SwitchB-rip-100] undo summary
[SwitchB-rip-100] quit
[SwitchB] rip 200
[SwitchB-rip-200] network 12.0.0.0
[SwitchB-rip-200] version 2
[SwitchB-rip-200] undo summary
[SwitchB-rip-200] quit
```

Enable RIP 200, and configure RIPv2 on Switch C.

```
<SwitchC> system-view
[SwitchC] rip 200
[SwitchC-rip-200] network 12.0.0.0
[SwitchC-rip-200] network 16.0.0.0
[SwitchC-rip-200] version 2
[SwitchC-rip-200] undo summary
[SwitchC-rip-200] quit
```

Display the IP routing table on Switch C.

```
[SwitchC] display ip routing-table

Destinations : 13        Routes : 13

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
12.3.1.0/24         Direct 0    0            12.3.1.2        Vlan200
12.3.1.0/32         Direct 0    0            12.3.1.2        Vlan200
12.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
12.3.1.255/32       Direct 0    0            12.3.1.2        Vlan200
16.4.1.0/24         Direct 0    0            16.4.1.1        Vlan400
16.4.1.0/32         Direct 0    0            16.4.1.1        Vlan400
16.4.1.1/32         Direct 0    0            127.0.0.1       InLoop0
16.4.1.255/32       Direct 0    0            16.4.1.1        Vlan400
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

3. Configure route redistribution:

Configure RIP 200 to redistribute routes from RIP 100 and direct routes on Switch B.

```
[SwitchB] rip 200
[SwitchB-rip-200] import-route rip 100
[SwitchB-rip-200] import-route direct
[SwitchB-rip-200] quit
```

Display the IP routing table on Switch C.

```
[SwitchC] display ip routing-table

Destinations : 15        Routes : 15

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
```

```
10.2.1.0/24         RIP    100  1           12.3.1.1        Vlan200
11.1.1.0/24         RIP    100  1           12.3.1.1        Vlan200
12.3.1.0/24         Direct 0    0           12.3.1.2        Vlan200
12.3.1.0/32         Direct 0    0           12.3.1.2        Vlan200
12.3.1.2/32         Direct 0    0           127.0.0.1       InLoop0
12.3.1.255/32       Direct 0    0           12.3.1.2        Vlan200
16.4.1.0/24         Direct 0    0           16.4.1.1        Vlan400
16.4.1.0/32         Direct 0    0           16.4.1.1        Vlan400
16.4.1.1/32         Direct 0    0           127.0.0.1       InLoop0
16.4.1.255/32       Direct 0    0           16.4.1.1        Vlan400
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0           127.0.0.1       InLoop0
```

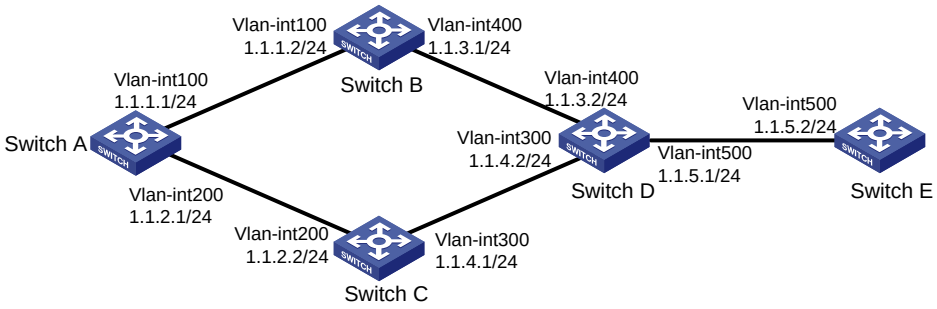
Example: Configuring an additional metric for a RIP interface

Network configuration

As shown in Figure 5, run RIPv2 on all the interfaces of Switch A, Switch B, Switch C, Switch D, and Switch E.

Switch A has two links to Switch D. The link from Switch B to Switch D is more stable than that from Switch C to Switch D. Configure an additional metric for RIP routes received from VLAN-interface 200 on Switch A so Switch A prefers route 1.1.5.0/24 learned from Switch B.

Figure 5 Network diagram



Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Configure basic RIP settings:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] rip 1
   [SwitchA-rip-1] network 1.0.0.0
   [SwitchA-rip-1] version 2
   [SwitchA-rip-1] undo summary
   [SwitchA-rip-1] quit
   ```

   # Configure Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] rip 1
   [SwitchB-rip-1] network 1.0.0.0
   ```

```
[SwitchB-rip-1] version 2
[SwitchB-rip-1] undo summary
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchB] rip 1
[SwitchC-rip-1] network 1.0.0.0
[SwitchC-rip-1] version 2
[SwitchC-rip-1] undo summary
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] rip 1
[SwitchD-rip-1] network 1.0.0.0
[SwitchD-rip-1] version 2
[SwitchD-rip-1] undo summary
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] rip 1
[SwitchE-rip-1] network 1.0.0.0
[SwitchE-rip-1] version 2
[SwitchE-rip-1] undo summary
```

Display all active routes in the RIP database on Switch A.

```
[SwitchA] display rip 1 database
   1.0.0.0/8, auto-summary
       1.1.1.0/24, cost 0, nexthop 1.1.1.1, RIP-interface
       1.1.2.0/24, cost 0, nexthop 1.1.2.1, RIP-interface
       1.1.3.0/24, cost 1, nexthop 1.1.1.2
       1.1.4.0/24, cost 1, nexthop 1.1.2.2
       1.1.5.0/24, cost 2, nexthop 1.1.1.2
       1.1.5.0/24, cost 2, nexthop 1.1.2.2
```

The output shows two RIP routes destined for network 1.1.5.0/24, with the next hops as Switch B (1.1.1.2) and Switch C (1.1.2.2), and with the same cost of 2.

3. Configure an additional metric for a RIP interface:

Configure an inbound additional metric of 3 for RIP-enabled interface VLAN-interface 200 on Switch A.

```
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] rip metricin 3
```

Display all active routes in the RIP database on Switch A.

```
[SwitchA-Vlan-interface200] display rip 1 database
   1.0.0.0/8, auto-summary
       1.1.1.0/24, cost 0, nexthop 1.1.1.1, RIP-interface
       1.1.2.0/24, cost 0, nexthop 1.1.2.1, RIP-interface
       1.1.3.0/24, cost 1, nexthop 1.1.1.2
       1.1.4.0/24, cost 2, nexthop 1.1.1.2
       1.1.5.0/24, cost 2, nexthop 1.1.1.2
```

The output shows that only one RIP route reaches network 1.1.5.0/24, with the next hop as Switch B (1.1.1.2) and a cost of 2.

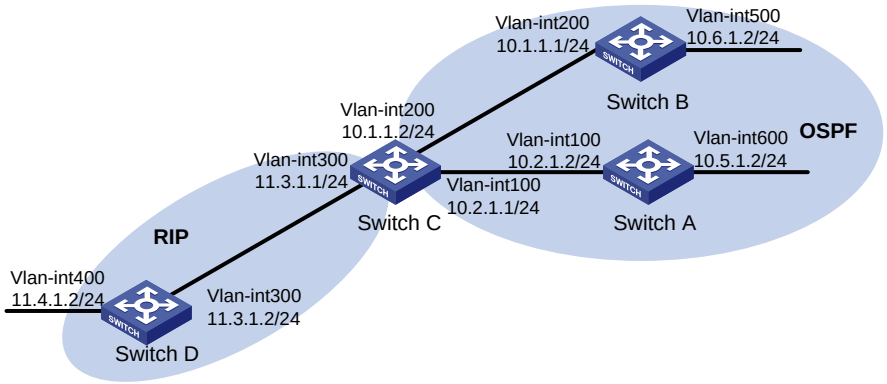
Example: Configuring RIP to advertise a summary route

Network configuration

As shown in Figure 6, Switch A and Switch B run OSPF, Switch D runs RIP, and Switch C runs OSPF and RIP. Configure RIP to redistribute OSPF routes on Switch C so Switch D can learn routes destined for networks 10.1.1.0/24, 10.2.1.0/24, 10.5.1.0/24, and 10.6.1.0/24.

To reduce the routing table size of Switch D, configure route summarization on Switch C to advertise only the summary route 10.0.0.0/8 to Switch D.

Figure 6 Network diagram



Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Configure basic OSPF settings:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] ospf
   [SwitchA-ospf-1] area 0
   [SwitchA-ospf-1-area-0.0.0.0] network 10.5.1.0 0.0.0.255
   [SwitchA-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
   [SwitchA-ospf-1-area-0.0.0.0] quit
   ```

   # Configure Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] ospf
   [SwitchB-ospf-1] area 0
   [SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
   [SwitchB-ospf-1-area-0.0.0.0] network 10.6.1.0 0.0.0.255
   [SwitchB-ospf-1-area-0.0.0.0] quit
   ```

   # Configure Switch C.

   ```
   <SwitchC> system-view
   [SwitchC] ospf
   [SwitchC-ospf-1] area 0
   [SwitchC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
   [SwitchC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
   [SwitchC-ospf-1-area-0.0.0.0] quit
   [SwitchC-ospf-1] quit
   ```

3. Configure basic RIP settings:

Configure Switch C.

```
[SwitchC] rip 1
[SwitchC-rip-1] network 11.3.1.0
[SwitchC-rip-1] version 2
[SwitchC-rip-1] undo summary
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] rip 1
[SwitchD-rip-1] network 11.0.0.0
[SwitchD-rip-1] version 2
[SwitchD-rip-1] undo summary
[SwitchD-rip-1] quit
```

Configure RIP to redistribute routes from OSPF process 1 and direct routes on Switch C.

```
[SwitchC-rip-1] import-route direct
[SwitchC-rip-1] import-route ospf 1
[SwitchC-rip-1] quit
```

Display the IP routing table on Switch D.

```
[SwitchD] display ip routing-table

Destinations : 15        Routes : 15

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.1.1.0/24         RIP    100  1            11.3.1.1        Vlan300
10.2.1.0/24         RIP    100  1            11.3.1.1        Vlan300
10.5.1.0/24         RIP    100  1            11.3.1.1        Vlan300
10.6.1.0/24         RIP    100  1            11.3.1.1        Vlan300
11.3.1.0/24         Direct 0    0            11.3.1.2        Vlan300
11.3.1.0/32         Direct 0    0            11.3.1.2        Vlan300
11.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
11.4.1.0/24         Direct 0    0            11.4.1.2        Vlan400
11.4.1.0/32         Direct 0    0            11.4.1.2        Vlan400
11.4.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

4. Configure route summarization:

Configure route summarization on Switch C and advertise only the summary route 10.0.0.0/8.

```
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] rip summary-address 10.0.0.0 8
```

Display the IP routing table on Switch D.

```
[SwitchD] display ip routing-table

Destinations : 12        Routes : 12

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
```

```
10.0.0.0/8          RIP    100  1           11.3.1.1        Vlan300
11.3.1.0/24         Direct 0    0           11.3.1.2        Vlan300
11.3.1.0/32         Direct 0    0           11.3.1.2        Vlan300
11.3.1.2/32         Direct 0    0           127.0.0.1       InLoop0
11.4.1.0/24         Direct 0    0           11.4.1.2        Vlan400
11.4.1.0/32         Direct 0    0           11.4.1.2        Vlan400
11.4.1.2/32         Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0           127.0.0.1       InLoop0
```

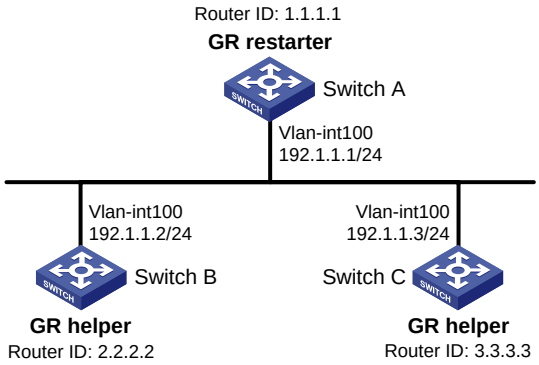
Example: Configuring RIP GR

Network configuration

As shown in Figure 7, Switch A, Switch B, and Switch C all run RIPv2.

- Enable GR on Switch A. Switch A acts as the GR restarter.
- Switch B and Switch C act as GR helpers to synchronize their routing tables with Switch A by using GR.

Figure 7 Network diagram



Procedure

1. Configure IP addresses and subnet masks for the interfaces on the switches. (Details not shown.)
2. Configure RIPv2 on the switches to ensure the following: (Details not shown.)
   - Switch A, Switch B, and Switch C can communicate with each other at Layer 3.
   - Dynamic route update can be implemented among them with RIPv2.
3. Enable RIP GR on Switch A.

```
<SwitchA> system-view
[SwitchA] rip
[SwitchA-rip-1] graceful-restart
```

Verifying the configuration

Restart RIP or trigger an active/standby switchover, and then display GR status on Switch A.

```
<SwitchA> display rip graceful-restart
 RIP process: 1
 Graceful Restart capability     : Enabled
```

```
Current GR state                : Normal
Graceful Restart period         : 60  seconds
Graceful Restart remaining time : 0   seconds
```

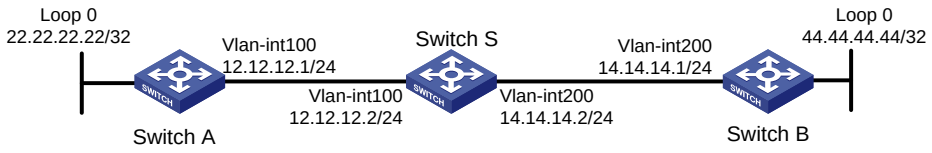
Example: Configuring RIP NSR

Network configuration

As shown in Figure 8, Switch A, Switch B, and Switch S all run RIPv2.

Enable RIP NSR on Switch S to ensure correct routing when an active/standby switchover occurs on Switch S.

Figure 8 Network diagram



Procedure

1. Configure IP addresses and subnet masks for the interfaces on the switches. (Details not shown.)
2. Configure RIPv2 on the switches to ensure the following: (Details not shown.)
   - Switch A, Switch B, and Switch S can communicate with each other at Layer 3.
   - Dynamic route update can be implemented among them with RIPv2.
3. Enable RIP NSR.

   # Enable RIP NSR on Switch S.

   ```
   <SwitchS> system-view
   [SwitchS] rip 100
   [SwitchS-rip-100] non-stop-routing
   [SwitchS-rip-100] quit
   ```

Verifying the configuration

Perform an active/standby switchover on Switch S.

```
[SwitchS] placement reoptimize
Predicted changes to the placement
Service Group(instance name)                          Cur location  New location
----------------------------------------------------------------------------
lb                                                    1/0           1/0
lsm                                                   1/0           1/0
slsp                                                  1/0           1/0
rib6                                                  1/0           1/0
routepolicy                                           1/0           1/0
rib                                                   1/0           1/0
staticroute6                                          1/0           1/0
staticroute                                           1/0           1/0
rip                                                   1/0           1/0
Continue? [y/n]:y
Re-optimization of the placement start. You will be notified on completion
```

```
Re-optimization of the placement complete. Use 'display placement' to view the new
placement
```

Display neighbor information and route information on Switch A.

```
[SwitchA] display rip 1 neighbor
 Neighbor Address: 12.12.12.2
     Interface  : Vlan-interface200
     Version    : RIPv2     Last update: 00h00m13s
     Relay nbr  : No        BFD session: None
     Bad packets: 0         Bad routes : 0
[SwitchA] display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
              D - Direct, O - Optimal, F - Flush to RIB
 ------------------------------------------------------------------------------
 Peer 12.12.12.2 on Vlan-interface200
      Destination/Mask        Nexthop           Cost    Tag     Flags   Sec
      14.0.0.0/8              12.12.12.2        1       0       RAOF    16
      44.0.0.0/8              12.12.12.2        2       0       RAOF    16
 Local route
      Destination/Mask        Nexthop           Cost    Tag     Flags   Sec
      12.12.12.0/24           0.0.0.0           0       0       RDOF    -
      22.22.22.22/32          0.0.0.0           0       0       RDOF    -
```

Display neighbor information and route information on Switch B.

```
[SwitchB] display rip 1 neighbor
 Neighbor Address: 14.14.14.2
     Interface  : Vlan-interface200
     Version    : RIPv2     Last update: 00h00m32s
     Relay nbr  : No        BFD session: None
     Bad packets: 0         Bad routes : 0
[SwitchB] display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
              D - Direct, O - Optimal, F - Flush to RIB
 ------------------------------------------------------------------------------
 Peer 14.14.14.2 on Vlan-interface200
      Destination/Mask        Nexthop           Cost    Tag     Flags   Sec
      12.0.0.0/8              14.14.14.2        1       0       RAOF    1
      22.0.0.0/8              14.14.14.2        2       0       RAOF    1
 Local route
      Destination/Mask        Nexthop           Cost    Tag     Flags   Sec
      44.44.44.44/32          0.0.0.0           0       0       RDOF    -
      14.14.14.0/24           0.0.0.0           0       0       RDOF    -
```

The output shows that the neighbor and route information on Switch A and Switch B keep unchanged during the active/standby switchover on Switch S. The traffic from Switch A to Switch B has not been impacted.

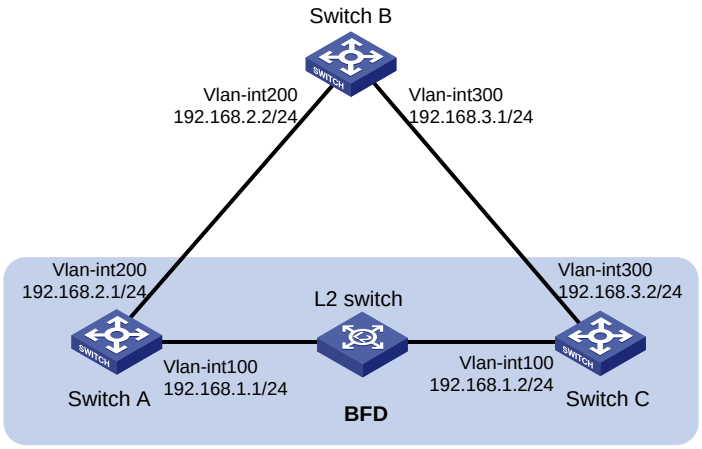
Example: Configuring BFD for RIP (single-hop echo detection for a directly connected neighbor)

Network configuration

As shown in Figure 9, VLAN-interface 100 of Switch A and Switch C runs RIP process 1. VLAN-interface 200 of Switch A runs RIP process 2. VLAN-interface 300 of Switch C and VLAN-interface 200 and VLAN-interface 300 of Switch B run RIP process 1.

- Configure a static route destined for 100.1.1.1/24 and enable static route redistribution into RIP on Switch C. This allows Switch A to learn two routes destined for 100.1.1.1/24 through VLAN-interface 100 and VLAN-interface 200 respectively, and uses the one through VLAN-interface 100.
- Enable BFD for RIP on VLAN-interface 100 of Switch A. When the link over VLAN-interface 100 fails, BFD can quickly detect the failure and notify RIP. RIP deletes the neighbor relationship and route information learned on VLAN-interface 100, and uses the route destined for 100.1.1.1 24 through VLAN-interface 200.

Figure 9 Network diagram



Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Configure basic RIP settings:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] rip 1
[SwitchA-rip-1] version 2
[SwitchA-rip-1] undo summary
[SwitchA-rip-1] network 192.168.1.0
[SwitchA-rip-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] rip bfd enable
[SwitchA-Vlan-interface100] quit
[SwitchA] rip 2
[SwitchA-rip-2] version 2
[SwitchA-rip-2] undo summary
[SwitchA-rip-2] network 192.168.2.0
[SwitchA-rip-2] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] rip 1
[SwitchB-rip-1] version 2
[SwitchB-rip-1] undo summary
[SwitchB-rip-1] network 192.168.2.0
[SwitchB-rip-1] network 192.168.3.0
[SwitchB-rip-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] rip 1
[SwitchC-rip-1] version 2
[SwitchC-rip-1] undo summary
[SwitchC-rip-1] network 192.168.1.0
[SwitchC-rip-1] network 192.168.3.0
[SwitchC-rip-1] import-route static
[SwitchC-rip-1] quit
```

3. Configure BFD parameters on VLAN-interface 100 of Switch A.

```
[SwitchA] bfd echo-source-ip 11.11.11.11
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] bfd min-echo-receive-interval 500
[SwitchA-Vlan-interface100] bfd detect-multiplier 7
[SwitchA-Vlan-interface100] quit
```

4. Configure a static route on Switch C.

```
[SwitchC] ip route-static 120.1.1.0 24 null 0
```

Verifying the configuration

Display the BFD session information on Switch A.

```
<SwitchA> display bfd session
 Total sessions: 1        Up sessions: 1        Init mode: Active

 IPv4 session working in echo mode:

 LD               SourceAddr      DestAddr        State  Holdtime    Interface
 4                192.168.1.1     192.168.1.2     Up     2000ms      Vlan100
```

Display RIP routes destined for 120.1.1.0/24 on Switch A.

```
<SwitchA> display ip routing-table 120.1.1.0 24

Summary count : 1

Destination/Mask   Proto   Pre Cost        NextHop         Interface
120.1.1.0/24       RIP     100 1           192.168.1.2     Vlan-interface100
```

The output shows that Switch A communicates with Switch C through VLAN-interface 100. Then the link over VLAN-interface 100 fails.

Display RIP routes destined for 120.1.1.0/24 on Switch A.

```
<SwitchA> display ip routing-table 120.1.1.0 24

Summary count : 1
```

```
Destination/Mask    Proto   Pre Cost        NextHop         Interface
120.1.1.0/24        RIP     100 1           192.168.2.2     Vlan-interface200
```

The output shows that Switch A communicates with Switch C through VLAN-interface 200.

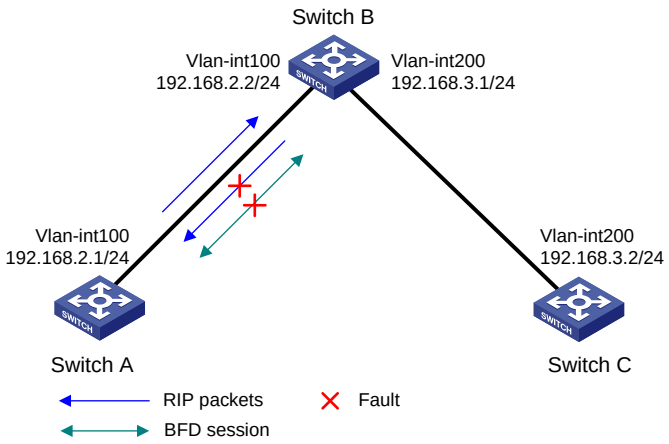
Example: Configuring BFD for RIP (single hop echo detection for a specific destination)

Network configuration

As shown in Figure 10, VLAN-interface 100 of Switch A and Switch B runs RIP process 1. VLAN-interface 200 of Switch B and Switch C runs RIP process 1.

- Configure a static route destined for 100.1.1.0/24 and enable static route redistribution into RIP on both Switch A and Switch C. This allows Switch B to learn two routes destined for 100.1.1.0/24 through VLAN-interface 100 and VLAN-interface 200. The route redistributed from Switch A has a smaller cost than that redistributed from Switch C, so Switch B uses the route through VLAN-interface 200.
- Enable BFD for RIP on VLAN-interface 100 of Switch A, and specify VLAN-interface 100 of Switch B as the destination. When a unidirectional link occurs between Switch A and Switch B, BFD can quickly detect the link failure and notify RIP. Switch B then deletes the neighbor relationship and the route information learned on VLAN-interface 100. It does not receive or send any packets from VLAN-interface 100. When the route learned from Switch A ages out, Switch B uses the route destined for 100.1.1.1 24 through VLAN-interface 200.

Figure 10 Network diagram



Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Configure basic RIP settings and enable BFD on the interfaces:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] rip 1
[SwitchA-rip-1] network 192.168.2.0
[SwitchA-rip-1] import-route static
[SwitchA-rip-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] rip bfd enable destination 192.168.2.2
```

```
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] rip 1
[SwitchB-rip-1] network 192.168.2.0
[SwitchB-rip-1] network 192.168.3.0
[SwitchB-rip-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] rip 1
[SwitchC-rip-1] network 192.168.3.0
[SwitchC-rip-1] import-route static cost 3
[SwitchC-rip-1] quit
```

3. Configure BFD parameters on VLAN-interface 100 of Switch A.

```
[SwitchA] bfd echo-source-ip 11.11.11.11
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] bfd min-echo-receive-interval 500
[SwitchA-Vlan-interface100] return
```

4. Configure static routes:

Configure a static route on Switch A.

```
[SwitchA] ip route-static 100.1.1.0 24 null 0
```

Configure a static route on Switch C.

```
[SwitchA] ip route-static 100.1.1.0 24 null 0
```

Verifying the configuration

Display BFD session information on Switch A.

```
<SwitchA> display bfd session
 Total sessions: 1        Up sessions: 1        Init mode: Active

 IPv4 session working in echo mode:

 LD               SourceAddr      DestAddr        State  Holdtime    Interface
 3                192.168.2.1     192.168.2.2     Up     2000ms      vlan100
```

Display routes destined for 100.1.1.0/24 on Switch B.

```
<SwitchB> display ip routing-table 100.1.1.0 24 verbose

Summary Count : 1

Destination: 100.1.1.0/24
   Protocol: RIP
 Process ID: 1
  SubProtID: 0x1                  Age: 00h02m47s
       Cost: 1              Preference: 100
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0x2                OrigAs: 0
      NibID: 0x12000002         LastAs: 0
```

```
     AttrID: 0xffffffff        Neighbor: 192.168.2.1
      Flags: 0x1008c        OrigNextHop: 192.168.2.1
      Label: NULL           RealNextHop: 192.168.2.1
    BkLabel: NULL             BkNextHop: N/A
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: vlan-interface 100
BkTunnel ID: Invalid        BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Display routes destined for 100.1.1.0/24 on Switch B when the link between Switch A and Switch B fails.

```
<SwitchB> display ip routing-table 100.1.1.0 24 verbose

Summary Count : 1

Destination: 100.1.1.0/24
   Protocol: RIP
 Process ID: 1
  SubProtID: 0x1                    Age: 00h21m23s
       Cost: 4               Preference: 100
      IpPre: N/A             QosLocalID: N/A
        Tag: 0                    State: Active Adv
  OrigTblID: 0x0                OrigVrf: default-vrf
    TableID: 0x2                 OrigAs: 0
      NibID: 0x12000002          LastAs: 0
     AttrID: 0xffffffff        Neighbor: 192.168.3.2
      Flags: 0x1008c        OrigNextHop: 192.168.3.2
      Label: NULL           RealNextHop: 192.168.3.2
    BkLabel: NULL             BkNextHop: N/A
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: vlan-interface 200
BkTunnel ID: Invalid        BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Example: Configuring BFD for RIP (bidirectional detection in BFD control packet mode)

Network configuration

As shown in Figure 11, VLAN-interface 100 of Switch A and VLAN-interface 200 of Switch C run RIP process 1.

VLAN-interface 300 of Switch A runs RIP process 2. VLAN-interface 400 of Switch C, and VLAN-interface 300 and VLAN-interface 400 of Switch D run RIP process 1.

- Configure a static route destined for 100.1.1.0/24 on Switch A.
- Configure a static route destined for 101.1.1.0/24 on Switch C.
- Enable static route redistribution into RIP on Switch A and Switch C. This allows Switch A to learn two routes destined for 100.1.1.0/24 through VLAN-interface 100 and VLAN-interface 300. It uses the route through VLAN-interface 100.
- Enable BFD on VLAN-interface 100 of Switch A and VLAN-interface 200 of Switch C.

When the link over VLAN-interface 100 fails, BFD can quickly detect the link failure and notify RIP. RIP deletes the neighbor relationship and the route information received learned on VLAN-interface 100. It uses the route destined for 100.1.1.0/24 through VLAN-interface 300.

Figure 11 Network diagram

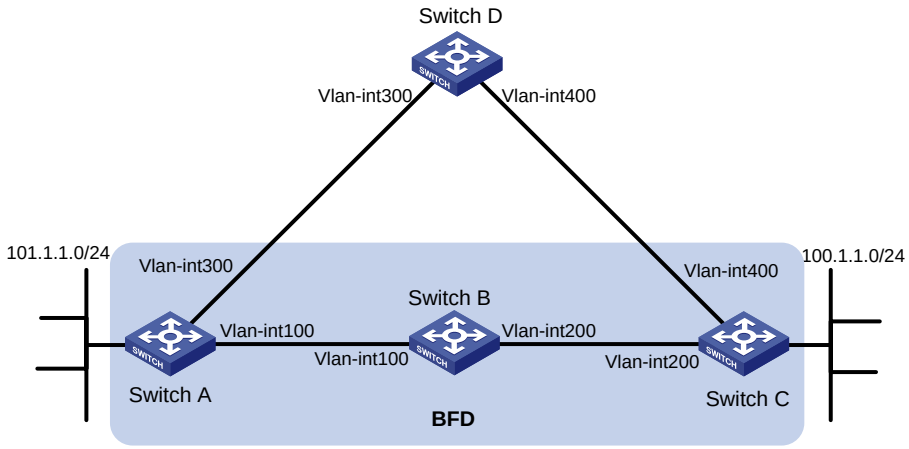


Table 1 Interface and IP address assignment

Device	Interface	IP address
Switch A	VLAN-interface 300	192.168.3.1/24
Switch A	VLAN-interface 100	192.168.1.1/24
Switch B	VLAN-interface 100	192.168.1.2/24
Switch B	VLAN-interface 200	192.168.2.1/24
Switch C	VLAN-interface 200	192.168.2.2/24
Switch C	VLAN-interface 400	192.168.4.2/24
Switch D	VLAN-interface 300	192.168.3.2/24
Switch D	VLAN-interface 400	192.168.4.1/24

Procedure

1. Configure IP addresses for the interfaces. (Details not shown.)
2. Configure basic RIP settings and enable static route redistribution into RIP so Switch A and Switch C have routes to send to each other:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] rip 1
   [SwitchA-rip-1] version 2
   ```

```
[SwitchA-rip-1] undo summary
[SwitchA-rip-1] network 192.168.1.0
[SwitchA-rip-1] network 101.1.1.0
[SwitchA-rip-1] peer 192.168.2.2
[SwitchA-rip-1] undo validate-source-address
[SwitchA-rip-1] import-route static
[SwitchA-rip-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] rip bfd enable
[SwitchA-Vlan-interface100] quit
[SwitchA] rip 2
[SwitchA-rip-2] version 2
[SwitchA-rip-2] undo summary
[SwitchA-rip-2] network 192.168.3.0
[SwitchA-rip-2] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] rip 1
[SwitchC-rip-1] version 2
[SwitchC-rip-1] undo summary
[SwitchC-rip-1] network 192.168.2.0
[SwitchC-rip-1] network 192.168.4.0
[SwitchC-rip-1] network 100.1.1.0
[SwitchC-rip-1] peer 192.168.1.1
[SwitchC-rip-1] undo validate-source-address
[SwitchC-rip-1] import-route static
[SwitchC-rip-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] rip bfd enable
[SwitchC-Vlan-interface200] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] rip 1
[SwitchD-rip-1] version 2
[SwitchD-rip-1] undo summary
[SwitchD-rip-1] network 192.168.3.0
[SwitchD-rip-1] network 192.168.4.0
```

3. Configure BFD parameters:

Configure Switch A.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] bfd min-transmit-interval 500
[SwitchA-Vlan-interface100] bfd min-receive-interval 500
[SwitchA-Vlan-interface100] bfd detect-multiplier 7
[SwitchA-Vlan-interface100] quit
```

Configure Switch C.

```
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] bfd min-transmit-interval 500
[SwitchC-Vlan-interface200] bfd min-receive-interval 500
```

```
[SwitchC-Vlan-interface200] bfd detect-multiplier 7
[SwitchC-Vlan-interface200] quit
```

4. Configure static routes:

   # Configure a static route to Switch C on Switch A.

   ```
   [SwitchA] ip route-static 192.168.2.0 24 vlan-interface 100 192.168.1.2
   ```

   # Configure a static route to Switch A on Switch C.

   ```
   [SwitchC] ip route-static 192.168.1.0 24 vlan-interface 200 192.168.2.1
   ```

Verifying the configuration

Display the BFD session information on Switch A.

```
<SwitchA> display bfd session
 Total sessions: 1        Up sessions: 1        Init mode: Active

 IPv4 session working in control packet mode:

 LD/RD            SourceAddr      DestAddr        State  Holdtime    Interface
 513/513          192.168.1.1     192.168.2.2     Up     1700ms      vlan100
```

Display RIP routes destined for 100.1.1.0/24 on Switch A.

```
<SwitchA> display ip routing-table 100.1.1.0 24

Summary count : 1

Destination/Mask   Proto   Pre Cost        NextHop         Interface
100.1.1.0/24       RIP     100 1           192.168.2.2     vlan-interface 100
```

The output shows that Switch A communicates with Switch C through VLAN-interface 100. Then the link over VLAN-interface 100 fails.

Display RIP routes destined for 100.1.1.0/24 on Switch A.

```
<SwitchA> display ip routing-table 100.1.1.0 24

Summary count : 1

Destination/Mask   Proto   Pre Cost        NextHop         Interface
100.1.1.0/24       RIP     100 2           192.168.3.2     vlan-interface 300
```

The output shows that Switch A communicates with Switch C through VLAN-interface 300.

Example: Configuring RIP FRR

Network configuration

As shown in Figure 12, Switch A, Switch B, and Switch C run RIPv2. Configure RIP FRR so that when Link A becomes unidirectional, services can be switched to Link B immediately.

Figure 12 Network diagram

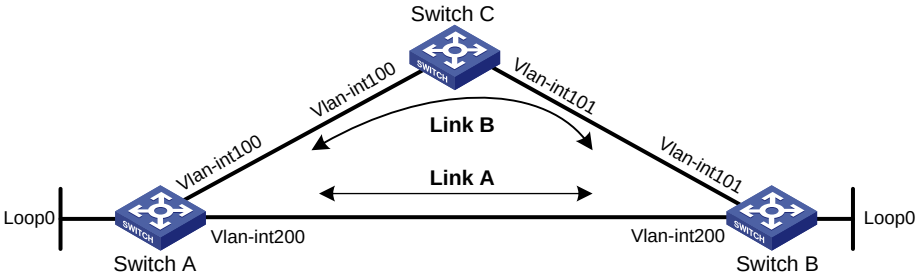


Table 2 Interface and IP address assignment

Device	Interface	IP address
Switch A	VLAN-interface 100	12.12.12.1/24
Switch A	VLAN-interface 200	13.13.13.1/24
Switch A	Loopback 0	1.1.1.1/32
Switch B	VLAN-interface 101	24.24.24.4/24
Switch B	VLAN-interface 200	13.13.13.2/24
Switch B	Loopback 0	4.4.4.4/32
Switch C	VLAN-interface 100	12.12.12.2/24
Switch C	VLAN-interface 101	24.24.24.2/24

Procedure

1. Configure IP addresses and subnet masks for the interfaces on the switches. (Details not shown.)
2. Configure RIPv2 on the switches to make sure Switch A, Switch B, and Switch C can communicate with each other at Layer 3. (Details not shown.)
3. Configure RIP FRR:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ip prefix-list abc index 10 permit 4.4.4.4 32
[SwitchA] route-policy frr permit node 10
[SwitchA-route-policy-frr-10] if-match ip address prefix-list abc
[SwitchA-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
100 backup-nexthop 12.12.12.2
[SwitchA-route-policy-frr-10] quit
[SwitchA] rip 1
[SwitchA-rip-1] fast-reroute route-policy frr
[SwitchA-rip-1] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ip prefix-list abc index 10 permit 1.1.1.1 32
[SwitchB] route-policy frr permit node 10
[SwitchB-route-policy-frr-10] if-match ip address prefix-list abc
[SwitchB-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
101 backup-nexthop 24.24.24.2
```

```
[SwitchB-route-policy-frr-10] quit
[SwitchB] rip 1
[SwitchB-rip-1] fast-reroute route-policy frr
[SwitchB-rip-1] quit
```

Verifying the configuration

Display route 4.4.4.4/32 on Switch A to view the backup next hop information.

```
[SwitchA] display ip routing-table 4.4.4.4 verbose

Summary Count : 1

Destination: 4.4.4.4/32
   Protocol: RIP
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 1               Preference: 100
      IpPre: N/A             QosLocalID: N/A
        Tag: 0                    State: Active Adv
  OrigTblID: 0x0                OrigVrf: default-vrf
    TableID: 0x2                 OrigAs: 0
      NibID: 0x26000002          LastAs: 0
     AttrID: 0xffffffff        Neighbor: 13.13.13.2
      Flags: 0x1008c        OrigNextHop: 13.13.13.2
      Label: NULL           RealNextHop: 13.13.13.2
    BkLabel: NULL             BkNextHop: 12.12.12.2
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface200
BkTunnel ID: Invalid        BkInterface: Vlan-interface100
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Display route 1.1.1.1/32 on Switch B to view the backup next hop information.

```
[SwitchB] display ip routing-table 1.1.1.1 verbose

Summary Count : 1

Destination: 1.1.1.1/32
   Protocol: RIP
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 1               Preference: 100
      IpPre: N/A             QosLocalID: N/A
        Tag: 0                    State: Active Adv
  OrigTblID: 0x0                OrigVrf: default-vrf
    TableID: 0x2                 OrigAs: 0
      NibID: 0x26000002          LastAs: 0
     AttrID: 0xffffffff        Neighbor: 13.13.13.1
```

```
      Flags: 0x1008c        OrigNextHop: 13.13.13.1
      Label: NULL           RealNextHop: 13.13.13.1
    BkLabel: NULL             BkNextHop: 24.24.24.2
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface200
BkTunnel ID: Invalid        BkInterface: Vlan-interface101
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0                MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Contents

Configuring OSPF 1

About OSPF 1
OSPF features 1
OSPF packets 1
LSA types 1
OSPF areas 2
Router types 4
Route types 5
Router ID 6
Route calculation 6
OSPF network types 6
DR and BDR 7
OSPF state machines 8
Protocols and standards 11
Restrictions and guidelines: OSPF configuration 12
OSPF tasks at a glance 12
Configuring basic OSPF functions 14
Enabling an OSPF process 14
Creating an OSPF area 14
Enabling OSPF 15
Enabling OSPF to establish neighbor relationships through the secondary IP addresses of an interface 16
Establishing OSPF neighbor relationships on an M-LAG network 17
Configuring OSPF stub and NSSA areas 18
About OSPF stub and NSSA area configuration 18
Configuring a stub area 18
Configuring an NSSA area 19
Configuring a virtual link 20
Configuring OSPF network types 20
Restrictions and guidelines for configuring OSPF network types 20
Configuring the broadcast network type for an interface 20
Configuring the NBMA network type for an interface 21
Configuring the P2MP network type for an interface 21
Configuring the P2P network type for an interface 22
Configuring OSPF route control 22
Configuring OSPF inter-area route summarization 22
Configuring redistributed route summarization 23
Configuring received OSPF route filtering 23
Configuring Type-3 LSA filtering 24
Setting an OSPF cost for an interface 24
Enabling OSPF to advertise the maximum link cost to neighbors 25
Associating OSPF with Track 25
Setting the maximum number of ECMP routes 27
Setting OSPF preference 28
Configuring discard routes for summary networks 28
Redistributing routes from another routing protocol 29
Redistributing a default route 30
Advertising a host route 31
Advertising OSPF link state information to BGP 31
Setting OSPF timers 32
About setting OSPF timers 32
Configuring OSPF packet timers 32
Setting the LSA transmission delay 33
Setting the OSPF route calculation interval 33
Setting the minimum LSA arrival interval 34
Setting the LSA update interval 34
Setting the OSPF exit overflow interval 35

Configuring OSPF packet parameters ····36
Disabling interfaces from receiving and sending OSPF packets ····36
Adding the interface MTU into DD packets ····36
Setting the DSCP value for outgoing OSPF packets ····37
Setting the maximum length of OSPF packets that can be sent by an interface ····37
Setting the LSU transmit rate ····37
Controlling LSA generation, advertisement, and reception ····38
Setting the maximum number of external LSAs in LSDB ····38
Filtering outbound LSAs on an interface ····38
Filtering LSAs for the specified neighbor ····39
Accelerating OSPF convergence speed ····39
Enabling OSPF ISPF ····39
Configuring prefix suppression ····39
Configuring prefix prioritization ····40
Configuring OSPF PIC ····41
Configuring advanced OSPF features ····42
Configuring stub routers ····42
Configuring OSPF isolation ····43
Configuring OSPF shutdown ····44
Enabling compatibility with RFC 1583 ····45
Configuring OSPF GR ····45
About OSPF GR ····45
Configuring OSPF GR restarter ····46
Configuring OSPF GR helper ····47
Triggering OSPF GR ····48
Configuring OSPF NSR ····48
Configuring BFD for OSPF ····49
About BFD for OSPF ····49
Restrictions and guidelines ····49
Procedure ····49
Configuring single-hop echo detection ····49
Configuring OSPF FRR ····50
About OSPF FRR ····50
Restrictions and guidelines for OSPF FRR ····50
Configuring OSPF FRR to use the LFA algorithm to calculate a backup next hop ····50
Configuring OSPF FRR to use a backup next hop specified in a routing policy ····51
Configuring BFD for OSPF FRR ····51
Configuring BFD echo packet mode for OSPF FRR ····51
Configuring OSPF authentication ····52
About OSPF area and interface authentication ····52
Restrictions and guidelines for configuring OSPF authentication ····52
Configuring OSPF area authentication ····52
Configuring OSPF interface authentication ····53
Configuring GTSM for OSPF ····53
About GTSM ····53
Restrictions and guidelines for GTSM ····54
Configuring GTSM in OSPF area view ····54
Configuring GTSM in interface view ····54
Configuring OSPF logging, SNMP notifications, and troubleshooting ····55
Logging neighbor state changes ····55
Configuring the OSPF logging feature ····55
Configuring OSPF network management ····55
Setting the maximum number of OSPF neighbor relationship troubleshooting entries ····56
Display and maintenance commands for OSPF ····56
OSPF configuration examples ····58
Example: Configuring basic OSPF ····58
Example: Configuring OSPF route redistribution ····61
Example: Configuring OSPF route summarization ····63
Example: Configuring OSPF stub area ····66
Example: Configuring OSPF NSSA area ····68
Example: Configuring OSPF DR election ····70
Example: Configuring OSPF virtual link ····74

Example: Configuring OSPF GR 76
Example: Configuring OSPF NSR 78
Example: Configuring BFD for OSPF 81
Example: Configuring OSPF FRR 84
Troubleshooting OSPF configuration 86
No OSPF neighbor relationship established 86
Incorrect routing information 87

Configuring OSPF

About OSPF

Open Shortest Path First (OSPF) is a link-state IGP developed by the OSPF working group of the IETF. OSPF version 2 is used for IPv4. OSPF refers to OSPFv2 throughout this chapter.

OSPF features

OSPF has the following features:

- **Wide scope**—Supports multiple network sizes and several hundred routers in an OSPF routing domain.
- **Fast convergence**—Advertises routing updates instantly upon network topology changes.
- **Loop free**—Computes routes with the SPF algorithm to avoid routing loops.
- **Area-based network partition**—Splits an AS into multiple areas to facilitate management. This feature reduces the LSDB size on routers to save memory and CPU resources, and reduces route updates transmitted between areas to save bandwidth.
- **ECMP routing**—Supports multiple equal-cost routes to a destination.
- **Routing hierarchy**—Supports a 4-level routing hierarchy that prioritizes routes into intra-area, inter-area, external Type-1, and external Type-2 routes.
- **Authentication**—Supports area- and interface-based packet authentication to ensure secure packet exchange.
- **Support for multicasting**—Multicasts protocol packets on some types of links to avoid impacting other devices.

OSPF packets

OSPF messages are carried directly over IP. The protocol number is 89.

OSPF uses the following packet types:

- **Hello**—Periodically sent to find and maintain neighbors, containing timer values, information about the DR, BDR, and known neighbors.
- **Database description (DD)**—Describes the digest of each LSA in the LSDB, exchanged between two routers for data synchronization.
- **Link state request (LSR)**—Requests needed LSAs from a neighbor. After exchanging the DD packets, the two routers know which LSAs of the neighbor are missing from their LSDBs. They then exchange LSR packets requesting the missing LSAs. LSR packets contain the digest of the missing LSAs.
- **Link state update (LSU)**—Transmits the requested LSAs to the neighbor.
- **Link state acknowledgment (LSAck)**—Acknowledges received LSU packets. It contains the headers of received LSAs (an LSAck packet can acknowledge multiple LSAs).

LSA types

OSPF advertises routing information in Link State Advertisements (LSAs). The following LSAs are commonly used:

- **Router LSA**—Type-1 LSA, originated by all routers and flooded throughout a single area only. This LSA describes the collected states of the router's interfaces to an area.
- **Network LSA**—Type-2 LSA, originated for broadcast and NBMA networks by the designated router, and flooded throughout a single area only. This LSA contains the list of routers connected to the network.
- **Network Summary LSA**—Type-3 LSA, originated by Area Border Routers (ABRs), and flooded throughout the LSA's associated area. Each summary-LSA describes a route to a destination outside the area, yet still inside the AS (an inter-area route).
- **ASBR Summary LSA**—Type-4 LSA, originated by ABRs and flooded throughout the LSA's associated area. Type 4 summary-LSAs describe routes to Autonomous System Boundary Router (ASBR).
- **AS External LSA**—Type-5 LSA, originated by ASBRs, and flooded throughout the AS (except stub and NSSA areas). Each AS-external-LSA describes a route to another AS.
- **NSSA LSA**—Type-7 LSA, as defined in RFC 1587, originated by ASBRs in NSSAs and flooded throughout a single NSSA. NSSA LSAs describe routes to other ASs.
- **Opaque LSA**—LSA for OSPF extensions. Its format consists of a standard LSA header and application specific information. The opaque LSA includes Type 9, Type 10, and Type 11. The Type 9 opaque LSA is flooded into the local subnet. Grace LSA, used by graceful restart, is Type 9 LSA. The Type 10 is flooded into the local area. The Type 11 is flooded throughout the AS.

OSPF areas

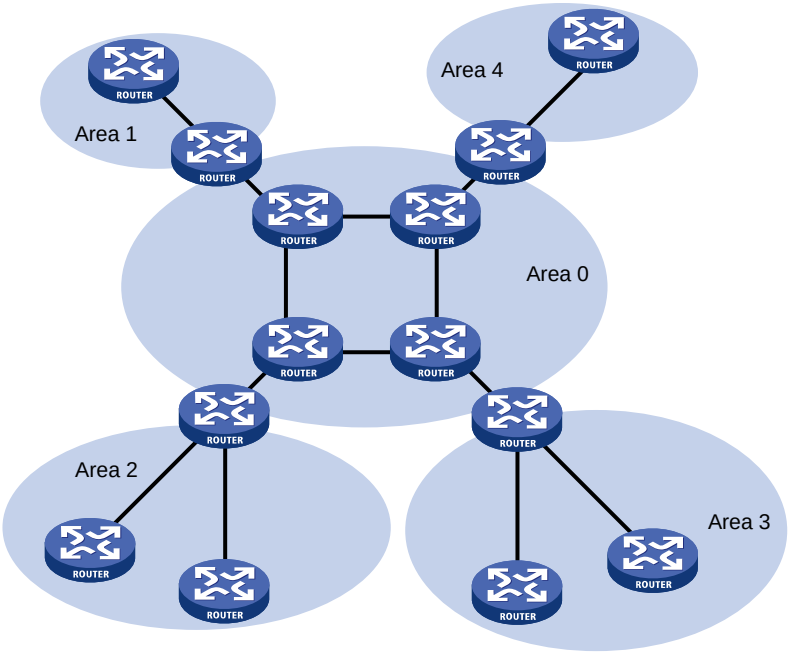
Area-based OSPF network partition

In large OSPF routing domains, SPF route computations consume too many storage and CPU resources, and enormous OSPF packets generated for route synchronization occupy excessive bandwidth.

To resolve these issues, OSPF splits an AS into multiple areas. Each area is identified by an area ID. The boundaries between areas are routers rather than links. A network segment (or a link) can only reside in one area as shown in Figure 1.

You can configure route summarization on ABRs to reduce the number of LSAs advertised to other areas and minimize the effect of topology changes.

Figure 1 Area-based OSPF network partition



Backbone area

Each AS has a backbone area that distributes routing information between non-backbone areas. Routing information between non-backbone areas must be forwarded by the backbone area. OSPF has the following requirements:

- All non-backbone areas must maintain connectivity to the backbone area.
- The backbone area must maintain connectivity within itself.

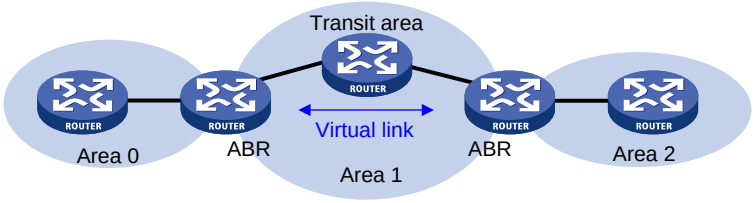
In practice, these requirements might not be met due to lack of physical links. OSPF virtual links can solve this issue.

Virtual links

A virtual link is established between two ABRs through a non-backbone area. It must be configured on both ABRs to take effect. The non-backbone area is called a transit area.

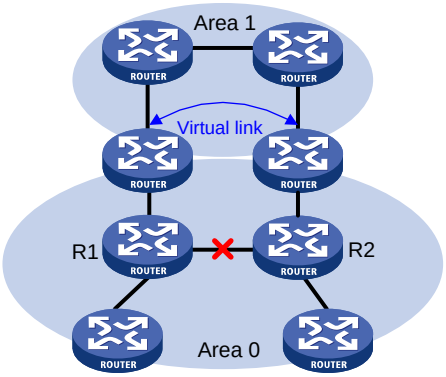
As shown in Figure 2, Area 2 has no direct physical link to the backbone Area 0. You can configure a virtual link between the two ABRs to connect Area 2 to the backbone area.

Figure 2 Virtual link application 1



Virtual links can also be used as redundant links. If a physical link failure breaks the internal connectivity of the backbone area, you can configure a virtual link to replace the failed physical link, as shown in Figure 3.

Figure 3 Virtual link application 2



The virtual link between the two ABRs acts as a point-to-point connection. You can configure interface parameters, such as hello interval, on the virtual link as they are configured on a physical interface.

The two ABRs on the virtual link unicast OSPF packets to each other, and the OSPF routers in between convey these OSPF packets as normal IP packets.

Stub area and totally stub area

A stub area does not distribute Type-5 LSAs to reduce the routing table size and LSAs advertised within the area. The ABR of the stub area advertises a default route in a Type-3 LSA so that the routers in the area can reach external networks through the default route.

To further reduce the routing table size and advertised LSAs, you can configure the stub area as a totally stub area. The ABR of a totally stub area does not advertise inter-area routes or external routes. It advertises a default route in a Type-3 LSA so that the routers in the area can reach external networks through the default route.

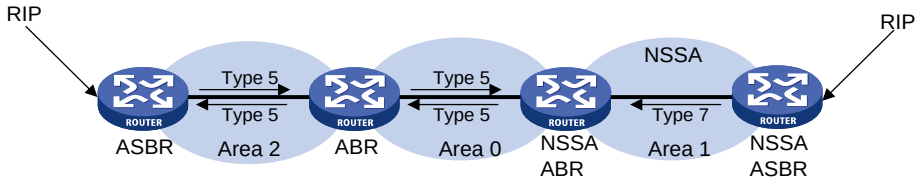
NSSA area and totally NSSA area

An NSSA area does not import AS external LSAs (Type-5 LSAs) but can import Type-7 LSAs generated by the NSSA ASBR. The NSSA ABR translates Type-7 LSAs into Type-5 LSAs and advertises the Type-5 LSAs to other areas.

As shown in Figure 4, the OSPF AS contains Area 1, Area 2, and Area 0. The other two ASs run RIP. Area 1 is an NSSA area where the ASBR redistributes RIP routes in Type-7 LSAs into Area 1. Upon receiving the Type-7 LSAs, the NSSA ABR translates them to Type-5 LSAs, and advertises the Type-5 LSAs to Area 0.

The ASBR of Area 2 redistributes RIP routes in Type-5 LSAs into the OSPF routing domain. However, Area 1 does not receive Type-5 LSAs because it is an NSSA area.

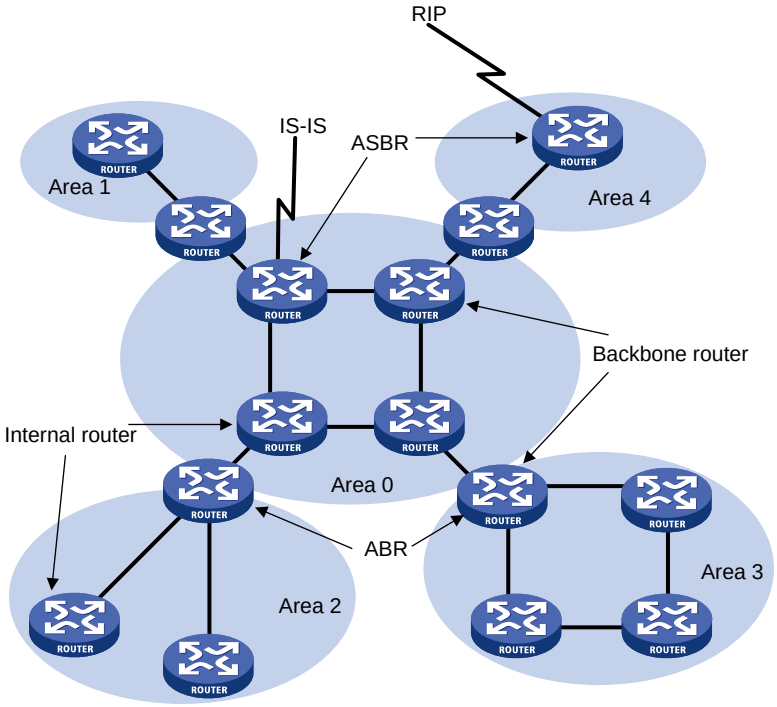
Figure 4 NSSA area



Router types

As shown in Figure 5, OSPF routers are classified into different types, including internal routers, ABRs, backbone routers, and ASBRs.

Figure 5 OSPF router types



Internal router

All interfaces on an internal router belong to one OSPF area.

ABR

An ABR belongs to more than two areas, one of which must be the backbone area. ABR connects the backbone area to a non-backbone area. An ABR and the backbone area can be connected through a physical or logical link.

Backbone router

No less than one interface of a backbone router must reside in the backbone area. All ABRs and internal routers in Area 0 are backbone routers.

ASBR

An ASBR exchanges routing information with another AS. An ASBR might not reside on the border of the AS. It can be an internal router or an ABR.

Route types

OSPF prioritizes routes into the following route levels:

- Intra-area route.
- Inter-area route.
- Type-1 external route.
- Type-2 external route.

The intra-area and inter-area routes describe the network topology of the AS. The external routes describe routes to external ASs.

A Type-1 external route has high credibility. The cost of a Type-1 external route = the cost from the router to the corresponding ASBR + the cost from the ASBR to the destination of the external route.

A Type-2 external route has low credibility. OSPF considers that the cost from the ASBR to the destination of a Type-2 external route is much greater than the cost from the ASBR to an OSPF internal router. The cost of a Type-2 external route = the cost from the ASBR to the destination of the Type-2 external route. If two Type-2 routes to the same destination have the same cost, OSPF takes the cost from the router to the ASBR into consideration to determine the best route.

Router ID

A router ID uniquely identifies a router in an AS. For a router to run OSPF, it must have a router ID. You can choose to manually specify a router ID or use the global router ID for an OSPF process.

Manual configuration

When you create an OSPF process, you can manually specify a router ID. To make sure the router ID is unique in the AS, you can specify the IP address of an interface on the router as the router ID.

Using the global router ID

If you do not specify a router ID when creating an OSPF process, the global router ID is used. As a best practice, manually specify a router ID or enable the OSPF process to automatically obtain a router ID when you create the OSPF process.

Route calculation

OSPF computes routes in an area as follows:

- Each router generates LSAs based on the network topology around itself, and sends them to other routers in update packets.
- Each OSPF router collects LSAs from other routers to compose an LSDB. An LSA describes the network topology around a router, and the LSDB describes the entire network topology of the area.
- Each router transforms the LSDB to a weighted directed graph that shows the topology of the area. All the routers within the area have the same graph.
- Each router uses the SPF algorithm to compute a shortest path tree that shows the routes to the nodes in the area. The router itself is the root of the tree.

OSPF network types

OSPF classifies networks into the following types, depending on different link layer protocols:

- **Broadcast**—If the link layer protocol is Ethernet or FDDI, OSPF considers the network type as broadcast by default. On a broadcast network, hello, LSU, and LSAck packets are multicast to 224.0.0.5 that identifies all OSPF routers or to 224.0.0.6 that identifies the DR and BDR. DD packets and LSR packets are unicast.
- **NBMA**—If the link layer protocol is Frame Relay, ATM, or X.25, OSPF considers the network type as NBMA by default. OSPF packets are unicast on an NBMA network.
- **P2MP**—No link is P2MP type by default. P2MP must be a conversion from other network types such as NBMA. On a P2MP network, OSPF packets are multicast to 224.0.0.5.
- **P2P**—If the link layer protocol is PPP or HDLC, OSPF considers the network type as P2P. On a P2P network, DD packets are unicast and other OSPF packets are multicast to 224.0.0.5.

The following are the differences between NBMA and P2MP networks:

- NBMA networks are fully meshed. P2MP networks are not required to be fully meshed.
- NBMA networks require DR and BDR election. P2MP networks do not have DR or BDR.

- On an NBMA network, OSPF packets are unicast, and neighbors are manually configured. On a P2MP network, OSPF packets are multicast by default, and you can configure OSPF to unicast protocol packets.

DR and BDR

DR and BDR mechanism

On a broadcast or NBMA network, any two routers must establish an adjacency to exchange routing information with each other. If n routers are present on the network, n(n-1)/2 adjacencies are established. Any topology change on the network results in an increase in traffic for route synchronization, which consumes a large amount of system and bandwidth resources.

Using the DR and BDR mechanisms can solve this problem.

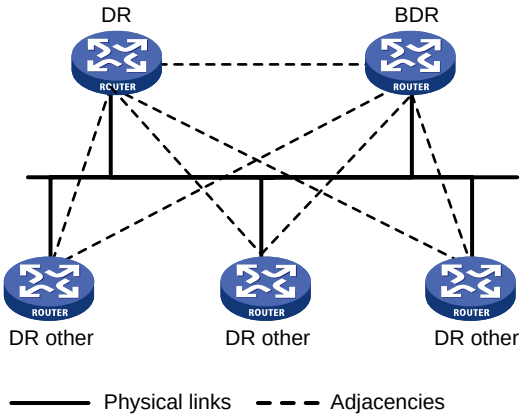
- **DR**—Elected to advertise routing information among other routers. If the DR fails, routers on the network must elect another DR and synchronize information with the new DR. Using this mechanism without BDR is time-consuming and is prone to route calculation errors.
- **BDR**—Elected along with the DR to establish adjacencies with all other routers. If the DR fails, the BDR immediately becomes the new DR, and other routers elect a new BDR.

Routers other than the DR and BDR are called DR Others. They do not establish adjacencies with one another, so the number of adjacencies is reduced.

The role of a router is subnet (or interface) specific. It might be a DR on one interface and a BDR or DR Other on another interface.

As shown in Figure 6, solid lines are Ethernet physical links, and dashed lines represent OSPF adjacencies. With the DR and BDR, only seven adjacencies are established.

Figure 6 DR and BDR in a network



NOTE:

In OSPF, neighbor and adjacency are different concepts. After startup, OSPF sends a hello packet on each OSPF interface. A receiving router checks parameters in the packet. If the parameters match its own, the receiving router considers the sending router an OSPF neighbor. Two OSPF neighbors establish an adjacency relationship after they synchronize their LSDBs through exchange of DD packets and LSAs.

DR and BDR election

DR election is performed on broadcast or NBMA networks but not on P2P and P2MP networks.

Routers in a broadcast or NBMA network elect the DR and BDR by router priority and ID. Routers with a router priority value higher than 0 are candidates for DR and BDR election.

The election votes are hello packets. Each router sends the DR elected by itself in a hello packet to all the other routers. If two routers on the network declare themselves as the DR, the router with the higher router priority wins. If router priorities are the same, the router with the higher router ID wins.

If a router with a higher router priority becomes active after DR and BDR election, the router cannot replace the DR or BDR until a new election is performed. Therefore, the DR of a network might not be the router with the highest priority, and the BDR might not be the router with the second highest priority.

OSPF state machines

Interface state machine

On receipt of link state information, OSPF establishes adjacency relationships with neighboring devices, and then exchanges LSAs with them. The state of an OSPF interface indicates the role of the device in an OSPF link. Two neighboring devices can correctly form an adjacency relationship by checking each other's interface state.

The interface state machine involves the following interface states:

- **Down**—Initial interface state.

  An OSPF interface in this state cannot be used for traffic sending or receiving.
- **Loopback**—The network-attached interface is looped back on the device.

  A loopback interface is not available for regular data transmission. However, to obtain link quality information on the interface, you might either send ICMP ping packets to the loopback interface or have a bit error test. To meet this potential requirement, each OSPF device advertises loopback interface information in Router LSAs.
- **Waiting**—The device is determining the DR and BDR for the network.

  When the device does not run for the DR or BDR, the interface starts a wait timer. The hello packets sent by the device does not contain any DR or BDR information unless the wait timer expires. The device cannot be elected as the DR or BDR on the network until it exits from **Waiting** state. The wait timer mechanism prevents unnecessary DR and BDR changes.

  This interface state is supported only on NBMA networks and broadcast networks.
- **Point-to-point**—The interface is connected to a physical P2P network or to a virtual link.

  If an interface enters this state, it attempts to establish an adjacency relationship with the neighboring device.

  This interface state is supported only on P2P networks and P2MP networks.
- **DR Other**—The interface is connected to a broadcast or NBMA network on which another device has been elected as the DR.

  A device in **DR Other** state is neither elected as the DR nor the BDR. When DR and BDR election is complete on the network, each DR Other device establishes adjacencies with both the DR and the BDR.
- **BDR**—The device is the BDR on the attached network.

  As the BDR, the device establishes adjacencies with all other devices on the attached network. It will immediately become the new DR if the existing DR fails.
- **DR**—The device is the DR on the attached network.

  As the DR, the device establishes adjacencies with all other devices on the attached network.

A number of input events (IE) can cause OSPF interface state changes. As shown in Figure 7, the relationships between IEs and state changes form the OSPF interface state machine.

Figure 7 OSPF interface state machine

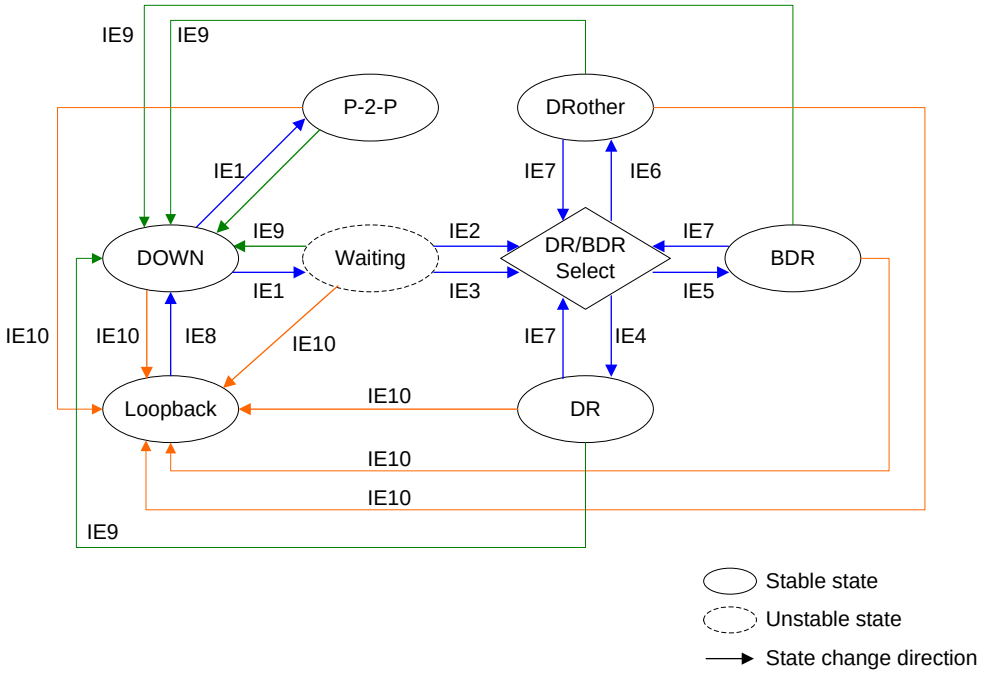


Table 1 shows the IEs that can cause interface state changes.

Table 1 Input events

Input event	Description
IE1	InterfaceUp event: Lower-level protocols have indicated that the interface is operational.
IE2	WaitTimer event: The wait timer has expired and the device can run for the DR or the BDR.
IE3	BackupSeen event: The device has detected the existence or non-existence of a BDR on the attached network through one of the following methods: • The interface receives a hello packet from a neighbor that claims itself to be the BDR. • The interface receives a hello packet from a neighbor that claims itself to be the DR. The hello packet also indicates that there is no BDR. This event marks an end to **Waiting** state and indicates that the device and the neighbor can have bidirectional communication.
IE4	The interface is elected as the DR.
IE5	The interface is elected as the BDR.
IE6	The interface is neither elected as the DR nor the BDR.
IE7	NeighborChange event: A neighbor associated with the interface has a change. The DR and BDR need to be re-elected. The following neighbor changes might lead to the re-election of DR and BDR: • The device can have bidirectional communication with a neighbor. • The device no longer has bidirectional communication with a neighbor. • A neighbor claims itself as DR or BDR. The device detects this event again by examining the hello packets from the neighbor. • A neighbor no longer claims itself as DR or BDR. The device detects this event again by examining the hello packets from the neighbor. • The DR priority for a neighbor has changed. The device detects this event again by

	examining the hello packets from the neighbor.
IE8	UnloopInd event: Network management or lower-level protocols has indicated that the interface is no longer looped back.
IE9	InterfaceDown event: Lower-level protocols has indicated that the interface is not operational. If this event occurs, the interface state might change to Down.
IE10	LoopInd event: Network management or lower-level protocols has indicated that the interface is looped back. If this event occurs, the interface state might change to Loopback.

Neighbor state machine

The process of OSPF adjacency establishment involves a series of neighbor state changes that transition from **Down** state to **Full** state.

The neighbor state machine involves the following neighbor states:

- **Down**—Initial state of a neighbor conversation.

  This state indicates that the local device does not receive any hello packets from the neighbor within the neighbor dead interval. An OSPF device sends hello packets at intervals of PollInterval.

  Whether an OSPF device sends hello packets to neighbors in **Down** state depends on its network type. The device sends hello packets to neighbors in **Down** state only when its network type is NBMA.
- **Attempt**—In this state, the local device attempts to establish neighbor relationships with the manually specified neighbors by sending hello packets at intervals of HelloInterval.

  This state is valid only for the neighbors on NBMA networks.
- **Init**—This state indicates that the local device has received a hello packet from a neighbor before the neighbor dead interval elapses. However, the local device has not established bidirectional communication with the neighbor, because the received hello packet does not contain the local device's router ID.
- **2-Way**—This state indicates that the local device has established a neighbor relationship with a neighboring device. Each of the two devices has received a hello packet from another and has seen its own router ID in the received hello packet.

  If the neighbor relationship between two devices does not prompt to an adjacency, the neighbor state will remain 2-Way.

  DR and BDR election starts only when the neighbor state is 2-Way or greater.
- **ExStart**—This state indicates that the local device and the neighbor is deciding which is the master. This step is to determine the sequence number for initial DD packets. The devices can then exchange DD packets in order.
- **Exchange**—This state indicates that the local device is exchanging DD packets with the neighbor. The local device describes its LSDB information in DD packets and sends them to the neighbor.
- **Loading**—This state indicates that the local device and the neighbor are having bidirectional LSDB synchronization. They send LSRs to each other to request the most recent LSAs, and then synchronize their LSDBs through exchanges of LSUs and LSAcks.
- **Full**—This state indicates that the local device and the neighbor has completed bidirectional LSDB synchronization and they are fully adjacent.

As shown in Figure 8, the relationships between IEs and state changes form the OSPF neighbor state machine.

Figure 8 OSPF neighbor state machine

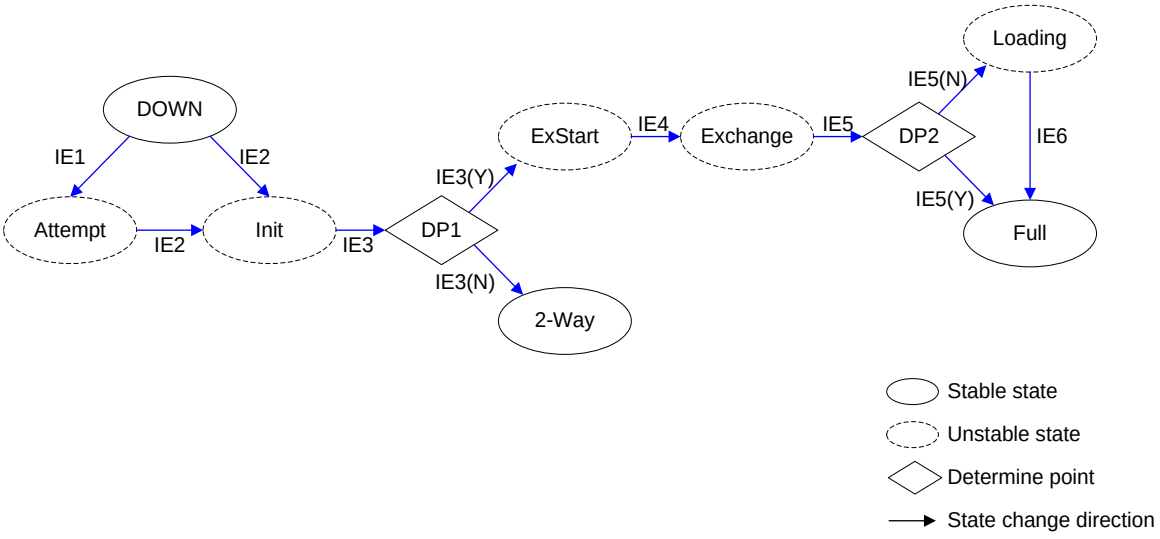


Table 2 shows the IEs that can cause neighbor state changes.

Table 2 Input events

Input event	Description
IE1	Start event: The OSPF device attempts to establish a neighbor relationship with the neighbor by sending hello packets at intervals of HelloInterval. This event is supported only on NBMA networks.
IE2	HelloReceived event: The OSPF device has received a hello packet from the neighbor.
IE3	2-WayReceived event: The OSPF device has received a hello packet from the neighbor and has seen its router ID in the received hello packet. The two neighboring devices have established bidirectional communication. The OSPF device then performs the following task: • If the OSPF device needs to form an adjacency with the neighbor, it changes the neighbor state to ExStart. • If the OSPF device should not form an adjacency with the neighbor, it changes the neighbor state to 2-Way.
IE4	NegotiationDone event: The OSPF device and the neighbor have completed master/secondary relationship negotiation and have determined the sequence number for initial DD packets.
IE5	ExchangeDone event: The OSPF device have exchanged DD packets with the neighbor. The OSPF device then performs the following task: • If the LSR list is empty, the device changes the neighbor state to Full. This neighbor state indicates that bidirectional LSDB synchronization is complete and the device has established an adjacency with the neighbor. • If the LSR list is not empty, the device changes the neighbor state to Loading, and then sends LSRs to the neighbor to request missing link state information.
IE6	LoadingDone event: The LSR list is already empty.

Protocols and standards

- RFC 1245, *OSPF protocol analysis*
- RFC 1246, *Experience with the OSPF protocol*

- RFC 1370, *Applicability Statement for OSPF*
- RFC 1403, *BGP OSPF Interaction*
- RFC 1745, *BGP4/IDRP for IP---OSPF Interaction*
- RFC 1765, *OSPF Database Overflow*
- RFC 1793, *Extending OSPF to Support Demand Circuits*
- RFC 2154, *OSPF with Digital Signatures*
- RFC 2328, *OSPF Version 2*
- RFC 3101, *OSPF Not-So-Stubby Area (NSSA) Option*
- RFC 3166, *Request to Move RFC 1403 to Historic Status*
- RFC 3509, *Alternative Implementations of OSPF Area Border Routers*
- RFC 3623, *Graceful OSPF Restart*
- RFC 4167, *Graceful OSPF Restart Implementation Report*
- RFC 4750, *OSPF Version 2 Management Information Base*
- RFC 4811, *OSPF Out-of-Band LSDB Resynchronization*
- RFC 4812, *OSPF Restart Signaling*
- RFC 5088, *OSPF Protocol Extensions for Path Computation Element (PCE) Discovery*
- RFC 5250, *The OSPF Opaque LSA Option*
- RFC 5613, *OSPF Link-Local Signaling*
- RFC 5642, *Dynamic Hostname Exchange Mechanism for OSPF*
- RFC 5709, *OSPFv2 HMAC-SHA Cryptographic Authentication*
- RFC 6571, *Loop-Free Alternate (LFA) Applicability in Service Provider (SP) Networks*
- RFC 6860, *Hiding Transit-Only Networks in OSPF*
- RFC 6987, *OSPF Stub Router Advertisement*

Restrictions and guidelines: OSPF configuration

To run OSPF, you must first enable OSPF on the router. Make a proper configuration plan to avoid incorrect settings that can result in route blocking and routing loops.

OSPF tasks at a glance

To configure OSPF, perform the following tasks:

1. Configuring basic OSPF functions
   - Enabling an OSPF process
   - Creating an OSPF area
   - Enabling OSPF
   - (Optional.) Enabling OSPF to establish neighbor relationships through the secondary IP addresses of an interface
   - (Optional.) Establishing OSPF neighbor relationships on an M-LAG network
2. (Optional.) Configuring OSPF stub and NSSA areas
   - Configuring a stub area
   - Configuring an NSSA area
   - Configuring a virtual link
3. (Optional.) Configuring OSPF network types

- Configuring the broadcast network type for an interface
- Configuring the NBMA network type for an interface
- Configuring the P2MP network type for an interface
- Configuring the P2P network type for an interface

4. (Optional.) Configuring OSPF route control
   - Configuring OSPF inter-area route summarization
   - Configuring redistributed route summarization
   - Configuring received OSPF route filtering
   - Configuring Type-3 LSA filtering
   - Setting an OSPF cost for an interface
   - Enabling OSPF to advertise the maximum link cost to neighbors
   - Associating OSPF with Track
   - Setting the maximum number of ECMP routes
   - Setting OSPF preference
   - Configuring discard routes for summary networks
   - Redistributing routes from another routing protocol
   - Redistributing a default route
   - Advertising a host route
   - Advertising OSPF link state information to BGP
5. (Optional.) Setting OSPF timers
   - Configuring OSPF packet timers
   - Setting the LSA transmission delay
   - Setting the OSPF route calculation interval
   - Setting the minimum LSA arrival interval
   - Setting the LSA update interval
   - Setting the OSPF exit overflow interval
6. (Optional.) Configuring OSPF packet parameters
   - Disabling interfaces from receiving and sending OSPF packets
   - Adding the interface MTU into DD packets
   - Setting the DSCP value for outgoing OSPF packets
   - Setting the maximum length of OSPF packets that can be sent by an interface
   - Setting the LSU transmit rate
7. (Optional.) Controlling LSA generation, advertisement, and reception
   - Setting the maximum number of external LSAs in LSDB
   - Filtering outbound LSAs on an interface
   - Filtering LSAs for the specified neighbor
8. (Optional.) Accelerating OSPF convergence speed
   - Enabling OSPF ISPF
   - Configuring prefix suppression
   - Configuring prefix prioritization
   - Configuring OSPF PIC
9. (Optional.) Configuring advanced OSPF features
   - Configuring stub routers
   - Configuring OSPF isolation

- Configuring OSPF shutdown
- Enabling compatibility with RFC 1583

10. (Optional.) Enhancing OSPF availability
    - Configuring OSPF GR
    - Configuring OSPF NSR
    - Configuring BFD for OSPF
    - Configuring OSPF FRR
11. (Optional.) Configuring OSPF security features
    - Configuring OSPF authentication
    - Configuring GTSM for OSPF
12. (Optional.) Configuring OSPF logging, SNMP notifications, and troubleshooting
    - Logging neighbor state changes
    - Configuring the OSPF logging feature
    - Configuring OSPF network management
    - Setting the maximum number of OSPF neighbor relationship troubleshooting entries

Configuring basic OSPF functions

Enabling an OSPF process

1. Enter system view.
   system-view
2. (Optional.) Configure a global router ID.
   router id *router-id*
   By default, no global router ID is configured.
   If no global router ID is configured, the highest loopback interface IP address, if any, is used as the router ID. If no loopback interface IP address is available, the highest physical interface IP address is used, regardless of the interface status (up or down).
3. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
   By default, OSPF is disabled.
4. (Optional.) Configure a description for the OSPF process.
   description *text*
   By default, no description is configured for the OSPF process.
   As a best practice, configure a description for each OSPF process.

Creating an OSPF area

1. Enter system view.
   system-view
2. (Optional.) Configure a global router ID.
   router id *router-id*
   By default, no global router ID is configured.
3. Enter OSPF view.

ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

By default, OSPF is disabled.

4. (Optional.) Configure a description for the OSPF process.

description *text*

By default, no description is configured for the OSPF process.

As a best practice, configure a description for each OSPF process.

5. Create an OSPF area and enter OSPF area view.

area *area-id*

6. (Optional.) Configure a description for the area.

description *text*

By default, no description is configured for the area.

As a best practice, configure a description for each OSPF area.

7. (Optional.) Exclude interfaces in the OSPF area from the base topology:

capability default-exclusion

By default, interfaces in an OSPF area belong to the base topology.

For correct neighbor relationship establishment, perform this task on both the local device and the neighbor device.

Enabling OSPF

About this task

To enable OSPF on a router, you must perform the following tasks:

1. Create an OSPF process.
2. Create an OSPF area for the process.
3. Specify a network in the area.

The interface attached to the network will run the OSPF process in the area. OSPF advertises direct routes of the interface.

OSPF supports multiple processes. To run multiple OSPF processes, you must specify an ID for each process. The process IDs take effect locally and has no influence on packet exchange between routers. Two routers with different process IDs can exchange packets.

OSPF supports multiple VPNs. You can configure an OSPF process to run in a VPN instance. An OSPF process with no VPN instance specified runs on the public network. For more information about VPN, see MCE configuration in *MCE Configuration Guide.*

Restrictions and guidelines for enabling OSPF

When you configure OSPF on an interface, follow these restrictions and guidelines:

- You can enable OSPF on the network where the interface resides or directly enable OSPF on that interface. If you configure both, the latter takes precedence.
- If the specified OSPF process and area do not exist, the operation creates an OSPF process and area for the interface. Disabling an OSPF process on an interface does not delete the OSPF process or the area.

Enabling OSPF on a network

1. Enter system view.

system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter OSPF area view.

   area *area-id*

4. Specify a network to enable the interface attached to the network to run the OSPF process in the area.

   network *ip-address wildcard-mask*

   By default, no network is specified to enable OSPF on the interface attached to the network.

   A network can be added to only one area.

Enabling OSPF on an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable an OSPF process on the interface.

   ospf *process-id* **area** *area-id* [**exclude-subip**]

   By default, OSPF is disabled on an interface.

Enabling OSPF to establish neighbor relationships through the secondary IP addresses of an interface

About this task

By default, OSPF uses only the primary IP address of an interface to establish neighbor relationships. This feature enables OSPF to establish neighbor relationships through the secondary IP addresses of an interface.

Restrictions and guidelines

- For an interface with both primary and secondary addresses, if you have advertised only the primary address in an area of an OSPF process, OSPF uses only the primary address to establish neighbor relationships. For OSPF to establish neighbor relationships through the secondary IP addresses, you must also advertise the secondary IP addresses in the same area of the same OSPF process.
- If an interface does not have a primary address but has multiple secondary addresses, OSPF uses the lowest secondary address to establish neighbor relationships. For OSPF to establish neighbor relationships through the other secondary addresses, you must advertise these addresses in the same area of the same OSPF process.
- After you execute the **ospf peer sub-address enable** command on a P2P OSPF interface, the device automatically checks the local and remote interface IP addresses before neighbor establishment. If the IP addresses belong to different networks, they cannot establish a neighbor relationship.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable OSPF to establish neighbor relationships through the secondary IP addresses of the interface.

   `ospf peer sub-address enable`

   By default, OSPF cannot establish neighbor relationships through the secondary IP addresses of an interface.

Establishing OSPF neighbor relationships on an M-LAG network

About this task

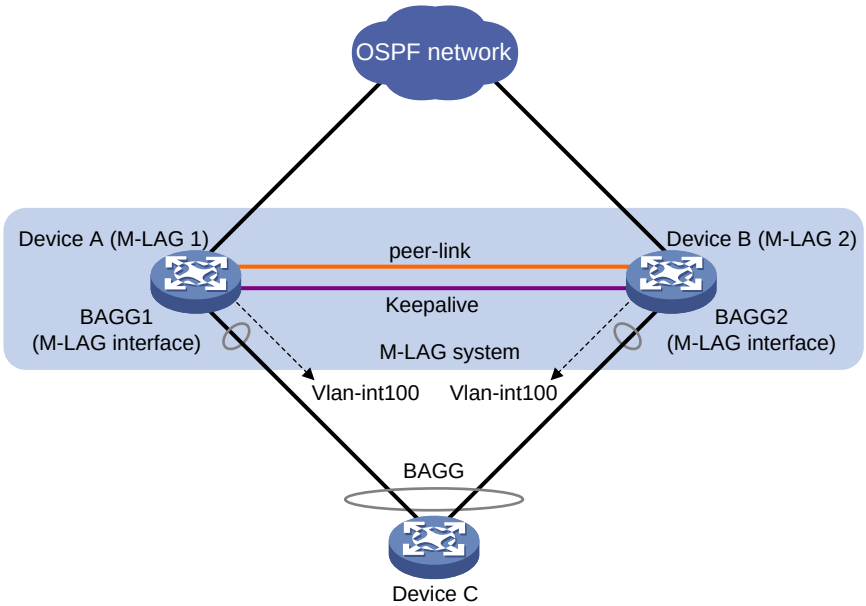
When a customer-side device accesses the OSPF network through both M-LAG member devices in an M-LAG system, the M-LAG member devices provide Layer 3 gateway services. The gateway interfaces (such as VLAN interfaces) on different M-LAG member devices might have the same IP address and the same MAC address. In this case, the M-LAG member devices cannot establish OSPF neighbor relationships with the customer-side device.

To resolve this issue, configure the gateway interfaces to use M-LAG virtual IPv4 addresses for OSPF neighbor relationship establishment. For more information about M-LAG, see M-LAG configuration in *Layer 2—LAN Switching Configuration Guide*.

As shown in Figure 9, Device C accesses the OSPF network through the M-LAG system that consists of Device A and Device B. Device A and Device B provide Layer 3 gateway services. The gateway interface VLAN-interface 100 on Device A and the gateway interface VLAN-interface 100 on Device B have the same IP address and the same MAC address. To ensure that both Device A and Device B can establish an OSPF neighbor relationship with Device C, perform the following tasks:

1. Configure different M-LAG virtual IPv4 addresses for the VLAN interfaces on Device A and Device B.
2. Configure the VLAN interface to use the M-LAG virtual IPv4 address for OSPF neighbor relationship establishment on both Device A and Device B.

Figure 9 OSPF neighbor relationship establishment on an M-LAG network



Procedure

1. Enter system view.

system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure an M-LAG virtual IPv4 address for the interface.

   port m-lag virtual-ip *ipv4-address* { *mask-length* | *mask* }

   By default, no M-LAG virtual IPv4 address is configured for an interface.

   An M-LAG virtual IPv4 address is similar to a secondary IP address. For more information about this command, see M-LAG commands in *Layer 2—LAN Switching Command Reference*.

4. Configure the interface to use the M-LAG virtual IPv4 address for OSPF neighbor relationship establishment.

   ospf peer sub-address enable *ip-address*

   By default, an interface uses the primary IP address for OSPF neighbor relationship establishment.

   The IP address specified in this command must be the same as the IPv4 address specified in the **port m-lag virtual-ip** command.

Configuring OSPF stub and NSSA areas

About OSPF stub and NSSA area configuration

This task allows you to configure an OSPF area as a stub area or NSSA area. It also allows you to create a virtual link if no connectivity can be achieved between a non-backbone area and backbone area, or in the backbone area.

Configuring a stub area

About this task

You can configure a non-backbone area at an AS edge as a stub area. To do so, execute the **stub** command on all routers attached to the area. The routing table size is reduced because Type-5 LSAs will not be flooded within the stub area. The ABR generates a default route into the stub area so all packets destined outside of the AS are sent through the default route.

To further reduce the routing table size and routing information exchanged in the stub area, configure a totally stub area by using the **stub no-summary** command on the ABR. AS external routes and inter-area routes will not be distributed into the area. All the packets destined for outside of the AS or area will be sent to the ABR for forwarding.

A stub or totally stub area cannot have an ASBR because external routes cannot be distributed into the area.

Restrictions and guidelines

Do not configure the backbone area as a stub area or totally stub area.

To configure an area as a stub area, execute the **stub** command on all routers attached to the area.

To configure an area as a totally stub area, execute the **stub** command on all routers attached to the area, and execute the **stub no-summary** command on the ABR.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter area view.

   area *area-id*

4. Configure the area as a stub area.

   stub [**default-route-advertise-always** | **no-summary**] *

   By default, no stub area is configured.

5. (Optional.) Set a cost for the default route advertised to the stub area.

   default-cost *cost-value*

   By default, the cost for the default route advertised to the stub area is 1.

   This command takes effect only on the ABR of a stub area or totally stub area.

Configuring an NSSA area

About this task

A stub area cannot import external routes, but an NSSA area can import external routes into the OSPF routing domain while retaining other stub area characteristics.

To configure an area as a totally NSSA area, use the **nssa no-summary** command. The ABR of the area does not advertise inter-area routes into the area.

Restrictions and guidelines

Do not configure the backbone area as an NSSA area or totally NSSA area.

To configure an NSSA area, configure the **nssa** command on all the routers attached to the area.

To configure a totally NSSA area, configure the **nssa** command on all the routers attached to the area and configure the **nssa no-summary** command on the ABR.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter area view.

   area *area-id*

4. Configure the area as an NSSA area.

   nssa [**default-route-advertise** [**cost** *cost-value* | **nssa-only** | **route-policy** *route-policy-name* | **type** *type*] * | **no-import-route** | **no-summary** | **suppress-fa** | [[[**translate-always**] [**translate-ignore-checking-backbone**]] | **translate-never**] | **translator-stability-interval** *value*] *

   By default, no area is configured as an NSSA area.

5. (Optional.) Set a cost for the default route advertised to the NSSA area.

   default-cost *cost-value*

   By default, the cost for the default route advertised to the NSSA area is 1.

   This command takes effect only on the ABR/ASBR on an NSSA area or totally NSSA area.

Configuring a virtual link

About this task

You can configure a virtual link to maintain connectivity between a non-backbone area and the backbone, or in the backbone itself.

Restrictions and guidelines

A virtual link cannot traverse a stub area, totally stub area, NSSA area, or totally NSSA area.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter OSPF area view.

   area *area-id*

4. Configure a virtual link.

   vlink-peer *router-id* [**dead** *seconds* | **hello** *seconds* | { **authentication-none** | { **hmac-md5** | **hmac-sha-256** | **hmac-sm3** | **md5** } [*key-id* { **cipher** | **plain** } *string*] | **keychain** *keychain-name* | **simple** [{ **cipher** | **plain** } *string*] } | **retransmit** *seconds* | **trans-delay** *seconds*] *

   Configure this command on both ends of a virtual link. The **hello** and **dead** intervals must be identical on both ends of the virtual link.

Configuring OSPF network types

Based on the link layer protocol, OSPF classifies networks into different types, including broadcast, NBMA, P2MP, and P2P.

Restrictions and guidelines for configuring OSPF network types

If any routers in a broadcast network do not support multicasting, change the network type to NBMA.

If only two routers running OSPF exist on a network segment, you can change the network type to P2P to save costs.

Two broadcast-, NBMA-, and P2MP-interfaces can establish a neighbor relationship only when they are on the same network segment.

Configuring the broadcast network type for an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the OSPF network type for the interface as broadcast.

ospf network-type broadcast

By default, the network type of an interface depends on the encapsulated link layer protocol.

4. (Optional.) Set a router priority for the interface.

ospf dr-priority *priority*

The default router priority is 1.

Configuring the NBMA network type for an interface

Restrictions and guidelines

After you configure the network type as NBMA, you must specify neighbors and their router priorities because NBMA interfaces cannot find neighbors by broadcasting hello packets.

Procedure

1. Enter system view.

system-view

2. Enter interface view.

interface *interface-type interface-number*

3. Configure the OSPF network type for the interface as NBMA.

ospf network-type nbma

By default, the network type of an interface depends on the encapsulated link layer protocol.

4. (Optional.) Set a router priority for the interface.

ospf dr-priority *priority*

The default router priority for an interface is 1.

The router priority configured with this command is for DR election.

5. Return to system view.

quit

6. Enter OSPF view.

ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

7. Specify an NBMA neighbor.

peer *ip-address* [**dr-priority** *priority*]

By default, no neighbor is specified.

The priority configured with this command indicates whether a neighbor has the election right or not. If you configure the router priority for a neighbor as 0, the local router determines the neighbor has no election right. It does not send hello packets to this neighbor. However, if the local router is the DR or BDR, it still sends hello packets to the neighbor for neighbor relationship establishment.

Configuring the P2MP network type for an interface

1. Enter system view.

system-view

2. Enter interface view.

interface *interface-type interface-number*

3. Configure the OSPF network type for the interface as P2MP.

ospf network-type p2mp [**unicast**]

By default, the network type of an interface depends on the encapsulated link layer protocol.

After you configure the OSPF network type for an interface as P2MP unicast, all packets are unicast over the interface. The interface cannot broadcast hello packets to discover neighbors, so you must manually specify the neighbors.

4. Return to system view.

   quit

5. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

6. Specify a P2MP neighbor.

   peer *ip-address* [**cost** *cost-value*]

   By default, no neighbor is specified

   This step is required if the interface network type is P2MP unicast.

Configuring the P2P network type for an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the OSPF network type for the interface as P2P.

   ospf network-type p2p [**multicast** | **peer-address-check**] *

   By default, the network type of an interface depends on the encapsulated link layer protocol.

Configuring OSPF route control

This section describes how to control the advertisement and reception of OSPF routing information, as well as route redistribution from other protocols.

Configuring OSPF inter-area route summarization

About this task

OSPF inter-area route summarization reduces the routing information exchanged between areas and the size of routing tables, and improves routing performance.

OSPF inter-area route summarization enables an ABR to summarize contiguous networks into a single network and advertise the network to other areas. For example, three internal networks 19.1.1.0/24, 19.1.2.0/24, and 19.1.3.0/24 are available within an area. You can configure the ABR to summarize the three networks into network 19.1.0.0/16, and advertise the summary network to other areas in a Type-3 LSA. This configuration reduces the scale of LSDBs on routers in other areas and the influence of topology changes.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter OSPF area view.

area *area-id*

4. Configure ABR route summarization.

   abr-summary *ip-address* { *mask-length* | *mask* } [**advertise** | **not-advertise**] [**cost** *cost-value*]

   By default, route summarization is not configured on an ABR.

Configuring redistributed route summarization

About this task

Perform this task to enable an ASBR to summarize external routes within the specified address range into a single route. The ASBR advertises only Type-5 LSAs to reduce the number of LSAs in the LSDB.

An ASBR can summarize routes in the following LSAs:

- Type-5 LSAs.
- Type-7 LSAs in an NSSA area.

Restrictions and guidelines

If an ASBR (also an ABR) is a translator in an NSSA area, it summarizes routes in Type-5 LSAs translated from Type-7 LSAs. If it is not a translator, it does not summarize routes in in Type-5 LSAs translated from Type-7 LSAs.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure ASBR route summarization.

   asbr-summary *ip-address* { *mask-length* | *mask* } [**cost** *cost-value* | **not-advertise** | **nssa-only** | **tag** *tag*] *

   By default, route summarization is not configured on an ASBR.

Configuring received OSPF route filtering

About this task

Perform this task to filter routes calculated using received LSAs.

The following filtering methods are available:

- Use an ACL or IP prefix list to filter routing information by destination address.
- Use the **gateway** *prefix-list-name* option to filter routing information by next hop.
- Use an ACL or IP prefix list to filter routing information by destination address. At the same time use the **gateway** *prefix-list-name* option to filter routing information by next hop.
- Use the **route-policy** *route-policy-name* option to filter routing information.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure OSPF to filter routes calculated using received LSAs.

   filter-policy { *ipv4-acl-number* [**gateway** *prefix-list-name*] | **gateway** *prefix-list-name* | **prefix-list** *prefix-list-name* [**gateway** *prefix-list-name*] | **route-policy** *route-policy-name* } **import**

   By default, OSPF accepts all routes calculated by using received LSAs.

Configuring Type-3 LSA filtering

About this task

Perform this task to filter Type-3 LSAs advertised into the local area or other areas on an ABR.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter OSPF area view.

   area *area-id*

4. Configure Type-3 LSA filtering.

   filter { *ipv4-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } { **export** | **import** }

   By default, the ABR does not filter Type-3 LSAs.

Setting an OSPF cost for an interface

About this task

Set an OSPF cost for an interface by using either of the following methods:

- Set the cost value in interface view.
- Set a bandwidth reference value for the interface. OSPF computes the cost with this formula: Interface OSPF cost = Bandwidth reference value (100 Mbps) / Expected interface bandwidth (Mbps). The expected bandwidth of an interface is configured with the **bandwidth** command (see *Interface Command Reference*).
  - If the calculated cost is greater than 65535, the value of 65535 is used. If the calculated cost is less than 1, the value of 1 is used.
  - If no cost or bandwidth reference value is configured for an interface, OSPF computes the interface cost based on the interface bandwidth and default bandwidth reference value.

Setting an OSPF cost for an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set an OSPF cost for the interface.

   ospf cost *cost-value*

By default, the OSPF cost is calculated according to the interface bandwidth. For a loopback interface, the OSPF cost is 0 by default.

Setting a bandwidth reference value

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Set a bandwidth reference value.
   bandwidth-reference *value*
   The default setting is 100 Mbps.

Enabling OSPF to advertise the maximum link cost to neighbors

About this task

On an OSPF network, when a link recovers from failures or the state of an interface changes, OSPF will re-establish neighbor relationships and perform route convergence. During the route convergence process, routing loops and traffic loss might occur because the convergence speeds of the nodes are different. To address this issue, enable OSPF to advertise the maximum link cost to neighbors within the specified period, so the traffic forwarding path remains unchanged. After the specified period, OSPF advertises the original link cost to neighbors and performs optimal route selection again.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable OSPF to advertise the maximum link cost to neighbors within the specified period.
   ospf peer hold-max-cost duration *time*
   By default, OSPF advertises the original link cost to neighbors during a route convergence.

Associating OSPF with Track

About this task

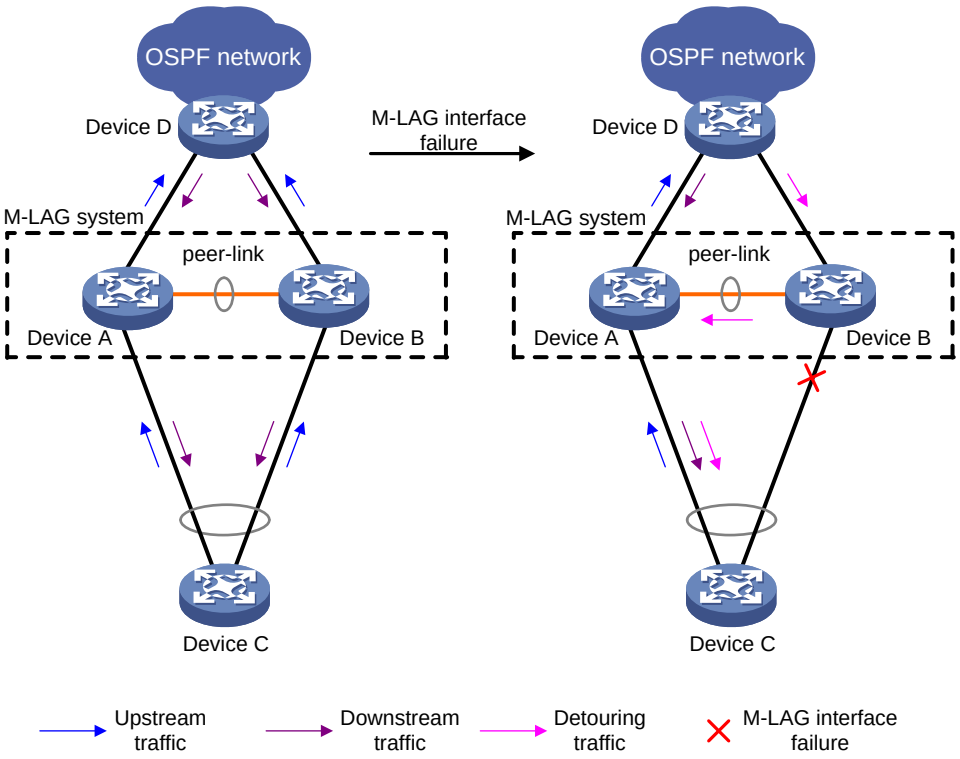
Perform this task to associate an OSPF interface with a track entry to adjust the cost of the interface as follows:

- When the track entry state becomes negative, OSPF increases the cost of the interface.
- When the track entry state becomes positive, OSPF restores the cost of the interface to the original value.

When a customer side device is dual-homed to an OSPF network through M-LAG, you can configure this feature to prevent the traffic from detouring through the M-LAG peer link. As shown in Figure 10, customer-side device Device C is dual-homed to the OSPF network through an M-LAG system formed by Device A and Device B. Device A and Device B share the traffic load between Device D and Device C. When the DR interface that connects Device B to Device C fails, the upstream traffic and downstream traffic are forwarded as follows:

- Device C detects the M-LAG interface failure, and traffic from Device C to Device D is switched from path Device C->Device B->Device D to path Device C->Device A->Device D.
- Device D cannot detect the M-LAG interface failure and still forwards traffic through Device B. Because the link between Device B and Device C fails, traffic from Device D to Device C is switched from path Device D->Device B->Device C to path Device D->Device B->Device A->Device C.

Figure 10 Traffic forwarding path after M-LAG interface failure



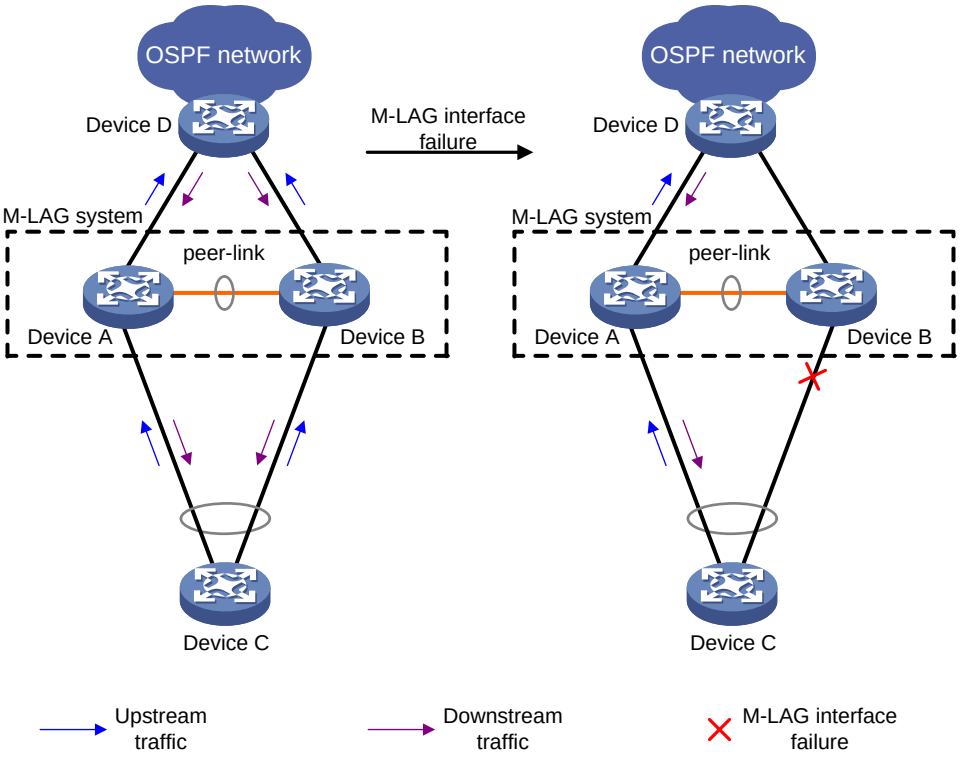
To prevent the downstream traffic from detouring through the M-LAG peer link, perform the following tasks on Device B:

1. Execute the **track interface** command to create a track entry and associate it with the link state of the M-LAG interface. For more information about the **track interface** command, see track commands in *High Availability Command Reference*.
2. Execute the **ospf track adjust-cost** command to associate the gateway interface with the track entry.

Then, Device B adjusts the cost of the gateway interface based on the state of the associated track entry:

- When the M-LAG interface fails, the track entry state becomes negative, and OSPF increases the cost of the gateway interface. Then, traffic from Device D to Device C is switched to path Device D->Device A->Device C.
- When the M-LAG interface recovers, the track entry state becomes positive, and OSPF restores the cost of the gateway interface to the original value. Then, traffic from Device D to Device C is switched back to path Device D->Device B->Device C.

Figure 11 Traffic forwarding path after associating OSPF with Track



Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Associate the OSPF interface with a track entry to adjust the cost of the interface based on the track entry state.

   ospf track *track-entry-number* **adjust-cost** { *cost-offset* | **max** }

   By default, an OSPF interface is not associated with any track entry.

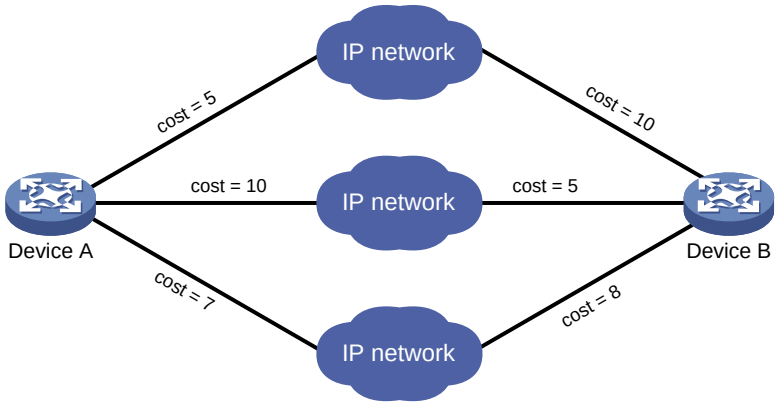
Setting the maximum number of ECMP routes

About this task

OSPF might find multiple optimal equal-cost routes to the same destination, which can be used to share the traffic load. This task allows you to set the maximum number of ECMP routes for OSPF.

As shown in Figure 12, Device A and Device B run OSPF. There are three equal-cost routes between Device A and Device B. You can use these routes to implement load sharing.

Figure 12 Equal-cost routes



Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the maximum number of ECMP routes.

   maximum load-balancing *number*

   By default, the maximum number of ECMP routes equals the maximum number of ECMP routes supported by the system.

Setting OSPF preference

About this task

A router can run multiple routing protocols, and each protocol is assigned a preference. If multiple routes are available to the same destination, the one with the highest protocol preference is selected as the best route.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Set a preference for OSPF.

   preference [**ase**] { *preference* | **route-policy** *route-policy-name* } *

   By default, the preference of OSPF internal routes is 10 and the preference of OSPF external routes is 150.

Configuring discard routes for summary networks

About this task

Perform this task on an ABR or ASBR to specify whether to generate discard routes for summary networks. You can also specify a preference for the discard routes.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure discard routes for summary networks.

   discard-route { **external** { *preference* | **suppression** } | **internal** { *preference* | **suppression** } } *

   By default, the ABR or ASBR generates discard routes for summary networks and the default preference of discard routes is 255.

Redistributing routes from another routing protocol

About this task

When an OSPF device does not have routes from a non-OSPF routing protocol, it cannot access any devices that run the non-OSPF routing protocol. To resolve this issue, you can configure OSPF to redistribute routes from other protocols, such as IS-IS and BGP. OSPF can then advertise the routes in Type-5 LSAs or Type-7 LSAs.

In addition, you can configure OSPF to filter redistributed routes so that OSPF advertises only permitted routes.

Restrictions and guidelines

Although OSPF is a dynamic routing protocol that provides loop-free intra-area/inter-area routes, this protocol is not good at guarding against the routing loops caused by redistributed external routes. When you configure route redistribution for OSPF, make sure you have understood the potential impact.

OSPF redistributes only active routes. To view route status information, use the **display ip routing-table protocol** command.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure route redistribution.

   import-route bgp [*as-number*] [**allow-ibgp**] [[**cost** *cost-value* | **inherit-cost**] | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   import-route { **direct** | **static** } [[**cost** *cost-value* | **inherit-cost**] | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   import-route { **isis** | **ospf** | **rip** } [*process-id* | **all-processes**] [**allow-direct** | [**cost** *cost-value* | **inherit-cost**] | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   By default, OSPF does not redistribute routes.

   The **import-route bgp** command redistributes only EBGP routes. The **import-route bgp allow-ibgp** command redistributes both EBGP and IBGP routes, which might cause routing loops. Therefore, use it with caution.

4. (Optional.) Configure OSPF to filter redistributed routes.

   filter-policy { *ipv4-acl-number* | **prefix-list** *prefix-list-name* } **export** [*protocol* [*process-id*]]

   By default, OSPF accepts all redistributed routes.

5. Configure the default parameters for redistributed routes (cost, tag, and type).

   default { **cost** *cost-value* | **tag** *tag* | **type** *type* } *

   By default, the cost is 1, the tag is 1, and the route type is 2

Redistributing a default route

About this task

On a OSPF network, multiple ABRs and ASBRs work together to implement multi-egress redundancy or load balancing. To ease the service burden on these devices and save routing table capacity, you can configure default routes.

OSPF default routes are mostly used in the following scenarios:

- After receiving a Type 3 LSA that describes a default route from the ABR, the intra-area router can use the route for packet forwarding to another area in the same AS.
- After receiving a Type 5 or 7 LSA that describes a default route from the ASBR, the intra-area router can use the route for packet forwarding to another AS.

When a router cannot find a exactly matching route, it can use the default route for packet forwarding. The default route originated from a Type 3 LSA takes precedence over that originated from a Type 5 LSA.

As shown in Table 3, the advertisement method for a redistributed default route varies by its area type.

Table 3 Advertisement of redistributed default routes

Area type	Route redistribution condition	Advertiser	LSA type	LSA flooding scope
Common area	Use the **default-route-advertise** command.	ASBR	Type 5 LSA	Common area
Stub area	Automatic route redistribution.	ABR	Type 3 LSA	Stub area
NSSA area	Use the **nssa** [**default-route-advertise**] command.	ASBR	Type 7 LSA	NSSA area
Totally NSSA area	Automatic route redistribution.	ABR	Type 3 LSA	NSSA area

The **import-route** command cannot redistribute a default route from an external routing protocol. To resolve this issue, perform this task on the ASBR to redistribute a default route.

Restrictions and guidelines

After receiving a default route from the ASBR, the device can add the default route to its routing table in one of the following scenarios:

- The device does not have a static route.
- The device has a static route, but the priority value for the default route is higher than that for the static route.

Procedure

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Redistribute a default route.
   default-route-advertise [[**always** | **permit-calculate-other**] | **cost** *cost-value* | **route-policy** *route-policy-name* | **type** *type*] *
   default-route-advertise [**summary cost** *cost-value*]
   By default, no default route is redistributed.
   This command is applicable only to VPNs. The PE router advertises a default route in a Type-3 LSA to a CE router.
4. Configure the default parameters for redistributed routes (cost, tag, and type).
   default { **cost** *cost-value* | **tag** *tag* | **type** *type* } *
   By default, the cost is 1, the tag is 1, and the route type is 2.

Advertising a host route

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter area view.
   area *area-id*
4. Advertise a host route.
   host-advertise *ip-address cost-value*
   By default, OSPF does not advertise host routes that are not in the area.

Advertising OSPF link state information to BGP

About this task

After the device advertises OSPF link state information to BGP, BGP can advertise the information for intended applications. For more information about BGP LS, see "Configuring BGP."

Procedure

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Advertise OSPF link state information to BGP.
   distribute bgp-ls [**strict-link-checking**]
   By default, the device does not advertise OSPF link state information to BGP.

Setting OSPF timers

About setting OSPF timers

This task allows you to change OSPF packet timers to adjust the convergence speed and network load and tune the delay time for sending LSAs on low-speed links.

Configuring OSPF packet timers

About this task

An OSPF interface includes the following timers:

- **Hello timer**—Interval for sending hello packets. It must be identical on OSPF neighbors.
- **Poll timer**—Interval for sending hello packets to a neighbor that is down on the NBMA network.
- **Dead timer**—Interval within which if the interface does not receive any hello packet from the neighbor, it declares the neighbor is down.
- **LSA retransmission timer**—Interval within which if the interface does not receive any acknowledgment packets after sending an LSA to the neighbor, it retransmits the LSA.

Restrictions and guidelines

The default value for the hello interval and neighbor dead interval depends on the network type. When the network type for an interface is changed, the default hello interval and neighbor dead interval are restored. Make sure two neighboring interfaces are configured with the same hello interval and neighbor dead interval. Inconsistent settings will affect the OSPF neighbor relationship establishment.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the hello interval.

   ospf timer hello *seconds*

   The default hello interval on P2P and broadcast interfaces is 10 seconds. The default hello interval on P2MP and NBMA interfaces is 30 seconds.

4. Set the poll interval.

   ospf timer poll *seconds*

   The default setting is 120 seconds.

   The poll interval is a minimum of four times the hello interval.

5. Set the dead interval.

   ospf timer dead *seconds*

   The default dead interval on P2P and broadcast interfaces is 40 seconds. The default dead interval on P2MP and NBMA interfaces is 120 seconds.

   The dead interval must be a minimum of four times the hello interval on an interface.

6. Set the retransmission interval.

   ospf timer retransmit *interval*

   The default retransmission interval is 5 seconds.

A retransmission interval setting that is too small can cause unnecessary LSA retransmissions. Typically set a bigger interval than the round-trip time of a packet between two neighbors.

Setting the LSA transmission delay

About this task

To avoid LSAs from aging out during transmission, set an LSA retransmission delay especially for low speed links.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the LSA transmission delay.
   ospf trans-delay *seconds*
   The default LSA transmission delay is 1 second.

Setting the OSPF route calculation interval

About this task

LSDB changes result in SPF calculations. When the topology changes frequently, a large amount of network and router resources are occupied by SPF calculation. You can adjust the route calculation interval to reduce the impact.

The default route calculation interval settings are as follows:

- The interval for the first route calculation is 50 milliseconds.
- The interval for the nth (n ≥ 2) route calculation is 50 + (200 x 2^{n-2}) milliseconds, and cannot exceed 5 seconds.

The **spf-schedule-interval** command functions as follows:

- If the **spf-schedule-interval** *maximum-interval* command is executed:
  - The values for the *minimum-interval* and *incremental-interval* arguments are both 0.
  - The route calculation interval is fixed at *maximum-interval*.
- If the **spf-schedule-interval** *maximum-interval minimum-interval* command is executed:
  - The value for the *incremental-interval* argument is 0.
  - The interval for the first route calculation is *minimum-interval*.
  - The interval for subsequent route calculations is *maximum-interval*.
- If the **spf-schedule-interval** *maximum-interval minimum-interval incremental-interval* command is executed:
  - The interval for the first route calculation is *minimum-interval*.
  - The interval for the nth (n ≥ 2) route calculation is *minimum-interval* + *incremental-interval* x 2^{n-2}, and cannot exceed *maximum-interval*.

Procedure

1. Enter system view.

system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the route calculation interval for OSPF.

   spf-schedule-interval *maximum-interval* [*minimum-interval* [*incremental-interval*]]

   By default, the maximum interval is 5 seconds, the minimum interval is 50 milliseconds, and the incremental interval is 200 milliseconds.

Setting the minimum LSA arrival interval

About this task

OSPF drops any duplicate LSAs (with the same LSA type, LS ID, and router ID) within the minimum LSA arrival interval. This helps avoid overuse of bandwidth and router resources due to frequent network changes.

Restrictions and guidelines

On a stable network, you can set the LSA arrival interval to 0 to fasten route convergence.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the minimum LSA arrival interval.

   lsa-arrival-interval *interval*

   By default, the minimum LSA arrival interval is 1000 milliseconds.

Setting the LSA update interval

About this task

If network changes become frequent, bandwidth resources and router resources might be overconsumed. To avoid this issue, use the **lsa-generation-interval** command to adjust the LSA update interval settings for OSPF.

The default LSA update interval settings are as follows:

- The interval for the first LSA update is 50 milliseconds.
- The interval for the nth (n ≥ 2) LSA update is 50 + (200 x 2^{n-2}) milliseconds, and cannot exceed 5 seconds.

The **lsa-generation-interval** command functions as follows:

- If the **lsa-generation**-*interval maximum-interval* command is executed:
  - The values for the *minimum-interval* and *incremental-interval* arguments are both 0.
  - The LSA update interval is fixed at *maximum-interval*.
- If the **lsa-generation-interval** *maximum-interval minimum-interval* command is executed:

- The value for the *incremental-interval* argument is 0.
- The interval for the first LSA update is *minimum-interval*.
- The interval for subsequent LSA updates is *maximum-interval*.

- If the **lsa-generation-interval** *maximum-interval minimum-interval incremental-interval* command is executed:
  - The interval for the first LSA update is *minimum-interval*.
  - The interval for the nth (n ≥ 2) LSA update is *minimum-interval* + *incremental-interval* x 2^{n-2}, and cannot exceed *maximum-interval*.

Procedure

1. Enter system view.

   system-view
2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Set the LSA update interval.

   lsa-generation-interval *maximum-interval* [*minimum-interval* [*incremental-interval*]]

   By default, the maximum interval is 5 seconds, the minimum interval is 50 milliseconds, and the incremental interval is 200 milliseconds.

Setting the OSPF exit overflow interval

About this task

When the number of LSAs in the LSDB exceeds the upper limit, the LSDB is in an overflow state. In this state, OSPF does not receive any external LSAs and deletes the external LSAs generated by itself to save system resources.

This task allows you to configure the interval that OSPF exits overflow state.

Procedure

1. Enter system view.

   system-view
2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Set the interval that OSPF exits overflow state.

   lsdb-overflow-interval *interval*

   By default, the OSPF exit overflow interval is 300 seconds. An interval of 0 means that OSPF does not exit overflow state.

Configuring OSPF packet parameters

Disabling interfaces from receiving and sending OSPF packets

About this task

To enhance OSPF adaptability and reduce resource consumption, you can set an OSPF interface to "silent." A silent OSPF interface blocks OSPF packets and cannot establish any OSPF neighbor relationship. However, other interfaces on the router can still advertise direct routes of the interface in Router LSAs.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Disable interfaces from receiving and sending OSPF packets.

   silent-interface { *interface-type interface-number* | **all** }

   By default, an OSPF interface can receive and send OSPF packets.

   This command disables only the interfaces associated with the current process rather than other processes. Multiple OSPF processes can disable the same interface from receiving and sending OSPF packets.

Adding the interface MTU into DD packets

About this task

By default, an OSPF interface adds a value of 0 into the interface MTU field of a DD packet rather than the actual interface MTU. On receipt of a DD packet, the interface does not check the interface MTU field of that packet. Such a mechanism ensures that two interfaces in different devices can establish a neighbor relationship regardless of the interface MTU. However, the device cannot establish neighbor relationships with the devices from another manufacturer that discard the DD packets whose interface MTU value is 0.

To resolve this issue, perform this task to enable an OSPF interface to add its interface MTU into DD packets and check the interface MTU in received DD packets. If the interface MTU in a received DD packet is higher than the MTU of the local interface, the interface will discard the DD packet. To ensure that two interfaces in different devices can establish a neighbor relationship, set the same MTU for these interfaces.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the interface to add its MTU into DD packets.

   ospf mtu-enable

   By default, the interface adds an MTU value of 0 into DD packets.

Setting the DSCP value for outgoing OSPF packets

About this task

The DSCP value specifies the precedence of outgoing packets.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the DSCP value for outgoing OSPF packets.

   dscp *dscp-value*

   By default, the DSCP value for outgoing OSPF packets is 48.

Setting the maximum length of OSPF packets that can be sent by an interface

About this task

This task allows you to limit the length of OSPF packets sent over an interface. In some scenarios, for example, when you establish OSPF neighbors over a tunnel, you can perform this task to prevent OSPF packet fragmentation on the outgoing tunnel interface. Make sure the maximum length of the OSPF packets plus the encapsulated header length is no greater than the outgoing tunnel interface's MTU. For more information about tunnels, see *Layer 3—IP Services Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the maximum length of OSPF packets that can be sent by an interface.

   ospf packet-size *value*

   By default, the maximum length of OSPF packets that an interface can send equals the interface's MTU.

Setting the LSU transmit rate

About this task

During LSDB synchronization, if the local router has multiple neighbors, it must send many LSUs to each neighbor. When a neighbor receives excessive LSUs within a short time period, the following events might occur:

- The neighbor experiences degraded performance because it uses too many system resources to process the received LSU packets.
- The neighbor drops hello packets used for maintaining the neighbor relationship because it is busy dealing with the LSUs. As a result, the neighbor relationship is torn down. To reestablish a relationship to the neighbor, the local router must send more LSUs to the neighbor. This exacerbates the performance degradation.

This task allows you to limit the LSU transmit rate by setting the LSU transmit interval and the maximum number of LSUs that can be sent at each interval.

Procedure

1. Enter system view.

   system-view

2. Enable OSPF to limit the LSU transmit rate.

   ospf lsu-flood-control [*interval count*]

   By default, OSPF does not limit the LSU transmit rate.

   Inappropriate use of this command might cause abnormal routing. As a best practice, execute this command with the default values.

3. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

4. (Optional.) Set the LSU transmit interval and the maximum number of LSUs that can be sent at each interval.

   transmit-pacing interval *interval* **count** *count*

   By default, an OSPF interface sends a maximum of three LSU packets every 20 milliseconds.

Controlling LSA generation, advertisement, and reception

Setting the maximum number of external LSAs in LSDB

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the maximum number of external LSAs in the LSDB.

   lsdb-overflow-limit *number*

   By default, the maximum number of external LSAs in the LSDB is not limited.

Filtering outbound LSAs on an interface

About this task

To reduce the LSDB size for the neighbor and save bandwidth, you can perform this task on an interface to filter LSAs to be sent to the neighbor.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Filter outbound LSAs on the interface.

ospf database-filter { **all** | { **ase** [**acl** *ipv4-acl-number*] | **nssa** [**acl** *ipv4-acl-number*] | **summary** [**acl** *ipv4-acl-number*] } * }

By default, the outbound LSAs are not filtered on the interface.

Filtering LSAs for the specified neighbor

About this task

On a P2MP network, a router might have multiple P2MP type OSPF neighbors. Perform this task to prevent the router from sending LSAs to the specified P2MP neighbor.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Filter LSAs for the specified P2MP neighbor.

   database-filter peer *ip-address* { **all** | { **ase** [**acl** *ipv4-acl-number*] | **nssa** [**acl** *ipv4-acl-number*] | **summary** [**acl** *ipv4-acl-number*] } * }

   By default, the LSAs for the specified P2MP neighbor are not filtered.

Accelerating OSPF convergence speed

Enabling OSPF ISPF

About this task

When the topology changes, Incremental Shortest Path First (ISPF) computes only the affected part of the SPT, instead of the entire SPT.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable OSPF ISPF.

   ispf enable

   By default, OSPF ISPF is enabled.

Configuring prefix suppression

About this task

By default, an OSPF interface advertises all of its prefixes in LSAs. To speed up OSPF convergence, you can suppress interfaces from advertising all of their prefixes. This feature helps improve network security by preventing IP routing to the suppressed networks.

When prefix suppression is enabled:

- On P2P and P2MP networks, OSPF does not advertise Type-3 links in Type-1 LSAs. Other routing information can still be advertised to ensure traffic forwarding.
- On broadcast and NBMA networks, the DR generates Type-2 LSAs with a mask length of 32 to suppress network routes. Other routing information can still be advertised to ensure traffic forwarding. If no neighbors exist, the DR does not advertise Type-3 links in Type-1 LSAs.

Restrictions and guidelines for prefix suppression

As a best practice, configure prefix suppression on all OSPF routers if you want to use prefix suppression.

Configuring prefix suppression for an OSPF process

1. Enter system view.

   system-view
2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable prefix suppression for the OSPF process.

   prefix-suppression

   By default, prefix suppression is disabled for an OSPF process.

   This feature does not suppress the prefixes of secondary IP addresses, loopback interfaces, and passive interfaces.

Configuring prefix suppression for an interface

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable prefix suppression for the interface.

   ospf prefix-suppression [**disable**]

   By default, prefix suppression is disabled on an interface.

   This feature does not suppress prefixes of secondary IP addresses.

Configuring prefix prioritization

About this task

OSPF calculates intra-area routes, inter-area routes, and AS external routes in descending convergence priority order: critical, high, medium, and low. When a route matches multiple prefix priorities, it prefers to use the highest priority.

To prioritize the convergence of specific routes, perform this task to set a higher convergence priority value for those routes.

Procedure

1. Enter system view.

   system-view
2. Configure an IPv4 prefix list.

   ip prefix-list *prefix-list-name* [**index** *index-number*] { **deny** | **permit** } *ip-address mask-length* [**greater-equal** *min-mask-length*] [**less-equal** *max-mask-length*]
3. Configure a routing policy.

route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*

4. Use the IPv4 prefix list as an IPv4 route filter.

   if-match ip { **address** | **next-hop** | **route-source** } **prefix-list** *prefix-list-name*

   By default, no IPv4 route filters are configured.

5. Set a prefix priority for matching routes.

   apply prefix-priority { **critical** | **high** | **medium** }

   By default, no prefix priority is set.

6. Return to system view.

   quit

7. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

8. Enable prefix prioritization.

   prefix-priority route-policy *route-policy-name*

   By default, prefix prioritization is disabled.

Configuring OSPF PIC

About this task

Prefix Independent Convergence (PIC) enables the device to speed up network convergence by ignoring the number of prefixes.

Restrictions and guidelines for OSPF PIC

When both OSPF PIC and OSPF FRR are configured, OSPF FRR takes effect.

OSPF PIC applies only to inter-area routes and external routes.

Enabling OSPF PIC

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable PIC for OSPF.

   pic [**additional-path-always**]

   By default, OSPF PIC is enabled.

Configuring BFD for OSPF PIC

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable BFD control packet mode for OSPF PIC.

   ospf primary-path-detect bfd ctrl

   By default, BFD control packet mode is disabled for OSPF PIC.

   This mode requires BFD configuration on both OSPF routers on the link.

Configuring BFD echo packet mode for OSPF PIC

1. Enter system view.
   system-view
2. Configure the source IP address of BFD echo packets.
   bfd echo-source-ip *ip-address*
   By default, the source IP address of BFD echo packets is not configured.
   The source IP address cannot be on the same network segment as any local interfaces.
   For more information about this command, see *High Availability Command Reference*.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable BFD echo packet mode for OSPF PIC.
   ospf primary-path-detect bfd echo
   By default, BFD echo packet mode is disabled for OSPF PIC.
   This mode requires BFD configuration on one OSPF router on the link.

Configuring advanced OSPF features

Configuring stub routers

About this task

A stub router is used for traffic control. It reports its status as a stub router to neighboring OSPF routers. The neighboring routers can have a route to the stub router, but they do not use the stub router to forward data.

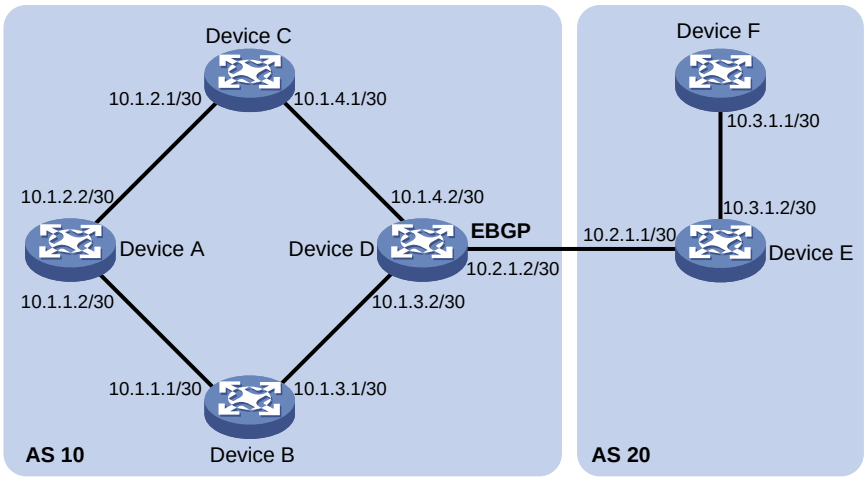
Router LSAs from the stub router might contain different link type values. A value of 3 means a link to a stub network, and the cost of the link will not be changed by default. To set the cost of the link to 65535, specify the **include-stub** keyword in the **stub-router** command. A value of 1, 2 or 4 means a point-to-point link, a link to a transit network, or a virtual link. On such links, a maximum cost value of 65535 is used. Neighbors do not send packets to the stub router as long as they have a route with a smaller cost.

As shown in Figure 13, Device A, Device B, Device C, and Device D run OSPF. Device C is the backup device for Device B. Device A and Device B, Device A and Device C, Device B and Device D, and Device C and Device D are IBGP peers.

When the network is stable, BGP and OSPF can finish route convergence completely. Device A forwards traffic destined for 10.3.1.0/30 to Device B. If Device B fails, Device C will take over to forward the traffic from Device A.

After Device B recovers, Device A immediately forwards the traffic Device B. However, BGP might have not finished route convergence when OSPF has finished route convergence, because OSPF has a higher route convergence speed than BGP. In this situation, Device B does not have a route to 10.3.1.0/30 and will discard the traffic destined for 10.3.1.0/30. To resolve this issue, configure Device B to act as a stub router during BGP route convergence. Device A will forward traffic to Device C instead of Device B until Device B finishes BGP route convergence.

Figure 13 Network diagram



Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure the router as a stub router.

   stub-router [**external-lsa** [*max-metric-value*] | **include-stub** | **on-startup** { *seconds* | **wait-for-bgp** [*seconds*] } | **summary-lsa** [*max-metric-value*]] *

   By default, the router is not configured as a stub router.

   A stub router is not related to a stub area.

Configuring OSPF isolation

About this task

Isolation is a method used for network device maintenance. It gracefully removes a device from the packet forwarding path for maintenance and gracefully adds the device to the network after maintenance.

To reduce impact on traffic forwarding, you can isolate a device before upgrading it. OSPF isolation works as follows:

1. After OSPF isolation is enabled for a device, OSPF increases the link cost in LSAs advertised by the device based on the following rules:
   - The link cost in Type-1 LSAs (Router LSAs) is increased to 65535.
   - The link cost in the following LSAs is increased to 16711680:
     - Type-3 LSAs (Network summary LSAs).
     - Type-5 LSAs (AS external LSAs).
     - Type-7 LSAs (NSSA external LSAs).
2. Each neighbor of the device reselects an optimal route based on the LSAs and stops forwarding traffic to the device. The device is fully isolated from the network and you can upgrade the device.

3. After the maintenance, disable OSPF isolation on the device to restore its link cost and gracefully add it back to the network.

Restrictions and guidelines

Both the **isolate enable** and **stub-router external-lsa 16711680 summary-lsa 16711680 include-stub** commands can isolate the device from the network.

When you execute both the **isolate enable** and **stub-router** commands, follow these restrictions and guidelines:

- If you specify the **external-lsa** and **summary-lsa** keywords in the **stub-router** command, the higher one of the two link costs provided by the isolation feature and the stub router feature takes effect.
- If the **on-startup** keyword is specified in the **stub-router** command, traffic forwarding path selection is affected only by the isolation feature when the stub router feature does not take effect.

Procedure

1. Enter system view.

   system-view
2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Configure OSPF isolation.
   - Enable OSPF isolation to gracefully isolate the device from the network.

     isolate enable
   - Disable OSPF isolation to gracefully add the device back to the network.

     undo isolate enable

   By default, OSPF isolation is disabled.

Configuring OSPF shutdown

About this task

For maintenance purposes, you can use this feature to shut down OSPF processes on the device with small impact on the network. If you shut down a process, it will perform the following operations:

- Send 1-way hello packets to its neighbors.

  On receipt of the packets, the neighbors disconnect from the OSPF process and use the backup path to forward traffic.
- Stop receiving and sending OSPF packets.
- Clear its neighbor, LSDB, and OSPF route information.

After maintenance, you can use the **undo shutdown process** command to restart the process for neighbor relationship re-establishment.

This feature shuts down an OSPF process while retaining the process-associated settings to facilitate your maintenance.

Procedure

1. Enter system view.

   system-view
2. Enter OSPF view.

ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure OSPF shutdown.
   - Shut down an OSPF process to isolate it from the network.

     shutdown process
   - Restart the OSPF process to add it back to the network.

     undo shutdown process

   By default, the OSPF process is not shut down.

Enabling compatibility with RFC 1583

About this task

RFC 1583 specifies a different method than RFC 2328 for selecting the optimal route to a destination in another AS. When multiple routes are available to the ASBR, OSPF selects the optimal route by using the following procedure:

1. Selects the route with the highest preference.
   - If RFC 2328 is compatible with RFC 1583, all these routes have equal preference.
   - If RFC 2328 is not compatible with RFC 1583, the intra-area route in a non-backbone area is preferred to reduce the burden of the backbone area. The inter-area route and intra-area route in the backbone area have equal preference.
2. Selects the route with the lower cost if two routes have equal preference.
3. Selects the route with the larger originating area ID if two routes have equal cost.

Restrictions and guidelines

To avoid routing loops, set identical RFC 1583-compatibility on all routers in a routing domain.

Procedure

1. Enter system view.

   system-view
2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable compatibility with RFC 1583.

   rfc1583 compatible

   By default, compatibility with RFC 1583 is enabled.

Configuring OSPF GR

About OSPF GR

GR ensures forwarding continuity when a routing protocol restarts or an active/standby switchover occurs.

Two routers are required to complete a GR process. The following are router roles in a GR process:

- **GR restarter**—Graceful restarting router. It must have GR capability.
- **GR helper**—A neighbor of the GR restarter. It helps the GR restarter to complete the GR process.

OSPF GR has the following types:

- **IETF GR**—Uses Opaque LSAs to implement GR.
- **Non-IETF GR**—Uses link local signaling (LLS) to advertise GR capability and uses out of band synchronization to synchronize the LSDB.

A device can act as a GR restarter and GR helper at the same time.

Configuring OSPF GR restarter

Configuring the IETF OSPF GR restarter

1. Enter system view.
   system-view
2. Enable OSPF and enter its view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable opaque LSA reception and advertisement capability.
   opaque-capability enable
   By default, opaque LSA reception and advertisement capability is enabled.
4. Enable the IETF GR.
   graceful-restart ietf [**global** | **planned-only**] *
   By default, the IETF GR capability is disabled.
5. (Optional.) Set the GR interval.
   graceful-restart interval *interval*
   By default, the GR interval is 120 seconds.

Configuring the non-IETF OSPF GR restarter

1. Enter system view.
   system-view
2. Enable OSPF and enter its view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable the link-local signaling capability.
   enable link-local-signaling
   By default, the link-local signaling capability is disabled.
4. Enable the out-of-band re-synchronization capability.
   enable out-of-band-resynchronization
   By default, the out-of-band re-synchronization capability is disabled.
5. Enable non-IETF GR.
   graceful-restart [**nonstandard**] [**global** | **planned-only**] *
   By default, non-IETF GR capability is disabled.
6. (Optional.) Set the GR interval.
   graceful-restart interval *interval*
   By default, the GR interval is 120 seconds.

Configuring OSPF GR helper

Configuring the IETF OSPF GR helper

1. Enter system view.

   system-view

2. Enable OSPF and enter its view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable opaque LSA reception and advertisement capability.

   opaque-capability enable

   By default, opaque LSA reception and advertisement capability is enabled.

4. Enable GR helper capability.

   graceful-restart helper enable [**planned-only**]

   By default, GR helper capability is enabled.

5. (Optional.) Enable strict LSA checking for the GR helper.

   graceful-restart helper strict-lsa-checking

   By default, strict LSA checking for the GR helper is disabled.

   When an LSA change on the GR helper is detected, the GR helper device exits the GR helper mode.

Configuring the non-IETF OSPF GR helper

1. Enter system view.

   system-view

2. Enable OSPF and enter its view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable the link-local signaling capability.

   enable link-local-signaling

   By default, the link-local signaling capability is disabled.

4. Enable the out-of-band re-synchronization capability.

   enable out-of-band-resynchronization

   By default, the out-of-band re-synchronization capability is disabled.

5. Enable GR helper.

   graceful-restart helper enable

   By default, GR helper is enabled.

6. (Optional.) Enable strict LSA checking for the GR helper.

   graceful-restart helper strict-lsa-checking

   By default, strict LSA checking for the GR helper is disabled.

   When an LSA change on the GR helper is detected, the GR helper device exits the GR helper mode.

Triggering OSPF GR

About this task

You can trigger OSPF GR by performing an active/standby switchover or using the **reset ospf process** command.

Procedure

To trigger OSPF GR, execute the **reset ospf** [*process-id*] **process graceful-restart** command in user view.

Configuring OSPF NSR

About this task

Nonstop routing (NSR) backs up OSPF link state information from the active process to the standby process. After an active/standby switchover, NSR can complete link state recovery and route regeneration without tearing down adjacencies or impacting forwarding services.

Choose OSPF NSR or OSPF GR according to the actual situation and make sure data forwarding is not interrupted. The differences between OSPF NSR and OSPF GR as follows:

- **Device requirements**—The active and standby processes of OSPF must run on different IRF member devices. To run OSPF NSR, you must set up an IRF fabric that contains a minimum of two member devices. To run OSPF GR, IRF fabric setup is not required. If an IRF fabric already exists, you can run OSPF GR when the IRF fabric contains only one member device.
- **OSPF neighbor requirements**—When you use OSPF NSR, OSPF neighbors cannot detect process restart or active/standby switchover on the local device. OSPF NSR can recover OSPF route information without requiring the cooperation of OSPF neighbors. OSPF GR requires OSPF neighbors to have the GR helper capability to help the GR restarter complete the GR process.

Restrictions and guidelines

When both GR and NSR are configured for OSPF on a device, the following rules apply:

- OSPF NSR has a higher priority than OSPF GR. When the active process goes down, the device does not act as the GR restarter and performs NSR to ensure nonstop forwarding.
- When the following conditions exist, the GR process might fail even if OSPF NSR is enabled on the GR helper:
  - OSPF is restarted through GR.
  - An active/standby switchover is triggered on the GR helper before the GR process finishes.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable OSPF NSR.

   non-stop-routing

   By default, OSPF NSR is enabled.

   This command takes effect only for the current process. As a best practice, enable OSPF NSR for each process if multiple OSPF processes exist.

Configuring BFD for OSPF

About BFD for OSPF

OSPF sends hello packets to neighbors at specific intervals to detect their state. If OSPF does not receive a hello packet from a neighbor within a specific period, OSPF considers the neighbor down. Such a failure detection method cannot achieve high data reliability, because it takes a long time and causes data loss.

To resolve this issue, configure Bidirectional Forwarding Detection (BFD) for OSPF. This feature improves the route convergence speed against link state changes through fast link state detection. For more information about BFD, see *High Availability Configuration Guide*.

OSPF supports the following BFD detection modes:

- **Bidirectional control detection**—Requires BFD configuration to be made on both OSPF routers on the link.
- **Single-hop echo detection**—Requires BFD configuration to be made on one OSPF router on the link.

Restrictions and guidelines

In the current software version, OSPF BFD is not supported on VSI interfaces.

Procedure

1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Enable BFD bidirectional control detection.
   `ospf bfd enable`
   By default, BFD bidirectional control detection is disabled.
   Both ends of a BFD session must be on the same network segment and in the same area.

Configuring single-hop echo detection

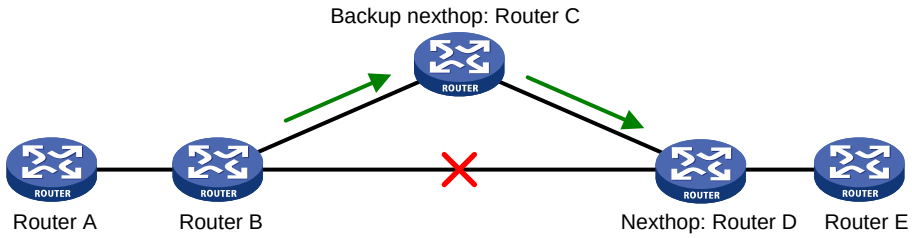
1. Enter system view.
   `system-view`
2. Configure the source address of echo packets.
   `bfd echo-source-ip` *`ip-address`*
   By default, the source address of echo packets is not configured.
   The source IP address cannot be on the same network segment as any local interfaces.
   For more information about this command, see *High Availability Command Reference*.
3. Enter interface view.
   `interface` *`interface-type interface-number`*
4. Enable BFD single-hop echo detection.
   `ospf bfd enable echo`
   By default, BFD single-hop echo detection is disabled.

Configuring OSPF FRR

About OSPF FRR

A link or router failure on a path can cause packet loss until OSPF completes routing convergence based on the new network topology. FRR enables fast rerouting to minimize the impact of link or node failures.

Figure 14 Network diagram for OSPF FRR



As shown in Figure 14, configure FRR on Router B by using a routing policy to specify a backup next hop. When the primary link fails, OSPF directs packets to the backup next hop. At the same time, OSPF calculates the shortest path based on the new network topology. It forwards packets over the path after network convergence.

You can configure OSPF FRR to calculate a backup next hop by using the loop free alternate (LFA) algorithm, or specify a backup next hop by using a routing policy.

Restrictions and guidelines for OSPF FRR

When both OSPF PIC and OSPF FRR are configured, OSPF FRR takes effect.

Configuring OSPF FRR to use the LFA algorithm to calculate a backup next hop

Restrictions and guidelines

Do not use the **fast-reroute lfa** command together with the **vlink-peer** command.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. (Optional.) Enable LFA on an interface.
   ospf fast-reroute lfa-backup
   By default, the interface is enabled with LFA and it can be selected as a backup interface.
4. Return to system view.
   quit
5. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

6. Enable OSPF FRR to use the LFA algorithm to calculate a backup next hop.

   fast-reroute lfa [**abr-only** | **ecmp-shared**]

   By default, OSPF FRR is disabled.

   If **abr-only** is specified, the route to the ABR is selected as the backup path.

Configuring OSPF FRR to use a backup next hop specified in a routing policy

About this task

Before you perform this task, use the **apply fast-reroute backup-interface** command to specify a backup next hop in a routing policy for OSPF FRR. For more information about the **apply fast-reroute backup-interface** command and routing policy configuration, see "Configuring routing policies."

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable OSPF FRR to use a backup next hop specified in a routing policy.

   fast-reroute route-policy *route-policy-name*

   By default, OSPF FRR is disabled.

Configuring BFD for OSPF FRR

About this task

By default, OSPF FRR does not use BFD to detect primary link failures. To speed up OSPF convergence, enable BFD control packet mode for OSPF FRR to detect primary link failures. This mode requires BFD configuration on both OSPF routers on the link.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable BFD control packet mode for OSPF FRR.

   ospf primary-path-detect bfd ctrl

   By default, BFD control packet mode is disabled for OSPF FRR.

Configuring BFD echo packet mode for OSPF FRR

About this task

By default, OSPF FRR does not use BFD to detect primary link failures. To speed up OSPF convergence, enable BFD echo packet mode for OSPF FRR to detect primary link failures. This mode requires BFD configuration on one OSPF router on the link.

Procedure

1. Enter system view.
   system-view
2. Configure the source IP address of BFD echo packets.
   bfd echo-source-ip *ip-address*
   By default, the source IP address of BFD echo packets is not configured.
   The source IP address cannot be on the same network segment as any local interface's IP address.
   For more information about this command, see *High Availability Command Reference*.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable BFD echo packet mode for OSPF FRR.
   ospf primary-path-detect bfd echo
   By default, BFD echo packet mode is disabled for OSPF FRR.

Configuring OSPF authentication

About OSPF area and interface authentication

Perform this task to configure OSPF area and interface authentication.

OSPF adds the authentication information into sent packets, and uses the key to authenticate received packets. Only packets that pass the authentication can be received. If a packet fails the authentication, the OSPF neighbor relationship cannot be established.

If you configure OSPF authentication for both an area and an interface in that area, the interface uses the OSPF authentication configured on it.

Restrictions and guidelines for configuring OSPF authentication

OSPF supports the MD5, HMAC-MD5, HMAC-SHA-256, and HMAC-SM3 authentication algorithms.

The ID of keys used for authentication can only be in the range of 0 to 255.

Configuring OSPF area authentication

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter area view.
   area *area-id*
4. Configure area authentication mode.
   - Configure HMAC-MD5/HMAC-SHA-256/HMAC-SM3/MD5 authentication.
     authentication-mode { **hmac-md5** | **hmac-sha-256** | **hmac-sm3** | **md5** } [*key-id* { **cipher** | **plain** } *string*]

- Configure simple authentication.

  authentication-mode simple [{ **cipher** | **plain** } *string*]
- Configure keychain authentication.

  authentication-mode keychain *keychain-name*

  For information about keychains, see *Security Configuration Guide*.

By default, no authentication is configured.

All routers in an area must use the same authentication mode. If a key is configured, all routers must be configured with the same key.

Configuring OSPF interface authentication

About this task

To prevent an interface from authenticating OSPF packets, configure the None authentication mode on the interface. Then, the interface will not inherit the authentication configuration of the area.

Restrictions and guidelines

The local interface and its peer interface must use the same authentication mode. If a key is configured, the local and peer interfaces must be configured with the same key.

If you configure the None authentication mode on an interface, make sure its peer interface uses the None authentication mode or is not configured with any authentication modes.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure interface authentication mode.
   - Configure simple authentication.

     ospf authentication-mode simple [{ **cipher** | **plain** } *string*]
   - *Configure* the None authentication mode.

     ospf authentication-mode none
   - Configure HMAC-MD5/HMAC-SHA-256/HMAC-SM3/MD5 authentication.

     ospf authentication-mode { **hmac-md5** | **hmac-sha-256** | **hmac-sm3** | **md5** } [*key-id* { **cipher** | **plain** } *string*]
   - Configure keychain authentication.

     ospf authentication-mode keychain *keychain-name*

     For information about keychains, see *Security Configuration Guide*.

   By default, no authentication is configured.

Configuring GTSM for OSPF

About GTSM

The Generalized TTL Security Mechanism (GTSM) protects the device by comparing the TTL value in the IP header of incoming OSPF packets against a valid TTL range. If the TTL value is within the valid TTL range, the packet is accepted. If not, the packet is discarded.

The valid TTL range is from 255 – the configured hop count + 1 to 255.

When GTSM is configured, the OSPF packets sent by the device have a TTL of 255.

Restrictions and guidelines for GTSM

To use GTSM, you must configure GTSM on both the local and peer devices. You can specify different *hop-count* values for them.

The GTSM configuration in OSPF area view applies to all OSPF interfaces in the area. The GTSM configuration in interface view takes precedence over the configuration in OSPF area view.

Configuring GTSM in OSPF area view

1. Enter system view.

   system-view
2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter OSPF area view.

   area *area-id*
4. Enable GTSM for the OSPF area.

   ttl-security [**hops** *hop-count*]

   By default, GTSM is disabled for the OSPF area.

Configuring GTSM in interface view

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure GTSM for the interface.
   - Enable GTSM for the interface.

     ospf ttl-security [**hops** *hop-count*]
   - Disable GTSM for the interface.

     ospf ttl-security disable

     Disable GTSM for an interface when the following conditions exist:
     - The area to which the interface belongs is enabled with GTSM.
     - The neighbor of the interface does not support GTSM.

   By default, an interface uses the GTSM configuration of the area to which the interface belongs.

Configuring OSPF logging, SNMP notifications, and troubleshooting

Logging neighbor state changes

About this task

Perform this task to enable output of neighbor state change logs to the information center. The information center processes the logs according to user-defined output rules (whether and where to output logs). For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable the logging of neighbor state changes.

   log-peer-change

   By default, this feature of logging neighbor state changes is enabled.

Configuring the OSPF logging feature

About this task

OSPF logs include LSA aging logs, route calculation logs, neighbor logs, OSPF route logs, and self-originated and received LSA logs.

Procedure

1. Enter system view.

   system-view

2. Enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the number of OSPF logs.

   event-log { **hello** { **received** [**abnormal** | **dropped**] | **sent** [**abnormal** | **failed**] } | **lsa-flush** | **peer** | **spf** } **size** *count*

   By default, the device can generate a maximum of 100 OSPF logs for each type.

Configuring OSPF network management

About this task

This task involves the following configurations:

- Bind an OSPF process to MIB so that you can use network management software to manage the specified OSPF process.
- Enable SNMP notifications for OSPF to report important events.

- Configure the SNMP notification output interval and the maximum number of SNMP notifications that can be output at each interval.

To report critical OSPF events to an NMS, enable SNMP notifications for OSPF. For SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.
   system-view
2. Bind MIB to an OSPF process.
   ospf mib-binding *process-id*
   By default, MIB is bound to the process with the smallest process ID.
3. Enable SNMP notifications for OSPF.
   snmp-agent trap enable ospf [**authentication-failure** | **bad-packet** | **config-error** | **dr-ip-address-conflict** | **grhelper-status-change** | **grrestarter-status-change** | **if-networktype-mismatch** | **if-state-change** | **inter-area-router-id-conflict** | **intra-area-router-id-conflict** | **lsa-maxage** | **lsa-originate** | **lsdb-approaching-overflow** | **lsdb-overflow** | **neighbor-state-change** | **nssatranslator-status-change** | **retransmit** | **virt-authentication-failure** | **virt-bad-packet** | **virt-config-error** | **virt-retransmit** | **virtgrhelper-status-change** | **virtif-state-change** | **virtneighbor-state-change** | **vlinkid-need-change**] *
   By default, SNMP notifications for OSPF are enabled.
4. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
5. Configure the SNMP notification output interval and the maximum number of SNMP notifications that can be output at each interval.
   snmp trap rate-limit interval *trap-interval* **count** *trap-number*
   By default, OSPF outputs a maximum of seven SNMP notifications within 10 seconds.

Setting the maximum number of OSPF neighbor relationship troubleshooting entries

1. Enter system view.
   system-view
2. Set the maximum number of OSPF neighbor relationship troubleshooting entries.
   ospf troubleshooting max-number *number*
   By default, OSPF can record a maximum of 100 neighbor relationship troubleshooting entries.

Display and maintenance commands for OSPF

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display summary route information	**display ospf** [*process-id*] [**area** *area-id*]

Task	Command
on the OSPF ABR.	**abr-summary** [*ip-address* { *mask-length* | *mask* }] [**verbose**]
Display OSPF FRR backup next hop information.	**display ospf** [*process-id*] [**area** *area-id*] **fast-reroute lfa-candidate**
Display OSPF topology information.	**display ospf** [*process-id*] [**area** *area-id*] **spf-tree** [**verbose**]
Display OSPF process information.	**display ospf** [*process-id*] [**verbose**]
Display OSPF ABR and ASBR information.	**display ospf** [*process-id*] **abr-asbr** [**verbose**]
Display OSPF ASBR route summarization information.	**display ospf** [*process-id*] **asbr-summary** [*ip-address* { *mask-length* | *mask* }]
Display OSPF log information.	**display ospf** [*process-id*] **event-log** { **lsa-flush** | **peer** [*neighbor-id*] [**slot** *slot-number*] | **spf** }
Display OSPF log information about received or sent hello packets.	**display ospf** [*process-id*] **event-log hello** { **received** [**abnormal** | **dropped**] | **sent** } [*neighbor-id*] [**slot** *slot-number*] **display ospf** [*process-id*] **event-log hello sent** { **abnormal** | **failed** } [*neighbor-address*] [**slot** *slot-number*]
Display OSPF GR information.	**display ospf** [*process-id*] **graceful-restart** [**verbose**]
Display OSPF interface information.	**display ospf** [*process-id*] **interface** [*interface-type interface-number* | **verbose**]
Display information about hello packets sent by OSPF interfaces.	**display ospf** [*process-id*] **interface** [*interface-type interface-number*] **hello**
Display OSPF LSDB information.	**display ospf** [*process-id*] [**area** *area-id*] **lsdb** { **asbr** | **network** | **nssa** | **opaque-area** | **opaque-link** | **router** | **summary** } [*link-state-id*] [**originate-router** *advertising-router-id* | **self-originate**] **display ospf** [*process-id*] **lsdb** [**brief** | **originate-router** *advertising-router-id* | **self-originate**] **display ospf** [*process-id*] **lsdb** { **ase** | **opaque-as** } [*link-state-id*] [**originate-router** *advertising-router-id* | **self-originate**]
Display OSPF next hop information.	**display ospf** [*process-id*] **nexthop**
Display OSPF NSR information.	**display ospf** [*process-id*] **non-stop-routing status**
Display OSPF neighbor information.	**display ospf** [*process-id*] **peer** [**hello** | **verbose**] [*interface-type interface-number*] [*neighbor-id*]
Display neighbor statistics for OSPF	**display ospf** [*process-id*] **peer statistics**

Task	Command
areas.	
Display OSPF request queue information.	**display ospf** [*process-id*] **request-queue** [*interface-type interface-number*] [*neighbor-id*]
Display OSPF retransmission queue information.	**display ospf** [*process-id*] **retrans-queue** [*interface-type interface-number*] [*neighbor-id*]
Display OSPF routing table information.	**display ospf** [*process-id*] **routing** [*ip-address* { *mask-length* | *mask* }] [**interface** *interface-type interface-number*] [**nexthop** *nexthop-address*] [**verbose**]
Display OSPF statistics.	**display ospf** [*process-id*] **statistics** [**error** | **packet** [**hello** | *interface-type interface-number*]]
Display OSPF virtual link information.	**display ospf** [*process-id*] **vlink**
Display the global router ID.	**display router id**
Display OSPF neighbor relationship troubleshooting information.	**display ospf troubleshooting**
Clear OSPF log information.	**reset ospf** [*process-id*] **event-log** [**lsa-flush** | **peer** | **spf**]
Clear OSPF log information about received or sent hello packets.	**reset ospf** [*process-id*] **event-log hello** { **received** [**abnormal** | **dropped**] | **sent** [**abnormal** | **failed**] } [**slot** *slot-number*]
Restart an OSPF process.	**reset ospf** [*process-id*] **process** [**graceful-restart**]
Re-enable OSPF route redistribution.	**reset ospf** [*process-id*] **redistribution**
Clear OSPF statistics.	**reset ospf** [*process-id*] **statistics**
Clear OSPF neighbor relationship troubleshooting information.	**reset ospf troubleshooting**

OSPF configuration examples

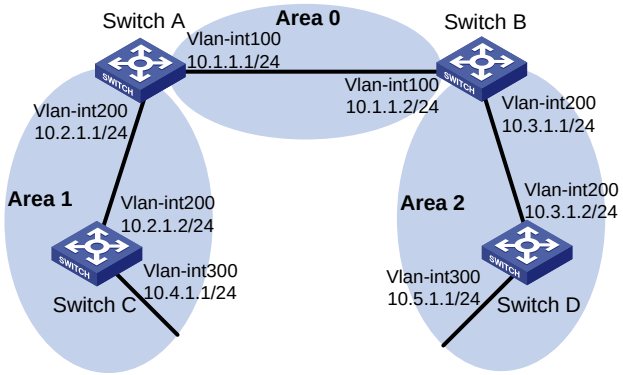
Example: Configuring basic OSPF

Network configuration

As shown in Figure 15:

- Enable OSPF on all switches, and split the AS into three areas.
- Configure Switch A and Switch B as ABRs.

Figure 15 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable OSPF:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] router id 10.2.1.1
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] router id 10.3.1.1
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] area 2
[SwitchB-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.2] quit
[SwitchB-ospf-1] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] router id 10.4.1.1
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] network 10.4.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

   # Configure Switch D.

```
<SwitchD> system-view
[SwitchD] router id 10.5.1.1
[SwitchD] ospf
[SwitchD-ospf-1] area 2
[SwitchD-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] network 10.5.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] quit
[SwitchD-ospf-1] quit
```

Verifying the configuration

Display information about neighbors on Switch A.

```
[SwitchA] display ospf peer verbose

          OSPF Process 1 with Router ID 10.2.1.1
                  Neighbors

 Area 0.0.0.0 interface 10.1.1.1(Vlan-interface100)'s neighbors
 Router ID: 10.3.1.1         Address: 10.1.1.2         GR State: Normal
   State: Full  Mode: Nbr is master  Priority: 1
   DR: 10.1.1.1  BDR: 10.1.1.2  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 37  sec
   Neighbor is up for 06:03:59
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 5
   BFD status: Disabled

Area 0.0.0.1 interface 10.2.1.1(Vlan-interface200)'s neighbors
 Router ID: 10.4.1.1         Address: 10.2.1.2         GR State: Normal
   State: Full  Mode: Nbr is master  Priority: 1
   DR: 10.2.1.1  BDR: 10.2.1.2  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 32  sec
   Neighbor is up for 06:03:12
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 5
   BFD status: Disabled
```

Display OSPF routing information on Switch A.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 10.2.1.1
                  Routing Table

                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        1        Transit 10.2.1.1        10.2.1.1        0.0.0.1
 10.3.1.0/24        2        Inter   10.1.1.2        10.3.1.1        0.0.0.0
```

```
 10.4.1.0/24        2        Stub    10.2.1.2        10.4.1.1        0.0.0.1
 10.5.1.0/24        3        Inter   10.1.1.2        10.3.1.1        0.0.0.0
 10.1.1.0/24        1        Transit 10.1.1.1        10.2.1.1        0.0.0.0

 Total nets: 5
 Intra area: 3  Inter area: 2  ASE: 0  NSSA: 0
```

Display OSPF routing information on Switch D.

```
[SwitchD] display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
                  Routing Table

                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        3        Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        1        Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24        4        Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24        1        Stub    10.5.1.1        10.5.1.1        0.0.0.2
 10.1.1.0/24        2        Inter   10.3.1.1        10.3.1.1        0.0.0.2

 Total nets: 5
 Intra area: 2  Inter area: 3  ASE: 0  NSSA: 0
```

On Switch D, ping the IP address 10.4.1.1 to test reachability.

```
[SwitchD] ping 10.4.1.1

Ping 10.4.1.1 (10.4.1.1): 56 data bytes, press CTRL+C to break
56 bytes from 10.4.1.1: icmp_seq=0 ttl=253 time=1.549 ms
56 bytes from 10.4.1.1: icmp_seq=1 ttl=253 time=1.539 ms
56 bytes from 10.4.1.1: icmp_seq=2 ttl=253 time=0.779 ms
56 bytes from 10.4.1.1: icmp_seq=3 ttl=253 time=1.702 ms
56 bytes from 10.4.1.1: icmp_seq=4 ttl=253 time=1.471 ms
--- Ping statistics for 10.4.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.779/1.408/1.702/0.323 ms
```

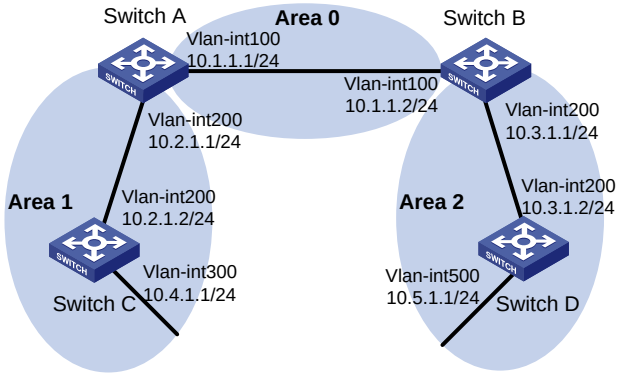
Example: Configuring OSPF route redistribution

Network configuration

As shown in Figure 16:

- Enable OSPF on all the switches.
- Split the AS into three areas.
- Configure Switch A and Switch B as ABRs.
- Configure Switch C as an ASBR to redistribute external routes (static routes).

Figure 16 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable OSPF (see "Example: Configuring basic OSPF").
3. Configure OSPF to redistribute routes:

   # On Switch C, configure a static route destined for network 3.1.2.0/24.

   ```
   <SwitchC> system-view
   [SwitchC] ip route-static 3.1.2.1 24 10.4.1.2
   ```

   # On Switch C, configure OSPF to redistribute static routes.

   ```
   [SwitchC] ospf 1
   [SwitchC-ospf-1] import-route static
   ```

Verifying the configuration

Display the ABR/ASBR information on Switch D.

```
<SwitchD> display ospf abr-asbr

          OSPF Process 1 with Router ID 10.5.1.1
                  Routing Table to ABR and ASBR


              Topology base (MTID 0)
 Type        Destination     Area            Cost  Nexthop         RtType
 Intra       10.3.1.1        0.0.0.2         10    10.3.1.1        ABR
 Inter       10.4.1.1        0.0.0.2         22    10.3.1.1        ASBR
```

Display the OSPF routing table on Switch D.

```
<SwitchD> display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
                   Routing Table

               Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        22       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        10       Transit 10.3.1.2        10.3.1.1        0.0.0.2
```

```
10.4.1.0/24        25       Inter   10.3.1.1        10.3.1.1        0.0.0.2
10.5.1.0/24        10       Stub    10.5.1.1        10.5.1.1        0.0.0.2
10.1.1.0/24        12       Inter   10.3.1.1        10.3.1.1        0.0.0.2

Routing for ASEs
Destination        Cost     Type    Tag             NextHop         AdvRouter
3.1.2.0/24         1        Type2   1               10.3.1.1        10.4.1.1

Total nets: 6
Intra area: 2  Inter area: 3  ASE: 1  NSSA: 0
```

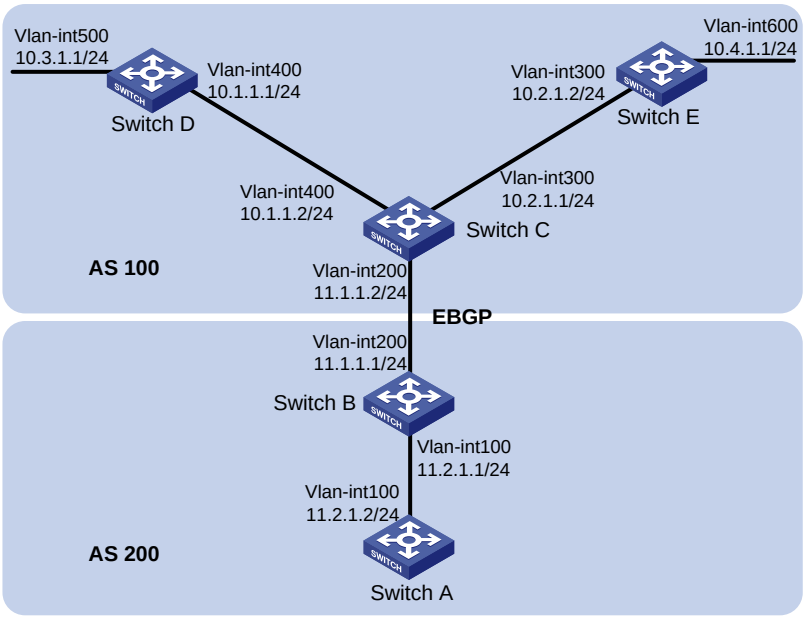
Example: Configuring OSPF route summarization

Network configuration

As shown in Figure 17:

- Configure OSPF on Switch A and Switch B in AS 200.
- Configure OSPF on Switch C, Switch D, and Switch E in AS 100.
- Configure an EBGP connection between Switch B and Switch C. Configure Switch B and Switch C to redistribute OSPF routes and direct routes into BGP and BGP routes into OSPF.
- Configure Switch B to advertise only summary route 10.0.0.0/8 to Switch A.

Figure 17 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable OSPF:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] router id 11.2.1.2
   [SwitchA] ospf
   [SwitchA-ospf-1] area 0
   ```

```
[SwitchA-ospf-1-area-0.0.0.0] network 11.2.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] router id 11.2.1.1
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 11.2.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] router id 11.1.1.2
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] router id 10.3.1.1
[witchD] ospf
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] router id 10.4.1.1
[SwitchE] ospf
[SwitchE-ospf-1] area 0
[SwitchE-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchE-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[SwitchE-ospf-1-area-0.0.0.0] quit
[SwitchE-ospf-1] quit
```

3. Configure BGP to redistribute OSPF routes and direct routes:

Configure Switch B.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 11.1.1.2 as 100
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 11.1.1.2 enable
[SwitchB-bgp-default-ipv4] import-route ospf
[SwitchB-bgp-default-ipv4] import-route direct
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
[SwitchC] bgp 100
[SwitchC-bgp-default] peer 11.1.1.1 as 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 11.1.1.1 enable
[SwitchC-bgp-default-ipv4] import-route ospf
[SwitchC-bgp-default-ipv4]import-route direct
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

4. Configure Switch B and Switch C to redistribute BGP routes into OSPF:

Configure OSPF to redistribute routes from BGP on Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] import-route bgp
```

Configure OSPF to redistribute routes from BGP on Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] import-route bgp
```

Display the OSPF routing table on Switch A.

```
[SwitchA] display ip routing-table
         Destinations : 16        Routes : 16
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.1.1.0/24         O_ASE2 150  1            11.2.1.1        Vlan100
10.2.1.0/24         O_ASE2 150  1            11.2.1.1        Vlan100
10.3.1.0/24         O_ASE2 150  1            11.2.1.1        Vlan100
10.4.1.0/24         O_ASE2 150  1            11.2.1.1        Vlan100
11.2.1.0/24         Direct 0    0            11.2.1.2        Vlan100
11.2.1.0/32         Direct 0    0            11.2.1.2        Vlan100
11.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
11.2.1.255/32       Direct 0    0            11.2.1.2        Vlan100
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

5. Configure route summarization:

Configure route summarization on Switch B to advertise a summary route 10.0.0.0/8.

```
[SwitchB-ospf-1] asbr-summary 10.0.0.0 8
```

Display the IP routing table on Switch A.

```
[SwitchA] display ip routing-table
         Destinations : 13        Routes : 13
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.0.0.0/8          O_ASE2 150  1            11.2.1.1        Vlan100
11.2.1.0/24         Direct 0    0            11.2.1.2        Vlan100
11.2.1.0/32         Direct 0    0            11.2.1.2        Vlan100
11.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
```

```
11.2.1.255/32       Direct 0    0           11.2.1.2        Vlan100
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0           127.0.0.1       InLoop0
224.0.0.0/4         Direct 0    0           0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0           0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0           127.0.0.1       InLoop0
```

The output shows that routes 10.1.1.0/24, 10.2.1.0/24, 10.3.1.0/24 and 10.4.1.0/24 are summarized into a single route 10.0.0.0/8.

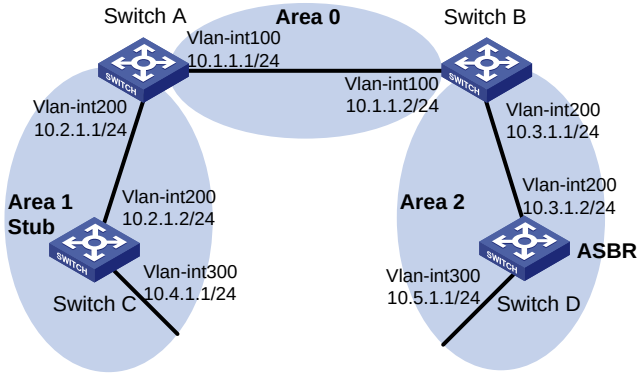
Example: Configuring OSPF stub area

Network configuration

As shown in Figure 18:

- Enable OSPF on all switches, and split the AS into three areas.
- Configure Switch A and Switch B as ABRs to forward routing information between areas.
- Configure Switch D as the ASBR to redistribute static routes.
- Configure Area 1 as a stub area to reduce advertised LSAs without influencing reachability.

Figure 18 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable OSPF (see "Example: Configuring basic OSPF").
3. Configure route redistribution:

   # Configure Switch D to redistribute static routes.

```
<SwitchD> system-view
[SwitchD] ip route-static 3.1.2.1 24 10.5.1.2
[SwitchD] ospf
[SwitchD-ospf-1] import-route static
[SwitchD-ospf-1] quit
```

   # Display ABR/ASBR information on Switch C.

```
<SwitchC> display ospf abr-asbr

        OSPF Process 1 with Router ID 10.4.1.1
                Routing Table to ABR and ASBR
```

```
                Topology base (MTID 0)
 Type        Destination     Area            Cost  Nexthop         RtType
 Intra       10.2.1.1        0.0.0.1         3     10.2.1.1        ABR
 Inter       10.5.1.1        0.0.0.1         7     10.2.1.1        ASBR
```

Display OSPF routing table on Switch C.

```
<SwitchC> display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Table

                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        3        Transit 0.0.0.0         10.2.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop         AdvRouter
3.1.2.0/24          1        Type2   1           10.2.1.1        10.5.1.1

 Total nets: 6
 Intra area: 2  Inter area: 3  ASE: 1  NSSA: 0
```

The output shows that Switch C's routing table contains an AS external route.

4. Configure Area 1 as a stub area:

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] stub
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] stub
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

Display OSPF routing information on Switch C

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Table
```

```
                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 0.0.0.0/0          4        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.2.1.0/24        3        Transit 0.0.0.0         10.2.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Total nets: 6
 Intra area: 2  Inter area: 4  ASE: 0  NSSA: 0
```

The output shows that a default route replaces the AS external route.

Configure Area 1 as a totally stub area.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] stub no-summary
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

Display OSPF routing information on Switch C.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Table

                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 0.0.0.0/0          4        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.2.1.0/24        3        Transit 0.0.0.0         10.4.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1

 Total nets: 3
 Intra area: 2  Inter area: 1  ASE: 0  NSSA: 0
```

The output shows that inter-area routes are removed, and only one external route (a default route) exists on Switch C.

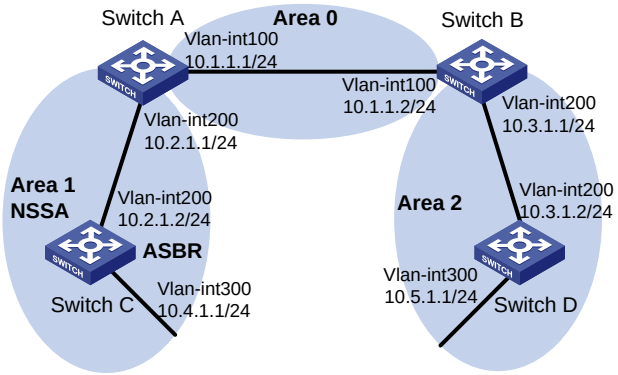
Example: Configuring OSPF NSSA area

Network configuration

As shown in Figure 19:

- Configure OSPF on all switches and split AS into three areas.
- Configure Switch A and Switch B as ABRs to forward routing information between areas.
- Configure Area 1 as an NSSA area and configure Switch C as an ASBR to redistribute static routes into the AS.

Figure 19 Network diagram



Procedure

1. Configure IP addresses for interfaces.
2. Enable OSPF (see "Example: Configuring basic OSPF").
3. Configure Area 1 as an NSSA area:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] nssa
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] nssa
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

   # Display OSPF routing information on Switch C.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Table

              Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        3        Transit 10.2.1.2        10.4.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Total nets: 5
```

```
 Intra area: 2  Inter area: 3  ASE: 0  NSSA: 0
```

4. Configure route redistribution:

   # Configure Switch C to redistribute static routes.

```
[SwitchC] ip route-static 3.1.3.1 24 10.4.1.2
[SwitchC] ospf
[SwitchC-ospf-1] import-route static
[SwitchC-ospf-1] quit
```

   # Display OSPF routing information on Switch D.

```
<SwitchD> display ospf routing
                Topology base (MTID 0)
          OSPF Process 1 with Router ID 10.5.1.1
                   Routing Table
 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        22       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        10       Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24        25       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24        10       Stub    10.5.1.1        10.5.1.1        0.0.0.2
 10.1.1.0/24        12       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop         AdvRouter
 3.1.3.0/24         1        Type2   1           10.3.1.1        10.2.1.1
 Total nets: 6
 Intra area: 2  Inter area: 3  ASE: 1  NSSA: 0
```

   The output shows that an external route imported from the NSSA area exists on Switch D.

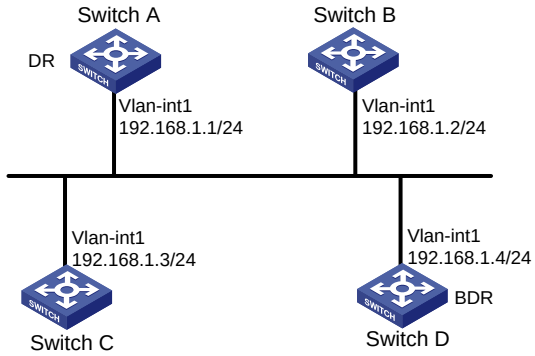
Example: Configuring OSPF DR election

Network configuration

As shown in Figure 20:

- Enable OSPF on Switches A, B, C, and D on the same network.
- Configure Switch D as the DR, and configure Switch C as the BDR.
- Change the router priorities on the interfaces to configure Switch A as the DR and Switch C as the BDR.

Figure 20 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure basic OSPF settings on all switches. (Details not shown.)
   For more information, see "Example: Configuring basic OSPF."
3. Display OSPF neighbor information on Switch A.

```
[SwitchA] display ospf peer verbose

          OSPF Process 1 with Router ID 1.1.1.1
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.1(Vlan-interface1)'s neighbors
 Router ID: 2.2.2.2          Address: 192.168.1.2      GR State: Normal
   State: 2-Way  Mode: None  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 38  sec
   Neighbor is up for 00:01:31
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled

 Router ID: 3.3.3.3          Address: 192.168.1.3      GR State: Normal
   State: Full  Mode: Nbr is master  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 31  sec
   Neighbor is up for 00:01:28
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled

 Router ID: 4.4.4.4          Address: 192.168.1.4      GR State: Normal
   State: Full  Mode: Nbr is master  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 31  sec
   Neighbor is up for 00:01:28
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled
```

   The output shows that Switch D is the DR and Switch C is the BDR.
4. Configure router priorities on interfaces:
   # Configure Switch A.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ospf dr-priority 100
[SwitchA-Vlan-interface1] quit
```

   # Configure Switch B.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ospf dr-priority 0
[SwitchB-Vlan-interface1] quit
```

Configure Switch C.

```
[SwitchC] interface vlan-interface 1
[SwitchC-Vlan-interface1] ospf dr-priority 2
[SwitchC-Vlan-interface1] quit
```

Display neighbor information on Switch D.

```
<SwitchD> display ospf peer verbose

          OSPF Process 1 with Router ID 4.4.4.4
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.4(Vlan-interface1)'s neighbors
 Router ID: 1.1.1.1      Address: 192.168.1.1      GR State: Normal
   State: Full  Mode:Nbr is  slave  Priority: 100
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 31  sec
   Neighbor is up for 00:11:17
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled

 Router ID: 2.2.2.2      Address: 192.168.1.2      GR State: Normal
   State: Full  Mode:Nbr is  slave  Priority: 0
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 35  sec
   Neighbor is up for 00:11:19
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled

 Router ID: 3.3.3.3      Address: 192.168.1.3      GR State: Normal
   State: Full  Mode:Nbr is  slave  Priority: 2
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 33  sec
   Neighbor is up for 00:11:15
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled
```

The output shows that the DR and BDR are not changed, because the priority settings do not take effect immediately.

5. Restart OSPF processes:

Restart the OSPF process of Switch A.

```
<SwitchA> reset ospf 1 process
```

```
Reset OSPF process? [Y/N]:y
```

Restart the OSPF process of Switch B.

```
<SwitchB> reset ospf 1 process
Reset OSPF process? [Y/N]:y
```

Restart the OSPF process of Switch C.

```
<SwitchC> reset ospf 1 process
Reset OSPF process? [Y/N]:y
```

Restart the OSPF process of Switch D.

```
<SwitchD> reset ospf 1 process
Reset OSPF process? [Y/N]:y
```

Display neighbor information on Switch D.

```
<SwitchD> display ospf peer verbose

          OSPF Process 1 with Router ID 4.4.4.4
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.4(Vlan-interface1)'s neighbors
 Router ID: 1.1.1.1          Address: 192.168.1.1      GR State: Normal
   State: Full  Mode: Nbr is slave  Priority: 100
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 39  sec
   Neighbor is up for 00:01:40
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled

 Router ID: 2.2.2.2          Address: 192.168.1.2      GR State: Normal
   State: 2-Way  Mode: None  Priority: 0
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 35  sec
   Neighbor is up for 00:01:44
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled

 Router ID: 3.3.3.3          Address: 192.168.1.3      GR State: Normal
   State: Full  Mode: Nbr is slave  Priority: 2
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Options is 0x02 (-|-|-|-|-|-|E|-)
   Dead timer due in 39  sec
   Neighbor is up for 00:01:41
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 6
   BFD status: Disabled
```

The output shows that Switch A becomes the DR and Switch C becomes the BDR.

If the neighbor state is *full,* Switch D has established an adjacency with the neighbor. If the neighbor state is *2-way*, the two switches are not the DR or the BDR, and they do not exchange LSAs.

Display OSPF interface information.

```
<SwitchA> display ospf interface
                 OSPF Process 1 with Router ID 1.1.1.1
                 Interfaces

 Area: 0.0.0.0
 IP Address      Type      State   Cost  Pri   DR             BDR
 192.168.1.1     Broadcast DR      1     100   192.168.1.1    192.168.1.3

<SwitchB> display ospf interface

          OSPF Process 1 with Router ID 2.2.2.2
                 Interfaces

 Area: 0.0.0.0
 IP Address      Type      State    Cost  Pri   DR             BDR
 192.168.1.2     Broadcast DROther  1     0     192.168.1.1    192.168.1.3
```

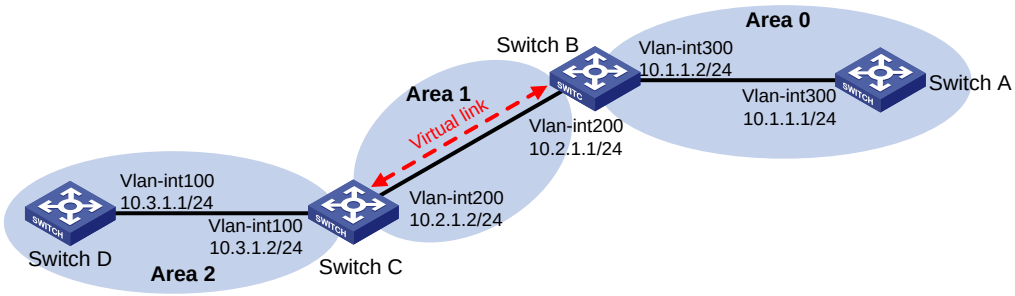
The interface state *DROther* means the interface is not the DR or BDR.

Example: Configuring OSPF virtual link

Network configuration

As shown in Figure 21, configure a virtual link between Switch B and Switch C to connect Area 2 to the backbone area. After configuration, Switch B can learn routes to Area 2.

Figure 21 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable OSPF:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf 1 router-id 1.1.1.1
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf 1 router-id 2.2.2.2
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] area 1
[SwitchB–ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchB–ospf-1-area-0.0.0.1] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf 1 router-id 3.3.3.3
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] area 2
[SwitchC–ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchC–ospf-1-area-0.0.0.2] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ospf 1 router-id 4.4.4.4
[SwitchD-ospf-1] area 2
[SwitchD-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] quit
[SwitchD-ospf-1] quit
```

Display the OSPF routing table on Switch B.

```
[SwitchB] display ospf routing

          OSPF Process 1 with Router ID 2.2.2.2
                  Routing Table

                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        2        Transit 10.2.1.1        3.3.3.3         0.0.0.1
 10.1.1.0/24        2        Transit 10.1.1.2        2.2.2.2         0.0.0.0
 Total nets: 2
 Intra area: 2  Inter area: 0  ASE: 0  NSSA: 0
```

The output shows that Switch B does not have routes to Area 2 because Area 0 is not directly connected to Area 2.

3. Configure a virtual link:

Configure Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] area 1
[SwitchB-ospf-1-area-0.0.0.1] vlink-peer 3.3.3.3
```

```
[SwitchB-ospf-1-area-0.0.0.1] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] vlink-peer 2.2.2.2
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

Display the OSPF routing table on Switch B.

```
[SwitchB] display ospf routing

          OSPF Process 1 with Router ID 2.2.2.2
                  Routing Table

                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        2        Transit 10.2.1.1        3.3.3.3         0.0.0.1
 10.3.1.0/24        5        Inter   10.2.1.2        3.3.3.3         0.0.0.0
 10.1.1.0/24        2        Transit 10.1.1.2        2.2.2.2         0.0.0.0

 Total nets: 3
 Intra area: 2  Inter area: 1  ASE: 0  NSSA: 0
```

The output shows that Switch B has learned the route 10.3.1.0/24 to Area 2.

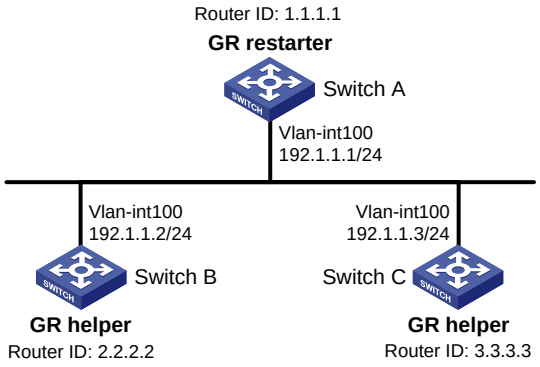
Example: Configuring OSPF GR

Network configuration

As shown in Figure 22:

- Switch A, Switch B, and Switch C that belong to the same AS and the same OSPF routing domain are GR capable.
- Switch A acts as the non-IETF GR restarter. Switch B and Switch C are the GR helpers, and synchronize their LSDBs with Switch A through OOB communication of GR.

Figure 22 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable OSPF:

   # Configure Switch A.

   ```
   SwitchA> system-view
   [SwitchA] router id 1.1.1.1
   [SwitchA] ospf 100
   [SwitchA-ospf-100] area 0
   [SwitchA-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
   [SwitchA-ospf-100-area-0.0.0.0] quit
   [SwitchA-ospf-1] quit
   ```

   # Configure Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] router id 2.2.2.2
   [SwitchB] ospf 100
   [SwitchB-ospf-100] area 0
   [SwitchB-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
   [SwitchB-ospf-100-area-0.0.0.0] quit
   [SwitchB-ospf-1] quit
   ```

   # Configure Switch C.

   ```
   <SwitchC> system-view
   [SwitchC] router id 3.3.3.3
   [SwitchC] ospf 100
   [SwitchC-ospf-100] area 0
   [SwitchC-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
   [SwitchC-ospf-100-area-0.0.0.0] quit
   [SwitchC-ospf-1] quit
   ```

3. Configure OSPF GR:

   # Configure Switch A as the non-IETF OSPF GR restarter: enable the link-local signaling capability, the out-of-band re-synchronization capability, and non-IETF GR capability for OSPF process 100.

   ```
   [SwitchA-ospf-100] enable link-local-signaling
   [SwitchA-ospf-100] enable out-of-band-resynchronization
   [SwitchA-ospf-100] graceful-restart
   [SwitchA-ospf-100] quit
   ```

   # Configure Switch B as the GR helper: enable the link-local signaling capability and the out-of-band re-synchronization capability for OSPF process 100.

   ```
   [SwitchB-ospf-100] enable link-local-signaling
   [SwitchB-ospf-100] enable out-of-band-resynchronization
   ```

   # Configure Switch C as the GR helper: enable the link-local signaling capability and the out-of-band re-synchronization capability for OSPF process 100.

   ```
   [SwitchC-ospf-100] enable link-local-signaling
   [SwitchC-ospf-100] enable out-of-band-resynchronization
   ```

Verifying the configuration

Enable OSPF GR event debugging and restart the OSPF process by using GR on Switch A.

```
<SwitchA> debugging ospf event graceful-restart
<SwitchA> terminal monitor
```

```
<SwitchA> terminal logging level 7
<SwitchA> reset ospf 100 process graceful-restart
Reset OSPF process? [Y/N]:y
%Oct 21 15:29:28:727 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.2(Vlan-interface100) from Full to Down.
%Oct 21 15:29:28:729 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.3(Vlan-interface100) from Full to Down.
*Oct 21 15:29:28:735 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 nonstandard GR Started for OSPF Router
*Oct 21 15:29:28:735 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 created GR wait timer,timeout interval is 40(s).
*Oct 21 15:29:28:735 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 created GR Interval timer,timeout interval is 120(s).
*Oct 21 15:29:28:758 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 created OOB Progress timer for neighbor 192.1.1.3.
*Oct 21 15:29:28:766 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 created OOB Progress timer for neighbor 192.1.1.2.
%Oct 21 15:29:29:902 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.2(Vlan-interface100) from Loading to Full.
*Oct 21 15:29:29:902 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 deleted OOB Progress timer for neighbor 192.1.1.2.
%Oct 21 15:29:30:897 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.3(Vlan-interface100) from Loading to Full.
*Oct 21 15:29:30:897 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 deleted OOB Progress timer for neighbor 192.1.1.3.
*Oct 21 15:29:30:911 2019 SwitchA OSPF/7/DEBUG:
OSPF GR: Process 100 Exit Restart,Reason : DR or BDR change,for neighbor : 192.1.1.3.
*Oct 21 15:29:30:911 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 deleted GR Interval timer.
*Oct 21 15:29:30:912 2019 SwitchA OSPF/7/DEBUG:
OSPF 100 deleted GR wait timer.
%Oct 21 15:29:30:920 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.2(Vlan-interface100) from Full to Down.
%Oct 21 15:29:30:921 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.3(Vlan-interface100) from Full to Down.
%Oct 21 15:29:33:815 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.3(Vlan-interface100) from Loading to Full.
%Oct 21 15:29:35:578 2019 SwitchA OSPF/5/OSPF_NBR_CHG: OSPF 100 Neighbor
192.1.1.2(Vlan-interface100) from Loading to Full.
```

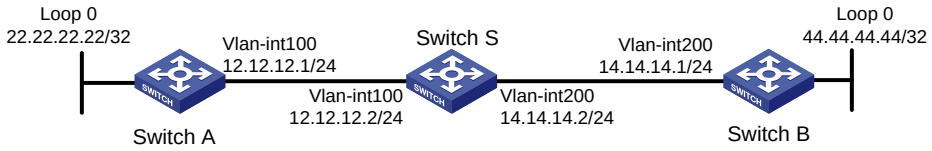
The output shows that Switch A completes GR.

Example: Configuring OSPF NSR

Network configuration

As shown in Figure 23, Switch S, Switch A, and Switch B belong to the same OSPF routing domain. Enable OSPF NSR on Switch S to ensure correct routing when an active/standby switchover occurs on Switch S.

Figure 23 Network diagram



Procedure

1. Configure IP addresses and subnet masks for interfaces on the switches. (Details not shown.)
2. Configure OSPF on the switches to ensure the following: (Details not shown.)
   - Switch S, Switch A, and Switch B can communicate with each other at Layer 3.
   - Dynamic route update can be implemented among them with OSPF.
3. Configure OSPF NSR:

   # Enable OSPF NSR on Switch S.

```
<SwitchS> system-view
[SwitchS] ospf 100
[SwitchS-ospf-100] non-stop-routing
[SwitchS-ospf-100] quit
```

Verifying the configuration

Perform an active/standby switchover on Switch S.

```
[SwitchS] placement reoptimize
Predicted changes to the placement
Service Group(instance name)                    Cur location  New location
------------------------------------------------------------------------------
rib                                             0/0           0/0
staticroute                                     0/0           0/0
ospf                                            0/0           1/0
Continue? [y/n]:y
Re-optimization of the placement start. You will be notified on completion.
Re-optimization of the placement complete. Use 'display placement' to view the new
placement.
```

During the switchover period, display OSPF neighbors on Switch A to verify the neighbor relationship between Switch A and Switch S.

```
<SwitchA> display ospf peer

          OSPF Process 1 with Router ID 2.2.2.1
               Neighbor Brief Information

 Area: 0.0.0.0
 Router ID       Address         Pri Dead-Time  State             Interface
 3.3.3.1         12.12.12.2      1   37         Full/BDR          Vlan100
```

Display OSPF routes on Switch A to verify if Switch A has a route to the loopback interface on Switch B.

```
<SwitchA> display ospf routing

          OSPF Process 1 with Router ID 2.2.2.1
                   Routing Table
```

```
                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 44.44.44.44/32     2        Stub    12.12.12.2      4.4.4.1         0.0.0.0
 14.14.14.0/24      2        Transit 12.12.12.2      4.4.4.1         0.0.0.0
 22.22.22.22/32     0        Stub    22.22.22.22     2.2.2.1         0.0.0.0
 12.12.12.0/24      1        Transit 12.12.12.1      2.2.2.1         0.0.0.0

 Total nets: 4
 Intra area: 4  Inter area: 0  ASE: 0  NSSA: 0
```

Display OSPF neighbors on Switch B to verify the neighbor relationship between Switch B and Switch S.

```
<SwitchB> display ospf peer

          OSPF Process 1 with Router ID 4.4.4.1
               Neighbor Brief Information

 Area: 0.0.0.0
 Router ID       Address         Pri Dead-Time  State             Interface
 3.3.3.1         14.14.14.2      1   39         Full/BDR          Vlan200
```

Display OSPF routes on Switch B to verify if Switch B has a route to the loopback interface on Switch A.

```
<SwitchB> display ospf routing

          OSPF Process 1 with Router ID 4.4.4.1
                   Routing Table

                Topology base (MTID 0)

 Routing for network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 44.44.44.44/32     0        Stub    44.44.44.44     4.4.4.1         0.0.0.0
 14.14.14.0/24      1        Transit 14.14.14.1      4.4.4.1         0.0.0.0
 22.22.22.22/32     2        Stub    14.14.14.2      2.2.2.1         0.0.0.0
 12.12.12.0/24      2        Transit 14.14.14.2      2.2.2.1         0.0.0.0

 Total nets: 4
 Intra area: 4  Inter area: 0  ASE: 0  NSSA: 0
```

The output shows the following when an active/standby switchover occurs on Switch S:

- The neighbor relationships and routing information on Switch A and Switch B have not changed.
- The traffic from Switch A to Switch B has not been impacted.

Example: Configuring BFD for OSPF

Network configuration

As shown in Figure 24, run OSPF on Switch A, Switch B, and Switch C so that they are reachable to each other at the network layer.

- When the link over which Switch A and Switch B communicate through a Layer 2 switch fails, BFD can quickly detect the failure and notify OSPF of the failure.
- Switch A and Switch B then communicate through Switch C.

Figure 24 Network diagram

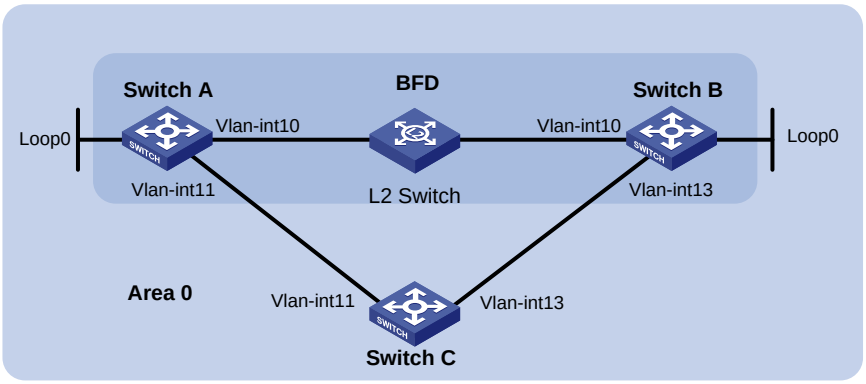


Table 4 Interface and IP address assignment

Device	Interface	IP address
Switch A	Vlan-int10	192.168.0.102/24
Switch A	Vlan-int11	10.1.1.102/24
Switch A	Loop0	121.1.1.1/32
Switch B	Vlan-int10	192.168.0.100/24
Switch B	Vlan-int13	13.1.1.1/24
Switch B	Loop0	120.1.1.1/32
Switch C	Vlan-int11	10.1.1.100/24
Switch C	Vlan-int13	13.1.1.2/24

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable OSPF:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 121.1.1.1 0.0.0.0
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.168.0.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 13.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 120.1.1.1 0.0.0.0
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 13.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

3. Configure BFD:

Enable BFD on Switch A and configure BFD parameters.

```
[SwitchA] bfd session init-mode active
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ospf bfd enable
[SwitchA-Vlan-interface10] bfd min-transmit-interval 500
[SwitchA-Vlan-interface10] bfd min-receive-interval 500
[SwitchA-Vlan-interface10] bfd detect-multiplier 7
[SwitchA-Vlan-interface10] quit
```

Enable BFD on Switch B and configure BFD parameters.

```
[SwitchB] bfd session init-mode active
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ospf bfd enable
[SwitchB-Vlan-interface10] bfd min-transmit-interval 500
[SwitchB-Vlan-interface10] bfd min-receive-interval 500
[SwitchB-Vlan-interface10] bfd detect-multiplier 6
[SwitchB-Vlan-interface10] quit
```

Verifying the configuration

Display the BFD information on Switch A.

```
<SwitchA> display bfd session

 Total sessions: 1     Up sessions: 1     Init mode: Active

 IPv4 session working in control packet mode:

 LD/RD            SourceAddr      DestAddr        State  Holdtime    Interface
3/1              192.168.0.102   192.168.0.100   Up     1700ms      Vlan10
```

Display routes destined for 120.1.1.1/32 on Switch A.

```
<SwitchA> display ip routing-table 120.1.1.1 verbose
Summary count : 1
```

```
Destination: 120.1.1.1/32
   Protocol: O_INTRA
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 1              Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                 State: Active Adv
  OrigTblID: 0x0              OrigVrf: default-vrf
    TableID: 0x2               OrigAs: 0
      NibID: 0x26000002        LastAs: 0
     AttrID: 0xffffffff       Neighbor: 0.0.0.0
      Flags: 0x1008c       OrigNextHop: 192.168.0.100
      Label: NULL          RealNextHop: 192.168.0.100
    BkLabel: NULL            BkNextHop: N/A
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface10
BkTunnel ID: Invalid       BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0              MicroSegID: 0
RealFIRType: Normal          RealThres: 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 10. Then the link over VLAN-interface 10 fails.

Display routes destined for 120.1.1.1/32 on Switch A.

```
<SwitchA> display ip routing-table 120.1.1.1 verbose
Summary count : 1
Destination: 120.1.1.1/32
   Protocol: O_INTRA
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 2              Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                 State: Active Adv
  OrigTblID: 0x0              OrigVrf: default-vrf
    TableID: 0x2               OrigAs: 0
      NibID: 0x26000002        LastAs: 0
     AttrID: 0xffffffff       Neighbor: 0.0.0.0
      Flags: 0x1008c       OrigNextHop: 10.1.1.100
      Label: NULL          RealNextHop: 10.1.1.100
    BkLabel: NULL           B kNextHop: N/A
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface11
BkTunnel ID: Invalid       BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0              MicroSegID: 0
```

```
RealFIRType: Normal          RealThres: 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 11.

Example: Configuring OSPF FRR

Network configuration

As shown in Figure 25, Switch A, Switch B, and Switch C reside in the same OSPF domain. Configure OSPF FRR so that when the link between Switch A and Switch B fails, traffic is immediately switched to Link B.

Figure 25 Network diagram

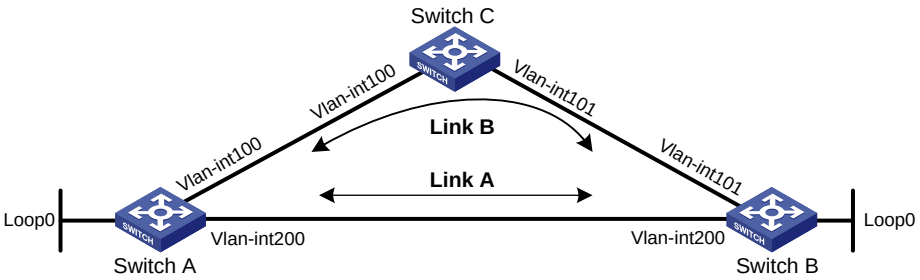


Table 5 Interface and IP address assignment

Device	Interface	IP address
Switch A	Vlan-int100	12.12.12.1/24
Switch A	Vlan-int200	13.13.13.1/24
Switch A	Loop0	1.1.1.1/32
Switch B	Vlan-int101	24.24.24.4/24
Switch B	Vlan-int200	13.13.13.2/24
Switch B	Loop0	4.4.4.4/32
Switch C	Vlan-int100	12.12.12.2/24
Switch C	Vlan-int101	24.24.24.2/24

Procedure

1. Configure IP addresses and subnet masks for interfaces on the switches. (Details not shown.)
2. Configure OSPF on the switches to ensure that Switch A, Switch B, and Switch C can communicate with each other at the network layer. (Details not shown.)
3. Configure OSPF FRR to automatically calculate the backup next hop:

   You can enable OSPF FRR to either calculate a backup next hop by using the LFA algorithm, or specify a backup next hop by using a routing policy.

   - (Method 1.) Enable OSPF FRR to calculate the backup next hop by using the LFA algorithm:

     # Configure Switch A.

     ```
     <SwitchA> system-view
     [SwitchA] ospf 1
     [SwitchA-ospf-1] fast-reroute lfa
     [SwitchA-ospf-1] quit
     ```

     # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf 1
[SwitchB-ospf-1] fast-reroute lfa
[SwitchB-ospf-1] quit
```

- (Method 2.) Enable OSPF FRR to designate a backup next hop by using a routing policy.

  # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ip prefix-list abc index 10 permit 4.4.4.4 32
[SwitchA] route-policy frr permit node 10
[SwitchA-route-policy-frr-10] if-match ip address prefix-list abc
[SwitchA-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
100 backup-nexthop 12.12.12.2
[SwitchA-route-policy-frr-10] quit
[SwitchA] ospf 1
[SwitchA-ospf-1] fast-reroute route-policy frr
[SwitchA-ospf-1] quit
```

  # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ip prefix-list abc index 10 permit 1.1.1.1 32
[SwitchB] route-policy frr permit node 10
[SwitchB-route-policy-frr-10] if-match ip address prefix-list abc
[SwitchB-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
101 backup-nexthop 24.24.24.2
[SwitchB-route-policy-frr-10] quit
[SwitchB] ospf 1
[SwitchB-ospf-1] fast-reroute route-policy frr
[SwitchB-ospf-1] quit
```

Verifying the configuration

Display route 4.4.4.4/32 on Switch A to view the backup next hop information.

```
[SwitchA] display ip routing-table 4.4.4.4 verbose

Summary count : 1

Destination: 4.4.4.4/32
   Protocol: O_INTRA
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 1              Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                 State: Active Adv
  OrigTblID: 0x0              OrigVrf: default-vrf
    TableID: 0x2               OrigAs: 0
      NibID: 0x26000002        LastAs: 0
     AttrID: 0xffffffff       Neighbor: 0.0.0.0
      Flags: 0x1008c        OrigNextHop: 13.13.13.2
      Label: NULL           RealNextHop: 13.13.13.2
    BkLabel: NULL             BkNextHop: 12.12.12.2
    SRLabel: NULL             BkSRLabel: NULL
```

```
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface200
BkTunnel ID: Invalid        BkInterface: Vlan-interface100
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Display route 1.1.1.1/32 on Switch B to view the backup next hop information.

```
[SwitchB] display ip routing-table 1.1.1.1 verbose

Summary count : 1

Destination: 1.1.1.1/32
   Protocol: O_INTRA
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 1              Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                 State: Active Adv
  OrigTblID: 0x0              OrigVrf: default-vrf
    TableID: 0x2               OrigAs: 0
      NibID: 0x26000002        LastAs: 0
     AttrID: 0xffffffff       Neighbor: 0.0.0.0
      Flags: 0x1008c        OrigNextHop: 13.13.13.1
      Label: NULL           RealNextHop: 13.13.13.1
    BkLabel: NULL             BkNextHop: 24.24.24.2
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface200
BkTunnel ID: Invalid        BkInterface: Vlan-interface101
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Troubleshooting OSPF configuration

No OSPF neighbor relationship established

Symptom

No OSPF neighbor relationship can be established.

Analysis

If the physical link and lower layer protocols work correctly, verify OSPF parameters configured on interfaces. Two neighbors must have the same parameters, such as the area ID, network segment, and mask. (A P2P or virtual link can have different network segments and masks.)

Solution

To resolve the problem:

1. Use the **display ospf peer** command to verify OSPF neighbor information.
2. Use the **display ospf interface** command to verify OSPF interface information.
3. Ping the neighbor router's IP address to verify that the connectivity is normal.
4. Verify OSPF timers. The dead interval on an interface must be a minimum of four times the hello interval.
5. On an NBMA network, use the **peer ip-address** command to manually specify the neighbor.
6. A minimum of one interface must have a router priority higher than 0 on an NBMA or a broadcast network.
7. If the problem persists, contact H3C Support.

Incorrect routing information

Symptom

OSPF cannot find routes to other areas.

Analysis

The backbone area must maintain connectivity to all other areas. If a router connects to more than one area, a minimum of one area must be connected to the backbone. The backbone cannot be configured as a stub area.

In a stub area, all routers cannot receive external routes, and all interfaces connected to the stub area must belong to the stub area.

Solution

To resolve the problem:

1. Use the **display ospf peer** command to verify neighbor information.
2. Use the **display ospf interface** command to verify OSPF interface information.
3. Use the **display ospf lsdb** command to verify the LSDB.
4. Use the **display current-configuration configuration ospf** command to verify area configuration. If more than two areas are configured, a minimum of one area is connected to the backbone.
5. In a stub area, all routers attached are configured with the **stub** command. In an NSSA area, all routers attached are configured with the **nssa** command.
6. If a virtual link is configured, use the **display ospf vlink** command to verify the state of the virtual link.
7. If the problem persists, contact H3C Support.

Contents

Configuring IS-IS 1

About IS-IS 1
Terminology 1
IS-IS address 1
NET 2
IS-IS area 3
IS-IS topology 3
Route leaking 4
IS-IS network types 4
DIS and pseudonodes 4
IS-IS PDUs 5
IPv6 IS-IS 7
Protocols and standards 7
IPv4 IS-IS tasks at a glance 8
IPv6 IS-IS tasks at a glance 10
Configuring basic IS-IS 12
Enabling IPv4 IS-IS 12
Enabling IPv6 IS-IS 12
Enabling a lightweight IS-IS process 13
Setting the IS level and circuit level 13
Configuring P2P network type for an interface 14
Configuring IS-IS multi-instance processes 14
Configuring IPv6 IS-IS MTR 15
Configuring IS-IS route control 16
Configuring IS-IS link cost 16
Enabling IS-IS to advertise the maximum link cost to neighbors 19
Specifying a preference for IS-IS 19
Configuring the maximum number of ECMP routes 20
Configuring IS-IS route summarization 20
Advertising a default route 21
Configuring IS-IS route redistribution 22
Filtering routes calculated from received LSPs 23
Filtering redistributed routes 24
Excluding redistributed routes from route calculation 25
Configuring IS-IS route leaking 25
Advertising IS-IS link state information to other protocols 27
Adjusting IS-IS routing convergence speed 27
Specifying the interval for sending IS-IS hello packets 27
Specifying the interval for sending IS-IS CSNP packets 28
Setting the maximum age of LSPs 28
Setting the initial remaining lifetime of IS-IS LSPs in the LSDB 28
Setting the LSP refresh interval and generation interval 29
Setting LSP sending intervals 30
Setting the LSP receiving interval 31
Setting the SPF calculation interval 31
Configuring convergence priorities for specific routes 32
Configuring IS-IS packet-related features 33
Configuring a DIS priority for an interface 33
Configuring the administrative tag value for an interface 33
Specifying the IS-IS hello multiplier 35
Disabling an interface from sending/receiving IS-IS packets 36
Enabling an interface to send small hello packets 36
Setting LSP lengths 36
Enabling LSP flash flooding 37
Enabling LSP fragment extension 37
Limiting LSP flooding 38
Configuring IS-IS route update suppression 39

Configuring advanced IS-IS features ····40
Enabling source address check for hello packets on a P2P interface ····40
Setting the LSDB overload bit ····40
Configuring IS-IS isolation ····42
Configuring IS-IS shutdown ····43
Configuring the ATT bit ····43
Configuring system ID to host name mappings ····44
Configuring IS-IS logging and SNMP notifications ····45
Enabling the logging of neighbor state changes ····45
Setting the maximum number of log entries that can be recorded ····46
Setting the maximum number of advertisement sources that trigger LSP update logging ····46
Setting the maximum number of IS-IS route change log messages that can be recorded ····47
Setting the maximum number of IS-IS neighbor relationship troubleshooting entries ····47
Configuring IS-IS network management ····47
Configuring IS-IS fast convergence ····48
Enabling ISPF ····48
Enabling prefix suppression ····49
Configuring IS-IS PIC ····49
Enhancing IS-IS network security ····50
Configuring neighbor relationship authentication ····50
Configuring area authentication ····51
Configuring routing domain authentication ····52
Configuring IS-IS GR ····52
Configuring IS-IS NSR ····53
Configuring BFD for IS-IS ····54
Configuring IS-IS FRR ····55
About IS-IS FRR ····55
Configuring IS-IS LFA FRR ····55
Setting the priority for FRR backup path selection policies ····59
Configuring IS-IS neighbor flapping suppression ····60
Configuring IS-IS member port flapping suppression ····61
Display and maintenance commands for IS-IS ····63
Displaying and maintaining IPv4 IS-IS ····64
Displaying and maintaining IPv6 IS-IS ····66
IS-IS configuration examples ····68
Example: Configuring basic IS-IS ····68
Example: Configuring DIS election ····72
Example: Configuring IS-IS route redistribution ····76
Example: Configuring IS-IS authentication ····80
Example: Configuring IS-IS GR ····82
Example: Configuring IS-IS NSR ····83
Example: Configuring BFD for IS-IS ····86
Example: Configuring IS-IS FRR ····89
Example: Configuring IS-IS multi-instance processes ····92
IPv6 IS-IS configuration examples ····96
Example: Configuring IPv6 IS-IS basics ····96
Example: Configuring BFD for IPv6 IS-IS ····101
Example: Configuring IPv6 IS-IS FRR ····103

Configuring IS-IS

About IS-IS

IS-IS is an IGP used within an AS. It uses the SPF algorithm for route calculation.

Terminology

- **Intermediate system**—Similar to a router in TCP/IP, IS is the basic unit used in an IS-IS routing domain to generate and propagate routing information. Throughout this chapter, an IS refers to a router.
- **End system**—Similar to a host in TCP/IP, an ES does not run IS-IS. ISO defines the ES-IS protocol for communication between an ES and an IS.
- **Routing domain**—An RD comprises a group of ISs that exchange routing information with each other by using the same routing protocol.
- **Area**—An IS-IS routing domain can be split into multiple areas.
- **Link State Database**—All link states in the network form the LSDB. Each IS has a minimum of one LSDB. An IS uses the SPF algorithm and LSDB to generate IS-IS routes.
- **Link State Protocol Data Unit or Link State Packet**—An IS advertises link state information in an LSP.
- **Network Protocol Data Unit**—An NPDU is a network layer protocol packet in OSI, similar to an IP packet in TCP/IP.
- **Designated IS**—A DIS is elected on a broadcast network.
- **Network service access point**—An NSAP is an OSI network layer address. The NSAP identifies an abstract network service access point and describes the network address format in the OSI reference model.

IS-IS address

As shown in Figure 1, an NSAP address comprises the Initial Domain Part (IDP) and the Domain Specific Part (DSP). The IDP is analogous to the network ID of an IP address, and the DSP is analogous to the subnet and host ID.

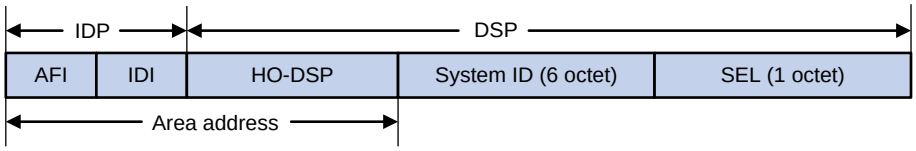
The IDP includes the Authority and Format Identifier (AFI) and the Initial Domain Identifier (IDI).

The DSP includes:

- **High Order Part of DSP (HO-DSP)**—Identifies the area.
- **System ID**—Identifies the host.
- **SEL**—Also known as the N-SEL or the NSAP selector (SEL). It is similar to the protocol identifier in IP and is used to identify the type of service. Different transport layer protocols correspond to different SELs.

The IDP and DSP are variable in length. The length of an NSAP address is in the range of 8 to 20 bytes.

Figure 1 NSAP address format



An IS-IS address contains the following components:

- Area address

  The area address comprises the IDP and the HO-DSP of the DSP, which identify the area and the routing domain. Different routing domains cannot have the same area address.

  Typically, a router only needs one area address, and all nodes in the same area must have the same area address. To support smooth area merging, partitioning, and switching, a router can have a maximum of three area addresses.

- System ID

  A system ID uniquely identifies a host or router. It has a fixed length of 48 bits (6 bytes).

  The system ID of a device can be generated from the router ID. For example, suppose a router uses the IP address 168.10.1.1 of Loopback 0 as the router ID. The system ID can be obtained in the following steps:

  a. Extend each decimal number of the IP address to three digits by adding 0s from the left, such as 168.010.001.001.

  b. Divide the extended IP address into three sections that each has four digits to get the system ID 1680.1000.1001.

  If you use other methods to define a system ID, make sure that it can uniquely identify the host or router.

- SEL

  An SEL is used to identify the type of service. Different transport layer protocols correspond to different SELs. It has a fixed length of 8 bits. All SELs in IP are 00.

NET

A network entity title (NET) identifies the network layer information of an IS. It does not include transport layer information. A NET is a special NSAP address with the SEL being 0. The length of a NET is in the range of 8 to 20 bytes, same as a NSAP address.

A NET includes the following parts:

- **Area ID**—Has a length of 1 to 13 bytes.
- **System ID**—A system ID uniquely identifies a host or router in the area and has a fixed length of 6 bytes.
- **SEL**—Has a value of 0 and a fixed length of 1 byte.

For example, for a NET ab.cdef.1234.5678.9abc.00, the area ID is ab.cdef, the system ID is 1234.5678.9abc, and the SEL is 00.

Typically, a router only needs one NET, but it can have a maximum of three NETs for smooth area merging and partitioning. When you configure multiple NETs, make sure the system IDs are the same.

IS-IS area

IS-IS has a 2-level hierarchy to support large-scale networks. A large-scale routing domain is divided into multiple areas. Typically, a Level-1 router is deployed within an area. A Level-2 router is deployed between areas. A Level-1-2 router is deployed between Level-1 and Level-2 routers.

Level-1 router

A Level-1 router establishes neighbor relationships with Level-1 and Level-1-2 routers in the same area. It maintains an LSDB comprising intra-area routing information. A Level-1 router forwards packets destined for external areas to the nearest Level-1-2 router. Level-1 routers in different areas cannot establish neighbor relationships.

Level-2 router

A Level-2 router establishes neighbor relationships with Level-2 and Level-1-2 routers in the same area or in different areas. It maintains a Level-2 LSDB containing inter-area routing information. All the Level-2 and Level-1-2 routers must be contiguous to form the backbone of the IS-IS routing domain. Level-2 routers can establish neighbor relationships even if they are in different areas.

Level-1-2 router

A router with both Level-1 and Level-2 router functions is a Level-1-2 router. It can establish Level-1 neighbor relationships with Level-1 and Level-1-2 routers in the same area. It can establish Level-2 neighbor relationships with Level-2 and Level-1-2 routers in different areas. A Level-1 router can reach other areas only through a Level-1-2 router. The Level-1-2 router maintains two LSDBs, a Level-1 LSDB for intra-area routing and a Level-2 LSDB for inter-area routing.

IS-IS topology

Figure 2 shows one IS-IS network topology. Area 1 is the backbone that comprises a set of Level-2 routers. The other four areas are non-backbone areas connected to the backbone through Level-1-2 routers.

Figure 2 IS-IS topology 1

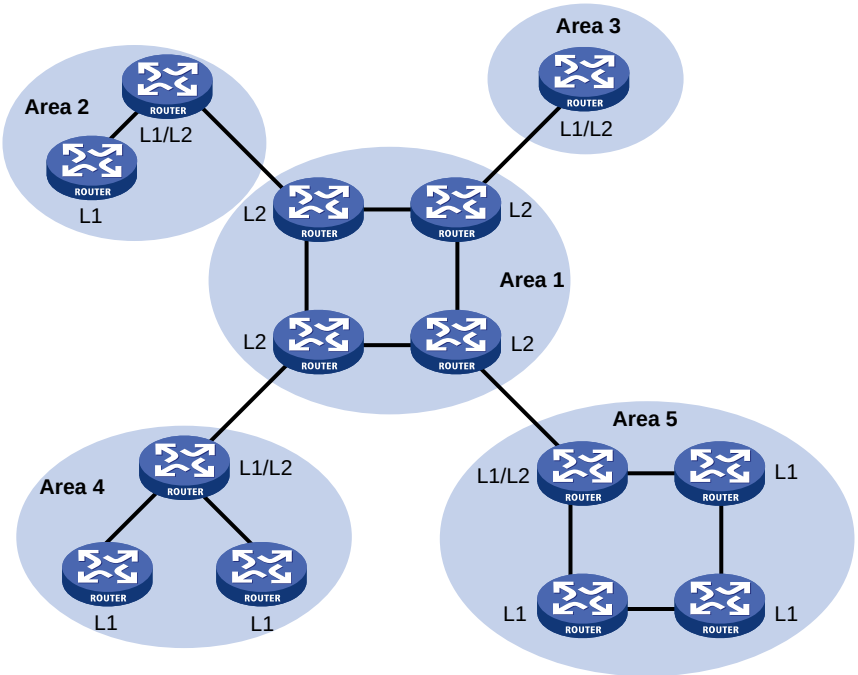
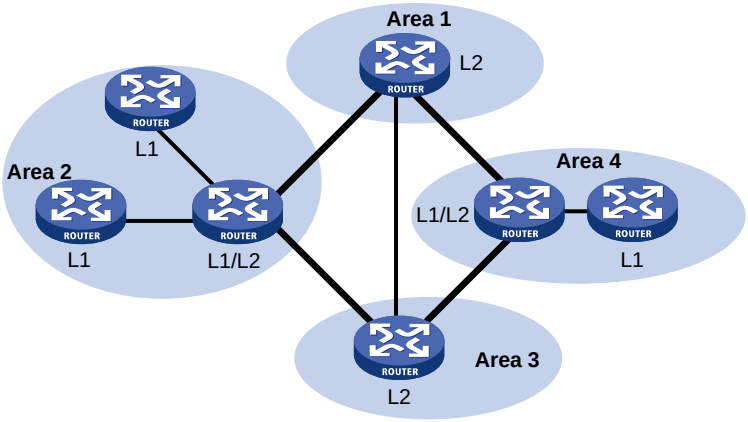


Figure 3 shows another IS-IS topology. No area is defined as the backbone in this topology. The backbone comprises all contiguous Level-2 and Level-1-2 routers in different areas. The IS-IS backbone does not need to be a specific area.

Figure 3 IS-IS topology 2



Both the Level-1 and Level-2 routers use the SPF algorithm to generate the shortest path tree.

Route leaking

Level-2 and Level-1-2 routers form a Level-2 area. An IS-IS routing domain comprises only one Level-2 area and multiple Level-1 areas. A Level-1 area must connect to the Level-1-2 area rather than another Level-1 area.

Level-1-2 routers send the routing information of Level-1 areas to the Level-2 area. Level-2 routers know the routing information of the entire IS-IS routing domain. By default, a Level-2 router does not advertise the routing information of other areas to a Level-1 area. A Level-1 router simply sends packets destined for other areas to the nearest Level-1-2 router. The path passing through the Level-1-2 router might not be the best. To solve this problem, IS-IS provides the route leaking feature.

Route leaking enables a Level-1-2 router to advertise the routes of other areas to the connected Level-1 area so that the Level-1 routers can select the optimal routes.

IS-IS network types

IS-IS supports broadcast networks (for example, Ethernet and Token Ring) and point-to-point networks (for example, PPP and HDLC).

IS-IS cannot run on P2MP links.

DIS and pseudonodes

IS-IS routers on a broadcast network must elect a DIS.

The Level-1 and Level-2 DISs are elected separately. You can assign different priorities to a router for different level DIS elections. The higher the router priority, the more likely the router becomes the DIS. If multiple routers with the same highest DIS priority exist, the one with the highest Subnetwork Point of Attachment (SNPA) address will be elected. On a broadcast network, the SNPA address is the MAC address. A router can be the DIS for different levels.

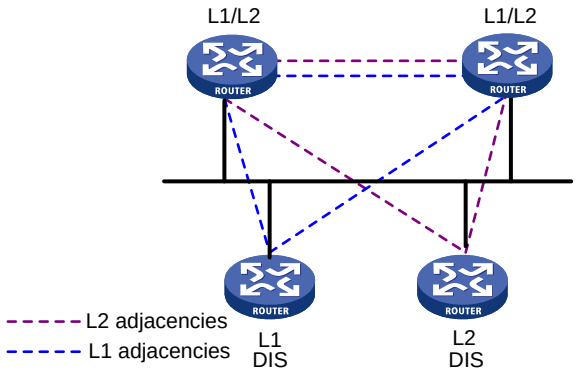
IS-IS DIS election differs from OSPF DIS election in the following ways:

- A router with priority 0 can also participate in the DIS election.

- When a router with a higher priority is added to the network, an LSP flooding process is performed to elect the router as the new DIS.

As shown in Figure 4, the same level routers on a network, including non-DIS routers, establish adjacency with each other.

Figure 4 DIS in the IS-IS broadcast network



The DIS creates and updates pseudonodes, and generates LSPs for the pseudonodes, to describe all routers on the network.

A pseudonode represents a virtual node on the broadcast network. It is not a real router. In IS-IS, it is identified by the system ID of the DIS and a 1-byte Circuit ID (a non-zero value).

Using pseudonodes simplifies network topology and can reduce the amount of resources consumed by SPF.

NOTE:

On an IS-IS broadcast network, all routers establish adjacency relationships, but they synchronize their LSDBs through the DIS.

On a broadcast network, the device can establish up to 255 neighbor relationships, because the Circuit ID is only one byte long. If the number of broadcast links that require neighbor relationship establishment exceeds 255, the device does not establish IS-IS neighbor relationships on the excess broadcast links.

IS-IS PDUs

PDU

IS-IS PDUs are encapsulated into link layer frames. An IS-IS PDU has two parts, the headers and the variable length fields. The headers comprise the PDU common header and the PDU specific header. All PDUs have the same PDU common header. The specific headers vary by PDU type.

Figure 5 PDU format

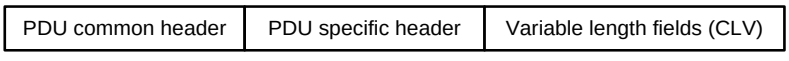


Table 1 PDU types

Type	PDU Type	Acronym
15	Level-1 LAN IS-IS hello PDU	L1 LAN IIH
16	Level-2 LAN IS-IS hello PDU	L2 LAN IIH
17	Point-to-Point IS-IS hello PDU	P2P IIH

Type	PDU Type	Acronym
18	Level-1 Link State PDU	L1 LSP
20	Level-2 Link State PDU	L2 LSP
24	Level-1 Complete Sequence Numbers PDU	L1 CSNP
25	Level-2 Complete Sequence Numbers PDU	L2 CSNP
26	Level-1 Partial Sequence Numbers PDU	L1 PSNP
27	Level-2 Partial Sequence Numbers PDU	L2 PSNP

Hello PDU

IS-to-IS hello (IIH) PDUs are used by routers to establish and maintain neighbor relationships. On broadcast networks, Level-1 routers use Level-1 LAN IIHs, and Level-2 routers use Level-2 LAN IIHs. The P2P IIHs are used on point-to-point networks.

LSP

The LSPs carry link state information. LSPs include Level-1 LSPs and Level-2 LSPs. The Level-2 LSPs are sent by the Level-2 routers, and the Level-1 LSPs are sent by the Level-1 routers. The Level-1-2 router can send both types of LSPs.

SNP

A sequence number PDU (SNP) describes the complete or partial LSPs for LSDB synchronization.

SNPs include CSNP and PSNP, which are further divided into Level-1 CSNP, Level-2 CSNP, Level-1 PSNP, and Level-2 PSNP.

A CSNP describes the summary of all LSPs for LSDB synchronization between neighboring routers. On broadcast networks, CSNPs are sent by the DIS periodically (every 10 seconds by default). On point-to-point networks, CSNPs are sent only during the first adjacency establishment.

A PSNP only contains the sequence numbers of one or multiple latest received LSPs. It can acknowledge multiple LSPs at one time. When LSDBs are not synchronized, a PSNP is used to request missing LSPs from a neighbor.

CLV

The variable fields of PDU comprise multiple Code-Length-Value (CLV) triplets.

Figure 6 CLV format

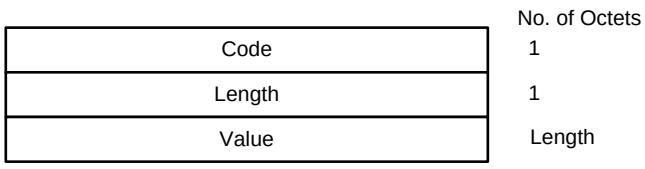


Table 2 shows that different PDUs contain different CLVs. Codes 1 through 10 are defined in ISO 10589 (code 3 and 5 are not shown in the table). Codes 128 through 132 are defined in RFC 1195. Codes 222 through 237 are defined in RFC 5120.

Table 2 CLV codes and PDU types

CLV Code	Name	PDU Type
1	Area Addresses	IIH, LSP
2	IS Neighbors (LSP)	LSP
4	Partition Designated Level 2 IS	L2 LSP

CLV Code	Name	PDU Type
6	IS Neighbors (MAC Address)	LAN IIH
7	IS Neighbors (SNPA Address)	LAN IIH
8	Padding	IIH
9	LSP Entries	SNP
10	Authentication Information	IIH, LSP, SNP
128	IP Internal Reachability Information	LSP
129	Protocols Supported	IIH, LSP
130	IP External Reachability Information	L2 LSP
131	Inter-Domain Routing Protocol Information	L2 LSP
132	IP Interface Address	IIH, LSP
222	MT-ISN	LSP
229	M-Topologies	IIH, LSP
235	MT IP. Reach	LSP
237	MT IPv6 IP. Reach	LSP

IPv6 IS-IS

IS-IS supports multiple network protocols, including IPv6. To support IPv6, the IETF added two type-length-values (TLVs) and a new network layer protocol identifier (NLPID).

The TLVs are as follows:

- **IPv6 Reachability**—Contains routing prefix and metric information to describe network reachability and has a type value of 236 (0xEC).
- **IPv6 Interface Address**—Same as the "IP Interface Address" TLV in IPv4 ISIS, except that the 32-bit IPv4 address is translated to the 128-bit IPv6 address.

The new NLPID is an 8-bit field that identifies which network layer protocol is supported. For IPv6, the NLPID is 142 (0x8E).

Protocols and standards

- ISO 8348, *Ad2 Network Services Access Points*
- ISO 9542, *ES-IS Routing Protocol*
- ISO 10589, *ISO IS-IS Routing Protocol*
- RFC 1195, *Use of OSI IS-IS for Routing in TCP/IP and Dual Environments*
- RFC 3277, *IS-IS Transient Blackhole Avoidance*
- RFC 3358, *Optional Checksums in ISIS*
- RFC 3359, *Reserved Type, Length and Value (TLV) Codepoints in Intermediate System to Intermediate System*
- RFC 3563, *Cooperative Agreement Between the ISOC/IETF and ISO/IEC Joint Technical Committee 1/Sub Committee 6 (JTC1/SC6) on IS-IS Routing Protocol Development*
- RFC 3719, *Recommendations for Interoperable Networks using Intermediate System to Intermediate System (IS-IS)*

- RFC 3787, *Recommendations for Interoperable IP Networks using Intermediate System to Intermediate System (IS-IS)*
- RFC 4444, *Management Information Base for Intermediate System to Intermediate System (IS-IS)*
- RFC 5029, *Definition of an IS-IS Link Attribute Sub-TLV*
- RFC 5089, *IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery*
- RFC 5120, *Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)*
- RFC 5130, *A Policy Control Mechanism in IS-IS Using Administrative Tags*
- RFC 5301, *Dynamic Hostname Exchange Mechanism for IS-IS*
- RFC 5302, *Domain-Wide Prefix Distribution with Two-Level IS-IS*
- RFC 5303, *Three-Way Handshake for IS-IS Point-to-Point Adjacencies*
- RFC 5304, *IS-IS Cryptographic Authentication*
- RFC 5306, *Restart Signaling for IS-IS*
- RFC 5308, *Routing IPv6 with IS-IS*
- RFC 5310, *IS-IS Generic Cryptographic Authentication*
- RFC 5311, *Simplified Extension of Link State PDU (LSP) Space for IS-IS*
- RFC 6165, *Extensions to IS-IS for Layer-2 Systems*
- RFC 6213, *IS-IS BFD-Enabled TLV*
- RFC 6232, *Purge Originator Identification TLV for IS-IS*
- RFC 6233, *IS-IS Registry Extension for Purges*
- RFC 6329, *IS-IS Extensions Supporting IEEE 802.1aq Shortest Path Bridging*
- RFC 6571, *Loop-Free Alternate (LFA) Applicability in Service Provider (SP) Networks*
- RFC 6823, *Advertising Generic Information in IS-IS*
- RFC 7142, *OSI IS-IS Intra-domain Routing Protocol*
- RFC 7356, *IS-IS Flooding Scope Link State PDUs (LSPs)*
- RFC 7370, *Updates to the IS-IS TLV Codepoints Registry*
- RFC 7602, *IS-IS Extended Sequence Number TLV*
- RFC 7645, *The Keying and Authentication for Routing Protocol (KARP) IS-IS Security Analysis*
- RFC 7775, *IS-IS Route Preference for Extended IP and IPv6 Reachability*
- RFC 7794, *IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability*
- RFC 7813, *IS-IS Path Control and Reservation*
- RFC 7917, *Advertising Node Administrative Tags in IS-IS*
- RFC 7981, *IS-IS Extensions for Advertising Router Information*
- RFC 7987, *IS-IS Minimum Remaining Lifetime*

IPv4 IS-IS tasks at a glance

To configure IPv4 IS-IS, perform the following tasks:

1. Configuring basic IS-IS
   a. Enabling IPv4 IS-IS
   b. (Optional.) **Enabling a** lightweight IS-IS process
   c. (Optional.) Setting the IS level and circuit level

d. (Optional.) Configuring P2P network type for an interface

2. (Optional.) Configuring IS-IS multi-instance processes
3. (Optional.) Configuring IS-IS route control
   - Configuring IS-IS link cost
   - Specifying a preference for IS-IS
   - Configuring the maximum number of ECMP routes
   - Configuring IS-IS route summarization
   - Advertising a default route
   - Configuring IS-IS route redistribution
   - Filtering routes calculated from received LSPs
   - Filtering redistributed routes
   - Excluding redistributed routes from route calculation
   - Configuring IS-IS route leaking
   - Advertising IS-IS link state information to other protocols
4. (Optional.) Adjusting IS-IS routing convergence speed
   - Specifying the interval for sending IS-IS hello packets
   - Specifying the interval for sending IS-IS CSNP packets
   - Setting the maximum age of LSPs
   - Setting the initial remaining lifetime of IS-IS LSPs in the LSDB
   - Setting the LSP refresh interval and generation interval
   - Setting LSP sending intervals
   - Setting the LSP receiving interval
   - Setting the SPF calculation interval
   - Configuring convergence priorities for specific routes
5. (Optional.) Configuring IS-IS packet-related features
   - Configuring a DIS priority for an interface
   - Configuring the administrative tag value for an interface
   - Specifying the IS-IS hello multiplier
   - Disabling an interface from sending/receiving IS-IS packets
   - Enabling an interface to send small hello packets
   - Setting LSP lengths
   - Enabling LSP flash flooding
   - Enabling LSP fragment extension
6. (Optional.) Limiting LSP flooding
7. (Optional.) Configuring IS-IS route update suppression
8. (Optional.) Configuring advanced IS-IS features
   - Enabling source address check for hello packets on a P2P interface
   - Setting the LSDB overload bit
   - Configuring IS-IS isolation
   - Configuring IS-IS shutdown
   - Configuring the ATT bit
   - Configuring system ID to host name mappings
9. (Optional.) Configuring IS-IS logging and SNMP notifications
   - Enabling the logging of neighbor state changes

- Setting the maximum number of log entries that can be recorded
- Setting the maximum number of advertisement sources that trigger LSP update logging
- Setting the maximum number of IS-IS route change log messages that can be recorded
- Setting the maximum number of IS-IS neighbor relationship troubleshooting entries
- Configuring IS-IS network management

10. (Optional.) Configuring IS-IS fast convergence
    - Enabling ISPF
    - Enabling prefix suppression
    - Configuring IS-IS PIC
11. (Optional.) Enhancing IS-IS network security
    - Configuring neighbor relationship authentication
    - Configuring area authentication
    - Configuring routing domain authentication
12. (Optional.) Enhancing IS-IS network reliability:
    - Configuring IS-IS GR
    - Configuring IS-IS NSR
    - Configuring BFD for IS-IS
    - Configuring IS-IS FRR
    - Configuring IS-IS neighbor flapping suppression
    - Configuring IS-IS member port flapping suppression
    - Setting the priority for FRR backup path selection policies

IPv6 IS-IS tasks at a glance

To configure IPv6 IS-IS, perform the following tasks:

1. Configuring basic IS-IS
   a. Enabling IPv6 IS-IS
   b. (Optional.) **Enabling a** lightweight IS-IS process
   c. (Optional.) Setting the IS level and circuit level
   d. (Optional.) Configuring P2P network type for an interface
2. (Optional.) Configuring IS-IS multi-instance processes
3. (Optional.) Configuring IPv6 IS-IS MTR
4. (Optional.) Configuring IS-IS route control
    - Configuring IS-IS link cost
    - Specifying a preference for IS-IS
    - Configuring the maximum number of ECMP routes
    - Configuring IS-IS route summarization
    - Advertising a default route
    - Configuring IS-IS route redistribution
    - Filtering routes calculated from received LSPs
    - Filtering redistributed routes
    - Excluding redistributed routes from route calculation
    - Configuring IS-IS route leaking
    - Advertising IS-IS link state information to other protocols

5. (Optional.) Adjusting IS-IS routing convergence speed
    - Specifying the interval for sending IS-IS hello packets
    - Specifying the interval for sending IS-IS CSNP packets
    - Setting the maximum age of LSPs
    - Setting the initial remaining lifetime of IS-IS LSPs in the LSDB
    - Setting the LSP refresh interval and generation interval
    - Setting LSP sending intervals
    - Setting the LSP receiving interval
    - Setting the SPF calculation interval
    - Configuring convergence priorities for specific routes
6. (Optional.) Configuring IS-IS packet-related features
7. Configuring a DIS priority for an interface
    - Configuring the administrative tag value for an interface
    - Specifying the IS-IS hello multiplier
    - Disabling an interface from sending/receiving IS-IS packets
    - Enabling an interface to send small hello packets
    - Setting LSP lengths
    - Enabling LSP flash flooding
    - Enabling LSP fragment extension
8. (Optional.) Limiting LSP flooding
9. (Optional.) Configuring IS-IS route update suppression
10. (Optional.) Configuring advanced IS-IS features
    - Enabling source address check for hello packets on a P2P interface
    - Setting the LSDB overload bit
    - Configuring IS-IS isolation
    - Configuring IS-IS shutdown
    - Configuring the ATT bit
    - Configuring system ID to host name mappings
11. (Optional.) Configuring IS-IS logging and SNMP notifications
    - Enabling the logging of neighbor state changes
    - Setting the maximum number of log entries that can be recorded
    - Setting the maximum number of advertisement sources that trigger LSP update logging
    - Setting the maximum number of IS-IS route change log messages that can be recorded
    - Setting the maximum number of IS-IS neighbor relationship troubleshooting entries
    - Configuring IS-IS network management
12. (Optional.) Configuring IS-IS fast convergence
    - Enabling ISPF
    - Enabling prefix suppression
    - Configuring IS-IS PIC
13. (Optional.) Enhancing IS-IS network security
    - Configuring neighbor relationship authentication
    - Configuring area authentication
    - Configuring routing domain authentication
14. (Optional.) Enhancing IS-IS network reliability:

- Configuring IS-IS GR
- Configuring IS-IS NSR
- Configuring BFD for IS-IS
- Configuring IS-IS FRR
- Configuring IS-IS neighbor flapping suppression
- Configuring IS-IS member port flapping suppression

Configuring basic IS-IS

Enabling IPv4 IS-IS

1. Enter system view.
   system-view
2. Enable IS-IS and enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   By default, IS-IS is disabled.
3. Assign a NET.
   network-entity *net*
   By default, NET is not assigned.

CAUTION:
When you execute the **network-entity** command together with the **cost-style** and **is-level** commands for the same IS-IS process, execute the **network-entity** command at last. Incorrect configuration order might cause data loss because the IS-IS process will restart.

4. Return to system view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Enable IS-IS on the interface.
   isis enable [*process-id*]
   By default, IS-IS is disabled.

Enabling IPv6 IS-IS

1. Enter system view.
   system-view
2. Enable an IS-IS process and enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   By default, no IS-IS process is enabled.
3. Configure the NET for the IS-IS process.
   network-entity *net*
   By default, the NET is not configured.

CAUTION:

When you execute the **network-entity** command together with the **cost-style** and **is-level** commands for the same IS-IS process, execute the **network-entity** command at last. Incorrect configuration order might cause data loss because the IS-IS process will restart.

4. Create the IPv6 address family and enter its view.
   address-family ipv6 [**unicast**]
5. Return to IS-IS view.
   quit
6. Return to system view.
   quit
7. Enter interface view.
   interface *interface-type interface-number*
8. Enable IPv6 for IS-IS on the interface.
   isis ipv6 enable [*process-id*]
   By default, IPv6 is disabled for IS-IS on an interface.

Enabling a lightweight IS-IS process

About this task

Enabling multiple IS-IS processes on a device might cause memory insufficiency. To resolve this issue, perform this task to enable lightweight IS-IS processes.

Restrictions and guidelines

Although lightweight processes use less memory resources than traditional processes, device performance might degrade when a large number of neighbors and routes exist. As a best practice, plan the number of lightweight IS-IS processes to be enabled on the basis of the device resource condition.

Procedure

1. Enter system view.
   system-view
2. Enable a lightweight IS-IS process and enter its view.
   isis [*process-id*] **lite** [**vpn-instance** *vpn-instance-name*]

Setting the IS level and circuit level

About this task

Follow these guidelines when you configure the IS level for routers in only one area:

- Set the IS level of all routers to Level-1 or Level-2 rather than different levels because the routers do not need to maintain two identical LSDBs.
- Set the IS level to Level-2 on all routers in an IP network for good scalability.

For an interface of a Level-1 or Level-2 router, the circuit level can only be Level-1 or Level-2. For an interface of a Level-1-2 router, the default circuit level is Level-1-2. If the router only needs to form Level-1 or Level-2 neighbor relationships, set the circuit level for its interfaces to Level-1 or Level-2. This will limit neighbor relationship establishment.

Procedure

1. Enter system view.

system-view

2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Specify the IS level.
   is-level { **level-1** | **level-1-2** | **level-2** }
   By default, the IS level is Level-1-2.
4. Return to system view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Specify the circuit level.
   isis circuit-level [**level-1** | **level-1-2** | **level-2**]
   By default, an interface can establish either the Level-1 or Level-2 adjacency.

Configuring P2P network type for an interface

About this task

Interfaces with different network types operate differently. For example, broadcast interfaces on a network must elect the DIS and flood CSNP packets to synchronize the LSDBs. However, P2P interfaces on a network do not need to elect the DIS, and have a different LSDB synchronization mechanism.

If only two routers exist on a broadcast network, set the network type of attached interfaces to P2P. This avoids DIS election and CSNP flooding, saving network bandwidth and speeding up network convergence.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure P2P network type for an interface.
   isis [**process-id** *process-id*] **circuit-type p2p**
   By default, the network type of a VLAN interface is broadcast.
   Perform this task only for a broadcast network that has up to two attached routers.

Configuring IS-IS multi-instance processes

About this task

NOTE:

IS-IS processes not enabled with the multi-instance process feature are called traditional IS-IS processes. IS-IS processes enabled with the multi-instance process feature are called IS-IS multi-instance processes.

By default, an interface supports only one IS-IS process. To configure multiple IS-IS processes on a device, you must add more interfaces to the device and configure the interfaces manually. To simplify configuration, use the IS-IS multi-instance process feature to configure multiple IS-IS multi-instance processes as well as a traditional IS-IS process on an interface.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable the IS-IS multi-instance process and specify an instance ID for the process.
   multi-instance enable iid *iid-value*
   By default, IS-IS multi-instance process is disabled.
4. Configure IS-IS multi-instance processes on an interface:
   a. Return to system view.
      quit
   b. Enter interface view.
      interface *interface-type interface-number*
   c. Enable an IS-IS multi-instance process on the interface:
      IPv4:
      isis enable [*process-id*]
      By default, IPv4 IS-IS is disabled on an interface, and the interface is not enabled with any IS-IS processes.
      IPv6:
      isis ipv6 enable [*process-id*]
      By default, IPv6 IS-IS is disabled on an interface, and the interface is not enabled with any IS-IS processes.

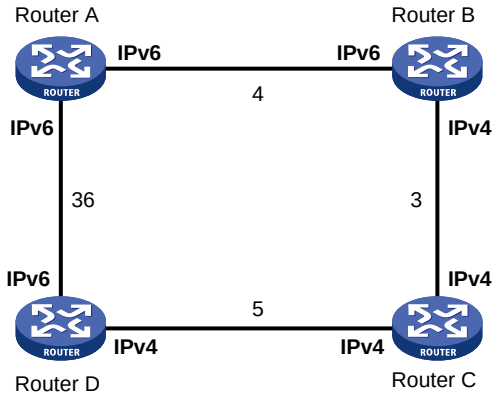
Configuring IPv6 IS-IS MTR

About this task

On a network, IPv4 and IPv6 topologies must be consistent so that both IPv6 IS-IS and IPv4 IS-IS can use the SPF algorithm to perform route calculation. If they are different, routers supporting both IPv4 and IPv6 might send IPv6 packets to routers that do not support IPv6, resulting in packet loss.

To resolve this issue, configure IPv6 IS-IS MTR to perform route calculation separately in IPv4 and IPv6 topologies.

Figure 7 Network diagram



As shown in Figure 7, the numbers refer to the link costs. Router A, Router B, and Router D support both IPv4 and IPv6. Router C supports only IPv4 and cannot forward IPv6 packets.

Enable IPv6 IS-IS MTR on Router A, Router B, Router C, and Router D to make them perform route calculation separately in IPv4 and IPv6 topologies. With this configuration, Router A does not forward IPv6 packets destined to Router D through Router B, avoiding packet loss.

Restrictions and guidelines

As a best practice to avoid route calculation failures, configure this feature when both IPv4 and IPv6 topologies exist in the network.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Specify an IS-IS cost style.
   cost-style { **compatible** | **wide** | **wide-compatible** }
   By default, IS-IS only transmits and receives packets using the **narrow** cost style.
4. Enter IPv6 address family view.
   address-family ipv6 [**unicast**]
5. Enable IPv6 IS-IS MTR.
   multi-topology [**compatible**]
   By default, IPv6 IS-IS MTR is disabled.

Configuring IS-IS route control

Configuring IS-IS link cost

About this task

The IS-IS cost of an interface is determined in the following order:

1. IS-IS cost specified in interface view.
2. IS-IS cost specified in system view.
   The cost is applied to the interfaces associated with the IS-IS process.
3. Automatically calculated cost.
   If the cost style is **wide** or **wide-compatible**, IS-IS automatically calculates the cost using the formula: Interface cost = (Bandwidth reference value / Expected interface bandwidth) × 10, in the range of 1 to 16777214. For other cost styles, Table 3 applies.
   Configure the expected bandwidth of an interface with the **bandwidth** command.

Table 3 Automatic cost calculation scheme for cost styles other than wide and wide-compatible

Interface bandwidth	Interface cost
≤ 10 Mbps	60
≤ 100 Mbps	50
≤ 155 Mbps	40
≤ 622 Mbps	30
≤ 2500 Mbps	20

Interface bandwidth	Interface cost
> 2500 Mbps	10

4. If none of the above costs is used, a default cost of 10 applies.

Configuring an IPv4 IS-IS cost for an interface

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. (Optional.) Specify an IS-IS cost style.
   cost-style { **narrow** | **wide** | **wide-compatible** | { **compatible** | **narrow-compatible** } [**relax-spf-limit**] }
   By default, the IS-IS cost type is **narrow**.
4. Return to system view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Specify a cost for the IS-IS interface.
   isis [**process-id** *process-id*] **cost** *cost-value* [**level-1** | **level-2**]
   By default, no cost for the interface is specified.

Configuring a global IPv4 IS-IS cost

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Specify a global IS-IS cost.
   circuit-cost *cost-value* [**level-1** | **level-2**]
   By default, no global cost is specified.

Enabling automatic IPv4 IS-IS cost calculation

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable automatic IS-IS cost calculation.
   auto-cost enable
   By default, automatic IS-IS cost calculation is disabled.
4. (Optional.) Configure a bandwidth reference value for automatic IS-IS cost calculation.
   bandwidth-reference *value*
   The default setting is 100 Mbps.

Configuring an IPv6 IS-IS cost for an interface

1. Enter system view.
   system-view

2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. (Optional.) Specify an IS-IS cost style.
   cost-style { **narrow** | **wide** | **wide-compatible** | { **compatible** | **narrow-compatible** } [**relax-spf-limit**] }
   By default, the IS-IS cost type is **narrow**.
4. Enter IPv6 address family view.
   address-family ipv6 [**unicast**]
5. Return to IS-IS view.
   quit
6. Return to system view.
   quit
7. Enter interface view.
   interface *interface-type interface-number*
8. Enable IPv6 for IS-IS on the interface.
   isis ipv6 enable [*process-id*]
   By default, IPv6 is disabled for IS-IS on an interface.
9. Specify an IPv6 cost for the IS-IS interface.
   isis [**process-id** *process-id*] **ipv6 cost** *cost-value* [**level-1** | **level-2**]
   By default, no IPv6 cost is specified for the interface.

Configuring a global IPv6 IS-IS cost

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Specify a global IPv6 IS-IS cost.
   circuit-cost *cost-value* [**level-1** | **level-2**]
   By default, no global IPv6 cost is specified.

Enabling automatic IPv6 IS-IS cost calculation

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Specify an IS-IS cost style.
   cost-style { **wide** | **wide-compatible** }
   By default, the IS-IS cost style is **narrow**.
4. Enter IPv6 address family view.
   address-family ipv6 [**unicast**]
5. Enable automatic IPv6 IS-IS cost calculation.
   auto-cost enable

By default, automatic IPv6 IS-IS cost calculation is disabled.

6. (Optional.) Configure a bandwidth reference value for automatic IPv6 IS-IS cost calculation.
   bandwidth-reference *value*

   By default, the bandwidth reference value is 100 Mbps.

Enabling IS-IS to advertise the maximum link cost to neighbors

About this task

On an IS-IS network, when a link recovers from failures or the state of an interface changes, IS-IS will re-establish neighbor relationships and perform route convergence. During the route convergence process, routing loops and traffic loss might occur because the convergence speeds of the nodes are different. To address this issue, enable IS-IS to advertise the maximum link cost to neighbors within the specified period, so the traffic forwarding path remains unchanged. After the specified period, IS-IS advertises the original link cost to neighbors and performs optimal route selection again.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IS-IS to advertise the maximum link cost to neighbors within the specified period.
   isis peer hold-max-cost duration *time*

   By default, IS-IS advertises the original link cost to neighbors during a route convergence.

Specifying a preference for IS-IS

About this task

If multiple routing protocols find routes to the same destination, the route found by the routing protocol that has the highest preference is selected as the optimal route.

Perform this task to assign a preference to IS-IS directly or by using a routing policy. For more information about the routing policy, see "Configuring routing policies."

Configuring a preference for IPv4 IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS IPv4 unicast address family view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   address-family ipv4 [**unicast**]
3. Configure a preference for IPv4 IS-IS.
   preference { *preference* | **route-policy** *route-policy-name* } *

   The default setting is 15.

Configuring a preference for IPv6 IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS view.

isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enter IS-IS IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Configure a preference for IPv6 IS-IS.
   preference { **route-policy** *route-policy-name* | *preference* } *
   The default setting is 15.

Configuring the maximum number of ECMP routes

About this task

Perform this task to implement load sharing over ECMP routes.

Configuring the maximum number of ECMP routes for IPv4 IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS IPv4 unicast address family view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   address-family ipv4 [**unicast**]
3. Specify the maximum number of ECMP routes.
   maximum load-balancing *number*
   By default, the maximum number of ECMP routes supported by IPv4 IS-IS equals the maximum number of ECMP routes supported by the system.

Configuring the maximum number of ECMP routes for IPv6 IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IS-IS IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Specify the maximum number of ECMP routes.
   maximum load-balancing *number*
   By default, the maximum number of ECMP routes supported by IPv6 IS-IS equals the maximum number of ECMP routes supported by the system.

Configuring IS-IS route summarization

About this task

Perform this task to summarize specific routes, including IS-IS routes and redistributed routes, into a single route. Route summarization can reduce the routing table size and the LSDB scale.

Route summarization applies only to locally generated LSPs.

Configuring IPv4 IS-IS route summarization

1. Enter system view.
   system-view
2. Enter IS-IS IPv4 unicast address family view.

isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

address-family ipv4 [**unicast**]

3. Configure IPv4 IS-IS route summarization.

   summary *ip-address* { *mask-length* | *mask* } [**avoid-feedback** | **generate_null0_route** | [**level-1** | **level-1-2** | **level-2**] | **tag** *tag*] *

   By default, IPv4 IS-IS route summarization is not configured.

   The cost of the summary route is the lowest one among the costs of the more-specific routes.

Configuring IPv6 IS-IS route summarization

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enter IS-IS IPv6 address family view.

   address-family ipv6 [**unicast**]

4. Configure IPv6 IS-IS route summarization.

   summary *ipv6-prefix prefix-length* [**avoid-feedback** | **generate_null0_route** | [**level-1** | **level-1-2** | **level-2**] | **tag** *tag*] *

   By default, IPv6 IS-IS route summarization is not configured.

Advertising a default route

About this task

IS-IS cannot redistribute a default route to its neighbors. This task enables IS-IS to advertise a default route of 0.0.0.0/0 in an LSP to the same-level neighbors. Upon receiving the default route, the neighbors add it into their routing table.

When multiple devices exist in an IS-IS network, configuring default route advertisement is more flexible than configuring default static routes. For example, if an IS-IS routing domain has multiple edge devices, you can configure a routing policy on an edge device to control the advertisement of default routes and to prevent routing blackholes.

Advertising an IPv4 IS-IS default route

1. Enter system view.

   system-view

2. Enter IS-IS IPv4 unicast address family view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

   address-family ipv4 [**unicast**]

3. Advertise a Level-1 or Level-2 default route.

   default-route-advertise [[**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name*] *

   By default, IPv4 IS-IS does not advertise a Level-1 or Level-2 default route.

Advertising an IPv6 IS-IS default route

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enter IS-IS IPv6 address family view.

   address-family ipv6 [**unicast**]

4. Advertise a Level-1 or Level-2 default route.

   default-route-advertise [**avoid-learning** | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   By default, IPv6 IS-IS does not advertise a Level-1 or Level-2 default route.

Configuring IS-IS route redistribution

About this task

After you enable default route advertisement for an edge device in an IS-IS routing domain, the device acts as a gateway to external routing domains for other devices in the routing domain. The service burden on the edge device will be heavy when it receives a large number of packets from other devices. Besides, selecting the optimal route to an external routing domain might be an issue when multiple edge devices exist in the IS-IS routing domain. To resolve these issues, you can perform this task to redistribute routes from other routing protocols into IS-IS.

You can configure the cost value for redistributed IGP routes and BGP routes. With an ACL, IP prefix list, or routing policy, IS-IS can filter redistributed routes. IS-IS adds only matching routes into its routing table, and then advertises them in LSPs.

To avoid the potential impact on the performance of some network devices, you can specify the maximum number of redistributed routes.

Restrictions and guidelines

This command redistributes only active routes. To display active routes, use the **display ip routing-table protocol** command.

Configuring IPv4 IS-IS route redistribution

1. Enter system view.

   system-view

2. Enter IS-IS IPv4 unicast address family view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

   address-family ipv4 [**unicast**]

3. Redistribute routes from other routing protocols or other IS-IS processes.

   import-route bgp [*as-number*] [**allow-ibgp**] [**cost** *cost-value* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   import-route bgp [*as-number*] [**allow-ibgp**] **inherit-cost** [[**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   import-route { **direct** | **static** } [**cost** *cost-value* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   import-route { **direct** | **static** } **inherit-cost** [[**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   import-route { **isis** | **ospf** | **rip** } [*process-id* | **all-processes**] [**allow-direct** | **cost** *cost-value* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   import-route { **isis** | **ospf** | **rip** } [*process-id* | **all-processes**] **inherit-cost** [**allow-direct** | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

By default, IS-IS does not redistribute routes.

4. (Optional.) Configure the maximum number of redistributed Level 1/Level 2 IPv4 routes.

   import-route limit *number*

   By default, the maximum number of redistributed Level 1/Level 2 IPv4 routes is 65536.

Configuring IPv6 IS-IS route redistribution

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enter IS-IS IPv6 address family view.

   address-family ipv6 [**unicast**]

4. Redistribute routes from other routing protocols or other IS-IS processes.

   import-route bgp4+ [*as-number*] [**allow-ibgp**] [[**cost** *cost-value* | **inherit-cost**] | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   import-route { **direct** | **static** } [[**cost** *cost-value* | **inherit-cost**] | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   import-route { **isisv6** | **ospfv3** | **ripng** } [*process-id*] [**allow-direct** | [**cost** *cost-value* | **inherit-cost**] | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   By default, IPv6 IS-IS does not redistribute routes.

5. (Optional.) Configure the maximum number of redistributed Level 1/Level 2 IPv6 routes.

   import-route limit *number*

   By default, the maximum number of redistributed Level 1/Level 2 IPv6 routes is 32768.

Filtering routes calculated from received LSPs

About this task

IS-IS saves LSPs received from neighbors in the LSDB, and uses the SPF algorithm to calculate the shortest path tree with itself as the root. IS-IS installs the calculated routes to the IS-IS routing table and the optimal routes to the IP routing table.

Perform this task to filter calculated routes. Only routes that are not filtered can be added to the IP routing table. The filtered routes retain in the IS-IS routing table and can be advertised to neighbors.

Filtering IPv4 IS-IS routes calculated from received LSPs

1. Enter system view.

   system-view

2. Enter IS-IS IPv4 unicast address family view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

   address-family ipv4 [**unicast**]

3. Filter routes calculated using received LSPs.

   filter-policy { *ipv4-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } **import**

   By default, IPv4 IS-IS route filtering is not configured.

Filtering IPv6 IS-IS routes calculated from received LSPs

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IS-IS IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Filter routes calculated using received LSPs.
   filter-policy { *ipv6-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } **import**

   By default, IPv6 IS-IS route filtering is not configured.

Filtering redistributed routes

About this task

IS-IS can redistribute routes from other routing protocols or other IS-IS processes, add them to the IS-IS routing table, and advertise them in LSPs.

Perform this task to filter redistributed routes. Only routes that are not filtered can be added to the IS-IS routing table and advertised to neighbors.

Restrictions and guidelines

Use this command together with the **import-route** command.

Filtering redistributed IPv4 IS-IS routes

1. Enter system view.
   system-view
2. Enter IS-IS IPv4 unicast address family view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   address-family ipv4 [**unicast**]
3. Filter routes redistributed from other routing protocols or IS-IS processes.
   filter-policy { *ipv4-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } **export** [*protocol process-id*]

   By default, IPv4 IS-IS route filtering is not configured.

Filtering redistributed IPv6 IS-IS routes

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IS-IS IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Filter routes redistributed from other routing protocols or IS-IS processes.
   filter-policy { *ipv6-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } **export** [*protocol process-id*]

   By default, IPv6 IS-IS route filtering is not configured.

Excluding redistributed routes from route calculation

About this task

By default, IS-IS adds only the redistributed route into the routing table if the following conditions exist:

- The redistributed route and the learned route have the same route prefix.
- The redistributed route has a higher preference value than the learned route.

In such a scenario, SR-MPLS and SRv6 cannot run normally because IS-IS does not generate any LSP for the learned route. To resolve this issue, use the **import-route no-route-calculate** command to exclude the redistributed route from route calculation. Then, IS-IS adds only the learned route into the routing table and generates an LSP for the learned route.

Restrictions and guidelines

This feature can take effect on a redistributed route only when this route has the same route prefix as a learned route.

Excluding IPv4 redistributed routes from route calculation

1. Enter system view.
   system-view
2. Exclude redistributed routes from route calculation
   import-route no-route-calculate
   By default, redistributed routes can participate in route calculation.

Excluding IPv6 redistributed routes from route calculation

1. Enter system view.
   system-view
2. Enter IS-IS IPv6 unicast address family view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   address-family ipv6 [**unicast**]
3. Exclude redistributed routes from route calculation
   import-route no-route-calculate
   By default, redistributed routes can participate in route calculation.

Configuring IS-IS route leaking

About this task

By default, the Level-2 router does not advertise the routes of the Level-2 area and other Level-1 areas to a Level-1 area. The Level-1 routers in that Level-1 area sends packets destined for other areas to the nearest Level-1-2 router. As a result, IS-IS cannot use the optimal route to forward these packets. To resolve this issue, perform this task to configure IS-IS route leaking.

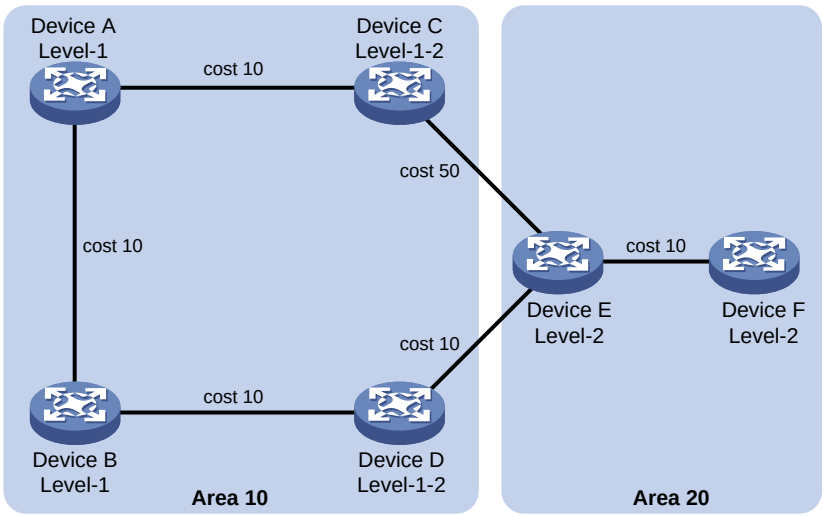
After you configure route leaking on a Level-1-2 router, this feature functions as follows:

1. Filters routes with the specified criteria.
2. Imports the desired routes from other Level-1 areas and the Level-2 area into the Level-1 area to which the Level-1-2 router belongs.

As shown in Figure 8, the optimal path from Device A to Device F is Device A > Device B > Device D > Device E > Device F and the total cost is 40. By default, the actual path to Device F on Device A is Device A > Device C > Device E > Device F and the total cost is 70. After you configure route leaking

on Device C and Device D, IS-IS will use the Device A > Device B > Device D > Device E > Device F path to forward traffic from Device A to Device F.

Figure 8 Route leaking



Perform this task to control route advertisement (route leaking) between Level-1 and Level-2.

You can configure IS-IS to advertise routes from Level-2 to Level-1, and to not advertise routes from Level-1 to Level-2.

Configuring IPv4 IS-IS route leaking

1. Enter system view.

   system-view

2. Enter IS-IS IPv4 unicast address family view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

   address-family ipv4 [**unicast**]

3. Configure route leaking from Level-1 to Level-2.

   import-route isis level-1 into level-2 [**filter-policy** { *ipv4-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } | **tag** *tag*] *

   By default, IS-IS advertises routes from Level-1 to Level-2.

4. Configure route leaking from Level-2 to Level-1.

   import-route isis level-2 into level-1 [**filter-policy** { *ipv4-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } | **tag** *tag*] *

   By default, IS-IS does not advertise routes from Level-2 to Level-1.

Configuring IPv6 IS-IS route leaking

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enter IS-IS IPv6 address family view.

   address-family ipv6 [**unicast**]

4. Configure route leaking from Level-2 to Level-1.

import-route isisv6 level-2 into level-1 [**filter-policy** { *ipv6-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } | **tag** *tag*] *

By default, IS-IS does not advertise routes from Level-2 to Level-1.

5. Configure route leaking from Level-1 to Level-2.

import-route isisv6 level-1 into level-2 [**filter-policy** { *ipv6-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } | **tag** *tag*] *

By default, IS-IS advertises routes from Level-1 to Level-2.

Advertising IS-IS link state information to other protocols

About this task

After the device advertises IS-IS link state information to other protocols, such as BGP, these protocols can then advertise the information for intended applications.

Restrictions and guidelines

When multiple IS-IS P2P neighbors exist between two devices, use one of the following commands to add the related local interface IP addresses into link state information. This operation ensures that other protocols can distinguish PPP links.

- **advertise link-attributes**
- **router-id** (IS-IS IPv6 address family view)

Procedure

1. Enter system view.

system-view

2. Enter IS-IS view.

isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Advertise IS-IS link state information to other protocols.

distribute { **bgp-ls** | **link-state** | **sr-db** } [**instance-id** *id*] [**level-1** | **level-2**]

By default, the device does not advertise IS-IS link state information to any other protocol.

Adjusting IS-IS routing convergence speed

Specifying the interval for sending IS-IS hello packets

About this task

IS-IS uses hello packets to maintain neighbor relationships. If a neighbor does not receive any hello packets from the router within the advertised hold time, it considers the router down and recalculates the routes. The hold time is the hello multiplier multiplied by the hello interval.

Restrictions and guidelines

The interval between hello packets sent by the DIS is 1/3 the hello interval set with the **isis timer hello** command.

Procedure

1. Enter system view.

system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the interval for sending hello packets.
   isis timer hello *seconds* [**level-1** | **level-2**]
   The default setting is 10 seconds.

Specifying the interval for sending IS-IS CSNP packets

About this task

On a broadcast network, perform this task on the DIS that uses CSNP packets to synchronize LSDBs.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the interval for sending CSNP packets on the DIS of a broadcast network.
   isis timer csnp *seconds* [**level-1** | **level-2**]
   The default setting is 10 seconds.

Setting the maximum age of LSPs

About this task

Each LSP has an age that decreases in the LSDB. Any LSP with an age of 0 is deleted from the LSDB. You can adjust the age value based on the scale of a network.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the maximum LSP age.
   timer lsp-max-age *seconds*
   The default setting is 1200 seconds.

Setting the initial remaining lifetime of IS-IS LSPs in the LSDB

About this task

A large number of LSPs might queue up for flooding in the LSDB of an advertisement source, which inevitably shortens their remaining lifetime. In this situation, the remaining lifetime of some LSPs might be too short when they are flooded. Such LSPs soon expire on the recipient side, causing LSP flapping or temporary unreachable routes. To resolve this issue, perform this task to set the initial remaining lifetime of IS-IS LSPs in the LSDB.

The **timer lsp-remaining-lifetime refresh**{ *refresh-value* | **lsp-max-age** } command takes effect only when the following requirements are met:

- The local device receives an LSP update from another device.
- The remaining lifetime of the LSP is not 0, and is smaller than that specified by the *refresh-value* argument or the **lsp-max-age** keyword in this command.

Restrictions and guidelines

As a best practice to prevent LSP flapping, execute the **timer lsp-remaining-lifetime refresh**{ *refresh-value* | **lsp-max-age** } command to adjust the remaining lifetime of IS-IS LSPs.

To have the local device use the original remaining lifetime of an updated LSP, execute the **timer lsp-remaining-lifetime refresh disable** command.

Setting the initial remaining lifetime of IS-IS LSPs in the LSDB

1. Enter system view.

   system-view
2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the initial remaining lifetime of IS-IS LSPs in the LSDB.

   timer lsp-remaining-lifetime refresh { *refresh-value* | **lsp-max-age** | **disable** }

   By default, when IS-IS receives an updated LSP, it sets the maximum LSP age as the initial remaining lifetime of this LSP.

Disabling configuring the initial remaining lifetime of IS-IS LSPs in the LSDB

1. Enter system view.

   system-view
2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Disable configuring the initial remaining lifetime of IS-IS LSPs in the LSDB.

   timer lsp-remaining-lifetime refresh disable

   By default, when IS-IS receives an updated LSP, it sets the maximum LSP age as the initial remaining lifetime of this LSP.

   After you execute this command, the local device uses the original remaining lifetime of an LSP when it receives an update of that LSP.

Setting the LSP refresh interval and generation interval

About this task

Each router needs to refresh its LSPs at a configurable interval and send them to other routers to prevent valid routes from aging out. A smaller refresh interval speeds up network convergence but consumes more bandwidth.

When network topology changes such as neighbor state, interface metric, system ID, or area ID changes occur, the router generates an LSP after a configurable interval. If such a change occurs frequently, excessive LSPs are generated, consuming a large amount of router resources and bandwidth. To solve the problem, you can adjust the LSP generation interval.

Restrictions and guidelines

Follow these restrictions and guidelines when you configure the **timer lsp-generation** command:

- If you specify only the *maximum-interval* argument, the LSP generation interval is *maximum-interval*.
- If you do not specify the *incremental-interval* argument, the LSP generation interval is in the range of *minimum-interval* to *maximum-interval*.
- If you specify the *incremental-interval* argument, the LSP generation interval is as follows:
  - When network changes are not frequent, the *minimum-interval* is adopted.
  - If network changes are frequent, the LSP generation interval increases by *incremental-interval* × 2^{n-2} (n is the number of calculation times) each time a generation occurs until the *maximum-interval* is reached.

Procedure

1. Enter system view.

   system-view
2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the LSP refresh interval.

   timer lsp-refresh *seconds*

   By default, the LSP refresh interval is 900 seconds.
4. Set the LSP generation interval.

   timer lsp-generation *maximum-interval* [*minimum-interval* [*incremental-interval*]] [**level-1** | **level-2**]

   By default:
   - The maximum interval is 5 seconds.
   - The minimum interval is 50 milliseconds.
   - The incremental interval is 200 milliseconds.

Setting LSP sending intervals

About this task

If a change occurs in the LSDB, IS-IS advertises the changed LSP to neighbors. You can specify the minimum interval for sending these LSPs to control the amount of LSPs on the network.

On a P2P link, IS-IS requires an advertised LSP be acknowledged. If no acknowledgment is received within a configurable interval, IS-IS will retransmit the LSP.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify the minimum interval for sending LSPs and the maximum LSP number that can be sent at a time.

   isis timer lsp *time* [**count** *count*]

   By default, the minimum interval is 33 milliseconds, and the maximum LSP number that can be sent at a time is 5.
4. Specify the LSP retransmission interval on a P2P link.

   isis timer retransmit *seconds*

By default, the LSP retransmission interval on a P2P link is 5 seconds.

Setting the LSP receiving interval

About this task

When the network is stable, IS-IS uses the minimum receiving interval. When the network changes frequently, the receiving interval increases by the incremental interval each time the same LSP is received until the maximum receiving interval is reached.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the LSP receiving interval.
   timer lsp-arrival *maximum-interval* [*minimum-interval* [*incremental-interval*]] [**level-1** | **level-2**]
   By default, the LSP receiving interval is not set.

Setting the SPF calculation interval

About this task

Based on the LSDB, an IS-IS router uses the SPF algorithm to calculate the shortest path tree with itself being the root, and uses the shortest path tree to determine the next hop to a destination network. By adjusting the SPF calculation interval, you can prevent bandwidth and router resources from being over consumed due to frequent topology changes.

When network changes are not frequent, the *minimum-interval* is adopted. If network changes become frequent, the SPF calculation interval increases by *incremental-interval* × 2^{n-2} (n is the number of calculation times) each time a calculation occurs until the *maximum-interval* is reached.

Setting the IPv4 SPF calculation interval

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the SPF calculation interval.
   timer spf { *maximum-interval* [*minimum-interval* [*incremental-interval* [**conservative**]]] | **millisecond** *millisecond-interval* } [**exclude-prc**]
   By default:
   - The maximum interval is 5 seconds.
   - The minimum interval is 50 milliseconds.
   - The incremental interval is 200 milliseconds.

Setting the IPv6 SPF calculation interval

1. Enter system view.
   system-view

2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Set the SPF calculation interval.
   timer spf { *maximum-interval* [*minimum-interval* [*incremental-interval* [**conservative**]]] | **millisecond** *millisecond-interval* } [**exclude-prc**]
   By default:
   - The maximum interval is 5 seconds.
   - The minimum interval is 50 milliseconds.
   - The incremental interval is 200 milliseconds.

Configuring convergence priorities for specific routes

About this task

A topology change causes IS-IS routing convergence. To improve convergence speed, you can assign convergence priorities to IS-IS routes. Convergence priority levels are critical, high, medium, and low. The higher the convergence priority, the faster the convergence speed.

After you use the **prefix-priority** command, IS-IS sets the convergence priorities for IS-IS routes as follows:

- IS-IS sets the convergence priorities for routes that already exist according to the configuration of the **prefix-priority** command.
- IS-IS sets the convergence priorities for subsequent routes according to the filter result of the **prefix-priority** command.
- When a route matches the rules of multiple convergence priorities, it uses the highest priority.

Configuring convergence priorities for specific IPv4 IS-IS routes

1. Enter system view.
   system-view
2. Enter IS-IS IPv4 unicast address family view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   address-family ipv4 [**unicast**]
3. Assign convergence priorities to specific IPv4 IS-IS routes.
   - Assign a convergence priority to IPv4 IS-IS routes matching the specified prefix list.
     prefix-priority { **critical** | **high** | **medium** } { **prefix-list** *prefix-list-name* | **tag** *tag-value* }
   - Assign a convergence priority to IPv4 IS-IS routes by using a route policy.
     prefix-priority route-policy *route-policy-name*

   By default, IPv4 IS-IS routes, except IS-IS host routes, have the low convergence priority.

Configuring convergence priorities for specific IPv6 IS-IS routes

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IPv6 address family view.

address-family ipv6 [**unicast**]

4. Assign convergence priorities to specific IPv6 IS-IS routes.

   prefix-priority { **critical** | **high** | **medium** } { **prefix-list** *prefix-list-name* | **tag** *tag-value* }

   prefix-priority route-policy *route-policy-name*

   By default, IPv6 IS-IS routes, except IS-IS host routes, have the low convergence priority.

Configuring IS-IS packet-related features

Configuring a DIS priority for an interface

About this task

On a broadcast network, IS-IS must elect a router as the DIS at a routing level. You can specify a DIS priority at a level for an interface. The greater the interface's priority, the more likely it becomes the DIS. If multiple routers in the broadcast network have the same highest DIS priority, the router with the highest MAC address becomes the DIS.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure a DIS priority for the interface.

   isis dis-priority *priority* [**level-1** | **level-2**]

   The default setting is 64.

Configuring the administrative tag value for an interface

About this task

This task simplifies route management by allowing IS-IS to control IP address prefixes with the administrative tag.

As shown in Figure 9, Device A needs to communicate with Device B, Device C, and Device D. For security purposes, other devices in areas 2, 3, and 5 should not receive the packets sent by Device A. To resolve this issue, perform the following task:

1. Configure the same administrative tag value for the IS-IS interfaces on Device B, Device C, and Device D.
2. Configure route leaking from Level-2 to Level-1 on the Level-1-2 device in Area 4 and specify the same tag value as the route filter.

As shown in Figure 10, the IS-IS network topology on Device A is updated after you perform the above task.

Figure 9 IS-IS network diagram without the administrative tag

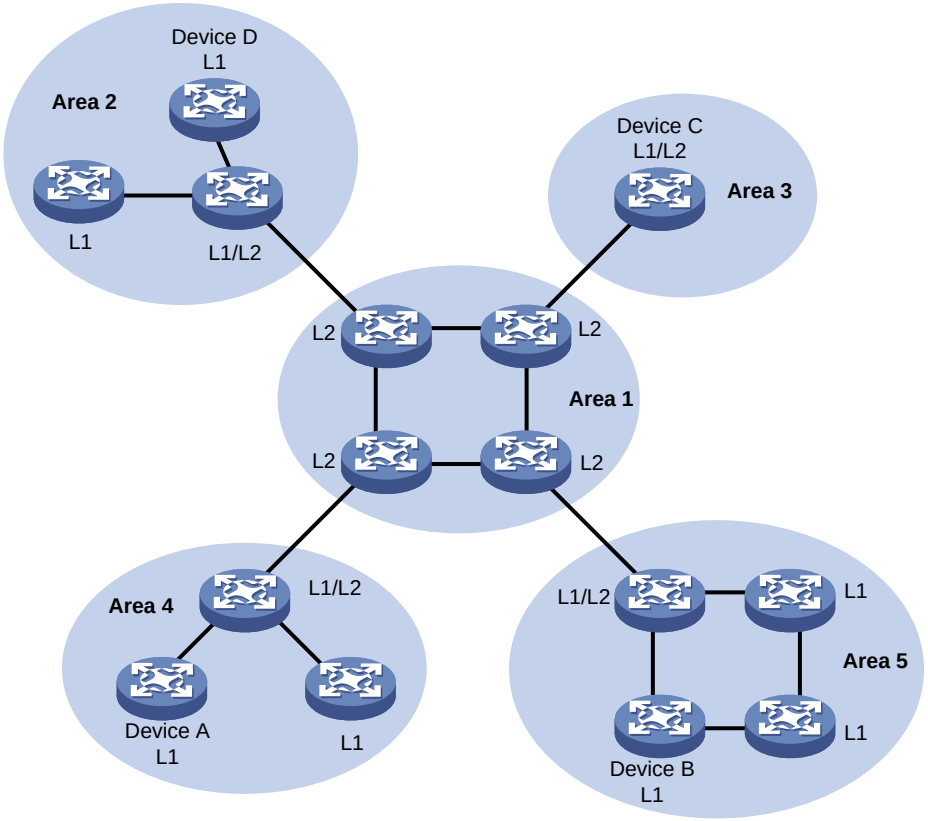
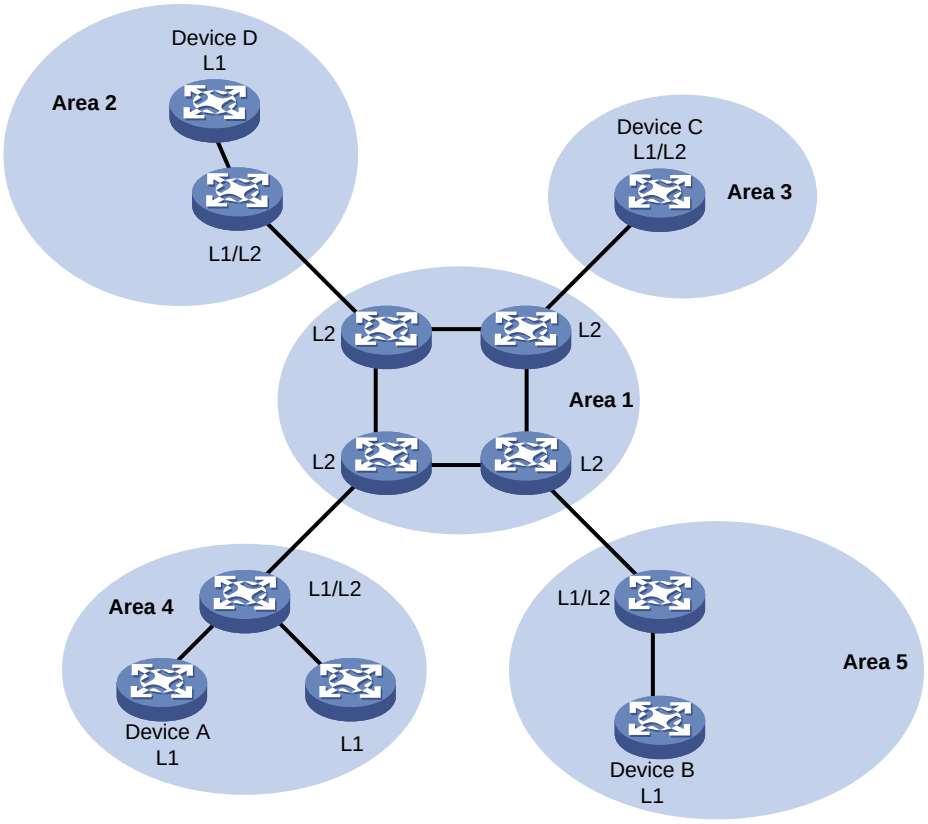


Figure 10 IS-IS network diagram with the administrative tag



Configuring the IPv4 IS-IS administrative tag value for an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the IPv4 IS-IS administrative tag value for the interface.
   isis tag *tag*

   By default, the IPv4 IS-IS administrative tag value of the interface is not configured.

   IS-IS adds the administrative tag to the network address information in LSPs only if the following conditions exist:
   - An administrative tag value is specified.
   - The link cost style is **wide**, **wide-compatible**, or **compatible**.

Configuring the IPv6 IS-IS administrative tag value for an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the IPv6 IS-IS administrative tag value for the interface.
   isis ipv6 tag *tag*

   By default, the IPv6 IS-IS administrative tag value of the interface is not configured.

   After you specify an administrative tag value, IS-IS always adds the specified tag value to the IPv6 prefix information in LSPs, regardless of the link cost style.

Specifying the IS-IS hello multiplier

About this task

The hello multiplier is the number of hello packets a neighbor must miss before it declares that the router is down.

If a neighbor receives no hello packets from the router within the advertised hold time, it considers the router down and recalculates the routes. The hold time is the hello multiplier multiplied by the hello interval.

On a broadcast link, Level-1 and Level-2 hello packets are advertised separately. You must set a hello multiplier for each level.

On a P2P link, Level-1 and Level-2 hello packets are advertised in P2P hello packets. You do not need to specify Level-1 or Level-2.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the hello multiplier.
   isis timer holding-multiplier *value* [**level-1** | **level-2**]

   The default setting is 3.

Disabling an interface from sending/receiving IS-IS packets

About this task

After being disabled from sending and receiving hello packets, an interface cannot form any neighbor relationship, but can advertise directly connected networks in LSPs through other interfaces. This can save bandwidth and CPU resources, and ensures that other routers know networks directly connected to the interface.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Disable the interface from sending and receiving IS-IS packets.
   isis silent
   By default, the interface can send and receive IS-IS packets.

Enabling an interface to send small hello packets

About this task

IS-IS messages cannot be fragmented at the IP layer because they are directly encapsulated in frames. Any two IS-IS neighboring routers must negotiate a common MTU. To avoid sending big hellos to save bandwidth, enable the interface to send small hello packets without CLVs.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the interface to send small hello packets without CLVs.
   isis small-hello
   By default, the interface sends standard hello packets.

Setting LSP lengths

About this task

IS-IS messages cannot be fragmented at the IP layer because they are directly encapsulated in frames. IS-IS routers in an area must send LSPs smaller than the smallest interface MTU in the area.

If the IS-IS routers have different interface MTUs, configure the maximum size of generated LSP packets to be smaller than the smallest interface MTU in the area. Without the configuration, the routers must dynamically adjust the LSP packet size to fit the smallest interface MTU, which takes time and affects other services.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Specify the maximum length of generated Level-1 LSPs or Level-2 LSPs.

   lsp-length originate *size* [**level-1** | **level-2**]

   By default, the maximum length of generated Level-1 LSPs or Level-2 LSPs is 1497 bytes.

4. Specify the maximum length of received LSPs.

   lsp-length receive *size*

   By default, the maximum length of received LSPs is 1497 bytes.

Enabling LSP flash flooding

About this task

Changed LSPs can trigger SPF recalculation. To advertise the changed LSPs before the router recalculates routes for faster network convergence, enable LSP flash flooding.

Procedure

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enable LSP flash flooding.

   flash-flood [**flood-count** *flooding-count* | **max-timer-interval** *flooding-interval* | [**level-1** | **level-2**]] *

   By default, LSP flash flooding is disabled.

Enabling LSP fragment extension

About this task

When the LSP capacity of a device is insufficient, enable LSP fragment extension to expand the LSP capacity. These LSPs can then carry more IS-IS information, including redistributed routes and TLVs. If IS-IS fails to add some redistributed routes or TLVs into an LSP, it will automatically try to re-add those redistributed routes and TLVs into the LSP after LSP fragment extension.

The LSP fragment extension feature involves the following concepts:

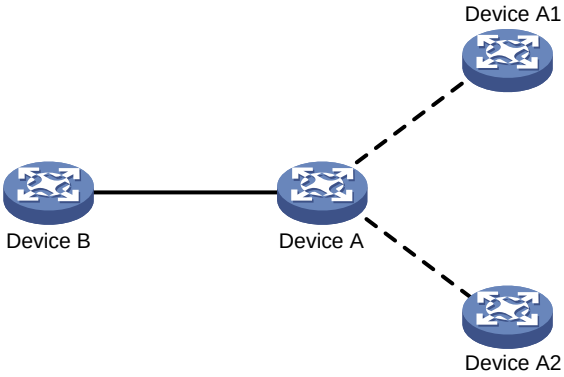
- **Originating System**—IS-IS process.
- **Normal System ID**—System ID of the originating system.
- **Additional System ID**—Unique identifier assigned to a virtual system.
- **Virtual System**—System used to generate extended LSP fragments. Each extended LSP fragment carries the additional system ID of the virtual system in its LSP ID.
- **IS Alias ID TLV**—TLV that describes the relationship between the originating system and the virtual system.

By default, an IS-IS process can generate a maximum of 256 LSP fragments. The number of routes that these LSP fragments can carry is limited. To resolve this issue, the LSP fragment extension feature introduces the concept of additional system ID. Each additional system ID can generate 256 extended LSP fragments. When the LSP capacity of the originating system is insufficient, IS-IS adds the overflowing route information together with the IS Alias ID TLV into the extended LSPs of a virtual system.

As shown in Figure 11, both Device A and Device B support LSP fragment extension. With LSP fragment extension enabled, Device A advertises some route information in the LSPs generated by Device A1 and Device A2. On receipt of these LSPs, Device B reads the IS Alias ID TLVs and determines that Device A is the originating system. The route information received from Device A1

and Device A2 is actually generated by Device A. In this situation, virtual systems do not participate in route calculation.

Figure 11 IS-IS LSP fragment extension



Procedure

2. Enter system view.

   system-view

3. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

4. Enable LSP fragment extension.

   lsp-fragments-extend [**level-1** | **level-1-2** | **level-2**]

   By default, LSP fragment extension is disabled.

   The MTUs of all interfaces running the IS-IS process must not be less than 512. Otherwise, LSP fragment extension does not take effect.

5. Configure a virtual system ID.

   virtual-system *virtual-system-id*

   By default, no virtual system ID is configured.

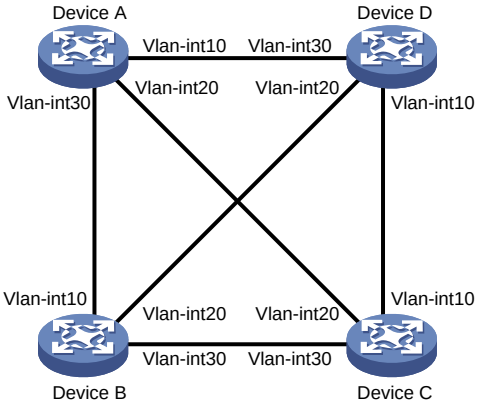
   Configure a minimum of one virtual system to generate extended LSP fragments.

Limiting LSP flooding

About this task

On a fully meshed network, many P2P links exist. As shown in Figure 12, Devices A, B, C and D run IS-IS. When Device A generates an LSP, it floods the LSP out of VLAN-interface 10, VLAN-interface 20, and VLAN-interface 30. After Device D receives the LSP from VLAN-interface 30, it floods it out of VLAN-interface 10 and VLAN-interface 20 to Device B and Device C. However, Device B and Device C have already received the LSP from Device A. Repeated LSP flooding consumes extra bandwidth.

Figure 12 Network diagram of a fully meshed network



To avoid this problem, you can add interfaces to a mesh group or block some interfaces.

- An interface in a mesh group floods a received LSP only to interfaces not in the mesh group.
- A blocked interface sends LSPs only after receiving LSP requests.

Restrictions and guidelines

Before you configure this task, you must consider redundancy for interfaces in case LSP packets cannot be flooded because of link failures.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Limit LSP flooding.
   - Add the interface to a mesh group.
     isis mesh-group *mesh-group-number*
   - Block the interface.
     isis mesh-group mesh-blocked

   By default, the interface does not belong to any mesh group and is not blocked.

   The commands take effect only on P2P interfaces.

Configuring IS-IS route update suppression

About this task

Perform this task to protect network resources and routers from being overwhelmed by LSPs when route flapping occurs.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Configure IS-IS route update suppression.
   - Suppress LSP generation.

timer lsp-generation suppress-flapping *delay-interval* [**threshold** *threshold-value*] [**level-1** | **level-2**]

By default, IS-IS does not suppress LSP generation.

- Suppress LSP flooding.

  timer lsp-flood suppress-flapping *delay-interval* [**threshold** *threshold-value*] [**level-1** | **level-2**]

  By default, IS-IS does not suppress LSP flooding.

- Suppress route calculation.

  timer route-calculate suppress-flapping *delay-interval* [**threshold** *threshold-value*] [**level-1** | **level-2**]

  By default, IS-IS does not suppress route calculation.

- Suppress route calculation after fragment 0 of a purge LSP is received.

  timer purge-zero-lsp route-calculate-delay *delay-interval* [**level-1** | **level-2**]

  By default, IS-IS suppresses route calculation for 10 seconds after fragment 0 of a purge LSP is received.

Configuring advanced IS-IS features

Enabling source address check for hello packets on a P2P interface

About this task

An IS-IS P2P interface can have a peer on a different network. Perform this task to configure an IS-IS P2P interface to establish neighbor relationship only with a peer on the same network.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable source address check for hello packets on a P2P interface.

   isis peer-ip-check

   By default, an IS-IS P2P interface can have a peer on a different network.

Setting the LSDB overload bit

About this task

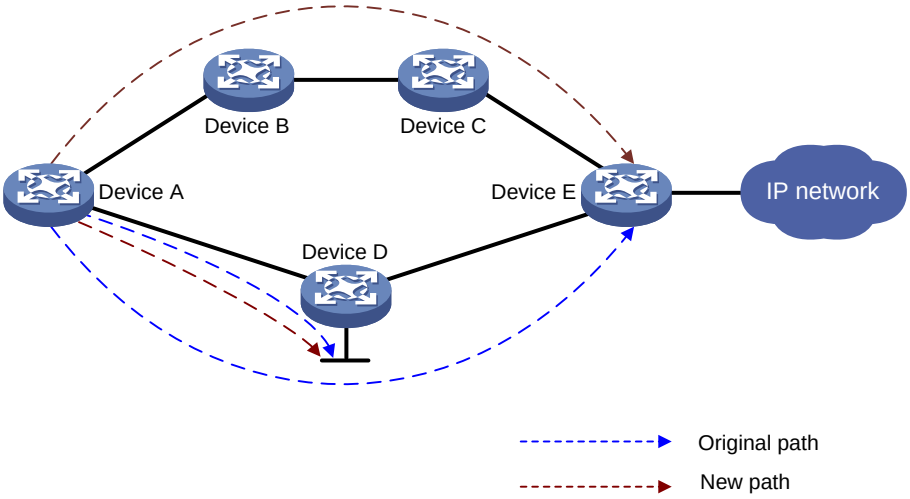
By setting the overload bit in sent LSPs, an IS-IS device informs other devices of failures that make it unable to select routes and forward packets. The device is then excluded from the SPF calculation by other devices except its direct routes.

When an IS-IS device cannot record the complete LSDB, for example, because of memory insufficiency, it will calculate routes incorrectly. To locate the faulty device, you can temporarily isolate the suspects from the IS-IS network by setting the overload bit.

As shown in Figure 13, the path from Device A to the IP network should be Device A > Device D > Device E > IP network. After you set the overload bit in the LSPs sent by Device D, the path from Device A to the IP network becomes Device A > Device B > Device C > Device E > IP network. If the

packets from Device A are destined for Device D's directly-connected networks, the forwarding path does not change.

Figure 13 Overload bit



Restrictions and guidelines

When both IPv4 and IPv6 topologies exist, the following rules apply to the **set-overload** command and the **multi-topology** command:

- If the **set-overload** command is executed in IS-IS view and the **multi-topology** command is not executed, the **set-overload** command takes effect on both IPv4 IS-IS and IPv6 IS-IS.
- If the following conditions exist, the **set-overload** command does not take effect on IPv6 IS-IS after it is executed in IS-IS IPv6 unicast address family view:
  - The **set-overload** command is not executed in IS-IS view.
  - The **multi-topology** command is not executed.
- If the **set-overload** command is executed in IS-IS view and the **multi-topology** command is executed, the **set-overload** command takes effect only on IPv4 IS-IS. To have the **set-overload** command take effect on IPv6 IS-IS, you must execute it in IS-IS IPv6 unicast address family view.

Setting the LSDB overload bit for IPv4 IS-IS

1. Enter system view.

   system-view
2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the overload bit.

   set-overload [**on-startup** [[**start-from-nbr** *system-id* [*timeout1* [*nbr-timeout*]]] | *timeout2* | **wait-for-bgp** [*timeout3*]]] [**allow** { **external** | **interlevel** } *]

   By default, the overload bit is not set.

Setting the LSDB overload bit for IPv6 IS-IS

1. Enter system view.

   system-view
2. Enter IS-IS view.

isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Specify an IS-IS cost style.

   cost-style { **compatible** | **wide** | **wide-compatible** }

   By default, IS-IS only transmits and receives packets using the **narrow** cost style.

4. Enter IPv6 address family view.

   address-family ipv6 [**unicast**]

5. Enable IPv6 IS-IS MTR.

   multi-topology [**compatible**]

   By default, IPv6 IS-IS MTR is disabled.

6. Set the overload bit.

   set-overload [**on-startup** [[**start-from-nbr** *system-id* [*timeout1* [*nbr-timeout*]]] | *timeout2* | **wait-for-bgp4+** [*timeout3*]]] [**allow** { **external** | **interlevel** } *]

   By default, the overload bit is not set.

Configuring IS-IS isolation

About this task

Isolation is a method used for network device maintenance. It gracefully removes a device from the packet forwarding path for maintenance and gracefully adds the device to the network after maintenance.

To reduce impact on traffic forwarding, you can isolate a device before upgrading it. IS-IS isolation works as follows:

1. After IS-IS isolation is enabled for a device, IS-IS sets the overload bit in the LSPs advertised by the device and sets the link cost to the maximum value.
2. Each neighbor of the device reselects an optimal route based on the LSPs and stops forwarding traffic to the device. The device is fully isolated from the network and you can upgrade the device.
3. After the maintenance, disable IS-IS isolation on the device to gracefully add it back to the network by clearing its overload bit and restoring its link cost.

Procedure

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Configure IS-IS isolation.

   a. Enable IS-IS isolation to gracefully remove the device from the network.

      isolate enable

   b. Disable IS-IS isolation to gracefully add the device to the network.

      undo isolate enable

   By default, IS-IS isolation is disabled.

Configuring IS-IS shutdown

About this task

For maintenance purposes, you can use this feature to shut down IS-IS processes on the device with small impact on the network. If you shut down an IS-IS process, it will perform the following operations:

- Change the state of all neighbors to down.
- Stop receiving and sending IS-IS packets.
- Clear its neighbor, LSDB, and IS-IS route information.

After maintenance, you can use the **undo shutdown process** command to restart the IS-IS process for neighbor relationship re-establishment.

This feature shuts down an IS-IS process while retaining the process-associated settings to facilitate your maintenance.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Configure IS-IS shutdown.
   a. Shut down an IS-IS process to isolate it from the network.
      shutdown process
   b. Restart the IS-IS process to add it back to the network.
      undo shutdown process

   By default, the IS-IS process is not shut down.

Configuring the ATT bit

About this task

The ATT bit is used to identify the connection status between a Level-1 area and other areas. By default, a Level-1-2 router sets the ATT bit for Level-1 LSPs as follows:

- The Level-1-2 router sets the ATT bit in Level-1 LSPs to inform the Level-1 devices that it can reach other areas. After a Level-1 device receives a Level-1 LSP with the ATT bit set, it generates a default route destined for the Level-1-2 router.
- The Level-1-2 router does not set the ATT bit in Level-1 LSPs if it can reach only one area.

To edit the default ATT bit setting rule for a Level-1-2 router, perform the following tasks as needed:

- To enable ATT bit setting for all Level-1 LSPs, execute the **set-att always** command on the Level-1-2 router.
- To disable a Level-1 device from generating a default route upon receiving an ATT-bit-set Level-1 LSP from the Level-1-2 router, you can perform one of the following tasks:
  - Execute the **ignore-att** command on the Level-1 device.
  - Execute the **set-att never** command on the Level-1-2 router.

Configuring IS-IS not to calculate the default route through the ATT bit

1. Enter system view.
   system-view

2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Configure IS-IS not to calculate the default route through the ATT bit.
   ignore-att
   By default, IS-IS uses the ATT bit to calculate the default route.

Setting the ATT bit of IPv4 Level-1 LSPs

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the ATT bit of IPv4 Level-1 LSPs.
   set-att { **always** | **never** }
   By default, the Level-1-2 router sets the ATT bit for IPv4 Level-1 LSPs in accordance with the default ATT bit setting rule.

Setting the ATT bit of IPv6 Level-1 LSPs

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Set the ATT bit of IPv6 Level-1 LSPs.
   set-att { **always** | **never** }
   By default, the Level-1-2 router sets the ATT bit for IPv6 Level-1 LSPs in accordance with the default ATT bit setting rule.

Configuring system ID to host name mappings

About this task

A 6-byte system ID in hexadecimal notation uniquely identifies a router or host in an IS-IS network. To make a system ID easy to read, the system allows you to use host names to identify devices. It also provides mappings between system IDs and host names.

The mappings can be configured manually or dynamically.

- **Static system ID to host name mapping**—You must manually configure a mapping for each router in the network. When a new router is added to the network or a mapping must be modified, you must configure all routers manually.
- **Dynamic system ID to host name mapping**—You only need to configure a host name for each router in the network. Each router advertises the host name in a dynamic host name CLV to other routers so all routers in the network can have all mappings. To help check the origin of LSPs in the LSDB, you can configure a name for the DIS in a broadcast network.

Restrictions and guidelines

Follow these guidelines when you configure the mappings:

- To view host names rather than system IDs by using the **display isis lsdb** command, you must enable dynamic system ID to host name mapping.

- If you configure both dynamic mapping and static mapping on a router, the host name specified for dynamic mapping applies.

Configuring a static system ID to host name mapping

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Configure a system ID to host name mapping for a remote IS.
   is-name map *sys-id map-sys-name*

   By default, no system ID to host name mapping is configured for a remote IS.

   A system ID can correspond to only one host name.

Configuring dynamic system ID to host name mapping

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Specify a host name for the IS and enable dynamic system ID to host name mapping.
   is-name *sys-name*

   By default, dynamic system ID to host name mapping is disabled and no host name is specified for the router.
4. Return to system view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Configure a DIS name.
   isis dis-name *symbolic-name*

   By default, no DIS name is configured.

   This command takes effect only on a router enabled with dynamic system ID to host name mapping.

   This command is not available on P2P interfaces.

Configuring IS-IS logging and SNMP notifications

Enabling the logging of neighbor state changes

About this task

Neighbor flappings cause a series of issues, such as network instability, unnecessary resource consumption, and even data loss. You can enable the logging of neighbor state changes for fast fault location and troubleshooting.

With this feature enabled, the device delivers logs about neighbor state changes to its information center. The information center processes the logs according to user-defined output rules (whether to output logs and where to output). For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable the logging of neighbor state changes.
   log-peer-change
   By default, the logging of neighbor state changes is enabled.

Setting the maximum number of log entries that can be recorded

About this task

Perform this task to set the maximum number of log entries that IS-IS can record for each log type.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the maximum number of log entries that IS-IS can record.
   event-log { **hello** { **peer-change** | **received-abnormal** | **received-dropped** | **sent-abnormal** | **sent-failed** } | **peer** } **size** *count*
   By default, IS-IS can record 100 log entries for each log type.

Setting the maximum number of advertisement sources that trigger LSP update logging

About this task

After the device starts running, it records LSP-update log information every 15 minutes. By default, it can record LSP-update log information for up to 200 advertisement sources. LSP update logging consumes memory resources. To control the memory usage by LSP update logging, you can perform this task.

Procedure

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the maximum number of advertisement sources that trigger LSP update logging.
   event-log updated-lsp size *count*
   By default, a maximum of 200 advertisement sources can trigger LSP update logging.

Setting the maximum number of IS-IS route change log messages that can be recorded

About this task

Supported IS-IS route change log messages include:

- **Route adding**—The device generates a route adding log message when IS-IS adds a route into the IP routing table.
- **Route deletion**—The device generates a route deletion log message when IS-IS deletes a route from the IP routing table.
- **Route update**—The device generates a route update log message when IS-IS updates a route in the IP routing table.

After the device starts running, it records IS-IS route change log messages. This operation consumes memory resources. To save memory resources, perform this task to adjust the maximum number of IS-IS route change log messages that the device can record.

Procedure

1. Enter system view.

   system-view
2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the maximum number of IS-IS route change log messages that the device can record.

   event-log route-record { **add** | **delete** | **modify** } **size** *count*

   By default, the device can record up to 5000 IPv4 IS-IS route change log messages for each log type and up to 5000 IPv6 IS-IS route change log messages for each log type.

Setting the maximum number of IS-IS neighbor relationship troubleshooting entries

1. Enter system view.

   system-view
2. Set the maximum number of IS-IS neighbor relationship troubleshooting entries.

   isis troubleshooting max-number *number*

   By default, IS-IS can record a maximum of 100 neighbor relationship troubleshooting entries.

Configuring IS-IS network management

About this task

This task includes the following configurations:

- Bind an IS-IS process to MIB so that you can use network management software to manage the specified IS-IS process.
- Enable IS-IS notifications to report important events.

To report critical IS-IS events to an NMS, enable SNMP notifications for IS-IS. For SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.
   system-view
2. Bind MIB to an IS-IS process.
   isis mib-binding *process-id*
   By default, MIB is bound to the IS-IS process with the smallest process ID.
3. Enable IS-IS notification sending.
   snmp-agent trap enable isis [**adjacency-protocol-change** | **adjacency-state-change** | **area-mismatch** | **authentication** | **authentication-type** | **buffsize-mismatch** | **id-length-mismatch** | **lsdboverload-state-change** | **lsp-corrupt** | **lsp-parse-error** | **lsp-size-exceeded** | **manual-address-drop** | **max-seq-exceeded** | **maxarea-mismatch** | **own-lsp-purge** | **protocol-support** | **rejected-adjacency** | **skip-sequence-number** | **version-skew**] *
   By default, IS-IS notification sending is enabled.
4. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
5. Configure the context name for the SNMP object for managing IS-IS.
   snmp context-name *context-name*
   By default, no context name is set for the SNMP object for managing IS-IS.

Configuring IS-IS fast convergence

Enabling ISPF

About this task

When the network topology changes, Incremental Shortest Path First (ISPF) computes only the affected part of the SPT, instead of the entire SPT.

Enabling IPv4 IS-IS ISPF

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable IPv4 IS-IS ISPF.
   ispf enable
   By default, IPv4 IS-IS ISPF is enabled.

Enabling IPv6 IS-IS ISPF

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enter IPv6 address family view.
   address-family ipv6 [**unicast**]
4. Enable IPv6 IS-IS ISPF.

ispf enable

By default, IPv6 IS-IS ISPF is enabled.

Enabling prefix suppression

About this task

Perform this task to disable an interface from advertising its prefix in LSPs. This enhances network security by preventing IP routing to the interval nodes and speeds up network convergence.

Enabling IPv4 IS-IS prefix suppression

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IPv4 IS-IS prefix suppression on the interface.
   isis prefix-suppression

   By default, IPv4 IS-IS prefix suppression is disabled on the interface.

   This command is also applicable to the secondary IP address of the interface.

Enabling IPv6 IS-IS prefix suppression

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IPv6 IS-IS prefix suppression on the interface.
   isis ipv6 prefix-suppression

   By default, IPv6 IS-IS prefix suppression is disabled on the interface.

Configuring IS-IS PIC

About this task

Prefix Independent Convergence (PIC) enables the device to speed up network convergence by ignoring the number of prefixes.

Restrictions and guidelines for IS-IS PIC

Follow these restrictions and guidelines when you configure IS-IS PIC:

- When both IS-IS PIC and IS-IS FRR are configured, IS-IS FRR takes effect.
- IS-IS PIC applies only to LSPs sent by neighbors.

Enabling IS-IS PIC

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable PIC for IS-IS.
   pic [**additional-path-always**]

By default, IS-IS PIC is enabled.

Enabling BFD for IS-IS PIC

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable BFD control packet mode for IS-IS PIC.

   isis primary-path-detect bfd ctrl

   By default, BFD control packet mode is disabled for IS-IS PIC.

   To use BFD (control packet mode) to detect primary link failures, you must enable BFD control packet mode on both ends of the link.

Enabling BFD echo packet mode for IS-IS PIC

1. Enter system view.

   system-view
2. Configure the source IP address of BFD echo packets.

   bfd echo-source-ip *ip-address*

   By default, the source IP address of BFD echo packets is not configured.

   The source IP address cannot be on the same network segment as any local interface's IP address.

   For more information, see *High Availability Command Reference*.
3. Enter interface view.

   interface *interface-type interface-number*
4. Enable BFD echo packet mode for IS-IS PIC.

   isis primary-path-detect bfd echo

   By default, BFD echo packet mode is disabled for IS-IS PIC.

   To use BFD (echo packet mode) to detect primary link failures, you only need to enable BFD echo packet mode on one end of the link.

Enhancing IS-IS network security

To enhance the security of an IS-IS network, you can configure IS-IS authentication. IS-IS authentication involves neighbor relationship authentication, area authentication, and routing domain authentication.

Configuring neighbor relationship authentication

About this task

With neighbor relationship authentication configured, an interface adds the key in the specified mode into hello packets to the peer and checks the key in the received hello packets. If the authentication succeeds, it forms the neighbor relationship with the peer.

The authentication mode and key at both ends must be identical.

To prevent packet exchange failure in case of an authentication key change, configure the interface not to check the authentication information in the received packets.

Restrictions and guidelines

The authentication mode and key at both ends must be identical.

The keychain authentication mode supports HMAC-MD5 and HMAC-SM3 algorithms.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify the authentication mode and key.

   isis [**process-id** *process-id*] **authentication-mode** { { **gca** *key-id* { **hmac-sha-1** | **hmac-sha-224** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** } [**nonstandard**] | **md5** | **simple** } { **cipher** | **plain** } *string* | **keychain** *keychain-name* } [**level-1** | **level-2**] [**ip** | **osi**]

   By default, the authentication mode and key are not configured.

4. (Optional.) Configure the interface not to check the authentication information in the received hello packets.

   isis authentication send-only [**level-1** | **level-2**]

   When the authentication mode and key are configured, the interface checks the authentication information in the received packets by default.

Configuring area authentication

About this task

Area authentication prevents the router from installing routing information from untrusted routers into the Level-1 LSDB. The router encapsulates the authentication key in the specified mode in Level-1 packets (LSP, CSNP, and PSNP). It also checks the key in received Level-1 packets.

To prevent packet exchange failure in case of an authentication key change, configure IS-IS not to check the authentication information in the received packets.

Restrictions and guidelines

Routers in an area must have the same authentication mode and key.

The keychain authentication mode supports HMAC-MD5 and HMAC-SM3 algorithms.

Procedure

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Specify the area authentication mode and key.

   area-authentication-mode { { **gca** *key-id* { **hmac-sha-1** | **hmac-sha-224** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** } [**nonstandard**] | **md5** | **simple** } { **cipher** | **plain** } *string* | **keychain** *keychain-name* } [**ip** | **osi**]

   By default, the area authentication mode and key are not configured.

4. (Optional.) Configure the interface not to check the authentication information in the received Level-1 packets, including LSPs, CSNPs, and PSNPs.

   area-authentication send-only

   When the authentication mode and key are configured, the interface checks the authentication information in the received packets by default.

Configuring routing domain authentication

About this task

Routing domain authentication prevents untrusted routing information from entering into a routing domain. A router with the authentication configured encapsulates the key in the specified mode into Level-2 packets (LSP, CSNP, and PSNP) and check the key in received Level-2 packets.

To prevent packet exchange failure in case of an authentication key change, configure IS-IS not to check the authentication information in the received packets.

Restrictions and guidelines

A device configured with routing domain authentication can establish adjacencies with other routers that are not configured with routing domain authentication, but it will discard LSPs not carrying the authentication information. In this situation, IS-IS might fail to learn the expected routes. To resolve this issue, configure the same authentication mode and key on all devices in the backbone area.

The keychain authentication mode supports only HMAC-MD5 and HMAC-SM3 algorithms.

Procedure

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Specify the routing domain authentication mode and key.

   domain-authentication-mode { { **gca** *key-id* { **hmac-sha-1** | **hmac-sha-224** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** } [**nonstandard**] | **md5** | **simple** } { **cipher** | **plain** } *string* | **keychain** *keychain-name* } [**ip** | **osi**]

   By default, the routing domain authentication mode and key are not configured.

4. (Optional.) Configure the interface not to check the authentication information in the received Level-2 packets, including LSPs, CSNPs, and PSNPs.

   domain-authentication send-only

   When the authentication mode and key are configured, the interface checks the authentication information in the received packets by default.

Configuring IS-IS GR

About this task

GR ensures forwarding continuity when a routing protocol restarts or an active/standby switchover occurs.

Two routers are required to complete a GR process. The following are router roles in a GR process.

- **GR restarter**—Graceful restarting router. It must have GR capability.
- **GR helper**—A neighbor of the GR restarter. It assists the GR restarter to complete the GR process. By default, the device acts as the GR helper.

Configure IS-IS GR on the GR restarter.

GR restarter uses the following timers:

- **T1 timer**—Specifies the times that GR restarter can send a Restart TLV with the RR bit set. When rebooted, the GR restarter sends a Restart TLV with the RR bit set to its neighbor. If the GR restarter receives a Restart TLV with the RA set from its neighbor before the T1 timer expires, the GR process starts. Otherwise, the GR process fails.

- **T2 timer**—Specifies the LSDB synchronization interval. Each LSDB has a T2 timer. The Level-1-2 router has a Level-1 timer and a Level-2 timer. If the LSDBs have not synchronized before the two timers expire, the GR process fails.
- **T3 timer**—Specifies the GR interval. The GR interval is set as the holdtime in hello PDUs. Within the interval, the neighbors maintain their adjacency with the GR restarter. If the GR process has not completed within the holdtime, the neighbors tear down the neighbor relationship and the GR process fails.

Restrictions and guidelines

Follow these restrictions and guidelines when you configure IS-IS GR timers:

- The product of the T1 timer and the number of times that the T1 timer can expire must be smaller than the T2 timer.
- The T2 timer must be smaller than the T3 timer.

Procedure

1. Enter system view.
   system-view
2. Enable IS-IS and enter IS-IS view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable IS-IS GR.
   graceful-restart
   By default, the GR capability for IS-IS is disabled.
4. (Optional.) Suppress the SA bit during restart.
   graceful-restart suppress-sa
   By default, the SA bit is not suppressed.
   By enabling the GR restarter to suppress the Suppress-Advertisement (SA) bit in the hello PDUs, the neighbors will still advertise their adjacency with the GR restarter.
5. (Optional.) Configure the T1 timer.
   graceful-restart t1 *seconds* **count** *count*
   By default, the T1 timer is 3 seconds and can expire 10 times.
6. (Optional.) Configure the T2 timer.
   graceful-restart t2 *seconds*
   By default, the T2 timer is 60 seconds.
7. (Optional.) Configure the T3 timer.
   graceful-restart t3 *seconds*
   By default, the T2 timer is 300 seconds.

Configuring IS-IS NSR

About this task

After an active/standby switchover, the GR restarter obtains routing information from its neighbors, and the IS-IS process must learn all the routes. If the network topology changes during the switchover, removed routes cannot be updated to the device, which can result in blackhole routes.

Nonstop Routing (NSR) solves the problem by backing up IS-IS link state information from the active process to the standby process. After an active/standby switchover, NSR can complete link state recovery and route regeneration without requiring the cooperation of other devices.

Choose IS-IS NSR or IS-IS GR according to the actual situation and make sure data forwarding is not interrupted. The differences between IS-IS NSR and IS-IS GR as follows:

- **Device requirements**—The active and standby processes of IS-IS must run on different IRF member devices. To run IS-IS NSR, you must set up an IRF fabric that contains a minimum of two member devices. To run IS-IS GR, IRF fabric setup is not required. If an IRF fabric already exists, you can run IS-IS GR when the IRF fabric contains only one member device.
- **IS-IS neighbor requirements**—When you use IS-IS NSR, IS-IS neighbors cannot detect process restart or active/standby switchover on the local device. IS-IS NSR can recover IS-IS route information without requiring the cooperation of IS-IS neighbors. IS-IS GR requires IS-IS neighbors to have the GR helper capability to help the GR restarter complete the GR process.

Restrictions and guidelines

When both GR and NSR are configured for IS-IS on a device, the following rules apply:

- IS-IS NSR has a higher priority than IS-IS GR. When the active process goes down, the device does not act as the GR restarter and performs NSR to ensure nonstop forwarding.
- When the following conditions exist, the GR process might fail even if NSR is enabled on the GR helper:
  - IS-IS is restarted through GR.
  - An active/standby switchover is triggered on the GR helper before the GR process finishes.

Procedure

1. Enter system view.

   system-view
2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Enable IS-IS NSR.

   non-stop-routing

   By default, IS-IS NSR is enabled.

   IS-IS NSR takes effect on a per-process basis. As a best practice, enable NSR for each IS-IS process.

Configuring BFD for IS-IS

About this task

BFD provides a single mechanism to quickly detect and monitor the connectivity of links between IS-IS neighbors, reducing network convergence time. For more information about BFD, see *High Availability Configuration Guide*.

Configuring BFD for IPv4 IS-IS

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable BFD on an IPv4 IS-IS interface.

   isis bfd enable

   By default, an IPv4 IS-IS interface is not enabled with BFD.

Configuring BFD for IPv6 IS-IS

1. Enter system view.

   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Enable BFD on an IPv6 IS-IS interface.
   isis ipv6 bfd enable
   By default, an IPv6 IS-IS interface is not enabled with BFD.

Configuring IS-IS FRR

About IS-IS FRR

IS-IS Fast Reroute (FRR) calculates a backup path based on the LSDB and saves the backup path information to the FIB. When the primary path fails, the system immediately switches traffic to the backup path to prevent traffic loss and reduce the route convergence time.

IS-IS supports Loop Free Alternate (LFA) FRR.

The following IS-IS FRR traffic protection types are available:

- **Link protection**—Protects traffic that traverses a specific link.
- **Node protection**—Protects traffic that traverses a specific node.

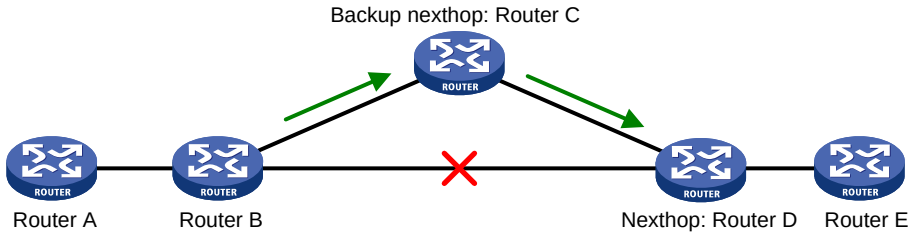
Node protection takes precedence over link protection.

Configuring IS-IS LFA FRR

About IS-IS LFA FRR

A link or router failure on a path can cause packet loss. IS-IS LFA FRR enables fast rerouting to minimize the failover time.

Figure 14 Network diagram for IS-IS LFA FRR



In Figure 14, after you enable LFA FRR on Router B, IS-IS automatically calculates or designates a backup next hop when a link failure is detected. In this way, packets are directed to the backup next hop to reduce traffic recovery time. Meanwhile, IS-IS calculates the shortest path based on the new network topology, and forwards packets over the path after network convergence.

You can assign a backup next hop for IS-IS FRR through the following ways:

- Enable IS-IS LFA FRR to calculate a backup next hop through LFA calculation.
- Designate a backup next hop with a routing policy for routes matching specific criteria.

Restrictions and guidelines

IS-IS LFA FRR and IS-IS TE are mutually exclusive.

Configuring IPv4 IS-IS LFA FRR to calculate a backup next hop through LFA calculation

1. Enter system view.
   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. (Optional.) Disable LFA calculation on the interface.
   isis fast-reroute lfa-backup exclude [**level-1** | **level-2**]
   By default, the interface participates in LFA calculation, and can be elected as a backup interface.
4. Return to system view.
   quit
5. Enter IS-IS IPv4 unicast address family view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   address-family ipv4 [**unicast**]
6. Enable IS-IS LFA FRR to calculate a backup next hop through LFA calculation.
   fast-reroute lfa [**ecmp-shared** | **level-1** | **level-2**]
   By default, IS-IS LFA FRR is disabled.

Configuring IPv4 IS-IS LFA FRR using a routing policy

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. (Optional.) Disable LFA calculation on the interface.
   isis fast-reroute lfa-backup exclude [**level-1** | **level-2**]
   By default, the interface participates in LFA calculation, and can be elected as a backup interface.
4. Return to system view.
   quit
5. Enter IS-IS IPv4 unicast address family view.
   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]
   address-family ipv4 [**unicast**]
6. Enable IPv4 IS-IS LFA FRR using a routing policy.
   - Create a routing policy and specify a backup next hop.
     apply fast-reroute backup-interface
     For more information about the apply fast-reroute backup-interface command and routing policy, see *Layer 3—IP Routing Configuration Guide*.
   - Configure IPv4 IS-IS LFA FRR.
     fast-reroute route-policy *route-policy-name*

   By default, IPv4 IS-IS LFA FRR is disabled.

Enabling BFD for IPv4 IS-IS LFA FRR

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable BFD control packet mode for IPv4 IS-IS FRR.
   isis primary-path-detect bfd ctrl

By default, BFD control packet mode is disabled for IPv4 IS-IS FRR.

To use BFD (control packet mode) to detect primary link failures, you must enable BFD control packet mode on both ends of the link.

Enabling BFD echo packet mode for IPv4 IS-IS LFA FRR

1. Enter system view.

   system-view

2. Configure the source IP address of BFD echo packets.

   bfd echo-source-ip *ip-address*

   By default, the source IP address of BFD echo packets is not configured.

   The source IP address cannot be on the same network segment as any local interface's IP address.

   For more information, see *High Availability Command Reference*.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable BFD echo packet mode for IPv4 IS-IS LFA FRR.

   isis primary-path-detect bfd echo

   By default, BFD echo packet mode is disabled for IPv4 IS-IS LFA FRR.

   To use BFD (echo packet mode) to detect primary link failures, you only need to enable BFD echo packet mode on one end of the link.

Configuring IPv6 IS-IS LFA FRR to calculate a backup next hop through LFA calculation

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. (Optional.) Disable LFA calculation on the interface.

   isis ipv6 fast-reroute lfa-backup exclude [**level-1** | **level-2**]

   By default, the interface participates in LFA calculation, and can be elected as a backup interface.

4. Return to system view.

   quit

5. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

6. Enter IS-IS IPv6 address family view.

   address-family ipv6 [**unicast**]

7. Enable IPv6 IS-IS LFA FRR to calculate a backup next hop through LFA calculation.

   fast-reroute lfa [**level-1** | **level-2**]

   By default, IPv6 IS-IS LFA FRR is disabled.

Configuring IPv6 IS-IS LFA FRR using a routing policy

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. (Optional.) Disable LFA calculation on the interface.

   isis ipv6 fast-reroute lfa-backup exclude [**level-1** | **level-2**]

By default, the interface participates in LFA calculation, and can be elected as a backup interface.

4. Return to system view.

   quit

5. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

6. Enter IS-IS IPv6 address family view.

   address-family ipv6 [**unicast**]

7. Enable IPv6 IS-IS LFA FRR using a routing policy.

   - Create a routing policy and specify a backup next hop.

     apply ipv6 fast-reroute backup-interface

     For more information about the **apply ipv6 fast-reroute backup-interface** command and routing policy, see *Layer 3—IP Routing Configuration Guide*.

   - Configure IPv6 IS-IS LFA FRR.

     fast-reroute route-policy *route-policy-name*

   By default, IPv6 IS-IS LFA FRR is disabled.

Enabling BFD for IPv6 IS-IS LFA FRR

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable BFD control packet mode for IPv6 IS-IS LFA FRR.

   isis ipv6 primary-path-detect bfd ctrl

   By default, BFD control packet mode is disabled for IPv6 IS-IS LFA FRR.

   To use BFD (control packet mode) to detect primary link failures, you must enable BFD control packet mode on both ends of the link.

Enabling BFD echo packet mode for IPv6 IS-IS LFA FRR

1. Enter system view.

   system-view

2. Configure the source IPv6 address of BFD echo packets.

   bfd echo-source-ipv6 *ip-address*

   By default, the source IPv6 address of BFD echo packets is not configured.

   The source IPv6 address cannot be on the same network segment as any local interface's IPv6 address.

   For more information, see *High Availability Command Reference*.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable BFD echo packet mode for IPv6 IS-IS LFA FRR.

   isis ipv6 primary-path-detect bfd echo

   By default, BFD echo packet mode is disabled for IPv6 IS-IS LFA FRR.

   To use BFD (echo packet mode) to detect primary link failures, you only need to enable BFD echo packet mode on one end of the link.

Setting the priority for FRR backup path selection policies

About this task

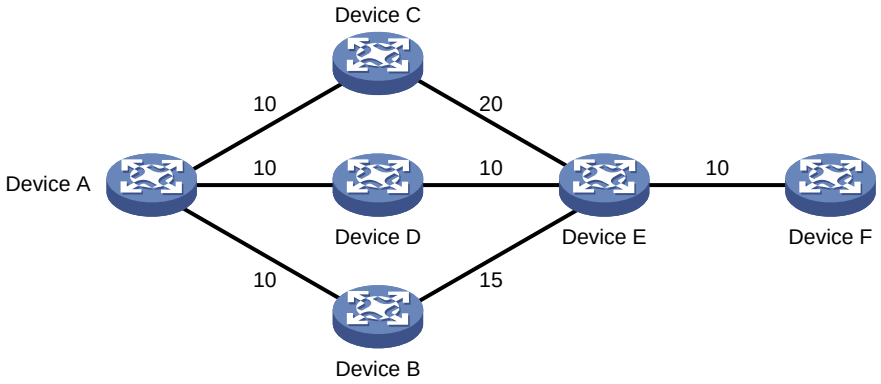
IS-IS FRR selects a backup path based on the following policies:

- **Node-protection policy**—Selects the path that involves a specific node.
- **Lowest-cost policy**—Selects the path that has the lowest cost.

As shown in Figure 15, traffic forwarding path **Device A->Device D->Device E->Device F** has two backup paths. Configure this feature as follows to control backup path selection:

- To select path **Device A->Device C->Device E->Device F** that involves Device C, set a higher priority for the node-protection policy than the lowest-cost policy.
- To select path **Device A->Device B->Device E->Device F** that has the lowest cost, set a higher priority for the lowest-cost policy than the node-protection policy.

Figure 15 IS-IS FRR backup path selection



Restrictions and guidelines

If the node-protection policy has a higher priority but the backup path calculation fails, IS-IS uses the lowest-cost policy for further calculation.

If the lowest-cost policy has a higher priority but the backup path calculation fails, IS-IS does not perform further backup path calculation.

Procedure

1. Enter system view.

   system-view

2. Enter IS-IS view.

   isis [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enter IS-IS IPv4 address family view or IS-IS IPv6 address family view.
   - Enter IS-IS IPv4 address family view.

     address-family ipv4 [**unicast**]
   - Enter IS-IS IPv6 address family view.

     address-family ipv6 [**unicast**]

4. Set the priority for the node-protection or lowest-cost backup path selection policy.

   fast-reroute tiebreaker { **lowest-cost** | **node-protecting** | **non-ecmp** | **srlg-disjoint** } **preference** *preference* [**level-1** | **level-2**]

   By default, the priority values of the node-protection and lowest-cost backup path selection policies are 40 and 20, respectively.

Configuring IS-IS neighbor flapping suppression

About this task

When the state of an IS-IS neighbor flaps frequently, IS-IS also frequently performs neighbor relationship re-establishment, LSDB synchronization, and route calculation. Such tasks interrupt IS-IS services and other services that depend on IS-IS, because performing these tasks requires a large number of packet exchanges. To resolve this issue, configure IS-IS neighbor flapping suppression. This feature enables IS-IS to suppress neighbor flapping on an interface by increasing the cost value for the interface. Traffic then does not pass through the interface during neighbor flapping suppression. When the resume timer for neighbor flapping suppression expires, IS-IS restores the cost value to the original for the interface.

IS-IS supports the conservative and non-conservative modes for neighbor flapping suppression.

The conservative mode functions as follows:

1. After you enable neighbor flapping suppression on an interface, the interface starts a flapping event counter to collect statistics about neighbor flapping events. When the time gap between two neighbor down events in a row is shorter than the detection timer, IS-IS increases the number of neighbor flapping events by one.

   IS-IS resets and restarts the flapping event counter when the following conditions exist:

   - The number of neighbor flapping events is less than the threshold that triggers neighbor flapping suppression.
   - The time gap between two neighbor down events in a row is longer than the resume timer.
2. IS-IS compares the number of neighbor flapping events with the threshold that triggers neighbor flapping suppression:
   - If the number of neighbor flapping events is less than the threshold, IS-IS does not suppress neighbor flapping.
   - If the number of neighbor flapping events is not less than the threshold, IS-IS starts neighbor flapping suppression. It increases the cost value for the interface and resets the flapping event counter to zero.
3. The resume timer starts as IS-IS starts neighbor flapping suppression. The suppression duration equals the resume timer.

   If a neighbor down event occurs on the interface during neighbor flapping suppression, IS-IS restarts the resume timer.
4. After the resume timer expires, IS-IS restores the cost value to the original for the interface if no neighbor flapping events occur on the interface before the resume timer expires.

The non-conservative mode functions as follows:

5. After you enable neighbor flapping suppression on an interface, the interface starts a flapping event counter to collect statistics about neighbor flapping events. IS-IS increases the number of neighbor flapping events by one each time a neighbor down event occurs on the interface.

   When a neighbor down event occurs on the interface, IS-IS performs the following tasks:

   a. Record the timestamp of the neighbor down event.

   NOTE:

   For brief description, the following arguments are introduced:

   - The *T* argument represents the timestamp that equals the neighbor down event timestamp minus the detection timer.
   - The *flapping-count* argument represents the number of neighbor flapping events that occur later than *T* and not later than the neighbor down event timestamp.

   b. Compare *flapping-count* with the threshold that triggers neighbor flapping suppression:

- If *flapping-count* is less than the threshold, IS-IS does not suppress neighbor flapping or reset the flapping event counter.
- If *flapping-count* is not less than the threshold, IS-IS starts neighbor flapping suppression. It increases the cost value for the interface and resets the flapping event counter.

6. The resume timer starts as IS-IS starts neighbor flapping suppression. The suppression duration equals the resume timer.

   If a neighbor down event occurs on the interface during neighbor flapping suppression, IS-IS restarts the resume timer.
7. After the resume timer expires, IS-IS restores the cost value to the original for the interface if no neighbor flapping events occur on the interface before the resume timer expires.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable neighbor flapping suppression on the interface and specify the cost adjustment value for the interface.
   - IPv4 IS-IS:

     isis peer suppress-flapping adjust-cost { *cost-offset* | **max** }

     By default, neighbor flapping suppression is disabled on an IPv4 IS-IS interface.
   - IPv6 IS-IS:

     isis ipv6 peer suppress-flapping adjust-cost { *cost-offset* | **max** }

     By default, neighbor flapping suppression is disabled on an IPv6 IS-IS interface.
4. Configure the detection parameters for IS-IS neighbor flapping suppression.
   - IPv4 IS-IS:

     isis peer suppress-flapping { [**detect-interval** *detect-interval* | **threshold** *threshold* | **resume-interval** *resume-interval*] * | **non-conservative** }

     By default, the detection timer is 60 seconds, the threshold is 10, and the resume timer is 120 seconds.
   - IPv6 IS-IS:

     isis ipv6 peer suppress-flapping { [**detect-interval** *detect-interval* | **threshold** *threshold* | **resume-interval** *resume-interval*] * | **non-conservative** }

     By default, the detection timer is 60 seconds, the threshold is 10, and the resume timer is 120 seconds.

Configuring IS-IS member port flapping suppression

About this task

When two IS-IS neighbors are connected by Layer 3 aggregate links and the links between member ports flap frequently, the neighbors perform link switchover frequently for traffic forwarding. To resolve this issue, enable member port flapping suppression on the aggregate interfaces.

This feature enables IS-IS to suppress member port flapping on an aggregate interface by increasing the cost value for the aggregate interface. Traffic then does not pass through the aggregate interface during member port flapping suppression. When the resume timer for member port flapping suppression expires, IS-IS restores the cost value to the original for the aggregate interface.

IS-IS supports the conservative and non-conservative modes for member port flapping suppression.

The conservative mode functions as follows:

1. After you enable member port flapping suppression on an aggregate interface, the aggregate interface starts a flapping event counter to collect statistics about the flapping events that occur on its member ports. When the time gap between two link failures in a row is shorter than the detection timer, IS-IS increases the number of member port flapping events by one.

   IS-IS resets and restarts the flapping event counter when the following conditions exist:

   - The number of member port flapping events is less than the threshold that triggers member port flapping suppression.
   - The time gap between two link failures in a row is longer than the resume timer.

2. IS-IS compares the number of member port flapping events with the threshold that triggers member port flapping suppression:

   - If the number of member port flapping events is less than the threshold, IS-IS does not suppress member port flapping.
   - If the number of member port flapping events is not less than the threshold, IS-IS starts member port flapping suppression. It increases the cost value for the aggregate interface and resets the flapping event counter to zero.

3. The resume timer starts as IS-IS starts member port flapping suppression. The suppression duration equals the resume timer.

   If a link failure occurs on a member port during member port flapping suppression, IS-IS restarts the resume timer.

4. After the resume timer expires, IS-IS restores the cost value to the original for the aggregate interface if no link failures occur on member ports before the resume timer expires.

The non-conservative mode functions as follows:

5. After you enable member port flapping suppression on an aggregate interface, the aggregate interface starts a flapping event counter to collect statistics about the flapping events that occur on its member ports. IS-IS increases the number of member port flapping events by one each time a link failure occurs on a member port.

   When a link failure occurs on a member port, IS-IS performs the following tasks:

   a. Record the timestamp of the link failure.

   NOTE:

   For brief description, the following arguments are introduced:

   - The *T* argument represents the timestamp that equals the link failure timestamp minus the detection timer.
   - The *flapping-count* argument represents the number of member port flapping events that occur later than *T* and not later than the link failure timestamp.

   b. Compare *flapping-count* with the threshold that triggers member port flapping suppression:

      - If *flapping-count* is less than the threshold, IS-IS does not suppress member port flapping or reset the flapping event counter.
      - If *flapping-count* is not less than the threshold, IS-IS starts member port flapping suppression. It increases the cost value for the aggregate interface and resets the flapping event counter to zero.

6. The resume timer starts as IS-IS starts member port flapping suppression. The suppression duration equals the resume timer.

If a link failure occurs on a member port during member port flapping suppression, IS-IS restarts the resume timer.

7. After the resume timer expires, IS-IS restores the cost value to the original for the aggregate interface if no link failures occur on member ports before the resume timer expires.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number* [**pex**]
   - Enter Layer 3 aggregate subinterface view.

     interface route-aggregation *interface-number.subnumber*

3. Enable member port flapping suppression on the interface and specify the cost adjustment value for the interface.
   - IPv4 IS-IS:

     isis member-port suppress-flapping adjust-cost { *cost-offset* | **max** }

     By default, member port flapping suppression is disabled on an IPv4 IS-IS aggregate interface.
   - IPv6 IS-IS:

     isis ipv6 member-port suppress-flapping adjust-cost { *cost-offset* | **max** }

     By default, member port flapping suppression is disabled on an IPv6 IS-IS aggregate interface.

4. Configure the detection parameters for IS-IS member port flapping suppression.
   - IPv4 IS-IS:

     isis member-port suppress-flapping { [**detect-interval** *detect-interval* | **threshold** *threshold* | **resume-interval** *resume-interval*] * | **non-conservative** }

     By default, the detection timer is 60 seconds, the threshold is 10, and the resume timer is 120 seconds.
   - IPv6 IS-IS:

     isis ipv6 member-port suppress-flapping { [**detect-interval** *detect-interval* | **threshold** *threshold* | **resume-interval** *resume-interval*] * | **non-conservative** }

     By default, the detection timer is 60 seconds, the threshold is 10, and the resume timer is 120 seconds.

Display and maintenance commands for IS-IS

Execute **display** commands in any view and the **reset** command in user view.

Displaying and maintaining IPv4 IS-IS

Task	Command
Display IS-IS process information.	**display isis** [*process-id*]
Display IS-IS GR log information.	**display isis event-log graceful-restart slot** *slot-number*
Display IS-IS log information about received or sent hello packets.	**display isis event-log hello** { **peer-change** | **received-abnormal** | **received-dropped** | **sent-abnormal** | **sent-failed** } [*process-id*]
Display IS-IS LSP log information.	**display isis event-log lsp** { **purged** [**generated** | **received**] | **refreshed** } [**level-1** | **level-2**] [*process-id*] [**from** *time1 date1* [**to** *time2 date2*]]
Display IS-IS NSR log information.	**display isis event-log non-stop-routing slot** *slot-number*
Display neighbor state change log information.	**display isis event-log peer** [*process-id*]
Display log information about route changes in the IP routing table.	**display isis event-log route-record** { **add** | **delete** | **modify** } [**ipv4** | **ipv6**] [**level-1** | **level-2**] [*process-id*] [**verbose**]
Display IPv4 IS-IS route calculation log information.	**display isis event-log spf** [**ipv4**] [[**level-1** | **level-2**] | **verbose**] * [*process-id*]
Display IS-IS LSP update log information.	**display isis event-log updated-lsp** [**history** [**hour-range** *hour-range1* **to** *hour-range2*]] [**system-id** *system-id*] [**level-1** | **level-2**] [*process-id*]
Display the IS-IS GR status.	**display isis graceful-restart status** [**level-1** | **level-2**] [*process-id*]
Display IS-IS interface information.	**display isis interface** [[*interface-type interface-number*] [**verbose**] | **statistics**] [*process-id*]
Display information about hello packets sent by an interface.	**display isis interface** [*interface-type interface-number*] **hello-sent** [*process-id*]
Display IS-IS LSDB information.	**display isis lsdb** [[**level-1** | **level-2**] | **local** | **lsp-id** *lspid* | [**lsp-name** *lspname*] | **verbose**] * [*process-id*]
Display IS-IS LSDB statistics.	**display isis lsdb statistics** [**level-1** | **level-2**] [*process-id*]
Display IS-IS mesh group information.	**display isis mesh-group** [*process-id*]
Display the host name to system ID mapping table.	**display isis name-table** [*process-id*]
Display the IS-IS NSR status.	**display isis non-stop-routing status**
Display IS-IS packet statistics.	**display isis packet** { **csnp** | **hello** | **lsp** | **psnp** } [**verbose**] [*interface-type interface-number*]

Task	Command
	[*process-id*]
Display IS-IS neighbor information.	**display isis peer** [**statistics** | **verbose**] [*process-id*]
Display information about hello packets received from neighbors.	**display isis peer hello-received** [*process-id*]
Display IPv4 IS-IS redistributed route information.	**display isis redistribute** [**ipv4** [*ip-address mask-length*]] [**level-1** | **level-2**] [*process-id*]
Display IS-IS IPv4 routing information.	**display isis route** [**ipv4** [*ip-address mask-length*]] [[**level-1** | **level-2**] | **verbose**] * [*process-id*]
Display IS-IS IPv4 topology information.	**display isis spf-tree** [**ipv4**] [[**level-1** | **level-2**] | [**source-id** *source-id* | **verbose**]] * [*process-id*]
Display IPv4 IS-IS statistics.	**display isis statistics** [**ipv4**] [**level-1** | **level-1-2** | **level-2**] [*process-id*]
Display IS-IS neighbor relationship troubleshooting information.	**display isis troubleshooting**
Display OSI connection information.	**display osi** [**slot** *slot-number*]
Display OSI connection statistics.	**display osi statistics** [**slot** *slot-number*]
Clear IS-IS process data structure information.	**reset isis all** [*process-id*] [**graceful-restart**]
Clear IS-IS log information.	**reset isis event-log** { **hello** { **peer-change** | **received-abnormal** | **received-dropped** | **sent-abnormal** | **sent-failed** } | **peer** } [*process-id*]
Clear IS-IS GR log information.	**reset isis event-log graceful-restart slot** *slot-number*
Clear IS-IS LSP log information.	**reset isis event-log lsp** { **purged** [**generated** | **received**] | **refreshed** } [*process-id*]
Clear IS-IS NSR log information.	**reset isis event-log non-stop-routing slot** *slot-number*
Clear IS-IS route change log messages.	**reset isis event-log route-record** { **add** | **delete** | **modify** }
Clear IS-IS route calculation log information.	**reset isis event-log spf** [*process-id*]
Clear IS-IS packet statistics.	**reset isis packet** [**csnp** | **hello** | **lsp** | **psnp**] **by-interface** [*interface-type interface-number*] [*process-id*]
Clear the data structure information of an IS-IS neighbor.	**reset isis peer** *system-id* [*process-id*]
Clear IS-IS neighbor relationship troubleshooting information.	**reset isis troubleshooting**

Task	Command
Clear OSI connection statistics.	**reset osi statistics**

Displaying and maintaining IPv6 IS-IS

Task	Command
Display IS-IS process information.	**display isis** [*process-id*]
Display IS-IS log information about received or sent hello packets.	**display isis event-log hello** { **peer-change** | **received-abnormal** | **received-dropped** | **sent-abnormal** | **sent-failed** } [*process-id*]
Display IS-IS LSP log information.	**display isis event-log lsp** { **purged** [**generated** | **received**] | **refreshed** } [**level-1** | **level-2**] [*process-id*] [**from** *time1 date1* [**to** *time2 date2*]]
Display neighbor state change log information.	**display isis event-log peer** [*process-id*]
Display log information about route changes in the IP routing table.	**display isis event-log route-record** { **add** | **delete** | **modify** } [**ipv4** | **ipv6**] [**level-1** | **level-2**] [*process-id*] [**verbose**]
Display IPv6 IS-IS route calculation log information.	**display isis event-log spf ipv6** [[**level-1** | **level-2**] | **verbose**] * [*process-id*]
Display IS-IS LSP update log information.	**display isis event-log updated-lsp** [**history** [**hour-range** *hour-range1* **to** *hour-range2*]] [**system-id** *system-id*] [**level-1** | **level-2**] [*process-id*]
Display IS-IS interface information.	**display isis interface** [[*interface-type interface-number*] [**verbose**] | **statistics**] [*process-id*]
Display information about hello packets sent by an interface.	**display isis interface** [*interface-type interface-number*] **hello-sent** [*process-id*]
Display IS-IS LSDB information.	**display isis lsdb** [[**level-1** | **level-2**] | **local** | **lsp-id** *lspid* | [**lsp-name** *lspname*] | **verbose**] * [*process-id*]
Display IS-IS LSDB statistics.	**display isis lsdb statistics** [**level-1** | **level-2**] [*process-id*]
Display IS-IS mesh group information.	**display isis mesh-group** [*process-id*]
Display the host name to system ID mapping table.	**display isis name-table** [*process-id*]
Display IS-IS packet statistics.	**display isis packet** { **csnp** | **hello** | **lsp** | **psnp** } [**verbose**] [*interface-type interface-number*] [*process-id*]
Display IS-IS neighbor information.	**display isis peer** [**statistics** | **verbose**] [*process-id*]

Task	Command
Display information about hello packets received from neighbors.	**display isis peer hello-received** [*process-id*]
Display IPv6 IS-IS redistributed route information.	**display isis redistribute ipv6** [*ipv6-address mask-length*] [**level-1** | **level-2**] [*process-id*]
Display IPv6 IS-IS routing information.	**display isis route ipv6** [*ipv6-address*] [[**level-1** | **level-2**] | **verbose**] * [*process-id*]
Display IPv6 IS-IS topology information.	**display isis spf-tree ipv6** [[**level-1** | **level-2**] [**source-id** *source-id* | | **verbose**]] * [*process-id*]
Display IPv6 IS-IS statistics.	**display isis statistics ipv6** [**level-1** | **level-1-2** | **level-2**] [*process-id*]
Display IS-IS neighbor relationship troubleshooting information.	**display isis troubleshooting**
Display OSI connection information.	**display osi** [**slot** *slot-number*]
Display OSI connection statistics.	**display osi statistics** [**slot** *slot-number*]
Clear IS-IS process data structure information.	**reset isis all** [*process-id*] [**graceful-restart**]
Clear IS-IS log information.	**reset isis event-log** { **hello** { **peer-change** | **received-abnormal** | **received-dropped** | **sent-abnormal** | **sent-failed** } | **peer** } [*process-id*]
Clear IS-IS LSP log information.	**reset isis event-log lsp** { **purged** [**generated** | **received**] | **refreshed** } [*process-id*]
Clear IS-IS route change log messages.	**reset isis event-log route-record** { **add** | **delete** | **modify** }
Clear IS-IS route calculation log information.	**reset isis event-log spf** [*process-id*]
Clear IS-IS packet statistics.	**reset isis packet** [**csnp** | **hello** | **lsp** | **psnp**] **by-interface** [*interface-type interface-number*] [*process-id*]
Clear the data structure information of an IS-IS neighbor.	**reset isis peer** *system-id* [*process-id*]
Clear IS-IS neighbor relationship troubleshooting information.	**reset isis troubleshooting**
Clear OSI connection statistics.	**reset osi statistics**

IS-IS configuration examples

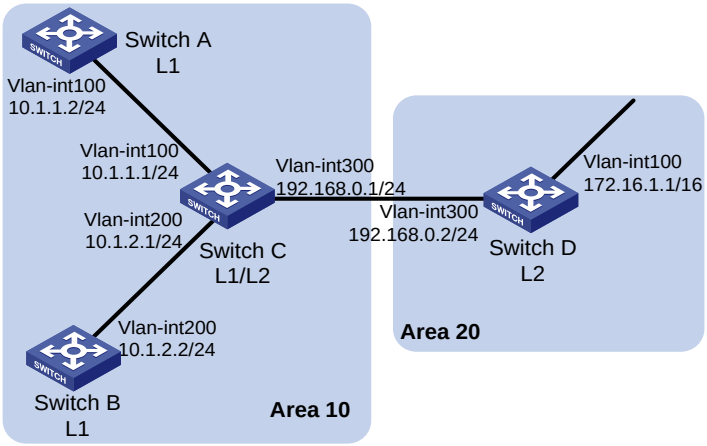
Example: Configuring basic IS-IS

Network configuration

As shown in Figure 16, Switch A, Switch B, Switch C, and Switch D reside in an IS-IS AS.

Switch A and B are Level-1 switches, Switch D is a Level-2 switch, and Switch C is a Level-1-2 switch. Switch A, Switch B, and Switch C are in Area 10, and Switch D is in Area 20.

Figure 16 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure IS-IS:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
```

```
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] isis enable 1
[SwitchD-Vlan-interface100] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis enable 1
[SwitchD-Vlan-interface300] quit
```

Verifying the configuration

Display the IS-IS LSDB on each switch to verify the LSPs.

```
[SwitchA] display isis lsdb

                      Database information for IS-IS(1)
                      ---------------------------------

                         Level-1 Link State Database
                         ---------------------------

LSPID                 Seq Num      Checksum   Holdtime   Length  ATT/P/OL
-------------------------------------------------------------------------------
0000.0000.0001.00-00* 0x00000004   0xdf5e     1096       68      0/0/0
0000.0000.0002.00-00  0x00000004   0xee4d     1102       68      0/0/0
0000.0000.0002.01-00  0x00000001   0xdaaf     1102       55      0/0/0
0000.0000.0003.00-00  0x00000009   0xcaa3     1161       111     1/0/0
0000.0000.0003.01-00  0x00000001   0xadda     1112       55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload
[SwitchB] display isis lsdb

                      Database information for IS-IS(1)
                      ---------------------------------

                         Level-1 Link State Database
```

```
                         --------------------------

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------------
0000.0000.0001.00-00  0x00000006   0xdb60     988         68      0/0/0
0000.0000.0002.00-00* 0x00000008   0xe651     1189        68      0/0/0
0000.0000.0002.01-00* 0x00000005   0xd2b3     1188        55      0/0/0
0000.0000.0003.00-00  0x00000014   0x194a     1190        111     1/0/0
0000.0000.0003.01-00  0x00000002   0xabdb     995         55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload
[SwitchC] display isis lsdb

                      Database information for IS-IS(1)
                      ---------------------------------

                         Level-1 Link State Database
                         --------------------------

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------------
0000.0000.0001.00-00  0x00000006   0xdb60     847         68      0/0/0
0000.0000.0002.00-00  0x00000008   0xe651     1053        68      0/0/0
0000.0000.0002.01-00  0x00000005   0xd2b3     1052        55      0/0/0
0000.0000.0003.00-00* 0x00000014   0x194a     1051        111     1/0/0
0000.0000.0003.01-00* 0x00000002   0xabdb     854         55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

                         Level-2 Link State Database
                         --------------------------

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------------
0000.0000.0003.00-00* 0x00000012   0xc93c     842         100     0/0/0
0000.0000.0004.00-00  0x00000026   0x331      1173        84      0/0/0
0000.0000.0004.01-00  0x00000001   0xee95     668         55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload
[SwitchD] display isis lsdb

                      Database information for IS-IS(1)
                      ---------------------------------

                         Level-2 Link State Database
                         --------------------------

LSPID                 Seq Num      Checksum      Holdtime      Length  ATT/P/OL
```

```
--------------------------------------------------------------------------------
0000.0000.0003.00-00  0x00000013   0xc73d        1003          100     0/0/0
0000.0000.0004.00-00* 0x0000003c   0xd647        1194          84      0/0/0
0000.0000.0004.01-00* 0x00000002   0xec96        1007          55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload
```

Display the IS-IS routing information on each switch.

```
[SwitchA] display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
--------------------------------------------------------------------------------
 10.1.1.0/24          10         NULL    Vlan100         Direct          D/L/-
 10.1.2.0/24          20         NULL    Vlan100         10.1.1.1        R/-/-
 192.168.0.0/24       20         NULL    Vlan100         10.1.1.1        R/-/-
 0.0.0.0/0            10         NULL    Vlan100         10.1.1.1        R/-/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
[SwitchC] display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
--------------------------------------------------------------------------------
 192.168.0.0/24       10         NULL    Vlan300         Direct          D/L/-
 10.1.1.0/24          10         NULL    Vlan100         Direct          D/L/-
 10.1.2.0/24          10         NULL    Vlan200         Direct          D/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                        Level-2 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
--------------------------------------------------------------------------------
 192.168.0.0/24       10         NULL                                    D/L/-
 10.1.1.0/24          10         NULL                                    D/L/-
 10.1.2.0/24          10         NULL                                    D/L/-
 172.16.0.0/16        20         NULL    Vlan300         192.168.0.2     R/-/-
```

```
      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
[SwitchD] display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-2 IPv4 Forwarding Table
                        -----------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 192.168.0.0/24       10         NULL    Vlan300         Direct          D/L/-
 10.1.1.0/24          20         NULL    Vlan300         192.168.0.1     R/-/-
 10.1.2.0/24          20         NULL    Vlan300         192.168.0.1     R/-/-
 172.16.0.0/16        10         NULL    Vlan100         Direct          D/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

The output shows that the routing table of Level-1 switches contains a default route with the next hop as the Level-1-2 switch. The routing table of Level-2 switch contains both routing information of Level-1 and Level-2.

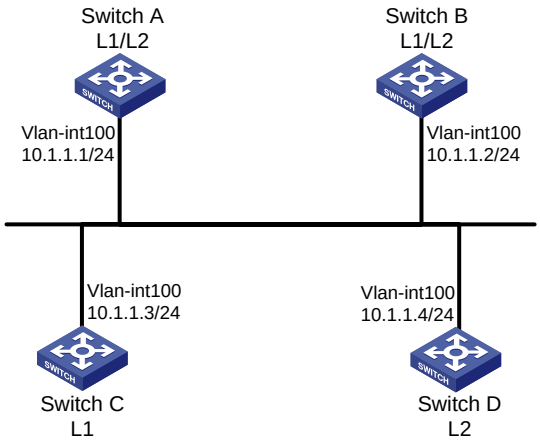
Example: Configuring DIS election

Network configuration

As shown in Figure 17, Switches A, B, C, and D reside in IS-IS area 10 on a broadcast network (Ethernet). Switch A and Switch B are Level-1-2 switches, Switch C is a Level-1 switch, and Switch D is a Level-2 switch.

Change the DIS priority of Switch A to make it elected as the Level-1-2 DIS router.

Figure 17 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Enable IS-IS:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] isis enable 1
[SwitchB-Vlan-interface100] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] is-level level-1
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] network-entity 10.0000.0000.0004.00
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] isis enable 1
[SwitchD-Vlan-interface100] quit
```

Display information about IS-IS neighbors on Switch A.

```
[SwitchA] display isis peer

                      Peer information for IS-IS(1)
                      ------------------------------

  System ID: 0000.0000.0002
 Interface: Vlan100                Circuit Id: 0000.0000.0003.01
 State: Up     HoldTime: 21s       Type: L1(L1L2)     PRI: 64

  System ID: 0000.0000.0003
 Interface: Vlan100                Circuit Id: 0000.0000.0003.01
 State: Up     HoldTime: 27s       Type: L1           PRI: 64

  System ID: 0000.0000.0002
```

```
 Interface: Vlan100                Circuit Id: 0000.0000.0004.01
 State: Up     HoldTime: 28s       Type: L2(L1L2)     PRI: 64

  System ID: 0000.0000.0004
 Interface: Vlan100                Circuit Id: 0000.0000.0004.01
 State: Up     HoldTime: 30s       Type: L2          PRI: 64
```

Display information about IS-IS interfaces on Switch A.

```
[SwitchA] display isis interface

                     Interface information for IS-IS(1)
                     ----------------------------------

  Interface:  Vlan-interface100
  Index     IPv4 state      IPv6 state     Circuit ID   MTU   Type   DIS
  00001     Up              Down           1            1497  L1/L2  No/No
```

Display information about IS-IS interfaces on Switch C.

```
[SwitchC] display isis interface

                     Interface information for IS-IS(1)
                     ----------------------------------

  Interface:  Vlan-interface100
  Index     IPv4 state      IPv6 state     Circuit ID   MTU   Type   DIS
  00001     Up              Down           1            1497  L1/L2  Yes/No
```

Display information about IS-IS interfaces on Switch D.

```
[SwitchD] display isis interface

                     Interface information for IS-IS(1)
                     ----------------------------------

  Interface:  Vlan-interface100
  Index     IPv4 state      IPv6 state     Circuit ID   MTU   Type   DIS
  00001     Up              Down           1            1497  L1/L2  No/Yes
```

The output shows that when the default DIS priority is used, Switch C is the DIS for Level-1, and Switch D is the DIS for Level-2. The pseudonodes of Level-1 and Level-2 are 0000.0000.0003.01 and 0000.0000.0004.01.

#Configure the DIS priority of Switch A.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis dis-priority 100
[SwitchA-Vlan-interface100] quit
```

Display IS-IS neighbors on Switch A.

```
[SwitchA] display isis peer

                       Peer information for IS-IS(1)
                       ------------------------------

 System ID: 0000.0000.0002
 Interface: Vlan100                Circuit Id: 0000.0000.0001.01
```

```
 State: Up     HoldTime: 21s        Type: L1(L1L2)     PRI: 64

 System ID: 0000.0000.0003
 Interface: Vlan100                 Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime: 27s        Type: L1           PRI: 64

 System ID: 0000.0000.0002
 Interface: Vlan100                 Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime: 28s        Type: L2(L1L2)     PRI: 64

 System ID: 0000.0000.0004
 Interface: Vlan100                 Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime: 30s        Type: L2           PRI: 64
```

Display information about IS-IS interfaces on Switch A.

```
[SwitchA] display isis interface

                     Interface information for IS-IS(1)
                    ----------------------------------
  Interface:  Vlan-interface100
  Index     IPv4 state      IPv6 state     Circuit ID   MTU   Type   DIS
  00001     Up              Down           1            1497  L1/L2  Yes/Yes
```

The output shows that after the DIS priority configuration, Switch A becomes the DIS for Level-1-2, and the pseudonode is 0000.0000.0001.01.

Display information about IS-IS neighbors and interfaces on Switch C.

```
[SwitchC] display isis peer

                       Peer information for IS-IS(1)
                       ------------------------------

 System ID: 0000.0000.0002
 Interface: Vlan100                 Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime: 25s        Type: L1           PRI: 64

 System ID: 0000.0000.0001
 Interface: Vlan100                 Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime: 7s         Type: L1           PRI: 100
[SwitchC] display isis interface

                     Interface information for IS-IS(1)
                     ----------------------------------

  Interface: Vlan-interface100
  Index     IPv4 state      IPv6 state     Circuit ID   MTU   Type   DIS
  00001     Up              Down           1            1497  L1/L2  No/No
```

Display information about IS-IS neighbors and interfaces on Switch D.

```
[SwitchD] display isis peer

                       Peer information for IS-IS(1)
```

```
                          ----------------------------

 System ID: 0000.0000.0001
 Interface: Vlan100                Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime: 9s        Type: L2           PRI: 100

 System ID: 0000.0000.0002
 Interface: Vlan100                Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime: 28s       Type: L2           PRI: 64
[SwitchD] display isis interface

                     Interface information for IS-IS(1)
                     ----------------------------------

  Interface:  Vlan-interface100
  Index     IPv4 state      IPv6 state     Circuit ID   MTU   Type   DIS
  00001     Up              Down           1            1497  L1/L2  No/No
```

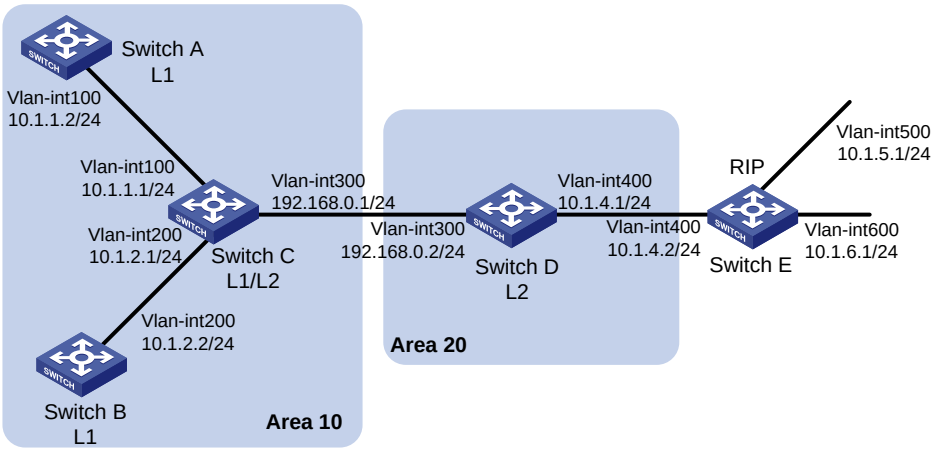
Example: Configuring IS-IS route redistribution

Network configuration

As shown in Figure 18, Switch A, Switch B, Switch C, and Switch D reside in the same AS. They use IS-IS to interconnect. Switch A and Switch B are Level-1 routers, Switch D is a Level-2 router, and Switch C is a Level-1-2 router.

Redistribute RIP routes into IS-IS on Switch D.

Figure 18 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure basic IS-IS:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
```

```
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis enable 1
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 400
[SwitchD-Vlan-interface400] isis enable 1
[SwitchD-Vlan-interface400] quit
```

Display IS-IS routing information on each switch.

```
[SwitchA] display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        -----------------------------
```

```
 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 10.1.1.0/24          10         NULL    VLAN100         Direct          D/L/-
 10.1.2.0/24          20         NULL    VLAN100         10.1.1.1        R/-/-
 192.168.0.0/24       20         NULL    VLAN100         10.1.1.1        R/-/-
 0.0.0.0/0            10         NULL    VLAN100         10.1.1.1        R/-/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
[SwitchC] display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 10.1.1.0/24          10         NULL    VLAN100         Direct          D/L/-
 10.1.2.0/24          10         NULL    VLAN200         Direct          D/L/-
 192.168.0.0/24       10         NULL    VLAN300         Direct          D/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                        Level-2 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 10.1.1.0/24          10         NULL                                    D/L/-
 10.1.2.0/24          10         NULL                                    D/L/-
 192.168.0.0/24       10         NULL                                    D/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
[SwitchD] display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-2 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 192.168.0.0/24       10         NULL    VLAN300         Direct          D/L/-
 10.1.1.0/24          20         NULL    VLAN300         192.168.0.1     R/-/-
 10.1.2.0/24          20         NULL    VLAN300         192.168.0.1     R/-/-
```

```
      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

3. Run RIPv2 between Switch D and Switch E, and configure IS-IS to redistribute RIP routes on Switch D:

Configure RIPv2 on Switch D.

```
[SwitchD] rip 1
[SwitchD-rip-1] network 10.0.0.0
[SwitchD-rip-1] version 2
[SwitchD-rip-1] undo summary
```

Configure RIPv2 on Switch E.

```
[SwitchE] rip 1
[SwitchE-rip-1] network 10.0.0.0
[SwitchE-rip-1] version 2
[SwitchE-rip-1] undo summary
[SwitchE-rip-1] quit
```

Configure IS-IS to redistribute RIP routes on Switch D.

```
[SwitchD-rip-1] quit
[SwitchD] isis 1
[SwitchD–isis-1] address-family ipv4
[SwitchD–isis-1-ipv4] import-route rip level-2
[SwitchD–isis-1-ipv4] quit
[SwitchD–isis-1] quit
```

Display IS-IS routing information on Switch C.

```
[SwitchC] display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
--------------------------------------------------------------------------------
 10.1.1.0/24          10         NULL    VLAN100         Direct          D/L/-
 10.1.2.0/24          10         NULL    VLAN200         Direct          D/L/-
 192.168.0.0/24       10         NULL    VLAN300         Direct          D/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                        Level-2 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
--------------------------------------------------------------------------------
 10.1.1.0/24          10         NULL                                    D/L/-
 10.1.2.0/24          10         NULL                                    D/L/-
 192.168.0.0/24       10         NULL                                    D/L/-
 10.1.4.0/24          20         NULL    VLAN300         192.168.0.2     R/L/-
 10.1.5.0/24          10         0       VLAN300         192.168.0.2     R/L/-
```

```
10.1.6.0/24          10          0        VLAN300         192.168.0.2      R/L/-

     Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

Example: Configuring IS-IS authentication

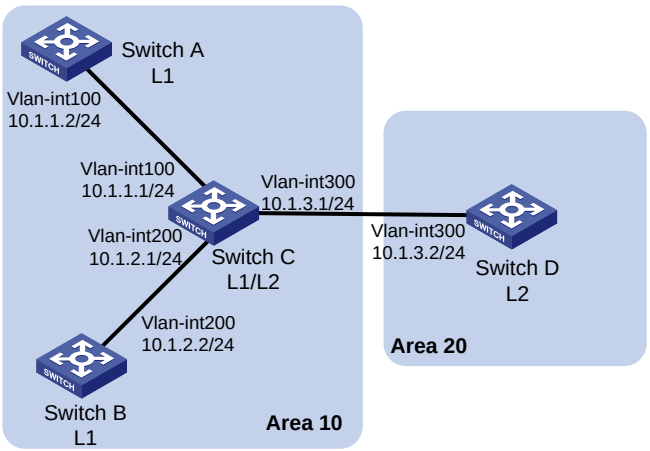
Network configuration

As shown in Figure 19, Switch A, Switch B, Switch C, and Switch D reside in the same IS-IS routing domain. Run IS-IS among them.

Switch A, Switch B, and Switch C belong to Area 10, and Switch D belongs to Area 20.

- Configure neighbor relationship authentication between neighbors.
- Configure area authentication in Area 10 to prevent untrusted routes from entering into the area.
- Configure routing domain authentication on Switch C and Switch D to prevent untrusted routes from entering the routing domain.

Figure 19 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure basic IS-IS:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
```

```
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] network-entity 20.0000.0000.0001.00
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis enable 1
[SwitchD-Vlan-interface300] quit
```

3. Configure neighbor relationship authentication between neighbors:

Set the authentication mode to MD5 and set the plaintext key to **eRq** on VLAN-interface 100 of Switch A and on VLAN-interface 100 of Switch C.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis authentication-mode md5 plain eRg
[SwitchA-Vlan-interface100] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis authentication-mode md5 plain eRg
[SwitchC-Vlan-interface100] quit
```

Set the authentication mode to MD5 and set the plaintext key to **t5Hr** on VLAN-interface 200 of Switch B and on VLAN-interface 200 of Switch C.

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis authentication-mode md5 plain t5Hr
[SwitchB-Vlan-interface200] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis authentication-mode md5 plain t5Hr
[SwitchC-Vlan-interface200] quit
```

Set the authentication mode to MD5 and set the plaintext key to **hSec** on VLAN-interface 300 of Switch D and on VLAN-interface 300 of Switch C.

```
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis authentication-mode md5 plain hSec
[SwitchC-Vlan-interface300] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis authentication-mode md5 plain hSec
```

```
[SwitchD-Vlan-interface300] quit
```

4. Set the area authentication mode to MD5 and set the plaintext key to **10Sec** on Switch A, Switch B, and Switch C.

```
[SwitchA] isis 1
[SwitchA-isis-1] area-authentication-mode md5 plain 10Sec
[SwitchA-isis-1] quit
[SwitchB] isis 1
[SwitchB-isis-1] area-authentication-mode md5 plain 10Sec
[SwitchB-isis-1] quit
[SwitchC] isis 1
[SwitchC-isis-1] area-authentication-mode md5 plain 10Sec
[SwitchC-isis-1] quit
```

5. Set routing domain authentication mode to MD5 and set the plaintext key to **1020Sec** on Switch C and Switch D.

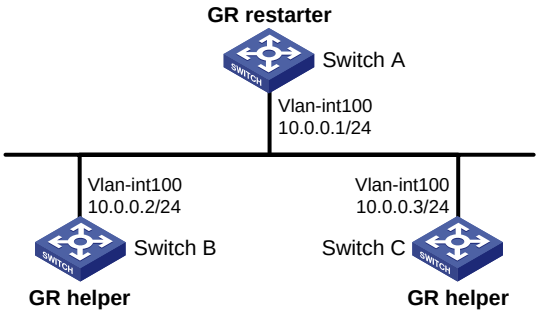
```
[SwitchC] isis 1
[SwitchC-isis-1] domain-authentication-mode md5 plain 1020Sec
[SwitchC-isis-1] quit
[SwitchD] isis 1
[SwitchD-isis-1] domain-authentication-mode md5 plain 1020Sec
```

Example: Configuring IS-IS GR

Network configuration

As shown in Figure 20, Switch A, Switch B, and Switch C belong to the same IS-IS routing domain.

Figure 20 Network diagram



Procedure

1. Configure IP addresses and subnet masks for interfaces. (Details not shown.)
2. Configure IS-IS on the switches to make sure Switch A, Switch B, and Switch C can communicate with each other at layer 3 and dynamic route update can be implemented among them with IS-IS. (Details not shown.)
3. Enable IS-IS GR on Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] graceful-restart
[SwitchA-isis-1] quit
[SwitchA] quit
```

Verifying the configuration

Restart the IS-IS process on Switch A.

```
<SwitchA> reset isis all 1 graceful-restart
Reset IS-IS process? [Y/N]:y
```

Check the GR state of the IS-IS process on Switch A.

```
<SwitchA> display isis graceful-restart status

                        Restart information for IS-IS(1)
                        --------------------------------
Restart status: COMPLETE
Restart phase: Finish
Restart t1: 3, count 10; Restart t2: 60; Restart t3: 300
SA Bit: supported

                          Level-1 restart information
                          ---------------------------
Total number of interfaces: 1
Number of waiting LSPs: 0

                          Level-2 restart information
                          ---------------------------
Total number of interfaces: 1
Number of waiting LSPs: 0
```

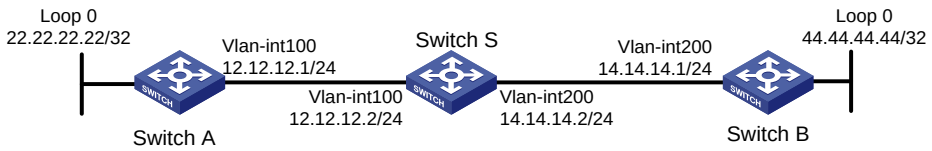
Example: Configuring IS-IS NSR

Network configuration

As shown in Figure 21, Switch S, Switch A, and Switch B belong to the same IS-IS routing domain.

- Run IS-IS on all the switches to interconnect them with each other.
- Enable IS-IS NSR on Switch S to ensure forwarding continuity between Switch A and Switch B when an active/standby switchover occurs on Switch S.

Figure 21 Network diagram



Procedure

1. Configure the IP addresses and subnet masks for interfaces on the switches. (Details not shown.)
2. Configure IS-IS on the switches to make sure Switch S, Switch A, and Switch B can communicate with each other at Layer 3 and dynamic route update can be implemented among them with IS-IS. (Details not shown.)
3. Enable IS-IS NSR on Switch S.

```
<SwitchS> system-view
[SwitchS] isis 1
```

```
[SwitchS-isis-1] non-stop-routing
[SwitchS-isis-1] quit
```

Verifying the configuration

Reoptimize process placement on Switch S to trigger an active/standby switchover.

```
[SwitchS] placement reoptimize
Predicted changes to the placement
Service Group(instance name)                    Cur location  New location
-------------------------------------------------------------------------
rib                                             0/0           0/0
staticroute                                     0/0           0/0
rib6                                            0/0           0/0
staticroute6                                    0/0           0/0
isis                                            0/0           1/0
Continue? [y/n]:y
Re-optimization of the placement start. You will be notified on completion.
Re-optimization of the placement complete. Use 'display placement' to view the new
placement.
```

During the switchover period, display IS-IS neighbor information on Switch A to verify the neighborship between Switch A and Switch S.

```
<SwitchA> display isis peer

                        Peer information for IS-IS(1)
                        ----------------------------

 System ID: 0000.0000.0001
 Interface: Vlan100                Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime:  25s      Type: L1(L1L2)     PRI: 64

 System ID: 0000.0000.0001
 Interface: Vlan100                Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime:  27s      Type: L2(L1L2)     PRI: 64
```

Display IS-IS routing information on Switch A to verify that Switch A has a route to the loopback interface of Switch B.

```
<SwitchA> display isis route

                        Route information for IS-IS(1)
                        ----------------------------

                        Level-1 IPv4 Forwarding Table
                        ----------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 12.12.12.0/24        10         NULL    vlan100         Direct          D/L/-
 22.22.22.22/32       10         NULL    Loop0           Direct          D/-/-
 14.14.14.0/32        10         NULL    vlan100         12.12.12.2      R/L/-
 44.44.44.44/32       10         NULL    vlan100         12.12.12.2      R/L/-
```

```
      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                        Level-2 IPv4 Forwarding Table
                        -----------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
--------------------------------------------------------------------------------
 12.12.12.0/24        10         NULL    vlan100         Direct          D/L/-
 22.22.22.22/32       10         NULL    Loop0           Direct          D/-/-
 14.14.14.0/32        10         NULL
 44.44.44.44/32       10         NULL

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

Display IS-IS neighbor information on Switch B to verify the neighborship between Switch B and Switch S.

```
<SwitchB> display isis peer

                        Peer information for IS-IS(1)
                        ------------------------------

 System ID: 0000.0000.0001
 Interface: Vlan200                Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime:  25s      Type: L1(L1L2)     PRI: 64

 System ID: 0000.0000.0001
 Interface: Vlan200                Circuit Id: 0000.0000.0001.01
 State: Up     HoldTime:  27s      Type: L2(L1L2)     PRI: 64
```

Display IS-IS routing information on Switch B to verify that Switch B has a route to the loopback interface of Switch A.

```
<SwitchB> display isis route

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        -----------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
--------------------------------------------------------------------------------
 14.14.14.0/24        10         NULL    vlan200         Direct          D/L/-
 44.44.44.44/32       10         NULL    Loop0           Direct          D/-/-
 12.12.12.0/32        10         NULL    vlan200         14.14.14.4      R/L/-
 22.22.22.22/32       10         NULL    vlan200         14.14.14.4      R/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                        Level-2 IPv4 Forwarding Table
                        -----------------------------
```

```
 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 14.14.14.0/24        10         NULL    vlan200         Direct          D/L/-
 44.44.44.44/32       10         NULL    Loop0           Direct          D/-/-
 12.12.12.0/32        10         NULL
 22.22.22.22/32       10         NULL

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

The output shows that the neighbor information and routing information on Switch A and Switch B have not changed during the active/standby switchover on Switch S. The neighbors are unaware of the switchover.

Example: Configuring BFD for IS-IS

Network configuration

- As shown in Figure 22, run IS-IS on Switch A, Switch B and Switch C so that they can reach each other at the network layer.
- After the link over which Switch A and Switch B communicate through the Layer-2 switch fails, BFD can quickly detect the failure and notify IS-IS of the failure. Switch A and Switch B then communicate through Switch C.

Figure 22 Network diagram

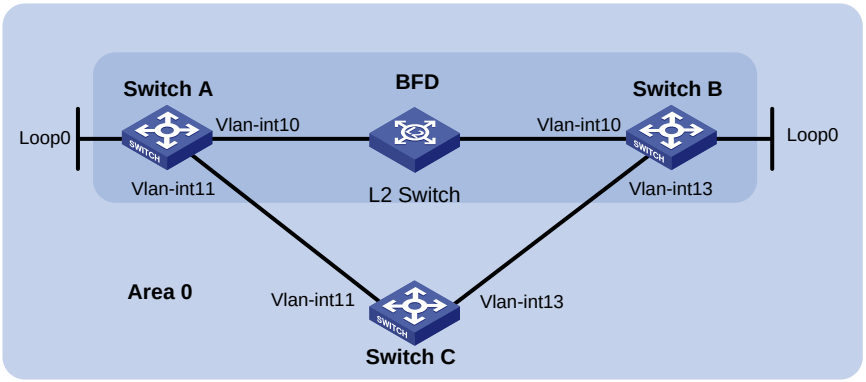


Table 4 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int10	10.1.0.102/24	Switch B	Vlan-int10	10.1.0.100/24
	Vlan-int11	11.1.1.1/24		Vlan-int13	13.1.1.1/24
	Loop0	121.1.1.1/32		Loop0	120.1.1.1/32
Switch C	Vlan-int11	11.1.1.2/24			
	Vlan-int13	13.1.1.2/24			

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure basic IS-IS:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   ```

```
[SwitchA] isis
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] isis enable
[SwitchA-LoopBack0] quit
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] isis enable
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 11
[SwitchA-Vlan-interface11] isis enable
[SwitchA-Vlan-interface11] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] isis enable
[SwitchB-LoopBack0] quit
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] isis enable
[SwitchB-Vlan-interface10] quit
[SwitchB] interface vlan-interface 13
[SwitchB-Vlan-interface13] isis enable
[SwitchB-Vlan-interface13] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 11
[SwitchC-Vlan-interface11] isis enable
[SwitchC-Vlan-interface11] quit
[SwitchC] interface vlan-interface 13
[SwitchC-Vlan-interface13] isis enable
[SwitchC-Vlan-interface13] quit
```

3. Configure BFD functions:

Enable BFD and configure BFD parameters on Switch A.

```
[SwitchA] bfd session init-mode passive
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] isis bfd enable
[SwitchA-Vlan-interface10] bfd min-receive-interval 500
[SwitchA-Vlan-interface10] bfd min-transmit-interval 500
[SwitchA-Vlan-interface10] bfd detect-multiplier 7
[SwitchA-Vlan-interface10] quit
```

Enable BFD and configure BFD parameters on Switch B.

```
[SwitchB] bfd session init-mode active
```

```
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] isis bfd enable
[SwitchB-Vlan-interface10] bfd min-receive-interval 500
[SwitchB-Vlan-interface10] bfd min-transmit-interval 500
[SwitchB-Vlan-interface10] bfd detect-multiplier 8
[SwitchB-Vlan-interface10] quit
```

Verifying the configuration

Display the BFD session information on Switch A.

```
[SwitchA] display bfd session

 Total sessions: 1     Up sessions: 1     Init mode: Active

 IPv4 session working in control packet mode:

 LD/RD            SourceAddr      DestAddr        State  Holdtime    Interface
 3/1              192.168.0.102   192.168.0.100   Up     1700ms      Vlan10
```

Display routes destined for 120.1.1.1/32 on Switch A.

```
[SwitchA] display ip routing-table 120.1.1.1 verbose

Summary count : 1

Destination: 120.1.1.1/32
   Protocol: IS_L1
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 10             Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                  State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0x2                OrigAs: 0
      NibID: 0x26000002         LastAs: 0
     AttrID: 0xffffffff        Neighbor: 0.0.0.0
      Flags: 0x1008c        OrigNextHop: 192.168.0.100
      Label: NULL           RealNextHop: 192.168.0.100
    BkLabel: NULL             BkNextHop: N/A
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface10
BkTunnel ID: Invalid        BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

The output shows that Switch A and Switch B communicate through VLAN-interface 10. Then the link over VLAN-interface 10 fails.

Display routes destined for 120.1.1.1/32 on Switch A.

```
[SwitchA] display ip routing-table 120.1.1.1 verbose
```

```
Summary count : 1

Destination: 120.1.1.1/32
   Protocol: IS_L1
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 20             Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0x2                OrigAs: 0
      NibID: 0x26000002         LastAs: 0
     AttrID: 0xffffffff       Neighbor: 0.0.0.0
      Flags: 0x1008c        OrigNextHop: 10.1.1.100
      Label: NULL           RealNextHop: 10.1.1.100
    BkLabel: NULL             BkNextHop: N/A
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface11
BkTunnel ID: Invalid        BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

The output shows that Switch A and Switch B communicate through VLAN-interface 11.

Example: Configuring IS-IS FRR

Network configuration

As shown in Figure 23, Switch A, Switch B, and Switch C belong to the same IS-IS routing domain. Configure IS-IS FRR so that when the Link A fails, traffic can be switched to Link B immediately.

Figure 23 Network diagram

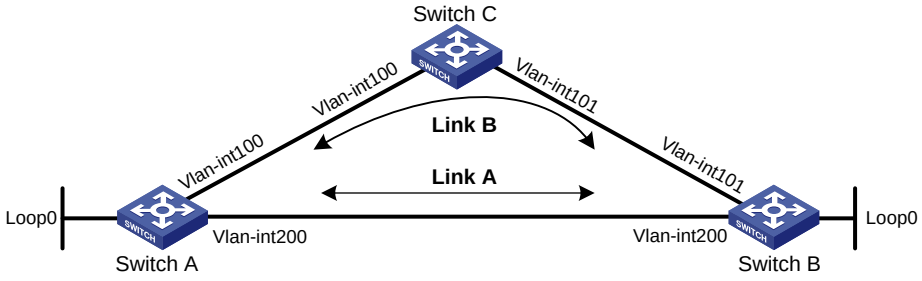


Table 5 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	12.12.12.1/24	Switch B	Vlan-int101	24.24.24.4/24
	Vlan-int200	13.13.13.1/24		Vlan-int200	13.13.13.2/24

Device	Interface	IP address	Device	Interface	IP address
	Loop0	1.1.1.1/32		Loop0	4.4.4.4/32
Switch C	Vlan-int100	12.12.12.2/24			
	Vlan-int101	24.24.24.2/24			

Procedure

1. Configure IP addresses and subnet masks for interfaces on the switches. (Details not shown.)
2. Configure IS-IS on the switches to make sure Switch A, Switch B, and Switch C can communicate with each other at Layer 3. (Details not shown.)
3. Configure IS-IS FRR:

   Enable IS-IS FRR to calculate a backup next hop through LFA calculation, or designate a backup next hop by using a referenced routing policy.

   - (Method 1.) Enable IS-IS FRR to calculate a backup next hop through LFA calculation:

     # Configure Switch A.

     ```
     <SwitchA> system-view
     [SwitchA] isis 1
     [SwitchA-isis-1] address-family ipv4
     [SwitchA-isis-1-ipv4] fast-reroute lfa
     [SwitchA-isis-1-ipv4] quit
     [SwitchA-isis-1] quit
     ```

     # Configure Switch B.

     ```
     <SwitchB> system-view
     [SwitchB] isis 1
     [SwitchB-isis-1] address-family ipv4
     [SwitchB-isis-1-ipv4] fast-reroute lfa
     [SwitchB-isis-1-ipv4] quit
     [SwitchB-isis-1] quit
     ```

   - (Method 2.) Enable IS-IS FRR to designate a backup next hop by using a referenced routing policy:

     # Configure Switch A.

     ```
     <SwitchA> system-view
     [SwitchA] ip prefix-list abc index 10 permit 4.4.4.4 32
     [SwitchA] route-policy frr permit node 10
     [SwitchA-route-policy-frr-10] if-match ip address prefix-list abc
     [SwitchA-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
     100 backup-nexthop 12.12.12.2
     [SwitchA-route-policy-frr-10] quit
     [SwitchA] isis 1
     [SwitchA-isis-1] address-family ipv4
     [SwitchA-isis-1-ipv4] fast-reroute route-policy frr
     [SwitchA-isis-1-ipv4] quit
     [SwitchA-isis-1] quit
     ```

     # Configure Switch B.

     ```
     <SwitchB> system-view
     [SwitchB] ip prefix-list abc index 10 permit 1.1.1.1 32
     [SwitchB] route-policy frr permit node 10
     [SwitchB-route-policy-frr-10] if-match ip address prefix-list abc
     ```

```
[SwitchB-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
101 backup-nexthop 24.24.24.2
[SwitchB-route-policy-frr-10] quit
[SwitchB] isis 1
[SwitchB-isis-1] address-family ipv4
[SwitchB-isis-1-ipv4] fast-reroute route-policy frr
[SwitchB-isis-1-ipv4] quit
[SwitchB-isis-1] quit
```

Verifying the configuration

Display route 4.4.4.4/32 on Switch A to view the backup next hop information.

```
[SwitchA] display ip routing-table 4.4.4.4 verbose

Summary count : 1

Destination: 4.4.4.4/32
   Protocol: IS_L1
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 10             Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0x2                OrigAs: 0
      NibID: 0x26000002         LastAs: 0
     AttrID: 0xffffffff       Neighbor: 0.0.0.0
      Flags: 0x1008c       OrigNextHop: 13.13.13.2
      Label: NULL          RealNextHop: 13.13.13.2
    BkLabel: NULL            BkNextHop: 12.12.12.2
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface200
BkTunnel ID: Invalid       BkInterface: Vlan-interface100
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0              MicroSegID: 0
RealFIRType: Normal          RealThres: 0
```

Display route 1.1.1.1/32 on Switch B to view the backup next hop information.

```
[SwitchB] display ip routing-table 1.1.1.1 verbose

Summary count : 1

Destination: 1.1.1.1/32
   Protocol: IS_L1
 Process ID: 1
  SubProtID: 0x1                  Age: 04h20m37s
       Cost: 10             Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
```

```
  OrigTblID: 0x0                OrigVrf: default-vrf
    TableID: 0x2                 OrigAs: 0
      NibID: 0x26000002          LastAs: 0
     AttrID: 0xffffffff        Neighbor: 0.0.0.0
      Flags: 0x1008c        OrigNextHop: 13.13.13.1
      Label: NULL           RealNextHop: 13.13.13.1
    BkLabel: NULL             BkNextHop: 24.24.24.2
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface200
BkTunnel ID: Invalid        BkInterface: Vlan-interface101
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0                MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Example: Configuring IS-IS multi-instance processes

Network configuration

As shown in Figure 24, the IPv4 and IPv6 costs are different on an interface. Configure IS-IS multi-instance processes to isolate the IPv4 and IPv6 network and avoid IPv6 route calculation errors.

Figure 24 Network diagram

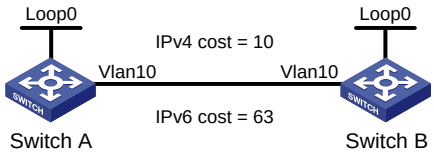


Table 6 Interface and IP address assignment

Device	Interface	IPv4 address	IPv6 address
Switch A	Vlan10	10.1.1.1/24	2001::1/64
	Loop0	1.1.1.1/32	10::1/128
Switch B	Vlan10	10.1.1.2/24	2001::2/64
	Loop0	2.2.2.2/32	20::1/128

Prerequisties

Configure IPv4 and IPv6 addresses for the interfaces. (Details not shown.)

Procedure

1. Configure traditional IPv4 IS-IS processes:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] isis 1
   [SwitchA-isis-1] network-entity 10.0000.0000.0001.00
   [SwitchA-isis-1] quit
   [SwitchA] interface loopback 0
   ```

```
[SwitchA-LoopBack0] isis enable 1
[SwitchA-LoopBack0] quit
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] isis enable 1
[SwitchA-Vlan-interface10] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] isis enable 1
[SwitchB-LoopBack0] quit
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] isis enable 1
[SwitchB-Vlan-interface10] quit
```

2. Configure IPv6 IS-IS multi-instance processes:

Configure Switch A.

```
[SwitchA] isis 2
[SwitchA-isis-2] network-entity 20.0000.0000.0010.00
[SwitchA-isis-2] multi-instance enable iid 1
[SwitchA-isis-2] address-family ipv6
[SwitchA-isis-2-ipv6] quit
[SwitchA-isis-2] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] isis ipv6 enable 2
[SwitchA-LoopBack0] quit
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] isis ipv6 enable 2
[SwitchA-Vlan-interface10] isis process-id 2 cost 63
[SwitchA-Vlan-interface10] quit
```

Configure Switch B.

```
[SwitchB] isis 2
[SwitchB-isis-2] network-entity 20.0000.0000.0020.00
[SwitchB-isis-2] multi-instance enable iid 1
[SwitchB-isis-2] address-family ipv6
[SwitchB-isis-2-ipv6] quit
[SwitchB-isis-2] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] isis ipv6 enable 2
[SwitchB-LoopBack0] quit
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] isis ipv6 enable 2
[SwitchB-Vlan-interface10] isis process-id 2 cost 63
[SwitchB-Vlan-interface10] quit
```

Verifying the configuration

View information about the IPv4 IS-IS routing table of Switch A.

```
[SwitchA] display isis route ipv4

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 1.1.1.1/32           0          NULL    Loop0           Direct          D/L/-
 10.1.1.0/24          10         NULL    Vlan10          Direct          D/L/-
 2.2.2.2/32           10         NULL    Vlan10          10.1.1.2        R/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                        Level-2 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 1.1.1.1/32           0          NULL    Loop0           Direct          D/L/-
 10.1.1.0/24          10         NULL    Vlan10          Direct          D/L/-
 2.2.2.2/32           10         NULL

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

The output shows that route calculation results are correct.

View information about the IPv4 IS-IS routing table of Switch B.

```
[SwitchB] display isis route ipv4

                        Route information for IS-IS(1)
                        ------------------------------

                        Level-1 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
-------------------------------------------------------------------------------
 1.1.1.1/32           10         NULL    Vlan10          10.1.1.1        R/L/-
 10.1.1.0/24          10         NULL    Vlan10          Direct          D/L/-
 2.2.2.2/32           0          NULL    Loop0           Direct          D/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                        Level-2 IPv4 Forwarding Table
                        ------------------------------

 IPv4 Destination     IntCost    ExtCost ExitInterface   NextHop         Flags
```

```
------------------------------------------------------------------------------
 1.1.1.1/32           10         NULL
 10.1.1.0/24          10         NULL    Vlan10          Direct          D/L/-
 2.2.2.2/32           0          NULL    Loop0           Direct          D/L/-

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

The output shows that route calculation results are correct.

View information about the IPv6 IS-IS routing table of Switch A.

```
[SwitchA] display isis route ipv6

                      Route information for IS-IS(2)
                      ------------------------------

                      Level-1 IPv6 forwarding table
                      ------------------------------

 Destination : 10::1                                 PrefixLen: 128
 Flag        : D/L/-                                 Cost     : 0
 Next hop    : Direct                                Interface: Loop0

 Destination : 2001::                                PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 63
 Next hop    : Direct                                Interface: Vlan10

 Destination : 20::1                                 PrefixLen: 128
 Flag        : R/L/-                                 Cost     : 63
 Next hop    : FE80::861F:31FF:FE6D:201              Interface: Vlan10

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                      Level-2 IPv6 forwarding table
                      ------------------------------

 Destination : 10::1                                 PrefixLen: 128
 Flag        : D/L/-                                 Cost     : 0
 Next hop    : Direct                                Interface: Loop0

 Destination : 2001::                                PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 63
 Next hop    : Direct                                Interface: Vlan10

 Destination : 20::1                                 PrefixLen: 128
 Flag        : -/-/-                                 Cost     : 63

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

The output shows that route calculation results are correct.

View information about the IPv6 IS-IS routing table of Switch B.

```
[SwitchB] display isis route ipv6

                      Route information for IS-IS(2)
                       ------------------------------

                       Level-1 IPv6 forwarding table
                       ------------------------------

 Destination : 10::1                                 PrefixLen: 128
 Flag        : R/L/-                                 Cost     : 63
 Next hop    : FE80::861F:29FF:FE93:101              Interface: Vlan10

 Destination : 2001::                                PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 63
 Next hop    : Direct                                Interface: Vlan10

 Destination : 20::1                                 PrefixLen: 128
 Flag        : D/L/-                                 Cost     : 0
 Next hop    : Direct                                Interface: Loop0

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                       Level-2 IPv6 forwarding table
                       ------------------------------

 Destination : 10::1                                 PrefixLen: 128
 Flag        : -/-/-                                 Cost     : 63

 Destination : 2001::                                PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 63
 Next hop    : Direct                                Interface: Vlan10

 Destination : 20::1                                 PrefixLen: 128
 Flag        : D/L/-                                 Cost     : 0
 Next hop    : Direct                                Interface: Loop0

      Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

The output shows that route calculation results are correct.

IPv6 IS-IS configuration examples

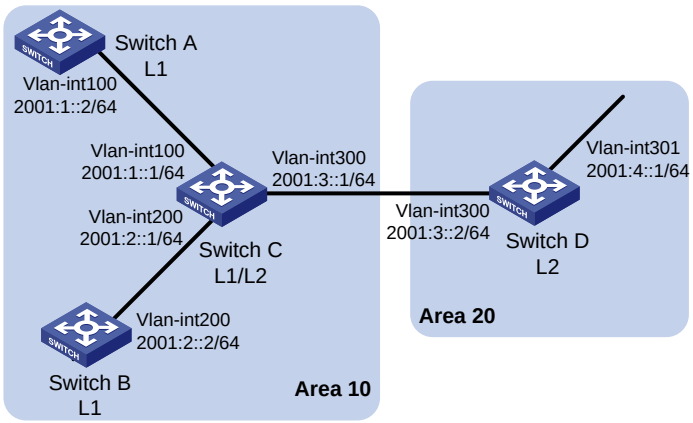
Example: Configuring IPv6 IS-IS basics

Network configuration

As shown in Figure 25, Switch A, Switch B, Switch C, and Switch D, all enabled with IPv6, reside in the same AS. Configure IPv6 IS-IS on the switches so that they can reach each other.

Switch A and Switch B are Level-1 switches, Switch D is a Level-2 switch, and Switch C is a Level-1-2 switch.

Figure 25 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure IPv6 IS-IS:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] address-family ipv6
[SwitchA-isis-1-ipv6] quit
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis ipv6 enable 1
[SwitchA-Vlan-interface100] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] address-family ipv6
[SwitchB-isis-1-ipv6] quit
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis ipv6 enable 1
[SwitchB-Vlan-interface200] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] address-family ipv6
[SwitchC-isis-1-ipv6] quit
[SwitchC-isis-1] quit
```

```
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis ipv6 enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis ipv6 enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis ipv6 enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] address-family ipv6
[SwitchD-isis-1-ipv6] quit
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis ipv6 enable 1
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 301
[SwitchD-Vlan-interface301] isis ipv6 enable 1
[SwitchD-Vlan-interface301] quit
```

Verifying the configuration

Display the IPv6 IS-IS routing table on Switch A.

```
[SwitchA] display isis route ipv6

                      Route information for IS-IS(1)
                      ------------------------------

                      Level-1 IPv6 Forwarding Table
                      ------------------------------

 Destination : ::                                  PrefixLen: 0
 Flag        : R/-/-                               Cost     : 10
 Next Hop    : FE80::200:FF:FE0F:4                 Interface: Vlan100

 Destination : 2001:1::                            PrefixLen: 64
 Flag        : D/L/-                               Cost     : 10
 Next Hop    : Direct                              Interface: Vlan100

 Destination : 2001:2::                            PrefixLen: 64
 Flag        : R/-/-                               Cost     : 20
 Next Hop    : FE80::200:FF:FE0F:4                 Interface: Vlan100

 Destination : 2001:3::                            PrefixLen: 64
 Flag        : R/-/-                               Cost     : 20
 Next Hop    : FE80::200:FF:FE0F:4                 Interface: Vlan100
```

```
     Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

Display the IPv6 IS-IS routing table on Switch B.

```
[SwitchB] display isis route ipv6

                      Route information for IS-IS(1)
                      ------------------------------

                      Level-1 IPv6 Forwarding Table
                      ------------------------------

 Destination : ::                                    PrefixLen: 0
 Flag        : R/-/-                                 Cost     : 10
 Next Hop    : FE80::200:FF:FE0F:4                   Interface: Vlan200

 Destination : 2001:1::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : FE80::200:FF:FE0F:4                   Interface: Vlan200

 Destination : 2001:2::                              PrefixLen: 64
 Flag        : R/-/-                                 Cost     : 20
 Next Hop    : Direct                                Interface: Vlan200

 Destination : 2001:3::                              PrefixLen: 64
 Flag        : R/-/-                                 Cost     : 20
 Next Hop    : FE80::200:FF:FE0F:4                   Interface: Vlan200

     Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

Display the IPv6 IS-IS routing table on Switch C.

```
[SwitchC] display isis route ipv6

                      Route information for IS-IS(1)
                      ------------------------------

                      Level-1 IPv6 Forwarding Table
                      ------------------------------

 Destination : 2001:1::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : Direct                                Interface: Vlan100

 Destination : 2001:2::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : Direct                                Interface: Vlan200

 Destination : 2001:3::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : Direct                                Interface: Vlan300
```

```
     Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set

                         Level-2 IPv6 Forwarding Table
                         -----------------------------

 Destination : 2001:1::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : Direct                                Interface: Vlan100

 Destination : 2001:2::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : Direct                                Interface: Vlan200

 Destination : 2001:3::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : Direct                                Interface: Vlan300

 Destination : 2001:4::1                             PrefixLen: 64
 Flag        : R/-/-                                 Cost     : 10
 Next Hop    : FE80::20F:E2FF:FE3E:FA3D              Interface: Vlan300

     Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

Display the IPv6 IS-IS routing table on Switch D.

```
[SwitchD] display isis route ipv6

                         Route information for IS-IS(1)
                         ------------------------------

                         Level-2 IPv6 Forwarding Table
                         -----------------------------

 Destination : 2001:1::                              PrefixLen: 64
 Flag        : R/-/-                                 Cost     : 20
 Next Hop    : FE80::200:FF:FE0F:4                   Interface: Vlan300

 Destination : 2001:2::                              PrefixLen: 64
 Flag        : R/-/-                                 Cost     : 20
 Next Hop    : FE80::200:FF:FE0F:4                   Interface: Vlan300

 Destination : 2001:3::                              PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 10
 Next Hop    : Direct                                Interface: Vlan300

 Destination : 2001:4::1                             PrefixLen: 64
 Flag        : D/L/-                                 Cost     : 0
 Next Hop    : Direct                                Interface: Loop1
```

```
Flags: D-Direct, R-Added to Rib, L-Advertised in LSPs, U-Up/Down Bit Set
```

Example: Configuring BFD for IPv6 IS-IS

Network configuration

As shown in Figure 26:

- Configure IPv6 IS-IS on Switch A and Switch B so that they can reach other.
- Enable BFD on VLAN-interface 10 of Switch A and Switch B.

After the link between Switch B and the Layer 2 switch fails, BFD can quickly detect the failure and notify IPv6 IS-IS of the failure. Then Switch A and Switch B communicate through Switch C.

Figure 26 Network diagram

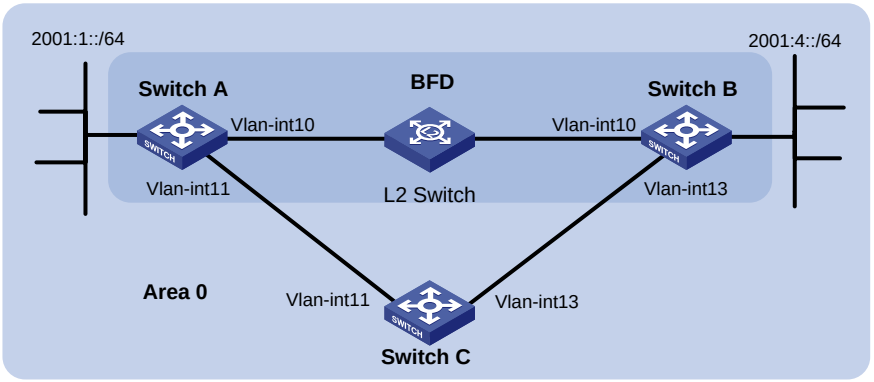


Table 7 Interface and IP address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Switch A	Vlan-int10	2001::1/64	Switch B	Vlan-int10	2001::2/64
	Vlan-int11	2001:2::1/64		Vlan-int13	2001:3::2/64
Switch C	Vlan-int11	2001:2::2/64			
	Vlan-int13	2001:3::1/64			

Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure IPv6 IS-IS:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] address-family ipv6
[SwitchA-isis-1-ipv6] quit
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] isis ipv6 enable 1
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 11
[SwitchA-Vlan-interface11] isis ipv6 enable 1
```

```
[SwitchA-Vlan-interface11] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] address-family ipv6
[SwitchB-isis-1-ipv6] quit
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] isis ipv6 enable 1
[SwitchB-Vlan-interface10] quit
[SwitchB] interface vlan-interface 13
[SwitchB-Vlan-interface13] isis ipv6 enable 1
[SwitchB-Vlan-interface13] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] address-family ipv6
[SwitchC-isis-1-ipv6] quit
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 11
[SwitchC-Vlan-interface11] isis ipv6 enable 1
[SwitchC-Vlan-interface11] quit
[SwitchC] interface vlan-interface 13
[SwitchC-Vlan-interface13] isis ipv6 enable 1
[SwitchC-Vlan-interface13] quit
```

3. Configure BFD functions:

Enable BFD and configure BFD parameters on Switch A.

```
[SwitchA] bfd session init-mode active
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] isis ipv6 bfd enable
[SwitchA-Vlan-interface10] bfd min-transmit-interval 500
[SwitchA-Vlan-interface10] bfd min-receive-interval 500
[SwitchA-Vlan-interface10] bfd detect-multiplier 7
[SwitchA-Vlan-interface10] quit
```

Enable BFD and configure BFD parameters on Switch B.

```
[SwitchB] bfd session init-mode active
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] isis ipv6 bfd enable
[SwitchB-Vlan-interface10] bfd min-transmit-interval 500
[SwitchB-Vlan-interface10] bfd min-receive-interval 500
[SwitchB-Vlan-interface10] bfd detect-multiplier 6
[SwitchB-Vlan-interface10] quit
```

Verifying the configuration

Display BFD session information on Switch A.

```
[SwitchA] display bfd session

 Total sessions: 1     Up sessions: 1     Init mode: Active

 IPv6 session working in control packet mode:

          Local discr: 1441                       Remote discr: 1450
            Source IP: FE80::20F:FF:FE00:1202 (link-local address of Vlan-interface10 on
Switch A)
       Destination IP: FE80::20F:FF:FE00:1200 (link-local address of Vlan-interface10 on
Switch B)
        Session state: Up                            Interface: Vlan10
            Hold time: 2319ms
```

Display routes destined for 2001:4::0/64 on Switch A.

```
[SwitchA] display ipv6 routing-table 2001:4::0 64

Summary count : 1

Destination: 2001:4::/64                                Protocol  : IS_L1
NextHop    : FE80::20F:FF:FE00:1200                     Preference: 15
Interface  : Vlan10                                     Cost      : 10
```

The output shows that Switch A and Switch B communicate through VLAN-interface 10. Then the link over VLAN-interface 10 fails.

Display routes destined for 2001:4::0/64 on Switch A.

```
[SwitchA] display ipv6 routing-table 2001:4::0 64

Summary count : 1

Destination: 2001:4::/64                                Protocol  : IS_L1
NextHop    : FE80::BAAF:67FF:FE27:DCD0                  Preference: 15
Interface  : Vlan11                                     Cost      : 20
```

The output shows that Switch A and Switch B communicate through VLAN-interface 11.

Example: Configuring IPv6 IS-IS FRR

Network configuration

As shown in Figure 27, Switch A, Switch B, and Switch C belong to the same IS-IS routing domain. Configure IPv6 IS-IS FRR so that when the Link A fails, traffic can be switched to Link B immediately.

Figure 27 Network diagram

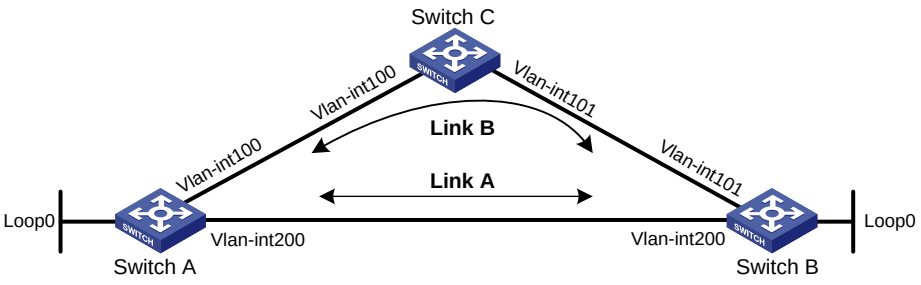


Table 8 Interface and IP address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Switch A	Vlan-int100	1::1/64	Switch B	Vlan-int101	3::1/64
	Vlan-int200	2::1/64		Vlan-int200	2::2/64
	Loop0	10::1/128		Loop0	20::1/128
Switch C	Vlan-int100	1::2/64			
	Vlan-int101	3::2/64			

Procedure

1. Configure IPv6 addresses for interfaces on the switches. (Details not shown.)
2. Configure IPv6 IS-IS:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] address-family ipv6
[SwitchA-isis-1-ipv6] quit
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis ipv6 enable 1
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] isis ipv6 enable 1
[SwitchA-Vlan-interface200] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] isis ipv6 enable 1
[SwitchA-LoopBack0] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] address-family ipv6
[SwitchB-isis-1-ipv6] quit
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] isis ipv6 enable 1
[SwitchB-Vlan-interface101] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis ipv6 enable 1
[SwitchB-Vlan-interface200] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] isis ipv6 enable 1
[SwitchB-LoopBack0] quit
```

3. Configure IPv6 IS-IS FRR:

Enable IPv6 IS-IS FRR to calculate a backup next hop through LFA calculation, or designate a backup next hop by using a routing policy.

- (Method 1.) Enable IPv6 IS-IS FRR to calculate a backup next hop through LFA calculation:

  # Configure Switch A.

  ```
  [SwitchA] isis 1
  [SwitchA-isis-1] address-family ipv6
  [SwitchA-isis-1-ipv6] fast-reroute lfa
  [SwitchA-isis-1-ipv6] quit
  [SwitchA-isis-1] quit
  ```

  # Configure Switch B.

  ```
  [SwitchB] isis 1
  [SwitchB-isis-1] address-family ipv6
  [SwitchB-isis-1-ipv6] fast-reroute lfa
  [SwitchB-isis-1-ipv6] quit
  [SwitchB-isis-1] quit
  ```

- (Method 2.) Enable IPv6 IS-IS FRR to designate a backup next hop by using a routing policy:

  # Configure Switch A.

  ```
  [SwitchA] ipv6 prefix-list abc index 10 permit 20::1 128
  [SwitchA] route-policy frr permit node 10
  [SwitchA-route-policy-frr-10] if-match ipv6 address prefix-list abc
  [SwitchA-route-policy-frr-10] apply ipv6 fast-reroute backup-interface
  vlan-interface 100 backup-nexthop fe80::6eab:8bff:fed4:205
  [SwitchA-route-policy-frr-10] quit
  [SwitchA] isis 1
  [SwitchA-isis-1] address-family ipv6
  [SwitchA-isis-1-ipv6] fast-reroute route-policy frr
  [SwitchA-isis-1-ipv6] quit
  [SwitchA-isis-1] quit
  ```

  # Configure Switch B.

  ```
  [SwitchB] ipv6 prefix-list abc index 10 permit 10::1 128
  [SwitchB] route-policy frr permit node 10
  [SwitchB-route-policy-frr-10] if-match ipv6 address prefix-list abc
  [SwitchB-route-policy-frr-10] apply ipv6 fast-reroute backup-interface
  vlan-interface 101 backup-nexthop fe80::6eab:8bff:fed4:206
  [SwitchB-route-policy-frr-10] quit
  [SwitchB] isis 1
  [SwitchB-isis-1] address-family ipv6
  [SwitchB-isis-1-ipv6] fast-reroute route-policy frr
  [SwitchB-isis-1-ipv6] quit
  [SwitchB-isis-1] quit
  ```

Verifying the configuration

Display route 20::1/128 on Switch A to view the backup next hop information.

```
[SwitchA] display ipv6 routing-table 20::1 128 verbose

Summary count : 1

Destination: 20::1/128
```

```
   Protocol: IS_L1
 Process ID: 1
  SubProtID: 0x1                  Age: 00h27m45s
       Cost: 10             Preference: 15
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0xa               OrigVrf: default-vrf
    TableID: 0xa                OrigAs: 0
      NibID: 0x24000005         LastAs: 0
     AttrID: 0xffffffff       Neighbor: ::
      Flags: 0x10041       OrigNextHop: FE80::6EAB:90FF:FEF3:305
      Label: NULL          RealNextHop: FE80::6EAB:90FF:FEF3:305
    BkLabel: NULL            BkNextHop: FE80::6EAB:8BFF:FED4:205
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface200
BkTunnel ID: Invalid       BkInterface: Vlan-interface100
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0              MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Display route 10::1/128 on Switch B to view the backup next hop information.

```
[SwitchB] display ipv6 routing-table 10::1 128 verbose

Summary count : 1

Destination: 10::1/128
   Protocol: IS_L1
 Process ID: 1
  SubProtID: 0x1                  Age: 00h33m23s
       Cost: 10             Preference: 15
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0xa               OrigVrf: default-vrf
    TableID: 0xa                OrigAs: 0
      NibID: 0x24000006         LastAs: 0
     AttrID: 0xffffffff       Neighbor: ::
      Flags: 0x10041       OrigNextHop: FE80::6EAB:81FF:FE67:106
      Label: NULL          RealNextHop: FE80::6EAB:81FF:FE67:106
    BkLabel: NULL            BkNextHop: FE80::6EAB:8BFF:FED4:206
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface200
BkTunnel ID: Invalid       BkInterface: Vlan-interface101
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0              MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Contents

BGP overview 1

BGP characteristics 1
BGP application scenarios 1
BGP speaker and BGP peer 2
BGP message types 2
BGP message formats 2
BGP message header 2
OPEN message 3
UPDATE message 4
NOTIFICATION message 6
KEEPALIVE message 9
ROUTE-REFRESH message 10
BGP finite state machine 10
BGP path attributes 11
Categories of BGP path attributes 12
ORIGIN 12
AS_PATH 12
NEXT_HOP 13
MED (MULTI_EXIT_DISC) 14
LOCAL_PREF 15
ATOMIC_AGGREGATE 15
AGGREGATOR 15
ORIGINATOR_ID 15
CLUSTER_LIST 16
COMMUNITY 16
EXT_COMMUNITY 17
4-byte AS numbers 17
BGP route processing mechanism 18
Route redistribution 18
Route summarization 19
BGP route selection 19
BGP route advertisement rules 20
BGP load balancing 21
Route reflector 22
Settlements for problems in large-scale BGP networks 23
Route dampening 23
Peer group and dynamic peer 24
Community 24
Confederation 24
BGP update group 25
MP-BGP 25
BGP route redistribution into local VPN instances 27
BGP multi-instance 28
About this feature 28
Examples 29
BGP configuration views 29
Protocols and standards 31

Building basic BGP networks 1

Basic BGP network configuration tasks at a glance (IPv4 unicast/IPv4 multicast) 1
Basic BGP network configuration tasks at a glance (IPv6 unicast/IPv6 multicast) 2
Configuring basic BGP 3
Enabling BGP 3
Configuring a BGP peer 4
Configuring dynamic BGP peers 8
Configuring an IBGP peer group 13
Configuring an EBGP peer group 17

Specifying the source address of TCP connections ····· 29
Controlling BGP route generation ····· 31
Injecting a local network ····· 31
Redistributing IGP routes ····· 32
Redistributing the local network specified in the public instance or a VPN instance ····· 34
Configuring BGP route summarization ····· 35
Advertising a default route to a peer or peer group ····· 37
Controlling BGP route advertisement ····· 38
Advertising optimal routes in the IP routing table ····· 38
Enabling prioritized withdrawal of the default route ····· 40
Enabling prioritized withdrawal of specific routes ····· 40
Configuring BGP route distribution filtering policies ····· 41
Enabling first-AS-number check for EBGP route advertisement ····· 45
Setting the BGP route sending rate ····· 46
Configuring BGP to delay sending route updates on reboot ····· 47
Configuring the interval for sending updates for the same route ····· 48
Configuring BGP route advertisement delay ····· 49
Setting the refresh route advertisement delay timer ····· 49
Configuring a startup policy for BGP route updates ····· 50
Controlling BGP route reception ····· 51
Limiting routes received from a peer or peer group ····· 51
Configuring BGP route reception filtering policies ····· 52
Configuring the SoO attribute ····· 56
Configuring BGP timers ····· 57
Configuring the keepalive interval and hold time ····· 57
Setting the session retry timer ····· 59
Setting an update delay for local MPLS labels ····· 60
Configuring BGP logging and notifications ····· 60
Enabling logging for session state changes ····· 60
Enabling logging for BGP route flapping ····· 61
Configuring BGP network management ····· 62
Display and maintenance commands for basic BGP network building ····· 64
Displaying BGP ····· 64
Resetting BGP sessions ····· 70
Clearing BGP information ····· 71
Basic IPv4 BGP network configuration examples ····· 72
Example: Configuring basic BGP ····· 72
Example: Configuring BGP and IGP route redistribution ····· 76
Example: Configuring dynamic BGP peers ····· 79
Example: Configuring BGP route summarization ····· 81
Example: Configuring multicast BGP ····· 84
Example: Separating unicast traffic and multicast traffic ····· 87
Example: Configuring multiple BGP instances on one switch ····· 92
Basic IPv6 BGP network configuration examples ····· 95
Example: Configuring IPv6 BGP basics ····· 95
Example: Configuring IPv6 multicast BGP ····· 98
Troubleshooting BGP ····· 102
State of the connection to a peer cannot become established ····· 102
Configuring large-scale BGP networks ····· 1
Large-scale BGP network configuration tasks at a glance ····· 1
Configuring BGP route dampening ····· 1
Configuring BGP communities ····· 2
Configuring BGP route reflection ····· 6
Configuring a BGP route reflector ····· 6
Ignoring the ORIGINATOR_ID attribute ····· 8
Configuring a route server ····· 9
Configuring BGP confederation settings ····· 10
About BGP confederation ····· 10
Configuring a BGP confederation ····· 10
Configuring confederation compatibility ····· 11
Display and maintenance commands for large-scale BGP network configuration ····· 11

Large-scale BGP network configuration examples ····14
Example: Configuring BGP communities ····14
Example: Configuring BGP route reflector ····17
Example: Configuring BGP confederation ····19
Controlling BGP path selection ····1
BGP path selection control tasks at a glance ····1
Configuring preferences for BGP routes ····2
Configuring the NEXT_HOP attribute ····3
Setting the local device IP as the next hop ····3
Advertising only the global unicast address in the NEXT_HOP attribute ····6
Setting a preferred value for received routes ····7
Configuring the default local preference ····8
Configuring the AS_PATH attribute ····9
Permitting local AS number to appear in routes from a peer or peer group ····9
Ignoring the AS_PATH attribute during optimal route selection ····11
Advertising a fake AS number to a peer or peer group ····11
Configuring AS number substitution ····12
Removing private AS numbers from sent updates ····14
Ignoring the first AS number of EBGP route updates ····15
Configuring the MED attribute ····16
About the MED attribute ····16
Configuring the default MED value ····16
Enabling MED comparison for routes from different ASs ····17
Enabling MED comparison for routes on a per-AS basis ····18
Enabling MED comparison for routes from confederation peers ····19
Configuring the link bandwidth attribute ····20
Minimizing the priority of BGP routes advertised to peers ····24
Minimizing the priority of BGP routes advertised to peers with a DOWN-to-UP state change ····24
Minimizing the priority of BGP routes advertised to peers after a device reboot ····31
Restoring the priority of advertised BGP routes ····33
Preferring BGP routes of a specific next hop IP type ····33
Ignoring IGP metrics during optimal route selection ····35
Ignoring router IDs during optimal route selection ····35
Preferring routes with an IPv6 next hop during optimal route selection ····36
Configuring the next hop type selection rule for optimal route selection ····36
About tunnel-type recursive next hops ····36
Configuring the condition for BGP routes to participate in optimal route selection ····37
Preferring BGP routes of a specific recursive next hop type during optimal route selection ····37
Associating a track entry with a BGP peer or peer group ····38
Display and maintenance commands for BGP path selection control ····39
BGP path selection control configuration examples ····40
Example: Configuring route selection based on the LOCAL_PREF attribute ····40
Example: Configuring route selection based on the MED attribute ····43
Example: Configuring route selection based on the AS_PATH attribute ····45
Example: Changing the preference value for a BGP route ····47
Tuning and optimizing BGP networks ····1
BGP network tuning and optimization tasks at a glance ····1
Setting the TCP MSS for TCP connections ····2
Enabling BGP to establish an EBGP session over multiple hops ····3
Enabling immediate re-establishment of direct EBGP connections upon link failure ····4
Enabling prefix-based BGP ORF capabilities negotiation ····4
About BGP ORF ····4
Restrictions and guidelines ····4
Prerequisites ····5
Enabling BGP ORF capabilities negotiation for a peer or peer group ····5
Enabling nonstandard BGP ORF capabilities negotiation for a peer or peer group ····6
Enabling 4-byte AS number suppression ····7
Enabling BGP to exchange withdrawal ROUTE-REFRESH messages with peers ····8
Enabling Track-based priority adjustment for BGP routes advertised to a peer or peer group ····10
Disabling BGP session establishment ····13

About disabling BGP session establishment 13
Restrictions and guidelines 14
Disabling BGP session establishment with a peer or peer group (IPv4 peers) 14
Disabling BGP session establishment with a peer or peer group (IPv6 peers) 14
Disabling BGP session establishment with all peers or peer groups 15
Configuring BGP soft-reset 16
About BGP soft-reset 16
Enabling route refresh (IPv4 peers) 16
Enabling route refresh (IPv6 peers) 17
Saving updates (IPv4 unicast/multicast address family) 17
Saving updates (IPv6 unicast/multicast address family) 18
Configuring manual soft-reset (IPv4 unicast/multicast address family) 18
Configuring manual soft-reset (IPv6 unicast/multicast address family) 19
Configuring BGP load balancing 20
Configuring the BGP Additional Paths feature 23
Protecting an EBGP peer when memory usage reaches level 2 threshold 25
Setting a DSCP value for outgoing BGP packets 26
Disabling route recursion policy control for routes received from a peer or peer group 27
Flushing the suboptimal BGP route to the RIB 27
Configuring a peer monitoring group 28
Disabling BGP from flushing specific routes to the routing table 29
Specifying a label allocation mode 30
Disabling optimal route selection for labeled routes without tunnel information 31
Configuring BGP optimal route selection delay 31
Configuring the BGP session state to collaborate with the monitor link downlink interfaces 34
Specifying the longest match principle for next hop recursion 38
Enabling conversational remote host route learning 40
Enabling peer unreachability detection 41
Displaying 4-byte AS numbers in asdot or asdot+ notation 42
Converting AS numbers in effective device configurations 43
Configuring the tag community attribute 43
About this task 43
Procedure (IPv4 unicast) 44
Procedure (IPv6 unicast) 45
Display and maintenance commands for BGP network tuning and optimization 46
Displaying BGP 46
Resetting BGP sessions 47
BGP network tuning and optimization configuration examples 47
Example: Configuring BGP load balancing 47
Example: Configuring the BGP Additional Paths feature 50
Example: Configuring prefix-based BGP ORF 53
Configuring BGP security features 1
BGP security feature configuration tasks at a glance 1
Enabling MD5 authentication for BGP peers 1
Configuring GTSM for BGP 2
Enabling keychain authentication for BGP peers 3
Configuring IPsec for IPv6 BGP 4
Configuring BGP RPKI 5
About BGP RPKI 5
Configuring RPKI connection parameters 6
Enabling BGP RPKI validation 7
Applying the BGP RPKI validation state to optimal route selection 8
Advertising BGP RPKI validation state to a peer or peer group 9
Resetting BGP RPKI sessions 10
Display and maintenance commands for BGP security features 10
IPv4 BGP security feature configuration examples 11
Example: Configuring BGP RPKI 11
Example: Configuring BGP keychain authentication 13
Example: Configuring BGP GTSM 15
IPv6 BGP security feature configuration examples 18
Example: Configuring IPsec for IPv6 BGP packets 18

Example: Configuring BGP RPKI ····· 23
Improving BGP network reliability ····· 1
BGP network reliability improvement tasks at a glance ····· 1
Configuring BGP GR ····· 1
Configuring BGP NSR ····· 5
Configuring BFD for BGP ····· 6
Configuring BGP FRR ····· 7
About BGP FRR ····· 7
Configuring BGP FRR by using a routing policy (IPv4 unicast address family) ····· 8
Configuring BGP FRR by using a routing policy (IPv6 unicast address family) ····· 9
Configuring BGP FRR through PIC (IPv4 unicast address family) ····· 10
Configuring BGP FRR through PIC (IPv6 unicast address family) ····· 10
Configuring BFD-powered nexthop connectivity detection for BGP FRR ····· 11
Configuring BGP maintenance features ····· 12
About this task ····· 12
Configuring BGP isolation ····· 12
Configuring BGP shutdown ····· 13
Display and maintenance commands for BGP network reliability improvement ····· 13
IPv4 BGP network reliability improvement configuration examples ····· 14
Example: Configuring BGP GR ····· 14
Example: Configuring BFD for BGP ····· 15
Example: Configuring BGP FRR ····· 19
IPv6 BGP network reliability improvement configuration examples ····· 23
Example: Configuring BFD for IPv6 BGP ····· 23
Example: Configuring IPv6 BGP FRR ····· 27
Configuring extended BGP features ····· 1
Extended BGP feature configuration tasks at a glance ····· 1
Configuring BMP ····· 1
BMP configuration tasks at a glance ····· 3
Establishing a TCP connection between the BMP client and BMP server ····· 3
Enabling a BMP server to monitor a peer or peer group ····· 4
Sending routes to the BMP server ····· 5
Configuring the BMP client to send notifications with mode flags to the BMP server ····· 6
Sending statistics information to the BMP server ····· 7
Configuring BGP LS ····· 7
About BGP LS ····· 7
BGP LS extensions ····· 8
Configuring basic BGP LS ····· 8
Configuring BGP LS route reflection ····· 9
Specifying an AS number and a router ID for BGP LS messages ····· 9
Performing manual soft-reset for BGP sessions of LS address family ····· 10
Configuring BGP dedicated routes ····· 10
About BGP dedicated routes ····· 10
Configuring basic BGP dedicated route settings ····· 11
Configuring BGP dedicated route reflection ····· 11
Resetting BGP sessions in dedicated address family ····· 12
Configuring 6PE ····· 12
About 6PE ····· 12
Prerequisites ····· 13
Configuring basic 6PE ····· 13
Controlling path selection for 6PE ····· 14
Controlling 6PE route advertisement and reception ····· 14
Tuning and optimizing 6PE ····· 15
Configuring 6PE route reflection ····· 16
Resetting 6PE connections ····· 16
Display and maintenance commands for extended BGP features ····· 16
Displaying BGP ····· 16
Resetting BGP sessions ····· 18
Clearing BGP information ····· 19
Extended BGP feature configuration examples ····· 19

Example: Configuring BGP LS 19
Example: Configuring BGP dedicated routes 21
Example: Configuring 6PE 25

BGP overview

Border Gateway Protocol (BGP) is an exterior gateway protocol (EGP). It is called internal BGP (IBGP) when it runs within an AS and called external BGP (EBGP) when it runs between ASs.

BGP characteristics

BGP has the following characteristics:

- Focuses on route control and selection rather than route discovery and calculation. In addition, it provides rich routing policies, allowing flexible route filtering and selection.
- Uses TCP (port number 179) to enhance reliability. This is because BGP needs to handle a large volume of routing information to implement inter-AS route advertisement and selection. Devices can exchange routing information even if they are not directly connected.
- Has good scalability and can adapt to network developments. An increasing number of new technologies use BGP as their control plane protocol for information transmission.
- Measures the distance of a route by using a list of ASs that the route must travel through to reach the destination. BGP is also called a path-vector protocol.
- Supports CIDR.
- Reduces bandwidth consumption by advertising only incremental updates. BGP is very suitable to advertise large numbers of routes on the Internet.
- Eliminates routing loops by adding AS path information to BGP route updates. A device discards a route when it finds that the route contains the local AS number.

BGP application scenarios

BGP helps an increasing number of emerging technologies establish information pathways. It inserts information of the control plane into BGP routes and transmits the information across networks. BGP is used in the following scenarios:

- The control plane is deployed with BGP and works in conjunction with forwarding-plane technologies to carry services.
  - Scenarios where different users need to unicast packets to each other across ASs in the public network.
  - Scenarios where unicast routing information used for RPF checks and multicast topology setup requires cross-AS advertisement.
  - Scenarios where BGP and MPLS must work together to implement BGP MPLS L3VPN or L2VPN:
    - VPN routing information needs to be transmitted over the public network. For more information, see MPLS L3VPN configuration in *MPLS Configuration Guide*.
    - Label block information needs to be transmitted. For more information, see MPLS L2VPN configuration and VPLS configuration in *MPLS Configuration Guide*.
  - Scenarios where SRv6 tunnels are used to forward VPN traffic over the public network. For more information, see SRv6 VPN overview, IP L3VPN over SRv6, EVPN L3VPN over SRv6, EVPN VPWS over SRv6, EVPN VPLS over SRv6, and IP public network over SRv6 in *Segment Routing Configuration Guide*.
  - Scenarios where multicast is required in VPNs. For more information, see multicast VPN configuration in *IP Multicast Configuration Guide*.
- The implementation of new features requires BGP extensions:

- Scenarios where devices need to transmit IPSG bindings.
- Ethernet Virtual Private Network (EVPN) scenarios. For more information, see *EVPN Configuration Guide*.
- Scenarios where inter-AS paths require SIDs to support SR cross-domain forwarding. These scenarios are based on BGP Egress Peer Engineering (BGP-EPE). For more information about BGP-EPE, see SR-MPLS configuration or SRv6 configuration in *Segment Routing Configuration Guide*.

BGP speaker and BGP peer

A router running BGP is a BGP speaker. A BGP speaker establishes peer relationships with other BGP speakers to exchange routing information over TCP connections.

Based on the AS, a BGP peer can be either of the following types:

- **IBGP peer**—Resides in the same AS as the local router.
- **EBGP peer**—Resides in different ASs from the local router.

Based on the IP version, a BGP peer can be either of the following types:

- **IPv4 peer**—Uses an IPv4 address to establish a peer relationship with the local router.
- **IPv6 peer**—Uses an IPv6 address to establish a peer relationship with the local router.

BGP message types

BGP uses the following message types:

- **OPEN**—After establishing a TCP connection, BGP sends an OPEN message to establish a session to the peer.
- **UPDATE**—BGP sends UPDATE messages to exchange routing information between peers. Each UPDATE message can advertise a group of routes with reachable next hops and identical attributes. An UPDATE message can also contain multiple withdrawn routes. You can identify those routes by their destination addresses in the Network Layer Reachability Information (NLRI) of that update message.
- **KEEPALIVE**—BGP sends KEEPALIVE messages between peers to maintain connectivity.
- **ROUTE-REFRESH**—BGP sends a ROUTE-REFRESH message to request the routing information for a specific address family from a peer. The local device might send a ROUTE-REFRESH message to a peer if both of them are enabled with the ROUTE-REFRESH capability. When the BGP route reception policy changes, the local device sends a ROUTE-REFRESH message to its peers. On receipt of that message, a peer sends its route information to the local device, so the local device can update BGP routes and apply the new reception policy to them.
- **NOTIFICATION**—BGP sends a NOTIFICATION message upon detecting an error and immediately closes the connection.

BGP message formats

A BGP message consists of message header and message content. The five BGP messages have the same message header, but the content of each message differs.

BGP message header

The format of the BGP message header is as shown in Figure 1. For more information about fields in the BGP message header, see Table 1.

Figure 1 BGP message header

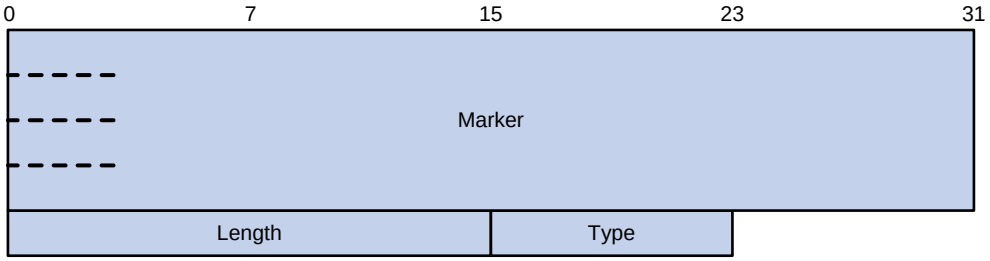


Table 1 Fields in the BGP message header

Field	Length (in bytes)	Description
Marker	16	Synchronization state of the BGP connection. If the value of this field is not a hexadecimal string that contains only Fs, the BGP connection has a synchronization error.
Length	2	Total length of the BGP message, including the message header. The length range is 19 to 4096 bytes.
Type	1	BGP message type: • **1**—OPEN • **2**—UPDATE • **3**—NOTIFICATION • **4**—KEEPALIVE • **5**—ROUTE-REFRESH

OPEN message

OPEN messages are used for BGP connection establishment and BGP capability negotiation. The format of OPEN messages is as shown in Figure 2. For more information about fields in an OPEN message, see Table 2.

Figure 2 OPEN message

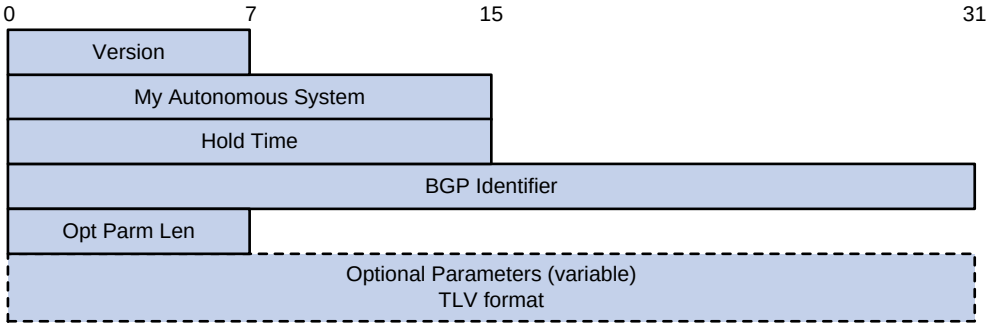


Table 2 Fields in an OPEN message

Field	Length (in bytes)	Description
Version	1	BGP version number.
My Autonomous System	2	AS number of the message sender.

Hold Time	2	Hold timer set for the message sender, in seconds. This timer is for peer connection maintenance purposes, and specifies the interval for sending KEEPALIVE or UPDATE messages. After both parties in a BGP session receive the hold timer set for the other, they use the smaller timer.
BGP Identifier	4	Router ID of the message sender.
Opt Parm Len	1	Total length of the **Optional Parameters** field, in bytes. A value of 0 indicates that they message does not contain optional parameters.
Option Parameters	Variable	Optional parameters, each of which is a unit in TLV format. They are typically used to list the capabilities that the message sender can negotiate. For more information about the capabilities supported by the device, see "Tuning and optimizing BGP networks."

UPDATE message

UPDATE messages are used for route advertisement. The format of UPDATE messages is as shown in Figure 3. For more information about fields in an UPDATE message, see Table 3.

Figure 3 UPDATE message

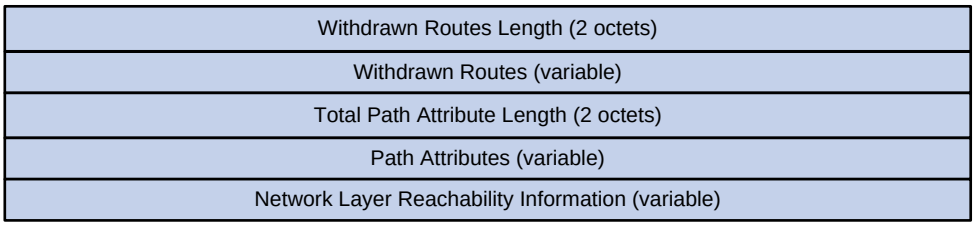
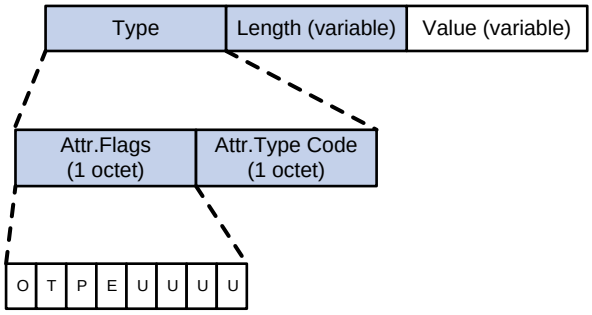


Table 3 Fields in an UPDATE message

Field	Length (in bytes)	Description
Withdrawn Routes Length	2	Total length of the **Withdrawn Routes** field, in bytes. A value of 0 indicates that the message does not contain routes being withdrawn.
Withdrawn Routes	Variable	List of routes being withdrawn. Each route is represented in Length/Prefix format: • **Length**—Length of the IP address prefix. This portion is one byte long. • **Prefix**—IP address prefix of the route. This portion is variable.
Total Path Attribute Length	2	Total length of the **Path Attributes** field, in bytes.
Path Attributes	Variable	List of path attributes carried by routes in the message. Those attributes are listed in ascending order of their type numbers and each of them is encoded in TLV format.

		 The **Type** field indicates the type of a route attribute. It is two bytes long and consists of the **Attr.Flag** and **Attr.Type Code** subfields. • **Attr.Flag**: Type flags of the route attribute: ○ **O**—Indicates whether the route attribute is a well-known attribute. If the route attribute is well-known, this flag is set to 0. If the route attribute is optional, this flag is set to 1. ○ **T**—Indicates whether the route attribute is a transitive attribute. If the route attribute is transitive, this flag is set to 1. If not, this flag is set to 0. ○ **P**—Indicates whether the route attribute is partially effective. If a route attribute has this flag set to 1, it takes effect only within the same AS and will not be transmitted to other ASs. This flag is set to 1 only if the route attribute is an optional transitive attribute and is partially effective . ○ **E**—Indicates whether the **Length** field of the route attribute needs to be extended. If field extension is required, this flag is set to 1, and the **Length** field occupies two bytes. If field extension is not required, this flag is not set, and the **Length** field occupies only one byte. ○ **U**—This flag is reserved and thus set to 0. • **Attr.Type Code**: Type code of the route attribute, which indicates the type of the route attribute. ○ **1**—ORIGIN ○ **2**—AS_PATH ○ **3**—NEXT_HOP ○ **4**—MED ○ **5**—LOCAL_PREF ○ **6**—ATOMIC_AGGREGATE ○ **7**—AGGREGATOR ○ **8**—COMMUNITY ○ **9**—ORIGINATOR_ID ○ **10**—CLUSTER_LIST ○ **14**—MP_REACH_NLRI ○ **15**—MP_UNREACH_NLRI ○ **16**—EXT-COMMUNITY The **Length** field indicates the length of the route attribute, in bytes. The **Value** field indicates the content of the route attribute.
Network Layer Reachability Information (NLRI)	Variable	List of routes to be updated. Each route is represented in Length/Prefix format: • **Length**—Length of the IP address prefix. This portion is one byte long. • **Prefix**—IP address prefix of the route. This portion is

		variable.

NOTIFICATION message

NOTIFICATION messages are used to notify BGP session errors. The format of NOTIFICATION messages is as shown in Figure 4. For more information about fields in a NOTIFICATION message, see Table 4.

Figure 4 NOTIFICATION message

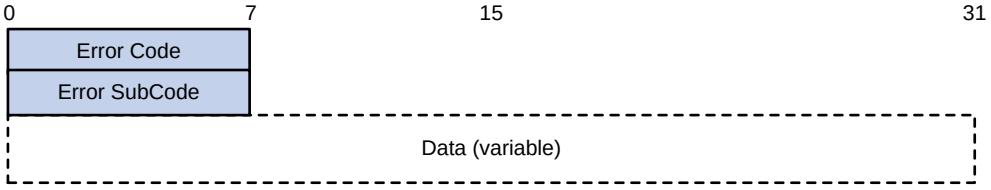


Table 4 Fields in a NOTIFICATION message

Field	Length (in bytes)	Description
Error Code	1	Error code.
Error SubCode	1	Error subcode.
Data	Variable	Detailed error information.

Table 5 introduces supported error codes and subcodes.

Table 5 Error codes and subcodes

Error code/subcode	Reason for peer disconnection	Description
1/1	connection not synchronized	The two ends of the connection were not synchronized. The current implementation is that the first 16 bytes of the received message's header do not contain only Fs.
1/2	bad message length	Invalid message length.
1/3	bad message type	Invalid message type.
3/1	the withdrawn length is too large	The length of routing information to be withdrawn was too long.
	the attribute length is too large	The attribute length was too long.
	one attribute appears more than once	A path attribute appeared multiple times in an UPDATE message.
	the attribute length is too small	The attribute length was less than two bytes.
	exntended length field is less than two octets	The attribute length was extensible, but it was less than two bytes.
	the length field is less than one octet	The attribute length was not extensible, but it was less than one byte.
	link-state attribute error	The link-state attribute was in incorrect form.
3/2	unrecognized well-known attribute	Unknown well-known attribute.

3/3	*attribute-type* attribute missed	The *attribute-type* attribute was lost. The values for the *attribute-type* argument include: • ORIGIN • AS_PATH • LOCAL_PREF • NEXT_HOP
3/4	attribute flags error	Incorrect attribute flags.
3/5	*attribute-type* attribute length error	The length of the *attribute-type* attribute was invalid. The values for the *attribute-type* argument include: • AS_PATH • AS4_PATH • CLUSTER_LIST • AGGREGATOR • AS4_AGGREGATOR • ORIGIN • NEXT_HOP • MED • LOCAL_PREF • ATOMIC_AGGREGATE • ORIGINATOR_ID • MP_REACH_NLRI • COMMUNITIES • EXT-COMMUNITIES
	attribute length exceeds	The attribute length crossed the limit.
3/6	invalid ORIGIN attribute	Invalid ORIGIN attribute.
3/8	invalid NEXT_HOP attribute	Invalid NEXT_HOP attribute.
3/9	invalid nexthop length in MP_REACH_NLRI (*address-family*)	The **Nexthop** length in the MP_REACH_NLRI attribute was invalid for the *address-family* address family. The values for the *address-family* argument include: • **4u**—IPv4 unicast address family. • **MPLS**—MPLS address family. • **VPNv4**—VPNv4 address family. • **6u**—IPv6 unicast address family. • **VPNv6**—VPNv6 address family. • **L2VPN**—L2VPN address family.
	the length of MP_UNREACH_NLRI is too small	The MP_UNREACH_NLRI attribute was shorter than three bytes.
	the MP NLRI attribute length exceeds	The length of the MP_REACH_NLRI or MP_UNREACH_NLRI attribute crossed the limit.
	erroneous MP NLRI attribute end position	The reachable or unreachable prefix and the path attribute ended at different positions.
3/10	invalid network field	Invalid network field.
3/11	malformed AS_PATH	The AS_PATH attribute was malformed.
4/0	Keepalive last triggered time	Most recent time when KEEPALIVE message

		sending was triggered.
	Keepalive last sent time	Most recent time when a KEEPALIVE message was sent.
	Update last sent time	Most recent time when an UPDATE message was sent.
	EPOLLOUT last occurred time	Most recent time when an EPOLLOUT event occurred.
	Keepalive last received time	Most recent time when a KEEPALIVE message was received.
	Update last received time	Most recent time when an UPDATE message was received.
	EPOLLIN last occurred time	Most recent time when an EPOLLIN event occurred.
5/0	connection retry timer expires	The ConnectRetry timer expired.
	TCP_CR_Acked event received	A TCP_CR_Acked event was received.
	TCP_Connection_Confirmed event received	A TCP_Connection_Confirmed event was received.
5/3	open message received	An OPEN message was received.
6/0	manualstop event received	A manualstop event was received.
	physical interface configuration changed	Physical configurations changed, such as interface settings.
	session down event received from BFD	A BFD session down event was received.
6/1	maximum number of prefixes reached	The number of route prefixes has exceeded the upper limit specified by using the **peer route-limit** command.
	maximum number of *address-family* prefixes reached	The number of route prefixes in the *address-family* address family has exceeded the upper limit specified by using the **peer route-limit** command. The values for the *address-family* argument include: • **IPv4 unicast**—IPv4 unicast address family. • **IPv6 unicast**—IPv6 unicast address family. • **VPNv4**—VPNv4 address family. • **VPNv6**—VPNv6 address family.
6/2	configuration of peer ignore changed	The **peer ignore** command was configured.
6/3	address family deleted	An address family was deleted.
	peer disabled	A peer was disabled.
6/4	administrative reset	The BGP session was reset because of the **reset bgp** command or configuration changes.
6/5	connection rejected	The connection request was rejected.
6/6	other configuration change	Other configurations changed.

6/7	connection collision resolution	A connection conflict occurred.
	two connections exist and MD5 authentication is configured for the neighbor	Two connections existed and MD5 authentication was configured for one of them.
6/8	no memory to process the attribute	The memory was insufficient for attribute parsing.
	no memory for the route	Failed to obtain memory resources for route or label block generation.
	no memory to generate unreachable NLRI	Failed to obtain memory resources for MP_UNREACH_NLRI encapsulation.
	no memory to generate a message	Failed to obtain memory resources for message encapsulation.
	can't get the VPN RD	Failed to obtain RDs upon prefix parsing.
	can't get the VPN routing table	Failed to obtain the VPN routing table upon prefix parsing.
	can't get the attributes	Failed to obtain attributes upon prefix parsing.
	entered severe memory state	A severe memory usage alarm was triggered.
	entered critical memory state	A critical memory usage alarm was triggered.

KEEPALIVE message

KEEPALIVE messages are transmitted between BGP peers to maintain connectivity. The format of KEEPALIVE messages is as shown in Figure 5. For more information about fields in a KEEPALIVE message, see Table 6.

Figure 5 KEEPALIVE message

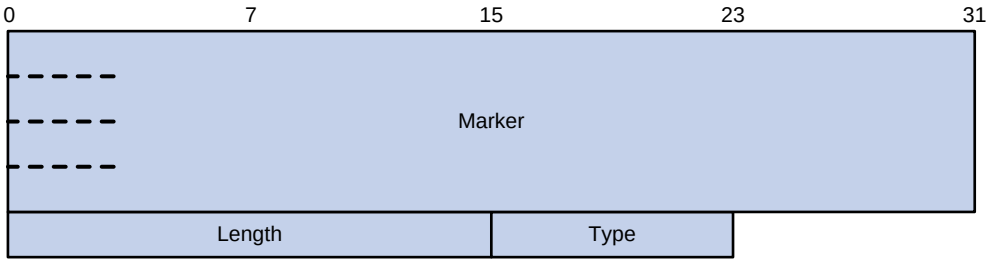


Table 6 Fields in a KEEPALIVE message

Field	Length (in bytes)	Description
Marker	16	This field indicates the completeness of information synchronization between BGP peers, and can be used for BGP authentication. When BGP authentication is not configured, the value for this field is a hexadecimal string that contains only Fs.
Length	2	Total length of the KEEPALIVE message, in bytes.
Type	1	BGP message type. The type value is 4 for KEEPALIVE messages.

ROUTE-REFRESH message

BGP sends a ROUTE-REFRESH message to request the routing information of a specific address family from a peer. On receipt of the message, the peer re-advertises routes in the specified address family. The format of ROUTE-REFRESH messages is as shown in Figure 6. For more information about fields in a ROUTE-REFRESH message, see Table 7.

Figure 6 ROUTE-REFRESH message

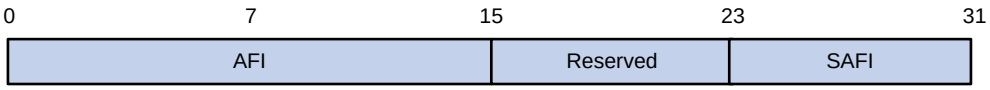


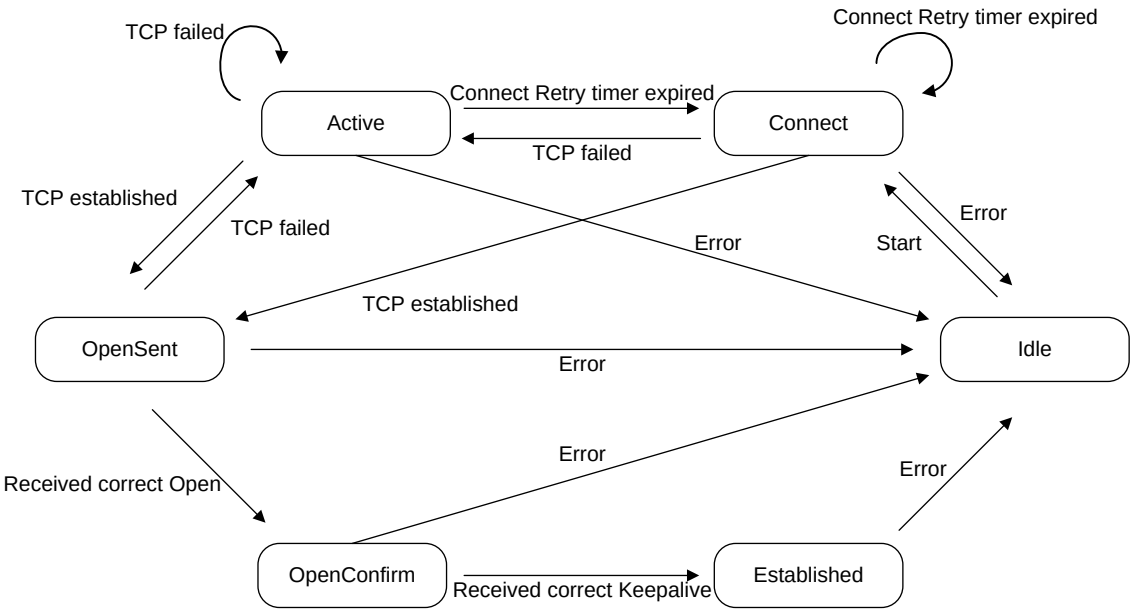
Table 7 Fields in a ROUTE-REFRESH message

Field	Length (in bytes)	Description
AFI	2	Identifier of the address family.
Reserved	1	Reserved field.
SAFI	1	Identifier of the subsequent address family.

BGP finite state machine

During session establishment, the devices at the two ends transition between different BGP states, as shown in Figure 7.

Figure 7 BGP finite state machine



Idle

The idle state is the initial BGP state. In this state, BGP rejects any connection requests. Only upon receiving a Start event, the device assigns resources to BGP, attempts to establish a TCP connection, and transitions to Connect state.

BGP transitions to Idle state from the remaining states upon a TCP disconnection error, packet error, connection closed due to configuration issues, and reception of a NOTIFICATION message.

Connect

In this state, BGP starts a Connect Retry timer and waits for TCP connection establishment to be completed.

- If TCP connection establishment succeeds before the Connect Retry timer expires, BGP closes the Connect Retry timer, sends an OPEN message to the peer, and transitions to OpenSent state.
- If TCP connection establishment fails before the Connect Retry timer expires, BGP transitions to Active state.
- If the Connect Retry timer expires, and no connection response is received, BGP resets the Connect Retry timer, tries to establish a TCP connection with the peer again, and stays in Connect state.

Active

In this state, BGP continuously attempts to establish a TCP connection.

- If TCP connection establishment succeeds, BGP closes the Connect Retry timer, sends an OPEN message to the peer, and transitions to OpenSent state.
- If TCP connection establishment fails, BGP resets the Connect Retry timer and stays in Active state.
- If the Connect Retry timer expires and no connection response is received, BGP resets the Connect Retry timer and transitions to Connect state.

OpenSent

In this state, BGP waits for an OPEN message from the peer, and checks the BGP version number and AS number in the OPEN message.

- If the OPEN message is correct, BGP sends a KEEPALIVE message to the peer, and transitions to OpenConfirm state.
- If the OPEN message is incorrect, BGP sends a NOTIFICATION message to the peer, and transitions to Idle state.

If the TCP connection is disconnected in this state, BGP resets the Connect Retry timer, transitions to Active state, and tries to establish a TCP connection again.

OpenConfirm

In this state, BGP waits for a KEEPALIVE or NOTIFICATION message.

- Upon receiving a KEEPALIVE message, BGP transitions to Established state.
- Upon receiving a NOTIFICATION message, BGP transitions to Idle state.

Established

In this state, BGP can exchange UPDATE, KEEPALIVE, ROUTE-REFRESH, and NOTIFICATION messages with the peer.

- Upon receiving a correct UPDATE or KEEPALIVE message, BGP stays in Established state.
- Upon receiving an incorrect UPDATE or KEEPALIVE message, BGP sends a NOTIFICATION message to the peer and transitions to Idle state.
- Upon receiving a ROUTE-REFRESH message, BGP does not change its state.
- Upon receiving a NOTIFICATION message, BGP transitions to Idle state.

BGP path attributes

BGP path attributes are a set of parameters that are advertised along with routes. They provide further description of specific routes, enabling the route receiver to filter and select routes.

BGP uses a group of path attributes in UPDATE messages for route filtering and selection.

Categories of BGP path attributes

BGP path attributes are categorized as follows:

- **Well-known mandatory**—Attributes that all BGP devices must recognize and UPDATE messages must contain. UPDATE messages lacking these attributes are invalid.
- **Well-known discretionary**—Attributes that all BGP devices must recognize, but UPDATE messages do not need to contain.
- **Optional transitive**—Attributes that BGP devices might not be able to recognize, but can receive and advertise to other BGP devices.
- **Optional non-transitive**—If a BGP device does not support these attributes, the device will ignore them in UPDATE messages and will not forward them to other peers.

ORIGIN

ORIGIN is a well-known mandatory attribute that specifies the origin of BGP routes. This attribute has the following types:

- **IGP**—Has the highest priority. Routes generated in the local AS have the IGP attribute.
- **EGP**—Has the second highest priority. Routes obtained through EGP have the EGP attribute.
- **INCOMPLETE**—Has the lowest priority. The source of routes with this attribute is unknown. Routes redistributed from other routing protocols have the INCOMPLETE attribute.

AS_PATH

AS_PATH is a well-known mandatory attribute that records the ASs a route must pass through from the local network to the destination address.The route receiver can then determine the ASs required to route traffic to the destination address based on the AS_PATH attribute.

When a BGP device advertises local routes, it processes them as follows:

- Before advertising routes to EBGP peers, it appends the local AS number to the end of each route's AS_PATH list.
- When it advertises routes to IBGP peers, it does not add any AS number to the AS_PATH list of each route. The AS_PATH list of each advertised route is empty.

When a BGP device forwards routes, it processes them as follows:

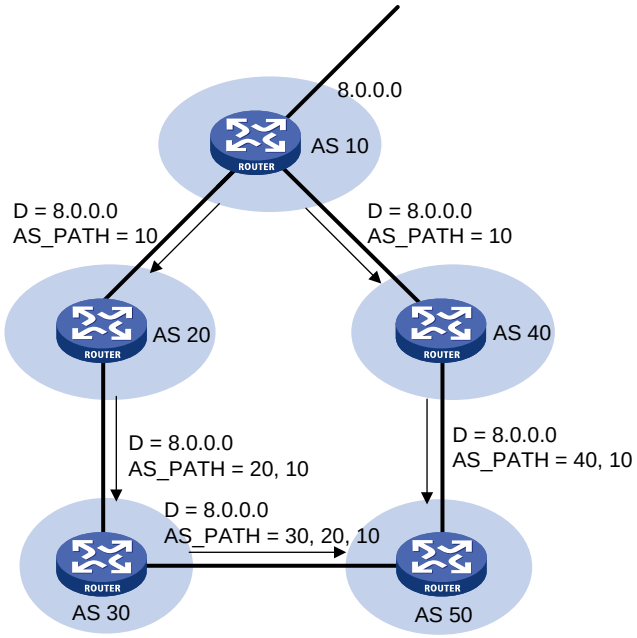
- Before forwarding routes to EBGP peers, it appends the local AS number to the head of each route's AS_PATH list.
- Before forwarding routes to IBGP peers, it does not change the AS_PATH list of each route.

AS numbers in the AS_PATH attribute are arranged in one of the following methods:

- **AS_SEQUENCE**—Arranges AS numbers in sequence. As shown in Figure 8, the number of the AS closest to the receiver's AS is leftmost.
- **AS_SET**—Arranges AS numbers randomly. This method is applicable to route summarization (also called route aggregation) scenarios. After routes are summarized, the system cannot sequentially arrange the AS_PATH attributes of summarized routes. Instead, it records the AS numbers in an unordered manner. When BGP calculates routes, the length of AS_PATH attributes in AS_SET format is counted as one AS number.
- **AS_CONFED_SEQUENCE**—Arranges sub-AS numbers within a confederation in sequence.
- **AS_CONFED_SET**—Arranges sub-AS numbers within a confederation randomly. This method is applicable to intra-confederation route summarization scenarios.

The AS_CONFED_SEQUENCE and AS_CONFED_SET methods are used only in BGP routes forwarded inside a confederation.

Figure 8 AS_PATH attribute



BGP uses the AS_PATH attribute to implement the following functions:

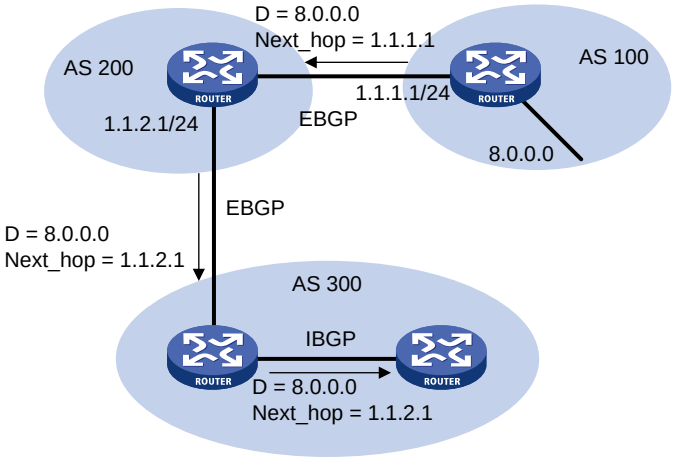
- **Avoid routing loops**—A BGP router does not receive routes containing the local AS number to avoid routing loops.
- **Affect route selection**—BGP gives priority to the route with the shortest AS_PATH length if other factors are the same. As shown in Figure 8, the BGP router in AS 50 gives priority to the route passing AS 40 for sending data to the destination 8.0.0.0. In some applications, you can apply a routing policy to control BGP route selection by modifying the AS_PATH length. For more information about routing policy, see "Configuring routing policies."
- **Filter routes**—By using an AS path list, you can filter routes based on AS numbers contained in the AS_PATH attribute. For more information about AS path list, see "Configuring routing policies."

NEXT_HOP

NEXT_HOP is a well-known mandatory attribute that might not contain the IP address of a directly connected router. Its value is determined as follows:

- When a BGP speaker advertises a self-originated route to a BGP peer, it sets the address of the sending interface as the NEXT_HOP.
- When a BGP speaker sends a received route to an EBGP peer, it sets the address of the sending interface as the NEXT_HOP.
- When a BGP speaker sends a route received from an EBGP peer to an IBGP peer, it does not modify the NEXT_HOP attribute. If load balancing is configured, BGP modifies the NEXT_HOP attribute for the equal-cost routes. For load balancing information, see "BGP load balancing."

Figure 9 NEXT_HOP attribute

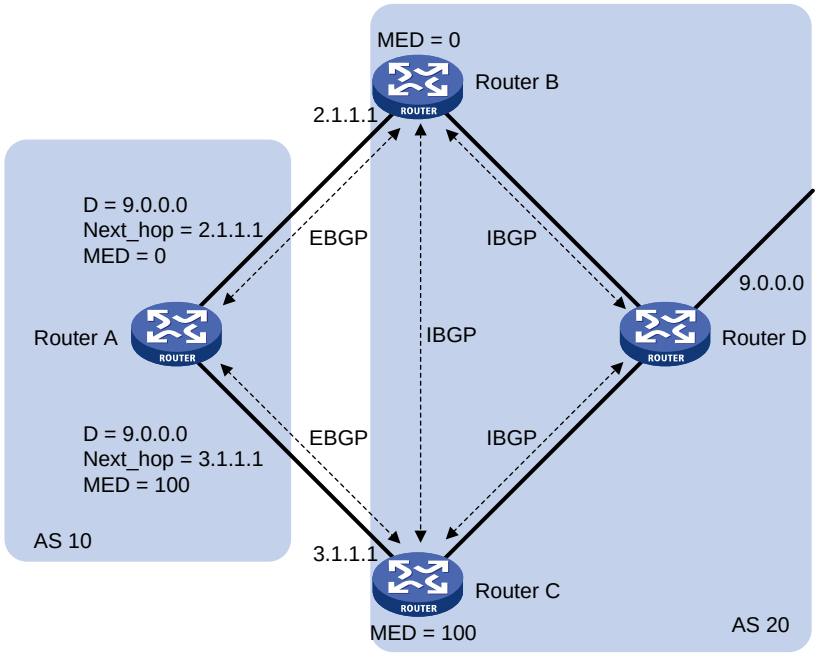


MED (MULTI_EXIT_DISC)

BGP advertises the MED attribute between two neighboring ASs, each of which does not advertise the attribute to any other AS. The MED attribute is an optional non-transitive attribute.

Similar to metrics used by IGPs, MED is used to determine the optimal route for traffic going into an AS. When a BGP router obtains multiple routes to the same destination but with different next hops, it selects the route with the smallest MED value as the optimal route. As shownin Figure 10, traffic from AS 10 to AS 20 travels through Router B that is selected according to MED.

Figure 10 MED attribute



Generally, BGP only compares MEDs of routes received from the same AS. You can also use the **`compare-different-as-med`** command to force BGP to compare MED values of routes received from different ASs.

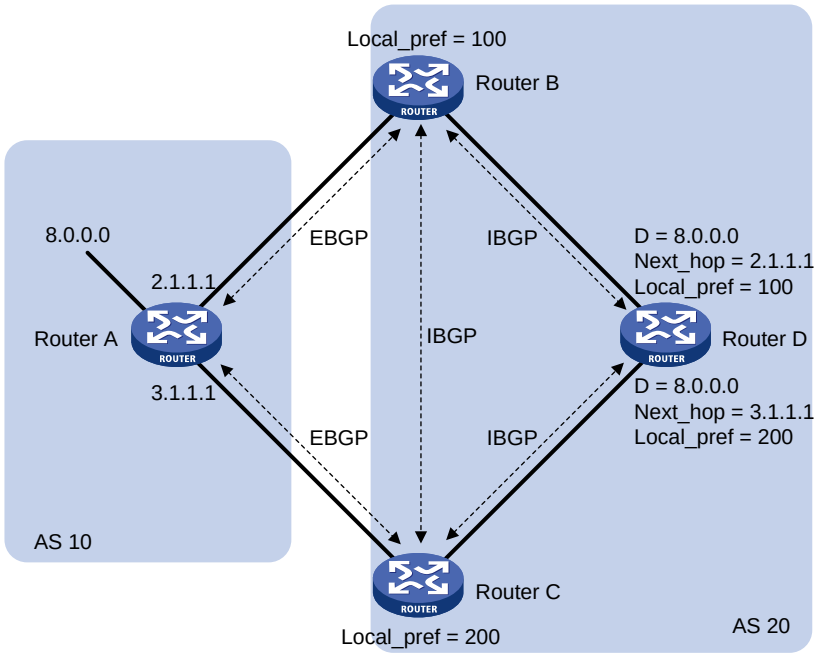
LOCAL_PREF

LOCAL_PREF is a well-known discretionary attribute exchanged between IBGP peers only, and is not advertised to any other AS. It indicates the priority of a BGP router.

BGP uses LOCAL_PREF to determine the optimal route for traffic leaving the local AS. When a BGP router obtains multiple routes to the same destination but with different next hops, it selects the route with the highest LOCAL_PREF value as the optimal route. As shown in Figure 11, traffic from AS 20 to AS 10 travels through Router C that is selected according to LOCAL_PREF.

The LOCAL_PREF attribute is transmitted only between IBGP peers, and cannot be advertised to other ASs.

Figure 11 LOCAL_PREF attribute



ATOMIC_AGGREGATE

ATOMIC_AGGREGATE is a well-known discretionary attribute that identifies the aggregation status of a route. A route with the ATOMIC_AGGREGATE attribute is an aggregate route. The route receiver should be aware that the member routes of the aggregate route might be dropped during route advertisement.

AGGREGATOR

AGGREGATOR is an optional transitive attribute supplementary to the ATOMIC_AGGREGATE attribute. It contains the IP address and the AS number of the BGP device that initiated the route aggregation. When a device aggregates BGP routes, it adds both the ATOMIC_AGGREGATE and AGGREGATOR attributes to the aggregate route.

ORIGINATOR_ID

ORIGINATOR_ID is an optional non-transitive attribute that identifies the router ID of the route generator within the local AS. Only route reflectors can add this attribute to routes.

The ORIGINATOR_ID attribute prevents routing loops caused by the introduction of route reflectors. When a BGP route is reflected for the first time, the route reflector adds the ORIGINATOR_ID attribute to the route. The attribute contains the router ID of the device that sent the route to the reflector. When the route reflector receives a route carrying the ORIGINATOR_ID attribute, it does not modify the route. Upon receiving a route with the ORIGINATOR_ID attribute, a device identifies whether the router ID in the ORIGINATOR_ID attribute is the same as its own router ID. If they are the same, a routing loop exists within the current AS, and the receiver discards the route.

CLUSTER_LIST

CLUSTER_LIST is an optional non-transitive attribute that prevents routing loops by recording the route reflector clusters a route has passed through.

This attribute is generated by route reflectors. When a route reflector a route, it adds the local cluster ID to the CLUSTER_LIST attribute of that route. If the route does not carry a CLUSTER_LIST attribute, the route reflector creates one for the route. When a route reflector receives a BGP route, it identifies whether the route's CLUSTER_LIST attribute contains the local cluster ID. If the local router ID is found, the route will be discarded to prevent routing loops.

COMMUNITY

COMMUNITY is an optional transitive attribute that identifies the communities of BGP routes. A BGP community is a group of routes with the same characteristics. It has no geographical boundaries. Routes of different ASs can belong to the same community.

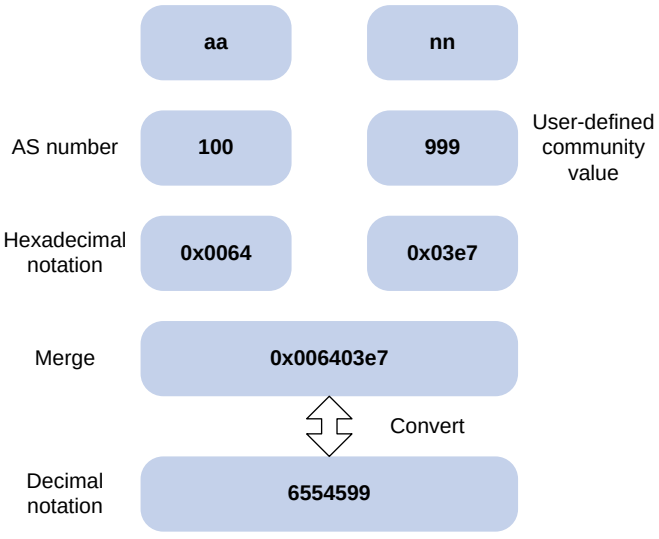
A route can carry one or more COMMUNITY attribute values (each of which is represented by a 4-byte integer). A router uses the COMMUNITY attribute to determine whether to advertise the route and the advertising scope without using complex filters such as ACLs. This mechanism simplifies routing policy configuration, management, and maintenance.

The COMMUNITY attribute supports the following formats:

- *aa*:*nn*—Both the *aa* and *nn* arguments are in the range of 0 to 65535. The *aa* argument represents the AS number from which the route comes, and the *nn* argument represents the user-defined community value. For example, if the community value of a route from AS 100 is 999, the COMMUNITY attribute of that route is 100:999.
- **Integer**—An integer in the range of 0 to 4294967295.

Figure 12 shows the relationship between the above two formats.

Figure 12 Relationship between the aa:nn and integer formats



Well-known COMMUNITY attributes involve the following:

- **INTERNET**—By default, all routes belong to the Internet community. Routes with this attribute can be advertised to all BGP peers. The value for this attribute is 0.
- **NO_EXPORT**—Routes with this attribute cannot be advertised out of the local AS or out of the local confederation, but can be advertised to other sub-ASs in the confederation. For confederation information, see "Settlements for problems in large-scale BGP networks." The value for this attribute is 4294967041.
- **No_ADVERTISE**—Routes with this attribute cannot be advertised to other BGP peers. The value for this attribute is 4294967042.
- **No_EXPORT_SUBCONFED**—Routes with this attribute cannot be advertised out of the local AS or other sub-ASs in the local confederation. The value for this attribute is 4294967043.

You can configure BGP community lists to filter BGP routes based on the BGP COMMUNITY attribute.

EXT_COMMUNITY

To meet new demands, the EXT_COMMUNITY attribute is introduced, which is an optional non-transitive attribute. This attribute has the following advantages over the COMMUNITY attribute:

- Provides more attribute values by extending the attribute length to eight bytes.
- Allows for using different types of extended community attributes in different scenarios to enhance route filtering and control and simplify configuration and management.

The device supports the link bandwidth extended community attribute, the route target extended community attribute, the tag community attribute, and the Site of Origin (SoO) extended community attribute. For information about route target, see *MCE Configuration Guide.*

The link bandwidth attribute is used for BGP route-based UCMP load balancing. For information about the link bandwidth attribute, see "Configuring the link bandwidth attribute."

The SoO attribute specifies the site where the route originated. It prevents advertising a route back to the originating site. If the AS-path attribute is lost, the router can use the SoO attribute to avoid routing loops.

The SoO attribute has the following formats:

- *16-bit integer AS number*:*32-bit user-defined number*. For example, 101:3.
- *16-bit dotted AS number*:*32-bit user-defined number*. For example, 0.1:1. The value range is 0.1 to 0.65535 for a dotted AS number.
- *32-bit IP address*:*16-bit user-defined number*. For example, 192.168.122.15:1.
- *32-bit integer AS number*:*16-bit user-defined number*. For example, 65536:1. The minimum integer AS number is 65536.
- *32-bit dotted AS number*:*16-bit user-defined number*. For example, 10.1:1. The minimum dotted AS number is 1.0.

The tag attribute is used to transmit IGP tag information.

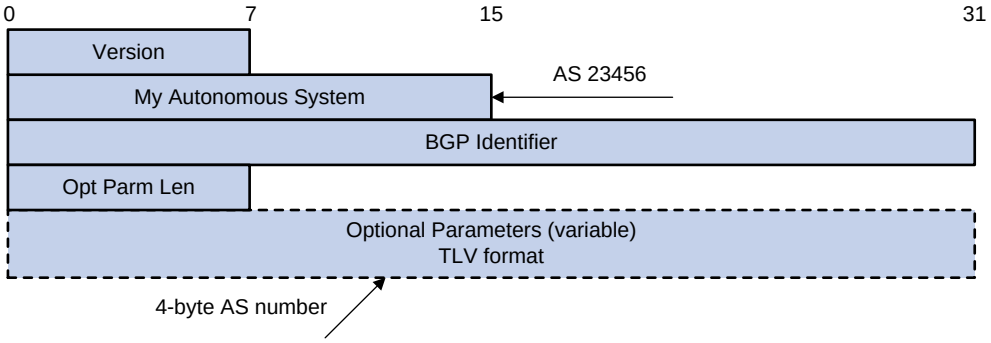
4-byte AS numbers

In earlier BGP versions, each AS number was 2 bytes long and was in the range of 1 to 65535. However, as the network size grew, 2-byte AS numbers became insufficient. To avoid AS number exhaustion, AS numbers were lengthened to 4 bytes and their value range became 1 to 4294967295.

Before establishing a session using 4-byte AS numbers, two devices must negotiate with each other about the 4-byte AS number capability through OPEN messages. The session can be established successfully only if both devices support 4-byte AS numbers and the capability negotiation succeeds.

During the capability negotiation, the format of OPEN messages is as shown in Figure 13. The **My Autonomous System** field does not support 4-byte AS numbers because the formats of fields in the OPEN message header are fixed. Therefore, a device using a 4-byte AS number inserts AS number 23456 into the **My Autonomous System** field, and places the real 4-byte AS number in the **Optional Parameters** field. With this approach, two peering devices can determine whether the other supports 4-byte AS numbers and obtain the other's 4-byte AS number during capability negotiation.

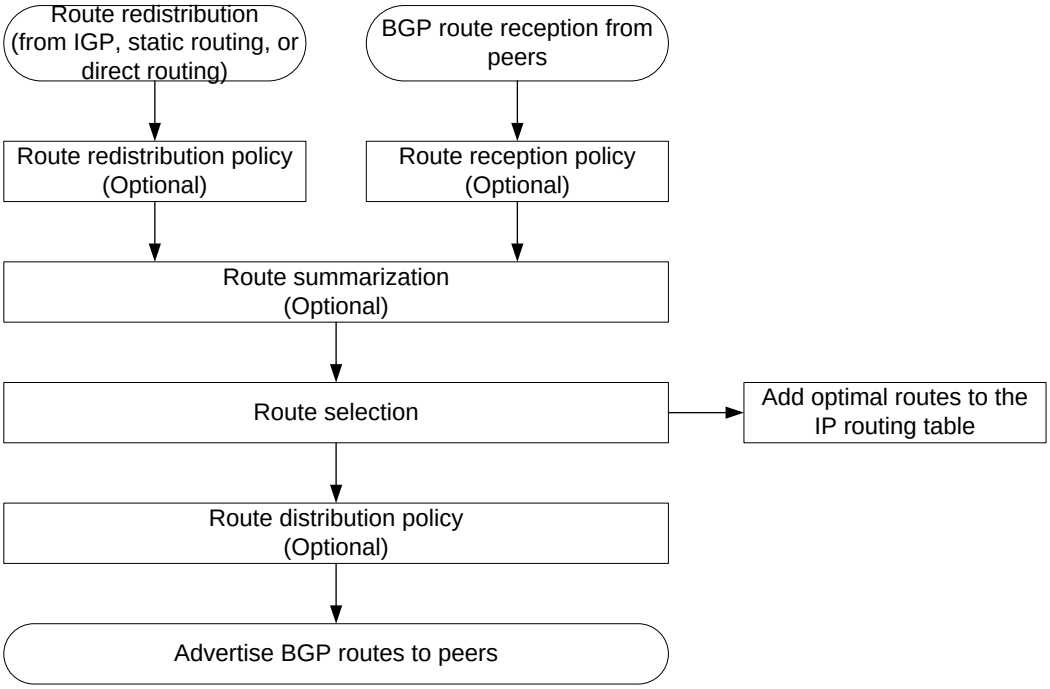
Figure 13 OPEN messages used for 4-byte AS number capability negotiation



BGP route processing mechanism

Figure 14 shows how BGP routes are generated, redistributed, and advertised.

Figure 14 BGP route processing mechanism



Route redistribution

BGP cannot discover routes by itself. To have BGP generate and advertise routes, you must redistribute routes from other protocols into the routing table of BGP. BGP supports the following route redistribution methods:

- **Per-protocol redistribution**—This method redistributes routes in bulk on a per routing protocol basis into the routing table of BGP. The routes that can be redistributed include IGP routes, direct routes, and static routes.
- **Per-route redistribution**—This method can identify a specific route by its prefix and redistribute it into the routing table of BGP.

Route summarization

Route summarization can summarize multiple routes into one summary route. This method can shrink the BGP routing table, because BGP only advertises summary routes rather than summarized routes to peers.

The system supports both manual and automatic route summarization.

- **Automatic summarization**—Summarizes routes redistributed into BGP according to natural network segments. After automatic summarization, BGP suppresses the summarized routes.
- **Manual summarization**—Summarizes BGP routes according to the user-defined summarization settings. This method allows you to determine the attribute of a summary route and whether to advertise the summarized routes.

BGP route selection

BGP discards routes with unreachable next hops and routes without IP-type recursive next hops. If the **bestroute nexthop-resolved none** command is executed, BGP does not discard routes with unreachable next hops. If the **bestroute nexthop-resolved tunnel** command is executed, BGP does not discard routes with only tunnel-type recursive next hops. In this situation, BGP still discards routes with next hops recursed to neither tunnels nor IP addresses.

If multiple routes to the same destination are available, BGP selects the optimal route in the following sequence:

1. The route with the highest-priority RPKI validation result if the **bestroute origin-as-validation** command is executed.

   If the **bestroute origin-as-validation** command is not executed, BGP skips this route selection rule.
2. The route with the highest Preferred_value.
3. The route of the nexthop IP type specified in the **bestroute nexthop-priority preferred** command.

   If the **bestroute nexthop-priority preferred** command is not executed, BGP skips this route selection rule.
4. The route of the encapsulation type specified in the **bestroute encap-type preferred** command in the BGP VPN instance.

   This route selection rule does not take effect in BGP public instances.

   If the **bestroute encap-type preferred** command is not executed, BGP skips this route selection rule.
5. The route with the highest LOCAL_PREF.
6. The route of the encapsulation type specified in the **bestroute encap-type** command in the BGP VPN instance.

   This route selection rule does not take effect in BGP public instances.

   If the **bestroute encap-type** command is not executed, BGP skips this route selection rule.
7. The route generated by the **network** command, the route redistributed by the **import-route** command, or the summary route in turn.

8. The route with the shortest AS_PATH.
9. The IGP, EGP, or INCOMPLETE route in turn.
10. The route with the lowest MED value.
11. The route learned from EBGP, confederation EBGP, confederation IBGP, or IBGP in turn.
12. The route of the nexthop IP type specified in the **bestroute nexthop-priority** command.

    If the **bestroute nexthop-priority** command is not executed, BGP skips this route selection rule.
13. The route with the smallest IGP metric.

    If the **bestroute igp-metric-ignore** command is executed, BGP skips this route selection rule.
14. The route with the smallest recursion depth.
15. The route of the nexthop type specified in the **bestroute nexthop-type** command.

    If the **bestroute nexthop-type** command is not executed, BGP skips this route selection rule.
16. If a route received from an EBGP peer is the current optimal route, BGP does not change the optimal route when it receives routes from other EBGP peers.
17. The route advertised by the router with the smallest router ID.

    If one of the routes is advertised by a route reflector, BGP compares the ORIGINATOR_ID of the route with the router IDs of other routers. Then, BGP selects the route with the smallest ID as the optimal route.

    If the **bestroute router-id-ignore** command is executed, BGP skips this route selection rule.
18. The route with the shortest CLUSTER_LIST.
19. The route with an IPv6 next hop if the **bestroute ipv6-nexthop** command is executed.

    If the **bestroute ipv6-nexthop** command is not executed, BGP skips this route selection rule.
20. The route with an IPv4 next hop.
21. The route advertised by the peer with the lowest IP address.

The CLUSTER_IDs of route reflectors form a CLUSTER_LIST. If a route reflector receives a route that contains its own CLUSTER ID in the CLUSTER_LIST, the router discards the route to avoid routing loops.

If load balancing is configured, the system selects available routes to implement load balancing.

BGP route advertisement rules

BGP follows these rules for route advertisement:

- When multiple feasible routes to a destination exist, BGP advertises only the optimal route to its peers. If the **advertise-rib-active** command is configured, BGP advertises the optimal route in the IP routing table. If not, BGP advertises the optimal route in the BGP routing table.
- BGP advertises only routes that it uses.
- BGP advertises routes learned from an EBGP peer to all BGP peers, including both EBGP and IBGP peers.
- BGP advertises routes learned from an IBGP peer to EBGP peers, rather than other IBGP peers.
- After establishing a session to a new BGP peer, BGP advertises all the routes matching the above rules to the peer. After that, BGP advertises only incremental updates to the peer.

BGP load balancing

BGP load balancing is applicable between EBGP peers, between IBGP peers, and between confederations.

BGP implements load balancing through route recursion and route selection.

BGP load balancing through route recursion

The next hop of a BGP route might not be directly connected. One of the reasons is that the next hop information exchanged between IBGP peers is not modified. The BGP router must find the directly connected next hop through IGP. The matching route with the direct next hop is called the recursive route. The process of finding a recursive route is route recursion.

If multiple recursive routes to the same destination are load balanced, BGP generates the same number of next hops to forward packets.

BGP load balancing based on route recursion is always enabled in the system.

BGP load balancing through route selection

IGP routing protocols, such as RIP and OSPF, can use route metrics as criteria to load balance between routes that have the same metric. BGP cannot load balance between routes by route metrics as an IGP protocol does, because BGP does not have a route computation algorithm.

BGP uses the following load balancing criteria to determine load balanced routes:

- The routes have the same ORIGIN, LOCAL_PREF, and MED attributes.
- The routes meet the following requirements on the AS_PATH attribute:
  - If the **`balance as-path-neglect`** command is configured, the routes can have different AS_PATH attributes.
  - If only the **`balance as-path-relax`** command is configured, the routes can have different AS_PATH attributes, but the length of the AS_PATH attributes must be the same.
  - If neither the **`balance as-path-neglect`** nor the **`balance as-path-relax`** command is configured, the routes must have the same AS_PATH attribute.
- The next hops of the routes meet the following requirements on IGP metrics:
  - If neither the **`bestroute igp-metric-ignore`** command nor the **`balance igp-metric-ignore`** command is configured, the next hops of the routes must have the same IGP metric value.
  - If either the **`bestroute igp-metric-ignore`** command or the **`balance igp-metric-ignore`** command is configured, the next hops of the routes can have different IGP metric values.
- The routes have the same MPLS label assignment status (labeled or not labeled).

BGP does not use the route selection rules described in "BGP route selection" for load balancing.

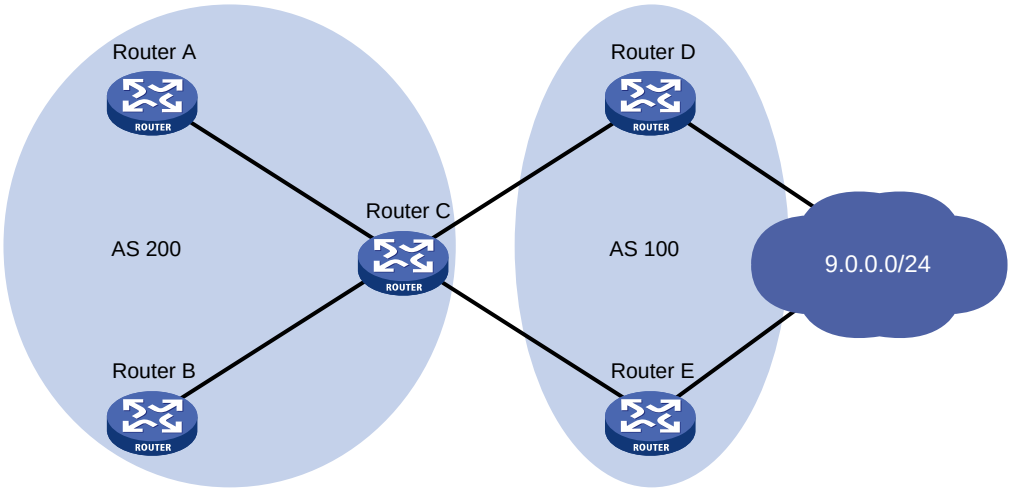
As shown in Figure 15, Router A and Router B are IBGP peers of Router C. Router C allows a maximum number of two ECMP routes for load balancing.

Router D and Router E both advertise a route 9.0.0.0 to Router C. Router C installs the two routes to its routing table for load balancing if the routes meet the BGP load balancing criteria. After that, Router C forwards to Router A and Router B a single route whose attributes are changed as follows:

- AS_PATH attribute:
  - If the **`balance as-path-neglect`** and **`balance as-path-relax`** commands are not configured, the AS_PATH attribute does not change.
  - If the **`balance as-path-neglect`** or **`balance as-path-relax`** command is configured, the AS_PATH attribute is changed to the attribute of the optimal route.
- The NEXT_HOP attribute is changed to the IP address of Router C.

- Other attributes are changed to be the same as the optimal route.

Figure 15 Network diagram



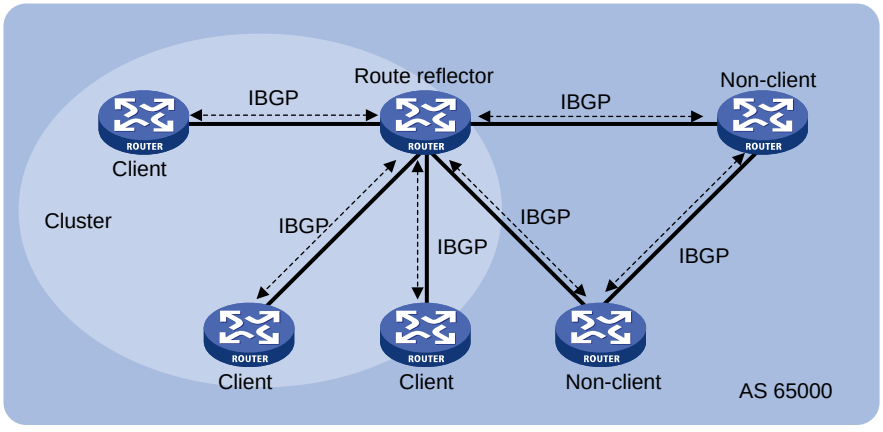
Route reflector

IBGP peers must be fully meshed to maintain connectivity. If n routers exist in an AS, the number of IBGP connections is n(n-1)/2. If a large number of IBGP peers exist, large amounts of network and CPU resources are consumed to maintain sessions.

Using route reflectors can solve this issue. In an AS, a router acts as a route reflector, and other routers act as clients connecting to the route reflector. The route reflector forwards routing information received from a client to other clients. In this way, all clients can receive routing information from one another without establishing BGP sessions.

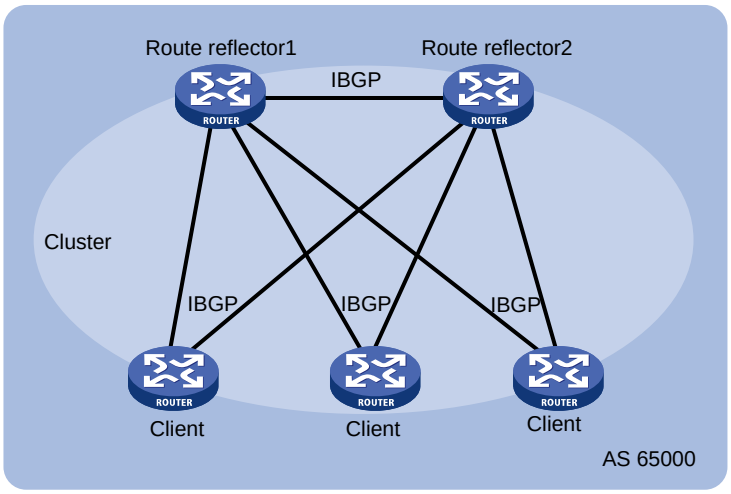
A router that is neither a route reflector nor a client is a non-client, which, as shown in Figure 16, must establish BGP sessions to the route reflector and other non-clients.

Figure 16 Network diagram for a route reflector



The route reflector and clients form a cluster. Typically a cluster has one route reflector. The ID of the route reflector is the Cluster_ID. You can configure more than one route reflector in a cluster to improve availability, as shown in Figure 17. The configured route reflectors must have the same cluster_ID to avoid routing loops.

Figure 17 Network diagram for route reflectors



When the BGP routers in an AS are fully meshed, route reflection is unnecessary because it consumes more bandwidth resources. You can use commands to disable route reflection between clients instead of modifying the network configuration or changing the network topology.

NOTE:

After route reflection is disabled between clients, routes can still be reflected between a client and a non-client.

Settlements for problems in large-scale BGP networks

You can use the following methods to facilitate management and improve route distribution efficiency on a large-scale BGP network.

Route dampening

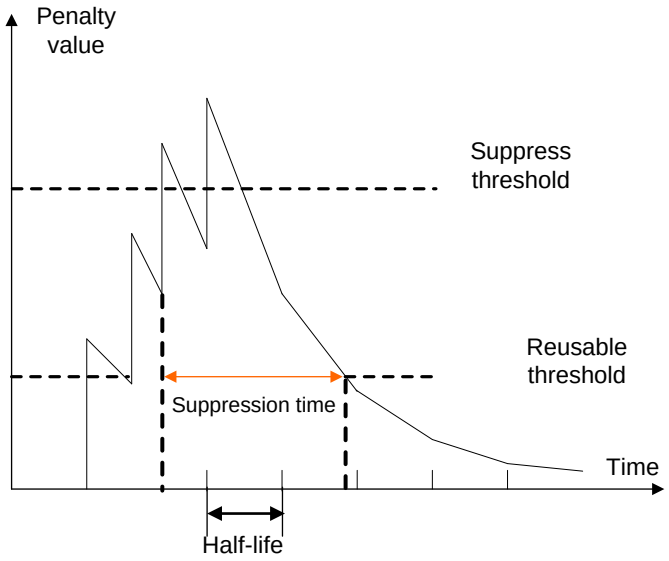
Route flapping (a route comes up and disappears in the routing table frequently) causes BGP to send many routing updates. It can consume too many resources and affect other operations.

In most cases, BGP runs in complex networks where route changes are more frequent. To solve the problem caused by route flapping, you can use BGP route dampening to suppress unstable routes.

BGP route dampening uses a penalty value to judge the stability of a route. The bigger the value, the less stable the route. Each time a route state changes from reachable to unreachable, or a reachable route's attribute changes, BGP adds a penalty value of 1000 to the route. When the penalty value of the route exceeds the suppress value, the route is suppressed and cannot become the optimal route. When the penalty value reaches the upper limit, no penalty value is added.

If the suppressed route does not flap, its penalty value gradually decreases to half of the suppress value after a period of time. This period is called "Half-life." When the value decreases to the reusable threshold value, the route is usable again.

Figure 18 BGP route dampening



Peer group and dynamic peer

You can organize BGP peers with the same attributes into a group to simplify their configurations.

When a peer joins the peer group, the peer obtains the same configuration as the peer group. If the configuration of the peer group is changed, the configuration of group members is changed.

In some BGP networks, peers might change frequently and cause changes to BGP sessions. Using the regular peering method requires frequent BGP configuration adjustments, which is laborious. To resolve this issue, you can configure the neighbors of a device in the same network segment as a dynamic peer. The device accepts the peering request of a device as long as the requesting device is in the configured network segment.

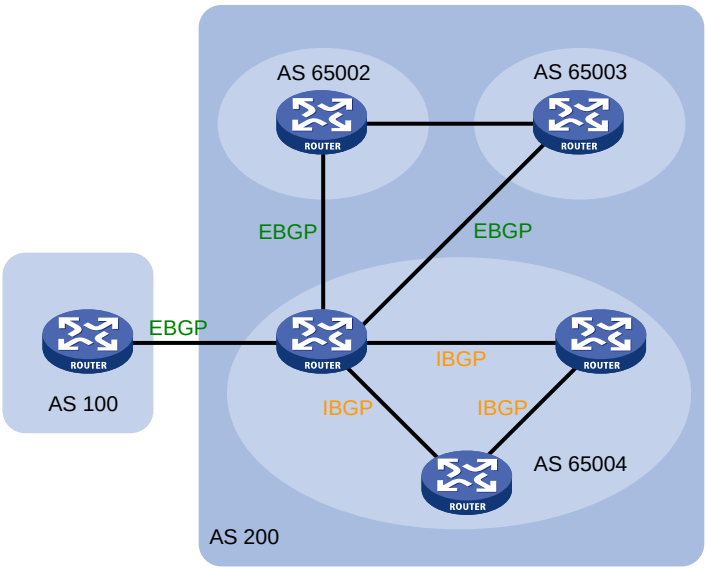
Community

You can apply a community list or an extended community list to a routing policy for route control. For more information, see "BGP path attributes."

Confederation

Confederation is another method to manage growing IBGP connections in an AS. It splits an AS into multiple sub-ASs. In each sub-AS, IBGP peers are fully meshed. As shown in Figure 19, intra-confederation EBGP connections are established between sub-ASs in AS 200.

Figure 19 Confederation network diagram



A non-confederation BGP speaker does not need to know sub-ASs in the confederation. To the BGP speaker, the confederation is one AS and the confederation ID is the AS number. In the above figure, AS 200 is the confederation ID.

Confederation has a deficiency. When you change an AS into a confederation, you must reconfigure the routers, and the topology will be changed.

In large-scale BGP networks, you can use both route reflector and confederation.

BGP update group

In large-scale BGP networks, a device needs to advertise routes to a large number of BGP peers, which requires high performance in bulk packet packaging and advertisement.

To conserve device resources, BGP can treat all peers with the same configurations as an update group. It packages a route only once and advertises that route to all peers in an update group. If the BGP update group feature is not supported, BGP must package routes on a per peer basis. For example, a route reflector has 10 clients and needs to reflect 10000 routes:

- If the route reflector does not support BGP update groups, it must package routes 100000 times.
- If the route reflector supports BGP update groups, it only needs to package routes 10000 times.

In the current software version, the BGP update group feature is supported by default..

MP-BGP

Supported address families

BGP-4 can only advertise IPv4 unicast routing information. Multiprotocol Extensions for BGP-4 (MP-BGP) can advertise routing information for the following address families:

- IPv6 unicast address family.
- IPv4 multicast address family and IPv6 multicast address family.

  PIM uses static and dynamic unicast routes to perform RPF check before creating multicast routing entries. When the multicast and unicast topologies are different, you can use MP-BGP

to advertise the routes for RPF check. MP-BGP stores the routes in the BGP multicast routing table. For more information about PIM and RPF check, see *IP Multicast Configuration Guide*.

- VPNv4 address family.

  For more information, see MPLS L3VPN configuration in *MPLS Configuration Guide.*

- Labeled IPv4 unicast address family and IPv6 unicast address family.

  MP-BGP advertises IPv4 unicast/IPv6 unicast routes and MPLS labels assigned for the routes. Labeled IPv4 unicast routes apply to inter-AS Option C for MPLS L3VPN. Labeled IPv6 unicast routes apply to 6PE and inter-AS Option C for MPLS L3VPN. For more information about inter-AS Option C, see MPLS L3VPN configruation in *MPLS Configuration Guide*.

- L2VPN address family.

  L2VPN information includes label block information and remote peer information. For more information about L2VPN and VPLS, see MPLS L2VPN configruation and VPLS configruation in *MPLS Configuration Guide*.

- EVPN address family.

  MP-BGP advertises EVPN routes to implement automatic VTEP discovery, VXLAN tunnel establishment and assignment, and MAC and ARP information advertisement. For more information about EVPN, see EVPN configruation in *EVPN Configuration Guide*.

MP-BGP extended attributes

Prefixes and next hops are key routing information. BGP-4 uses UPDATE messages to carry the following information:

- Feasible route prefixes in the Network Layer Reachability Information (NLRI) field.
- Unfeasible route prefixes in the withdrawn routes field.
- Next hops in the NEXT_HOP attribute.

BGP-4 cannot carry routing information for multiple network layer protocols.

To support multiple network layer protocols, MP-BGP defines the following path attributes:

- **MP_REACH_NLRI**—Carries feasible route prefixes and next hops for multiple network layer protocols.
- **MP_UNREACH_NLRI**—Carries unfeasible route prefixes for multiple network layer protocols.

MP-BGP uses these two attributes to advertise feasible and unfeasible routes for different network layer protocols. BGP speakers not supporting MP-BGP ignore updates containing these attributes and do not forward them to its peers.

Address family

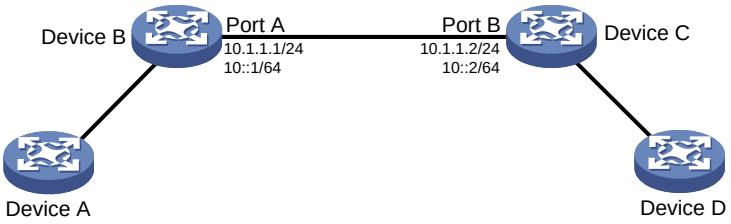
MP-BGP uses address families and subsequent address families to identify different network layer protocols for routes contained in the MP_REACH_NLRI and MP_UNREACH_NLRI attributes. For example, an Address Family Identifier (AFI) of 2 and a Subsequent Address Family Identifier (SAFI) of 1 identify IPv6 unicast routing information carried in the MP_REACH_NLRI attribute. For address family values, see RFC 1700.

Exchanging IPv4 and IPv6 routes in both IPv4 and IPv6 address families

MP-BGP supports IPv4 route exchange between IPv6 peers and IPv6 route exchange between IPv4 peers as follows:

- When the next hop of an IPv6 route is an IPv4 address, MP-BGP maps the IPv4 address to an IPv6 address encapsulated in the NEXT_HOP attribute of update messages. In this scenario, you must specify a routing policy to change the next hop of the IPv6 route to the IPv6 address of the peer.
- When the next hop of an IPv4 route is an IPv6 address, BGP negotiates the extended next hop encoding capability with its peer. Then, BGP encapsulates the IPv4 NLRI in the MP_REACH_NLRI attribute of update messages. In this scenario, you must specify a routing policy to change the next hop of the IPv4 route to the IPv4 address of the peer.

Figure 20 Exchanging IPv4 and IPv6 routes in both IPv4 and IPv6 address families



As shown in Figure 20, an IPv6 BGP peer relationship is established between Device A and Device B, between Device B and Device C, and between Device C and Device D. An IPv4 BGP peer relationship is established between Device A and Device B and between Device C and Device D. Device A and Device D can learn both IPv4 and IPv6 routes from each other and traffic is forwarded correctly in both IPv4 and IPv6 address families. For Device C to correctly receive IPv4 routes using the IPv6 address of Device B as the next hop, configure a routing policy on Device C. Use the routing policy to change the next hop of these routes to the IPv4 address of Device B.

BGP route redistribution into local VPN instances

In a private BGP network, BGP sessions might transmit routes that belong to different VPN instances. On receipt of those VPN routes, BGP needs to redistribute them into local VPN instances. During the route redistribution, BGP redistributes routes of non-local VPN instances or routes of another local VPN instance to the BGP routing table of a local VPN instance. Route redistribution into local VPN instances can be divided into local route redistribution and remote route redistribution based on the origins of routes:

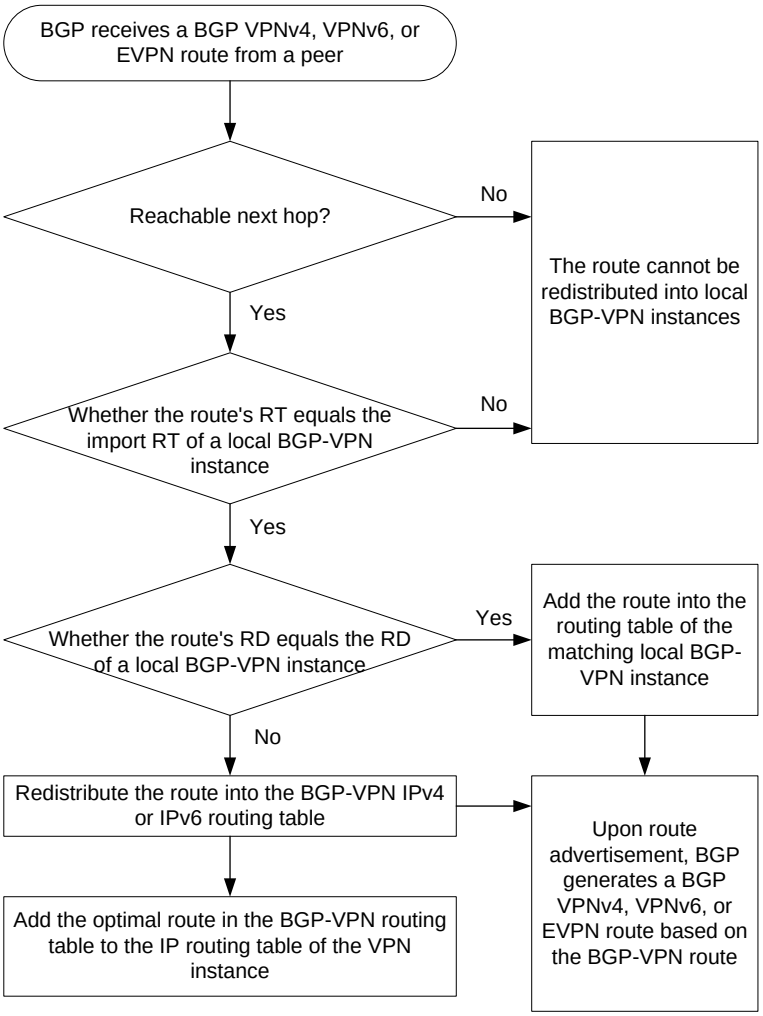
- **Local route redistribution**—Redistributes the BGP routes of a local BGP-VPN instance to another local BGP-VPN instance. BGP VPN routes can be redistributed from a local VPN instance to another only if their export RTs equal the import RT of the destination VPN instance.
- **Remote route redistribution**—Redistributes the BGP routes received from remote peers into local VPN instances. BGP VPN routes can be redistributed into a local VPN instance only if their RTs equal the import RT of that VPN instance. If a BGP VPN route matches multiple local VPN instances, BGP will copy the route and redistribute it into those matching VPN instances separately.

In the current software version, only the following routes support route redistribution between VPN instances:

- **Local route redistribution**: BGP-VPN IPv4 unicast routes and BGP-VPN IPv6 unicast routes.
- **Remote route redistribution**: BGP VPNv4 routes, BGP EVPN routes, and BGP VPNv6 routes.

After BGP receives a BGP VPNv4, VPNv6, or EVPN route from a peer, it redistributes that route into local VPN instances as shown in Figure 21.

Figure 21 Remote route redistribution



BGP multi-instance

About this feature

A BGP router can run multiple BGP processes. Each BGP process corresponds to a BGP instance. BGP maintains an independent routing table for each BGP instance.

When you create multiple BGP instances, follow these guidelines:

- You can create multiple public address families for a BGP instance. However, each public address family (except for IPv4 unicast address family, IPv6 unicast address family, VPNv4 address family, and IPv4 RT-filter address family) can belong to only one BGP instance.
- You can create multiple VPN instances for a BGP instance, and each VPN instance can have multiple address families. A VPN instance can belong to only one BGP instance.
- The IPv4 and IPv6 multicast address families must belong to the same BGP instance.
- Different BGP instances can have the same AS number but cannot have the same name.

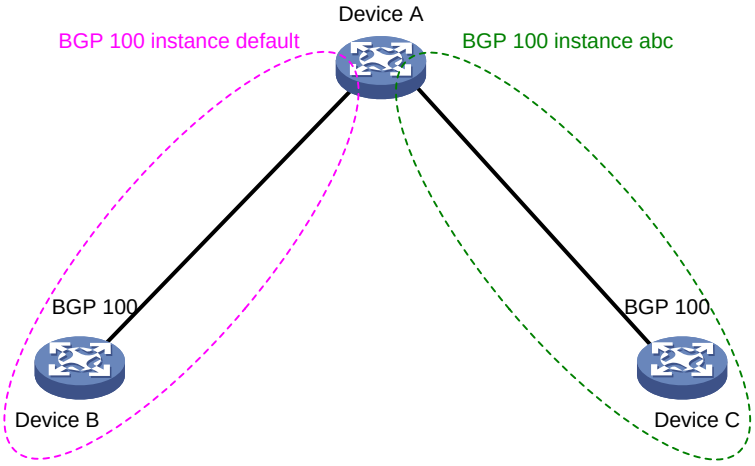
Examples

As shown in Figure 22, Device A, Device B, and Device C all run in AS 100. To enable Device A to communicate with Device B and Device C and implement service isolation between Device B and Device C, perform the following task on Device A:

1. Create two BGP instances, for example, default and abc.
2. Use BGP instance **default** to establish a session to Device B.
3. Use BGP instance **abc** to establish a session to Device C.

After you complete this task, Device A can learn routes from Device B and Device C. However, Device B and Device C cannot learn routes from each other, because BGP instance **default** and **abc** are two independent instances.

Figure 22 Service isolation with BGP multi-instance



BGP configuration views

BGP uses different views to manage routing information for different BGP instances, VPN instances, and address families. Most BGP commands are available in all BGP views. BGP supports multiple VPN instances by establishing a separate routing table for each VPN instance.

If a command is supported only in BGP instance view, this command takes effect for both the public instance and all VPN instances in the BGP instance.

If a command is supported both in BGP instance view and BGP-VPN instance view, this command takes effect as follows:

- In BGP instance view, this command takes effect only for the public instance in the BGP instance.
- In BGP-VPN instance view, this command takes effect only for VPN instances associated with in the BGP-VPN instance.

Table 8 describes different BGP configuration views.

Table 8 BGP configuration views

View names	Ways to enter the views	Remarks

BGP instance view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc]`
		You can create a BGP instance and enter its view by specifying the **`instance`** keyword in the **`bgp`** command.

View names	Ways to enter the views	Remarks
		Configurations in this view apply to all public address families for the specified BGP instance. Some configurations (such as confederation, GR, and logging configurations) also apply to the address families of VPN instances.

BGP IPv4 unicast address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family ipv4 unicast`
	`[Sysname-bgp-abc-ipv4]`
		Configurations in this view apply to public IPv4 unicast routes and peers of the specified BGP instance.

BGP IPv6 unicast address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family ipv6 unicast`
	`[Sysname-bgp-abc-ipv6]`
		Configurations in this view apply to public IPv6 unicast routes and peers of the specified BGP instance.

BGP IPv4 multicast address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family ipv4 multicast`
	`[Sysname-bgp-abc-mul-ipv4]`
		Configurations in this view apply to IPv4 multicast routes and peers of the specified BGP instance.

BGP IPv6 multicast address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family ipv6 multicast`
	`[Sysname-bgp-abc-mul-ipv6]`
		Configurations in this view apply to IPv6 multicast routes and peers of the specified BGP instance.

BGP VPNv4 address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family vpnv4`
	`[Sysname-bgp-abc-vpnv4]`
		Configurations in this view apply to VPNv4 routes and peers of the specified BGP instance.
		For more information about BGP VPNv4 address family view, see *MPLS Configuration Guide*.

BGP L2VPN address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family l2vpn`
	`[Sysname-bgp-abc-l2vpn]`
		Configurations in this view apply to L2VPN information and L2VPN peers of the specified BGP instance.
		For more information about BGP L2VPN address family view, see *MPLS Configuration Guide*.

BGP EVPN address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family l2vpn evpn`
	`[Sysname-bgp-abc-evpn]`
		Configurations in this view apply to EVPN routes and peers of the specified BGP instance.
		For more information about BGP EVPN address family view, see *EVPN Configuration Guide*.

BGP-VPN instance view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] ip vpn-instance vpn1`
	`[Sysname-bgp-abc-vpn1]`
		Configurations in this view apply to all address families in the specified VPN instance of the specified BGP instance.
		For more information about VPN instances, see MPLS L3VPN configuration in *MPLS Configuration*

View names	Ways to enter the views	Remarks
		Guide.

BGP-VPN IPv4 unicast address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] ip vpn-instance vpn1`
	`[Sysname-bgp-abc-vpn1] address-family ipv4 unicast`
	`[Sysname-bgp-abc-ipv4-vpn1]`
		Configurations in this view apply to IPv4 unicast routes and peers in the specified VPN instance of the specified BGP instance.
		For more information about VPN instances, see MPLS L3VPN configuration in *MPLS Configuration Guide*.

BGP-VPN IPv6 unicast address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] ip vpn-instance vpn1`
	`[Sysname-bgp-abc-vpn1] address-family ipv6 unicast`
	`[Sysname-bgp-abc-ipv6-vpn1]`
		Configurations in this view apply to IPv6 unicast routes and peers in the specified VPN instance of the specified BGP instance.
		For more information about VPN instances, see MPLS L3VPN configuration in *MPLS Configuration Guide*.

BGP-VPN VPNv4 address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] ip vpn-instance vpn1`
	`[Sysname-bgp-abc-vpn1] address-family vpnv4`
	`[Sysname-bgp-abc-vpnv4-vpn1]`
		Configurations in this view apply to VPNv4 routes and peers in the specified VPN instance of the specified BGP instance.
		For more information about BGP-VPN VPNv4 address family view, see *MPLS Configuration Guide*.

BGP LS address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family link-state`
	`[Sysname-bgp-abc-ls]`
		Configurations in this view apply to LS messages and peers of the specified BGP instance.

BGP IPv4 RT filter address family view	`<Sysname> system-view`
	`[Sysname] bgp 100 instance abc`
	`[Sysname-bgp-abc] address-family ipv4 rtfilter`
	`[Sysname-bgp-abc-rtf-ipv4]`
		Configurations in this view apply to IPv4 RT filter routes and peers of the specified BGP instance.
		For more information about BGP IPv4 RT filter address family view, see *MPLS L3VPN Configuration Guide*.

Protocols and standards

- RFC 1700, *ASSIGNED NUMBERS*
- RFC 1997, *BGP Communities Attribute*
- RFC 2439, *BGP Route Flap Damping*
- RFC 2545, *Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing*
- RFC 2918, *Route Refresh Capability for BGP-4*
- RFC 4271, *A Border Gateway Protocol 4 (BGP-4)*
- RFC 4275, *BGP-4 MIB Implementation Survey*

- RFC 4277, *Experience with the BGP-4 Protocol*
- RFC 4360, *BGP Extended Communities Attribute*
- RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*
- RFC 4382, *MPLS/BGP Layer 3 Virtual Private Network (VPN) Management*
- RFC 4451, *BGP MULTI_EXIT_DISC (MED) Consideration*
- RFC 4456, *BGP Route Reflection: An Alternative to Full Mesh Internal BGP*
- RFC 4486, *Subcodes for BGP Cease Notification Message*
- RFC 4659, *BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN*
- RFC 4684, *Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)*
- RFC 4724, *Graceful Restart Mechanism for BGP*
- RFC 4760, *Multiprotocol Extensions for BGP-4*
- RFC 4781, *Graceful Restart Mechanism for BGP with MPLS*
- RFC 5004, *Avoid BGP Best Path Transitions from One External to Another*
- RFC 5065, *Autonomous System Confederations for BGP*
- RFC 5082, *The Generalized TTL Security Mechanism (GTSM)*
- RFC 5291, *Outbound Route Filtering Capability for BGP-4*
- RFC 5292, *Address-Prefix-Based Outbound Route Filter for BGP-4*
- RFC 5549, *Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop*
- RFC 5492, *Capabilities Advertisement with BGP-4*
- RFC 5668, *4-Octet AS Specific BGP Extended Community*
- RFC 6037, *Cisco Systems' Solution for Multicast in BGP MPLS IP VPNs*
- RFC 6198, *Requirements for the Graceful Shutdown of BGP Sessions*
- RFC 6368, *Internal BGP as the Provider/Customer Edge Protocol for BGP/MPLS IP*
- RFC 6513, *Multicast in MPLS/BGP IP VPNs*
- RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs*
- RFC 6793, *BGP Support for Four-Octet Autonomous System (AS) Number Space*
- RFC 7432, *BGP MPLS-Based Ethernet VPN*
- RFC 7752, *North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP*
- RFC 7854, *BGP Monitoring Protocol (BMP)*
- RFC 7911, *Advertisement of Multiple Paths in BGP*
- RFC 7947, *Internet Exchange BGP Route Server*
-

Building basic BGP networks

Basic BGP network configuration tasks at a glance (IPv4 unicast/IPv4 multicast)

To build basic BGP networks for the IPv4 unicast or IPv4 multicast address family, perform the following tasks:

1. Configuring basic BGP
   a. Enabling BGP
   b. Configuring a BGP peer
   c. Configuring dynamic BGP peers
   d. Configuring an IBGP peer group

      Configure BGP peer groups on large-scale BGP networks for easy configuration and maintenance.
   e. Configuring an EBGP peer group

      Configure BGP peer groups on large-scale BGP networks for easy configuration and maintenance.
   f. (Optional.) Specifying the source address of TCP connections
2. Controlling BGP route generation

   Choose the following tasks as needed:
   - Injecting a local network
   - Redistributing IGP routes
   - Redistributing the local network specified in the public instance or a VPN instance
   - (Optional.) Configuring BGP route summarization
   - (Optional.) Advertising a default route to a peer or peer group
3. (Optional.) Controlling BGP route advertisement
   - Advertising optimal routes in the IP routing table

     BGP cannot advertise optimal routes in the IP routing table for IPv4 multicast address family.
   - Enabling prioritized withdrawal of the default route
   - Enabling prioritized withdrawal of specific routes
   - Configuring BGP route distribution filtering policies
   - Enabling first-AS-number check for EBGP route advertisement
   - Setting the BGP route sending rate

     This task applies only to IPv4 unicast routes.
   - Configuring BGP to delay sending route updates on reboot
   - Configuring the interval for sending updates for the same route
   - Configuring BGP route advertisement delay
   - Setting the refresh route advertisement delay timer
   - Configuring a startup policy for BGP route updates
4. (Optional.) Controlling BGP route reception
   - Limiting routes received from a peer or peer group
   - Configuring BGP route reception filtering policies

- Configuring the SoO attribute

5. (Optional.) Configuring BGP timers
   - Configuring the keepalive interval and hold time
   - Setting the session retry timer
   - Setting an update delay for local MPLS labels
6. (Optional.) Configuring BGP logging and notifications
   - Enabling logging for session state changes
   - Enabling logging for BGP route flapping
   - Configuring BGP network management

Basic BGP network configuration tasks at a glance (IPv6 unicast/IPv6 multicast)

To build basic BGP networks for the IPv6 unicast or IPv6 multicast address family, perform the following tasks:

1. Configuring basic BGP
   a. Enabling BGP
   b. Configuring a BGP peer
   c. Configuring dynamic BGP peers
   d. Configuring an IBGP peer group

      Configure BGP peer groups on large-scale BGP networks for easy configuration and maintenance.
   e. Configuring an EBGP peer group

      Configure BGP peer groups on large-scale BGP networks for easy configuration and maintenance.
   f. (Optional.) Specifying the source address of TCP connections
2. Controlling BGP route generation

   Choose the following tasks as needed:
   - Injecting a local network
   - Redistributing IGP routes
   - (Optional.) Configuring BGP route summarization
   - (Optional.) Advertising a default route to a peer or peer group
3. (Optional.) Controlling BGP route advertisement
   - Advertising optimal routes in the IP routing table

     BGP cannot advertise optimal routes in the IP routing table for IPv6 multicast address family.
   - Enabling prioritized withdrawal of the default route
   - Enabling prioritized withdrawal of specific routes
   - Configuring BGP route distribution filtering policies
   - Setting the BGP route sending rate

     This task applies only to IPv6 unicast routes.
   - Configuring BGP to delay sending route updates on reboot
   - Configuring the interval for sending updates for the same route
   - Configuring BGP route advertisement delay

- Setting the refresh route advertisement delay timer
- Configuring a startup policy for BGP route updates

4. (Optional.) Controlling BGP route reception
   - Limiting routes received from a peer or peer group
   - Configuring BGP route reception filtering policies
   - Configuring the SoO attribute
5. (Optional.) Configuring BGP timers
   - Configuring the keepalive interval and hold time
   - Setting the session retry timer
   - Setting an update delay for local MPLS labels
6. (Optional.) Configuring BGP logging and notifications
   - Enabling logging for session state changes
   - Enabling logging for BGP route flapping
   - Configuring BGP network management

Configuring basic BGP

Enabling BGP

Restrictions and guidelines

A router ID is the unique identifier of a BGP router in an AS.

- To ensure the uniqueness of a router ID and enhance availability, specify in BGP instance view the IP address of a local loopback interface as the router ID. Different BGP instances can have the same router ID.
- If no router ID is specified in BGP instance view, the global router ID is used.
- To modify a non-zero router ID of a BGP instance, use the **`router-id`** command in BGP instance view, rather than the **`router id`** command in system view.
- If you specify a router ID in BGP instance view and then remove the interface that owns the router ID, the router does not select a new router ID. To select a new router ID, use the **`undo router-id`** command in BGP instance view.

Procedure

1. Enter system view.

   `system-view`
2. Configure a global router ID.

   `router id` *`router-id`*

   By default, no global router ID is configured.

   If no global router ID is configured, the following rules apply:
   - If loopback interfaces configured with an IP address exist, BGP uses the highest loopback interface IP address as the router ID.
   - If no loopback interface IP address is available, BGP uses the highest physical interface IP address as the route ID regardless of the interface status.
3. Enable BGP and enter BGP instance view.

   `bgp` *`as-number`* [**`instance`** *`instance-name`*]

   By default, BGP is disabled and no BGP instances exist.
4. (Optional.) Configure a router ID for the BGP instance.

router-id *router-id*

By default, no router ID is configured for a BGP instance, and the BGP instance uses the global router ID configured by the **router-id** command in system view.

5. (Optional.) Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*

   The specified VPN instance must have been created and have an RD.

6. (Optional.) Configure a router ID for the BGP VPN instance.

   router-id { *router-id* | **auto-select** }

   By default, no router ID is configured for a BGP VPN instance.

   The BGP VPN instance uses the router ID configured in BGP instance view. If no router ID is configured in BGP instance view, the BGP VPN instance uses the global router ID configured in system view.

Configuring a BGP peer

Restrictions and guidelines

You can configure BGP peers for the local device by using one of the following methods:

- Specify a BGP peer by its global unicast address.
- Specify a BGP peer by its link-local address. A BGP peer at an IPv6 link-local address must be directly connected to the local router. If you do not specify the interface connected to the peer, you must specify the interface directly connected to the BGP peer as the source interface of TCP connections on the local router.
- Specify a local interface and establish peer relationships with link-local addresses that the interface learned in ND entries.

To exchange IPv4 routes with an IPv6 peer or exchange IPv6 routes with an IPv4 peer, you must configure a routing policy to perform the following tasks:

- Change the next hop of IPv4 routes received from the IPv6 peer to the IPv4 address of the interface that connects the IPv6 peer to the local router.
- Change the next hop of IPv6 routes received from the IPv4 peer to the IPv6 address of the interface that connects the IPv4 peer to the local router.

Procedure (Exchanging IPv4 unicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an IPv4 BGP peer and specify its AS number.

   peer *ipv4-address* **as-number** *as-number*

4. (Optional.) Configure a description for a peer.

   peer *ipv4-address* **description** *text*

   By default, no description is configured for a peer.

5. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

address-family ipv4 [**unicast**]

6. Enable the router to exchange IPv4 unicast routing information with the specified peer.
   peer *ipv4-address* **enable**
   By default, the router cannot exchange IPv4 unicast routing information with the peer.

Procedure (Exchanging IPv6 unicast routes with IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Create an IPv4 BGP peer and specify its AS number.
   peer *ipv4-address* **as-number** *as-number*
4. (Optional.) Configure a description for the IPv4 peer.
   peer *ipv4-address* **description** *text*
   By default, no description is configured for an IPv4 peer.
5. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.
   address-family ipv6 [**unicast**]
6. Enable BGP to exchange IPv6 unicast routing information with the IPv4 peer.
   peer *ipv4-address* **enable**
   By default, BGP cannot exchange IPv6 unicast routing information with an IPv4 peer.
7. Use a routing policy to modify the next hop of routes received from the IPv4 peer.
   peer *ipv4-address* **route-policy** *route-policy-name* **import**

Procedure (Exchanging IPv6 unicast routes with IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Create an IPv6 BGP peer and specify its AS number.
   peer { *ipv6-address* | *link-local-address* **interface** *interface-type interface-number* } **as-number** *as-number*
4. (Optional.) Configure a description for a peer.
   peer { *ipv6-address* | *link-local-address* **interface** *interface-type interface-number* } **description** *text*
   By default, no description is configured for a peer.
5. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

address-family ipv6 [**unicast**]

6. Enable the router to exchange IPv6 unicast routing information with the specified peer.

   peer { *ipv6-address* | *link-local-address* **interface** *interface-type interface-number* } **enable**

   By default, the router cannot exchange IPv6 unicast routing information with the peer.

Procedure (Exchanging IPv6 unicast routes with IPv6 peers created by specifying an interface)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create peer relationships with link-local addresses that the specified interface learned in ND entries, and specify an AS number for the peers.

   interface-peer *interface-type interface-number* **as-number** *as-number*

4. (Optional.) Configure a description for the peers.

   interface-peer *interface-type interface-number* **description** *text*

   By default, no description is configured for a peer.

5. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

6. Enable the router to exchange IPv6 unicast routing information with the specified peers.

   interface-peer *interface-type interface-number* **enable**

   By default, the router cannot exchange IPv6 unicast routing information with a peer.

Procedure (Exchanging IPv4 unicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an IPv6 BGP peer and specify its AS number.

   peer *ipv6-address* **as-number** *as-number*

4. (Optional.) Configure a description for the IPv6 peer.

   peer *ipv6-address* **description** *text*

   By default, no description is configured for an IPv6 peer.

5. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

6. Enable BGP to exchange IPv4 unicast routing information with the IPv6 peer.

   peer *ipv6-address* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information with an IPv6 peer.

7. Use a routing policy to modify the next hop of routes received from the IPv6 peer.

   peer *ipv6-address* **route-policy** *route-policy-name* **import**

Procedure (Exchanging IPv4 unicast routes with IPv6 peers created by specifying an interface)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create peer relationships with link-local addresses that the specified interface learned in ND entries, and specify an AS number for the peers.

   interface-peer *interface-type interface-number* **as-number** *as-number*

4. (Optional.) Configure a description for the IPv6 peers.

   interface-peer *interface-type interface-number* **description** *text*

   By default, no description is configured for an IPv6 peer.

5. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

6. Enable BGP to exchange IPv4 unicast routing information with the specified peers.

   interface-peer *interface-type interface-number* **enable**

   By default, the router cannot exchange IPv4 unicast routing information with a peer.

7. Use a routing policy to modify the next hop of routes received from the IPv6 peer.

   interface-peer *interface-type interface-number* **route-policy** *route-policy-name* **import**

Procedure (Exchanging IPv4 multicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Create an IPv4 BGP peer and specify its AS number.

   peer *ipv4-address* **as-number** *as-number*

4. (Optional.) Configure a description for the peer.

   peer *ipv4-address* **description** *text*

   By default, no description is configured for a peer.

5. Create the BGP IPv4 multicast address family and enter its view.

   address-family ipv4 multicast

6. Enable the router to exchange IPv4 unicast routing information used for RPF check with the specified peer.

peer *ipv4-address* **enable**

By default, the router cannot exchange IPv4 unicast routing information used for RPF check with the peer.

Procedure (Exchanging IPv6 multicast routes with IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Create an IPv6 BGP peer and specify its AS number.

   peer *ipv6-address* **as-number** *as-number*
4. (Optional.) Configure a description for the peer.

   peer *ipv6-address* **description** *text*

   By default, no description is configured for a peer.
5. Create the BGP IPv6 multicast address family and enter its view.

   address-family ipv6 multicast
6. Enable the router to exchange IPv6 unicast routing information used for RPF check with the specified peer.

   peer *ipv6-address* **enable**

   By default, the router cannot exchange IPv6 unicast routing information used for RPF check with the peer.

Configuring dynamic BGP peers

About this task

This feature enables BGP to establish dynamic BGP peer relationships with devices in a network. BGP accepts connection requests from the network but it does not initiate connection requests to the network.

After a device in the network initiates a connection request, BGP establishes a dynamic peer relationship with the device.

If multiple BGP peers reside in the same network, you can use this feature to simplify BGP peer configuration.

If BGP establishes a large number of dynamic peer sessions, these sessions might consume system resources excessively. To control the number of dynamic BGP peer sessions, use the **dynamic-peer-limit** command to set an upper limit on the number of dynamic peer sessions. BGP will not establish new dynamic peer sessions if the number of established dynamic peer sessions exceeds the specified limit.

Restrictions and guidelines

For a remote device to establish a peer relationship with the local device, you must specify the IP address of the local device on the remote device.

A BGP peer at an IPv6 link-local address must be directly connected to the local router. If you do not specify the interface connected to the peer, you must specify the interface directly connected to the BGP peer as the source interface of TCP connections on the local router.

To exchange IPv4 routes with an IPv6 peer or exchange IPv6 routes with an IPv4 peer, you must configure a routing policy to perform the following tasks:

- Change the next hop of IPv4 routes received from the IPv6 peer to the IPv4 address of the interface that connects the IPv6 peer to the local router.

- Change the next hop of IPv6 routes received from the IPv4 peer to the IPv6 address of the interface that connects the IPv4 peer to the local router.

When you configure the **dynamic-peer-limit** command, the number of established dynamic peer sessions might already exceed the specified limit. In this scenario, BGP does not terminate the existing dynamic peer sessions. However, if BGP resets and re-establishes peer sessions later, the number of established dynamic peer sessions cannot exceed the specified limit.

When you use the **dynamic-peer-limit** command in a scenario where BGP needs to re-establish peer sessions, make sure the specified limit is not smaller than the number of existing dynamic peer sessions. Otherwise, traffic interruption will occur because BGP cannot re-establish some of the dynamic peer sessions.

If the local device has established a dynamic peer relationship with an IP address through network segment matching, you cannot configure that IP address as a static peer. Before you configure an IP address as a static peer, make sure that IP address is not in the network segment to which any existing dynamic peer belongs. If the IP address belongs to a predefined dynamic-peer network segment, create that network segment as a dynamic peer only after you finish configuring the IP address as a static peer. For example, if you configure network segment 10.1.1.0/24 as a dynamic peer and the local device establishes a BGP connection with 10.1.1.1, you cannot configure 10.1.1.1 as a static peer. In contrast, if you first configure 10.1.1.1 as a static peer, you can continue to configure network segment 10.1.1.0/24 as a dynamic peer.

Exchanging IPv4 unicast routes with dynamic IPv4 peers

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Specify devices in a network as a dynamic BGP peer and specify one or multiple AS numbers for the peer. Choose one of the following options:
   - Specify a single AS number for the dynamic peer.

     peer *ipv4-address* *mask-length* **as-number** *as-number*
   - Specify multiple AS numbers for the dynamic peer.

     peer *ipv4-address* *mask-length* **listen-as** *as-number-list*
   - Specify an AS number range for the dynamic peer.

     peer *ipv4-address* *mask-length* **listen-as-segment begin-as** *as-number* **end-as** *as-number*
4. (Optional.) Configure a description for dynamic BGP peers.

   peer *ipv4-address* *mask-length* **description** *text*

   By default, no description is configured for dynamic BGP peers.
5. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]
6. Enable BGP to exchange IPv4 unicast routing information with dynamic BGP peers in the specified network.

   peer *ipv4-address mask-length* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information with dynamic BGP peers.

Exchanging IPv6 unicast routes with dynamic IPv4 peers

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Specify devices in an IPv4 network as a dynamic IPv4 BGP peer and specify one or multiple AS numbers for the peer. Choose one of the following options:
   - Specify a single AS number for the dynamic peer.
     peer *ipv4-address mask-length* **as-number** *as-number*
   - Specify multiple AS numbers for the dynamic peer.
     peer *ipv4-address mask-length* **listen-as** *as-number-list*
   - Specify an AS number range for the dynamic peer.
     peer *ipv4-address mask-length* **listen-as-segment begin-as** *as-number* **end-as** *as-number*
4. (Optional.) Configure a description for the dynamic IPv4 peers.
   peer *ipv4-address mask-length* **description** *text*
   By default, no description is configured for dynamic IPv4 peers.
5. Create the BGP IPv6 unicast address family and enter its view.
   address-family ipv6 [**unicast**]
6. Enable BGP to exchange IPv6 unicast routing information with the dynamic IPv4 peers.
   peer *ipv4-address mask-length* **enable**
   By default, BGP cannot exchange IPv6 unicast routing information with dynamic IPv4 peers.
7. Use a routing policy to modify the next hop of routes received from the dynamic IPv4 peers.
   peer *ipv4-address mask-length* **route-policy** *route-policy-name* **import**

Exchanging IPv6 unicast routes with dynamic IPv6 peers

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Specify devices in a network as a dynamic BGP peer and specify one or multiple AS numbers for the peer. Choose one of the following options:
   - Specify a single AS number for the dynamic peer.
     peer { *ipv6-address prefix-length* | *link-local-address* **interface** *interface-type interface-number* } **as-number** *as-number*
   - Specify multiple AS numbers for the dynamic peer.
     peer *ipv6-address prefix-length* **listen-as** *as-number-list*

- Specify an AS number range for the dynamic peer.

  peer *ipv6-address prefix-length* **listen-as-segment begin-as** *as-number* **end-as** *as-number*

4. (Optional.) Configure a description for dynamic BGP peers.

   peer { *ipv6-address prefix-length* | *link-local-address* **interface** *interface-type interface-number* } **description** *text*

   By default, no description is configured for dynamic BGP peers.

5. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

6. Enable BGP to exchange IPv6 unicast routing information with dynamic BGP peers in the specified network.

   peer { *ipv6-address prefix-length* | *link-local-address* **interface** *interface-type interface-number* } **enable**

   By default, BGP cannot exchange IPv6 unicast routing information with dynamic BGP peers.

Exchanging IPv4 unicast routes with dynamic IPv6 peers

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Specify devices in an IPv6 network as a dynamic IPv6 BGP peer and specify one or multiple AS numbers for the peer. Choose one of the following options:
   - Specify a single AS number for the dynamic peer.

     peer *ipv6-address prefix-length* **as-number** *as-number*
   - Specify multiple AS numbers for the dynamic peer.

     peer *ipv6-address prefix-length* **listen-as** *as-number-list*
   - Specify an AS number range for the dynamic peer.

     peer *ipv6-address prefix-length* **listen-as-segment begin-as** *as-number* **end-as** *as-number*

4. (Optional.) Configure a description for the dynamic IPv6 peers.

   peer *ipv6-address prefix-length* **description** *text*

   By default, no description is configured for dynamic IPv6 peers.

5. Create the BGP IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

6. Enable BGP to exchange IPv4 unicast routing information with the dynamic IPv6 peers.

   peer *ipv6-address prefix-length* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information with dynamic IPv6 peers.

7. Use a routing policy to modify the next hop of routes received from the dynamic IPv6 peers.

   peer *ipv6-address mask-length* **route-policy** *route-policy-name* **import**

Exchanging IPv4 multicast routes with dynamic IPv4 peers

1. Enter system view.

system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Specify devices in a network as a dynamic BGP peer and specify one or multiple AS numbers for the peer. Choose one of the following options:
   - Specify a single AS number for the dynamic peer.

     peer *ipv4-address mask-length* **as-number** *as-number*
   - Specify multiple AS numbers for the dynamic peer.

     peer *ipv4-address mask-length* **listen-as** *as-number-list*
   - Specify an AS number range for the dynamic peer.

     peer *ipv4-address mask-length* **listen-as-segment begin-as** *as-number* **end-as** *as-number*

4. (Optional.) Configure a description for dynamic BGP peers.

   peer *ipv4-address mask-length* **description** *text*

   By default, no description is configured for dynamic BGP peers.

5. Create the BGP IPv4 multicast address family and enter its view.

   address-family ipv4 multicast

6. Enable BGP to exchange IPv4 unicast routing information used for RPF check with dynamic BGP peers in the specified network.

   peer *ipv4-address mask-length* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information used for RPF check with dynamic BGP peers.

Exchanging IPv6 multicast routes with dynamic IPv6 peers

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Specify devices in a network as a dynamic BGP peer and specify one or multiple AS numbers for the peer. Choose one of the following options:
   - Specify a single AS number for the dynamic peer.

     peer *ipv6-address prefix-length* **as-number** *as-number*
   - Specify multiple AS numbers for the dynamic peer.

     peer *ipv6-address prefix-length* **listen-as** *as-number-list*
   - Specify an AS number range for the dynamic peer.

     peer *ipv6-address prefix-length* **listen-as-segment begin-as** *as-number* **end-as** *as-number*

4. (Optional.) Configure a description for dynamic BGP peers.

   peer *ipv6-address prefix-length* **description** *text*

   By default, no description is configured for dynamic BGP peers.

5. Create the BGP IPv6 multicast address family and enter its view.

   address-family ipv6 multicast

6. Enable BGP to exchange IPv6 unicast routing information used for RPF check with dynamic BGP peers in the specified network.

   peer *ipv6-address prefix-length* **enable**

By default, BGP cannot exchange IPv6 unicast routing information used for RPF check with dynamic BGP peers.

Setting the maximum number of dynamic BGP peer sessions that BGP can establish

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Configure the maximum number of dynamic BGP peer sessions that BGP can establish.
   dynamic-peer-limit *max-peer-number*

   By default, the number of dynamic BGP peer sessions is not limited.

   If you execute this command multiple times for a BGP instance, the most recent configuration takes effect.

Configuring an IBGP peer group

About this task

A peer group is an IBGP peer group if peers in it belong to the same AS as the local router.

After you create an IBGP peer group and then add a peer into it, the system creates the peer in BGP instance view and specifies the local AS number for the peer.

Restrictions and guidelines

A BGP peer at an IPv6 link-local address must be directly connected to the local router. If you do not specify the interface connected to the peer, you must specify the interface directly connected to the BGP peer as the source interface of TCP connections on the local router.

To exchange IPv4 routes with an IPv6 peer or exchange IPv6 routes with an IPv4 peer, you must configure a routing policy to perform the following tasks:

- Change the next hop of IPv4 routes received from the IPv6 peer to the IPv4 address of the interface that connects the IPv6 peer to the local router.
- Change the next hop of IPv6 routes received from the IPv4 peer to the IPv6 address of the interface that connects the IPv4 peer to the local router.

If you configure a BGP setting at both the peer group and the peer level, the most recent configuration takes effect on the peer.

Procedure (Exchanging IPv4 unicast routes with IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Create an IBGP peer group.
   group *group-name* [**internal**]
4. Add a peer into the IBGP peer group.
   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

The **as-number** *as-number* option must specify the local AS number.

5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

6. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

7. Enable the router to exchange IPv4 unicast routing information with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv4 unicast routing information with the peers.

Procedure (Exchanging IPv6 unicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.

   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]

   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an IBGP peer group.

   group *group-name* [**internal**]

4. Add an IPv4 peer into the IBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the local AS number.

5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.

6. Create the BGP IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

7. Enable BGP to exchange IPv6 unicast routing information with peers in the peer group.

   peer *group-name* **enable**

   By default, BGP cannot exchange IPv6 unicast routing information with peers in a peer group.

8. Use a routing policy to modify the next hop of routes received from peers in the peer group.

   peer *group-name* **route-policy** *route-policy-name* **import**

Procedure (Exchanging IPv4 multicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Create an IBGP peer group.

   group *group-name* [**internal**]

4. Add an IPv4 peer into the IBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the local AS number.

5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

6. Create the BGP IPv4 multicast address family and enter its view.

   address-family ipv4 multicast

7. Enable the router to exchange IPv4 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv4 unicast routing information used for RPF check with the peers in the peer group.

Procedure (Exchanging IPv6 unicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an IBGP peer group.

   group *group-name* [**internal**]

4. Add a peer into the IBGP peer group. Choose the options to configure as needed:
   - Specify a peer by its address.

     peer { *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **group** *group-name* [**as-number** *as-number*]
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

     interface-peer *interface-type interface-number* **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the local AS number.

5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

6. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

7. Enable the router to exchange IPv6 unicast routing information with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information with the peers.

Procedure (Exchanging IPv4 unicast routes with IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Create an IBGP peer group.
   group *group-name* [**internal**]
4. Add an IPv6 peer into the IBGP peer group. Choose the options to configure as needed:
   - Specify a peer by its address.
     peer *ipv6-address* [*prefix-length*] **group** *group-name* [**as-number** *as-number*]
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.
     interface-peer *interface-type interface-number* **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the local AS number.
5. (Optional.) Configure a description for the peer group.
   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.
6. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.
   address-family ipv4 [**unicast**]
7. Enable BGP to exchange IPv4 unicast routing information with peers in the peer group.
   peer *group-name* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information with peers in a peer group.
8. Use a routing policy to modify the next hop of routes received from peers in the peer group.
   peer *group-name* **route-policy** *route-policy-name* **import**

Procedure (Exchanging IPv6 multicast routes with IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Create an IBGP peer group.
   group *group-name* [**internal**]
4. Add a peer into the IBGP peer group.
   peer *ipv6-address* [*prefix-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the local AS number.
5. (Optional.) Configure a description for the peer group.
   peer *group-name* **description** *text*

By default, no description is configured for the peer group.

6. Create the BGP IPv6 multicast address family and enter its view.

   address-family ipv6 multicast

7. Enable the router to exchange IPv6 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information used for RPF check with the peers in the peer group.

Configuring an EBGP peer group

About this task

A peer group is an EBGP peer group if peers in it belong to different ASs.

If peers in an EBGP group belong to the same external AS, the EBGP peer group is a pure EBGP peer group. If not, it is a mixed EBGP peer group.

Restrictions and guidelines

Use one of the following methods to configure an EBGP peer group:

- **Method 1**—Create an EBGP peer group, specify its AS number, and add peers into it. All the added peers have the same AS number. All peers in the peer group have the same AS number as the peer group. You can specify an AS number for a peer before adding it into the peer group. The AS number must be the same as that of the peer group.
- **Method 2**—Create an EBGP peer group, specify an AS number for a peer, and add the peer into the peer group. Peers added in the group can have different AS numbers.
- **Method 3**—Create an EBGP peer group and add a peer with an AS number into it. Peers added in the group can have different AS numbers.

To exchange IPv4 routes with an IPv6 peer or exchange IPv6 routes with an IPv4 peer, you must configure a routing policy to perform the following tasks:

- Change the next hop of IPv4 routes received from the IPv6 peer to the IPv4 address of the interface that connects the IPv6 peer to the local router.
- Change the next hop of IPv6 routes received from the IPv4 peer to the IPv6 address of the interface that connects the IPv4 peer to the local router.

If you configure a BGP setting at both the peer group and the peer level, the most recent configuration takes effect on the peer.

Configuring an EBGP peer group by using Method 1 (Exchanging IPv4 unicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an EBGP peer group.

   group *group-name* **external**

4. Specify the AS number of the group.

peer *group-name* **as-number** *as-number*

By default, no AS number is specified.

If a peer group contains peers, you cannot remove or change its AS number.

5. Add a peer into the EBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the same AS number as the **peer** *group-name* **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

7. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

8. Enable the router to exchange IPv4 unicast routing information with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv4 unicast routing information with the peers.

Configuring an EBGP peer group by using Method 2 (Exchanging IPv4 unicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an EBGP peer group.

   group *group-name* **external**

4. Create an IPv4 BGP peer and specify its AS number.

   peer *ipv4-address* [*mask-length*] **as-number** *as-number*

5. Add the peer into the EBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the same AS number as the **peer** *ipv4-address* [*mask-length*] **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

7. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

8. Enable the router to exchange IPv4 unicast routing information with peers in the specified peer group.

peer *group-name* **enable**

By default, the router cannot exchange IPv4 unicast routing information with the peers.

Configuring an EBGP peer group by using Method 3 (Exchanging IPv4 unicast routes with IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Create an EBGP peer group.

   group *group-name* **external**
4. Add a peer into the EBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* **as-number** *as-number*
5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.
6. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]
7. Enable the router to exchange IPv4 unicast routing information with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv4 unicast routing information with the peers.

Configuring an EBGP peer group by using Method 1 (Exchanging IPv6 unicast routes with IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Create an EBGP peer group.

   group *group-name* **external**
4. Specify an AS number for the peer group.

   peer *group-name* **as-number** *as-number*

   By default, no AS number is specified for a peer group.

   If a peer group contains peers, you cannot remove or change its AS number.
5. Add an IPv4 peer into the EBGP peer group.

peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

The **as-number** *as-number* option must specify the same AS number as the **peer** *group-name* **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.

7. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

8. Enable BGP to exchange IPv6 unicast routing information with peers in the peer group.

   peer *group-name* **enable**

   By default, BGP cannot exchange IPv6 unicast routing information with peers in a peer group.

9. Use a routing policy to modify the next hop of routes received from peers in the peer group.

   peer *group-name* **route-policy** *route-policy-name* **import**

Configuring an EBGP peer group by using Method 2 (Exchanging IPv6 unicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an EBGP peer group.

   group *group-name* **external**

4. Create an IPv4 BGP peer and specify its AS number.

   peer *ipv4-address* [*mask-length*] **as-number** *as-number*

5. Add the IPv4 peer into the EBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the same AS number as the **peer** *ipv4-address* [*mask-length*] **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.

7. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

8. Enable BGP to exchange IPv6 unicast routing information with peers in the peer group.

   peer *group-name* **enable**

   By default, BGP cannot exchange IPv6 unicast routing information with peers in a peer group.

9. Use a routing policy to modify the next hop of routes received from peers in the peer group.

   peer *group-name* **route-policy** *route-policy-name* **import**

Configuring an EBGP peer group by using Method 3 (Exchanging IPv6 unicast routes with IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Create an EBGP peer group.
   group *group-name* **external**
4. Add an IPv4 peer into the EBGP peer group.
   peer *ipv4-address* [*mask-length*] **group** *group-name* **as-number** *as-number*
5. (Optional.) Configure a description for the peer group.
   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.
6. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.
   address-family ipv6 [**unicast**]
7. Enable BGP to exchange IPv6 unicast routing information with peers in the peer group.
   peer *group-name* **enable**

   By default, BGP cannot exchange IPv6 unicast routing information with peers in a peer group.
8. Use a routing policy to modify the next hop of routes received from peers in the peer group.
   peer *group-name* **route-policy** *route-policy-name* **import**

Configuring an EBGP peer group by using Method 1 (Exchanging IPv4 multicast routes with IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Create an EBGP peer group.
   group *group-name* **external**
4. Specify the AS number of the group.
   peer *group-name* **as-number** *as-number*

   By default, no AS number is specified.

   If a peer group contains peers, you cannot remove or change its AS number.
5. Add an IPv4 BGP peer into the EBGP peer group.
   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the same AS number as the **peer** *group-name* **as-number** *as-number* command.
6. (Optional.) Configure a description for the peer group.
   peer *group-name* **description** *text*

By default, no description is configured for the peer group.

7. Create the BGP IPv4 multicast address family and enter its view.

   address-family ipv4 multicast

8. Enable the router to exchange IPv4 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv4 unicast routing information used for RPF check with the peers in the group.

Configuring an EBGP peer group by using Method 2 (Exchanging IPv4 multicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Create an EBGP peer group.

   group *group-name* **external**

4. Create an IPv4 BGP peer and specify its AS number.

   peer *ipv4-address* [*mask-length*] **as-number** *as-number*

5. Add the peer into the EBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the same AS number as the **peer** *ipv4-address* [*mask-length*] **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

7. Create the BGP IPv4 multicast address family and enter its view.

   address-family ipv4 multicast

8. Enable the router to exchange IPv4 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv4 unicast routing information used for RPF check with the peers in the group.

Configuring an EBGP peer group by using Method 3 (Exchanging IPv4 multicast routes with IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Create an EBGP peer group.

   group *group-name* **external**

4. Add an IPv4 BGP peer into the EBGP peer group.

   peer *ipv4-address* [*mask-length*] **group** *group-name* **as-number** *as-number*

5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

By default, no description is configured for the peer group.

6. Create the BGP IPv4 multicast address family and enter its view.

   address-family ipv4 multicast

7. Enable the router to exchange IPv4 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv4 unicast routing information used for RPF check with the peers.

Configuring an EBGP peer group by using Method 1 (Exchanging IPv6 unicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an EBGP peer group.

   group *group-name* **external**

4. Specify the AS number of the group.

   peer *group-name* **as-number** *as-number*

   By default, no AS number is specified.

   If a peer group contains peers, you cannot remove or change its AS number.

5. Add a peer into the EBGP peer group. Choose the options to configure as needed:
   - Specify a peer by its address.

     peer { *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **group** *group-name* [**as-number** *as-number*]
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

     interface-peer *interface-type interface-number* **group** *group-name* [**as-number** *as-number*]

     The **as-number** *as-number* option must specify the same AS number as the **peer** *group-name* **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

7. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

8. Enable the router to exchange IPv6 unicast routing information with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information with the peers.

Configuring an EBGP peer group by using Method 2 (Exchanging IPv6 unicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an EBGP peer group.

   group *group-name* **external**

4. Create an IPv6 BGP peer and specify its AS number. Choose the options to configure as needed:
   - Specify a peer by its address.

     peer { *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **as-number** *as-number*
   - Specify a peer by its directly connected interface. The address of the peer is the link-local address on an ND entry learned by the interface.

     interface-peer *interface-type interface-number* **as-number** *as-number*

5. Add the peer into the EBGP peer group. Choose the options to configure based on the configuration of the above step:
   - Specify a peer by its address.

     peer { *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **group** *group-name* [**as-number** *as-number*]

     The **as-number** *as-number* option must specify the same AS number as the **peer** *ipv6-address* [*prefix-length*] **as-number** *as-number* command.
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

     interface-peer *interface-type interface-number* **group** *group-name* [**as-number** *as-number*]

     The **as-number** *as-number* option must specify the same AS number as the **interface-peer** *interface-type interface-number* **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

7. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

8. Enable the router to exchange IPv6 unicast routing information with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information with the peers.

Configuring an EBGP peer group by using Method 3 (Exchanging IPv6 unicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an EBGP peer group.

   group *group-name* **external**

4. Add the peer into the EBGP peer group. Choose the options to configure as needed:
   - Specify a peer by its address.

     peer { *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **group** *group-name* **as-number** *as-number*
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

     interface-peer *interface-type interface-number* **group** *group-name* [**as-number** *as-number*]

5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

6. Create the BGP IPv6 unicast address family or BGP-VPN IPv6 unicast address family and enter its view.

   address-family ipv6 [**unicast**]

7. Enable the router to exchange IPv6 unicast routing information with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information with the peers.

Configuring an EBGP peer group by using Method 1 (Exchanging IPv4 unicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Create an EBGP peer group.

   group *group-name* **external**

4. Specify an AS number for the peer group.

   peer *group-name* **as-number** *as-number*

By default, no AS number is specified for a peer group.

If a peer group contains peers, you cannot remove or change its AS number.

5. Add an IPv6 peer into the EBGP peer group. Choose the options to configure as needed:
   - Specify a peer by its address.

     peer *ipv6-address* [*prefix-length*] **group** *group-name* [**as-number** *as-number*]
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

     interface-peer *interface-type interface-number* **group** *group-name* [**as-number** *as-number*]

     The **as-number** *as-number* option must specify the same AS number as the **peer** *group-name* **as-number** *as-number* command.
6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.
7. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]
8. Enable BGP to exchange IPv4 unicast routing information with peers in the peer group.

   peer *group-name* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information with peers in a peer group.
9. Use a routing policy to modify the next hop of routes received from peers in the peer group.

   peer *group-name* **route-policy** *route-policy-name* **import**

Configuring an EBGP peer group by using Method 2 (Exchanging IPv4 unicast routes with IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Create an EBGP peer group.

   group *group-name* **external**
4. Create an IPv6 BGP peer and specify its AS number. Choose the options to configure as needed:
   - Specify a peer by its address.

     peer *ipv6-address* [*prefix-length*] **as-number** *as-number*
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

     interface-peer *interface-type interface-number* **as-number** *as-number*
5. Add the IPv6 peer into the EBGP peer group. Choose the options to configure based on the configuration of the above step:

- Specify a peer by its address.

  peer *ipv6-address* [*prefix-length*] **group** *group-name* [**as-number** *as-number*]

  The **as-number** *as-number* option must specify the same AS number as the **peer** *ipv6-address* [*prefix-length*] **as-number** *as-number* command.
- Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

  interface-peer *interface-type interface-number* **group** *group-name* [**as-number** *as-number*]

  The **as-number** *as-number* option must specify the same AS number as the **interface-peer** *interface-type interface-number* **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.
7. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]
8. Enable BGP to exchange IPv4 unicast routing information with peers in the peer group.

   peer *group-name* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information with peers in a peer group.
9. Use a routing policy to modify the next hop of routes received from peers in the peer group.

   peer *group-name* **route-policy** *route-policy-name* **import**

Configuring an EBGP peer group by using Method 3 (Exchanging IPv4 unicast routes with IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Create an EBGP peer group.

   group *group-name* **external**
4. Add an IPv6 peer into the EBGP peer group. Choose the options to configure as needed:
   - Specify a peer by its address.

     peer *ipv6-address* [*prefix-length*] **group** *group-name* **as-number** *as-number*
   - Specify peers that are established through link-local addresses that the specified interface learned in ND entries.

     interface-peer *interface-type interface-number* **group** *group-name* **as-number** *as-number*
5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for a peer group.

6. Create the BGP IPv4 unicast address family or BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

7. Enable BGP to exchange IPv4 unicast routing information with peers in the peer group.

   peer *group-name* **enable**

   By default, BGP cannot exchange IPv4 unicast routing information with peers in a peer group.

8. Use a routing policy to modify the next hop of routes received from peers in the peer group.

   peer *group-name* **route-policy** *route-policy-name* **import**

Configuring an EBGP peer group by using Method 1 (Exchanging IPv6 multicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Create an EBGP peer group.

   group *group-name* **external**

4. Specify the AS number of the group.

   peer *group-name* **as-number** *as-number*

   By default, no AS number is specified.

   If a peer group contains peers, you cannot remove or change its AS number.

5. Add an IPv6 BGP peer into the EBGP peer group.

   peer *ipv6-address* [*prefix-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the same AS number as the **peer** *group-name* **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

7. Create the BGP IPv6 multicast address family and enter its view.

   address-family ipv6 multicast

8. Enable the router to exchange IPv6 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information used for RPF check with the peers in the group.

Configuring an EBGP peer group by using Method 2 (Exchanging IPv6 multicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Create an EBGP peer group.

   group *group-name* **external**

4. Create an IPv6 BGP peer and specify its AS number.

peer *ipv6-address* [*prefix-length*] **as-number** *as-number*

5. Add the peer into the EBGP peer group.

   peer *ipv6-address* [*prefix-length*] **group** *group-name* [**as-number** *as-number*]

   The **as-number** *as-number* option must specify the same AS number as the **peer** *ipv6-address* [*prefix-length*] **as-number** *as-number* command.

6. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

7. Create the BGP IPv6 multicast address family and enter its view.

   address-family ipv6 multicast

8. Enable the router to exchange IPv6 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information used for RPF check with the peers in the group.

Configuring an EBGP peer group by using Method 3 (Exchanging IPv6 multicast routes with IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Create an EBGP peer group.

   group *group-name* **external**

4. Add an IPv6 BGP peer into the EBGP peer group.

   peer *ipv6-address* [*prefix-length*] **group** *group-name* **as-number** *as-number*

5. (Optional.) Configure a description for the peer group.

   peer *group-name* **description** *text*

   By default, no description is configured for the peer group.

6. Create the BGP IPv6 multicast address family and enter its view.

   address-family ipv6 multicast

7. Enable the router to exchange IPv6 unicast routing information used for RPF check with peers in the specified peer group.

   peer *group-name* **enable**

   By default, the router cannot exchange IPv6 unicast routing information used for RPF check with the peers in the group.

Specifying the source address of TCP connections

About this task

BGP uses TCP as the transport layer protocol. Perform this task in the following scenarios to specify the source address or source interface of TCP connections to a peer or peer group:

- The peer's IPv4/IPv6 address does not belong to the interface directly connected to the local router. To ensure successful TCP connection establishment, use one of the following methods:

- Specify the interface to which the IPv4/IPv6 address belongs as the source interface on the peer.
- Specify the IPv4/IPv6 address of the interface directly connected to the local router as the source address on the peer.

- A BGP peer at an IPv6 link-local address must be directly connected to the local router. If you do not specify the interface connected to the peer, you must specify the interface directly connected to the BGP peer as the source interface of TCP connections on the local router.
- On a BGP router that has multiple links to a peer, the source interface for TCP connection changes because the primary source interface fails. To avoid this problem, specify a loopback interface as the source interface or specify the IP address of a loopback interface as the source address.
- You want to establish multiple BGP sessions to a router. In this case, BGP might fail to determine the source address for each TCP connection based on the optimal route to the peer. To prevent this problem, use one of the following methods:
  - If the BGP sessions use IP addresses of different interfaces, specify a source interface or source address for each session.
  - If the BGP sessions use different IP addresses of the same interface, specify a source address for each session.

Restrictions and guidelines

BGP immediately tears down the session to an IBGP peer or peer group when the following conditions exist:

- The source interface of TCP connections to the IBGP peer or peer group is a physical interface.
- The source interface fails and the link to the IBGP peer or peer group goes down.

Procedure (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Specify the source address or source interface of TCP connections to a peer or peer group.
   - Specify the source address of TCP connections to a peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **source-address** *source-ipv4-address*
   - Specify the source interface of TCP connections to a peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **connect-interface** *interface-type interface-number*

   By default, BGP uses the primary IPv4 address of the output interface in the optimal route to a peer or peer group as the source address of TCP connections to the peer or peer group.

Procedure (IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

bgp *as-number* [**instance** *instance-name*]

- Enter BGP-VPN instance view.

  bgp *as-number* [**instance** *instance-name*]

  ip vpn-instance *vpn-instance-name*

3. Specify the source IPv6 address or source interface of TCP connections to a peer or peer group.
   - Specify the source IPv6 address of TCP connections to a peer or peer group.

     peer { *group-name* | *ipv6-address* [*prefix-length*] } **source-address** *source-ipv6-address*
   - Specify the source interface of TCP connections to a peer or peer group.

     peer { *group-name* | *ipv6-address* [*prefix-length*] } **connect-interface** *interface-type interface-number*

   By default, BGP uses the IPv6 address of the output interface in the optimal route to the BGP peer or peer group as the source address of TCP connections to the peer or peer group.

Controlling BGP route generation

Injecting a local network

About this task

Perform this task to inject a network in the local routing table to the BGP routing table, so BGP can advertise the network to BGP peers. The ORIGIN attribute of BGP routes advertised in this way is IGP. You can also use a routing policy to control route advertisement.

The specified network must be available and active in the local IP routing table.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast
3. Configure BGP to advertise a local network.

   network *ipv4-address* [*mask-length* | *mask*] [**route-policy** *route-policy-name*]

   By default, BGP does not advertise local networks.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Configure BGP to advertise a local network.

   network *ipv6-address prefix-length* [**route-policy** *route-policy-name*]

   By default, BGP does not advertise local networks.

Redistributing IGP routes

About this task

Perform this task to configure route redistribution from an IGP to BGP.

By default, BGP does not redistribute default IGP routes. You can use the **default-route imported** command to redistribute default IGP routes into the BGP routing table.

The ORIGIN attribute of BGP routes redistributed from IGPs is INCOMPLETE.

Only active routes can be redistributed. To view route state information, use the **display ip routing-table protocol** or **display ipv6 routing-table protocol** command. For more information about the commands, see *Layer 3—IP Routing Command Reference*.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 multicast

3. Enable route redistribution from the specified IGP into BGP.
   - Redistribute **IS-IS, OSPF, or RIP routes.**

     import-route { **isis** | **ospf** | **rip** } [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]
   - Redistribute direct or static routes.

     import-route { **direct** | **static** } [**med** *med-value* | **route-policy** *route-policy-name*] *

   By default, BGP does not redistribute IGP routes.
4. (Optional.) Redistribute all routes with the same destination address.

   import-route multipath

   This command is not supported in BGP IPv4 multicast address family view.

   By default, BGP redistributes only the optimal routes to its routing table when multiple routes with the same destination address exist.

   This command takes effect only on the routes redistributed by the **import-route** command in the same address family.
5. (Optional.) Enable default route redistribution into BGP.

   default-route imported

   By default, BGP does not redistribute default routes.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Enable route redistribution from the specified IGP into BGP.
   - Redistribute IPv6 **IS-IS, OSPFv3, or RIPng routes.**

     import-route { **isisv6** | **ospfv3** | **ripng** } [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]

- Redistribute direct or static routes.

  import-route { **direct** | **static** } [**med** *med-value* | **route-policy** *route-policy-name*] *

By default, BGP does not redistribute IGP routes.

4. (Optional.) Redistribute all routes with the same destination address.

   import-route multipath

   This command is not supported in BGP IPv6 multicast address family view.

   By default, BGP redistributes only the optimal routes to its routing table when multiple routes with the same destination address exist.

   This command takes effect only on the routes redistributed by the **import-route** command in the same address family.

5. (Optional.) Enable default route redistribution into BGP.

   default-route imported

   By default, BGP does not redistribute default routes.

Redistributing the local network specified in the public instance or a VPN instance

About this task

Perform this task to redistribute the local network specified in the public instance or a VPN instance to the BGP routing table, so BGP can advertise the network. The ORIGIN attribute is IGP for routes advertised in this way.

The local network must exist and be active in the routing table of the public instance or VPN instance.

Procedure

1. Enter system view.

   system-view

2. Enter public instance IPv4 address family view or VPN instance IPv4 address family view.
   - Enter public instance IPv4 address family view.

     ip public-instance

     address-family ipv4
   - Enter VPN instance IPv4 address family view.

     ip vpn-instance *vpn-instance-name*

     address-family ipv4

3. Specify a local network in the public instance or VPN instance.

   network *ipv4-address* [*mask-length* | *mask*]

   By default, no local network is specified in the public instance or VPN instance.

   For more information about this command, see MPLS L3VPN in *MPLS Command Reference*.

4. Execute the **quit** command twice to return to system view.

   quit

5. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  ip vpn-instance *vpn-instance-name*

  address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 multicast

6. Redistribute the local network specified in the public instance or VPN instance.

   import-route local-aggregate [**route-policy** *route-policy-name*]

   By default, BGP does not redistribute the local network specified in the public instance or any VPN instances.

Configuring BGP route summarization

About this task

Route summarization can reduce the number of redistributed routes and the routing table size. IPv4 BGP supports automatic route summarization and manual route summarization. Manual summarization takes precedence over automatic summarization. IPv6 BGP supports only manual route summarization.

Automatic route summarization enables BGP to summarize IGP subnet routes redistributed by the **import-route** command, so BGP advertises only natural network routes.

By configuring manual route summarization, you can do the following:

- Summarize both redistributed routes and routes injected using the **network** command.
- Determine the mask length for a summary route.

Restrictions and guidelines for configuring BGP route summarization

The output interface of a BGP summary route is Null 0 on the originating router. Therefore, a summary route must not be an optimal route on the originating router. Otherwise, BGP will fail to forward packets matching the route. If a summarized specific route has the same mask as the summary route, but has a lower priority, the summary route becomes the optimal route. To ensure correct packet forwarding, change the priority of the summary or specific route to make the specific route the optimal route.

Configuring automatic route summarization (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 multicast

3. Configure automatic route summarization.

   summary automatic

   By default, automatic route summarization is not configured.

Configuring manual route summarization (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Create a summary route in the BGP routing table.

   aggregate *ipv4-address* { *mask-length* | *mask* } [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *

   By default, no summary routes are configured.

Configuring BGP manual route summarization (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

address-family ipv6 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv6 multicast

3. Create a summary route in the IPv6 BGP routing table.

   aggregate *ipv6-address prefix-length* [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *

   By default, no summary routes are configured.

Advertising a default route to a peer or peer group

About this task

Perform this task to advertise a default BGP route with the next hop being the advertising router to a peer or peer group.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Advertise a default route to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **default-route-advertise** [**route-policy** *route-policy-name*]

   interface-peer *interface-type interface-number* **default-route-advertise** [**route-policy** *route-policy-name*]

   By default, no default route is advertised.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Advertise a default route to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **default-route-advertise** [**route-policy** *route-policy-name*]

   interface-peer *interface-type interface-number* **default-route-advertise** [**route-policy** *route-policy-name*]

   By default, no default route is advertised.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Controlling BGP route advertisement

Advertising optimal routes in the IP routing table

About this task

By default, BGP advertises optimal routes in the BGP routing table, which may not be optimal in the IP routing table. This task allows you to advertise BGP routes that are optimal in the IP routing table

Restrictions and guidelines

The **advertise-rib-active** command does not apply to the following routes:

- Routes redistributed by using the **import-route** command.
- Routes advertised by using the **network** command.
- Default routes redistributed by using the **default-route imported** command.
- VPNv4 routes.
- VPNv6 routes.
- IPv4 multicast routes.
- IPv6 multicast routes.

The **advertise-rib-active** command takes effect only on the routes generated after you execute this command. To apply this command to existing routes, use the **reset bgp** command to reset BGP sessions.

The configuration of the **advertise-rib-active** command in BGP unicast address family view applies when it is different from that in BGP instance view.

The **advertise-rib-active** command is mutually exclusive with the **routing-table bgp-rib-only** command.

Procedure (IPv4 unicast)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP to advertise optimal routes in the IP routing table.

   advertise-rib-active

   By default, BGP advertises optimal routes in the BGP routing table.
4. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Enter BGP IPv4 unicast address family view.

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
5. Enable BGP to advertise optimal routes in the IP routing table of the address family in the VPN instance.

   advertise-rib-active

   By default, the setting is the same as that in BGP instance view.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP to advertise optimal routes in the IPv6 routing table.

   advertise-rib-active

   By default, BGP advertises optimal routes in the BGP routing table.
4. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Enter BGP IPv6 unicast address family view.

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
5. Enable BGP to advertise optimal routes in the IPv6 routing table of the address family in the VPN instance.

   advertise-rib-active

By default, the setting is the same as that in BGP instance view.

Enabling prioritized withdrawal of the default route

About this task

Typically a BGP router does not send withdrawal messages of the default route prior to other routes to its peers. If the peer relationship is down, the default route cannot be withdrawn first. Traffic interruption might occur. Perform this task to configure BGP to send the withdrawal messages of the default route prior to other routes. This can reduce the traffic interruption time when the peer relationship is down.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP to send withdrawal messages of the default route prior to other routes.

   default-route update-first

   By default, BGP does not send withdrawal messages of the default route prior to other routes.

Enabling prioritized withdrawal of specific routes

About this task

Perform this task to configure BGP to send the withdrawal messages of specific routes prior to other routes. This can achieve fast route switchover and reduce the traffic interruption time.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast
3. Enable BGP to send withdrawal messages of routes matching the specified routing policy prior to other routes.

   update-first route-policy *route-policy-name*

   By default, BGP does not send withdrawal messages of specific routes prior to other routes.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Enable BGP to send withdrawal messages of routes matching the specified routing policy prior to other routes.

   update-first route-policy *route-policy-name*

   By default, BGP does not send withdrawal messages of specific routes prior to other routes.

Configuring BGP route distribution filtering policies

About this task

To configure BGP route distribution filtering policies, use the following methods:

- Use an ACL or prefix list to filter routing information advertised to all peers.
- Use a routing policy, conditional advertisement policy (existent policy or nonexistent policy), ACL, AS path list, or prefix list to filter routing information advertised to a peer or peer group.

If you configure multiple filtering policies, apply them in the following sequence:

1. **peer prefix-list export**
2. **peer as-path-acl export**
3. **filter-policy export**
4. **peer filter-policy export**
5. **peer advertise-policy exist-policy**
6. **peer advertise-policy non-exist-policy**
7. **peer route-policy export**

Only routes passing all the configured policies can be advertised.

Prerequisites

Before you configure BGP routing filtering policies, configure the following filters used for route filtering as needed:

- ACL (see *ACL and QoS Configuration Guide*).
- Prefix list (see "Configuring routing policies").

- Routing policy (see "Configuring routing policies").
- AS path list (see "Configuring routing policies").

Restrictions and guidelines

When you specify an ACL, prefix list, or routing policy as a filtering policy to filter BGP routing information, follow these guidelines:

- If the specified ACL, prefix list, or routing policy does not exist, all BGP routes can pass the filtering policy.
- When you specify an ACL or prefix list as a BGP route filtering policy (for example, use **filter-policy export** to specify an ACL for route filtering), the following rules apply:
  - BGP routes can pass the filtering policy only when one of the following conditions exists:
    - They match a **permit** rule in the specified ACL.
    - They match a **permit** item in the specified prefix list.

    When the ACL has only **deny** rules or the prefix list has only **deny** items, no BGP routes can pass the filtering policy.
  - When the ACL match order is **config**, BGP routes match against rules in ascending order of rule ID. When BGP routes match against the items of the prefix list, an item with a smaller index number is matched first.
  - In either of the match modes above, a BGP route does not match against the next ACL rule or prefix list item when it already matches the current ACL rule or prefix list item. To avoid wasting filter resources, set a small match range for each ACL rule or prefix list item when you configure multiple ACL rules or prefix list items for an ACL or prefix list.
  - You can use the **rule** [*rule-id*] **permit ip source** *sour-addr sour-wildcard* **destination** *dest-addr dest-wildcard* command to create a rule for an advanced ACL to match specific routes. The **source** *sour-addr* option represents the destination network address and the **destination** *dest-addr* option represents the mask in dotted decimal notation. For example, to match route 3.3.3.0/24, the rule configuration should be **rule** [*rule-id*] **permit ip source** *3.3.3.0 0.0.0.255* **destination** *255.255.255.0 0.0.0.255*.
- When you specify a routing policy as a BGP route filtering policy, the following rules apply:
  - BGP routes can pass the filtering policy only when they match a **permit** node in the routing policy.
  - A BGP route matches against the nodes of the routing policy in ascending order of node number. It does not match against the next node when it already matches the current node (unless a continue clause exists).
  - When a BGP route matches against a node that has an ACL or prefix-list **if-match** clause, it can pass the filtering policy only when the following conditions exist:
    - The route matches a **permit** rule in the ACL or a **permit** item in the prefix list.
    - The route matches a **permit** node in the routing policy.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  ip vpn-instance *vpn-instance-name*

  address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 multicast

3. Configure BGP route distribution filtering policies. Choose the options to configure as needed:

- Reference an ACL or IP prefix list to filter advertised BGP routes.

  filter-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* | **prefix-list** *ipv4-prefix-list-name* } **export** [**direct** | { **isis** | **ospf** | **rip** } *process-id* | **static**]

- Specify a routing policy as the existent policy to control route advertisement. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise-policy** *advertise-policy-name* **exist-policy** *exist-policy-name*

  interface-peer *interface-type interface-number* **advertise-policy** *advertise-policy-name* **exist-policy** *exist-policy-name*

  This command is supported only in BGP IPv4 unicast address family view.

- Specify a routing policy as the nonexistent policy to control route advertisement. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise-policy** *advertise-policy-name* **non-exist-policy** *non-exist-policy-name*

  interface-peer *interface-type interface-number* **advertise-policy** *advertise-policy-name* **non-exist-policy** *non-exist-policy-name*

  This command is supported only in BGP IPv4 unicast address family view.

- Reference a routing policy to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-policy** *route-policy-name* **export**

  interface-peer *interface-type interface-number* **route-policy** *route-policy-name* **export**

- Reference an ACL to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **filter-policy** { *ipv4-acl-number* | **name** *ipv4-acl-name* } **export**

  interface-peer *interface-type interface-number* **filter-policy** { *ipv4-acl-number* | **name** *ipv4-acl-name* } **export**

- Reference an AS path list to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } { **as-path-acl** *as-path-acl-number* | *as-path-acl-name* } **export**

interface-peer *interface-type interface-number* **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **export**

- Reference an IPv4 prefix list to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **prefix-list** *ipv4-prefix-list-name* **export**

  interface-peer *interface-type interface-number* **prefix-list** *ipv4-prefix-list-name* **export**

By default, no BGP distribution filtering policy is configured.

The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Configure BGP route distribution filtering policies. Choose the options to configure as needed:

   - Reference an ACL or IPv6 prefix list to filter advertised BGP routes.

     filter-policy { *ipv6-acl-number* | **name** *ipv6-acl-name* | **prefix-list** *ipv6-prefix-list-name* } **export** [**direct** | { **isisv6** | **ospfv3** | **ripng** } *process-id* | **static**]

   - Specify a routing policy as the existent policy to control route advertisement. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **advertise-policy** *advertise-policy-name* **exist-policy** *exist-policy-name*

     interface-peer *interface-type interface-number* **advertise-policy** *advertise-policy-name* **exist-policy** *exist-policy-name*

     This command is supported only in BGP IPv6 unicast address family view.

   - Specify a routing policy as the nonexistent policy to control route advertisement. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type*

interface-number } **advertise-policy** *advertise-policy-name* **non-exist-policy** *non-exist-policy-name*

interface-peer *interface-type interface-number* **advertise-policy** *advertise-policy-name* **non-exist-policy** *non-exist-policy-name*

This command is supported only in BGP IPv6 unicast address family view.

- Reference a routing policy to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **route-policy** *route-policy-name* **export**

  interface-peer *interface-type interface-number* **route-policy** *route-policy-name* **export**

- Reference an ACL to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **filter-policy** { *ipv6-acl-number* | **name** *ipv6-acl-name* } **export**

  interface-peer *interface-type interface-number* **filter-policy** { *ipv6-acl-number* | **name** *ipv6-acl-name* } **export**

- Reference an AS path list to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **export**

  interface-peer *interface-type interface-number* **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **export**

- Reference an IPv6 prefix list to filter BGP routes advertised to a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **prefix-list** *ipv6-prefix-list-name* **export**

  interface-peer *interface-type interface-number* **prefix-list** *ipv6-prefix-list-name* **export**

By default, no BGP distribution filtering policy is configured.

The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Enabling first-AS-number check for EBGP route advertisement

About this task

By default, BGP does not check the first AS number of a received EBGP route and directly advertises it to all peers except the source peer. The default setting for EBGP route advertisement brings about a large number of BGP routes and is not friendly to network maintenance.

To resolve this issue, perform this task. When the first AS number of an EBGP route to be advertised is the same as that of an EBGP peer, BGP does not advertise the route to that peer. This can reduce the number of routes received by EBGP peers.

Procedure (IPv4 unicast)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
3. Enable BGP to perform first-AS-number check for EBGP routes before route advertisement.

   peer-as-check enable

   By default, BGP does not check the first AS number of a received EBGP route and directly advertises it to all peers except the source peer.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view
2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
3. Enable BGP to perform first-AS-number check for EBGP routes before route advertisement.

   peer-as-check enable

   By default, BGP does not check the first AS number of a received EBGP route and directly advertises it to all peers except the source peer.

Setting the BGP route sending rate

About this task

If a device sends many new routes within a short time period, it might be unable to add the routes to the FIB before the peer device adds them. This might result in traffic forwarding failure. To avoid this problem, you can perform this task to set an appropriate route sending rate for the device.

Restrictions and guidelines

For a device with high performance, you can set a high BGP route sending rate as needed. For a device without high performance, set a relatively low BGP route sending rate as a best practice.

As a best practice to avoid route withdrawal failures, do not set the BGP route sending rate to 0 or a small value when network flapping occurs.

This task applies only to IPv4 unicast routes and IPv6 unicast routes.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Set the BGP route sending rate.
   route-rate-limit *rate*
   By default, the BGP route sending rate is not limited.

Configuring BGP to delay sending route updates on reboot

About this task

This task reduces traffic loss. With this task performed, BGP delays sending route updates when it restores after a device reboot. During the delay time, BGP learns all routes from other neighbors, and then selects the optimal route. After the delay time elapses, BGP will advertise the optimal route.

You can specify a prefix list and enable BGP to immediately send route updates for routes that match the prefix list.

Restrictions and guidelines

For BGP EVPN routes, this feature takes effect only on IP prefix routes and MAC/IP advertisement routes that carry IP routing information.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Configure BGP to delay sending route updates when it restores after a device reboot.
   bgp update-delay on-startup *seconds*
   By default, BGP immediately sends route updates to BGP peers in established state when it restores after a device reboot.
4. (Optional.) Configure BGP to immediately send route updates for routes that match an IPv4 prefix list.
   bgp update-delay on-startup prefix-list *ipv4-prefix-list-name*
   By default, no IPv4 prefix list is specified to filter routes.
5. (Optional.) Configure BGP to immediately send route updates for routes that match an IPv6 prefix list.
   bgp update-delay on-startup ipv6-prefix-list *ipv6-prefix-list-name*
   By default, no IPv6 prefix list is specified to filter routes.

Configuring the interval for sending updates for the same route

About this task

A BGP router sends an UPDATE message to its peers when a route is changed. If the route changes frequently, the BGP router keeps sending update or withdrawal messages for the same route, resulting route flapping. To prevent this situation, perform this task to configure the interval for sending updates for the same route to a peer or peer group.

This feature takes effect on withdrawn routes only after you apply the route update interval setting to withdrawn routes.

Restrictions and guidelines

BGP compares the intervals specified for a BGP peer by using the **peer route-update-interval** and **route-update-delay** commands. Only the larger interval takes effect. If you did not execute the **peer route-update-interval** command, BGP uses the default value of this command to compare with the interval configured by the **route-update-delay** command.

Procedure (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. (Optional.) Apply route update interval setting to withdrawn routes.

   route-update-interval withdrawn enable

   By default, route update interval setting does not apply to withdrawn routes. BGP sends withdrawal messages for withdrawn routes immediately.
4. (Optional.) Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*
5. Configure the interval for sending updates for the same route to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] }
   route-update-interval *interval*

   By default, the interval is 15 seconds for an IBGP peer and 30 seconds for an EBGP peer.

Procedure (IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. (Optional.) Apply route update interval setting to withdrawn routes.

   route-update-interval withdrawn enable

   By default, route update interval setting does not apply to withdrawn routes. BGP sends withdrawal messages for withdrawn routes immediately.
4. (Optional.) Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*
5. Configure the interval for sending updates for the same route to a peer or peer group. Choose one command as needed:

peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **route-update-interval** *interval*

interface-peer *interface-type interface-number* **route-update-interval** *interval*

By default, the interval is 15 seconds for an IBGP peer and 30 seconds for an EBGP peer.

Configuring BGP route advertisement delay

About this task

BGP sends update messages to peers when route changes occur or peer relationship establishment is complete. If a large number of routes need to be updated within a short time period, BGP might send update messages to peers before issuing the updated routes to the FIB. As a result, some packets might be discarded because BGP cannot find the forwarding path for the packets.

To resolve this issue, perform this task to enable BGP route advertisement delay and set a delay timer. When route changes occur or peer relationship establishment is complete, BGP cannot send any update messages unless the delay timer expires.

Restrictions and guidelines

BGP compares the timer values specified for a BGP peer by using the **route-update-delay** command and the **peer route-update-interval** command. Only the larger timer value takes effect. If you did not execute the **peer route-update-interval** command, BGP compares the BGP route advertisement delay timer with the default value of the **peer route-update-interval** command.

This feature applies to the following routes:

- BGP IPv4 unicast routes.
- BGP IPv6 unicast routes.
- BGP VPNv4 routes.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP route advertisement delay and set a delay timer.

   route-update-delay *delay-value*

   By default, BGP route advertisement delay is disabled.

Setting the refresh route advertisement delay timer

About this task

BGP advertises all routes to a peer in the following scenarios:

- BGP establishes a session to the peer for the first time.
- The update group of the peer changes.
- BGP receives route-refresh messages from the peer.

Routes advertised in the above scenarios are called refresh routes.

Perform this task to set the refresh route advertisement delay timer that starts when BGP advertises refresh routes for the first time. BGP delays refresh route advertisement to all peers until the timer

expires. After the timer expires, BGP advertises refresh routes to the corresponding peers by update group, which improves route advertisement efficiency and saves system resources.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Set the refresh route advertisement delay timer.
   refresh delay *delay-value*
   By default, the refresh route advertisement delay timer is five seconds.

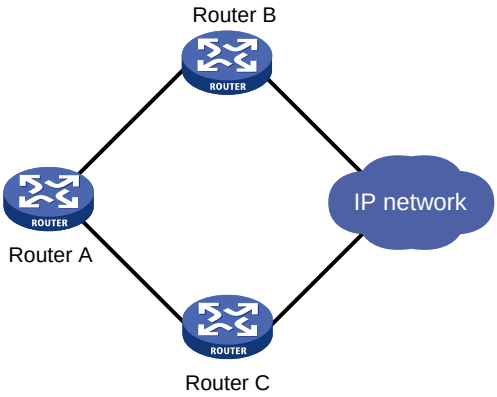
Configuring a startup policy for BGP route updates

About this task

Perform this task to configure BGP to send route updates with the specified attributes within the specified period after reboot.

As shown in Figure 23, if Router B restarts and sends route updates before route convergence completes, traffic sent from Router A through Router B might be lost. This feature enables Router B to send route updates with the specified attribute values within the specified period after reboot, so that Router A can forward traffic through Router C.

Figure 23 Network diagram



Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Specify the period after reboot within which the startup policy is effective.
   bgp apply-policy on-startup duration *seconds*
   By default, the startup policy does not take effect.
4. Specify a MED attribute value in the startup policy.
   bgp policy on-startup med *med-value*
   By default, the MED attribute value in the startup policy is 4294967295.

Controlling BGP route reception

Limiting routes received from a peer or peer group

About this task

This feature can prevent attacks that send a large number of BGP routes to the router.

If the number of routes received from a peer or peer group exceeds the upper limit, the router takes one of the following actions based on your configuration:

- Tears down the BGP session to the peer or peer group and does not attempt to re-establish the session.
- Continues to receive routes from the peer or peer group and generates a log message.
- Retains the session to the peer or peer group, but it discards excess routes and generates a log message.
- Tears down the BGP session to the peer or peer group and, after a specific period of time, re-establishes a BGP session to the peer or peer group.

You can specify a percentage threshold for the router to generate a log message. When the ratio of the number of received routes to the maximum number reaches the percentage value, the router generates a log message.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Specify the maximum number of routes that a router can receive from a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-limit** *prefix-number* [{ **alert-only** | **discard** | **reconnect** *reconnect-time* } | *percentage-value*] *

   interface-peer *interface-type interface-number* **route-limit** *prefix-number* [{ **alert-only** | **discard** | **reconnect** *reconnect-time* } | *percentage-value*] *

   By default, the number of routes that a router can receive from a peer or peer group is not limited.

The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Specify the maximum number of routes that a router can receive from a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **route-limit** *prefix-number* [{ **alert-only** | **discard** | **reconnect** *reconnect-time* } | *percentage-value*] *

   interface-peer *interface-type interface-number* **route-limit** *prefix-number* [{ **alert-only** | **discard** | **reconnect** *reconnect-time* } | *percentage-value*] *

   By default, the number of routes that a router can receive from a peer or peer group is not limited.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Configuring BGP route reception filtering policies

About this task

You can use the following methods to configure BGP route reception filtering policies:

- Use an ACL or prefix list to filter routing information received from all peers.
- Use a routing policy, ACL, AS path list, or prefix list to filter routing information received from a peer or peer group.

If you configure multiple filtering policies, apply them in the following sequence:

1. **filter-policy import**
2. **peer filter-policy import**
3. **peer as-path-acl import**

4. **peer prefix-list import**
5. **peer route-policy import**

Only routes passing all the configured policies can be received.

Prerequisites

Before you configure BGP route reception filtering policies, configure the following filters used for route filtering as needed:

- ACL (see *ACL and QoS Configuration Guide*).
- Prefix list (see "Configuring routing policies").
- Routing policy (see "Configuring routing policies").
- AS path list (see "Configuring routing policies").

Restrictions and guidelines

When you specify an ACL, prefix list, or routing policy as a filtering policy to filter BGP routing information, follow these guidelines:

- If the specified ACL, prefix list, or routing policy does not exist, all BGP routes can pass the filtering policy.
- When you specify an ACL or prefix list as a BGP route filtering policy (for example, use **filter-policy import** to specify an ACL for routing filtering), the following rules apply:
  - BGP routes can pass the filtering policy only when one of the following conditions exists:
    - They match a **permit** rule in the specified ACL.
    - They match a **permit** item in the specified prefix list.

    When the ACL has only **deny** rules or the prefix list has only **deny** items, no BGP routes can pass the filtering policy.
  - When the ACL match order is **config**, BGP routes match against rules in ascending order of rule ID. When BGP routes match against the items of the prefix list, an item with a smaller index number is matched first.

    In either of the match modes above, a BGP route does not match against the next ACL rule or prefix list item when it already matches the current ACL rule or prefix list item. To avoid wasting filter resources, set a small match range for each ACL rule or prefix list item when you configure multiple ACL rules or prefix list items for an ACL or prefix list.
  - You can use the **rule** [*rule-id*] **permit ip source** *sour-addr sour-wildcard* **destination** *dest-addr dest-wildcard* command to create a rule for an advanced ACL to match specific routes. The **source** *sour-addr* option represents the destination network address and the **destination** *dest-addr* option represents the mask in dotted decimal notation. For example, to match route 3.3.3.0/24, the rule configuration should be **rule** [*rule-id*] **permit ip source** *3.3.3.0 0.0.0.255* **destination** *255.255.255.0 0.0.0.255*.
- When you specify a routing policy as a BGP route filtering policy, the following rules apply:
  - BGP routes can pass the filtering policy only when they match a **permit** node in the routing policy.
  - A BGP route matches against the nodes of the routing policy in ascending order of node number. It does not match against the next node when it already matches the current node (unless a continue clause exists).
  - When a BGP route matches against a node that has an ACL or prefix-list **if-match** clause, it can pass the filtering policy only when the following conditions exist:
    - The route matches a **permit** rule in the ACL or a **permit** item in the prefix list.
    - The route matches a **permit** node in the routing policy.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Configure BGP route reception filtering policies. Choose the options to configure as needed:

   - Reference an ACL or IP prefix list to filter BGP routes received from all peers.

     filter-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* | **prefix-list** *ipv4-prefix-list-name* } **import**

   - Reference a routing policy to filter BGP routes received from a peer or peer group. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-policy** *route-policy-name* **import**

     interface-peer *interface-type interface-number* **route-policy** *route-policy-name* **import**

   - Reference an ACL to filter BGP routes received from a peer or peer group. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **filter-policy** { *ipv4-acl-number* | **name** *ipv4-acl-name* } **import**

     interface-peer *interface-type interface-number* **filter-policy** { *ipv4-acl-number* | **name** *ipv4-acl-name* } **import**

   - Reference an AS path list to filter BGP routes received from a peer or peer group. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **import**

     interface-peer *interface-type interface-number* **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **import**

   - Reference an IPv4 prefix list to filter BGP routes received from a peer or peer group. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **prefix-list** *ipv4-prefix-list-name* **import**

     interface-peer *interface-type interface-number* **prefix-list** *ipv4-prefix-list-name* **import**

By default, no route reception filtering is configured.

The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Configure BGP route reception filtering policies. Choose the options to configure as needed:
   - Reference ACL or IPv6 prefix list to filter BGP routes received from all peers.

     filter-policy { *ipv6-acl-number* | **name** *ipv6-acl-name* | **prefix-list** *ipv6-prefix-list-name* } **import**
   - Reference a routing policy to filter BGP routes received from a peer or peer group. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **route-policy** *route-policy-name* **import**

     interface-peer *interface-type interface-number* **route-policy** *route-policy-name* **import**
   - Reference an ACL to filter BGP routes received from a peer or peer group. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **filter-policy** { *ipv6-acl-number* | **name** *ipv6-acl-name* } **import**

     interface-peer *interface-type interface-number* **filter-policy** { *ipv6-acl-number* | **name** *ipv6-acl-name* } **import**
   - Reference an AS path list to filter BGP routes received from a peer or peer group. Choose one command as needed:

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **import**

     interface-peer *interface-type interface-number* **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **import**

- Reference an IPv6 prefix list to filter BGP routes received from a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **prefix-list** *ipv6-prefix-list-name* **import**

  interface-peer *interface-type interface-number* **prefix-list** *ipv6-prefix-list-name* **import**

  By default, no route reception filtering is configured.

  The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Configuring the SoO attribute

About this task

After you configure the SoO attribute for a BGP peer or peer group, BGP adds the SoO attribute into the route updates received from the BGP peer or peer group. In addition, before advertising route updates to the peer or peer group, BGP checks the SoO attribute of the route update against the configured SoO attribute. If they are the same, BGP does not advertise the route updates to the BGP peer or peer group.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast
3. Configure the SoO attribute for a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **soo** *site-of-origin*

   interface-peer *interface-type interface-number* **soo** *site-of-origin*

   By default, no SoO attribute is configured for a peer or peer group.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Configure the SoO attribute for a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **soo** *site-of-origin*

   interface-peer *interface-type interface-number* **soo** *site-of-origin*

   By default, no SoO attribute is configured for a peer or peer group.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Configuring BGP timers

Configuring the keepalive interval and hold time

About this task

BGP sends KEEPALIVE messages regularly to keep the BGP session between two routers.

If a router receives no KEEPALIVE or UPDATE message from a peer within the hold time, it tears down the session.

You can configure the keepalive interval and hold time globally or for a peer or peer group. The individual settings take precedence over the global settings.

The actual keepalive interval and hold time are determined as follows:

- If the local device is configured with a different hold time from a peer, the smaller hold time takes effect after peer negotiation. To ensure an appropriate negotiated hold time, use the **min-holdtime** *min-holdtime* option to specify the minimum hold time. If the negotiated hold time is smaller than the minimum hold time configured on the local device, it terminates the related BGP session. If either the local device or the peer device has a hold time of 0, the local device will not be affected by the configured minimum hold time. They can establish a BGP session regardless of the minimum hold time.

  If the hold time is set to 0, the local device does not send keepalive messages to the peer after negotiation. In this case, the peer session will never time out, because the keepalive mechanism will not be triggered. As a best practice, set the hold time to 0 only in certain scenarios, for example, debugging or testing.

- If the keepalive interval is not 0, the actual keepalive interval is the smaller one between 1/3 of the hold time and the keepalive interval.

Restrictions and guidelines

The hold time must be a minimum of three times the keepalive interval. The minimum hold time must be smaller than or equal to the hold time.

You can perform this task for all BGP peers or for a specific BGP peer or peer group. For a BGP peer or peer group, the configuration of the **peer timer** command takes precedence over that of the **timer** command. If you do not configure the **peer timer** command for the peer or peer group, the configuration of the **timer** command takes effect on the peer or peer group.

Procedure (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Configure the keepalive interval and hold time.
   - Configure the global keepalive interval and hold time.

     timer keepalive *keepalive* **hold** *holdtime* [**min-holdtime** *min-holdtime*]

     This command takes effect for new BGP sessions and does not affect existing sessions.
   - Configure the keepalive interval and hold time for a peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **timer keepalive** *keepalive* **hold** *holdtime* [**min-holdtime** *min-holdtime*]

   By default, the keepalive interval is 60 seconds, and hold time is 180 seconds.

   The timers configured with the **timer** and **peer timer** commands do not take effect until a session is re-established (for example, a session is reset).

Procedure (IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Configure the keepalive interval and hold time.
   - Configure the global keepalive interval and hold time.

     timer keepalive *keepalive* **hold** *holdtime*

     This command takes effect for new BGP sessions and does not affect existing sessions.
   - Configure the keepalive interval and hold time for a peer or peer group. Choose one command as needed:

peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **timer keepalive** *keepalive* **hold** *holdtime* [**min-holdtime** *min-holdtime*]

interface-peer *interface-type interface-number* **timer keepalive** *keepalive* **hold** *holdtime* [**min-holdtime** *min-holdtime*]

By default, the keepalive interval is 60 seconds, and hold time is 180 seconds.

The timers configured with the **timer** and **peer timer** commands do not take effect until a session is re-established (for example, a session is reset).

Setting the session retry timer

About this task

To speed up session establishment to a peer or peer group and route convergence, set a small session retry timer. If the BGP session flaps, you can set a large session retry timer to reduce the impact.

Restrictions and guidelines

The timer set by the **peer timer connect-retry** command takes precedence over the timer set by the **timer connect-retry** command.

Procedure (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Set the session retry timer.
   - Set the session retry timer for all peers or peer groups.

     timer connect-retry *retry-time*
   - Set the session retry timer for a peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **timer connect-retry** *retry-time*

   By default, the session retry timer is 32 seconds for a peer or peer group.

Procedure (IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Set the session retry timer.

- Set the session retry timer for all peers or peer groups.

  timer connect-retry *retry-time*

- Set the session retry timer for a peer or peer group. Choose one command as needed:

  peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **timer connect-retry** *retry-time*

  interface-peer *interface-type interface-number* **timer connect-retry** *retry-time*

By default, the session retry timer is 32 seconds for a peer or peer group.

Setting an update delay for local MPLS labels

About this task

BGP includes local MPLS labels in advertised VPNv4 routes, labeled IPv6 unicast routes, and labeled IPv4 unicast routes.

When a local label is changed, BGP removes the old label and advertises the new label. Traffic interruption occurs if BGP peers use the old label to forward packets before they learn the new label. To resolve this issue, set an update delay for local MPLS labels. BGP does not remove the old label before the update delay timer expires.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Set an update delay for local MPLS labels.

   retain local-label *retain-time*

   By default, the update delay is 60 seconds.

Configuring BGP logging and notifications

Enabling logging for session state changes

About this task

Perform this task to enable BGP to log BGP session establishment and disconnection events. To display the log information, use the **display bgp peer ipv4 unicast log-info** command or the **display bgp peer ipv6 unicast log-info** command. The logs are sent to the information center. The output rules of the logs (whether to output the logs and where to output) are determined by the information center configuration.

For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure (IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enable logging for session state changes globally.

log-peer-change

By default, logging for session state changes is enabled globally.

4. (Optional.) Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*

5. Enable logging for session state changes for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **log-change**

   By default, logging for session state changes is enabled for all peers and peer groups.

Procedure (IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enable logging for session state changes globally.

   log-peer-change

   By default, logging for session state changes is enabled globally.

4. (Optional.) Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*

5. Enable logging for session state changes for a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **log-change**

   interface-peer *interface-type interface-number* **log-change**

   By default, logging for session state changes is enabled for all peers and peer groups.

Enabling logging for BGP route flapping

About this task

This feature enables BGP to generate logs for BGP route flapping events that trigger log generation. The generated logs are sent to the information center. For the logs to be output correctly, you must also configure information center on the device. For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure (IPv4 unicast/IPv4 multicast)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 multicast

3. Enable logging for BGP route flapping.

   log-route-flap *monitor-time monitor-count* [*log-count-limit* | **route-policy** *route-policy-name*] *

   By default, logging for BGP route flapping is disabled.

Procedure (IPv6 unicast/IPv6 multicast)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Enable logging for BGP route flapping.

   log-route-flap *monitor-time monitor-count* [*log-count-limit* | **route-policy** *route-policy-name*] *

   By default, logging for BGP route flapping is disabled.

Configuring BGP network management

About this task

After you enable SNMP notifications for BGP, the device generates a notification when a BGP session state change occurs. The notification includes the neighbor address, the error code and subcode of the most recent error, and the current neighbor state. For BGP notifications to be sent correctly, you must also configure SNMP on the device.

BGP does not know the BGP instance to which a managed MIB node belongs. To resolve this issue, configure different SNMP contexts for different BGP instances.

The device selects a MIB for an SNMP packet according to the context (for SNMPv3) or community name (for SNMPv1/v2c) in the following ways:

- For an SNMPv3 packet:
  - The device selects the MIB of the default BGP instance if the packet does not carry a context and no SNMP context is configured for the default BGP instance.
  - The device selects the MIB of a BGP instance if the packet meets the following conditions:

- Carries a context that is configured with the **snmp-agent context** command in system view.
- Matches the context of the BGP instance.

- The device does not process any MIBs in other situations.

- For an SNMPv1/v2c packet:
  - The device selects the MIB of the default BGP instance if the following conditions are met:
    - No community name-to-SNMP context mapping is configured with the **snmp-agent community-map** command in system view.
    - No SNMP context is configured for the default BGP instance.
  - The device selects the MIB of a BGP instance if the community name is mapped to an SNMP context and the context matches the context of the BGP instance.
  - The device does not process any MIBs in other situations.

For more information about SNMP contexts and community names, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

After you configure the **snmp-agent trap enable bgp** command, the following rules apply:

- Whether notifications specified by the **peer-addrfamily-routeexceed**, **peer-addrfamily-routeexceed-clear**, **peer-addrfamily-routethreshold-exceed**, **peer-addrfamily-routethreshold-clear**, **peer-routeexceed**, **peer-routeexceed-clear**, **peer-routethreshold-clear**, or **peer-routethreshold-exceed** keyword can be generated depends on the configuration of the **peer route-limit** command. The **peer route-limit** command determines the number of routes that the router can receive and the log generation threshold. If the **peer route-limit** command is not executed, the maximum number of routes that triggers notifications is 4294967295 and the route quantity alarm threshold is 70%.
- Whether notifications specified by the **peer-rpkisession-roaexceed** or **peer-rpkisession-roaexceed-clear** keyword can be generated depends on the configuration of the **rpki-limit** command. If the **rpki-limit** command is not configured in RPKI server view, the device generates notifications when the number of ROAs received from the RPKI server reaches or exceeds 4294967295.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for BGP.

   snmp-agent trap enable bgp [**instance** *instance-name*] [**dynamic-peer-limit-exceed** | **dynamic-peer-limit-exceed-clear** | **originator-id-conflict** | **peer-addrfamily-routeexceed** | **peer-addrfamily-routeexceed-clear** | **peer-addrfamily-routethreshold-clear** | **peer-addrfamily-routethreshold-exceed** | **peer-backward-transition** | **peer-established** | **peer-gr-status-change** | **peer-routeexceed** | **peer-routeexceed-clear** | **peer-route-flap** | **peer-routethreshold-clear** | **peer-routethreshold-exceed** | **peer-rpkisession-roaexceed** | **peer-rpkisession-roaexceed-clear** | **remote-routerid-conflict**] *

   By default, SNMP notifications for BGP are enabled.

If you do not specify any notification types, this command enables all types of notifications for BGP. If you specify notification types, this command enables only the specified types of notifications for BGP.

3. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

4. Configure an SNMP context for the BGP instance.

   snmp context-name *context-name*

   By default, no SNMP context is configured for the BGP instance.

Display and maintenance commands for basic BGP network building

Displaying BGP

Execute **display** commands in any view.

Displaying BGP (IPv4 unicast address family)

Task	Command
Display BGP IPv4 unicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**group-name** *group-name*]
Display information about routes advertised by the **network** command and shortcut routes configured by the **network short-cut** command.	**display bgp** [**instance** *instance-name*] **network ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*]
Display information about a peer or peer group in BGP IPv4 unicast address family.	**display bgp** [**instance** *instance-name*] **peer ipv4** [**unicast**] [*ipv4-address mask-length* | { *ipv4-address* | **group-name** *group-name* } **log-info** | [*ipv4-address*] **verbose**] **display bgp** [**instance** *instance-name*] **peer ipv4** [**unicast**] [*ipv6-address prefix-length* | **interface** *interface-type interface-number* | *ipv6-address* [**interface** *interface-type interface-number*] { **log-info** | **verbose** }] **display bgp** [**instance** *instance-name*] **peer ipv4** [**unicast**] **vpn-instance-all** [**verbose**]
Display BGP IPv4 unicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv4-address* [{ *mask-length* | *mask* } [**longest-match**]] | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]

Task	Command
Display BGP IPv4 unicast routing information.	**display bgp** [**instance** *instance-name*] **routing-table ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv4-address* [{ *mask-length* | *mask* } [**longest-match**]] | *ipv4-address* [*mask-length* | *mask*] **advertise-info** | *ipv4-address* [*mask-length* | *mask*] { **as-path** | **cluster-list** | **community** | **ext-community** } | **peer** { *ipv4-address* | *ipv6-address* [**interface** *interface-type interface-number*] } { **advertised-routes** | **received-routes** } [*ipv4-address* [*mask-length* | *mask*] | **statistics**] | **statistics**] **display bgp** [**instance** *instance-name*] **routing-table ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **display bgp** [**instance** *instance-name*] **routing-table ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**statistics**] **community** [*community-number*&<1-32> | *aa:nn*&<1-32>] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**statistics**] **community-list** { *basic-community-list-number* | *adv-community-list-number* | *comm-list-name* } [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**statistics**] **ext-community** [**bandwidth** *link-bandwidth-value* | **rt** *route-target* | **soo** *site-of-origin*]&<1-32> [**whole-match**]
Display information about BGP peer relationship down events.	**display bgp** [**instance** *instance-name*] **troubleshooting** [*event-count*] [**reverse**]
Display BGP IPv4 unicast address family update group information.	**display bgp** [**instance** *instance-name*] **update-group ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv4-address*] **display bgp** [**instance** *instance-name*] **update-group ipv4** [**unicast**] [*ipv6-address*]

Task	Command
Display information about all BGP instances.	**display bgp instance-info**

Displaying BGP (IPv6 unicast address family)

Task	Command
Display BGP IPv6 unicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**group-name** *group-name*]
Display information about routes advertised by the **network** command and shortcut routes configured by the **network short-cut** command.	**display bgp** [**instance** *instance-name*] **network ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*]
Display information about a peer or peer group in BGP IPv6 unicast address family.	**display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv6-address prefix-length* | { *ipv6-address* | **group-name** *group-name* } **log-info** | [*ipv6-address*] **verbose**] **display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [[*link-local-address*] **interface** *interface-type interface-number* | *link-local-address* **interface** *interface-type interface-number* { **log-info** | **verbose** }] **display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [*ipv4-address mask-length* | *ipv4-address* **log-info** | [*ipv4-address*] **verbose**] **display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] **vpn-instance-all** [**verbose**]
Display BGP IPv6 unicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv6-address prefix-length* | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]
Display BGP IPv6 unicast routing information.	**display bgp** [**instance** *instance-name*] **routing-table ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv6-address prefix-length* [**advertise-info**] | *ipv6-address prefix-length* { **as-path** | **cluster-list** | **community** | **ext-community** } | **peer** *ipv6-address* { **advertised-routes** | **received-routes** } [*ipv6-address prefix-length* | **statistics**] |

Task	Command
	statistics] **display bgp** [**instance** *instance-name*] **routing-table ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] **peer** { *ipv4-address* | [*link-local-address*] **interface** *interface-type interface-number* } { **advertised-routes** | **received-routes** } [*ipv6-address prefix-length* | **statistics**] **display bgp** [**instance** *instance-name*] **routing-table ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **display bgp** [**instance** *instance-name*] **routing-table ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**statistics**] **community** [*community-number*&<1-32> | *aa:nn*&<1-32>] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**statistics**] **community-list** { *basic-community-list-number* | *adv-community-list-number* | *comm-list-name* } [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**statistics**] **ext-community** [**bandwidth** *link-bandwidth-value* | **rt** *route-target* | **soo** *site-of-origin*]&<1-32> [**whole-match**]
Display information about BGP peer relationship down events.	**display bgp** [**instance** *instance-name*] **troubleshooting** [*event-count*] [**reverse**]
Display BGP IPv6 unicast address family update group information.	**display bgp** [**instance** *instance-name*] **update-group ipv6** [**unicast**] [*ipv4-address* | *ipv6-address*] **display bgp** [**instance** *instance-name*] **update-group ipv6** [**unicast**] **vpn-instance** *vpn-instance-name* [*ipv6-address*]
Display information about all BGP instances.	**display bgp instance-info**

Displaying BGP (IPv4 multicast address family)

Task	Command
Display BGP IPv4 multicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv4 multicast** [**group-name** *group-name*]
Display information about routes advertised by the **network** command and shortcut routes configured by the **network short-cut** command.	**display bgp** [**instance** *instance-name*] **network ipv4 multicast**
Display BGP path attribute information.	**display bgp** [**instance** *instance-name*] **paths** [*as-regular-expression*]
Display BGP IPv4 multicast peer or peer group information.	**display bgp** [**instance** *instance-name*] **peer ipv4 multicast** [*ipv4-address mask-length* | { *ipv4-address* | **group-name** *group-name* } **log-info** | [*ipv4-address*] **verbose**]
Display BGP IPv4 multicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv4 multicast** [*ipv4-address* [{ *mask-length* | *mask* } [**longest-match**]] | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]
Display BGP IPv4 multicast routing information.	**display bgp** [**instance** *instance-name*] **routing-table ipv4 multicast** [*ipv4-address* [{ *mask-length* | *mask* } [**longest-match**]] | *ipv4-address* [*mask-length* | *mask*] **advertise-info** | *ipv4-address* [*mask-length* | *mask*] { **as-path** | **cluster-list** | **community** | **ext-community** } | **peer** *ipv4-address* { **advertised-routes** | **received-routes** } [*ipv4-address* [*mask-length* | *mask*] | **statistics**] | **statistics**] **display bgp** [**instance** *instance-name*] **routing-table ipv4 multicast as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **display bgp** [**instance** *instance-name*] **routing-table ipv4 multicast** [**statistics**] **community** [*community-number*&<1-32> | *aa:nn*&<1-32>] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv4 multicast** [**statistics**] **community-list** { *basic-community-list-number* | *adv-community-list-number* | *comm-list-name* } [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv4 multicast**

Task	Command
	[**statistics**] **ext-community** [**bandwidth** *link-bandwidth-value* | **rt** *route-target* | **soo** *site-of-origin*]&<1-32> [**whole-match**]
Display information about BGP peer relationship down events.	**display bgp** [**instance** *instance-name*] **troubleshooting** [*event-count*] [**reverse**]
Display BGP IPv4 multicast address family update group information.	**display bgp** [**instance** *instance-name*] **update-group ipv4 multicast** [*ipv4-address*]
Display information about all BGP instances.	**display bgp instance-info**

Displaying BGP (IPv6 multicast address family)

Task	Command
Display BGP IPv6 multicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv6 multicast** [**group-name** *group-name*]
Display information about routes advertised by the **network** command and shortcut routes configured by the **network short-cut** command.	**display bgp** [**instance** *instance-name*] **network ipv6 multicast**
Display BGP path attribute information.	**display bgp** [**instance** *instance-name*] **paths** [*as-regular-expression*]
Display BGP IPv6 multicast peer or peer group information.	**display bgp** [**instance** *instance-name*] **peer ipv6 multicast** [*ipv6-address prefix-length* | { *ipv6-address* | **group-name** *group-name* } **log-info** | [*ipv6-address*] **verbose**]
Display BGP IPv6 multicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv6 multicast** [*ipv6-address prefix-length* | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]
Display BGP IPv6 multicast routing information.	**display bgp** [**instance** *instance-name*] **routing-table ipv6 multicast** [*ipv6-address prefix-length* [**advertise-info**] | *ipv6-address prefix-length* { **as-path** | **cluster-list** | **community** | **ext-community** } | **peer** *ipv6-address* { **advertised-routes** | **received-routes** } [*ipv6-address prefix-length* | **statistics**] | **statistics**] **display bgp** [**instance** *instance-name*] **routing-table ipv6 multicast as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **display bgp** [**instance** *instance-name*]

Task	Command
	routing-table ipv6 multicast [**statistics**] **community** [*community-number*&<1-32> | *aa:nn*&<1-32>] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv6 multicast** [**statistics**] **community-list** { *basic-community-list-number* | *adv-community-list-number* | *comm-list-name* } [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table ipv6 multicast** [**statistics**] **ext-community** [**bandwidth** *link-bandwidth-value* | **rt** *route-target* | **soo** *site-of-origin*]&<1-32> [**whole-match**]
Display information about BGP peer relationship down events.	**display bgp** [**instance** *instance-name*] **troubleshooting** [*event-count*] [**reverse**]
Display BGP IPv6 multicast address family update group information.	**display bgp** [**instance** *instance-name*] **update-group ipv6 multicast** [*ipv6-address*]
Display information about all BGP instances.	**display bgp instance-info**

Resetting BGP sessions

CAUTION:

A reset operation tears down BGP sessions for a short period of time.

Execute **reset** commands in user view.

Task	Command
Reset BGP sessions for IPv4 unicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] **reset bgp** { *ipv6-address* [*prefix-length*] | **interface** *interface-type interface-number* } **ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*]
Reset BGP sessions for IPv4 multicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **ipv4 multicast**

Task	Command
Reset BGP sessions for IPv6 unicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv6-address* [*prefix-length*] | [*link-local-address*] **interface** *interface-type interface-number* | **all** | **external** | **group** *group-name* | **internal** } **ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] **reset bgp** *ipv4-address* [*mask-length*] **ipv6** [**unicast**]
Reset BGP sessions for IPv6 multicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } **ipv6 multicast**
Reset all BGP sessions.	**reset bgp** [**instance** *instance-name*] **all**

Clearing BGP information

Execute **reset** commands in user view.

Task	Command
Clear flap information for BGP IPv4 unicast routes.	**reset bgp** [**instance** *instance-name*] **flap-info ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv4-address* [*mask-length* | *mask*] | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } | **peer** *ipv4-address* [*mask-length*]]
Clear flap information for BGP IPv4 multicast routes.	**reset bgp** [**instance** *instance-name*] **flap-info ipv4 multicast** [*ipv4-address* [*mask-length* | *mask*] | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } | **peer** *ipv4-address* [*mask-length*]]
Clear flap information for BGP IPv6 unicast routes.	**reset bgp** [**instance** *instance-name*] **flap-info ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv6-address prefix-length* | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } | **peer** *ipv6-address* [*prefix-length*]]
Clear flap information for BGP IPv6 multicast routes.	**reset bgp** [**instance** *instance-name*] **flap-info ipv6 multicast** [*ipv6-address prefix-length* | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } | **peer** *ipv6-address* [*prefix-length*]]

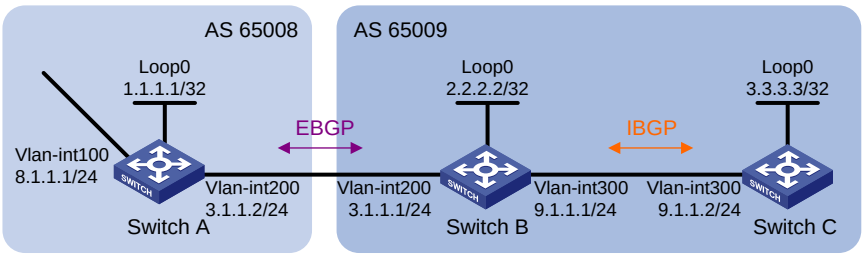
Basic IPv4 BGP network configuration examples

Example: Configuring basic BGP

Network configuration

As shown in Figure 24, all switches run BGP. Run EBGP between Switch A and Switch B, and run IBGP between Switch B and Switch C to allow Switch C to access network 8.1.1.0/24 connected to Switch A.

Figure 24 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure IBGP:
   - To prevent route flapping caused by port state changes, this example uses loopback interfaces to establish IBGP connections.
   - Loopback interfaces are virtual interfaces. Therefore, use the **peer connect-interface** command to specify the loopback interface as the source interface for establishing BGP connections.
   - Enable OSPF in AS 65009 to ensure that Switch B can communicate with Switch C through loopback interfaces.

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 3.3.3.3 as-number 65009
[SwitchB-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 3.3.3.3 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
[SwitchB] ospf 1
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[SwitchB-ospf-1-area-0.0.0.0] network 9.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65009
```

```
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 2.2.2.2 as-number 65009
[SwitchC-bgp-default] peer 2.2.2.2 connect-interface loopback 0
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 2.2.2.2 enable
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
[SwitchC] ospf 1
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[SwitchC-ospf-1-area-0.0.0.0] network 9.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
[SwitchC] display bgp peer ipv4

 BGP local router ID : 3.3.3.3
 Local AS number : 65009
 Total number of peers : 1                 Peers in established state : 1

  * - Dynamically created peer
  ^ - Peer created through link-local address
  Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

  2.2.2.2              65009        2        2    0       0 00:00:13 Established
```

The output shows that Switch C has established an IBGP peer relationship with Switch B.

3. Configure EBGP:

- The EBGP peers, Switch A and Switch B (usually belong to different carriers), are located in different ASs. Typically, their loopback interfaces are not reachable to each other, so directly connected interfaces are used for establishing EBGP sessions.
- To enable Switch C to access the network 8.1.1.0/24 connected directly to Switch A, inject network 8.1.1.0/24 to the BGP routing table of Switch A.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 3.1.1.1 as-number 65009
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 3.1.1.1 enable
[SwitchA-bgp-default-ipv4] network 8.1.1.0 24
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp-default] peer 3.1.1.2 as-number 65008
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 3.1.1.2 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Display BGP peer information on Switch B.

```
[SwitchB] display bgp peer ipv4

 BGP local router ID : 2.2.2.2
 Local AS number : 65009
 Total number of peers : 2                 Peers in established state : 2

  * - Dynamically created peer
  ^ - Peer created through link-local address
  Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

  3.3.3.3              65009        4        4    0       0 00:02:49 Established
  3.1.1.2              65008        2        2    0       0 00:00:05 Established
```

The output shows that Switch B has established an IBGP peer relationship with Switch C and an EBGP peer relationship with Switch A.

Display the BGP routing table on Switch A.

```
[SwitchA] display bgp routing-table ipv4

 Total number of routes: 1

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >  8.1.1.0/24         8.1.1.1         0                     32768   i
```

Display the BGP routing table on Switch B.

```
[SwitchB] display bgp routing-table ipv4

 Total number of routes: 1

 BGP local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 8.1.1.0/24         3.1.1.2         0                     0       65008i
```

Display the BGP routing table on Switch C.

```
[SwitchC] display bgp routing-table ipv4

 Total number of routes: 1
```

```
BGP local router ID is 3.3.3.3
Status codes: * - valid, > - best, d - dampened, h - history,
              s - suppressed, S - stale, i - internal, e - external
              a – additional-path
      Origin: i - IGP, e - EGP, ? - incomplete

    Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

  i 8.1.1.0/24         3.1.1.2         0          100        0       65008i
```

The outputs show that Switch A has learned no route to AS 65009, and Switch C has learned network 8.1.1.0, but the next hop 3.1.1.2 is unreachable. As a result, the route is invalid.

4. Redistribute direct routes:

Configure BGP to redistribute direct routes on Switch B, so that Switch A can obtain the route to 9.1.1.0/24, and Switch C can obtain the route to 3.1.1.0/24.

Configure Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] import-route direct
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Display the BGP routing table on Switch A.

```
[SwitchA] display bgp routing-table ipv4

 Total number of routes: 4

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 2.2.2.2/32         3.1.1.1         0                     0       65009?
* >e 3.1.1.0/24         3.1.1.1         0                     0       65009?
* >  8.1.1.0/24         8.1.1.1         0                     32768   i
* >e 9.1.1.0/24         3.1.1.1         0                     0       65009?
```

Two routes, 2.2.2.2/32 and 9.1.1.0/24, have been added in Switch A's routing table.

Display the BGP routing table on Switch C.

```
[SwitchC] display bgp routing-table ipv4

 Total number of routes: 4

 BGP local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete
```

```
     Network          NextHop        MED        LocPrf     PrefVal Path/Ogn

* >i 2.2.2.2/32      2.2.2.2        0          100        0       ?
* >i 3.1.1.0/24      2.2.2.2        0          100        0       ?
* >i 8.1.1.0/24      3.1.1.2        0          100        0       65008i
* >i 9.1.1.0/24      2.2.2.2        0          100        0       ?
```

The output shows that the route 8.1.1.0 becomes valid with the next hop as Switch A.

Verifying the configuration

Verify that Switch C can ping 8.1.1.1.

```
[SwitchC] ping 8.1.1.1
Ping 8.1.1.1 (8.1.1.1): 56 data bytes, press CTRL+C to break
56 bytes from 8.1.1.1: icmp_seq=0 ttl=254 time=10.000 ms
56 bytes from 8.1.1.1: icmp_seq=1 ttl=254 time=4.000 ms
56 bytes from 8.1.1.1: icmp_seq=2 ttl=254 time=4.000 ms
56 bytes from 8.1.1.1: icmp_seq=3 ttl=254 time=3.000 ms
56 bytes from 8.1.1.1: icmp_seq=4 ttl=254 time=3.000 ms
--- Ping statistics for 8.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 3.000/4.800/10.000/2.638 ms
```

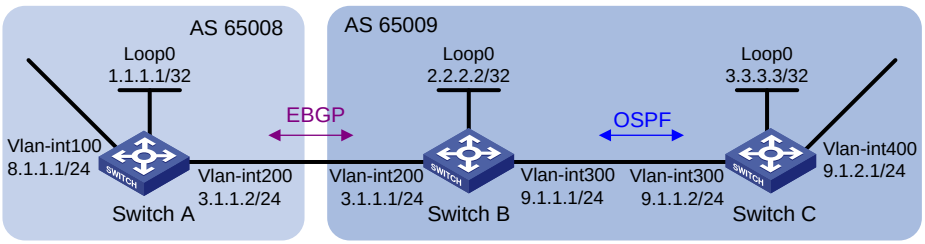
Example: Configuring BGP and IGP route redistribution

Network configuration

As shown in Figure 25, all devices of company A belong to AS 65008, and all devices of company B belong to AS 65009.

Configure BGP and IGP route redistribution to allow Switch A to access network 9.1.2.0/24 in AS 65009, and Switch C to access network 8.1.1.0/24 in AS 65008.

Figure 25 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure OSPF:

   Enable OSPF in AS 65009, so Switch B can obtain the route to 9.1.2.0/24.

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf 1
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[SwitchB-ospf-1-area-0.0.0.0] network 9.1.1.0 0.0.0.255
```

```
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf 1
[SwitchC-ospf-1] import-route direct
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 9.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

3. Configure the EBGP connection:

Configure the EBGP connection and inject network 8.1.1.0/24 to the BGP routing table of Switch A, so that Switch B can obtain the route to 8.1.1.0/24.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 3.1.1.1 as-number 65009
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 3.1.1.1 enable
[SwitchA-bgp-default-ipv4] network 8.1.1.0 24
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 3.1.1.2 as-number 65008
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 3.1.1.2 enable
```

4. Configure BGP and IGP route redistribution:

- Configure BGP to redistribute routes from OSPF on Switch B, so Switch A can obtain the route to 9.1.2.0/24.
- Configure OSPF to redistribute routes from BGP on Switch B, so Switch C can obtain the route to 8.1.1.0/24.

Configure route redistribution between BGP and OSPF on Switch B.

```
[SwitchB-bgp-default-ipv4] import-route ospf 1
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
[SwitchB] ospf 1
[SwitchB-ospf-1] import-route bgp
[SwitchB-ospf-1] quit
```

Display the BGP routing table on Switch A.

```
[SwitchA] display bgp routing-table ipv4

 Total number of routes: 3

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
```

```
                s - suppressed, S - stale, i - internal, e - external
                a – additional-path
        Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 3.3.3.3/32         3.1.1.1         1                     0       65009?
* >  8.1.1.0/24         8.1.1.1         0                     32768   i
* >e 9.1.2.0/24         3.1.1.1         1                     0       65009?
```

Display the OSPF routing table on Switch C.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 3.3.3.3
                  Routing Table

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 9.1.1.0/24         1        Transit 9.1.1.2         3.3.3.3         0.0.0.0
 2.2.2.2/32         1        Stub    9.1.1.1         2.2.2.2         0.0.0.0

 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop         AdvRouter
 8.1.1.0/24         1        Type2   1           9.1.1.1         2.2.2.2

 Total Nets: 3
 Intra Area: 2  Inter Area: 0  ASE: 1  NSSA: 0
```

Verifying the configuration

Use ping for verification.

```
[SwitchA] ping -a 8.1.1.1 9.1.2.1
Ping 9.1.2.1 (9.1.2.1) from 8.1.1.1: 56 data bytes, press CTRL+C to break
56 bytes from 9.1.2.1: icmp_seq=0 ttl=254 time=10.000 ms
56 bytes from 9.1.2.1: icmp_seq=1 ttl=254 time=12.000 ms
56 bytes from 9.1.2.1: icmp_seq=2 ttl=254 time=2.000 ms
56 bytes from 9.1.2.1: icmp_seq=3 ttl=254 time=7.000 ms
56 bytes from 9.1.2.1: icmp_seq=4 ttl=254 time=9.000 ms

--- Ping statistics for 9.1.2.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 2.000/8.000/12.000/3.406 ms
[SwitchC] ping -a 9.1.2.1 8.1.1.1
Ping 8.1.1.1 (8.1.1.1) from 9.1.2.1: 56 data bytes, press CTRL+C to break
56 bytes from 8.1.1.1: icmp_seq=0 ttl=254 time=9.000 ms
56 bytes from 8.1.1.1: icmp_seq=1 ttl=254 time=4.000 ms
56 bytes from 8.1.1.1: icmp_seq=2 ttl=254 time=3.000 ms
56 bytes from 8.1.1.1: icmp_seq=3 ttl=254 time=3.000 ms
56 bytes from 8.1.1.1: icmp_seq=4 ttl=254 time=3.000 ms

--- Ping statistics for 8.1.1.1 ---
```

```
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 3.000/4.400/9.000/2.332 ms
```

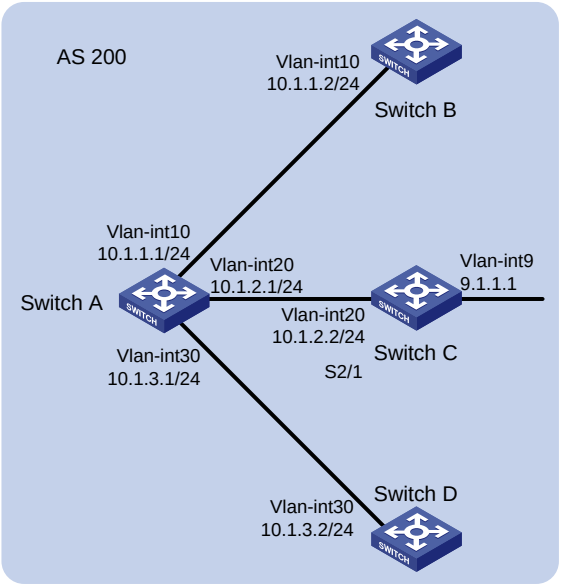
Example: Configuring dynamic BGP peers

Network configuration

As shown in Figure 26, Switch A needs to establish IBGP peer relationships with Switch B, Switch C, and Switch D in network 10.1.0.0/16. Configure dynamic BGP peers to simplify the configuration.

Configure Switch A as the route reflector, and configure Switch B, Switch C, and Switch D as its clients.

Figure 26 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure IBGP peer relationship:

   # Configure Switch A to establish dynamic BGP peer relationships with switches in network 10.1.0.0/16.

   ```
   <SwitchA> system-view
   [SwitchA] bgp 200
   [SwitchA-bgp-default] router-id 1.1.1.1
   [SwitchA-bgp-default] peer 10.1.0.0 16 as-number 200
   [SwitchA-bgp-default] address-family ipv4
   [SwitchA-bgp-default-ipv4] peer 10.1.0.0 16 enable
   ```

   # Configure Switch B to establish an IBGP peer relationship with Switch A.

   ```
   <SwitchB> system-view
   [SwitchB] bgp 200
   [SwitchB-bgp-default] router-id 2.2.2.2
   [SwitchB-bgp-default] peer 10.1.1.1 as-number 200
   [SwitchB-bgp-default] address-family ipv4
   [SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
   ```

   # Configure Switch C to establish an IBGP peer relationship with Switch A.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 10.1.2.1 as-number 200
[SwitchC-bgp-default] address-family ipv4
[SwitchC-bgp-default-ipv4] peer 10.1.2.1 enable
```

Configure Switch D to establish an IBGP peer relationship with Switch A.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] peer 10.1.3.1 as-number 200
[SwitchD-bgp-default] address-family ipv4
[SwitchD-bgp-default-ipv4] peer 10.1.3.1 enable
```

Display BGP peer information on Switch A. The output shows that Switch A has established IBGP peer relationships with Switch B, Switch C, and Switch D.

```
[SwitchA] display bgp peer ipv4

 BGP local router ID : 1.1.1.1
 Local AS number : 200

 Total number of peers : 3                 Peers in established state : 3

 * - Dynamically created peer
 ^ - Peer created through link-local address
 Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

*10.1.1.2               200        7       10    0       0 00:06:09 Established
*10.1.2.2               200        7       10    0       0 00:06:09 Established
*10.1.3.2               200        7       10    0       0 00:06:09 Established
```

3. Configure Switch A as the route reflector, and configure peers in network 10.1.0.0/16 as its clients.

```
[SwitchA-bgp-default-ipv4] peer 10.1.0.0 16 reflect-client
```

4. Configure Switch C to advertise network 9.1.1.0/24.

```
[SwitchC-bgp-default-ipv4] network 9.1.1.0 24
```

Verifying the configuration

Verify that route 9.1.1.0/24 exists in the BGP routing table on Switch A, Switch B, and Switch D. This example uses Switch A.

```
[SwitchA-bgp-default] display bgp routing-table ipv4

 Total Number of Routes: 1

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

    Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
```

```
*  i 9.1.1.0/24        10.1.2.2         0          100        0       ?
```

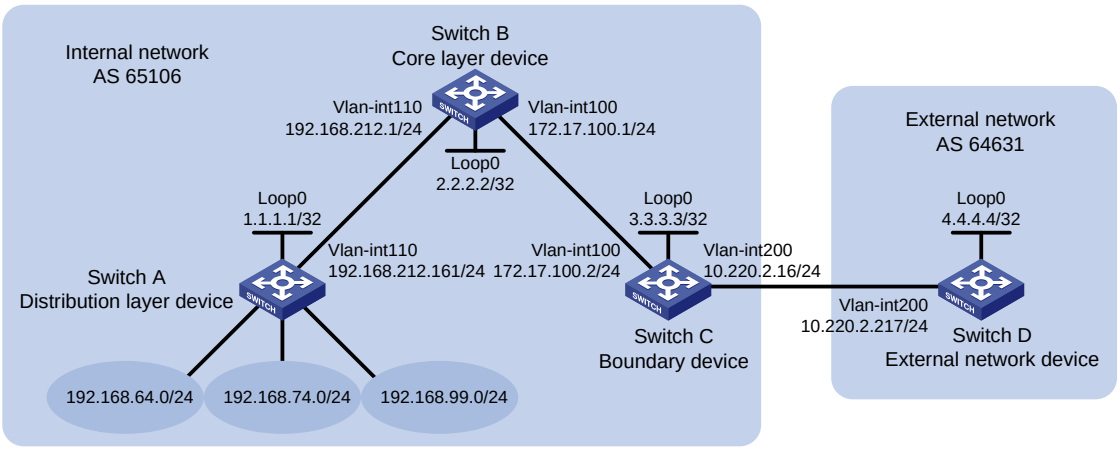
Example: Configuring BGP route summarization

Network configuration

As shown in Figure 27, run EBGP between Switch C and Switch D, so the internal network and external network can communicate with each other.

- In AS 65106, perform the following configurations so the devices in the internal network can communicate:
  - Configure static routing between Switch A and Switch B.
  - Configure OSPF between Switch B and Switch C.
  - Configure OSPF to redistribute static routes.
- Configure route summarization on Switch C so BGP advertises a summary route instead of advertising routes to the 192.168.64.0/24, 192.168.74.0/24, and 192.168.99.0/24 networks to Switch D.

Figure 27 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure static routing between Switch A and Switch B:

   # Configure a default route with the next hop 192.168.212.1 on Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] ip route-static 0.0.0.0 0 192.168.212.1
   ```

   # Configure static routes to 192.168.64.0/24, 192.168.74.0/24, and 192.168.99.0/24 with the same next hop 192.168.212.161 on Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] ip route-static 192.168.64.0 24 192.168.212.161
   [SwitchB] ip route-static 192.168.74.0 24 192.168.212.161
   [SwitchB] ip route-static 192.168.99.0 24 192.168.212.161
   ```
3. Configure OSPF between Switch B and Switch C and configure OSPF on Switch B to redistribute static routes:

   # Configure OSPF to advertise the local network and enable OSPF to redistribute static routes on Switch B.

   ```
   [SwitchB] ospf
   ```

```
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 172.17.100.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] import-route static
[SwitchB-ospf-1] quit
```

Configure OSPF to advertise the local networks on Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 172.17.100.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 10.220.2.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Display the IP routing table on Switch C.

```
[SwitchC] display ip routing-table protocol ospf

Summary count : 5

OSPF Routing table Status : <Active>
Summary count : 3

Destination/Mask    Proto   Pre  Cost         NextHop         Interface
192.168.64.0/24     O_ASE2  150  1            172.17.100.1    Vlan100
192.168.74.0/24     O_ASE2  150  1            172.17.100.1    Vlan100
192.168.99.0/24     O_ASE2  150  1            172.17.100.1    Vlan100

OSPF Routing table Status : <Inactive>
Summary count : 2

Destination/Mask    Proto    Pre  Cost         NextHop         Interface
10.220.2.0/24       O_INTRA  10   1            10.220.2.16     Vlan200
172.17.100.0/24     O_INTRA  10   1            172.17.100.2    Vlan100
```

The output shows that Switch C has learned routes to 192.168.64.0/24, 192.168.99.0/24, and 192.168.64.0/18 through OSPF.

4. Configure BGP between Switch C and Switch D and configure BGP on Switch C to redistribute OSPF routes:

On Switch C, enable BGP, specify Switch D as an EBGP peer, and configure BGP to redistribute OSPF routes.

```
[SwitchC] bgp 65106
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 10.220.2.217 as-number 64631
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 10.220.2.217 enable
[SwitchC-bgp-default-ipv4] import-route ospf
```

Enable BGP, and configure Switch C as an EBGP peer on Switch D.

```
[SwitchD] bgp 64631
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] peer 10.220.2.16 as-number 65106
[SwitchD-bgp-default] address-family ipv4 unicast
```

```
[SwitchD-bgp-default-ipv4] peer 10.220.2.16 enable
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

Display the IP routing table on Switch D.

```
[SwitchD] display ip routing-table protocol bgp

Summary count : 3

BGP Routing table Status : <Active>
Summary count : 3

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
192.168.64.0/24     BGP    255  1            10.220.2.16     Vlan200
192.168.74.0/24     BGP    255  1            10.220.2.16     Vlan200
192.168.99.0/24     BGP    255  1            10.220.2.16     Vlan200

BGP Routing table Status : <Inactive>
Summary count : 0
```

The output shows that Switch D has learned routes to 192.168.64.0/24, 192.168.74.0/24, and 192.168.99.0/24 through BGP.

Verify that Switch D can ping hosts on networks 192.168.74.0/24, 192.168.99.0/24, and 192.168.64.0/18. (Details not shown.)

5. Configure route summarization on Switch C to summarize 192.168.64.0/24, 192.168.74.0/24, and 192.168.99.0/24 into a single route 192.168.64.0/18, and disable advertisement of specific routes.

```
[SwitchC-bgp-default-ipv4] aggregate 192.168.64.0 18 detail-suppressed
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Verifying the configuration

Display IP routing table on Switch C.

```
[SwitchC] display ip routing-table | include 192.168
192.168.64.0/18    BGP     130  0            127.0.0.1       NULL0
192.168.64.0/24    O_ASE2  150  1            172.17.100.1    Vlan100
192.168.74.0/24    O_ASE2  150  1            172.17.100.1    Vlan100
192.168.99.0/24    O_ASE2  150  1            172.17.100.1    Vlan100
```

The output shows that Switch C has a summary route 192.168.64.0/18 with the output interface Null0.

Display IP routing table on Switch D.

```
[SwitchD] display ip routing-table protocol bgp

Summary count : 1

BGP Routing table Status : <Active>
Summary count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
192.168.64.0/18     BGP    255  0            10.220.2.16     Vlan200
```

```
BGP Routing table Status : <Inactive>
Summary count : 0
```

The output shows that Switch D has only one route 192.168.64.0/18 to AS 65106.

Verify that Switch D can ping the hosts on networks 192.168.64.0/24, 192.168.74.0/24, and 192.168.99.0/24. (Details not shown.)

Example: Configuring multicast BGP

Network configuration

As shown in Figure 28, OSPF runs within AS 100 and AS 200 to ensure intra-AS connectivity. MBGP runs between the two ASs to exchange IPv4 unicast routes used for RPF check.

- Configure the Loopback 0 interface of Switch A and Switch B as the C-BSR and C-RP.
- Configure Switch A and Switch B to establish a Multicast Source Discovery Protocol (MSDP) peer relationship through MBGP, so that the receiver can receive multicast traffic from the source.

Figure 28 Network diagram

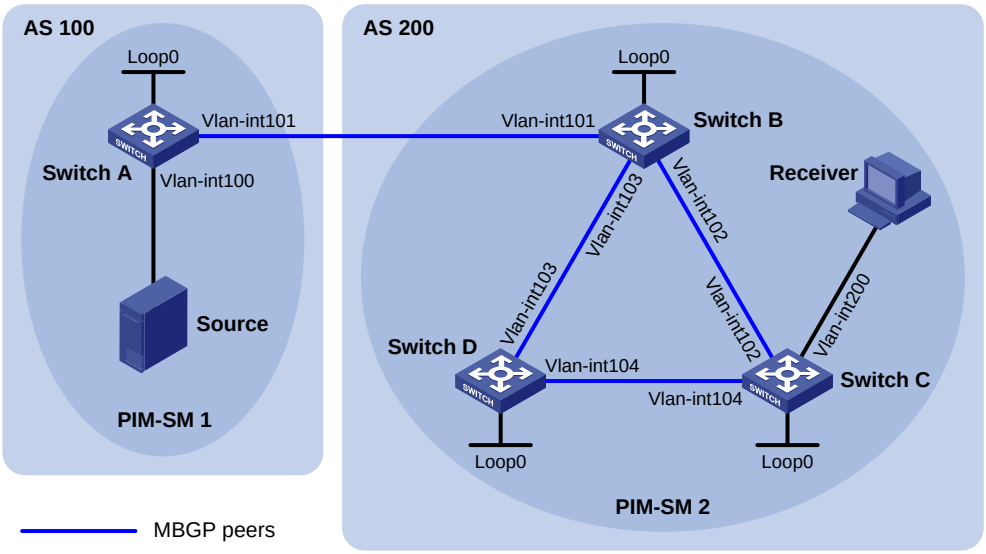


Table 9 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source	N/A	10.110.1.100/24	Switch C	Vlan-int200	10.110.2.1/24
Switch A	Vlan-int100	10.110.1.1/24		Vlan-int102	192.168.2.2/24
	Vlan-int101	192.168.1.1/24		Vlan-int104	192.168.4.1/24
	Loop0	1.1.1.1/32		Loop0	3.3.3.3/32
Switch B	Vlan-int101	192.168.1.2/24	Switch D	Vlan-int103	192.168.3.2/24
	Vlan-int102	192.168.2.1/24		Vlan-int104	192.168.4.2/24
	Vlan-int103	192.168.3.1/24		Loop0	4.4.4.4/32
	Loop0	2.2.2.2/32			

Procedure

1. Configure IP addresses for interfaces and configure OSPF (this example uses OSPF process 1) in AS 200 to ensure intra-AS connectivity. (Details not shown.)
2. Enable IP multicast routing, PIM-SM, and IGMP, and configure BSR boundaries:

   # On Switch A, enable multicast routing globally, and enable PIM-SM on interfaces.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
```

   # Configure Switch B and Switch D in the same way that Switch A was configured.

   # On Switch C, enable multicast routing globally.

```
<SwitchC> system-view
[SwitchC] multicast routing
[SwitchA-mrib] quit
```

   # Enable PIM-SM on interfaces, and enable IGMP on VLAN-interface 200.

```
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] pim sm
[SwitchC-Vlan-interface102] quit
[SwitchC] interface vlan-interface 104
[SwitchC-Vlan-interface104] pim sm
[SwitchC-Vlan-interface104] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] pim sm
[SwitchC-Vlan-interface200] igmp enable
[SwitchC-Vlan-interface200] quit
```

   # Configure the BSR boundary on Switch A.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim bsr-boundary
[SwitchA-Vlan-interface101] quit
```

   # Configure the BSR boundary on Switch B.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim bsr-boundary
[SwitchB-Vlan-interface101] quit
```

3. Configure Loopback 0, C-BSR, and C-RP:

   # Configure the Loopback 0 interface and specify it as the C-BSR and C-RP on Switch A.

```
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.1 32
[SwitchA-LoopBack0] pim sm
[SwitchA-LoopBack0] quit
[SwitchA] pim
[SwitchA-pim] c-bsr 1.1.1.1
[SwitchA-pim] c-rp 1.1.1.1
```

```
[SwitchA-pim] quit
```

Configure the Loopback 0 interface and specify it as the C-BSR and C-RP on Switch B.

```
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.2 32
[SwitchB-LoopBack0] pim sm
[SwitchB-LoopBack0] quit
[SwitchB] pim
[SwitchB-pim] c-bsr 2.2.2.2
[SwitchB-pim] c-rp 2.2.2.2
[SwitchB-pim] quit
```

4. Configure BGP to establish BGP IPv4 multicast peers and redistribute routes:

On Switch A, establish an EBGP session to Switch B.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 192.168.1.2 as-number 200
```

Enable exchange of IPv4 unicast routes used for RPF check with Switch B.

```
[SwitchA-bgp-default] address-family ipv4 multicast
[SwitchA-bgp-default-mul-ipv4] peer 192.168.1.2 enable
```

Redistribute direct routes into BGP.

```
[SwitchA-bgp-default-mul-ipv4] import-route direct
[SwitchA-bgp-default-mul-ipv4] quit
[SwitchA-bgp-default] quit
```

On Switch B, establish an EBGP session to Switch A.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 192.168.1.1 as-number 100
```

Enable exchange of IPv4 unicast routes used for RPF check with Switch B.

```
[SwitchB-bgp-default] address-family ipv4 multicast
[SwitchB-bgp-default-mul-ipv4] peer 192.168.1.1 enable
```

Redistribute OSPF routes into BGP.

```
[SwitchB-bgp-default-mul-ipv4] import-route ospf 1
[SwitchB-bgp-default-mul-ipv4] quit
[SwitchB-bgp-default] quit
```

5. Configure MSDP peers:

Configure an MSDP peer on Switch A.

```
[SwitchA] msdp
[SwitchA-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchA-msdp] quit
```

Configure an MSDP peer on Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchB-msdp] quit
```

Verifying the configuration

Verify the BGP IPv4 multicast peer information on Switch B.

```
[SwitchB] display bgp peer ipv4 multicast

 BGP local router ID : 2.2.2.2
```

```
 Local AS number : 200
 Total number of peers : 3                 Peers in established state : 3

  Peer            AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.1.1    100       56       56     0       0 00:40:54 Established
```

Verify the MSDP peer information on Switch B.

```
[SwitchB] display msdp brief
Configured   Established  Listen       Connect      Shutdown     Disabled
1            1            0            0            0            0

Peer address    State       Up/Down time    AS         SA count    Reset count
192.168.1.1     Established 00:07:17        100        1           0
```

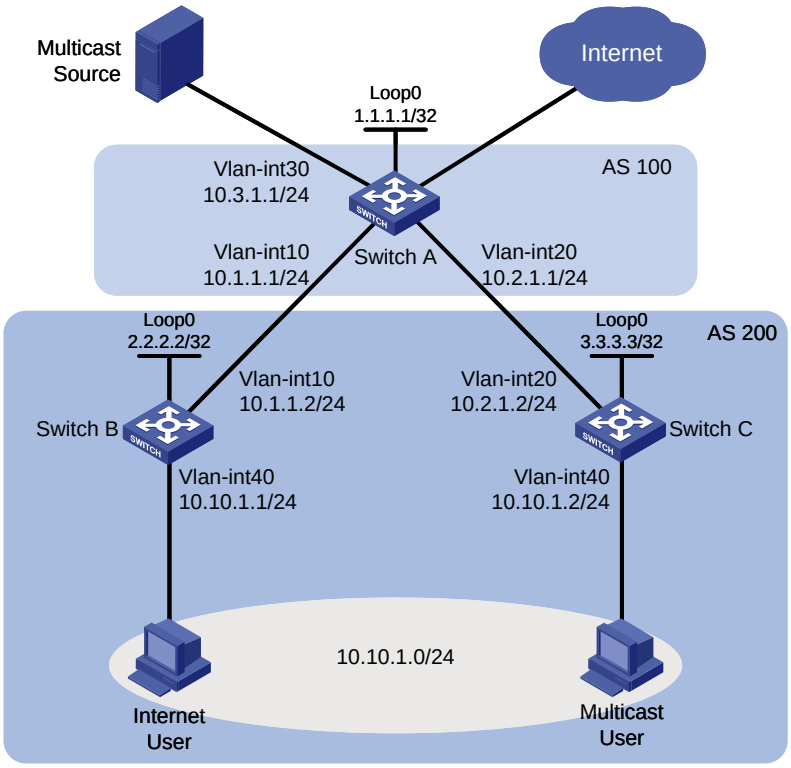
Example: Separating unicast traffic and multicast traffic

Network configuration

Switch A is located in AS 100 and is connected to a multicast source and the Internet. Switch B and Switch C run in AS 200 and are connected to Internet users and multicast users, respectively. Unicast traffic and multicast traffic need to be separated, which ensures that traffic from AS 100 to AS 200 is forwarded along the following paths:

- **Internet traffic**: **Switch A** > **Switch B** > **Internet user**
- **Multicast traffic**: **Switch A** > **Switch C** > **Multicast user**

Figure 29 Network diagram



Procedure

1. Configure IP addresses for interfaces and configure OSPF (this example uses OSPF process 1) in AS 200 to ensure intra-AS connectivity. (Details not shown.)
2. Establish BGP sessions.

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 10.1.1.2 as-number 200
[SwitchA-bgp-default] peer 10.2.1.2 as-number 200
[SwitchA-bgp-default] address-family ipv4 multicast
[SwitchA-bgp-default-mul-ipv4] peer 10.1.1.2 enable
[SwitchA-bgp-default-mul-ipv4] peer 10.2.1.2 enable
[SwitchA-bgp-default-mul-ipv4] quit
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
[SwitchA-bgp-default-ipv4] peer 10.2.1.2 enable
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 10.1.1.1 as-number 100
[SwitchB-bgp-default] peer 10.10.1.2 as-number 200
[SwitchB-bgp-default] address-family ipv4 multicast
[SwitchB-bgp-default-mul-ipv4] peer 10.1.1.1 enable
[SwitchB-bgp-default-mul-ipv4] peer 10.10.1.2 enable
[SwitchB-bgp-default-mul-ipv4] quit
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
[SwitchB-bgp-default-ipv4] peer 10.10.1.2 enable
[SwitchB-bgp-default-ipv4] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 10.2.1.1 as-number 100
[SwitchC-bgp-default] peer 10.10.1.1 as-number 200
[SwitchC-bgp-default] address-family ipv4 multicast
[SwitchC-bgp-default-mul-ipv4] peer 10.2.1.1 enable
[SwitchC-bgp-default-mul-ipv4] peer 10.10.1.1 enable
[SwitchC-bgp-default-mul-ipv4] quit
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 10.2.1.1 enable
[SwitchC-bgp-default-ipv4] peer 10.10.1.1 enable
[SwitchC-bgp-default-ipv4] quit
```

3. Configure BGP unicast routes used for unicast traffic forwarding, and BGP unicast routes used for multicast RPF check.

Configure Switch B.

```
[SwitchB-bgp-default] address-family ipv4 multicast
[SwitchB-bgp-default-mul-ipv4] network 2.2.2.2 32
[SwitchB-bgp-default-mul-ipv4] network 10.10.1.0 24
[SwitchB-bgp-default-mul-ipv4] quit
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] network 10.10.1.0 24
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
[SwitchC-bgp-default] address-family ipv4 multicast
[SwitchC-bgp-default-mul-ipv4] network 2.2.2.2 32
[SwitchC-bgp-default-mul-ipv4] network 10.10.1.0 24
[SwitchC-bgp-default-mul-ipv4] quit
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] network 10.10.1.0 24
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

4. Configure a routing policy on Switch B and Switch C separately, so the two routers can modify the MED values of routes before route advertisement. This operation ensures that Switch A selects the BGP unicast route from Switch B for unicast traffic forwarding and selects the BGP unicast route from Switch C for multicast RPF check.

Configure Switch B.

```
[SwitchB] acl basic 2077
[SwitchB-acl-ipv4-basic-2077] rule permit source 10.10.1.0 0.0.0.255
[SwitchB-acl-ipv4-basic-2077] quit
[SwitchB] route-policy b2a permit node 10
[SwitchB-route-policy-b2a-10] if-match ip address acl 2077
[SwitchB-route-policy-b2a-10] apply cost 200
[SwitchB-route-policy-b2a-10] quit
[SwitchB] route-policy b2a permit node 20
[SwitchB-route-policy-b2a-20] quit
[SwitchB] bgp 200
[SwitchB-bgp-default] address-family ipv4 multicast
[SwitchB-bgp-default-mul-ipv4] peer 10.1.1.1 route-policy b2a export
[SwitchB-bgp-default-mul-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
[SwitchC] acl basic 2077
[SwitchC-acl-ipv4-basic-2077] rule permit source 10.10.1.0 0.0.0.255
[SwitchC-acl-ipv4-basic-2077] quit
[SwitchC] route-policy b2a permit node 10
[SwitchC-route-policy-b2a-10] if-match ip address acl 2077
[SwitchC-route-policy-b2a-10] apply cost 200
[SwitchC-route-policy-b2a-10] quit
[SwitchC] route-policy b2a permit node 20
[SwitchC-route-policy-b2a-20] quit
```

```
[SwitchC] bgp 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 10.2.1.1 route-policy b2a export
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

5. Configure IP multicast.

Configure Switch A.

```
[SwitchA] multicast routing
[SwitchA-mrib] quit
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] pim sm
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] pim sm
[SwitchA-Vlan-interface20] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] pim sm
[SwitchA-LoopBack0] quit
[SwitchA] pim
[SwitchA-pim] c-bsr 1.1.1.1
[SwitchA-pim] c-rp 1.1.1.1
[SwitchA-pim] quit
```

Configure Switch B.

```
[SwitchB] multicast routing
[SwitchB-mrib] quit
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] pim sm
[SwitchB-Vlan-interface10] quit
[SwitchB] interface vlan-interface 40
[SwitchB-Vlan-interface40] pim sm
[SwitchB-Vlan-interface40] igmp enable
[SwitchB-Vlan-interface40] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] pim sm
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
[SwitchC] multicast routing
[SwitchC-mrib] quit
[SwitchC] interface vlan-interface 20
[SwitchC-Vlan-interface20] pim sm
[SwitchC-Vlan-interface20] quit
[SwitchC] interface vlan-interface 40
[SwitchC-Vlan-interface40] pim sm
[SwitchC-Vlan-interface40] igmp enable
[SwitchC-Vlan-interface40] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] pim sm
[SwitchC-LoopBack0] quit
```

```
[SwitchC] pim
[SwitchC-pim] c-bsr 3.3.3.3
[SwitchC-pim] c-rp 3.3.3.3
[SwitchC-pim] quit
```

6. Configure BSR boundaries.

Configure Switch A.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] pim bsr-boundary
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] pim bsr-boundary
[SwitchA-Vlan-interface20] quit
```

Configure Switch B.

```
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] pim bsr-boundary
[SwitchB-Vlan-interface10] quit
```

Configure Switch C.

```
[SwitchC] interface vlan-interface 20
[SwitchC-Vlan-interface20] pim bsr-boundary
[SwitchC-Vlan-interface20] quit
```

7. Configure MSDP peers.

Configure Switch A.

```
[SwitchA] msdp
[SwitchA-msdp] peer 10.1.1.2 connect-interface vlan-interface 10
[SwitchA-msdp] peer 10.2.1.2 connect-interface vlan-interface 20
[SwitchA-msdp] quit
```

Configure Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] peer 10.1.1.1 connect-interface vlan-interface 10
[SwitchB-msdp] quit
```

Configure Switch C.

```
[SwitchC] msdp
[SwitchC-msdp] peer 10.2.1.1 connect-interface vlan-interface 20
[SwitchC-msdp] quit
```

Verifying the configuration

View BGP IPv4 unicast routing information on Switch A.

```
[SwitchA] display bgp routing-table ipv4 unicast

 Total number of routes: 2

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
```

```
* >e  10.10.1.0/24      10.1.1.2        0                    0       200i
*  e                    10.2.1.2        200                  0       200i
```

View BGP IPv4 multicast routing information on Switch A.

```
[SwitchA] display bgp routing-table ipv4 multicast

 Total number of routes: 6

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e  2.2.2.2/32        10.1.1.2        0                     0       200i
*  e                    10.2.1.2                              0       200i
* >e  3.3.3.3/32        10.2.1.2        0                     0       200i
*  e                    10.1.1.2                              0       200i
* >e  10.10.1.0/24      10.2.1.2        0                     0       200i
*  e                    10.1.1.2        200                   0       200i
```

在 Switch A 上，查看 MSDP 对等体的建立情况。

According to the above command outputs, Switch A selects the BGP unicast route from Switch B for unicast traffic forwarding and selects the BGP unicast route from Switch C for multicast RPF check.

On Switch A, view MSDP peers.

```
[SwitchA] display msdp brief
Configured   Established  Listen       Connect      Shutdown     Disabled
2            2            0            0            0            0

Peer address    State       Up/Down time    AS         SA count   Reset count
10.1.1.2        Established 00:40:17        200        0          0
10.2.1.2        Established 00:41:17        200        0          0
```

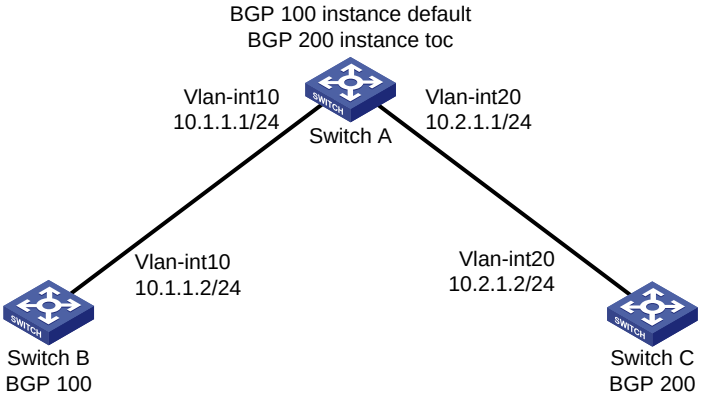
According to the above command output, Switch A has established an MSDP peer relationship with Switch B and Switch C separately.

Example: Configuring multiple BGP instances on one switch

Network configuration

To achieve separate management and maintenance of different switches' routes, Switch A uses different BGP instances to establish a BGP session to Switch B and Switch C separately. After Switch A receive routes from Switch B and Switch C, it does not advertise the routes of Switch B to Switch C or the routes of Switch C to Switch B.

Figure 30 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP connections.

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] router-id 10.1.1.1
[SwitchA-bgp-default] peer 10.1.1.2 as-number 100
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
[SwitchA] bgp 200 instance toc
[SwitchA-bgp-toc] router-id 10.2.1.1
[SwitchA-bgp-toc] peer 10.2.1.2 as-number 200
[SwitchA-bgp-toc] address-family ipv4 unicast
[SwitchA-bgp-toc-ipv4] peer 10.2.1.2 enable
[SwitchA-bgp-toc-ipv4] quit
[SwitchA-bgp-toc] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 100
[SwitchB-bgp-default] peer 10.1.1.1 as-number 100
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 10.2.1.1 as-number 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 10.2.1.1 enable
[SwitchC-bgp-default-ipv4] quit
```

```
[SwitchC-bgp-default] quit
```

3. Configure route advertisement.

Configure Switch B.

```
[SwitchB] ip route-static 192.168.10.10 32 null 0
[SwitchB] bgp 100
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] import-route static
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
[SwitchC] ip route-static 192.168.20.20 32 null 0
[SwitchC] bgp 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] import-route static
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Verifying the configuration

View the IP routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 12       Routes : 12

Destination/Mask   Proto   Pre Cost        NextHop         Interface
10.1.1.0/24        Direct  0   0           10.1.1.1        Vlan10
10.1.1.1/32        Direct  0   0           127.0.0.1       Vlan10
10.1.1.255/32      Direct  0   0           10.1.1.1        Vlan10
10.2.1.0/24        Direct  0   0           10.2.1.1        Vlan20
10.2.1.1/32        Direct  0   0           127.0.0.1       Vlan20
10.2.1.255/32      Direct  0   0           10.2.1.1        Vlan20
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
192.168.10.10/32   BGP     255 0           10.1.1.2        Vlan10
192.168.20.20/32   BGP     255 0           10.2.1.2        Vlan20
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

View the IP routing table of Switch B.

```
[SwitchB] display ip routing-table

Destinations : 8        Routes : 8

Destination/Mask   Proto   Pre Cost        NextHop         Interface
10.1.1.0/24        Direct  0   0           10.1.1.2        Vlan10
10.1.1.2/32        Direct  0   0           127.0.0.1       Vlan10
10.1.1.255/32      Direct  0   0           10.1.1.2        Vlan10
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

```
192.168.10.10/32   Static  60  0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

View the IP routing table of Switch C.

```
[SwitchC] display ip routing-table

Destinations : 8        Routes : 8

Destination/Mask    Proto   Pre Cost        NextHop         Interface
10.2.1.0/24         Direct  0   0           10.2.1.2        Vlan20
10.2.1.2/32         Direct  0   0           127.0.0.1       Vlan20
10.2.1.255/32       Direct  0   0           10.2.1.2        Vlan20
127.0.0.0/8         Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
192.168.20.20/32    Static  60  0           0.0.0.0         NULL0
255.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
```

According to the above command outputs, Switch A can receive routes from Switch B and Switch C, but the routing table of Switch B or Switch C only has their own routes.

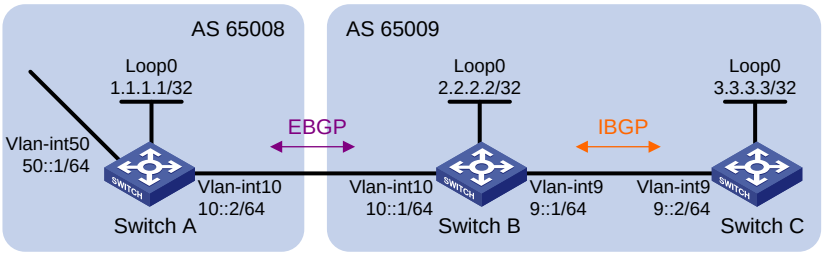
Basic IPv6 BGP network configuration examples

Example: Configuring IPv6 BGP basics

Network configuration

As shown in Figure 31, all switches run BGP. Run EBGP between Switch A and Switch B, and run IBGP between Switch B and Switch C to allow Switch C to access network 50::/64 connected to Switch A.

Figure 31 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure IBGP:

   # Configure Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] bgp 65009
   [SwitchB-bgp-default] router-id 2.2.2.2
   [SwitchB-bgp-default] peer 9::2 as-number 65009
   [SwitchB-bgp-default] address-family ipv6
   [SwitchB-bgp-default-ipv6] peer 9::2 enable
   ```

```
[SwitchB-bgp-default-ipv6] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 9::1 as-number 65009
[SwitchC-bgp-default] address-family ipv6
[SwitchC-bgp-default-ipv6] peer 9::1 enable
```

3. Configure EBGP:

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 10::1 as-number 65009
[SwitchA-bgp-default] address-family ipv6
[SwitchA-bgp-default-ipv6] peer 10::1 enable
```

Configure Switch B.

```
[SwitchB-bgp-default] peer 10::2 as-number 65008
[SwitchB-bgp-default] address-family ipv6
[SwitchB-bgp-default-ipv6] peer 10::2 enable
```

4. Inject network routes to the BGP routing table:

Configure Switch A.

```
[SwitchA-bgp-default-ipv6] network 10:: 64
[SwitchA-bgp-default-ipv6] network 50:: 64
[SwitchA-bgp-default-ipv6] quit
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
[SwitchB-bgp-default-ipv6] network 10:: 64
[SwitchB-bgp-default-ipv6] network 9:: 64
[SwitchB-bgp-default-ipv6] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
[SwitchC-bgp-default-ipv6] network 9:: 64
[SwitchC-bgp-default-ipv6] quit
[SwitchC-bgp-default] quit
```

Verifying the configuration

Display IPv6 BGP peer information on Switch B.

```
[SwitchB] display bgp peer ipv6

 BGP local router ID: 2.2.2.2
 Local AS number: 65009
 Total number of peers: 2                 Peers in established state: 2

  * - Dynamically created peer
  ^ - Peer created through link-local address
  Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State
```

```
 9::2                65009       41       43    0       1 00:29:00 Established
 10::2               65008       38       38    0       2 00:27:20 Established
```

The output shows that Switch A and Switch B have established an EBGP connection, and Switch B and Switch C have established an IBGP connection.

Display IPv6 BGP routing table information on Switch A.

```
[SwitchA] display bgp routing-table ipv6

 Total number of routes: 4

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

* >e Network : 9::                                      PrefixLen : 64
     NextHop : 10::1                                    LocPrf    :
     PrefVal : 0                                        OutLabel  : NULL
     MED     : 0
     Path/Ogn: 65009i

* >  Network : 10::                                     PrefixLen : 64
     NextHop : ::                                       LocPrf    :
     PrefVal : 32768                                    OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

*  e Network : 10::                                     PrefixLen : 64
     NextHop : 10::1                                    LocPrf    :
     PrefVal : 0                                        OutLabel  : NULL
     MED     : 0
     Path/Ogn: 65009i

* >  Network : 50::                                     PrefixLen : 64
     NextHop : ::                                       LocPrf    :
     PrefVal : 32768                                    OutLabel  : NULL
     MED     : 0
     Path/Ogn: i
```

The output shows that Switch A has learned routing information of AS 65009.

Display IPv6 BGP routing table information on Switch C.

```
[SwitchC] display bgp routing-table ipv6

 Total number of routes: 4

 BGP local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
```

```
       Origin: i - IGP, e - EGP, ? - incomplete

* >  Network : 9::                                        PrefixLen : 64
     NextHop : ::                                        LocPrf    :
     PrefVal : 32768                                     OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

*  i Network : 9::                                        PrefixLen : 64
     NextHop : 9::1                                      LocPrf    : 100
     PrefVal : 0                                         OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >i Network : 10::                                       PrefixLen : 64
     NextHop : 9::1                                      LocPrf    : 100
     PrefVal : 0                                         OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >i Network : 50::                                       PrefixLen : 64
     NextHop : 10::2                                     LocPrf    : 100
     PrefVal : 0                                         OutLabel  : NULL
     MED     : 0
     Path/Ogn: 65008i
```

The output shows that Switch C has learned the route 50::/64.

Verify that Switch C can ping hosts on network 50::/64. (Details not shown.)

Example: Configuring IPv6 multicast BGP

Network configuration

As shown in Figure 32, OSPFv3 runs within AS 100 and AS 200 to ensure intra-AS connectivity. IPv6 MBGP runs between the two ASs to exchange IPv6 unicast routes used for RPF check.

Enable Anycast RP on Switch A and Switch B.

Figure 32 Network diagram

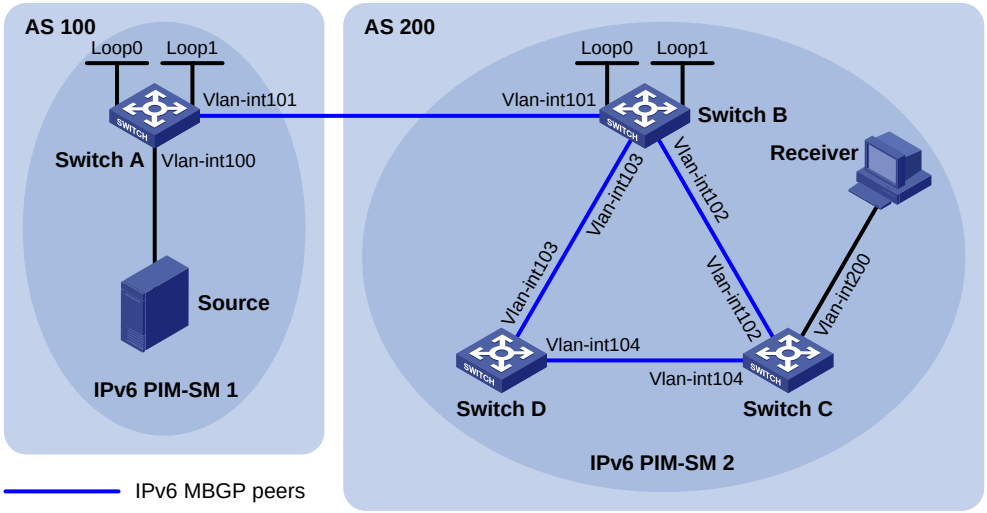


Table 10 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source	-	1002::100/64	Switch B	Vlan-int101	1001::2/64
Switch A	Vlan-int100	1002::1/64		Vlan-int102	2001::1/64
	Vlan-int101	1001::1/64		Vlan-int103	2002::1/64
	Loop0	1:1::1/128		Loop0	1:1::1/128
	Loop1	1:1::2/128		Loop1	2:2::2/128
Switch C	Vlan-int200	3002::1/64	Switch D	Vlan-int103	2002::2/64
	Vlan-int102	2001::2/64		Vlan-int104	3001::2/64
	Vlan-int104	3001::1/64			

Procedure

1. Configure IPv6 addresses for interfaces and configure OSPFv3 (this example uses OSPFv3 process 1) in AS 200 to ensure intra-AS connectivity. (Details not shown.)
2. Enable IPv6 multicast routing, IPv6 PIM-SM, and MLD, and configure BSR boundaries:

   # On Switch A, enable IPv6 multicast routing globally, and enable IPv6 PIM-SM on interfaces.

```
<SwitchA> system-view
[SwitchA] ipv6 multicast routing
[SwitchA-mrib6] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipv6 pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ipv6 pim sm
[SwitchA-LoopBack0] quit
```

   # Configure Switch B and Switch D in the same way that Switch A was configured.

On Switch C, enable IPv6 multicast routing globally.

```
<SwitchC> system-view
[SwitchC] ipv6 multicast routing
[SwitchC-mrib6] quit
```

Enable IPv6 PIM-SM on interfaces, and enable MLD on VLAN-interface 200.

```
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] ipv6 pim sm
[SwitchC-Vlan-interface102] quit
[SwitchC] interface vlan-interface 104
[SwitchC-Vlan-interface104] ipv6 pim sm
[SwitchC-Vlan-interface104] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] ipv6 pim sm
[SwitchC-Vlan-interface200] mld enable
[SwitchC-Vlan-interface200] quit
```

Configure the BSR boundary on Switch A.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 pim bsr-boundary
[SwitchA-Vlan-interface101] quit
```

Configure the BSR boundary on Switch B.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 pim bsr-boundary
[SwitchB-Vlan-interface101] quit
```

3. Enable Anycast RP, and specify C-BSR and C-RP:

Configure Switch A.

```
[SwitchA] ipv6 pim
[SwitchA-pim6] anycast-rp 1:1::1 1:1::2
[SwitchA-pim6] anycast-rp 1:1::1 2:2::2
[SwitchA-pim6] c-bsr 1:1::1
[SwitchA-pim6] c-rp 1:1::1
[SwitchA-pim6] quit
```

Configure Switch B.

```
[SwitchB] ipv6 pim
[SwitchB-pim6] anycast-rp 1:1::1 1:1::2
[SwitchB-pim6] anycast-rp 1:1::1 2:2::2
[SwitchB-pim6] c-bsr 1:1::1
[SwitchB-pim6] c-rp 1:1::1
[SwitchB-pim6] quit
```

4. Configure BGP to establish BGP IPv6 multicast peers and redistribute routes:

On Switch A, establish an EBGP session to Switch B.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 1001::2 as-number 200
```

Enable exchange of IPv6 unicast routes used for RPF check with Switch B.

```
[SwitchA-bgp-default] address-family ipv6 multicast
[SwitchA-bgp-default-mul-ipv6] peer 1001::2 enable
```

Redistribute direct routes into BGP.

```
[SwitchA-bgp-default-mul-ipv6] import-route direct
[SwitchA-bgp-default-mul-ipv6] quit
```

On Switch B, establish an EBGP session to Switch A.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 1001::1 as-number 100
```

Enable exchange of IPv6 unicast routes used for RPF check with Switch B.

```
[SwitchB-bgp-default] address-family ipv6 multicast
[SwitchB-bgp-default-mul-ipv6] peer 1001::1 enable
```

Redistribute OSPFv3 routes into BGP.

```
[SwitchB-bgp-default-mul-ipv6] import-route ospfv3 1
[SwitchB-bgp-default-mul-ipv6] quit
```

5. Establish BGP IPv6 unicast peer relationships between Switch A and Switch B and redistribute routes:

On Switch A, enable BGP to exchange IPv6 unicast routes with Switch B, and redistribute direct routes into BGP.

```
[SwitchA-bgp-default] address-family ipv6 unicast
[SwitchA-bgp-default-ipv6] peer 1001::2 enable
[SwitchA-bgp-default-ipv6] import-route direct
[SwitchA-bgp-default-ipv6] quit
[SwitchA-bgp-default] quit
```

On Switch B, enable BGP to exchange IPv6 unicast routes with Switch A, and redistribute direct routes into BGP.

```
[SwitchB-bgp-default] address-family ipv6 unicast
[SwitchB-bgp-default-ipv6] peer 1001::1 enable
[SwitchB-bgp-default-ipv6] import-route direct
[SwitchB-bgp-default-ipv6] quit
[SwitchB-bgp-default] quit
```

Verifying the configuration

Verify the BGP IPv6 multicast peer information on Switch B.

```
[SwitchB] display bgp peer ipv6 multicast

 BGP local router ID : 2.2.2.2
 Local AS number : 200
 Total number of peers : 3                 Peers in established state : 3

  Peer        AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  1001::1     100       56       56      0        0 00:40:54 Established
```

Verify the RPF information for the multicast source on Switch B.

```
[SwitchB] display ipv6 multicast rpf-info 1002::1
 RPF information about source 1002::1:
     RPF interface: Vlan-interface101, RPF neighbor: 1001::1
     Referenced prefix/prefix length: 1002::/64
     Referenced route type: mbgp
     Route selection rule: preference-preferred
     Load splitting rule: disable
```

Troubleshooting BGP

State of the connection to a peer cannot become established

Symptom

The **display bgp peer ipv4 unicast** or **display bgp peer ipv6 unicast** command output shows that the state of the connection to a peer cannot become established.

Analysis

To become BGP peers, any two routers must establish a TCP connection using port 179 and exchange OPEN messages successfully.

Solution

1. To resolve the problem:
   a. Use the **display current-configuration** command to verify the current configuration, and verify that the peer's AS number is correct.
   b. Use the **display bgp peer ipv4 unicast** or **display bgp peer ipv6 unicast** command to verify that the peer's IP/IPv6 address is correct.
   c. If a loopback interface is used, verify that the loopback interface is specified with the **peer connect-interface** command.
   d. If the peer is a non-direct EBGP peer, verify that the **peer ebgp-max-hop** command is configured.
   e. If the **peer ttl-security hops** command is configured, verify that the command is configured on the peer. Verify that the *hop-count* values configured on them are greater than the number of hops between them.
   f. Verify that a valid route to the peer is available.
   g. Use the **ping** command to verify the connectivity to the peer.
   h. Use the **display tcp verbose** or **display ipv6 tcp verbose** command to verify the TCP connection.
   i. Verify that no ACL rule is applied to disable TCP port 179.
2. If the problem persists, contact H3C Support.

Configuring large-scale BGP networks

Large-scale BGP network configuration tasks at a glance

To configure large-scale BGP networks, perform the following tasks:

- Configuring BGP route dampening
- Configuring BGP communities
- Configuring BGP route reflection
- Configuring a route server
- Configuring BGP confederation settings
  - Configuring a BGP confederation
  - (Optional.) Configuring confederation compatibility

Configuring BGP route dampening

About this task

Route dampening enables BGP to not select unstable routes as optimal routes.

Restrictions and guidelines

This feature applies to EBGP routes but not to IBGP routes.

If an EBGP peer goes down after you configure this feature, routes coming from the peer are dampened but not deleted.

After an active/standby switchover, BGP route dampening processes and information will be reset. The device needs to recalculate the dampening mechanism of BGP routes.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Configure BGP route dampening.

   dampening [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *

   By default, BGP route dampening is not configured.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Configure IPv6 BGP route dampening.

   dampening [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *

   By default, IPv6 BGP route dampening is not configured.

Configuring BGP communities

About this task

By default, a router does not advertise the COMMUNITY or extended community attribute to its peers or peer groups. When the router receives a route carrying the COMMUNITY or extended community attribute, it removes the attribute before advertising the route to other peers or peer groups.

Perform this task to enable a router to advertise the COMMUNITY or extended community attribute to its peers for route filtering and control. You can also use a routing policy to add or modify the COMMUNITY or extended community attribute for specific routes. For more information about routing policy, see "Configuring routing policies."

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Create a routing policy and a node, and enter routing policy node view.

   route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*

   For more information about this command, see routing policy commands in *Layer 3—IP Routing Command Reference*.

3. (Optional.) Configure **if-match** clauses as BGP route match criteria.
4. The routing policy can configure the COMMUNITY attribute of BGP routes only when these routes match a **permit** node.

   For more information about **if-match** clauses, see routing policy configuration in *Layer 3—IP Routing Configuration Guide.*
5. Configure the COMMUNITY attribute and the extended community attributes of BGP routes.
   - Set the COMMUNITY attribute for BGP routes.

     apply community { **none** | **additive** | { *community-number*&<1-32> | *aa:nn*&<1-32> | **internet** | **no-advertise** | **no-export** | **no-export-subconfed** } * [**additive**] }
   - Delete COMMUNITY attribute values from BGP routes.

     apply comm-list { *comm-list-number* | *comm-list-name* } **delete**

     By default, no COMMUNITY attribute value is deleted from BGP routes.
   - Set the link bandwidth extended community attribute for BGP routes.

     apply extcommunity bandwidth *link-bandwidth-value* [**additive**]
   - Set the RT extended community attribute for BGP routes.

     apply extcommunity { **rt** *route-target* }&<1-32> [**additive**]
   - Set the SoO extended community attribute for BGP routes.

     apply extcommunity soo *site-of-origin*&<1-32> [**additive**]

   For more information about these commands, see routing policy commands in *Layer 3—IP Routing Command Reference*.
6. Return to system view.

   quit
7. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast
8. Advertise the COMMUNITY attribute to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise-community**

   interface-peer *interface-type interface-number* **advertise-community**

   By default, the COMMUNITY attribute is not advertised.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

9. Advertise the extended community attributes to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise-ext-community**

   interface-peer *interface-type interface-number* **advertise-ext-community**

   By default, the extended community attributes is not advertised.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

10. Apply a routing policy to routes advertised to a peer or peer group. Choose one command as needed:

    peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-policy** *route-policy-name* **export**

    interface-peer *interface-type interface-number* **route-policy** *route-policy-name* **export**

    The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

    By default, no routing policy is applied.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Create a routing policy and a node, and enter routing policy node view.

   route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*

   For more information about this command, see routing policy commands in *Layer 3—IP Routing Command Reference*.

3. (Optional.) Configure **if-match** clauses as BGP route match criteria. The routing policy can configure the COMMUNITY attribute of BGP routes only when these routes match a **permit** node.

   For more information about **if-match** clauses, see routing policy configuration in *Layer 3—IP Routing Configuration Guide.*

4. Configure the COMMUNITY attribute and the extended community attributes of BGP routes.
   - Set the COMMUNITY attribute for BGP routes.

     apply community { **none** | **additive** | { *community-number*&<1-32> | *aa:nn*&<1-32> | **internet** | **no-advertise** | **no-export** | **no-export-subconfed** } * [**additive**] }
   - Delete COMMUNITY attribute values from BGP routes.

     apply comm-list { *comm-list-number* | *comm-list-name* } **delete**

     By default, no COMMUNITY attribute value is deleted from BGP routes.
   - Set the link bandwidth extended community attribute for BGP routes.

     apply extcommunity bandwidth *link-bandwidth-value* [**additive**]
   - Set the RT extended community attribute for BGP routes.

     apply extcommunity { **rt** *route-target* }&<1-32> [**additive**]
   - Set the SoO extended community attribute for BGP routes.

     apply extcommunity soo *site-of-origin*&<1-32> [**additive**]

   For more information about these commands, see routing policy commands in *Layer 3—IP Routing Command Reference*.

5. Return to system view.

quit

6. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
7. Advertise the COMMUNITY attribute to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **advertise-community**

   interface-peer *interface-type interface-number* **advertise-community**

   By default, the COMMUNITY attribute is not advertised.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.
8. Advertise the extended community attributes to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **advertise-ext-community**

   interface-peer *interface-type interface-number* **advertise-ext-community**

   By default, the extended community attributes is not advertised.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.
9. Apply a routing policy to routes advertised to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **route-policy** *route-policy-name* **export**

   interface-peer *interface-type interface-number* **route-policy** *route-policy-name* **export**

   By default, no routing policy is applied.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Configuring BGP route reflection

Configuring a BGP route reflector

About this task

Perform this task to configure a BGP route reflector and its clients. The route reflector and its clients automatically form a cluster identified by the router ID of the route reflector. The route reflector forwards route updates among its clients.

To improve availability, you can specify multiple route reflectors for a cluster. The route reflectors in the cluster must have the same cluster ID to avoid routing loops.

When a route reflector is connected to multiple clusters in the network, you can specify different cluster IDs for different peers and peer groups to control route reflection.

You only need to configure BGP route reflection on the device that acts as a route reflector. Other devices do not need to know the role of the local device in route reflection.

After you configure a device as a route reflector, it advertises routes as follows:

- Advertises routes received from a non-client IBGP peer to all clients.
- Advertises routes received from an IBGP peer that acts as a client to all peers.
- Advertises routes received from an EBGP peer to all peers.

Restrictions and guidelines

By default, the route reflector does not change the attributes of routes before route reflection to avoid routing loops. When you replan network traffic, you can perform this task to allow the route reflector to change the attributes of the routes to be reflected.

After using the **reflect change-path-attribute** command, you can use the **peer route-policy export** command to apply a routing policy that can change the attributes of the routes to be reflected. The following **apply** commands can define how the routing policy modifies route attributes:

- Use **apply origin** to change the ORIGIN attribute of routes.
- Use **apply ip-address next-hop** or **apply ipv6 next-hop** to change the NEXT_HOP attribute of routes.
- Use **apply cost** to change the MED attribute of routes.
- Use **apply local-preference** to change the local preference for routes.
- Use **apply community** to configure the COMMUNITY attribute of routes. For BGP to advertise the COMMUNITY attribute to a peer or peer group, use the **peer advertise-community** command.
- Use **apply comm-list delete** to delete COMMUNITY attribute values from routes.
- Use **apply extcommunity** to configure the link bandwidth extended community attribute or the RT extended community attribute of routes. For BGP to advertise the extended community attribute to a peer or peer group, use the **peer advertise-ext-community** command.
- Use **apply extcommunity soo** to configure the SoO extended community attribute of routes.

For more information about the **apply** commands, see routing policy commands in *Layer 3—IP Routing Command Reference*.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Configure the router as a route reflector and specify a peer or peer group as its client. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **reflect-client**

   interface-peer *interface-type interface-number* **reflect-client**

   By default, no route reflector or client is configured.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

4. (Optional.) Enable route reflection between clients.

   reflect between-clients

   By default, route reflection between clients is enabled.

5. (Optional.) Configure the cluster ID of the route reflector.

   reflector cluster-id { *cluster-id* | *ipv4-address* }

   By default, a route reflector uses its own router ID as the cluster ID.

6. (Optional.) Enable the route reflector to change the attributes of routes to be reflected.

   reflect change-path-attribute

   By default, the route reflector cannot change the attributes of routes to be reflected.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view or BGP IPv6 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

bgp *as-number* [**instance** *instance-name*]

address-family ipv6 multicast

3. Configure the router as a route reflector and specify a peer or peer group as its client. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **reflect-client**

   interface-peer *interface-type interface-number* **reflect-client**

   By default, no route reflector or client is configured.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

4. (Optional.) Enable route reflection between clients.

   reflect between-clients

   By default, route reflection between clients is enabled.

5. (Optional.) Configure the cluster ID of the route reflector.

   reflector cluster-id { *cluster-id* | *ipv4-address* }

   By default, a route reflector uses its own router ID as the cluster ID.

6. (Optional.) Enable the route reflector to change the attributes of routes to be reflected.

   reflect change-path-attribute

   By default, the route reflector cannot change the attributes of routes to be reflected.

Ignoring the ORIGINATOR_ID attribute

About this task

By default, BGP drops incoming route updates whose ORIGINATOR_ID attribute is the same as the local router ID. Some special networks such as firewall networks require BGP to accept such route updates. To meet the requirement, you must configure BGP to ignore the ORIGINATOR_ID attribute.

Restrictions and guidelines

Make sure this command does not result in a routing loop.

After you execute this command, BGP also ignores the CLUSTER_LIST attribute.

Procedure (IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Ignore the ORIGINATOR_ID attribute.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **ignore-originatorid**

   By default, BGP does not ignore the ORIGINATOR_ID attribute.

Procedure (IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Ignore the ORIGINATOR_ID attribute. Choose one command as needed:
   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **ignore-originatorid**
   interface-peer *interface-type interface-number* **ignore-originatorid**
   By default, BGP does not ignore the ORIGINATOR_ID attribute.

Configuring a route server

About this task

Some networks require the border devices in multiple ASs to establish full mesh connections, which consumes many network resources and reduces device performance. To resolve this issue, configure a device as a route server and the border devices in different ASs as its clients as needed.

Once a device is configured as a route server, it advertises routes to the clients without changing their AS path, next hop, and MED attributes. The clients can exchange routes without establishing EBGP connections and forward traffic to each other without the participation of the route server.

Restrictions and guidelines

The route server does not add its AS number in the EBGP routes advertised to a peer. To avoid peer establishment failures, execute the **peer ignore-first-as** command on the clients.

If you do not specify the **external** keyword, the route server advertises IBGP routes to its clients without changing the routes' AS path, next hop, and MED attributes. As a result, these routes might be unusable for EBGP peers. To resolve this issue, specify the **external** keyword when you use the **peer route-server-client** command. The route server can change the AS path, next hop, and MED attributes of IBGP routes before advertising them to the clients.

Procedure (IPv4 unicast)

1. Enter system view.
   system-view
2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
     address-family ipv4 [**unicast**]

3. Configure the device as a route server and specify a peer or peer group as a client.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **route-server-client** [**external**]

   By default, neither the route server nor the client is configured.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

3. Configure the device as a route server and specify a peer or peer group as a client.

   peer { *group-name* | *ipv6-address* [*prefix-length*] } **route-server-client** [**external**]

   By default, neither the route server nor the client is configured.

Configuring BGP confederation settings

About BGP confederation

BGP confederation provides another way to reduce IBGP connections in an AS.

A confederation contains sub-ASs. In each sub-AS, IBGP peers are fully meshed. Sub-ASs establish EBGP connections in between.

Configuring a BGP confederation

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Configure a confederation ID.

   confederation id *as-number*

   By default, no confederation ID is configured.

   From an outsider's perspective, the sub-ASs of the confederation is a single AS, which is identified by the confederation ID.

4. Specify confederation peer sub-ASs in the confederation.

   confederation peer-as *as-number-list*

   By default, no confederation peer sub-ASs are specified.

A confederation can contain a maximum of 32 sub-ASs. The AS number of a sub-AS is effective only in the confederation.

If the router needs to establish EBGP connections to other sub-ASs, you must specify the peering sub-ASs in the confederation.

Configuring confederation compatibility

About this task

If any routers in the confederation do not comply with RFC 3065, enable confederation compatibility to allow the router to work with those routers.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enable confederation compatibility.
   confederation nonstandard
   By default, confederation compatibility is disabled.

Display and maintenance commands for large-scale BGP network configuration

Execute **display** commands in any view.

Displaying BGP (IPv4 unicast address family)

Task	Command
Display BGP IPv4 unicast route dampening parameter information.	**display bgp** [**instance** *instance-name*] **dampening parameter ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*]
Display BGP IPv4 unicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**group-name** *group-name*]
Display information about a peer or peer group in BGP IPv4 unicast address family.	**display bgp** [**instance** *instance-name*] **peer ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv4-address mask-length* | { *ipv4-address* | **group-name** *group-name* } **log-info** | [*ipv4-address*] **verbose**] **display bgp** [**instance** *instance-name*] **peer ipv4** [**unicast**] [[*ipv6-address prefix-length* | **interface** *interface-type interface-number*] | [*ipv6-address* [**interface** *interface-type interface-number*] **log-info** | [*ipv6-address* [**interface** *interface-type interface-number*] **verbose**]

Task	Command
Display dampened BGP IPv4 unicast route information.	**display bgp** [**instance** *instance-name*] **routing-table dampened ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*]
Display BGP IPv4 unicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv4-address* [{ *mask-length* | *mask* } [**longest-match**]] | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]

Displaying BGP (IPv6 unicast address family)

Task	Command
Display BGP IPv6 unicast route dampening parameter information.	**display bgp** [**instance** *instance-name*] **dampening parameter ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*]
Display BGP IPv6 unicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [**group-name** *group-name*]
Display information about a peer or peer group in BGP IPv6 unicast address family.	**display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv6-address prefix-length* | { *ipv6-address* | **group-name** *group-name* } **log-info** | [*ipv6-address*] **verbose**] **display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [[*link-local-address*] **interface** *interface-type interface-number* | *link-local-address* **interface** *interface-type interface-number* { **log-info** | **verbose** }] **display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [*ipv4-address mask-length* | *ipv4-address* **log-info** | [*ipv4-address*] **verbose**]
Display dampened BGP IPv6 unicast route information.	**display bgp** [**instance** *instance-name*] **routing-table dampened ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*]
Display BGP IPv6 unicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] [*ipv6-address prefix-length* | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]

Displaying BGP (IPv4 multicast address family)

Task	Command
Display BGP IPv4 multicast route dampening parameter information.	**display bgp** [**instance** *instance-name*] **dampening parameter ipv4 multicast**
Display BGP IPv4 multicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv4 multicast** [**group-name** *group-name*]
Display BGP IPv4 multicast peer or peer group information.	**display bgp** [**instance** *instance-name*] **peer ipv4 multicast** [*ipv4-address mask-length* | { *ipv4-address* | **group-name** *group-name* } **log-info** | [*ipv4-address*] **verbose**]
Display dampened BGP IPv4 multicast route information.	**display bgp** [**instance** *instance-name*] **routing-table dampened ipv4 multicast**
Display BGP IPv4 multicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv4 multicast** [*ipv4-address* [{ *mask-length* | *mask* } [**longest-match**]] | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]

Displaying BGP (IPv6 multicast address family)

Task	Command
Display BGP IPv6 multicast route dampening parameter information.	**display bgp** [**instance** *instance-name*] **dampening parameter ipv6 multicast**
Display BGP IPv6 multicast peer group information.	**display bgp** [**instance** *instance-name*] **group ipv6 multicast** [**group-name** *group-name*]
Display BGP IPv6 multicast peer or peer group information.	**display bgp** [**instance** *instance-name*] **peer ipv6 multicast** [*ipv6-address prefix-length* | { *ipv6-address* | **group-name** *group-name* } **log-info** | [*ipv6-address*] **verbose**]
Display dampened BGP IPv6 multicast route information.	**display bgp** [**instance** *instance-name*] **routing-table dampened ipv6 multicast**
Display BGP IPv6 multicast route flapping statistics.	**display bgp** [**instance** *instance-name*] **routing-table flap-info ipv6 multicast** [*ipv6-address prefix-length* | **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* }]

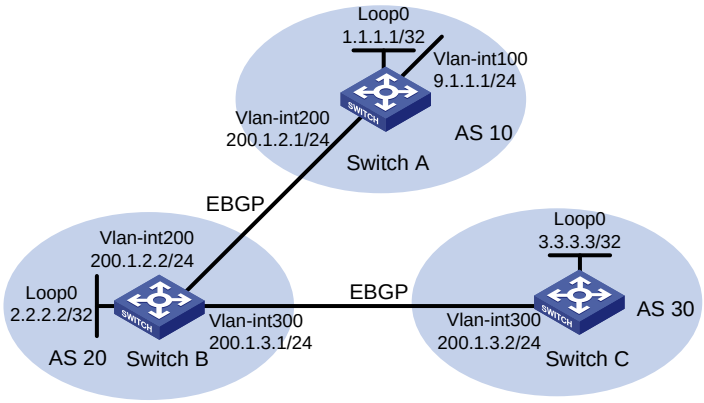
Large-scale BGP network configuration examples

Example: Configuring BGP communities

Network configuration

As shown in Figure 33, Switch B establishes EBGP connections with Switch A and Switch C. Configure NO_EXPORT community attribute on Switch A to make routes from AS 10 not advertised by AS 20 to any other AS.

Figure 33 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure EBGP connections:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 10
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 200.1.2.2 as-number 20
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 200.1.2.2 enable
[SwitchA-bgp-default-ipv4] network 9.1.1.0 255.255.255.0
[SwitchA-bgp-default] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 20
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 200.1.2.1 as-number 10
[SwitchB-bgp-default] peer 200.1.3.2 as-number 30
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 200.1.2.1 enable
[SwitchB-bgp-default-ipv4] peer 200.1.3.2 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 30
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 200.1.3.1 as-number 20
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 200.1.3.1 enable
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Display the BGP routing table on Switch B.

```
[SwitchB] display bgp routing-table ipv4 9.1.1.0

 BGP local router ID: 2.2.2.2
 Local AS number: 20

 Paths:   1 available, 1 best

 BGP routing table information of 9.1.1.0/24:
 From            : 200.1.2.1 (1.1.1.1)
 Rely nexthop    : 200.1.2.1
 Original nexthop: 200.1.2.1
 Route age       : 03h30m59s
 OutLabel        : NULL
 RxPathID        : 0x0
 TxPathID        : 0x0
 AS-path         : 10
 Origin          : igp
 Attribute value : pref-val 0
 State           : valid, external, best
 IP precedence   : N/A
 QoS local ID    : N/A
 Traffic index   : N/A
```

Display advertisement information of network 9.1.1.0 on Switch B.

```
[SwitchB] display bgp routing-table ipv4 9.1.1.0 advertise-info

 BGP local router ID: 2.2.2.2
 Local AS number: 20

 Paths:   1 best

 BGP routing table information of 9.1.1.0/24:
 Advertised to peers (1 in total):
    200.1.3.2
```

The output shows that Switch B can advertise the route with the destination 9.1.1.0/24 to other ASs through BGP.

Display the BGP routing table on Switch C.

```
[SwitchC] display bgp routing-table ipv4
```

```
Total number of routes: 1

BGP local router ID is 3.3.3.3
Status codes: * - valid, > - best, d - dampened, h - history,
              s - suppressed, S - stale, i - internal, e - external
              a – additional-path
      Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 9.1.1.0/24         200.1.3.1                              0       20 10i
```

The output shows that Switch C has learned route 9.1.1.0/24 from Switch B.

3. Configure a BGP community:

Configure a routing policy.

```
[SwitchA] route-policy comm_policy permit node 0
[SwitchA-route-policy-comm_policy-0] apply community no-export
[SwitchA-route-policy-comm_policy-0] quit
```

Apply the routing policy.

```
[SwitchA] bgp 10
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 200.1.2.2 route-policy comm_policy export
[SwitchA-bgp-default-ipv4] peer 200.1.2.2 advertise-community
```

Verifying the configuration

Display the routing table on Switch B.

```
[SwitchB] display bgp routing-table ipv4 9.1.1.0

 BGP local router ID: 2.2.2.2
 Local AS number: 20

 Paths:   1 available, 1 best

 BGP routing table information of 9.1.1.0/24:
 From            : 200.1.2.1 (1.1.1.1)
 Rely nexthop    : 200.1.2.1
 Original nexthop: 200.1.2.1
 Route age       : 03h30m59s
 OutLabel        : NULL
 RxPathID        : 0x0
 TxPathID        : 0x0
 Community       : No-Export
 AS-path         : 10
 Origin          : igp
 Attribute value : pref-val 0
 State           : valid, external, best
 IP precedence   : N/A
 QoS local ID    : N/A
 Traffic index   : N/A
```

Display advertisement information for the route 9.1.1.0 on Switch B.

```
[SwitchB] display bgp routing-table ipv4 9.1.1.0 advertise-info

 BGP local router ID: 2.2.2.2
 Local AS number: 20

 Paths:   1 best

 BGP routing table information of 9.1.1.0/24:
 Not advertised to any peers yet
```

Display the BGP routing table on Switch C.

```
[SwitchC] display bgp routing-table ipv4

 Total number of routes: 0
```

The output shows that BGP has not learned any route.

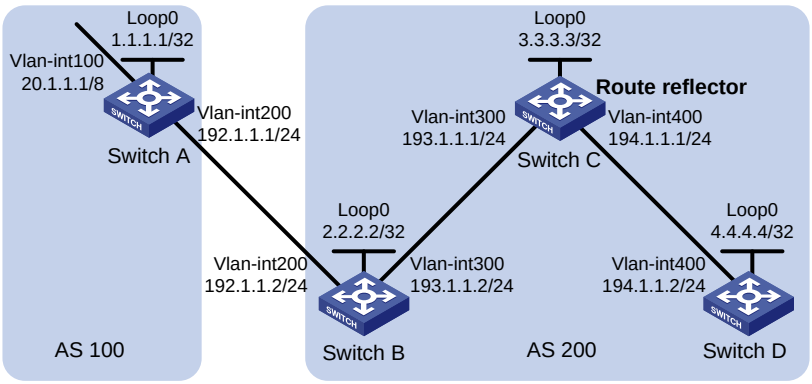
Example: Configuring BGP route reflector

Network configuration

As shown in Figure 34, all switches run BGP. Run EBGP between Switch A and Switch B, and run IBGP between Switch C and Switch B, and between Switch C and Switch D.

Configure Switch C as a route reflector with clients Switch B and Switch D to allow Switch D to learn route 20.0.0.0/8 from Switch C.

Figure 34 Network diagram



Procedure

1. Configure IP addresses for interfaces and configure OSPF in AS 200. (Details not shown.)
2. Configure BGP connections:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] bgp 100
   [SwitchA-bgp-default] router-id 1.1.1.1
   [SwitchA-bgp-default] peer 192.1.1.2 as-number 200
   [SwitchA-bgp-default] address-family ipv4 unicast
   [SwitchA-bgp-default-ipv4] peer 192.1.1.2 enable
   ```

   # Inject network 20.0.0.0/8 to the BGP routing table.

```
[SwitchA-bgp-default-ipv4] network 20.0.0.0
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 192.1.1.1 as-number 100
[SwitchB-bgp-default] peer 193.1.1.1 as-number 200
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 192.1.1.1 enable
[SwitchB-bgp-default-ipv4] peer 193.1.1.1 enable
[SwitchB-bgp-default-ipv4] peer 193.1.1.1 next-hop-local
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 193.1.1.2 as-number 200
[SwitchC-bgp-default] peer 194.1.1.2 as-number 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 193.1.1.2 enable
[SwitchC-bgp-default-ipv4] peer 194.1.1.2 enable
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] peer 194.1.1.1 as-number 200
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] peer 194.1.1.1 enable
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

3. Configure Switch C as the route reflector.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 193.1.1.2 reflect-client
[SwitchC-bgp-default-ipv4] peer 194.1.1.2 reflect-client
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Verifying the configuration

Display the BGP routing table on Switch B.

```
[SwitchB] display bgp routing-table ipv4

 Total number of routes: 1
```

```
 BGP local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 20.0.0.0/8         192.1.1.1       0                     0       100i
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table ipv4

 Total number of routes: 1

 BGP local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >i 20.0.0.0/8         193.1.1.2       0          100        0       100i
```

The output shows that Switch D has learned route 20.0.0.0/8 from Switch C.

Example: Configuring BGP confederation

Network configuration

As shown in Figure 35, split AS 200 into three sub-ASs (AS 65001, AS 65002, and AS 65003) to reduce IBGP connections. Switches in AS 65001 are fully meshed.

Figure 35 Network diagram

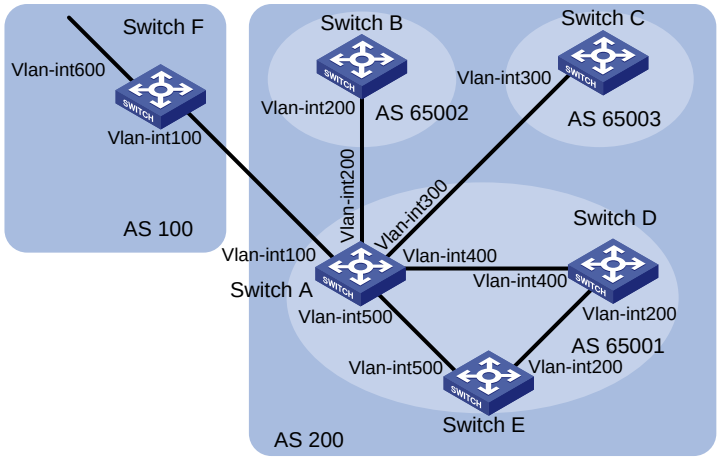


Table 11 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	200.1.1.1/24	Switch D	Vlan-int200	10.1.5.1/24
	Vlan-int200	10.1.1.1/24		Vlan-int400	10.1.3.2/24
	Vlan-int300	10.1.2.1/24	Switch E	Vlan-int200	10.1.5.2/24
	Vlan-int400	10.1.3.1/24		Vlan-int500	10.1.4.2/24
	Vlan-int500	10.1.4.1/24	Switch F	Vlan-int100	200.1.1.2/24
Switch B	Vlan-int200	10.1.1.2/24		Vlan-int600	9.1.1.1/24
Switch C	Vlan-int300	10.1.2.2/24			

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP confederation:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65001
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] confederation id 200
[SwitchA-bgp-default] confederation peer-as 65002 65003
[SwitchA-bgp-default] peer 10.1.1.2 as-number 65002
[SwitchA-bgp-default] peer 10.1.2.2 as-number 65003
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
[SwitchA-bgp-default-ipv4] peer 10.1.2.2 enable
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 next-hop-local
[SwitchA-bgp-default-ipv4] peer 10.1.2.2 next-hop-local
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65002
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] confederation id 200
[SwitchB-bgp-default] confederation peer-as 65001 65003
[SwitchB-bgp-default] peer 10.1.1.1 as-number 65001
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65003
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] confederation id 200
[SwitchC-bgp-default] confederation peer-as 65001 65002
```

```
[SwitchC-bgp-default] peer 10.1.2.1 as-number 65001
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 10.1.2.1 enable
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

3. Configure IBGP connections in AS 65001:

Configure Switch A.

```
[SwitchA] bgp 65001
[SwitchA-bgp-default] peer 10.1.3.2 as-number 65001
[SwitchA-bgp-default] peer 10.1.4.2 as-number 65001
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.3.2 enable
[SwitchA-bgp-default-ipv4] peer 10.1.4.2 enable
[SwitchA-bgp-default-ipv4] peer 10.1.3.2 next-hop-local
[SwitchA-bgp-default-ipv4] peer 10.1.4.2 next-hop-local
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 65001
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] confederation id 200
[SwitchD-bgp-default] peer 10.1.3.1 as-number 65001
[SwitchD-bgp-default] peer 10.1.5.2 as-number 65001
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] peer 10.1.3.1 enable
[SwitchD-bgp-default-ipv4] peer 10.1.5.2 enable
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] bgp 65001
[SwitchE-bgp-default] router-id 5.5.5.5
[SwitchE-bgp-default] confederation id 200
[SwitchE-bgp-default] peer 10.1.4.1 as-number 65001
[SwitchE-bgp-default] peer 10.1.5.1 as-number 65001
[SwitchE-bgp-default] address-family ipv4 unicast
[SwitchE-bgp-default-ipv4] peer 10.1.4.1 enable
[SwitchE-bgp-default-ipv4] peer 10.1.5.1 enable
[SwitchE-bgp-default-ipv4] quit
[SwitchE-bgp-default] quit
```

4. Configure the EBGP connection between AS 100 and AS 200:

Configure Switch A.

```
[SwitchA] bgp 65001
[SwitchA-bgp-default] peer 200.1.1.2 as-number 100
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 200.1.1.2 enable
[SwitchA-bgp-default-ipv4] quit
```

```
[SwitchA-bgp-default] quit
```

Configure Switch F.

```
<SwitchF> system-view
[SwitchF] bgp 100
[SwitchF-bgp-default] router-id 6.6.6.6
[SwitchF-bgp-default] peer 200.1.1.1 as-number 200
[SwitchF-bgp-default] address-family ipv4 unicast
[SwitchF-bgp-default-ipv4] peer 200.1.1.1 enable
[SwitchF-bgp-default-ipv4] network 9.1.1.0 255.255.255.0
[SwitchF-bgp-default-ipv4] quit
[SwitchF-bgp-default] quit
```

Verifying the configuration

Display the routing table on Switch B.

```
[SwitchB] display bgp routing-table ipv4

 Total number of routes: 1

 BGP local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >i 9.1.1.0/24         10.1.1.1        0          100        0       (65001)
                                                                      100i
[SwitchB] display bgp routing-table ipv4 9.1.1.0

 BGP local router ID: 2.2.2.2
 Local AS number: 65002

 Paths:   1 available, 1 best

 BGP routing table information of 9.1.1.0/24:
 From            : 10.1.1.1 (1.1.1.1)
 Rely nexthop    : 10.1.1.1
 Original nexthop: 10.1.1.1
 Route age       : 03h30m59s
 OutLabel        : NULL
 RxPathID        : 0x0
 TxPathID        : 0x0
 AS-path         : (65001) 100
 Origin          : igp
 Attribute value : MED 0, localpref 100, pref-val 0, pre 255
 State           : valid, external-confed, best
 IP precedence   : N/A
 QoS local ID    : N/A
```

```
 Traffic index   : N/A
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table ipv4

 Total number of routes: 1

 BGP local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >i 9.1.1.0/24         10.1.3.1        0          100        0       100i
[SwitchD] display bgp routing-table ipv4 9.1.1.0

 BGP local router ID: 4.4.4.4
 Local AS number: 65001

 Paths:   1 available, 1 best

 BGP routing table information of 9.1.1.0/24:
 From            : 10.1.3.1 (1.1.1.1)
 Rely nexthop    : 10.1.3.1
 Original nexthop: 10.1.3.1
 Route age       : 03h30m59s
 OutLabel        : NULL
 RxPathID        : 0x0
 TxPathID        : 0x0
 AS-path         : 100
 Origin          : igp
 Attribute value : MED 0, localpref 100, pref-val 0, pre 255
 State           : valid, internal-confed, best
 IP precedence   : N/A
 QoS local ID    : N/A
 Traffic index   : N/A
```

The output shows the following:

- Switch F can send route information to Switch B and Switch C through the confederation by establishing only an EBGP connection with Switch A.
- Switch B and Switch D are in the same confederation, but belong to different sub-ASs. They obtain external route information from Switch A, and generate identical BGP route entries although they have no direct connection in between.

Controlling BGP path selection

BGP path selection control tasks at a glance

By configuring BGP path attributes, you can control BGP path selection.

To control BGP path selection, perform the following tasks:

1. Configuring preferences for BGP routes
2. Configuring the NEXT_HOP attribute
   - Setting the local device IP as the next hop
   - Advertising only the global unicast address in the NEXT_HOP attribute
3. Setting a preferred value for received routes
4. Configuring the default local preference
5. Configuring the AS_PATH attribute
   - Permitting local AS number to appear in routes from a peer or peer group
   - Ignoring the AS_PATH attribute during optimal route selection
   - Advertising a fake AS number to a peer or peer group
   - Configuring AS number substitution
   - Removing private AS numbers from sent updates
   - Ignoring the first AS number of EBGP route updates
6. Configuring the MED attribute
   - Configuring the default MED value
   - Enabling MED comparison for routes from different ASs
   - Enabling MED comparison for routes on a per-AS basis
   -
   - Enabling MED comparison for routes from confederation peers
7. Configuring the link bandwidth attribute
8. Minimizing the priority of BGP routes advertised to peers
   - Minimizing the priority of BGP routes advertised to peers with a DOWN-to-UP state change
   - Minimizing the priority of BGP routes advertised to peers after a device reboot
   - (Optional.) Restoring the priority of advertised BGP routes
9. Configuring BGP route selection specifications

   To view the impact of this configuration step on BGP route selection rules, see "BGP route selection."
   - Preferring BGP routes of a specific next hop IP type
   - Ignoring IGP metrics during optimal route selection
   - Ignoring router IDs during optimal route selection
   - Preferring routes with an IPv6 next hop during optimal route selection

     Only the BGP EVPN address family supports this feature.
   - Configuring the next hop type selection rule for optimal route selection
10. Associating a track entry with a BGP peer or peer group

Configuring preferences for BGP routes

About this task

Routing protocols each have a default preference. If they find multiple routes destined for the same network, the route found by the routing protocol with the highest preference is selected as the optimal route. A smaller preference value indicates a higher preference.

You can use the **preference** command to modify preferences for EBGP, IBGP, and local BGP routes, or use a routing policy to set a preference for matching routes. For routes not matching the routing policy, the default preference applies.

If a device has an EBGP route and a local BGP route to reach the same destination, it does not select the EBGP route because the EBGP route has a lower preference than the local BGP route by default. You can use the **network short-cut** command to configure the EBGP route as a shortcut route that has the same preference as the local BGP route. The EBGP route will more likely become the optimal route.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Configure preferences for EBGP, IBGP, and local BGP routes.

   preference { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }

   By default, the preferences for EBGP, IBGP, and local BGP routes are 255, 255, and 130, respectively.

   You can use the **preference** command to configure preferences for BGP routes generated by the local device. In the current software version, this command takes effect only on BGP summary routes generated by using the **aggregate** command.

4. (Optional.) Configure an EBGP route as a shortcut route.

   network *ipv4-address* [*mask-length* | *mask*] **short-cut**

   By default, an EBGP route has a preference of 255.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Configure preferences for EBGP, IBGP, and local BGP routes.

   preference { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }

   By default, the preferences for EBGP, IBGP, and local BGP routes are 255, 255, and 130, respectively.

   You can use the **preference** command to configure preferences for BGP routes generated by the local device. In the current software version, this command takes effect only on BGP summary routes generated by using the **aggregate** command.
4. (Optional.) Configure an EBGP route as a shortcut route.

   network *ipv6-address prefix-length* **short-cut**

   By default, an EBGP route has a preference of 255.

Configuring the NEXT_HOP attribute

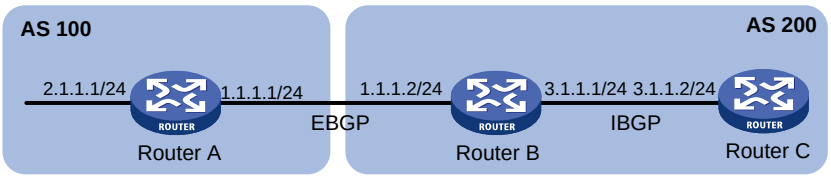
Setting the local device IP as the next hop

About this task

By default, a BGP router does not set itself as the next hop for routes advertised to an IBGP peer or peer group. In some cases, however, you must configure the advertising router as the next hop to ensure that the BGP peer can find the correct next hop.

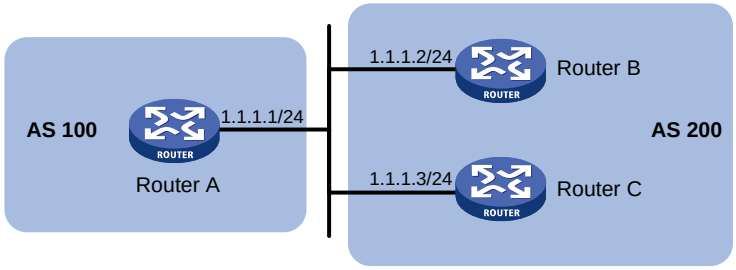
For example, as shown in Figure 36, Router A and Router B establish an EBGP neighbor relationship, and Router B and Router C establish an IBGP neighbor relationship. If Router C has no route destined for IP address 1.1.1.1/24, you must configure Router B to set itself 3.1.1.1/24 as the next hop for the network 2.1.1.1/24 advertised to Router C.

Figure 36 NEXT_HOP attribute configuration



If a BGP router has two peers on a broadcast network, it does not set itself as the next hop for routes sent to an EBGP peer by default. As shown in Figure 37, Router A and Router B establish an EBGP neighbor relationship, and Router B and Router C establish an IBGP neighbor relationship. They are on the same broadcast network 1.1.1.0/24. When Router B sends EBGP routes to Router A, it does not set itself as the next hop by default. However, you can configure Router B to set it (1.1.1.2/24) as the next hop for routes sent to Router A by using the **peer next-hop-local** command as needed.

Figure 37 NEXT_HOP attribute configuration



Restrictions and guidelines

If you have configured BGP load balancing, the router sets itself as the next hop for routes sent to an IBGP peer or peer group regardless of whether the **peer next-hop-local** command is configured.

A route reflector can change the NEXT_HOP attribute of reflected routes only if it is configured with both of the **peer next-hop-local** and **reflect change-path-attribute** commands.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Specify the local device IP as the next hop for routes sent to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **next-hop-local**

   interface-peer *interface-type interface-number* **next-hop-local**

   The default settings are as follows:

- When the device advertises a route to an EBGP peer, it processes the next hop of that route as follows:
  - If the EBGP peer is a directly connected peer and the next hop of the route is in the same network as the peer-facing interface of the local device, the device does not change the next hop of the route.
  - In other scenarios, the device sets the local device IP as the next hop of the route.
- The device does not change the next hops of routes sent to an IBGP peer or peer group. When the device sends ECMP BGP routes to IBGP peers, it sets the local device IP as the next hop of the optimal route among these routes.

The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Specify the local device IP as the next hop for routes sent to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **next-hop-local**

   interface-peer *interface-type interface-number* **next-hop-local**

   The default settings are as follows:
   - When the device advertises a route to an EBGP peer, it processes the next hop of that route as follows:
     - If the EBGP peer is a directly connected peer and the next hop of the route is in the same network as the peer-facing interface of the local device, the device does not change the next hop of the route.
     - In other scenarios, the device sets the local device IP as the next hop of the route.
   - The device does not change the next hops of routes sent to an IBGP peer or peer group. When the device sends ECMP BGP routes to IBGP peers, it sets the local device IP as the next hop of the optimal route among these routes.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Advertising only the global unicast address in the NEXT_HOP attribute

About this task

A directly-connected IPv6 EBGP peer might fail to learn routes if it cannot parse a route update that contains both the link-local address and the global unicast address. To resolve this issue, perform this task to enable the local device to advertise only the global unicast address in the NEXT_HOP attribute to its IPv6 peers.

Restrictions and guidelines

This feature might not apply to EBGP peers established through directly connected broadcast interfaces. If the next hop of the advertised route and the directly connected broadcast interfaces belong to the same subnet, this feature does not take effect.

Procedure (IPv4 unicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

3. Enable the device to advertise only the global unicast address in the NEXT_HOP attribute to its IPv6 peers.

   nexthop global-address-only

   By default, if the local device has a link-local address, it advertises both the link-local address and the global unicast address in the NEXT_HOP attribute to the directly-connected IPv6 EBGP peers established with its global unicast address.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv6 multicast

3. Enable the device to advertise only the global unicast address in the NEXT_HOP attribute to its IPv6 peers.

   nexthop global-address-only

   By default, if the local device has a link-local address, it advertises both the link-local address and the global unicast address in the NEXT_HOP attribute to the directly-connected IPv6 EBGP peers established with its global unicast address.

Setting a preferred value for received routes

About this task

Perform this task to set a preferred value for specific routes to control BGP path selection.

Among multiple routes that have the same destination/mask and are learned from different peers, the one with the greatest preferred value is selected as the optimal route.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Set a preferred value for routes received from a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **preferred-value** *value*

   interface-peer *interface-type interface-number* **preferred-value** *value*

   By default, the preferred value is 0 for routes received from a peer or peer group.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Set a preferred value for routes received from a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **preferred-value** *value*

   interface-peer *interface-type interface-number* **preferred-value** *value*

   By default, the preferred value is 0 for routes received from a peer or peer group.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Configuring the default local preference

About this task

The local preference is used to determine the optimal route for traffic leaving the local AS. When a BGP router obtains from several IBGP peers multiple routes to the same destination, but with different next hops, it selects the route with the highest local preference as the optimal route.

This task allows you to specify the default local preference for routes sent to IBGP peers.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 multicast

3. Configure the default local preference.

   default local-preference *value*

   The default local preference is 100.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Configure the default local preference.

   default local-preference *value*

   The default local preference is 100.

Configuring the AS_PATH attribute

Permitting local AS number to appear in routes from a peer or peer group

About this task

In general, BGP checks whether the AS_PATH attribute of a route from a peer contains the local AS number. If yes, it discards the route to avoid routing loops.

In certain network environments (for example, a Hub&Spoke network in MPLS L3VPN), however, the AS_PATH attribute of a route from a peer must be allowed to contain the local AS number. Otherwise, the route cannot be advertised correctly.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Permit the local AS number to appear in routes from a peer or peer group and set the appearance times. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **allow-as-loop** [*number*]

   interface-peer *interface-type interface-number* **allow-as-loop** [*number*]

   By default, the local AS number is not allowed in routes from a peer or peer group.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.

   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Permit the local AS number to appear in routes from a peer or peer group and set the appearance times. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **allow-as-loop** [*number*]

   interface-peer *interface-type interface-number* **allow-as-loop** [*number*]

   By default, the local AS number is not allowed in routes from a peer or peer group.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Ignoring the AS_PATH attribute during optimal route selection

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Configure BGP to ignore the AS_PATH attribute during optimal route selection.

   bestroute as-path-neglect

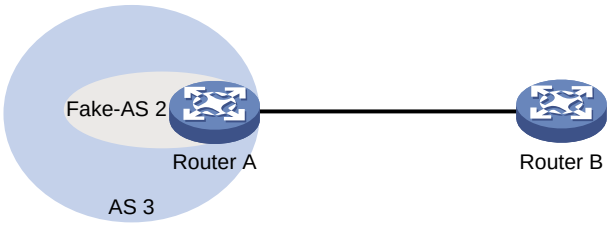
   By default, BGP includes the AS_PATH attribute in optimal route selection.

Advertising a fake AS number to a peer or peer group

About this task

After you move a BGP router from an AS to another AS (from AS 2 to AS 3 for example), you have to modify the AS number of the router on all its EBGP peers. To avoid such modifications, you can configure the router to advertise a fake AS number 2 to its EBGP peers so that the EBGP peers still think that Router A is in AS 2.

Figure 38 Network diagram



Restrictions and guidelines

The **peer fake-as** command is applicable only to EBGP peers and peer groups. When you set the AS number of the specified peer as the fake AS number in the **peer fake-as** command, follow these restrictions and guidelines:

- If the device successfully establishes a session to the specified peer by using the fake AS number, the device considers the peer as an IBGP peer.
- If the device successfully establishes a session to the specified peer by using the actual AS number, the device considers the peer as an EBGP peer.
- Both the features supported only by IBGP peers and the features supported only by EBGP peers are configurable for the peer. However, only features supported by the established peer session can take effect on the peer.
- BGP clears the configuration that takes effect on only IBGP peers for the peer if one of the following conditions exists:
  - The value for the *as-number* argument in the **peer fake-as** command is different from that in the **peer as-number** command.
  - The **undo peer fake-as** command is executed.

Procedure (IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Advertise a fake AS number to a peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **fake-as** *as-number*
   By default, no fake AS number is advertised to a peer or peer group.

Procedure (IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Advertise a fake AS number to a peer or peer group. Choose one command as needed:
   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **fake-as** *as-number*
   interface-peer *interface-type interface-number* **fake-as** *as-number*
   By default, no fake AS number is advertised to a peer or peer group.

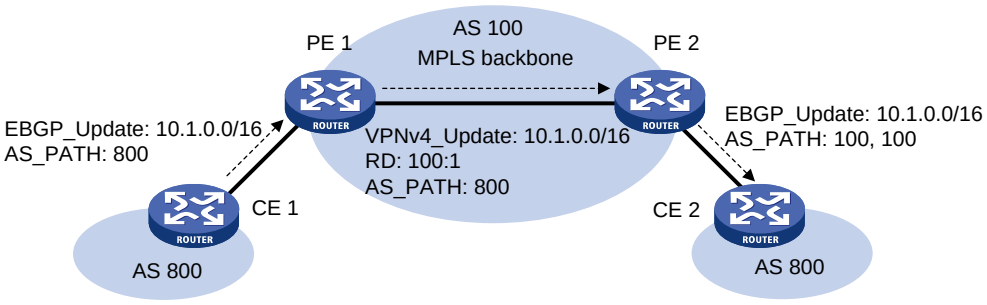
Configuring AS number substitution

About this task

To use EBGP between PE and CE in MPLS L3VPN, VPN sites in different geographical areas should have different AS numbers. Otherwise, BGP discards route updates containing the local AS number. If two CEs connected to different PEs use the same AS number, you must configure AS

number substitution on each PE. This substitution can replace the AS number in route updates originated by the remote CE as its own AS number before advertising them to the connected CE.

Figure 39 AS number substitution configuration (in an IPv4 network)



As shown in Figure 39, CE 1 and CE 2 use the same AS number 800. To ensure bidirectional communication between the two sites, configure AS number substitution on PE 2. PE 2 replaces AS 800 with AS 100 for the BGP route update originated from CE 1 before advertising it to CE 2. Perform the same configuration on PE 1.

Restrictions and guidelines

Do not configure AS number substitution in normal circumstances. Otherwise, routing loops might occur.

Procedure (IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Configure AS number substitution for a peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **substitute-as**
   By default, AS number substitution is not configured.

Procedure (IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Configure AS number substitution for a peer or peer group. Choose one command as needed:
   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **substitute-as**
   interface-peer *interface-type interface-number* **substitute-as**

By default, AS number substitution is not configured.

Removing private AS numbers from sent updates

About this task

Private AS numbers are typically used in test networks, and should not be transmitted in public networks. The range of private AS numbers is from 64512 to 65535.

Perform this task to enable BGP to remove or replace private AS numbers from the AS_PATH attribute of updates sent to a peer or peer group.

Restrictions and guidelines

This feature is applicable only to EBGP peers or peer groups.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Configure BGP to remove private AS numbers from the AS_PATH attribute of updates sent to an EBGP peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **public-as-only** [{ **force** | **limited** } [**replace**] [**include-peer-as**]]

   interface-peer *interface-type interface-number* **public-as-only** [{ **force** | **limited** } [**replace**] [**include-peer-as**]]

   By default, BGP updates sent to an EBGP peer or peer group can carry both public and private AS numbers.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.

- Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv6 [**unicast**]

- Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  ip vpn-instance *vpn-instance-name*

  address-family ipv6 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv6 multicast

3. Configure BGP to remove private AS numbers from the AS_PATH attribute of updates sent to an EBGP peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **public-as-only** [{ **force** | **limited** } [**replace**] [**include-peer-as**]]

   interface-peer *interface-type interface-number* **public-as-only** [{ **force** | **limited** } [**replace**] [**include-peer-as**]]

   By default, BGP updates sent to an EBGP peer or peer group can carry both public and private AS numbers.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

   This command applies only to EBGP peers and peer groups.

Ignoring the first AS number of EBGP route updates

About this task

By default, BGP checks the first AS number of an EBGP-learned route update. If the first AS number is neither the AS number of the BGP peer nor a private AS number, the BGP router disconnects the BGP session to the peer.

Ignoring the first AS number of EBGP route updates

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Configure BGP to ignore the first AS number of EBGP route updates.

   ignore-first-as

   By default, BGP checks the first AS number of EBGP-learned route updates.

Ignoring the first AS number of EBGP route updates received from a peer or peer group (IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enable BGP to ignore the first AS number of EBGP route updates received from the specified peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **ignore-first-as**
   By default, BGP checks the first AS number of route updates learned from EBGP peers or peer groups.

Ignoring the first AS number of EBGP route updates received from a peer or peer group (IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enable BGP to ignore the first AS number of EBGP route updates received from a peer or peer group. Choose one command as needed:
   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **ignore-first-as**
   interface-peer *interface-type interface-number* **ignore-first-as**
   By default, BGP checks the first AS number of route updates learned from EBGP peers or peer groups.

Configuring the MED attribute

About the MED attribute

BGP uses MED to determine the optimal route for traffic going into an AS. When a BGP router obtains multiple routes with the same destination but with different next hops, it selects the route with the smallest MED value as the optimal route if other conditions are the same.

Configuring the default MED value

Configuring the default MED value (IPv4 unicast/multicast address family)

1. Enter system view.
   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.

- Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 [**unicast**]
- Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  ip vpn-instance *vpn-instance-name*

  address-family ipv4 [**unicast**]
- Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv4 multicast

3. Configure the default MED value.

   default med *med-value*

   The default MED value is 0.

Configuring the default MED value (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Configure the default MED value.

   default med *med-value*

   The default MED value is 0.

Enabling MED comparison for routes from different ASs

About this task

By default, BGP only compares the MEDs of routes from the same AS. This task enables BGP to compare the MEDs of routes from different ASs.

Procedure

1. Enter system view.

system-view

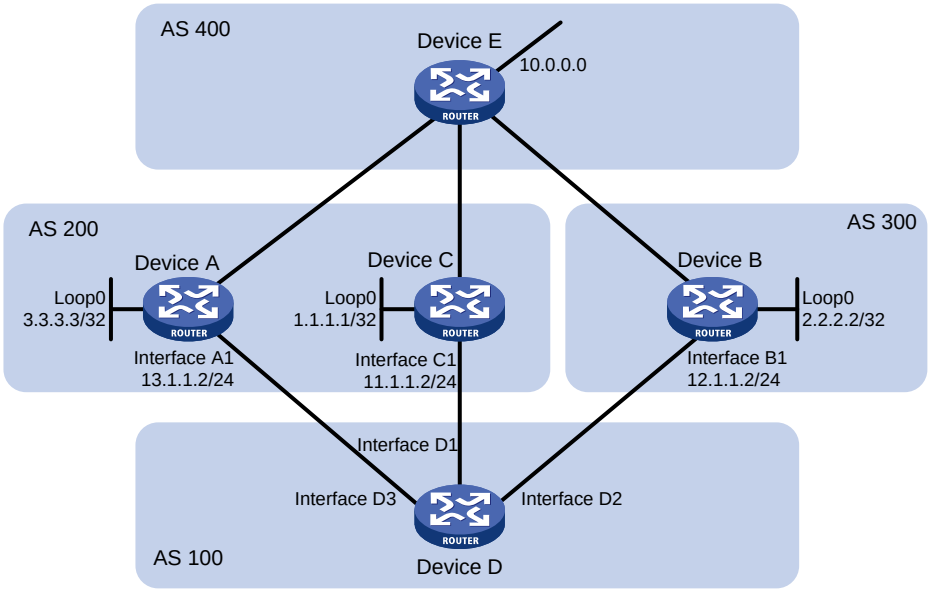
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enable MED comparison for routes from different ASs.
   compare-different-as-med
   By default, MED comparison for routes from different ASs is disabled.

Enabling MED comparison for routes on a per-AS basis

About this task

This task enables BGP to compare the MEDs of routes from an AS.

Figure 40 Route selection based on MED (in an IPv4 network)



As shown in Figure 40, Device D establishes indirect EBGP peer relationships with Device A, Device B, and Device C, and learns addresses 1.1.1.1/32, 2.2.2.2/32, and 3.3.3.3/32 through OSPF. The following output shows the routing information on Device D.

```
Destination/Mask    Proto   Pre Cost        NextHop         Interface
1.1.1.1/32          O_INTRA 10  10          11.1.1.2        Interface D1
2.2.2.2/32          O_INTRA 10  20          12.1.1.2        Interface D2
3.3.3.3/32          O_INTRA 10  30          13.1.1.2        Interface D3
```

Device D learns network 10.0.0.0 from both Device A and Device B. Because the route learned from Device B has a smaller IGP metric, the route is optimal. The following output shows the BGP routing table on Device D.

```
     Network           NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>e 10.0.0.0/8        2.2.2.2         50                     0       300 400e
 * e                   3.3.3.3         50                     0       200 400e
```

When Device D learns network 10.0.0.0 from Device C, it compares the route with the optimal route in its routing table. Because Device C and Device B reside in different ASs, BGP does not compare the MEDs of the two routes. The route from Device C has a smaller IGP metric than the route from Device B, so the route from Device C becomes optimal. The following output shows the BGP routing table on Device D.

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>e 10.0.0.0/8         1.1.1.1         60                     0       200 400e
 *  e 10.0.0.0/8         2.2.2.2         50                     0       300 400e
 *  e                    3.3.3.3         50                     0       200 400e
```

However, Device C and Device A reside in the same AS, and Device C has a greater MED, so network 10.0.0.0 learned from Device C should not be optimal.

To avoid this problem, you can configure the **bestroute compare-med** command to enable MED comparison for routes from the same AS on Device D. After that, Device D puts the routes received from each AS into a group, selects the route with the lowest MED from each group, and compares routes from different groups. Network 10.0.0.0 learned from Device B is the optimal route. The following output shows the BGP routing table on Device D.

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>e 10.0.0.0/8         2.2.2.2         50                     0       300 400e
 *  e                    3.3.3.3         50                     0       200 400e
 *  e                    1.1.1.1         60                     0       200 400e
```

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable MED comparison for routes on a per-AS basis.

   bestroute compare-med

   By default, MED comparison for routes on a per-AS basis is disabled.

Enabling MED comparison for routes from confederation peers

About this task

This task enables BGP to compare the MEDs of routes received from confederation peers. However, if a route received from a confederation peer has an AS number that does not belong to the confederation, BGP does not compare the route with other routes. For example, a confederation has three AS numbers 65006, 65007, and 65009. BGP receives three routes from different confederation peers. The AS_PATH attributes of these routes are 65006 65009, 65007 65009, and 65008 65009, and the MED values of them are 2, 3, and 1. Because the third route's AS_PATH attribute contains AS number 65008 that does not belong to the confederation, BGP does not compare it with other routes. As a result, the first route becomes the optimal route.

Procedure

1. Enter system view.

system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enable MED comparison for routes from confederation peers.
   bestroute med-confederation
   By default, MED comparison for routes from confederation peers is disabled.

Configuring the link bandwidth attribute

About this task

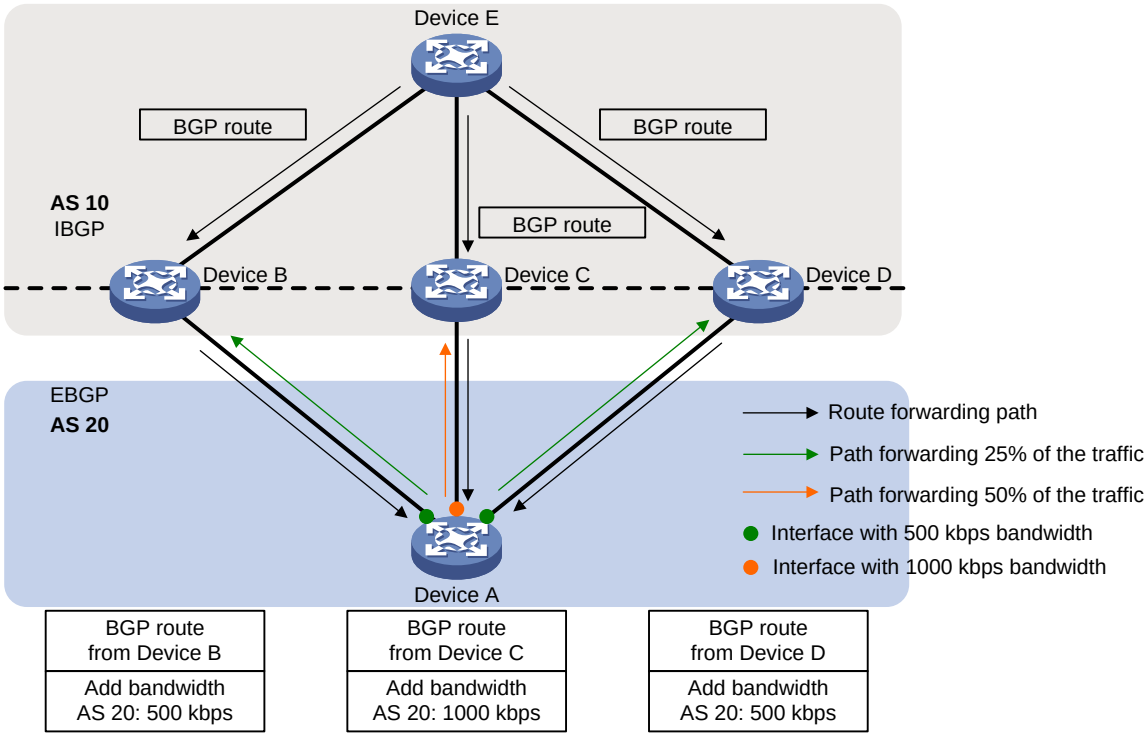
ECMP load balancing equally distributes traffic to the equal-cost links without considering link bandwidth differences. This might cause insufficient usage of high-speed links and traffic congestion on low-speed links. To resolve this issue, you can configure Unequal Cost Multiple Path (UCMP) load balancing based on the link bandwidth attributes of BGP routes.

The link bandwidth attribute of a route indicates the link bandwidth of the outgoing interface of the route. After UCMP load balancing is enabled for an EBGP peer on the device, the device adds the link bandwidth attribute into the route received from the EBGP peer. This attribute carries the AS number of the current BGP process and the interface bandwidth for direct EBGP session establishment. UCMP BGP routes are formed in the following situation:

- BGP route load balancing is configured by the **balance** command.
- All BGP routes used for load balancing carry a link bandwidth attribute.

During traffic forwarding, the proportion of the traffic carried on each BGP route is that of the route's link bandwidth to the total link bandwidth of all UCMP routes. As shown in Figure 41, Device A receives routes for the same destination address from three EBGP peers. Assume Device A adds link bandwidth attribute values 500 kbps, 1000 kbps, and 500 kbps for the three BGP routes. The traffic load proportions carried on the three BGP routes are 25%, 50%, and 25%, respectively.

Figure 41 UCMP load balancing based on BGP routes with link bandwidth attributes



By default, the BGP routes with link bandwidth attributes are advertised only to IBGP peers and the added link bandwidth attributes are optional non-transitive attributes. You can change the advertisement rule for BGP routes with link bandwidth attributes as needed:

- To advertise link bandwidth attributes to EBGP peers, execute the **peer advertise ebgp bandwidth** command.
- To make BGP peers that cannot identify link bandwidth attributes to carry link bandwidth attributes when they forward BGP routes, execute the **peer advertise bandwidth transitive** command. This command can convert the link bandwidth attributes carried in BGP routes into optional transitive attributes.

Restrictions and guidelines

When you configure this feature to add link bandwidth attributes to BGP routes, follow these restrictions and guidelines:

- This feature applies only to direct EBGP peers or peer groups. If the specified peer or peer group becomes an indirectly connected one, BGP stops adding the link bandwidth attribute to routes received from the EBGP peer or peer group.
- Only if all BGP routes for the same destination address carry the link bandwidth attributes, can these routes provide UCMP load balancing. Otherwise, the routes can only provide ECMP load balancing.
- A BGP route can carry a maximum of one link bandwidth attribute. If a received route already has a link bandwidth attribute, the device cannot add a link bandwidth attribute to the route any more.

In advertisement of BGP routes with link bandwidth attributes, the following restrictions apply:

- For a BGP peer or peer group, the configuration of the **peer route-policy export** command takes precedence over that of the **peer bandwidth** command when the following requirements are met:

- You use the **peer route-policy export** command to apply a routing policy to the BGP peer or peer group. The routing policy uses the **apply extcommunity** command to configure a link bandwidth attribute for BGP routes.
- You configure the **peer bandwidth** command for the BGP peer or peer group.

- Before advertisement, the device deletes the link bandwidth attribute of a received route if the device needs to change the next hop of that route. This is because the forwarding link instructed by this route has changed and thus the original link bandwidth no longer indicates the bandwidth of the real forwarding link. This restriction does not exist if the device receives a route without any link bandwidth attribute and adds a link bandwidth attribute to the route.
- For UCMP routes, if the Add-Path feature is configured to advertise only one optimal route, the link bandwidth attribute value for the optimal route equals the total link bandwidth of all UCMP routes. If the Add-Path feature is configured to advertise more than one optimal route, the advertised routes carry their respective link bandwidth attributes.

Prerequisites

Configure the **peer advertise-ext-community** command. BGP can advertise the link bandwidth attribute to a BGP peer or peer group only after you execute this command for the BGP peer or peer group.

Procedure (IPv4 unicast/IPv4 multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast
3. Configure a link bandwidth attribute for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **bandwidth** [*bandwidth-value*]

   By default, no link bandwidth attribute is configured for a peer or peer group.
4. (Optional.) Enable advertising the link bandwidth attribute to EBGP peers.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise ebgp bandwidth**

   By default, the device does not advertise the link bandwidth attribute to EBGP peers.
5. (Optional.) Convert the type of the link bandwidth attribute to optional transitive for BGP routes advertised to a peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise bandwidth transitive**

By default, the link bandwidth attribute of BGP routes is optional non-transitive.

Procedure (IPv6 unicast/IPv6 multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Configure a link bandwidth attribute for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **bandwidth** [*bandwidth-value*]

   By default, no link bandwidth attribute is configured for a peer or peer group.

   The *ipv4-address* [*mask-length*] option is supported only in BGP IPv6 unicast address family view.

4. (Optional.) Enable advertising the link bandwidth attribute to EBGP peers.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise ebgp bandwidth**

   By default, the device does not advertise the link bandwidth attribute to EBGP peers.

   The *ipv4-address* [*mask-length*] option is supported only in BGP IPv6 unicast address family view.

5. (Optional.) Convert the type of the link bandwidth attribute to optional transitive for BGP routes advertised to a peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise bandwidth transitive**

   By default, the link bandwidth attribute of BGP routes is optional non-transitive.

   The *ipv4-address* [*mask-length*] option is supported only in BGP IPv6 unicast address family view.

Minimizing the priority of BGP routes advertised to peers

Minimizing the priority of BGP routes advertised to peers with a DOWN-to-UP state change

About this task

By default, the device sends BGP routes immediately to a peer when the peer state changes to UP. If the device advertises routes before completing ARP or ND table convergence, the traffic matching these advertised routes might not be forwarded correctly, which causes traffic loss.

To resolve this issue, use this feature to minimize the priority of the routes to be advertised after the state of a peer changes from DOWN to UP. The peer will not select the routes received from the device as the optimal routes. After the device completes ARP or ND table convergence and the specified wait time elapses, the device restores the priority of the routes advertised to the peer.

After you use this feature, the device performs the following tasks when the state of a peer changes from DOWN to UP:

1. Set the local preference value to the minimum (zero) and the MED value to the maximum (4294967295) for the routes to be advertised.
2. After the specified wait time elapses, restore the local preference value and the MED value for the advertised routes, and then advertise these routes immediately.

   For the device to restore the priority of the advertised routes before the specified wait time elapses, use one of the following commands:

   - **`reset bgp advertise lowest-priority on-peer-up`**

     After you use this command, the device immediately sends BGP routes with the original local preference value and MED value to the peer. If the state of a peer changes from DOWN to UP later, the device still minimizes the priority of the routes advertised to that peer.
   - **`undo advertise lowest-priority on-peer-up duration`**

     After you use this command, the device immediately sends BGP routes with the original local preference value and MED value to the peer. The device does not change the original local preference value and MED value for the subsequent routes advertised to peers.

Restrictions and guidelines

You can use the **`advertise lowest-priority on-peer-up duration`** command in BGP instance view or BGP address family view:

- If you use this command in the view of a BGP instance, this command takes effect on all address families in the instance.
- If you use this command in the view of a BGP address family, this command takes effect only on the address family.
- For a BGP address family, the configuration of this command in BGP address family view takes precedence over that in BGP instance view.

You can use this command multiple times to change the wait time. After you change the wait time, the following rules apply:

- If the original wait time setting is taking effect, the new wait time setting will immediately take over and the device will update the time to wait for route priority restoration.

  For example, you can set the wait time to 10 seconds by using the **`advertise lowest-priority on-peer-up duration`** `10` command. Then, you use the **`advertise lowest-priority on-peer-up duration`** 6 command to set a new wait

time six seconds after the state of a peer changes from DOWN to UP. The device will still minimize the priority of the routes advertised to the peer during the following six seconds.

- If the original wait time is not taking effect, the new wait time will take effect after the state of a peer changes from DOWN to UP.

The following commands can influence each other:

- **advertise lowest-priority on-peer-up duration**
- **advertise lowest-priority on-startup duration**
- **bgp update-delay on-startup**
- **route-update-delay**

When you use two of the commands above together, the configuration result varies by command as follows:

Table 12 Configuration results

Command A	Command B	Configuration result
bgp update-delay on-startup	**advertise lowest-priority on-startup duration** in BGP instance view	Command A and command B overwrite each other. The command executed at last takes effect.
	advertise lowest-priority on-peer-up duration in BGP instance view	Provided that the delay time specified in command A is TA and the wait time specified in command B is TB, the two commands take effect as follows after the device restarts and the BGP process recovers: • If TA ≥ TB, only command A takes effect. • If TA < TB, the device performs the following tasks: **a.** The device does not send any BGP routes to peers within TA. **b.** After TA elapses, the device sends BGP routes to peers and minimizes the priority of these routes during the remaining time (TB minus TA). **c.** After TB elapses, the device restores the priority of the advertised routes, and then immediately sends them to peers. The device does not minimize the priority of the subsequent routes advertised to peers.
	route-update-delay	Both command A and command B take effect after the device restarts and the BGP process recovers. Provided that the delay time specified in command A is TA, the device can advertise BGP routes to a peer only after TA plus TS elapses. TS represents the larger interval between the intervals specified in the

		peer route-update-interval command and command B. If you did not execute the **peer route-update-interval** command for the peer, the device uses the default value of this command to compare with the interval configured in command B.
	advertise lowest-priority on-startup duration in the view of a BGP address family	• In the address family configured with command B, only command B takes effect. • In the address families not configured with command B, command A takes effect.
	advertise lowest-priority on-peer-up duration in the view of a BGP address family	Provided that the delay time specified in command A is TA and the wait time specified in command B is TB, the following rules apply: • In the address families not configured with command B, only command A takes effect. • In the address family configured with command B, the two commands take effect as follows after the device restarts and the BGP process recovers: ○ If TA ≥ TB, only command A takes effect. ○ If TA < TB, the device performs the following tasks: ○ i. The device does not send any BGP routes to peers within TA. ○ ii. The device sends BGP routes to peers and minimizes the priority of these routes during the remaining time (TB minus TA). ○ iii. After TB elapses, the device restores the priority of the advertised routes, and then immediately sends them to peers. The device does not minimize the priority of the subsequent routes advertised to peers.
route-update-delay	**advertise lowest-priority on-startup duration** in BGP instance view	The following description applies to BGP IPv4 unicast address family, BGP-VPN IPv4 unicast address family, BGP IPv6 unicast address family, and BGP-VPN IPv6 unicast address family. Provided that the wait time specified in command B is TB, the two commands take effect as follows after the device restarts and the BGP process recovers: • If TS ≥ TB, only command A takes effect.

		• If TS < TB, the device performs the following tasks: **d.** The device does not send any BGP routes to peers within TS. **e.** After TS elapses, the device sends BGP routes to peers and minimizes the priority of these routes during the remaining time (TB minus TS). **f.** After TB elapses, the device restores the priority of the advertised routes, and then immediately sends them to peers. The device does not minimize the priority of the subsequent routes advertised to peers. TS represents the larger interval between the intervals specified in the **peer route-update-interval** command and command A. If you did not execute the **peer route-update-interval** command for the peers, the device uses the default value of this command to compare with the interval configured in command A.
	advertise lowest-priority on-peer-up duration in BGP instance view	The following description applies to BGP IPv4 unicast address family, BGP-VPN IPv4 unicast address family, BGP IPv6 unicast address family, and BGP-VPN IPv6 unicast address family. Provided that the wait time specified in command B is TB, the two commands take effect as follows when the state of a peer changes from DOWN to UP: • If TS ≥ TB, only command A takes effect. • If TS < TB, the device performs the following tasks: **g.** The device does not send any BGP routes to the peer within TS. **h.** After TS elapses, the device sends BGP routes to the peer and minimizes the priority of these routes during the remaining time (TB minus TS). **i.** After TB elapses, the device restores the priority of the advertised routes, and then immediately sends them to the peer. The device does not minimize the priority of the subsequent routes advertised

		to the peer. TS represents the larger interval between the intervals specified in the **peer route-update-interval** command and command A. If you did not execute the **peer route-update-interval** command for the peer, the device uses the default value of this command to compare with the interval configured in command A.
	advertise lowest-priority on-startup duration in BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, BGP IPv6 unicast address family view, or BGP-VPN IPv6 unicast address family view	Provided that the wait time specified in command B is TB, the following rules apply: • In the address families not configured with command B, only command A takes effect. • In the address family configured with command B, the two commands take effect as follows after the device restarts and the BGP process recovers: o If TS ≥ TB, only command A takes effect. o If TS < TB, the device performs the following tasks: o i. The device does not send any BGP routes to peers within TS. o ii. After TS elapses, the device sends BGP routes to peers and minimizes the priority of these routes during the remaining time (TB minus TS). o iii. After TB elapses, the device restores the priority of the advertised routes, and then immediately sends them to peers. The device does not minimize the priority of the subsequent routes advertised to peers. • TS represents the larger interval between the intervals specified in the **peer route-update-interval** command and command A. If you did not execute the **peer route-update-interval** command for the peers, the device uses the default value of this command to compare with the interval configured in command A.
	advertise lowest-priority on-peer-up duration in BGP IPv4 unicast address	Provided that the wait time specified in command B is TB, the following rules apply: • In the address families not

	family view, BGP-VPN IPv4 unicast address family view, BGP IPv6 unicast address family view, or BGP-VPN IPv6 unicast address family view	configured with command B, only command A takes effect. • In the address family configured with command B, the two commands take effect as follows when the state of a peer changes from DOWN to UP: o If TS ≥ TB, only command A takes effect. o If TS < TB, the device performs the following tasks: o i. The device does not send any BGP routes to the peer within TS. o ii. After TS elapses, the device sends BGP routes to the peer and minimizes the priority of these routes during the remaining time (TB minus TS). o iii. After TB elapses, the device restores the priority of the advertised routes, and then immediately sends them to the peer. The device does not minimize the priority of the subsequent routes advertised to the peer. • TS represents the larger interval between the intervals specified in the **peer route-update-interval** command and command A. If you did not execute the **peer route-update-interval** command for the peer, the device uses the default value of this command to compare with the interval configured in command A.
advertise lowest-priority on-startup duration in BGP instance view	**advertise lowest-priority on-peer-up duration** in BGP instance view	Command A is mutually exclusive with command B.
	advertise lowest-priority on-peer-up duration in the view of a BGP address family	• In the address family configured with command B, only command B takes effect. • In the address families not configured with command B, only command A takes effect.
advertise lowest-priority on-peer-up duration in BGP instance view	**advertise lowest-priority on-startup duration** in the view of a BGP address family	• In the address family configured with command B, only command B takes effect. • In the address families not configured with command B, only command A takes effect.
advertise lowest-priority on-peer-up duration	**advertise lowest-priority on-startup duration** in	Command A is mutually exclusive with command B in the same address family.

in the view of a BGP address family	the view of a BGP address family	

Procedure (BGP instance view)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Minimize the priority of the routes advertised to a peer after the peer state changes from DOWN to UP.

   advertise lowest-priority on-peer-up duration *seconds*

   By default, BGP does not change the priority of the routes advertised to its peers.

Procedure (BGP IPv4 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

3. Minimize the priority of the routes advertised to a peer after the peer state changes from DOWN to UP.

   advertise lowest-priority on-peer-up duration *seconds*

   By default, BGP does not change the priority of the routes advertised to its peers.

4. (Optional.) Restore the route priority adjusted by the **advertise lowest-priority on-peer-up duration** command.

   reset bgp [**instance** *instance-name*] **advertise lowest-priority on-peer-up**

   This command is supported only in user view.

Procedure (BGP IPv6 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

address-family ipv6 [**unicast**]

3. Minimize the priority of the routes advertised to a peer after the peer state changes from DOWN to UP.

   advertise lowest-priority on-peer-up duration *seconds*

   By default, BGP does not change the priority of the routes advertised to its peers.

Minimizing the priority of BGP routes advertised to peers after a device reboot

About this task

By default, the device sends BGP routes immediately to peers after the device restarts and the BGP process recovers. If the device advertises routes before completing ARP or ND table convergence, the traffic matching these advertised routes might not be forwarded correctly, which causes traffic loss.

To resolve this issue, use this feature to minimize the priority of the routes to be advertised after the device restarts and the BGP process recovers. Peers will not select the routes received from the device as the optimal routes. After the device completes ARP or ND table convergence and the specified wait time elapses, the device restores the priority of the routes advertised to the peers.

After you use this feature, the device performs the following tasks after the device restarts and the BGP process recovers:

1. Set the local preference value to the minimum (zero) and the MED value to the maximum (4294967295) for the routes to be advertised.
2. After the specified wait time elapses, restore the local preference value and the MED value for the advertised routes, and then advertise these routes immediately.

   For the device to restore the priority of the advertised routes before the specified wait time elapses, use one of the following commands:

   - **reset bgp advertise lowest-priority on-startup**

     After you use this command, the device immediately sends BGP routes with the original local preference value and MED value to peers. The device does not minimize the priority of the subsequent routes advertised to the peers unless the device restarts again.
   - **undo advertise lowest-priority on-startup duration**

     After you use this command, the device immediately sends BGP routes with the original local preference value and MED value to peers. The device does not change the original local preference value and MED value for the subsequent routes advertised to the peers.

Restrictions and guidelines

When you use the **advertise lowest-priority on-startup duration** command, follows these guidelines:

- You can use this command in BGP instance view or BGP address family view:
  - If you use this command in the view of a BGP instance, this command takes effect on all address families in the instance.
  - If you use this command in the view of a BGP address family, this command takes effect only on the address family.
  - For a BGP address family, the configuration of this command in BGP address family view takes precedence over that in BGP instance view.
- The configuration of this command does not take effect immediately. It takes effect only after the device restarts.

The following commands can influence the configuration of each other:

- **advertise lowest-priority on-peer-up duration**
- **advertise lowest-priority on-startup duration**
- **bgp update-delay on-startup**
- **route-update-delay**

When you use two of the commands above together, the configuration result varies by command. For more information, see the related table in "Minimizing the priority of BGP routes advertised to peers with a DOWN-to-UP state change."

Procedure (BGP instance view)

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Minimize the priority of the routes advertised to peers after the device restarts and the BGP process recovers.
   advertise lowest-priority on-startup duration *seconds*
   By default, BGP does not change the priority of the routes advertised to its peers.

Procedure (BGP IPv4 unicast)

1. Enter system view.
   system-view
2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
     address-family ipv4 [**unicast**]
3. Minimize the priority of the routes advertised to peers after the device restarts and the BGP process recovers.
   advertise lowest-priority on-startup duration *seconds*
   By default, BGP does not change the priority of the routes advertised to its peers.

Procedure (BGP IPv6 unicast)

1. Enter system view.
   system-view
2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

address-family ipv6 [**unicast**]

3. Minimize the priority of the routes advertised to peers after the device restarts and the BGP process recovers.

   advertise lowest-priority on-startup duration *seconds*

   By default, BGP does not change the priority of the routes advertised to its peers.

Restoring the priority of advertised BGP routes

About this task

Before performing this task, make sure the following conditions exist:

- The device has completed ARP or ND table convergence and is ready to advertise routes with original priority.
- The **advertise lowest-priority on-peer-up duration** or **advertise lowest-priority on-startup duration** command still takes effect after a peer comes up or the device restarts.

After you perform this task, the device immediately sends BGP routes with the original local preference value and MED value to peers.

- If the state of a peer changes from DOWN to UP later, the device still minimizes the priority of the routes advertised to that peer.
- If the device restarts again later, the device still minimizes the priority of the routes advertised to peers.

Procedure

To restore the priority of the advertised BGP routes, use the **reset bgp** [**instance** *instance-name*] **advertise lowest-priority** { **on-peer-up** | **on-startup** } command in user view.

Preferring BGP routes of a specific next hop IP type

About this task

In scenarios where multiple networks coexist, the device might simultaneously learn BGP routes with the same destination network address but different types of next hops. Supported next hop types include IPv4 and IPv6. For example, the device might be able to reach the same destination address through either an SRv6 network or the public IPv4 network. In this case, it might learn a route with an IPv6 next hop from the SRv6 network and a route with an IPv4 next hop from the public IPv4 network. By default, the device selects the route learned from the public IPv4 network as the optimal route. To have the device select the route learned from the SRv6 network as the optimal route, you need to configure other route attributes to affect the route selection result. This task enables you to specify the next hop IP type that BGP prefers during optimal route selection, thus providing more flexibility and convenience in BGP route planning.

In scenarios where multiple networks coexist, the device might simultaneously learn BGP routes with the same destination network address but different types of next hops. Supported next hop types include IPv4 and IPv6. For example, the device might be able to reach the same destination address through either an MPLS network or the public IPv4 network. In this case, it might learn a route with an IPv6 next hop from the MPLS network and a route with an IPv4 next hop from the public IPv4 network. By default, the device selects the route learned from the public IPv4 network as the optimal route. To have the device select the route learned from the MPLS network as the optimal route, you need to configure other route attributes to affect the route selection result. This task enables you to

specify the next hop IP type that BGP prefers during optimal route selection, thus providing more flexibility and convenience in BGP route planning.

To view the priority of the nexthop IP type-based route selection rule among the BGP route selection rules, see "BGP route selection."

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Prefer BGP routes of a specific next hop IP type during optimal route selection.

   bestroute nexthop-priority { **ipv4** | **ipv6** } [**preferred**]

   By default, BGP prefers routes with IPv4 next hops during optimal route selection.

   If you execute this command multiple times in the same address family, the most recent configuration takes effect.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Prefer BGP routes of a specific next hop IP type during optimal route selection.

   bestroute nexthop-priority { **ipv4** | **ipv6** } [**preferred**]

   By default, BGP prefers routes with IPv4 next hops during optimal route selection.

   If you execute this command multiple times in the same address family, the most recent configuration takes effect.

Ignoring IGP metrics during optimal route selection

About this task

By default, BGP takes IGP metrics into account during optimal route selection. If multiple routes to the same destination are available, BGP selects the route with the smallest IGP metric as the optimal route.

Perform this task to enable BGP to ignore IGP metrics during optimal route selection.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Configure BGP to ignore IGP metrics during optimal route selection.

   bestroute igp-metric-ignore

   By default, BGP takes IGP metrics into account during optimal route selection.

Ignoring router IDs during optimal route selection

About this task

By default, BGP compares router IDs during optimal route selection. If multiple routes to the same destination are available, BGP selects the route with the smallest router ID as the optimal route.

Perform this task to enable BGP to ignore router IDs during optimal route selection.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Configure BGP to ignore router IDs during optimal route selection.

bestroute router-id-ignore

By default, BGP compares router IDs during optimal route selection.

Preferring routes with an IPv6 next hop during optimal route selection

About this task

Configure this feature for the VXLAN packets in an EVPN network to be forwarded through IPv6 routes when both IPv4 and IPv6 routes exist.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP EVPN address family view.
   address-family l2vpn evpn
4. Enable BGP to prefer routes with an IPv6 next hop during optimal route selection.
   bestroute ipv6-nexthop
   By default, BGP prefers routes with an IPv4 next hop during optimal route selection.

Configuring the next hop type selection rule for optimal route selection

About tunnel-type recursive next hops

NOTE:

In this section, a recursive next hop refers to a next hop obtained through route recursion. Recursive next hops include the following:

- **IP-type recursive next hops**: This type of recursive next hops are IP addresses.
- **Tunnel-type recursive next hops**: This type of recursive next hops are tunnels.

When the next hop of a BGP route is recursed to a tunnel, the tunnel can be one of the following:

- Label switched path (LSP) or MPLS TE tunnel.
- GRE or VXLAN tunnel.
- SR-MPLS TE policy or SRv6 TE policy tunnel.

Each tunnel-type recursive next hop has a unique tunnel ID. To view the tunnel-type recursive next hops of BGP routes, see the **Rely tunnel IDs** field in the output of the **display bgp routing-table** command.

Configuring the condition for BGP routes to participate in optimal route selection

About this task

The next hop of a BGP route can be recursed to an IP addresses and a tunnel synchronously. By default, a BGP route can participate in optimal route selection only when its next hop is recursed to an IP address. This command is applicable to the following scenarios:

- When the device acts as a route reflector, it only reflects routes to clients and does not provide traffic forwarding guidance. To ensure that the device can reflect all received routes to its clients, execute the **bestroute nexthop-resolved none** command. This command disables BGP from examining the reachability of next hops during optimal route selection.
- In some special networks, the next hop of a route might be a tunnel only after route recursion if its AS_PATH attribute is modified by an intermediate device. In this situation, the route cannot participate in optimal route selection though its next hop is reachable, which leads to traffic forwarding failure. To resolve this issue, use the **bestroute nexthop-resolved tunnel** command. This command allows BGP routes that have only tunnel-type recursive next hops to participate in optimal route selection.

Restrictions and guidelines

Although the **bestroute nexthop-resolved none** command enables BGP routes with unreachable next hops to participate in optimal route selection, those routes cannot be added into the IP routing table or FIB. As a result, they are not used to direct traffic forwarding.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Configure the condition for BGP routes to participate in optimal route selection

   bestroute nexthop-resolved { **none** | **tunnel** }

   By default, BGP routes can participate in optimal route selection only when their next hops are recursed to IP addresses.

   If you execute this command multiple times in the same address family, the most recent configuration takes effect.

Preferring BGP routes of a specific recursive next hop type during optimal route selection

About this task

By default, the recursive next hops of BGP routes does not affect their priorities during optimal route selection. If the next hop of a BGP route is recursed to an IP address and that of another BGP route is recursed to a tunnel, the two routes still have the same priority during optimal route selection.

This feature enables BGP to prefer routes with IP-type or tunnel-type recursive next hops during optimal route selection.

To view the priority of the route selection rule configured by this feature during optimal route selection, see "BGP route selection."

Restrictions and guidelines

A BGP route that does not have an IP-type recursive next hop cannot participate in optimal route selection. For BGP routes with only tunnel-type recursive next hops, they cannot participate in optimal route selection even if you have executed the **bestroute nexthop-type tunnel** command. To resolve this issue, execute the **bestroute nexthop-resolved tunnel** command.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Prefer routes with IP-type or tunnel-type recursive next hops during optimal route selection.
   bestroute nexthop-type { **ip** | **tunnel** } [**preferred**]

   By default, the recursive next hops of BGP routes does not affect their priorities during optimal route selection. If the next hop of a BGP route is recursed to an IP address and that of another BGP route is recursed to a tunnel, the two routes still have the same priority during optimal route selection.

   If you execute this command multiple times, the most recent configuration takes effect..

Associating a track entry with a BGP peer or peer group

About this task

This feature enables BGP to monitor the session to each peer, such as the link state of a peer interface. BGP can adjust the priority of routes received from a peer or peer group based on the track entry state. If a network state change occurs, BGP can response quickly to avoid traffic loss due to network issues, such as link failures.

After BGP associates a track entry with a peer, BGP adjusts the priority of routes received from the peer as follows:

- When the state of the associated track entry becomes negative, BGP increases the MED value of the routes to the maximum and set the route priority to the lowest.
- When the state of the associated track entry restores to positive, BGP restores the MED value and the priority of the routes to the original values.

Restrictions and guidelines

When you specify a track entry to be associated with a peer, the track entry must have been created. As a best practice, use a track entry associated with the link state of the peer interface. The track feature can obtain the link state of the peer interface easily and notify BGP to adjust the route priority when the link state changes to down.

For more information about track entries, see track configuration in *High Availability Configuration Guide*.

Procedure (IPv4 unicast)

1. Enter system view.
   system-view
2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
     address-family ipv4 [**unicast**]
3. Associate a track entry with a BGP peer or peer group for route priority adjustment.
   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-priority-track** *track-entry-number*
   By default, BGP does not use track entries for route priority adjustment.

Procedure (IPv6 unicast)

1. Enter system view.
   system-view
2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
     address-family ipv6 [**unicast**]
3. Associate a track entry with a BGP peer or peer group for route priority adjustment.
   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-priority-track** *track-entry-number*
   By default, BGP does not use track entries for route priority adjustment.

Display and maintenance commands for BGP path selection control

Execute **display** commands in any view.

Displaying BGP (IPv4 unicast address family)

Task	Command
Display BGP path attribute information.	**display bgp** [**instance** *instance-name*] **paths** [*as-regular-expression*]

Task	Command
Restore the priority of the routes to be advertised during the time to wait for route priority restoration.	**reset bgp** [**instance** *instance-name*] **advertise lowest-priority** { **on-peer-up** | **on-startup** }

Displaying BGP (IPv6 unicast address family)

Task	Command
Display BGP path attribute information.	**display bgp** [**instance** *instance-name*] **paths** [*as-regular-expression*]
Restore the priority of the routes to be advertised during the time to wait for route priority restoration.	**reset bgp** [**instance** *instance-name*] **advertise lowest-priority** { **on-peer-up** | **on-startup** }

Displaying BGP (IPv4 multicast address family)

Task	Command
Display BGP path attribute information.	**display bgp** [**instance** *instance-name*] **paths** [*as-regular-expression*]

Displaying BGP (IPv6 multicast address family)

Task	Command
Display BGP path attribute information.	**display bgp** [**instance** *instance-name*] **paths** [*as-regular-expression*]

BGP path selection control configuration examples

Example: Configuring route selection based on the LOCAL_PREF attribute

Network configuration

As shown in Figure 42, all switches run BGP.

- EBGP runs between Switch A and Switch B, and between Switch A and Switch C.
- IBGP runs between Switch B and Switch D, and between Switch D and Switch C. OSPF is the IGP protocol in AS 200.

Configure routing policies for LOCAL_PREF attribute adding, which makes Switch D select route 1.0.0.0/8 from Switch C as the optimal.

Figure 42 Network diagram

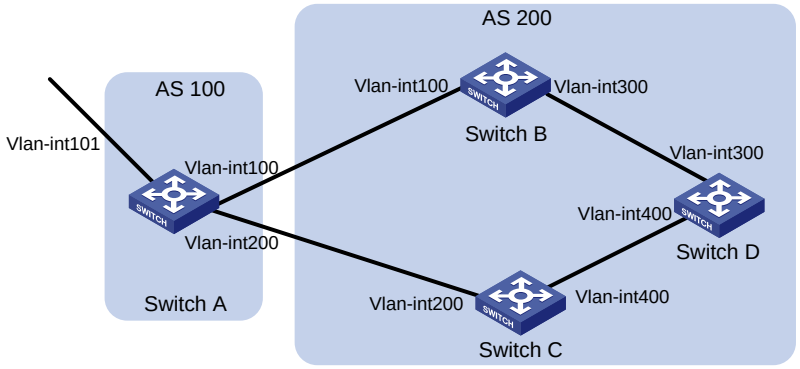


Table 13 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int101	1.0.0.1/8	Switch D	Vlan-int400	195.1.1.1/24
	Vlan-int100	192.1.1.1/24		Vlan-int300	194.1.1.1/24
	Vlan-int200	193.1.1.1/24	Switch C	Vlan-int400	195.1.1.2/24
Switch B	Vlan-int100	192.1.1.2/24		Vlan-int200	193.1.1.2/24
	Vlan-int300	194.1.1.2/24			

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure OSPF on Switch B, Switch C, and Switch D:

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 193.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

   # Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ospf
[SwitchD-ospf] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
```

```
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

3. Configure BGP connections:

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] peer 192.1.1.2 as-number 200
[SwitchA-bgp-default] peer 193.1.1.2 as-number 200
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 192.1.1.2 enable
[SwitchA-bgp-default-ipv4] peer 193.1.1.2 enable
```

Inject network 1.0.0.0/8 to the BGP routing table on Switch A.

```
[SwitchA-bgp-default-ipv4] network 1.0.0.0 8
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 192.1.1.1 as-number 100
[SwitchB-bgp-default] peer 194.1.1.1 as-number 200
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 192.1.1.1 enable
[SwitchB-bgp-default-ipv4] peer 194.1.1.1 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 193.1.1.1 as-number 100
[SwitchC-bgp-default] peer 195.1.1.1 as-number 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 193.1.1.1 enable
[SwitchC-bgp-default-ipv4] peer 195.1.1.1 enable
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Configure Switch D.

```
[SwitchD] bgp 200
[SwitchD-bgp-default] peer 194.1.1.2 as-number 200
[SwitchD-bgp-default] peer 195.1.1.2 as-number 200
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] peer 194.1.1.2 enable
[SwitchD-bgp-default-ipv4] peer 195.1.1.2 enable
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

4. Configure local preference for route 1.0.0.0/8, making Switch D give priority to the route learned from Switch C:

Define IPv4 basic ACL 2000 on Switch C to permit route 1.0.0.0/8.

```
[SwitchC] acl basic 2000
[SwitchC-acl-ipv4-basic-2000] rule permit source 1.0.0.0 0.255.255.255
[SwitchC-acl-ipv4-basic-2000] quit
```

Configure a routing policy named **localpref** on Switch C, setting the local preference of route 1.0.0.0/8 to 200 (the default is 100).

```
[SwitchC] route-policy localpref permit node 10
[SwitchC-route-policy-localpref-10] if-match ip address acl 2000
[SwitchC-route-policy-localpref-10] apply local-preference 200
[SwitchC-route-policy-localpref-10] quit
```

Apply routing policy **localpref** to routes from peer 193.1.1.1.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 193.1.1.1 route-policy localpref import
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table ipv4

 Total number of routes: 2

BGP local router ID is 195.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >i 1.0.0.0/8          193.1.1.1                  200        0       100i
*  i                    192.1.1.1                  100        0       100i
```

Route 1.0.0.0/8 learned from Switch C is the optimal.

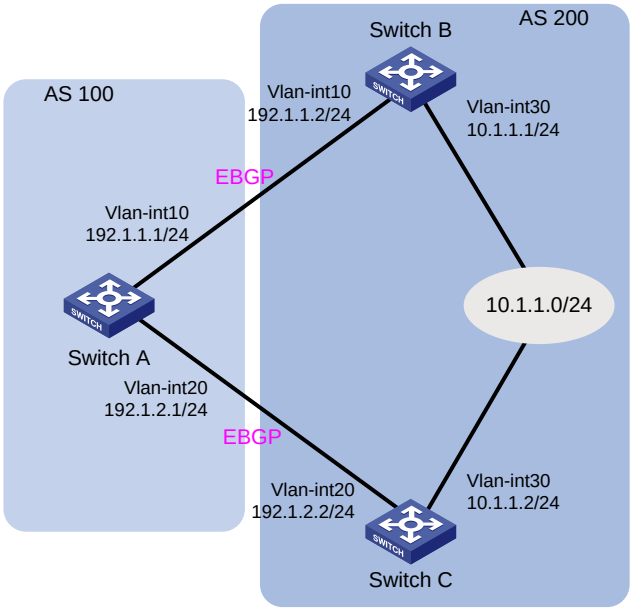
Example: Configuring route selection based on the MED attribute

Network configuration

As shown in Figure 43, all devices run BGP. EBGP runs between Switch A and Switch B, and between Switch A and Switch C. IBGP runs between Switch B and Switch C.

For Switch A, traffic destined for network segment 10.1.1.0/24 can reach AS 200 via Switch B or Switch. To achieve this goal, configure a routing policy that adds the MED attribute to BGP routes, so Switch A can select Switch C as the ingress of AS 200.

Figure 43 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP connections.

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] peer 192.1.1.2 as-number 200
[SwitchA-bgp-default] peer 192.1.2.2 as-number 200
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 192.1.1.2 enable
[SwitchA-bgp-default-ipv4] peer 193.1.2.2 enable
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 192.1.1.1 as-number 100
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 192.1.1.1 enable
[SwitchB-bgp-default-ipv4] network 10.1.1.0 24
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 192.1.2.1 as-number 100
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 192.1.2.1 enable
[SwitchC-bgp-default-ipv4] network 10.1.1.0 24
```

```
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

3. Configure the MED attribute, ensuring that Switch A selects the route from Switch C as the optimal route.

   # Configure IPv4 basic ACL 2000 on Switch A. This ACL only permits route 10.1.1.0/24.

```
[SwitchA] acl basic 2000
[SwitchA-acl-ipv4-basic-2000] rule permit source 10.1.1.0 0.0.0.255
[SwitchA-acl-ipv4-basic-2000] quit
```

   # Configure routing policy **med** on Switch A. This routing policy sets the local preference of route 10.1.1.0/24 to 100. The default MED value is 0.

```
[SwitchA] route-policy med permit node 10
[SwitchA-route-policy-med-10] if-match ip address acl 2000
[SwitchA-route-policy-med-10] apply cost 100
[SwitchA-route-policy-med-10] quit
```

   # On Switch A, apply routing policy **med** to BGP routes received from Switch B.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 192.1.1.2 route-policy med import
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Verifying the configuration

View the BGP routing table of Switch A.

```
<SwitchA> display bgp routing-table ipv4 unicast
 Total number of routes: 2
 BGP local router ID is 192.1.2.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
       Origin: i - IGP, e - EGP, ? - incomplete
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
* >e 10.1.1.0/24        192.1.2.2       0                     0       200i
*  e                    192.1.1.2       100                   0       200i
```

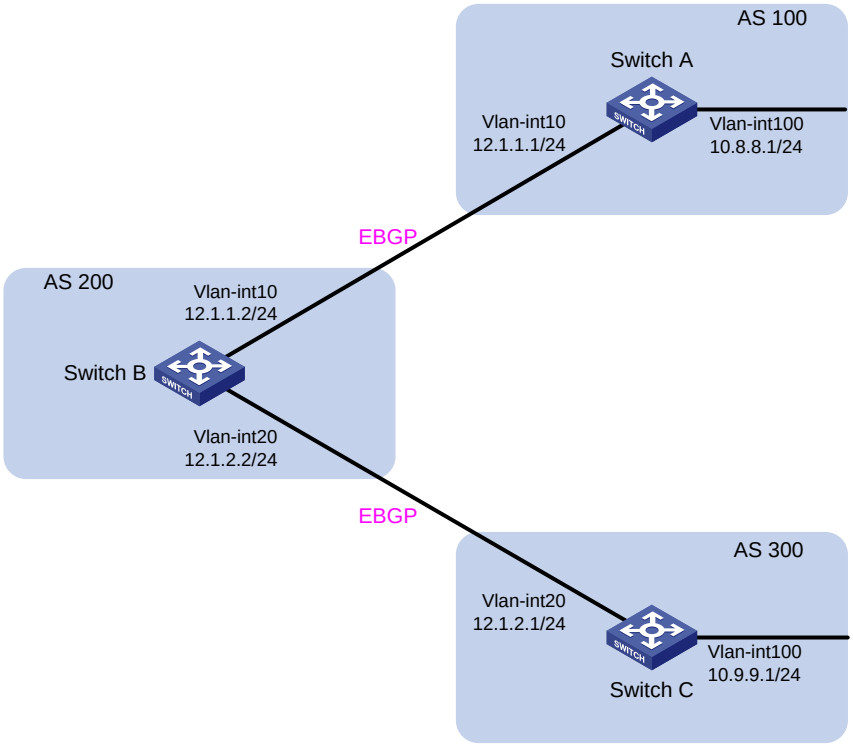
The command output shows that Switch A has selected route 10.1.1.0/24 learned from Switch C as the optimal route.

Example: Configuring route selection based on the AS_PATH attribute

Network configuration

As shown in Figure 44, Switch A, Switch B, and Switch C are located in different ASs. Switch B communicates with the other routers through EBGP. For security purposes, Switch A and Switch C should not learn the routes that the other advertises to Switch B. To meet this requirement, configure a BGP route filtering policy based on the AS_PATH attribute on Switch B. This policy prevents Switch B from advertising routes learned from Switch C to Switch A, and from advertising routes learned from Switch A to Switch C. With this route filtering policy, AS 100 and AS 300 are isolated from each other.

Figure 44 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP connections.

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] peer 12.1.1.2 as-number 200
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 12.1.1.2 enable
[SwitchA-bgp-default-ipv4] network 10.8.8.0 24
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 100
[SwitchB-bgp-default] peer 12.1.1.1 as-number 100
[SwitchB-bgp-default] peer 12.1.2.1 as-number 300
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 12.1.1.1 enable
[SwitchB-bgp-default-ipv4] peer 12.1.2.1 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 100
```

```
[SwitchC-bgp-default] peer 12.1.2.2 as-number 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 12.1.2.2 enable
[SwitchC-bgp-default-ipv4] network 10.9.9.0 24
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

3. Configure a BGP route distribution policy based on the AS_PATH attribute on Switch B.

   # Create AS path list **deny100** to deny BGP routes whose AS_PATH attribute contains AS 100.

```
[SwitchB] ip as-path deny100 deny _100_
[SwitchB] ip as-path deny100 permit .*
```

   # Create AS path list **deny300** to deny BGP routes whose AS_PATH attribute contains AS 300.

```
[SwitchB] ip as-path deny300 deny _300_
[SwitchB] ip as-path deny300 permit .*
```

   # Apply AS path lists **deny100** and **deny300** to routes advertised to Switch A and Switch C, respectively.

```
[SwitchB] bgp 100
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 12.1.1.1 as-path-acl deny 300 export
[SwitchB-bgp-default-ipv4] peer 12.1.2.1 as-path-acl deny 100 export
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Verifying the configuration

Execute the **`display bgp routing-table ipv4 unicast`** command on Switch A, Switch B, and Switch C separately. The command outputs show that:

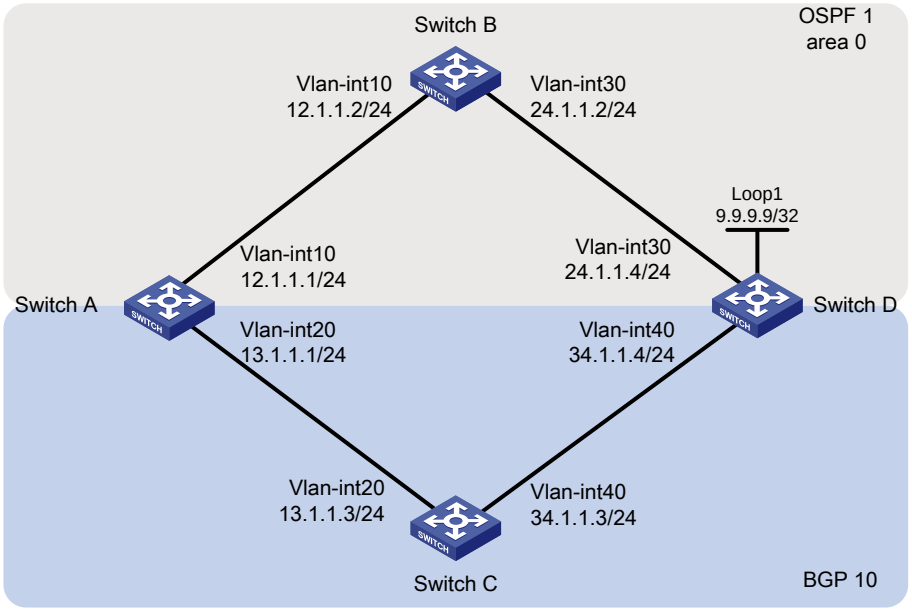
- Switch B has learned route 10.8.8.0/24 from Switch A and route 10.9.9.0/24 from Switch C.
- Switch A has not learned route 10.9.9.0/24 from Switch C, and Switch C has not learned route10.8.8.0/24 from Switch A. This indicates that Switch B does not forward Switch A's routes to Switch C or forward Switch C's routes to Switch A.

Example: Changing the preference value for a BGP route

Network configuration

As shown in Figure 45, Switches A, C, and D establish IBGP peer relationships, with Switch C acting as the route reflector. Switches A, B, and D establish OSPF neighbor relationships. Switch D advertises IPv4 unicast route 9.9.9.9/32 through OSPF and BGP. Upon receiving OSPF route 9.9.9.9/32 and BGP route 9.9.9.9/32, Switch A changes the preference value for BGP route 9.9.9.9/32, so the route can be selected as optimal route in Switch A's IP routing table.

Figure 45 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure Switch A.

   # Configure basic OSPF settings.

   ```
   [SwitchA] ospf 1
   [SwitchA-ospf-1] area 0.0.0.0
   [SwitchA-ospf-1] quit
   ```

   # Enable OSPF on VLAN-interface 10.

   ```
   [SwitchA] interface vlan-interface 10
   [SwitchA-Vlan-interface10] ospf 1 area 0.0.0.0
   [SwitchA-Vlan-interface10] quit
   ```

   # Configure BGP settings.

   ```
   [SwitchA] bgp 10
   [SwitchA-bgp-default] peer 13.1.1.3 as-number 10
   [SwitchA-bgp-default] address-family ipv4 unicast
   [SwitchA-bgp-default-ipv4] peer 13.1.1.3 enable
   [SwitchA-bgp-default-ipv4] quit
   [SwitchA-bgp-default] quit
   ```

3. Configure Switch B.

   # Configure basic OSPF settings.

   ```
   [SwitchB] ospf 1
   [SwitchB-ospf-1] area 0.0.0.0
   [SwitchB-ospf-1] quit
   ```

   # Enable OSPF on VLAN-interface 10 and VLAN-interface 30.

   ```
   [SwitchB] interface interface vlan-interface 10
   [SwitchB-Vlan-interface10] ospf 1 area 0.0.0.0
   [SwitchB-Vlan-interface10] quit
   [SwitchB] interface interface vlan-interface 30
   ```

```
[SwitchB-Vlan-interface30] ospf 1 area 0.0.0.0
[SwitchB-Vlan-interface30] quit
```

4. Configure Switch C:

Configure BGP settings.

```
[SwitchC] bgp 10
[SwitchC-bgp-default] peer 13.1.1.1 as-number 10
[SwitchC-bgp-default] peer 34.1.1.4 as-number 10
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 13.1.1.1 enable
[SwitchC-bgp-default-ipv4] peer 13.1.1.1 reflect-client
[SwitchC-bgp-default-ipv4] peer 13.1.1.1 next-hop-local
[SwitchC-bgp-default-ipv4] peer 34.1.1.4 enable
[SwitchC-bgp-default-ipv4] peer 34.1.1.4 reflect-client
[SwitchC-bgp-default-ipv4] peer 34.1.1.4 next-hop-local
[SwitchC-bgp-default-ipv4] reflect change-path-attribute
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

5. Configure Switch D:

Configure basic OSPF settings.

```
[SwitchD] ospf 1
[SwitchD-ospf-1] area 0.0.0.0
[SwitchD-ospf-1] quit
```

Enable OSPF on Loopback 1 and VLAN-interface 30.

```
[SwitchD] interface LoopBack1
[SwitchD-LoopBack1] ip address 9.9.9.9 255.255.255.255
[SwitchD-LoopBack1] ospf 1 area 0.0.0.0
[SwitchD-LoopBack1] quit
[SwitchD] interface interface vlan-interface 30
[SwitchD-Vlan-interface30] ospf 1 area 0.0.0.0
[SwitchD-Vlan-interface30] quit
```

Configure BGP settings.

```
[SwitchD] bgp 10
[SwitchD-bgp-default] peer 34.1.1.3 as-number 10
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] network 9.9.9.9 255.255.255.255
[SwitchD-bgp-default-ipv4] peer 34.1.1.3 enable
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

6. Enable Switch A to change the preference value for BGP route 9.9.9.9/32.

On Switch A, view detailed information about BGP route 9.9.9.9/32 in the IP routing table.

```
[SwitchA] displey ip routing-table 9.9.9.9 verbose

Summary count : 2

 Destination: 9.9.9.9/32
    Protocol: O_INTRA
  Process ID: 1
   SubProtID: 0x1                      Age: 06h34m02s
```

```
  FlushedAge: 00h01m35s
        Cost: 2                  Preference: 10
       IpPre: N/A                QosLocalID: N/A
         Tag: 0                       State: Active Adv
   OrigTblID: 0x2                   OrigVrf: default-vrf
     TableID: 0x2                    OrigAs: 0
       NibID: 0x13000002             LastAs: 0
      AttrID: 0xffffffffffffffff
    BkAttrID: 0xffffffffffffffff   Neighbor: 0.0.0.0
       Flags: 0x10041           OrigNextHop: 12.1.1.2
       Label: NULL              RealNextHop: 12.1.1.2
     BkLabel: NULL                BkNextHop: N/A
     SRLabel: NULL                Interface: Vlan-interface10
   BkSRLabel: NULL              BkInterface: N/A
   Tunnel ID: Invalid           IPInterface: Vlan-interface10
 BkTunnel ID: Invalid         BkIPInterface: N/A
     InLabel: NULL           ColorInterface: N/A
    SIDIndex: NULL         BkColorInterface: N/A
    FtnIndex: 0x0            TunnelInterface: N/A
TrafficIndex: N/A         BkTunnelInterface: N/A
   Connector: 0.0.0.0                PathID: 0x0
      UserID: 0xffffffff         SRTunnelID: Invalid
    SID Type: N/A                       NID: Invalid
    FlushNID: Invalid                BkNID: Invalid
  BkFlushNID: Invalid             StatFlags: 0x0
         SID: N/A
       BkSID: N/A
CommBlockLen: 0                    Priority: Medium
  MemberPort: N/A                  ExtFlags: 0x0
    UCMIndex: 0x0                  UserVLAN: 65535
      UCMMTU: 0                    BrasHash: 0
   SessionID: 0                 UserGroupID: 0
  IsoGroupID: 0               BkIsoGroupID: 0

 Destination: 9.9.9.9/32
    Protocol: BGP instance default
  Process ID: 0
   SubProtID: 0x1                      Age: 00h01m35s
  FlushedAge: 06h23m44s
        Cost: 0                  Preference: 255
       IpPre: N/A                QosLocalID: N/A
         Tag: 0                       State: Inactive Adv
   OrigTblID: 0x2                   OrigVrf: default-vrf
     TableID: 0x2                    OrigAs: 10
       NibID: 0x16000001             LastAs: 10
      AttrID: 0x0
    BkAttrID: 0xffffffffffffffff   Neighbor: 3.3.3.3
       Flags: 0x60              OrigNextHop: 4.4.4.4
```

```
       Label: NULL              RealNextHop: 13.1.1.3
     BkLabel: NULL                BkNextHop: N/A
     SRLabel: NULL                Interface: Vlan-interface20
   BkSRLabel: NULL              BkInterface: N/A
   Tunnel ID: Invalid           IPInterface: Vlan-interface20
 BkTunnel ID: Invalid         BkIPInterface: N/A
     InLabel: NULL            ColorInterface: N/A
    SIDIndex: NULL          BkColorInterface: N/A
    FtnIndex: 0x0             TunnelInterface: N/A
TrafficIndex: N/A          BkTunnelInterface: N/A
   Connector: N/A                     PathID: 0x0
      UserID: 0xffffffff          SRTunnelID: Invalid
    SID Type: N/A                        NID: Invalid
    FlushNID: Invalid                  BkNID: Invalid
  BkFlushNID: Invalid              StatFlags: 0x0
         SID: N/A
       BkSID: N/A
CommBlockLen: 0                     Priority: Low
  MemberPort: N/A                   ExtFlags: 0x0
    UCMIndex: 0x0                   UserVLAN: 65535
      UCMMTU: 0                     BrasHash: 0
   SessionID: 0                 UserGroupID: 0
  IsoGroupID: 256               BkIsoGroupID: 256
```

The IP routing table has two routes destined for 9.9.9.9/32. The preference value is 10 for OSPF route 9.9.9.9/32 and is 255 for BGP route 9.9.9.9/32. Consequently, only OSPF route 9.9.9.9/32 is active.

On Switch A, configure an ACL that match routes with prefix 9.9.9.9/32, and then apply the ACL to a routing policy. Then, configure the routing policy to change the preference value to 5 for matching routes in the IP routing table.

```
[SwitchA] acl basic 2077
[SwitchA-acl-ipv4-basic-2077] rule 5 permit source 9.9.9.9 0
[SwitchA-acl-ipv4-basic-2077] quit
[SwitchA] route-policy change-preference permit node 10
[SwitchA-route-policy-change-preference-10] if-match ip address acl 2077
[SwitchA-route-policy-change-preference-10] apply preference 5
[SwitchA-route-policy-change-preference-10] quit
```

On Switch A, apply the previous routing policy to BGP. Switch A can then set the preference value to 5 for BGP route 9.9.9.9/32.

```
[SwitchA] bgp 10
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] preference route-policy change-preference
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Verifying the configuration

On Switch A, view detailed information about BGP route 9.9.9.9/32 in the IP routing table.

```
[SwitchA] displey ip routing-table 9.9.9.9 verbose

Summary count : 2
```

```
 Destination: 9.9.9.9/32
    Protocol: BGP instance default
  Process ID: 0
   SubProtID: 0x1                      Age: 00h00m08s
  FlushedAge: 00h00m08s
        Cost: 0                  Preference: 5
       IpPre: N/A                QosLocalID: N/A
         Tag: 0                       State: Active Adv
   OrigTblID: 0x2                   OrigVrf: default-vrf
     TableID: 0x2                    OrigAs: 10
       NibID: 0x16000001             LastAs: 10
      AttrID: 0x0
    BkAttrID: 0xffffffffffffffff   Neighbor: 3.3.3.3
       Flags: 0x10060           OrigNextHop: 4.4.4.4
       Label: NULL              RealNextHop: 13.1.1.3
     BkLabel: NULL                BkNextHop: N/A
     SRLabel: NULL                Interface: Vlan-interface20
   BkSRLabel: NULL              BkInterface: N/A
   Tunnel ID: Invalid           IPInterface: Vlan-interface20
 BkTunnel ID: Invalid         BkIPInterface: N/A
     InLabel: NULL            ColorInterface: N/A
    SIDIndex: NULL          BkColorInterface: N/A
    FtnIndex: 0x0            TunnelInterface: N/A
TrafficIndex: N/A          BkTunnelInterface: N/A
   Connector: N/A                     PathID: 0x0
      UserID: 0xffffffff          SRTunnelID: Invalid
    SID Type: N/A                        NID: Invalid
    FlushNID: Invalid                  BkNID: Invalid
  BkFlushNID: Invalid              StatFlags: 0x0
         SID: N/A
       BkSID: N/A
CommBlockLen: 0                     Priority: Low
  MemberPort: N/A                   ExtFlags: 0x0
    UCMIndex: 0x0                   UserVLAN: 65535
      UCMMTU: 0                     BrasHash: 0
   SessionID: 0                  UserGroupID: 0
  IsoGroupID: 256               BkIsoGroupID: 256

 Destination: 9.9.9.9/32
    Protocol: O_INTRA
  Process ID: 1
   SubProtID: 0x1                      Age: 06h54m06s
  FlushedAge: 00h21m39s
        Cost: 2                  Preference: 10
       IpPre: N/A                QosLocalID: N/A
         Tag: 0                       State: Inactive Adv
   OrigTblID: 0x2                   OrigVrf: default-vrf
```

```
     TableID: 0x2                    OrigAs: 0
       NibID: 0x13000002             LastAs: 0
      AttrID: 0xffffffffffffffff
    BkAttrID: 0xffffffffffffffff   Neighbor: 0.0.0.0
       Flags: 0x41              OrigNextHop: 12.1.1.2
       Label: NULL              RealNextHop: 12.1.1.2
     BkLabel: NULL                BkNextHop: N/A
     SRLabel: NULL                Interface: Vlan-interface10
   BkSRLabel: NULL              BkInterface: N/A
   Tunnel ID: Invalid           IPInterface: Vlan-interface10
 BkTunnel ID: Invalid         BkIPInterface: N/A
     InLabel: NULL           ColorInterface: N/A
    SIDIndex: NULL         BkColorInterface: N/A
    FtnIndex: 0x0            TunnelInterface: N/A
TrafficIndex: N/A         BkTunnelInterface: N/A
   Connector: 0.0.0.0               PathID: 0x0
      UserID: 0xffffffff         SRTunnelID: Invalid
    SID Type: N/A                      NID: Invalid
    FlushNID: Invalid                BkNID: Invalid
  BkFlushNID: Invalid             StatFlags: 0x0
         SID: N/A
       BkSID: N/A
CommBlockLen: 0                   Priority: Medium
  MemberPort: N/A                 ExtFlags: 0x0
    UCMIndex: 0x0                 UserVLAN: 65535
      UCMMTU: 0                   BrasHash: 0
   SessionID: 0                UserGroupID: 0
  IsoGroupID: 0               BkIsoGroupID: 0
```

According to the command output, the preference value for BGP route 9.9.9.9/32 has been changed to 5. The BGP route is in active state, because its preference is higher than that of OSPF route 9.9.9.9/32.

Tuning and optimizing BGP networks

BGP network tuning and optimization tasks at a glance

To tune and optimize BGP networks, perform the following tasks:

- Setting the TCP MSS for TCP connections
- Establishing and resetting EBGP sessions
  - Enabling BGP to establish an EBGP session over multiple hops
  - Enabling immediate re-establishment of direct EBGP connections upon link failure
- Enabling BGP to negotiate special capabilities
  - Enabling prefix-based BGP ORF capabilities negotiation
  - Enabling 4-byte AS number suppression
  - Enabling BGP to exchange withdrawal ROUTE-REFRESH messages with peers
- Enabling Track-based priority adjustment for BGP routes advertised to a peer or peer group
- Establishing, terminating, and resetting BGP sessions
  - Disabling BGP session establishment
  - Configuring BGP soft-reset
- Configuring BGP load balancing
- Configuring the BGP Additional Paths feature
- Protecting an EBGP peer when memory usage reaches level 2 threshold
- Setting a DSCP value for outgoing BGP packets
- Disabling route recursion policy control for routes received from a peer or peer group
- Flushing the suboptimal BGP route to the RIB
- Configuring a peer monitoring group
- Disabling BGP from flushing specific routes to the routing table
- Configuring BGP optimal route selection delay
- Configuring the BGP session state to collaborate with the monitor link downlink interfaces
- Specifying the longest match principle for next hop recursion
- Enabling conversational remote host route learning
- Configuring label allocation and path selection
  - Specifying a label allocation mode
  - Disabling optimal route selection for labeled routes without tunnel information
- Enabling peer unreachability detection
- Displaying 4-byte AS numbers in asdot or asdot+ notation
- Converting AS numbers in effective device configurations
- Configuring the tag community attribute

Setting the TCP MSS for TCP connections

About this task

BGP typically establishes a multihop TCP connection with a peer. Multiple intermediate devices might fragment BGP packets due to their MTU settings, resulting in frequent packet encapsulations and decapsulations that reduce forwarding efficiency.

To resolve this issue, you can perform this task to set the TCP maximum segment size (MSS) for a peer or peer group. TCP segments sent by the source will not be re-fragmented along the path to the destination.

During TCP connection establishment with the peer or peer group specified in the **peer tcp-mss** command, the minimum of the following MSSs takes effect:

- The MSS calculated according to the IPv4 MTU set by the **ip mtu** command on the peer-facing interface.
- The MSS set by the **tcp mss** command for the peer-facing interface.
- The MSS calculated according to the path MTU detected by TCP path MTU discovery (enabled with the **tcp path-mtu-discovery** command).
- The MSS set by the **peer tcp-mss** command.

The MSS is calculated by using the following formula:

MSS = path MTU (or interface MTU) – IP header length – TCP header length

For more information about MTU and MSS, see IP performance optimization configuration in *Layer 3—IP Services Configuration Guide*.

Restrictions and guidelines

This task might cause BGP session re-establishment. Make sure you understand the potential impact before performing this task.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Set the TCP MSS for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **tcp-mss** *mss-value*

   By default, the TCP MSS is not set.

Enabling BGP to establish an EBGP session over multiple hops

About this task

To establish an EBGP session, two routers must have a direct physical link and use directly connected interfaces. If no direct link is available, you must use the **peer ebgp-max-hop** command to enable BGP to establish an EBGP session over multiple hops and specify the maximum hops.

Restrictions and guidelines

When the BGP GTSM feature is enabled, two peers can establish an EBGP session after passing GTSM check, regardless of whether the maximum number of hops is reached.

To ensure correct BGP route recursion, do not specify any direct EBGP peer or peer group when you configure the **peer ebgp-max-hop** command.

Procedure (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable BGP to establish an EBGP session to an indirectly connected peer or peer group and specify the maximum hop count.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **ebgp-max-hop** [*hop-count*]

   By default, BGP cannot establish an EBGP session to an indirectly connected peer or peer group.

Procedure (IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable BGP to establish an EBGP session to an indirectly connected peer or peer group and specify the maximum hop count.

   peer { *group-name* | *ipv6-address* [*prefix-length*] } **ebgp-max-hop** [*hop-count*]

   By default, BGP cannot establish an EBGP session to an indirectly connected peer or peer group.

Enabling immediate re-establishment of direct EBGP connections upon link failure

About this task

With this feature disabled, when the link to a directly connected EBGP peer goes down, the router does not re-establish a session to the peer until the hold time timer expires. This feature enables BGP to immediately discover the link failure and recreate the session in that situation. When this feature is disabled, route flapping does not affect EBGP session state.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable immediate re-establishment of direct EBGP connections upon link failure.

   ebgp-interface-sensitive

   By default, immediate re-establishment of direct EBGP connections is enabled.

Enabling prefix-based BGP ORF capabilities negotiation

About BGP ORF

Configure this feature in scenarios where the route sender is supposed to send only the desired BGP routes to the route receiver. Without BGP ORF configured, you must configure a route distribution policy on the route sender on a per receiver basis, which is complex and larborious. This feature helps you get rid of this situaiton.

BGP Outbound Route Filtering (ORF) reduces the bandwidth usage by reducing the route updates that are sent between BGP peers.

The BGP peers negotiate the ORF capabilities through OPEN messages. After completing the negotiation process, the BGP peers can exchange ORF information (local route reception filtering policy) through ROUTE-REFRESH messages. Then, only routes that pass both the local route distribution filtering policy and the received route reception filtering policy can be advertised.

ORF information includes the local address prefix list used for selective BGP route reception. After a peer receives the local device's ORF information, it will only send BGP routes that match the address prefix list to the local device. This approach reduces the number of UPDATE messages transmitted between BGP peers, saving network resources.

Restrictions and guidelines

You can enable the ORF information sending, receiving, or both sending and receiving capabilities on a BGP router. For two BGP peers to successfully negotiate the ORF capabilities, make sure one end has the sending capability and the other end has the receiving capability.

After you enable BGP ORF capabilities negotiation for a peer, the local device negotiates standard ORF capabilities as defined in RFC with the peer. If the peer uses nonstandard ORF, you must also enable nonstandard ORF capabilities negotiation for the peer.

Prerequisites

Before enabling BGP ORF capabilities negotiation for a peer, use the **peer prefix-list import** command to set a BGP route reception policy referencing an address prefix list. If you fail to do so, BGP does not send address prefix list information to that peer.

Enabling BGP ORF capabilities negotiation for a peer or peer group

Procedure (IPv4 unicast/IPv4 multicast)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

3. Reference an address prefix list to filter BGP routes received from a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **prefix-list** *ipv4-prefix-list-name* **import**

   By default, no address prefix list is referenced to filter BGP routes received from a peer or peer group.

4. Enable prefix-based BGP ORF capabilities negotiation for a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **capability-advertise orf prefix-list** { **both** | **receive** | **send** }

   interface-peer *interface-type interface-number* **capability-advertise orf prefix-list** { **both** | **receive** | **send** }

   By default, BGP ORF capabilities negotiation is disabled for a peer or peer group.

   The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

Procedure (IPv6 unicast/IPv6 multicast)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast
3. Reference an address prefix list to filter BGP routes received from a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **prefix-list** *ipv6-prefix-list-name* **import**

   By default, no address prefix list is referenced to filter BGP routes received from a peer or peer group.
4. Enable prefix-based BGP ORF capabilities negotiation for a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **capability-advertise orf prefix-list** { **both** | **receive** | **send** }

   interface-peer *interface-type interface-number* **capability-advertise orf prefix-list** { **both** | **receive** | **send** }

   By default, BGP ORF capabilities negotiation is disabled for a peer or peer group.

   The *ipv4-address* [*mask-length*] and *link-local-address* **interface** *interface-type interface-number* options and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Enabling nonstandard BGP ORF capabilities negotiation for a peer or peer group

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable nonstandard BGP ORF capabilities negotiation for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **capability-advertise orf non-standard**

By default, nonstandard BGP ORF capabilities negotiation is disabled for a peer or peer group.

Enabling 4-byte AS number suppression

About this task

BGP supports 4-byte AS numbers. The 4-byte AS number occupies four bytes, in the range of 1 to 4294967295. By default, a device sends an OPEN message to the peer device for session establishment. The OPEN message indicates that the device supports 4-byte AS numbers. If the peer device supports 2-byte AS numbers instead of 4-byte AS numbers, the session cannot be established. To resolve this issue, enable the 4-byte AS number suppression feature. The device then sends an OPEN message to inform the peer that it does not support 4-byte AS numbers, so the BGP session can be established.

Restrictions and guidelines

If the peer device supports 4-byte AS numbers and the peer AS number is greater than 65535, do not enable the 4-byte AS number suppression feature. Otherwise, the BGP session cannot be established.

Procedure (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable 4-byte AS number suppression.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **capability-advertise suppress-4-byte-as**

   By default, 4-byte AS number suppression is disabled.

Procedure (IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable 4-byte AS number suppression.

   peer { *group-name* | *ipv6-address* [*prefix-length*] } **capability-advertise suppress-4-byte-as**

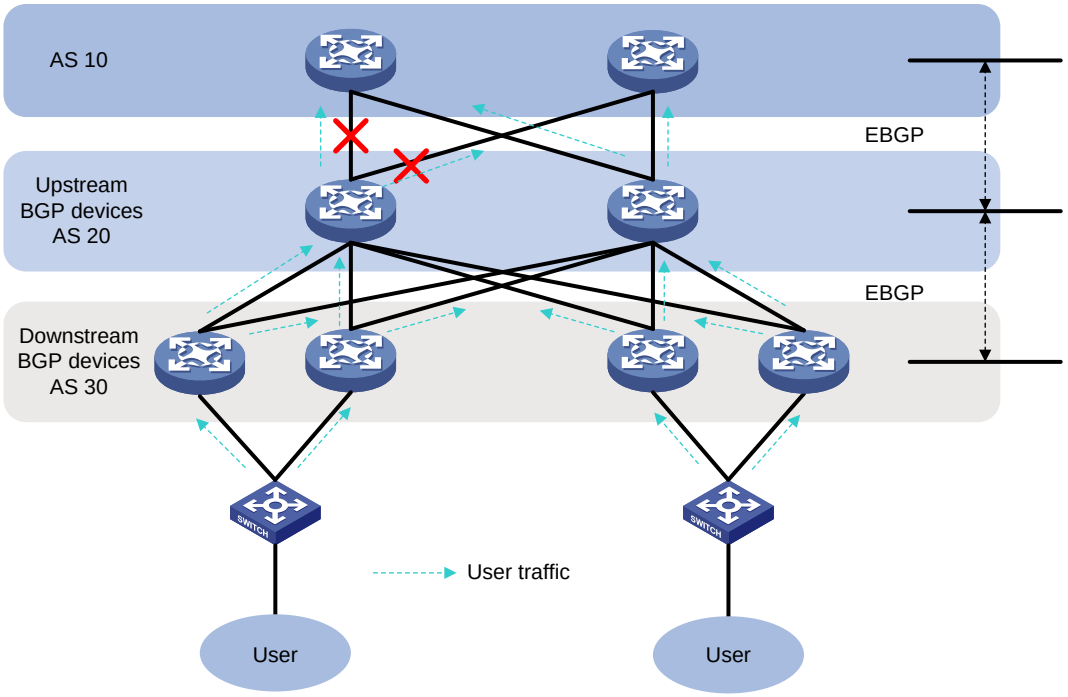
   By default, 4-byte AS number suppression is disabled.

Enabling BGP to exchange withdrawal ROUTE-REFRESH messages with peers

About this task

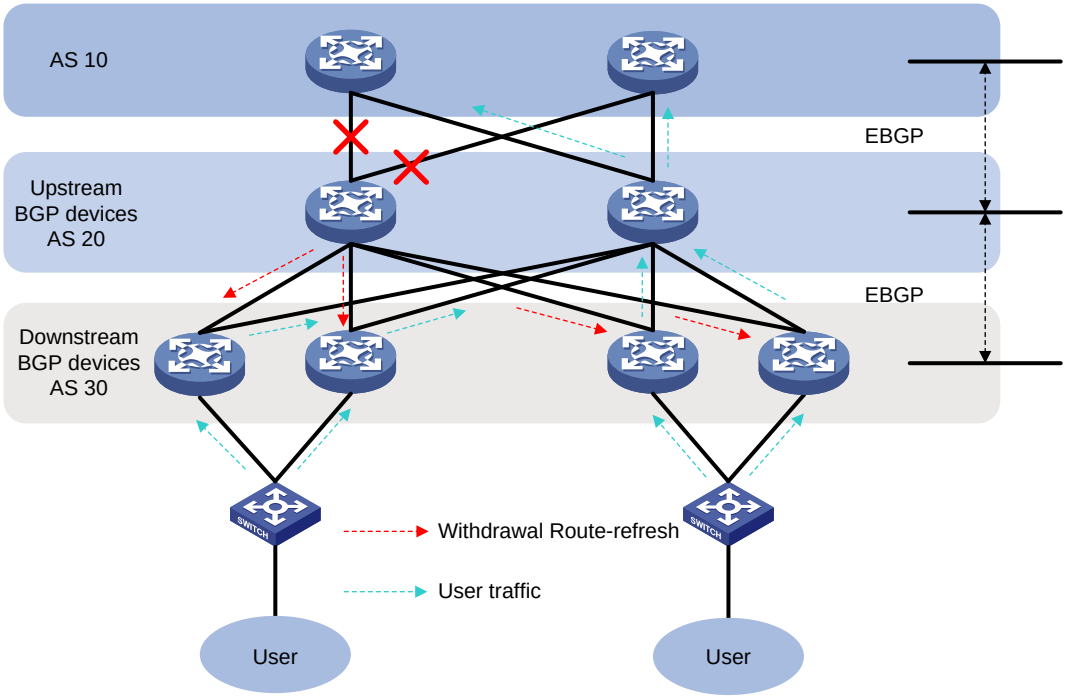
As shown in Figure 46, two upstream BGP devices are deployed in a multi-level BGP network. When all uplinks of an upstream device fail and the device cannot complete route convergence in time, the device might also fail to send route withdrawal messages to downstream devices in time. At this time, those downstream devices still forward some user traffic to the upstream device, resulting in traffic forwarding failure.

Figure 46 Multi-level BGP network



To avoid this issue, enable the upstream device to exchange withdrawal ROUTE-REFRESH messages with downstream devices. As shown in Figure 47, when all uplinks of the upstream device fail, it sends withdrawal ROUTE-REFRESH messages to the downstream devices. On receipt of the withdrawal ROUTE-REFRESH messages, the downstream devices immediately perform route switchover and forward traffic to another upstream device with normal uplinks.

Figure 47 Exchanging withdrawal ROUTE-REFRESH messages in the multi-level BGP network



After this feature is configured, the upstream device performs the following operations:

1. Terminates and re-establishes the BGP session to the specified downstream device.
2. Exchanges OPEN messages with the downstream device to negotiate the capability of exchanging withdrawal ROUTE-REFRESH messages.

   To ensure successful negotiation, you must configure this command at both ends of the BGP session, specifying the other end.

With the withdrawal ROUTE-REFRESH capability negotiated successfully, if the upstream device is disconnected from all peers in the same AS, it sends a withdrawal ROUTE-REFRESH message to the downstream device. The sent message carries the AS number from which the upstream device is disconnected. In this example, the disconnected AS is AS 10.

After receiving the withdrawal ROUTE-REFRESH message, the downstream device performs the following operations:

3. Searches in all BGP routes received from the upstream BGP device for routes with AS 10 as the first AS number in the NEXT_HOP attribute.

   To view AS number information in the NEXT_HOP attribute of a BGP route, see the **AS-path** field in the output of the **`display rib nib`** or **`display ipv6 rib nib`** command. For more information about the two commands, see IP routing basics commands in *Layer 3—IP Routing Command Reference*.
4. After finding routes with AS 10 as the first AS number in the NEXT_HOP attribute, marks the IP routes corresponding to those routes as backup routes.

   The downstream device no longer uses these routes for traffic forwarding, achieving fast route switchover.

After receiving a withdrawal ROUTE-REFRESH message, the device forwards the message to its peers as follows:

- If the message is received from an IBGP peer, the device forwards the message only to EBGP peers.

- If the message is received from an EBGP peer, the device forwards the message to IBGP peers without changing the original AS number of the message.

You can use the **peer capability-advertise withdraw-refresh** command in conjunction with the **peer tracking** command on an upstream device. The **peer tracking** command enables the upstream device to fast detect status of the specified BGP peer session. When the related peer is unreachable, the local device immediately terminates the BGP peer session in time and sends withdrawal ROUTE-REFRESH messages to downstream devices. Collaboration of the two commands expedites forwarding path convergence upon uplink failures.

Restrictions and guidelines

Execute the **peer capability-advertise withdraw-refresh** command with caution, because this operation will terminate and re-establish the BGP session between the local device and the specified peer.

This feature is only applicable to the scenario described in "About this task." It is not applicable to scenarios where upstream devices that advertise withdrawal ROUTE-REFRESH messages are connected to IBGP peers in the uplink direction.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable BGP to exchange withdrawal ROUTE-REFRESH messages with a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **capability-advertise withdraw-refresh**

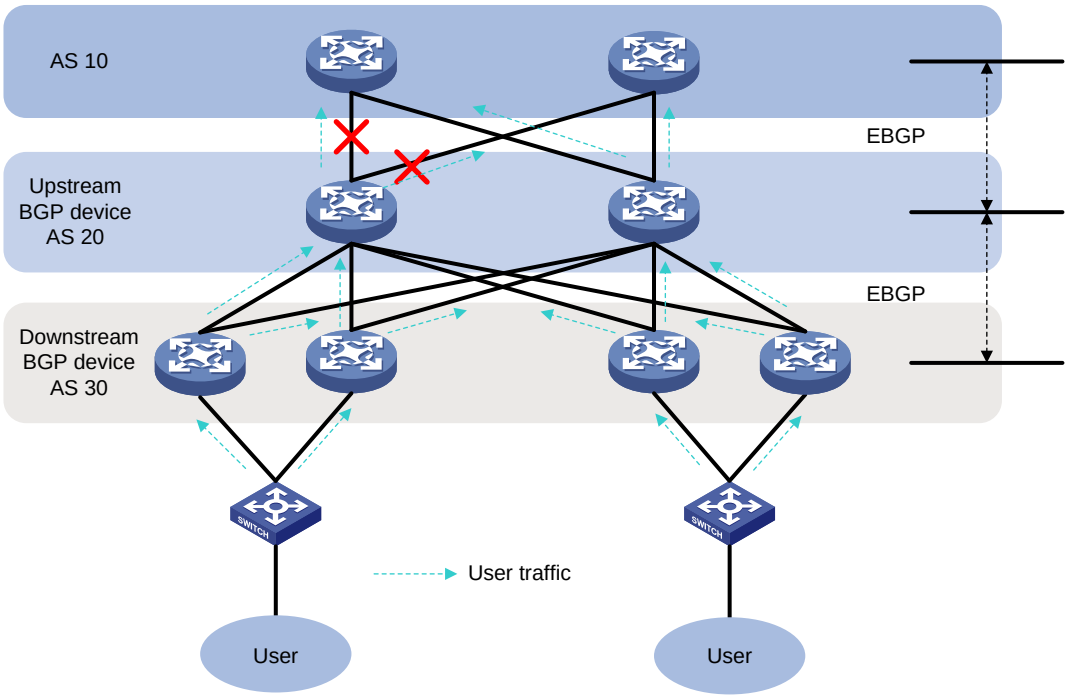
   By default, BGP cannot exchange withdrawal ROUTE-REFRESH messages with any peer or peer group.

Enabling Track-based priority adjustment for BGP routes advertised to a peer or peer group

About this task

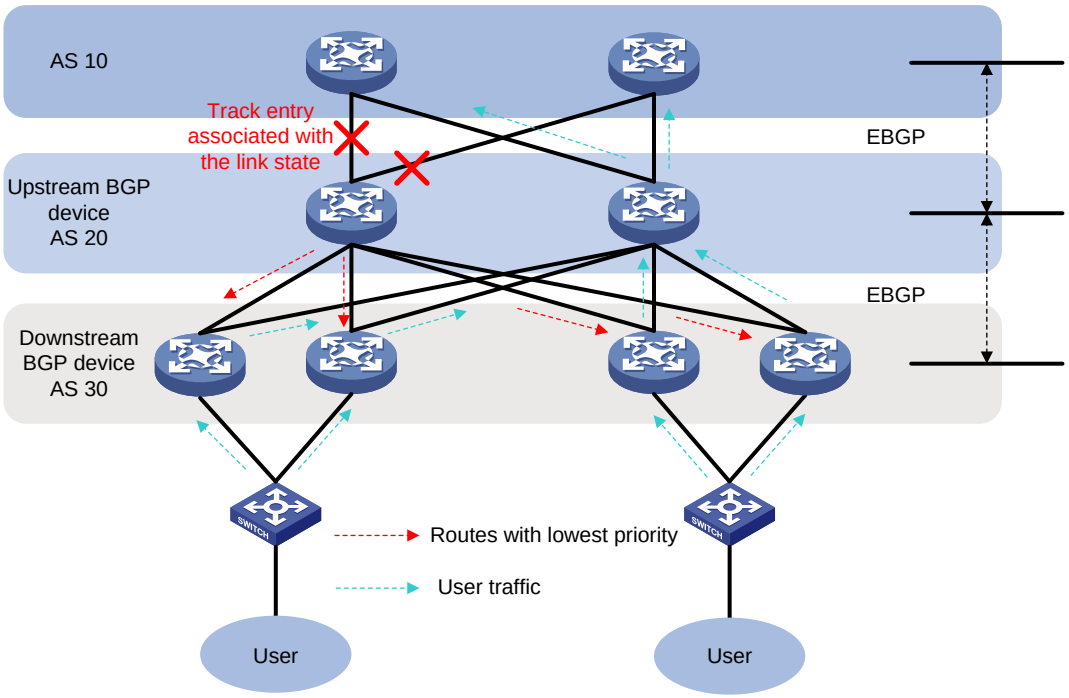
This feature collaborates BGP with Track to achieve fast switchover of traffic forwarding paths in multi-level BGP networks. As shown in Figure 48, two upstream BGP devices are deployed on the network, for example, Device 1 and Device 2. When all uplinks of Device 1 fail and the device cannot complete route convergence in time, the device might also fail to send route withdrawal messages to downstream devices in time. At this time, those downstream devices still forward some user traffic to Device 1, resulting in traffic forwarding failure.

Figure 48 Multi-level BGP network diagram



To avoid this issue, enable BGP to collaborate with Track on upstream BGP devices. When uplink failures occur on an upstream BGP device, route switchover is triggered immediately and user traffic can be forwarded to another upstream BGP device with normal uplinks.

Figure 49 Track-based priority adjustment for BGP routes advertised to a peer or peer group



To enable Track-based priority adjustment for BGP routes advertised to a downstream BGP device, perform the following operations on each upstream BGP device:

1. Associate the link state of its upstream interface with a track entry. For more information about Track, see track configuration in *High Availability Configuration Guide*.
2. Execute the **apply track** command to configure an apply clause that references the associated track entry.
3. Execute the **peer advertise lowest-priority track-route-policy** command to enable BGP to collaborate with Track. The { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } parameter specifies the desired downstream BGP device and the *route-policy-name* parameter specifies the routing policy configured in the previous step.

When certain uplinks of the upstream BGP device go down, the associated track entry transitions to Negative state. After perceiving the track entry state change, BGP immediately performs the following operations on the upstream BGP device:

4. Minimizes the priority of all routes advertised to the specified downstream BGP device by setting the local preference value to 0 and MED value to the maximum.
5. Advertises those routes again to the downstream BGP device.

The downstream BGP device will no longer select routes from the upstream BGP device as optimal routes and will forward user traffic to another upstream BGP device. This mechanism accelerates forwarding path switchover upon uplink failures on an upstream BGP device by skipping the process of hold timer timeout > route invalidation > route withdrawal.

For more information about routing policies and the **apply track** command, see routing policy configuration in *Layer 3—IP Routing Configuration Guide*.

Restrictions and guidelines

If a routing policy is specified in the **peer advertise lowest-priority track-route-policy** command, only the **apply track** clause of that routing policy can take effect.

Although the **peer advertise lowest-priority track-route-policy** command and the **peer capability-advertise withdraw-refresh** command function similarly, they are different in the following aspects:

- To use the feature associated with the **peer advertise lowest-priority track-route-policy** command, you only need to configure this command on the desired device. The device and devices from different vendors can co-locate in the same network. However, this command cannot provide selective traffic forwarding path control, because it minimizes the priority of all routes advertised to the specified peer or peer group.
- The **peer capability-advertise withdraw-refresh** command can take effect only if the specified downstream BGP device supports withdraw ROUTE-REFRESH messages. This command can provide selective traffic forwarding path control, because it can perform traffic forwarding path switchover in scenarios where only uplink failures occur.

Procedure (IPv4 unicast)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view.

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view.

     bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

address-family ipv4 [**unicast**]

3. Enable Track-based priority adjustment for BGP routes advertised to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise lowest-priority track-route-policy** *route-policy-name*

   By default, Track-based priority adjustment is disabled for BGP routes advertised to a peer or peer group.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view.

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

3. Enable Track-based priority adjustment for BGP routes advertised to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise lowest-priority track-route-policy** *route-policy-name*

   By default, Track-based priority adjustment is disabled for BGP routes advertised to a peer or peer group.

Disabling BGP session establishment

About disabling BGP session establishment

This task enables you to temporarily tear down BGP sessions to a peer/peer group or all peers/peer groups. You can perform network upgrade and maintenance without needing to delete and reconfigure the peers and peer groups. To recover the sessions, execute the **undo peer ignore** or **undo ignore all-peers** command.

To lower the priority of the routes advertised by BGP, you can specify the **graceful** keyword in the **peer ignore** command or the **ignore all-peers** command, and configure route attribures. BGP peers can then select other peers' routes as optimal routes, which avoids traffic interruption upon wait timer expiration or peer disconnection. To enable BGP to advertise low-priority routes without tearing down BGP sessions, you can set the value for the **graceful** *graceful-time* option to 0.

If you specify the **graceful** keyword in the **peer ignore** command, BGP performs the following tasks:

1. Starts the wait timer specified with the **graceful** keyword.
2. Advertises all routes to the specified peer or peer group and changes the attribute of the advertised routes to the specified value.

Advertises the routes from the specified peer or peer group to other IBGP peers and peer groups and changes the attribute of the advertised routes to the specified value.

3. Shuts down the session to the specified peer or peer group after the wait timer expires.

If you specify the **graceful** keyword in the **ignore all-peers** command, BGP performs the following tasks:

4. Starts the wait timer specified with the **graceful** keyword.
5. Advertises all routes to all peers and peer groups and changes the attribute of the advertised routes to the specified value.
6. Shuts down all sessions to peers and peer groups after the wait timer expires.

Restrictions and guidelines

For a BGP peer or peer group, the configuration made by the **peer ignore** command takes precedence over the configuration made by the **ignore all-peers** command.

Disabling BGP session establishment with a peer or peer group (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Disable BGP session establishment with a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **ignore** [**graceful** *graceful-time* { **community** { *community-number* | *aa:nn* } | **local-preference** *preference* | **med** *med* } *]

   By default, BGP can establish a session to a peer or peer group.

CAUTION:

- If a session has been established to a peer, executing this command for the peer tears down the session and clears all related routing information.
- If sessions have been established to a peer group, executing this command for the peer group tears down the sessions to all peers in the group and clears all related routing information.

Disabling BGP session establishment with a peer or peer group (IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Disable BGP session establishment with a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **ignore** [**graceful** *graceful-time* { **community** { *community-number* | *aa:nn* } | **local-preference** *preference* | **med** *med* } *]

   interface-peer *interface-type interface-number* **ignore** [**graceful** *graceful-time* { **community** { *community-number* | *aa:nn* } | **local-preference** *preference* | **med** *med* } *]

   By default, BGP can establish a session to a peer or peer group.

CAUTION:

- If a session has been established to a peer, executing the **peer ignore** command for the peer tears down the session and clears all related routing information.
- If sessions have been established to a peer group, executing the **peer ignore** command for the peer group tears down the sessions to all peers in the group and clears all related routing information.

Disabling BGP session establishment with all peers or peer groups

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Disable BGP session establishment with all peers or peer groups.

   ignore all-peers [**graceful** *graceful-time* { **community** { *community-number* | *aa:nn* } | **local-preference** *preference* | **med** *med* } *]

   By default, BGP can establish sessions to all peers and peer groups.

CAUTION:

Executing this command tears down all existing sessions to peers and peer groups and clears all related routing information.

Configuring BGP soft-reset

About BGP soft-reset

After you modify the route selection policy, for example, modify the preferred value, you must reset BGP sessions to apply the new policy. The reset operation tears down and re-establishes BGP sessions.

To avoid tearing down BGP sessions, you can use one of the following soft-reset methods to apply the new policy:

- **Enabling route refresh**—The BGP router advertises a ROUTE-REFRESH message to the specified peer, and the peer resends its routing information to the router. After receiving the routing information, the router filters the routing information by using the new policy.

  This method requires that both the local router and the peer support route refresh.
- **Saving updates**—Use the **peer keep-all-routes** command to save all route updates from the specified peer. After modifying the route selection policy, filter routing information by using the new policy.

  This method does not require that the local router and the peer support route refresh but it uses more memory resources to save routes.
- **Manual soft-reset**—Use the **refresh bgp** command to enable BGP to send local routing information or advertise a ROUTE-REFRESH message to the specified peer. The peer then resends its routing information. After receiving the routing information, the router filters the routing information by using the new policy.

  This method requires that both the local router and the peer support route refresh.

Enabling route refresh (IPv4 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable BGP route refresh for a peer or peer group.
   - Enable BGP route refresh for the specified peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] }
     capability-advertise route-refresh
   - Enable the BGP route refresh, multi-protocol extension, and 4-byte AS number features for the specified peer or peer group.

     undo peer { *group-name* | *ipv4-address* [*mask-length*] }
     capability-advertise conventional

   By default, the BGP route refresh, multi-protocol extension, and 4-byte AS number features are enabled.

Enabling route refresh (IPv6 peers)

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable BGP route refresh for a peer or peer group.
   - Enable BGP route refresh for the specified peer or peer group.

     peer { *group-name* | *ipv6-address* [*prefix-length*] } **capability-advertise route-refresh**
   - Enable the BGP route refresh, multi-protocol extension, and 4-byte AS number features for the specified peer or peer group.

     undo peer { *group-name* | *ipv6-address* [*prefix-length*] } **capability-advertise conventional**

   By default, the BGP route refresh, multi-protocol extension, and 4-byte AS number features are enabled.

Saving updates (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast
3. Save all route updates from the peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **keep-all-routes**

   interface-peer *interface-type interface-number* **keep-all-routes**

   By default, route updates from peers and peer groups are not saved.

The *ipv6-address* [*prefix-length*] argument and the **interface-peer** command are not supported in BGP IPv4 multicast address family view.

This command takes effect only for the routes received after this command is executed.

Saving updates (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 multicast

3. Save all route updates from the peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **keep-all-routes**

   interface-peer *interface-type interface-number* **keep-all-routes**

   By default, route updates from peers and peer groups are not saved.

   This command takes effect only for the routes received after this command is executed.

   The *ipv4-address* [*mask-length*] option, the *link-local-address* **interface** *interface-type interface-number* option, and the **interface-peer** command are not supported in BGP IPv6 multicast address family view.

Configuring manual soft-reset (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Enable BGP route refresh for a peer or peer group.

- Enable BGP route refresh for the specified peer or peer group.

  peer { *group-name* | *ipv4-address* [*mask-length*] }
  capability-advertise route-refresh

- Enable the BGP route refresh, multi-protocol extension, and 4-byte AS number features for the specified peer or peer group.

  undo peer { *group-name* | *ipv4-address* [*mask-length*] }
  capability-advertise conventional

By default, the BGP route refresh, multi-protocol extension, and 4-byte AS number features are enabled.

4. Return to user view.

   return

5. Perform manual soft-reset. Choose one option as needed:

   - Perform manual soft-reset on IPv4 sessions in BGP IPv4 address family.

     refresh bgp [**instance** *instance-name*] { *ipv4-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **ipv4** [**multicast** | [**unicast**] [**vpn-instance** *vpn-instance-name*]]

   - Perform manual soft-reset on IPv6 sessions in BGP IPv4 address family.

     refresh bgp [**instance** *instance-name*] { *ipv6-address* [*prefix-length*] | [*link-local-address*] **interface** *interface-type interface-number* } { **export** | **import** } **ipv4** [**unicast**]

Configuring manual soft-reset (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.

   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]

   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Enable BGP route refresh for a peer or peer group.

   - Enable BGP route refresh for the specified peer or peer group.

     peer { *group-name* | *ipv6-address* [*prefix-length*] }
     capability-advertise route-refresh

   - Enable the BGP route refresh, multi-protocol extension, and 4-byte AS number features for the specified peer or peer group.

     undo peer { *group-name* | *ipv6-address* [*prefix-length*] }
     capability-advertise conventional

   By default, the BGP route refresh, multi-protocol extension, and 4-byte AS number features are enabled.

4. Return to user view.

   return

5. Perform manual soft-reset. Choose one option as needed:

   - Perform manual soft-reset on IPv6 sessions in BGP IPv6 address family.

refresh bgp [**instance** *instance-name*] { *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **ipv6** [**multicast** | [**unicast**] [**vpn-instance** *vpn-instance-name*]]

- Perform manual soft-reset on IPv4 sessions in BGP IPv6 address family.

  refresh bgp [**instance** *instance-name*] *ipv4-address* [*mask-length*] { **export** | **import** } **ipv6** [**unicast**]

Configuring BGP load balancing

About this task

Perform this task to specify the maximum number of BGP ECMP routes for load balancing.

Restrictions and guidelines

When you use the **balance** command, follow these guidelines:

- If you specify the **ecmp-nexthop-local** keyword, BGP sets the local device IP as the next hop of each load balancing route.
- If you specify the **ecmp-nexthop-unchanged** keyword, BGP does not change the next hop of each load balancing route.
- If you specify neither the **ecmp-nexthop-local** keyword nor the **ecmp-nexthop-unchanged** keyword, BGP advertises the routes used for load balancing to IBGP peers as follows:
  - Without the Add-Path feature configured, BGP advertises only the optimal route among the routes used for load balancing and sets the local device IP as the next hop of that optimal route.
  - With the Add-Path feature configured, BGP performs the following tasks:
    - Advertises the routes used for load balancing. The number of routes that BGP can advertise equals the maximum number of Add-Path optimal routes that can be advertised.
    - Sets the local device IP as the next hop of the optimal route among the advertised routes.

Both the **balance igp-metric-ignore** and **bestroute igp-metric-ignore** commands can enable BGP to use routes with different IGP metrics for load balancing. The two commands are different as follows:

- The **balance igp-metric-ignore** command can be used only for load balancing. It does not change the rules of optimal route selection. BGP still compares the IGP metric value of each route during optimal route selection.
- The **bestroute igp-metric-ignore** command takes effect on optimal route selection and route selection for load balancing. With this command executed, BGP will ignore the IGP metric values of routes in both cases.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]

   - Execute the following commands in sequence to enter BGP-VPN instance view:

bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

3. (Optional.) Enable BGP to ignore IGP metrics during optimal route selection.

   bestroute igp-metric-ignore

   By default, BGP compares IGP metrics during optimal route selection, and selects the route with the smallest IGP metric as the optimal route.

   BGP cannot use routes with different IGP metrics to implement load balancing. To resolve this issue, you can use this command.

4. Return to system view.
   - In BGP instance view:

     quit
   - In BGP-VPN instance view:

     quit

     quit

5. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast

6. Specify the maximum number of BGP ECMP routes for load balancing.

   balance [**ebgp** | **eibgp** | **ibgp**] *number* [**ecmp-nexthop-local** | **ecmp-nexthop-unchanged**]

   By default, load balancing is disabled.

7. (Optional.) Enable BGP to ignore the AS_PATH attribute when it implements load balancing.

   balance as-path-neglect

   By default, BGP does not ignore the AS_PATH attribute when it implements load balancing.

8. (Optional.) Enable BGP to perform load balancing for routes that have different AS_PATH attributes of the same length.

   balance as-path-relax [**ebgp** | **ibgp**]

   By default, BGP cannot perform load balancing for routes that have different AS_PATH attributes of the same length.

9. (Optional.) Enable BGP to use routes with different IGP metrics for load balancing.

   balance igp-metric-ignore

   By default, BGP cannot use routes with different IGP metrics for load balancing.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. (Optional.) Enable BGP to ignore IGP metrics during optimal route selection.
   bestroute igp-metric-ignore

   By default, BGP compares IGP metrics during optimal route selection, and selects the route with the smallest IGP metric as the optimal route.

   BGP cannot use routes with different IGP metrics to implement load balancing. To resolve this issue, you can use this command.
4. Return to system view.
   - In BGP instance view:
     quit
   - In BGP-VPN instance view:
     quit
     quit
5. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv6 multicast address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family ipv6 multicast
6. Specify the maximum number of BGP ECMP routes for load balancing.
   balance [**ebgp** | **eibgp** | **ibgp**] *number* [**ecmp-nexthop-local** | **ecmp-nexthop-unchanged**]

   By default, load balancing is disabled.
7. (Optional.) Enable BGP to ignore the AS_PATH attribute when it implements load balancing.
   balance as-path-neglect

   By default, BGP does not ignore the AS_PATH attribute when it implements load balancing.
8. (Optional.) Enable BGP to perform load balancing for routes that have different AS_PATH attributes of the same length.

balance as-path-relax [**ebgp** | **ibgp**]

By default, BGP cannot perform load balancing for routes that have different AS_PATH attributes of the same length.

9. (Optional.) Enable BGP to use routes with different IGP metrics for load balancing.

   balance igp-metric-ignore

   By default, BGP cannot use routes with different IGP metrics for load balancing.

Configuring the BGP Additional Paths feature

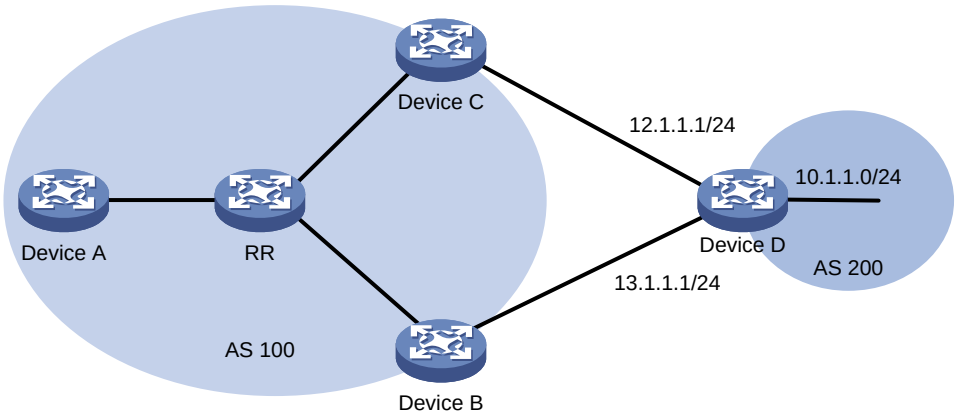
About this task

On a large-scale network, a route reflector can receive multiple BGP routes with the same prefix. By default, BGP advertises only one optimal route. When the optimal route fails, traffic forwarding will be interrupted until route convergence completes.

To resolve this issue, configure the BGP Additional Paths (Add-Path) feature on the route reflector. This feature enables the device to advertise multiple routes with the same prefix to a peer or peer group. On receipt of these routes, the peer generates multiple links to the same destination network. You can use these links to implement load balancing with the **balance** command or use them as backup routes. When the optimal link fails, another link can take over to forward traffic, which enhances the reliability of data transmission.

As shown in Figure 50, Device A, Device B, and Device C are located in AS 100, exchanging routing information within the domain through the RR. Both Device B and Device C establish an EBGP session to Device D and learn a route with prefix 10.1.1.0/24 from Device D.

Figure 50 Network diagram



In this situation, the RR can learn two routes destined for 10.1.1.0/24 from Device B and Device C. To have the RR advertise multiple routes with the same prefix to Device A, configure the BGP Additional Paths feature. Table 14 shows the effect of the BGP Additional Paths feature on route advertisement and route convergence upon link failure:

Table 14 Route advertisement and route convergence upon link failure

Whether the BGP Additional Paths feature is configured	Route advertisement	Route convergence upon failure of a single link
Not configured	The RR selects the optimal route from the two routes destined for 10.1.1.0/24, and only advertises the optimal route to Device A.	Device A needs to perform route convergence again.

Configured	RR advertises the two routes destined for 10.1.1.0/24 to Device A. The next hops of the two routes are 12.1.1.1 and 13.1.1.1, respectively.	• If Device A establishes an FRR relationship between the two routes and the primary path to 10.1.1.0/24 fails, Device A can quickly use the backup path for traffic forwarding. • If Device A establishes an ECMP relationship between the two routes and one of them fails, Device A switches all traffic to the other path.

Restrictions and guidelines

BGP Additional Paths capabilities include the sending capability and the receiving capability. You can enable either of the sending and receiving capabilities or both on a peer. For two BGP peers to successfully negotiate the Additional Paths capabilities, make sure one end has the sending capability and the other end has the receiving capability.

Procedure (IPv4 unicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

3. Configure the BGP Additional Paths capabilities. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **additional-paths** { **receive** | **send** } *

   interface-peer *interface-type interface-number* **additional-paths** { **receive** | **send** } *

   By default, no BGP Additional Paths capabilities are configured.

4. Set the maximum number of Add-Path optimal routes that can be advertised to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise additional-paths best** *number*

   interface-peer *interface-type interface-number* **advertise additional-paths best** *number*

   By default, a maximum of one Add-Path optimal route can be advertised to a peer or peer group.

5. Set the maximum number of Add-Path optimal routes that can be advertised to all peers.

   additional-paths select-best *best-number*

   By default, a maximum of one Add-Path optimal route can be advertised to all peers.

Procedure (IPv6 unicast address family)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

3. Configure the BGP Additional Paths capabilities. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **additional-paths** { **receive** | **send** } *

   interface-peer *interface-type interface-number* **additional-paths** { **receive** | **send** } *

   By default, no BGP Additional Paths capabilities are configured.

4. Set the maximum number of Add-Path optimal routes that can be advertised to a peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **advertise additional-paths best** *number*

   interface-peer *interface-type interface-number* **advertise additional-paths best** *number*

   By default, a maximum of one Add-Path optimal route can be advertised to a peer or peer group.

5. Set the maximum number of Add-Path optimal routes that can be advertised to all peers.

   additional-paths select-best *best-number*

   By default, a maximum of one Add-Path optimal route can be advertised to all peers.

Protecting an EBGP peer when memory usage reaches level 2 threshold

About this task

Memory usage includes the following threshold levels: normal, level 1, level 2, and level 3. When the level 2 threshold is reached, BGP periodically tears down an EBGP session to release memory resources until the memory usage falls below the level 2 threshold. You can configure this feature to avoid tearing down the EBGP session to an EBGP peer when the memory usage reaches the level 2 threshold.

For more information about memory usage thresholds, see device management configuration in *Fundamentals Configuration Guide*.

Procedure (IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Configure BGP to protect an EBGP peer or peer group when the memory usage reaches level 2 threshold.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **low-memory-exempt**

   By default, BGP periodically tears down an EBGP session to release memory resources when level 2 threshold is reached.

Procedure (IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Configure BGP to protect an EBGP peer or peer group when the memory usage reaches level 2 threshold. Choose one command as needed:
   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **low-memory-exempt**

   interface-peer *interface-type interface-number* **low-memory-exempt**

   By default, BGP tears down an EBGP session to release memory resources periodically when level 2 threshold is reached.

Setting a DSCP value for outgoing BGP packets

About this task

The DSCP value of an IP packet specifies the priority level of the packet and affects the transmission priority of the packet.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

3. Set a DSCP value for outgoing BGP packets. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **dscp** *dscp-value*

   interface-peer *interface-type interface-number* **dscp** *dscp-value*

   By default, the DSCP value for outgoing BGP packets is 48.

Disabling route recursion policy control for routes received from a peer or peer group

About this task

If you configure routing policy-based recursive lookup for BGP routes, this route recursion policy applies to BGP routes learned from all peers. This task allows you to disable route recursion policy control for routes learned from certain peers, for example, direct EBGP peers.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Disable route recursion policy control for routes received from the specified peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number*} **nexthop-recursive-policy disable**

   interface-peer *interface-type interface-number* **nexthop-recursive-policy disable**

   By default, the route recursion policy applies to routes received from the peer or peer group.

Flushing the suboptimal BGP route to the RIB

About this task

This feature flushes the suboptimal BGP route to the RIB when the following conditions are met:

- The optimal route is generated by the **network** command or is redistributed by the **import-route** command.
- The suboptimal route is received from a BGP peer.

After the suboptimal route is flushed to the RIB on a network, BGP immediately switches traffic to the suboptimal route when the optimal route fails.

For example, the device has a static route to the subnet 1.1.1.0/24 that has a higher priority than a BGP route. BGP redistributes the static route and receives a route to 1.1.1.0/24 from a peer. After

the **flush suboptimal-route** command is executed, BGP flushes the received BGP route to the RIB as the suboptimal route. When the static route fails, BGP immediately switches traffic to the suboptimal route if inter-protocol FRR is enabled. For more information about inter-protocol FRR, see "Configuring basic IP routing."

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Flush the suboptimal BGP route to the RIB.

   flush suboptimal-route

   By default, BGP is disabled from flushing the suboptimal BGP route to the RIB, and only the optimal route is flushed to the RIB.

Configuring a peer monitoring group

About this task

A peer monitoring group is mainly used in traffic switchover scenarios. For example, when the local device receives multiple BGP routes to the same destination from multiple BGP peers in AS 10 and AS 20, the traffic to that destination can be load shared between AS 10 and AS 20. If multiple BGP peers within AS 10 are disconnected from the local device, the network administrator might perceive a risk in the network status within AS 10 even if there are still running BGP sessions within the AS. The administrator might need to switch all traffic passing through AS 10 to other ASs. You can configure this feature to implement automatic traffic switchover in such scenarios.

Execute the **monitor-group** command to create a peer monitoring group, and then execute the **peer monitor-group** command to add shared-risk BGP peers to the group. The state of a BGP peer in a peer monitoring group depends on the state of the BGP session between the local device and that peer:

- If the BGP session is in Established state when you add the peer to the peer monitoring group, the peer is in UP state. If the BGP session is not in Established state, the peer is in INIT state.
- If the BGP session state transitions from INIT to Established after you add the peer to the peer monitoring group, the peer state will change to UP.
- If the BGP session state transitions from Established to another state, the peer state will change from UP to DOWN. This rule does not apply to the following scenarios where BGP configurations bring about BGP session state changes:
  - The BGP session is terminated during a normal GR process. In this situation, the state of the BGP peer does not change in the peer monitoring group.
  - The BGP session is terminated due to other BGP configurations (for example, the **peer ignore** command). In this situation, if the number of DOWN-state BGP peers in the peer monitoring group does not reach the specified threshold, the state of the disconnected peer will change to INIT. If the specified threshold is reached, the state of the disconnected peer remains DOWN.

When the number of DOWN-state BGP peers in the peer monitoring group reaches the specified threshold, the local device will disconnect from all BGP peers in the group to avoid potential traffic loss.

With the peer monitoring group feature configured, when the number of DOWN-state BGP peers reaches the specified threshold, the local device disconnects from all peers in the peer monitoring group. The local device will not proactively establish sessions to the devices in the group. To re-establish BGP sessions to the devices in the peer monitoring group, execute the **reset bgp monitor-group** command. This command places the entire peer monitoring group in INIT state.

The state of each normal BGP session will be restored to Established in the group, and the state of BGP peers corresponding to those BGP sessions will be restored to UP in the group.

Restrictions and guidelines

The number of BGP sessions in a peer monitoring group depends on the number of BGP peers in the group. Each BGP peer is counted as one BGP session, regardless of whether the peer has established sessions in multiple address families. The BGP sessions created by using the same IP address in different BGP-VPN instances and the public instance are considered as different sessions.

A BGP peer or peer group can be added to only one peer monitoring group.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Create a peer monitoring group.
   monitor-group *monitor-group-name* **threshold** *number*
   By default, no peer monitoring groups exist.
4. (Optional.) Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
5. Add the specified peer or peer group to the peer monitoring group.
   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **monitor-group** *monitor-group-name*
   By default, a peer or peer group does not belong to any peer monitoring group.
6. (Optional.) Restore the normal sessions in the peer monitoring group.
   reset bgp [**instance** *instance-name*] **monitor-group** *monitor-group-name*
   This command is supported in user view. Before you use this command, execute the **quit** *command once or more times as needed to return to user view.*

Disabling BGP from flushing specific routes to the routing table

About this task

Perform this task to enable BGP to not flush routes matching the specified routing policy to the routing table.

This feature applies to the following scenarios:

- In an MPLS L3VPN network, all the optimal private BGP routes will be flushed to the routing table of a PE by default. You can configure this feature to prevent unnecessary routes from being flushed to the IP routing table and improve the forwarding performance of the PE.
- In a route reflection network, configure this feature to not flush specific BGP routes to the routing table of the RR. Then, the RR only advertises and receives routes but does not forward service traffic, which saves the system resources of the RR.

Procedure (IPv4 unicast)

1. Enter system view.
   system-view

2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
3. Disable BGP from flushing routes matching the specified routing policy to the routing table.

   routing-table bgp-rib-only [[**all**] [**route-policy** *route-policy-name*] | **cross-pod-host**]

   By default, BGP flushes the optimal routes to the routing table.

   The **cross-pod-host** keyword is supported only in BGP IPv4 unicast address family view.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view
2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
3. Disable BGP from flushing routes matching the specified routing policy to the routing table.

   routing-table bgp-rib-only [[**all**] [**route-policy** *route-policy-name*] | **cross-pod-host**]

   By default, BGP flushes the optimal routes to the routing table.

   The **cross-pod-host** keyword is supported only in BGP IPv6 unicast address family view.

Specifying a label allocation mode

About this task

BGP supports the following label allocation modes:

- **Per-prefix**—Allocates a label to each route prefix.
- **Per-next-hop**—Allocates a label to each next hop. This mode is applicable when the number of labels required by the per-prefix mode exceeds the maximum number of labels supported by the device.
- **Per-VPN-instance**—Allocates a label to each VPN instance. This mode is applicable when the number of labels required by the per-next-hop mode exceeds the maximum number of labels supported by the device.

Restrictions and guidelines

When you specify the per-prefix or per-next-hop label allocation mode, you can execute the **vpn popgo** command to specify the POPGO forwarding mode on an egress PE. The egress PE will pop the label for each packet and forward the packet out of the interface corresponding to the label.

When you specify the per-VPN instance label allocation mode, do not execute the **vpn popgo** command because it is mutually exclusive with the **label-allocation-mode per-vrf** command. The egress PE will pop the label for each packet and forward the packet through the FIB table.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Specify a label allocation mode.
   label-allocation-mode { **per-prefix** | **per-vrf** }
   By default, BGP allocates labels on a per-next-hop basis.

CAUTION:
A change to the label allocation mode enables BGP to re-advertise all routes, which will cause temporary service interruption. Use this feature with caution.

Disabling optimal route selection for labeled routes without tunnel information

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Disable optimal route selection for labeled routes without tunnel information.
   labeled-route ignore-no-tunnel
   By default, labeled routes without tunnel information can participate in optimal route selection.

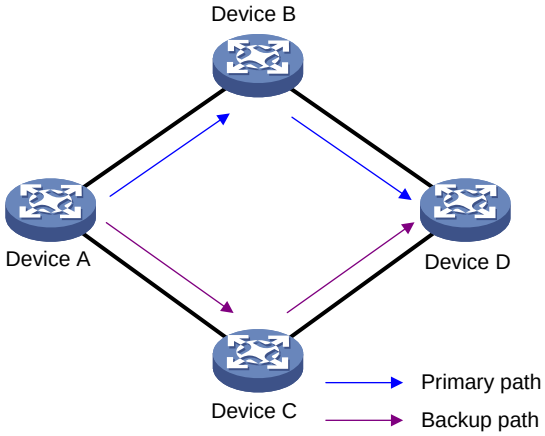
Configuring BGP optimal route selection delay

About this task

Typically BGP optimal route selection is triggered in real time by the events such as attribute change, configuration change, and route recursion. To avoid packet loss upon switchover between redundant links, you can perform this task to delay optimal route selection.

As shown in Figure 51, BGP runs on all devices in the network. Device A and Device D uses the primary path for communication. When the primary path fails, Device A and Device D switch to the backup path for communication and then back to the primary path after the primary path recovers. In such case, traffic loss might occur if Device A forwards packets through Device B before Device B completes route convergence. You can configure optimal route selection delay on Device A to resolve the problem.

Figure 51 Network diagram



To configure BGP optimal route selection delay, you can use either of the following commands:

- **`route-select delay`**

  If you execute this command for an address family, the device delays optimal route selection as long as the address family has a route change.

- **`route-select suppress on-peer-up`**

  When the state of a peer changes from DOWN to UP, the device delays optimal route selection only for BGP routes received from that peer.

If you configure the two commands synchronously, only the longer delay timer takes effect.

Restrictions and guidelines

Follow these restrictions and guidelines when you configure this feature:

- The optimal route selection delay setting applies only when multiple effective routes with the same prefix exist after a route change occurs.
- For routes being delayed for optimal route selection, modifying the optimal route selection delay timer has the following effects:
  - If you modify the delay timer to a non-zero value, the routes are not affected, and they still use the original delay timer.
  - If you execute the **undo** form of the **`route-select delay`** command or modify the delay timer to 0, the device performs optimal route selection immediately.
- If you configure this feature multiple times for an address family, the most recent configuration takes effect.
- The optimal route selection delay configuration does not apply to the following conditions:
  - A route change is caused by execution of a command or by route withdrawal.
  - After a route change occurs, only one route exists for a specific destination network.
  - An active/standby process switchover occurs.
  - A route change occurs among equal-cost routes.
  - Only the optimal and suboptimal routes exist when FRR is configured.
  - Optimal route selection is triggered by a redistributed route.
- To have the device immediately perform optimal route selection for routes being delayed, execute either of the following commands:
  - **`undo route-select suppress on-peer-up`**

    This command takes effect only on routes that are delayed by the **`route-select suppress on-peer-up`** command.

- **undo route-select delay**

  This command takes effect only on routes that are delayed by the **route-select suppress on-peer-up** command or the **route-select delay** command.

Procedure (IPv4 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view, BGP-VPN IPv4 unicast address family view, or BGP IPv4 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP IPv4 multicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 multicast
3. Set the optimal route selection delay timer. Choose one of the following option:
   - Enable optimal route selection delay for all routes in the address family.

     route-select delay *delay-value*

     With this command configured, the device delays optimal route selection as long as the address family has a route change.
   - Enable optimal route selection delay based on DOWN-to-UP peer state changes.

     route-select suppress on-peer-up *milliseconds*

     With this command configured, when the state of a peer changes from DOWN to UP, the device performs optimal route selection for BGP routes received from that peer only after the delay timer expires.

   By default, the device does not delay optimal route selection.

Procedure (IPv6 unicast/multicast address family)

1. Enter system view.

   system-view
2. Enter BGP IPv6 unicast address family view, BGP-VPN IPv6 unicast address family view, or BGP IPv6 multicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

address-family ipv6 [**unicast**]

- Execute the following commands in sequence to enter BGP IPv6 multicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv6 multicast

3. Set the optimal route selection delay timer. Choose one of the following option:
   - Enable optimal route selection delay for all routes in the address family.

     route-select delay *delay-value*

     With this command configured, the device delays optimal route selection as long as the address family has a route change.
   - Enable optimal route selection delay based on DOWN-to-UP peer state changes.

     route-select suppress on-peer-up *milliseconds*

     With this command configured, when the state of a peer changes from DOWN to UP, the device performs optimal route selection for BGP routes received from that peer only after the delay timer expires.

   By default, the device does not delay optimal route selection.

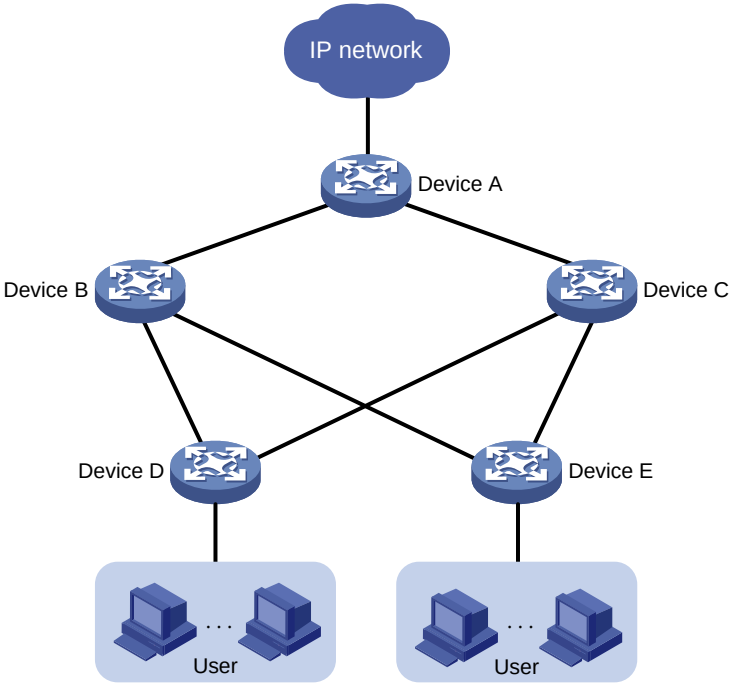
Configuring the BGP session state to collaborate with the monitor link downlink interfaces

About this task

On the network as shown in Figure 52, Device A establishes BGP sessions to Device B and Device C, and both Device B and Device C establish BGP sessions to Device D and Device E. Users accessing Device D and Device E communicate with the external network through the BGP routes deployed by Device A. In normal conditions, traffic from users to the external network is load-shared between Device B and Device C.

In this scenario, if the link between Device A and Device B fails, the BGP session between Device A and Device B will be terminated. Then, the BGP routes advertised by Device A cannot be forwarded to Device D and Device E through Device B. As a result, the traffic from users to the external network will not be forwarded through Device B. In this case, BGP sessions from Device B to Device D and Device E do not need to be established. You can disconnect these BGP sessions to save system resources.

Figure 52 Network diagram for the BGP session state to collaborate with the monitor link downlink interfaces



With this feature configured, when the BGP session between Device A and Device B is disconnected, the BGP sessions from Device B to Device D and Device E are also automatically disconnected. When the BGP session between Device A and Device B recovers, the BGP sessions from Device B to Device D and Device E also automatically recover.

After this feature is configured, the BGP session between the local device and the peer or peer group specified by this feature will be used as a type of uplink interface in the specified monitor link group. The uplink interfaces and downlink interfaces in a monitor link group collaborate as follows:

- Uplink interfaces are defined as follows:
  - In a monitor link group, an interface specified by using the **`port uplink`** or **`port monitor-link group uplink`** command is considered as an uplink interface.
  - In a monitor link group, each BGP process specified by using the **`monitor bgp`** command is considered as an uplink interface.
  - In a monitor link group, a BGP session specified by using the **`peer monitor-link group`** command is considered as an uplink interface.

  If the BGP sessions specified by using the **`peer monitor-link group`** command are established in the BGP process specified by using the **`monitor bgp`** command, and at least one of these BGP session is in **Established** state, the BGP process is not an uplink interface.

  The BGP sessions specified by using the **`peer monitor-link group`** command are not considered as uplink interfaces if the following conditions exist:
  - These BGP sessions are established in the BGP process specified by using the **`monitor bgp`** command.
  - None of these BGP sessions is in **Established** state.

  In this situation, only the BGP process is considered as an uplink interface.

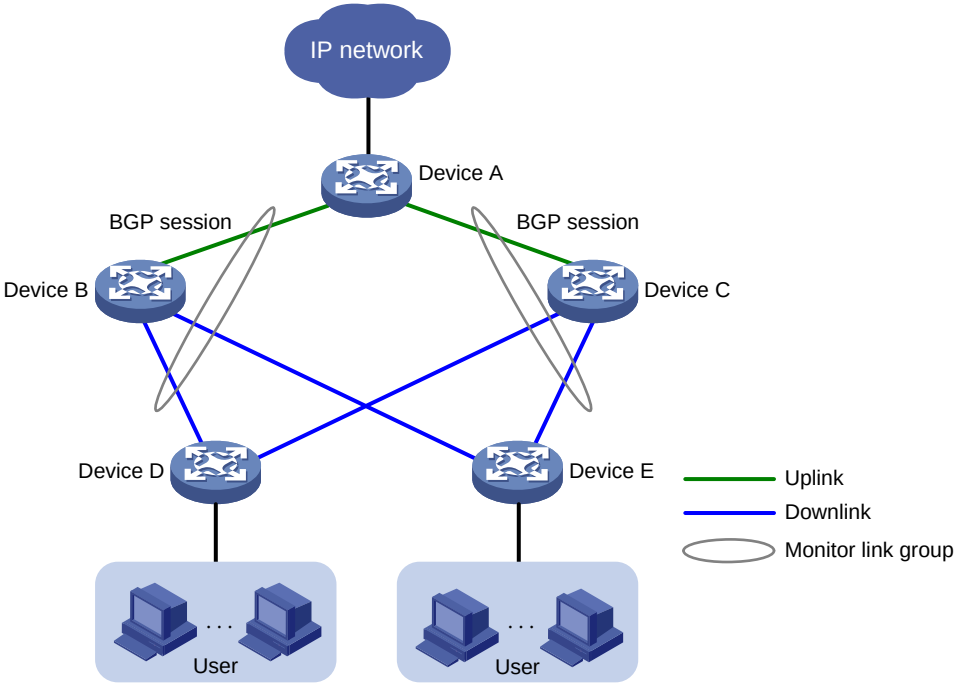
  For example, use the **`monitor bgp`** command to specify BGP instances **aaa** and **bbb** as two uplink interfaces in a monitor link group, use the **`peer monitor-link group`** command in BGP instance **aaa** to specify two BGP sessions as the uplink interfaces, and do not use the **`peer monitor-link group`** command in BGP instance **bbb** to specify uplink interfaces. In

this case, when at least one of the two specified BGP sessions are in **Established** state, the monitor link group has three uplink interfaces in total, including two BGP sessions and one BGP process without BGP sessions specified. When neither of the two specified BGP sessions is in **Established** state, the monitor link group has two uplink interfaces in total. One uplink interface is the BGP process with BGP sessions specified but neither of the BGP sessions is in **Established** state. The other is the BGP process without BGP sessions specified.

- Downlink interfaces are defined as follows: In a monitor link group, an interface specified by using the **port downlink** or **port monitor-link group downlink** command is considered as a downlink interface.
- Uplink interfaces and downlink interfaces collaborate as follows:
  - When the number of uplink interfaces in up state in a monitor link group is lower than the threshold configured by using the **uplink up-port-threshold** command, the physical state is set to down for all downlink interfaces in the monitor link group. An uplink interface is determined as up or down following these rules:
    - For an interface specified by using the **port uplink** or **port monitor-link group uplink** command, the uplink interface is considered as up if the interface is physically up, and the uplink interface is considered as down if the interface is physically down.
    - For a BGP process specified by using the **monitor bgp** command, the uplink interface is considered as up when the BGP instance exists and runs normally, and the uplink interface is considered as down when the BGP instance is deleted and corresponding process exits.
    - For a BGP session specified by using the **peer monitor-link group** command, the uplink interface is considered as up when the BGP session is in **Established** state, and the uplink interface is considered as down when the BGP session is in any other state.
  - When the number of uplink interfaces in up state in a monitor link group is greater than or equal to the threshold configured by using the **uplink up-port-threshold** command, the physical state is set to up for all downlink interfaces in the monitor link group.

As shown in Figure 53, use Device B as an example. Configure the monitor link group as follows: Configure the BGP session between Device A and Device B as the only uplink interface of the monitor link group. Configure the interfaces connecting Device B to Device D and Device E as the downlink interfaces of the monitor link group. When the BGP session between Device A and Device B is disconnected, the physical interfaces connecting Device B to Device D and Device E go down physically. As a result, the BGP sessions from Device B to Device D and Device E are automatically disconnected, and you do not need to modify the BGP-related settings. When the BGP session between Device A and Device B restores to the **Established** state, the physical interfaces connecting Device B to Device D and Device E come up physically. As a result, the BGP sessions from Device B to Device D and Device E automatically recover, and you still do not need to modify the BGP-related settings.

Figure 53 Schematic diagram for the BGP session state to collaborate with the monitor link downlink interfaces



For more information about monitor link, see monitor link configuration in *High Availability Configuration Guide*. For more information about the **monitor bgp**, **monitor-link group**, **port**, and **port monitor-link group** commands, see monitor link commands in *High Availability Command Reference*.

Restrictions and guidelines

To prevent BGP sessions flappings from causing frequent downlink switchovers in a monitor link group, execute the **downlink up-delay** command in the monitor link group to set the switchover delay for the downlink interfaces in the monitor link group.

For more information about the **downlink up-delay** command, see monitor link commands in *High Availability Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Create a monitor link group and enter its view.

   monitor-link group *group-id*

   For more information about this command, see monitor link commands in *High Availability Command Reference*.

3. Configure downlink interfaces for the monitor link group. Choose one of the following tasks:
   - Configure downlink interfaces for the monitor link group in monitor link group view.

     port *interface-type* { *interface-number* | *interface-number.subnumber* } **downlink**
   - Execute the following commands in sequence to configure downlink interfaces for the monitor link group in interface view.

     quit

interface *interface-type* { *interface-number* | *interface-number.subnumber* }

port monitor-link group *group-id* **downlink**

By default, a monitor link group does not have member ports.

For more information about this command, see monitor link commands in *High Availability Command Reference*.

4. (Optional.) Configure the uplink interface threshold for triggering state switchover in the monitor link group.

   uplink up-port-threshold *number-of-port*

   By default, the uplink interface threshold for triggering state switchover in the monitor link group is 1.

   This command is supported only in monitor link group view.

   For more information about this command, see monitor link commands in *High Availability Command Reference*.

5. Return to system view.

   quit

6. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

7. Configure the BGP session state to collaborate with the monitor link downlink interfaces.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **monitor-link group** *group-id*

   By default, the BGP session state does not collaborate with the monitor link downlink interfaces.

Specifying the longest match principle for next hop recursion

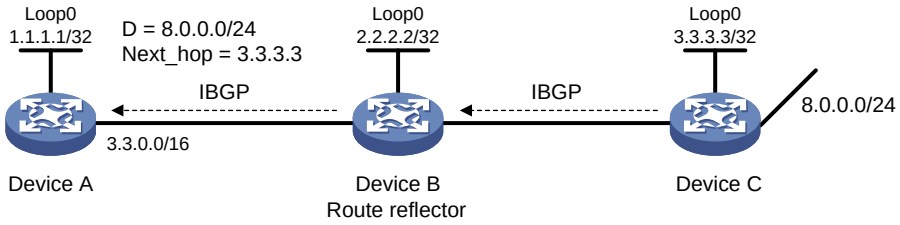
About this task

By default, BGP first looks up in its directly connected routes for a recursive route during a route recursion. If no matching routes exist, BGP will look up in routes generated by all routing protocols based on the longest match principle. The device considers the next hop of a received BGP route as directly connected when the next hop belongs to the same subnet as a direct route. As a result, packet forwarding failure will occur when the device uses this BGP route to forward matching packets.

As shown in Figure 54, Device B has established an IBGP session to Device A and Device C respectively. All devices use a loopback interface as the source port for TCP connection. On receipt of BGP route 8.0.0.0/24 from Device C, Device B reflects the route to Device A. The addresses of the interfaces that connect Device A and Device B belong to subnet 3.3.0.0/16.

If the next hop information exchanged between IBGP peers remains unchanged, the next hop of BGP route 8.0.0.0/24 is 3.3.3.3. Device A considers 3.3.3.3 as a directly connected next hop because 3.3.3.3 belongs to subnet 3.3.0.0/16 and a direct route to 3.3.0.0/16 exists. Packet forwarding failure will occur if Device A forwards packets destined for subnet 8.0.0.0/24 over 3.3.3.3.

Figure 54 Next hop recursion error



To resolve this issue, you can specify the longest match principle for next hop recursion to ensure the recursive route is reachable. BGP will directly look up in routes generated by all routing protocols and select a matching route based on the longest match principle.

Procedure (IPv4 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

3. Specify the longest match principle for BGP next hop recursion.

   nexthop recursive-lookup longest-match [**route-policy** *route-policy-name*]

   By default, BGP first looks up in its directly connected routes for a recursive route during route recursion.

   This command applies to all BGP routes received from peers in one of the following scenarios:
   - The **route-policy** *route-policy-name* option is not specified.
   - The specified routing policy does not exist.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

3. Specify the longest match principle for BGP next hop recursion.

   nexthop recursive-lookup longest-match [**route-policy** *route-policy-name*]

   By default, BGP first looks up in its directly connected routes for a recursive route during route recursion.

   This command applies to all BGP routes received from peers in one of the following scenarios:

   - The **route-policy** *route-policy-name* option is not specified.
   - The specified routing policy does not exist.

Enabling conversational remote host route learning

About this task

By default, BGP generates FIB entries for all host routes received from peers (remote host routes) and issues the routes to the FIB. To conserve device resources, you can enable conversational remote host route learning. This feature enables BGP to issue a remote host route to the FIB only if it is required for forwarding a packet.

Restrictions and guidelines

This feature takes effect only on remote host routes generated based on ARP or ND entries and does not take effect on local host routes and other BGP routes. Host routes generated based on ARP or ND entries carry a special mark advertised through the extended community attribute. Enabling this feature also enables BGP to advertise the extended community attribute to all peers and peer groups.

This feature takes effect only if you execute the **ip forwarding-conversational-learning** command.

The device can perform conversational learning for a remote host route only if a local interface belongs to a network segment that contains the destination address of the host route.

Procedure (IPv4 unicast)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP IPv4 unicast address family view.

   address-family ipv4 [**unicast**]

4. Enable conversational remote host route learning.

   forwarding-conversational-learning [**route-policy** *route-policy-name*]

   By default, conversational remote host route learning is disabled.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP IPv6 unicast address family view.

   address-family ipv6 [**unicast**]

4. Enable conversational remote host route learning.

   forwarding-conversational-learning [**route-policy** *route-policy-name*]

   By default, conversational remote host route learning is disabled.

Enabling peer unreachability detection

About this task

Peer unreachability detection achieves fast route convergence by enabling BGP to discover link failures in a timely manner. With this feature enabled, BGP will disconnect from the specified peer or peer group if that peer or peer group is detected unreachable.

BGP considers a peer unreachable if one of the following conditions exists:

- BGP cannot find a route to that peer in the routing table of the public network instance or VPN instance to which the peer belongs.
- BGP finds the route to that peer is a blackhole route in the routing table of the public network instance or VPN instance to which the peer belongs.

As shown in Figure 55, Device A, Device B, and DeviceC are IGP neighbors. Device A and Device C are BGP peers. When a link failure occurs between Devices A and B after peer unreachability detection is enabled on Device A, the IGP immediately performs route convergence. As a result, Device A cannot find the IP route destined for Device C and thus terminates its BGP session with Device C without waiting for the hold timer to expire.

Figure 55 Network diagram



Restrictions and guidelines

If you do not set the delay time of peer disconnection, BGP disconnects from the specified peer or peer group immediately if that peer or peer group is detected unreachable.

When you set the delay time of peer disconnection, follow these guidelines to improve network stability:

- To avoid IBGP session flappings caused by unstable traffic, make sure the delay time of peer disconnection specified for an IBGP peer or peer group is longer than the convergence time of IGP routes.
- With peer unreachability detection enabled, the GR helper can detect whether the GR restarter is reachable. To ensure successful GR, make sure the delay time of peer disconnection is longer than the GR timer plus the extra timer to wait.

For more information about the GR timer and the extra timer to wait, see "Configuring BGP GR."

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

3. Enable peer unreachability detection for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **tracking** [**delay** *delay-time*]

   By default, peer unreachability detection is disabled.

   This command does not take effect if you enter a link-local address for the *ipv6-address* argument.

Displaying 4-byte AS numbers in asdot or asdot+ notation

About this task

NOTE:

In the current software version, the following AS number notations are supported:

- **asplain**—Represents AS numbers with decimal integers, for example, 65535. AS numbers in this notation are called integer AS numbers.
- **asdot+**—Represents each AS number with two decimal integers separated by a period character: *<higher 16-bit value in decimal>.<lower 16-bit value in decimal>*, for example, 65535.65535. AS numbers in this notation are called dotted AS numbers.
- **asdot**—Represents AS numbers less than 65536 with the asplain notation and represents AS numbers equal to or greater than 65536 with the asdot+ notation.

By default, a BGP display command uses the asplain notation to represent 4-byte AS numbers. A 4-byte AS number in asplain notation tends to be long and reader-unfriendly. To resolve this issue, use this feature to display AS numbers in asdot or asdot+ notation.

The asdot+ notation refers to a method of representing AS numbers in **X.Y** format. The value range for argument *Y* is 0 to 65535, and the value range for a dotted AS number is 0.1 to 65535.65535. To convert a dotted AS number into an integer one, use the following formula: *integer* = *X x 65536* + *Y*. For example, if a 4-byte dotted AS number is 2.1, the corresponding integer AS number is 131073, which equals 2 x 65536 + 1.

Restrictions and guidelines

After you configure this feature, the AS path list can match only 4-byte dotted AS numbers. To avoid route filtering failure, make sure the AS path list used in the route filtering policy referenced by BGP, OSPF, IS-IS, RIP, OSPFv3, or RIPng is configured to match 4-byte dotted AS numbers.

This feature does not take effect on the 4-byte AS numbers in the system configuration file.

Before using this feature, configure an AS path list that matches 4-byte dotted AS numbers, avoiding route filtering failure.

Procedure

1. Enter system view.

   system-view

2. Display 4-byte AS numbers in asdot or asdot+ notation.

   as-notation dotted [**4-octet-only**]

   By default, 4-byte AS numbers are displayed in asplain notation.

Converting AS numbers in effective device configurations

About this task

Before upgrading a device to a version that supports 4-byte AS numbers, use the **as-notation convert asdot** command to convert all AS numbers (greater than 65535) in effective device configurations into dotted AS numbers. This operation improves configuration readability.

Before downgrading a device, use the **as-notation convert asplain** command to convert all AS numbers in effective device configurations into integer AS numbers. This operation ensures successful device downgrade.

Restrictions and guidelines

The **as-notation convert** command is an operational command and takes effect only on existing configurations. This command does not convert the AS numbers configured after it is executed. You can repeat this command to convert the AS numbers in effective device configurations multiple times as needed.

As a best practice, save the configurations after you convert the AS numbers in effective device configurations.

Procedure

1. Enter system view.

   system-view

2. Convert the AS numbers in effective device configurations between asplain notation and asdot notation.

   as-notation convert { **asdot** | **asplain** }

   By default, the format of an AS number in effective device configurations is the same as the format in which the AS number is configured.

Configuring the tag community attribute

About this task

When routes are redistributed from a routing protocol into an IGP, the IGP can add a tag for them. This tag can be used for route filtering methods such as routing policies. Routing information is distributed across ASs after it is redistributed between IGPs and BGP. As a result, different route filtering methods might be applied to an IGP route when it exists in different ASs. To remove the impact of route location on the selection of route filtering methods, you can enable BGP to transmit the tag of IGP routes. After IGP routes are distributed to other ASs, IGP routes with the same tag will match the same route filtering method, regardless of their locations.

The tag community attribute is an H3C-proprietary extended community attribute, which is designed for transmitting IGP route tags.

NOTE:

For brief introduction, a BGP route carrying the tag community attribute is called tagged route. When BGP tags a route, it adds the tag community attribute to that route. When BGP untags a route, it removes the tag community attribute from that route.

The **peer advertise tag** command functions as follows:

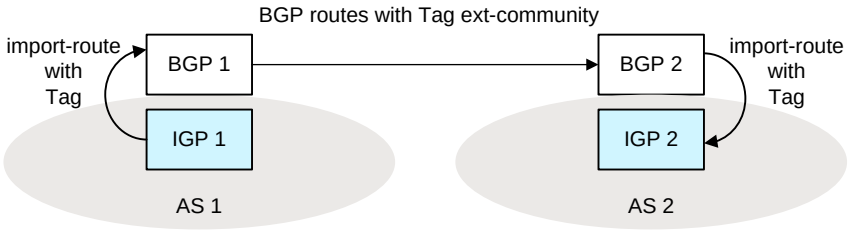
- Before advertising a BGP route generated from a redistributed IGP route to the specified peer or peer group, BGP tags that route. The value of the tag community attribute depends on the configuration of the **import-route** command.
- After receiving a tagged BGP route, BGP forwards that route to the specified peer or peer group without untagging.

If the **peer advertise tag** commands is not configured or the **undo peer advertise tag** command is configured, BGP untags BGP routes (if tagged) before advertising them to peers. In this situation, BGP does not support the following operations:

- Tags BGP routes before route advertisement.
- Advertises BGP routes without untagging.

Route tag information is distributed across ASs as shown in Figure 56.

Figure 56 Distributing route tag information



1. In the source AS, the network administrator redistributes IGP routes to BGP and enables BGP to inherit IGP route tags. To have BGP inherit IGP route tags, use one of the following methods:
   - Execute the **import-route** command with the **inherit-tag** keyword specified. BGP will directly inherit the original tags of redistributed IGP routes.
   - Execute the **import-route** command with the **inherit-tag** keyword and the **route-policy** *route-policy-name* option specified, and configure an **apply tag** clause for the specified routing policy. BGP will tag the BGP routes generated from redistributed IGP routes according to the configuration of that **apply tag** clause.
2. The inherited tag information forms a tag community attribute, which is inserted into BGP routes and transmitted to another AS.
3. The network administrator of the receiver AS redistributes the tagged BGP routes into the desired IGP. The generated IGP routes can inherit those BGP routes' tags. You can use multiple methods to set the tag of IGP routes derived from redistributed BGP routes. For an IGP route, the priorities of tags configured by using those methods are as follows:
   a. Tag configured in the **apply tag** clause of the specified routing policy.
   b. Tag specified by using the **tag** keyword in the **import-route** command configured for the IGP.
   c. Tag inherited from the tag community attribute of the redistributed BGP route.
   d. Tag specified by using the tag keyword in the **route-tag** command.

Procedure (IPv4 unicast)

1. Enter System view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. (Optional.) Configure the type value for the tag community attribute.
   extcommunity-type tag *tag-type-value*

By default, the type value is 83cf in hexadecimal notation.

4. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Enter BGP IPv4 unicast address family view.

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
5. Redistribute IGP routes into the BGP routing table.

   import-route { **isis** | **ospf** | **rip** } [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *] [**inherit-tag**]

   By default, IGP routes are not redistributed into BGP.
6. Enable BGP to advertise the tag community attribute to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise tag**

   By default, BGP does not advertise the tag community attribute to a peer or peer group.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. (Optional.) Configure the type value for the tag community attribute.

   extcommunity-type tag *tag-type-value*

   By default, the type value is 83cf in hexadecimal notation.
4. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Enter BGP IPv6 unicast address family view.

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]
5. Redistribute IGP routes into the BGP routing table.

   import-route { **isisv6** | **ospfv3** | **ripng** } [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *] [**inherit-tag**]

   By default, IGP routes are not redistributed into BGP.
6. Enable BGP to advertise the tag community attribute to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise tag**

   By default, BGP does not advertise the tag community attribute to a peer or peer group.

Display and maintenance commands for BGP network tuning and optimization

Displaying BGP

Execute **display** commands in any view.

Displaying BGP (IPv4 unicast address family)

Task	Command
Display ORF prefix information received by a peer.	**display bgp** [**instance** *instance-name*] **peer ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] *ipv4-address* **received prefix-list** **display bgp** [**instance** *instance-name*] **peer ipv4** [**unicast**] *ipv6-address* [**interface** *interface-type interface-number*] **received prefix-list**
Display peer monitoring group information.	**display bgp** [**instance** *instance-name*] **monitor-group** { **all** | **group-name** *monitor-group-name* }

Displaying BGP (IPv6 unicast address family)

Task	Command
Display ORF prefix information received by a peer.	**display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] *ipv6-address* **received prefix-list** **display bgp** [**instance** *instance-name*] **peer ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] { *ipv4-address* | [*link-local-address*] **interface** *interface-type interface-number* } **received prefix-list**

Displaying BGP (IPv4 multicast address family)

Task	Command
Display ORF prefix information received by a peer.	**display bgp** [**instance** *instance-name*] **peer ipv4 multicast** *ipv4-address* **received prefix-list**

Displaying BGP (IPv6 multicast address family)

Task	Command
Display ORF prefix information received by a peer.	**display bgp** [**instance** *instance-name*] **peer ipv6 multicast** *ipv6-address* **received prefix-list**

Resetting BGP sessions

CAUTION:
A reset operation tears down BGP sessions for a short period of time.

Execute **reset** commands in user view.

Task	Command
Reset BGP sessions for IPv4 unicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] **reset bgp** [**instance** *instance-name*] { *ipv6-address* [*prefix-length*] | **interface** *interface-type interface-number* } **ipv4** [*unicast*]
Reset BGP sessions for IPv4 multicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **ipv4 multicast**
Reset BGP sessions for IPv6 unicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv6-address* [*prefix-length*] | [*link-local-address*] **interface** *interface-type interface-number* | **all** | **external** | **group** *group-name* | **internal** } **ipv6** [**unicast**] [**vpn-instance** *vpn-instance-name*] **reset bgp** *ipv4-address* [*mask-length*] **ipv6** [**unicast**]
Reset BGP sessions for IPv6 multicast address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } **ipv6 multicast**
Reset all BGP sessions.	**reset bgp** [**instance** *instance-name*] **all**

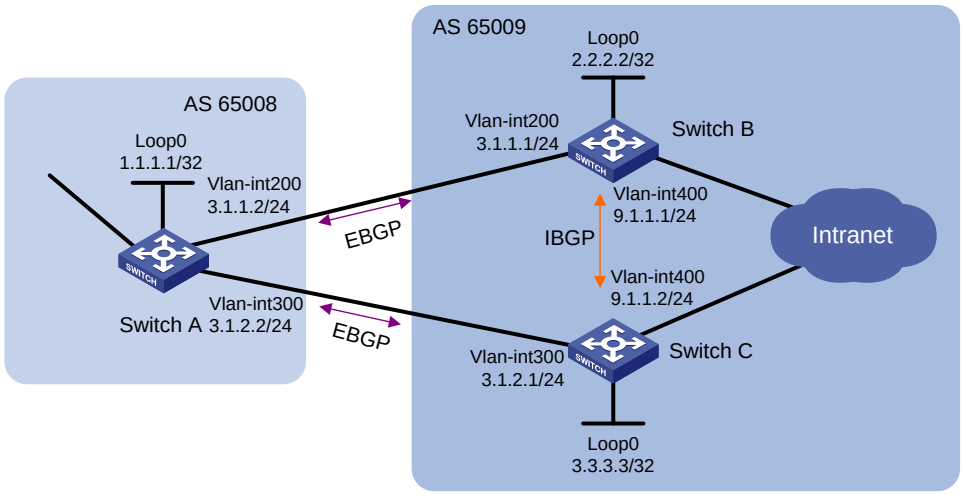
BGP network tuning and optimization configuration examples

Example: Configuring BGP load balancing

Network configuration

As shown in Figure 57, run EBGP between Switch A and Switch B, and between Switch A and Switch C. Run IBGP between Switch B and Switch C.

Figure 57 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP connections:
   - On Switch A, establish EBGP connections with Switch B and Switch C. Configure BGP to advertise network 8.1.1.0/24 to Switch B and Switch C, so that Switch B and Switch C can access the internal network connected to Switch A. Configure load balancing over the two EBGP links on Switch A.
   - On Switch B, establish an EBGP connection with Switch A and an IBGP connection with Switch C. Configure BGP to advertise network 9.1.1.0/24 to Switch A, so that Switch A can access the intranet through Switch B. Configure a static route to interface loopback 0 on Switch C (or use a routing protocol like OSPF) to establish the IBGP connection.
   - On Switch C, establish an EBGP connection with Switch A and an IBGP connection with Switch B. Configure BGP to advertise network 9.1.1.0/24 to Switch A, so that Switch A can access the intranet through Switch C. Configure a static route to interface loopback 0 on Switch B (or use another protocol like OSPF) to establish the IBGP connection.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 3.1.1.1 as-number 65009
[SwitchA-bgp-default] peer 3.1.2.1 as-number 65009
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 3.1.1.1 enable
[SwitchA-bgp-default-ipv4] peer 3.1.2.1 enable
[SwitchA-bgp-default-ipv4] network 8.1.1.0 24
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 3.1.1.2 as-number 65008
[SwitchB-bgp-default] peer 3.3.3.3 as-number 65009
```

```
[SwitchB-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 3.1.1.2 enable
[SwitchB-bgp-default-ipv4] peer 3.3.3.3 enable
[SwitchB-bgp-default-ipv4] network 9.1.1.0 24
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
[SwitchB] ip route-static 3.3.3.3 32 9.1.1.2
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 3.1.2.2 as-number 65008
[SwitchC-bgp-default] peer 2.2.2.2 as-number 65009
[SwitchC-bgp-default] peer 2.2.2.2 connect-interface loopback 0
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 3.1.2.2 enable
[SwitchC-bgp-default-ipv4] peer 2.2.2.2 enable
[SwitchC-bgp-default-ipv4] network 9.1.1.0 24
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
[SwitchC] ip route-static 2.2.2.2 32 9.1.1.1
```

Display the BGP routing table on Switch A.

```
[SwitchA] display bgp routing-table ipv4

 Total number of routes: 3

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >  8.1.1.0/24         8.1.1.1         0                     32768   i
* >e 9.1.1.0/24         3.1.1.1         0                     0       65009i
*  e                    3.1.2.1         0                     0       65009i
```

- The output shows two valid routes to destination 9.1.1.0/24. The route with next hop 3.1.1.1 is marked with a greater-than sign (>), indicating that it is the optimal route (because the ID of Switch B is smaller). The route with next hop 3.1.2.1 is marked with an asterisk (*), indicating that it is a valid route, but not the optimal route.
- By using the **`display ip routing-table`** command, you can find only one route to 9.1.1.0/24 with next hop 3.1.1.1 and output interface VLAN-interface 200.

3. Configure loading balancing:

Because Switch A has two routes to reach AS 65009, configuring load balancing over the two BGP routes on Switch A can improve link usage.

Configure Switch A.

```
[SwitchA] bgp 65008
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] balance 2
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Verifying the configuration

Display the BGP routing table on Switch A.

```
[SwitchA] display bgp routing-table ipv4

 Total number of routes: 3

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >  8.1.1.0/24         8.1.1.1         0                     32768   i
* >e 9.1.1.0/24         3.1.1.1         0                     0       65009i
* >e                    3.1.2.1         0                     0       65009i
```

- The route 9.1.1.0/24 has two next hops, 3.1.1.1 and 3.1.2.1, both of which are marked with a greater-than sign (>), indicating that they are the optimal routes.
- By using the **`display ip routing-table`** command, you can find two routes to 9.1.1.0/24. One has next hop 3.1.1.1 and output interface VLAN-interface 200, and the other has next hop 3.1.2.1 and output interface VLAN-interface 300.

Example: Configuring the BGP Additional Paths feature

Network configuration

As shown in Figure 58, all switches run BGP. EBGP runs between Switch A and Switch B, and between Switch A and Switch C. IBGP runs between Switch B and Switch D, between Switch C and Switch D, and between Switch D and Switch E. Switch D is a route reflector and Switch E is its client.

Configure the BGP Additional Paths feature to enable Switch E to learn routes with the same prefix and different next hops from Switch B and Switch C.

Figure 58 Network diagram

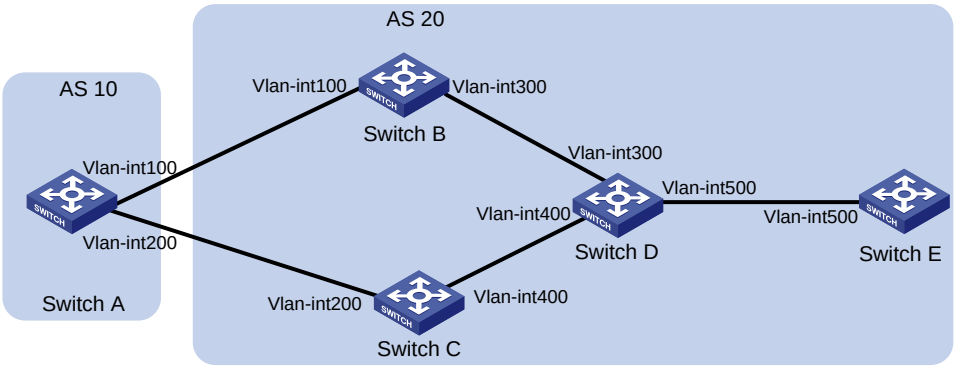


Table 15 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.1.1.1/24	Switch D	Vlan-int300	30.1.1.1/24
	Vlan-int200	20.1.1.1/24		Vlan-int400	40.1.1.1/24
Switch B	Vlan-int100	10.1.1.2/24		Vlan-int500	50.1.1.1/24
	Vlan-int300	30.1.1.2/24	Switch E	Vlan-int500	50.1.1.2/24
Switch C	Vlan-int200	20.1.1.2/24			
	Vlan-int400	40.1.1.2/24			

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP connections:

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 10
[SwitchA-bgp-default] peer 10.1.1.2 as-number 20
[SwitchA-bgp-default] peer 20.1.1.2 as-number 20
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
[SwitchA-bgp-default-ipv4] peer 20.1.1.2 enable
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 20
[SwitchB-bgp-default] peer 10.1.1.1 as-number 10
[SwitchB-bgp-default] peer 30.1.1.1 as-number 20
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
[SwitchB-bgp-default-ipv4] peer 30.1.1.1 enable
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 20
[SwitchC-bgp-default] peer 20.1.1.1 as-number 10
[SwitchC-bgp-default] peer 40.1.1.1 as-number 20
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 20.1.1.1 enable
[SwitchC-bgp-default-ipv4] peer 40.1.1.1 enable
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 20
[SwitchD-bgp-default] peer 30.1.1.2 as-number 20
[SwitchD-bgp-default] peer 40.1.1.2 as-number 20
[SwitchD-bgp-default] peer 50.1.1.2 as-number 20
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] peer 30.1.1.2 enable
[SwitchD-bgp-default-ipv4] peer 40.1.1.2 enable
[SwitchD-bgp-default-ipv4] peer 50.1.1.2 enable
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] bgp 20
[SwitchE-bgp-default] peer 50.1.1.1 as-number 20
[SwitchE-bgp-default] address-family ipv4 unicast
[SwitchE-bgp-default-ipv4] peer 50.1.1.1 enable
```

3. Configure Switch A to advertise network 10.1.1.0/24.

```
[SwitchA-bgp-default-ipv4] network 10.1.1.0 24
```

4. Set the local switch as the next hop for routes sent to a peer:

Configure Switch B.

```
[SwitchB-bgp-default-ipv4] peer 30.1.1.1 next-hop-local
```

Configure Switch C.

```
[SwitchC-bgp-default-ipv4] peer 40.1.1.1 next-hop-local
```

5. Configure Switch D as a route reflector.

```
[SwitchD-bgp-default-ipv4] peer 50.1.1.2 reflect-client
```

6. Configure the Additional Paths feature:

Enable the additional path sending capability on Switch D.

```
[SwitchD-bgp-default-ipv4] peer 50.1.1.2 additional-paths send
```

Set the maximum number to 2 for Add-Path optimal routes that can be advertised.

```
[SwitchD-bgp-default-ipv4] additional-paths select-best 2
```

Set the maximum number to 2 for Add-Path optimal routes that can be advertised to peer 50.1.1.2.

```
[SwitchD-bgp-default-ipv4] peer 50.1.1.2 advertise additional-paths best 2
```

Enable the additional path receiving capability on Switch E.

```
[SwitchE-bgp-default-ipv4] peer 50.1.1.1 additional-paths receive
```

Verifying the configuration

Display BGP routing information on Switch E.

```
[Switch E] display bgp routing-table ipv4

 Total number of routes: 2

 BGP local Switch ID is 50.1.1.2
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

   i 10.1.1.0/24        30.1.1.2        0          100        0       10i
   i                    40.1.1.2        0          100        0       10i
```

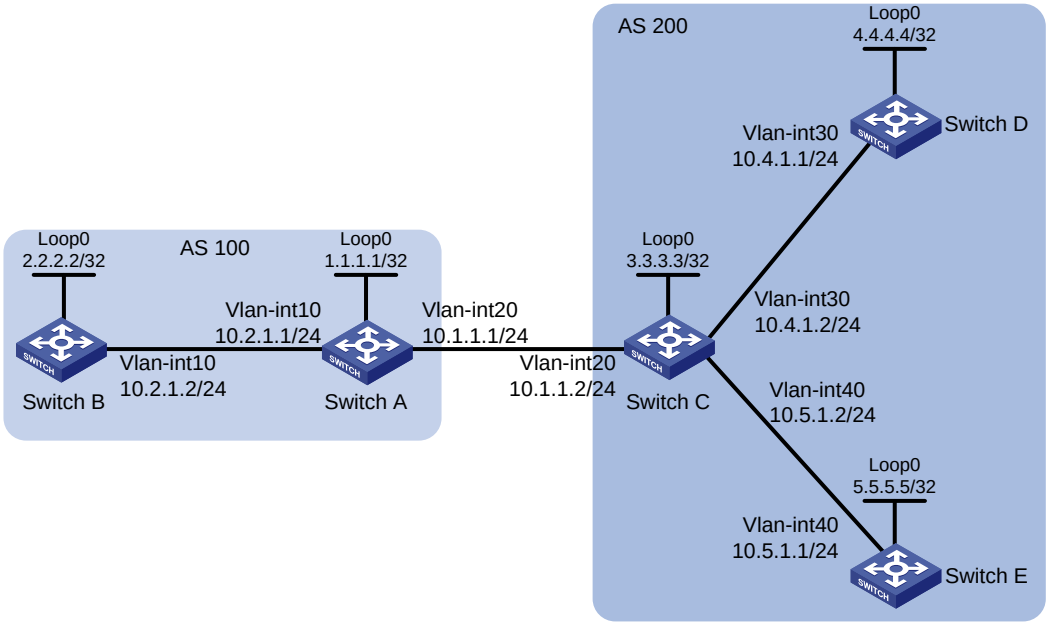
The output shows that Switch D has learned two routes with the same prefix and different next hops.

Example: Configuring prefix-based BGP ORF

Network configuration

As shown in Figure 59, AS 100 and AS 200 can exchange BGP routes. To ensure network security and optimize network performance, devices in AS 100 and AS 200 need to filter routes before route advertisement. The routes required by Switch A frequently change. Configure prefix-based BGP ORF on both Switch A and Switch C, so Switch C sends only routes matching Switch A's route reception policy to Switch A.

Figure 59 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP connections:
   - On Switch A, establish an IBGP connection with Switch B and an EBGP connection with Switch C.
   - On Switch C, establish an IBGP connection with Switch D and Switch E separately.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 10.1.1.2 as-number 200
[SwitchA-bgp-default] peer 10.2.1.2 as-number 100
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
[SwitchA-bgp-default-ipv4] peer 10.2.1.2 enable
[SwitchA-bgp-default-ipv4] import-route direct
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
<SwitchB> system-view
```

```
[SwitchB] bgp 100
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 10.2.1.1 as-number 100
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 10.2.1.1 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 10.1.1.1 as-number 100
[SwitchC-bgp-default] peer 10.4.1.1 as-number 200
[SwitchC-bgp-default] peer 10.5.1.1 as-number 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 10.1.1.1 enable
[SwitchC-bgp-default-ipv4] peer 10.4.1.1 enable
[SwitchC-bgp-default-ipv4] peer 10.5.1.1 enable
[SwitchC-bgp-default-ipv4] import-route direct
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] peer 10.4.1.2 as-number 200
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] peer 10.4.1.2 enable
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] bgp 200
[SwitchE-bgp-default] router-id 5.5.5.5
[SwitchE-bgp-default] peer 10.5.1.2 as-number 200
[SwitchE-bgp-default] address-family ipv4 unicast
[SwitchE-bgp-default-ipv4] peer 10.5.1.2 enable
[SwitchE-bgp-default-ipv4] quit
[SwitchE-bgp-default] quit
```

3. On Switch A, reference an address prefix list to filter BGP routes received from Switch C.

Configure Switch A.

```
[SwitchA] ip refix-list policy1 permit 10.4.1.0 24 less-equal 32
[SwitchA] bgp 100
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 prefix-list policy1 import
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

On Switch C, view the routes advertised to Switch A.

```
[SwitchC] display bgp routing-table ipv4 unicast peer 10.1.1.1 advertised-routes

 Total number of routes: 4

 BGP local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf             Path/Ogn

* >  3.3.3.3/32         127.0.0.1       0                             200?
* >  10.1.1.0/24        10.1.1.2        0                             200?
* >  10.4.1.0/24        10.4.1.2        0                             200?
* >  10.5.1.0/24        10.5.1.2        0                             200?
```

On Switch A, view the routes received from Switch C.

```
[SwitchA] display bgp routing-table ipv4 unicast peer 10.1.1.2 received-routes

 Total number of routes: 1

 BGP local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 10.4.1.0/24        10.1.1.2        0                     0       ?
```

According to the above command outputs, Switch C has advertised four routes to Switch A, but Switch A only received one of them under the restriction of its BGP route reception policy.

4. Configure prefix-based BGP ORF.

On Switch A, enable prefix-based BGP ORF capabilities negotiation for Switch C.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 capability-advertise orf prefix-list both
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

On Switch C, enable prefix-based BGP ORF capabilities negotiation for Switch A.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 10.1.1.1 capability-advertise orf prefix-list both
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Verifying the configuration

On Switch C, view address prefix list information in the ORF information received from Switch A.

```
[SwitchC] display bgp peer ipv4 10.1.1.1 received prefix-list

ORF prefix list entries: 2
ge: greater-equal   le: less-equal
  Index     Rule    Prefix                                  Ge    Le
  10        permit  10.4.1.0/24                             0     32
```

On Switch C, view the routes advertised to Switch A.

```
[SwitchC] display bgp routing-table ipv4 unicast peer 10.1.1.1 advertised-routes

 Total number of routes: 1

 BGP local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf             Path/Ogn

* >  10.4.1.0/24        10.4.1.2        0                             200?
```

According to the above command outputs, Switch C only advertised the route matching Switch A's route reception policy to Switch A.

Configuring BGP security features

BGP security feature configuration tasks at a glance

To configure BGP security features, perform the following tasks:

- Enabling MD5 authentication for BGP peers
- Configuring GTSM for BGP
- Enabling keychain authentication for BGP peers
- Configuring IPsec for IPv6 BGP
- Configuring BGP RPKI

Enabling MD5 authentication for BGP peers

About this task

MD5 authentication provides the following benefits:

- Peer authentication ensures that only BGP peers that have the same password can establish TCP connections.
- Integrity check ensures that BGP packets exchanged between peers are intact.

Procedure (IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enable MD5 authentication for a BGP peer group or peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **password** { **cipher** | **simple** } *password*
   By default, MD5 authentication is disabled.

Procedure (IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*

3. Enable MD5 authentication for a BGP peer group or peer. Choose one command as needed:

   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **password** { **cipher** | **simple** } *password*

   interface-peer *interface-type interface-number* **password** { **cipher** | **simple** } *password*

   By default, MD5 authentication is disabled.

Configuring GTSM for BGP

About this task

The Generalized TTL Security Mechanism (GTSM) protects a BGP session by comparing the TTL value in the IP header of incoming BGP packets against a valid TTL range. If the TTL value is within the valid TTL range, the packet is accepted. If not, the packet is discarded.

The valid TTL range is from 255 – the configured hop count + 1 to 255.

When GTSM is configured, the BGP packets sent by the device have a TTL of 255.

GTSM provides best protection for directly connected EBGP sessions, but not for multihop EBGP or IBGP sessions because the TTL of packets might be modified by intermediate devices.

Restrictions and guidelines

When GTSM is configured, the local device can establish an EBGP session to the peer after both devices pass GTSM check, regardless of whether the maximum number of hops is reached.

To use GTSM, you must configure GTSM on both the local and peer devices. You can specify different *hop-count* values for them.

Procedure (IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Configure GTSM for the specified BGP peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **ttl-security hops** *hop-count*

   By default, GTSM is disabled.

Procedure (IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

3. Configure GTSM for the specified BGP peer or peer group.

   peer { *group-name* | *ipv6-address* [*prefix-length*] } **ttl-security hops** *hop-count*

   By default, GTSM is disabled.

Enabling keychain authentication for BGP peers

About this task

Keychain authentication enhances the security of TCP connection establishment between BGP peers. It allows BGP peers to establish TCP connections only when the following conditions are met:

- Keychain authentication is enabled on both BGP peers.
- The keys used by the BGP peers at the same time must have the same ID.
- The keys with the same ID must use the same authentication algorithm and key string.

Before configuring keychain authentication, make sure the specified keychain has been created.

For more information about keychains, see *Security Configuration Guide*.

Restrictions and guidelines

Follow these restrictions and guidelines when you configure the algorithm and key ID in keychain authentication:

- BGP supports the HMAC-MD5, MD5, HMAC-SM3, SM3, and HMAC-SHA-256 algorithms. To configure an algorithm, execute the **authentication-algorithm** command.
- BGP supports key IDs in the range of 0 to 63. To configure a key ID, execute the **key** command.

Procedure (IPv4 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Enable keychain authentication for a BGP peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **keychain** *keychain-name*

   By default, keychain authentication is disabled.

Procedure (IPv6 peers)

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

bgp *as-number* [**instance** *instance-name*]

- Enter BGP-VPN instance view.

  bgp *as-number* [**instance** *instance-name*]

  ip vpn-instance *vpn-instance-name*

3. Enable keychain authentication for a BGP peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **keychain** *keychain-name*

   interface-peer *interface-type interface-number* **keychain** *keychain-name*

   By default, keychain authentication is disabled.

Configuring IPsec for IPv6 BGP

About this task

Perform this task to configure IPsec for IPv6 BGP. IPsec can provide privacy, integrity, and authentication for IPv6 BGP packets exchanged between BGP peers.

When two IPv6 BGP peers are configured with IPsec (for example, Device A and Device B), Device A encapsulates an IPv6 BGP packet with IPsec before sending it to Device B. If Device B successfully receives and de-encapsulates the packet, it establishes an IPv6 BGP peer relationship with Device A and learns IPv6 BGP routes from Device A. If Device B receives but fails to de-encapsulate the packet, or receives a packet not protected by IPsec, it discards the packet.

Procedure

1. Enter system view.

   system-view

2. Configure an IPsec transform set and a manual IPsec profile.

   See *Security Configuration Guide*.

3. Enter BGP instance view or BGP-VPN instance view.

   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]

   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

4. Apply the IPsec profile to an IPv6 BGP peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **ipsec-profile** *profile-name*

   interface-peer *interface-type interface-number* **ipsec-profile** *profile-name*

   By default, no IPsec profile is configured for any IPv6 BGP peer or peer group.

   This command supports only IPsec profiles in manual mode.

Configuring BGP RPKI

About BGP RPKI

About RPKI

The AS_PATH attribute identifies the ASs through which a route has passed, and the AS that originated the route is the origin AS of the route. If the origin AS number of a route is incorrect, traffic transmission failure or even network collapse might occur.

To avoid this problem, you can configure the BGP Resource Public Key Infrastructure (RPKI) feature. It enables BGP to validate the origin AS of a route and determine whether to use and advertise the route based on the validation state.

BGP RPKI can address the issue of traffic hijacking. For example, an ISP advertises a legitimate route with a prefix of 192.168.0.0/16 to its users, and an attacker advertises a route with a prefix of 192.168.10.0/24. After receiving the two routes, a user prefers route 192.168.10.0/24 during route selection, because this route is more specific. As a result, the user might send traffic to the attacker mistakenly. To avoid this issue, configure BGP RPKI.

About ROAs

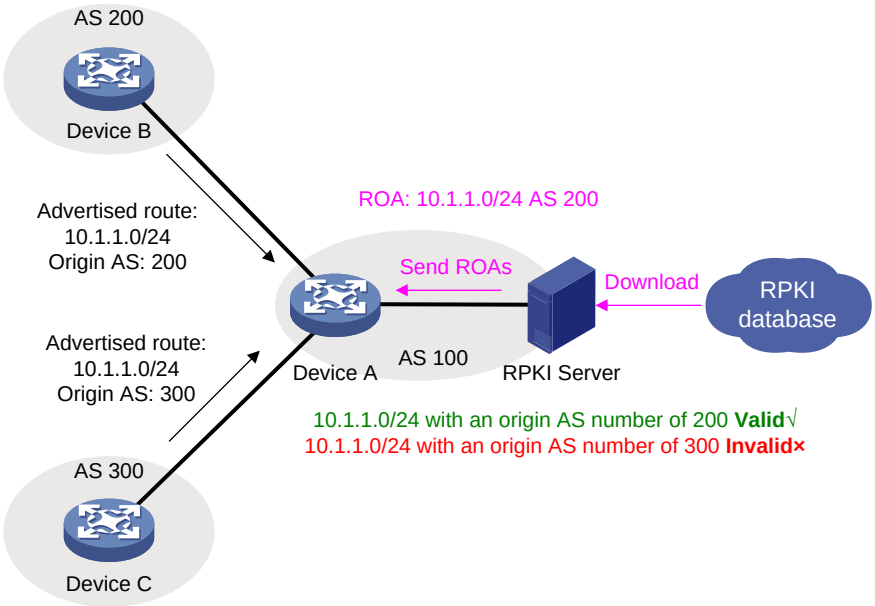
Route Origin Authorization (ROA) is a digital certificate that binds a route prefix to an origin AS number. With ROAs, the device can verify the validity of a route based on its origin AS number.

As shown in Figure 60, Device A and the RPKI server are connected. The RPKI server downloads ROA information from the RPKI database and sends the downloaded ROAs to Device A. Device A then stores the received ROAs in its local ROA database. Upon receiving a route, Device A searches the local ROA database for an ROA that matches the route prefix. If such an ROA exists, Device A compares the origin AS number of the route with that of the matching ROA.

- If the two AS numbers are the same, the route is valid.
- If the two AS numbers are different, the route is invalid.

In this example, Device A determines that route 10.1.1.0/24 from Device B is valid and route 10.1.1.0/24 from Device C is invalid.

Figure 60 Route validation with ROAs



Configuring RPKI connection parameters

About this task

A router establishes a TCP connection with an RPKI server to obtain the Route Origin Authorization (ROA) information used for RPKI validation. The router checks the connection to the RPKI server at the specified interval. If the router does not receive a response from the RPKI server within the specified time period, it tears down the connection to the RPKI server.

When the connection between a router and an RPKI server goes down (except when the **shutdown** command is executed), the router takes the following actions:

- Attempts to reconnect to the server.
- Places the ROA information obtained from the server in aging state, and starts the aging timer for the ROA information.

If the router reconnects to the server before the aging timer expires, it releases the ROA information from the aging state. If the router fails to reconnect to the server when the aging timer expires, it deletes the ROA information obtained from the server.

Restrictions and guidelines

Follow these restrictions and guidelines when you configure RPKI connection parameters:

- As a best practice, set an ROA information aging time longer than the time to wait for the response from the RPKI server.
- To tear down the connection to an RPKI server, execute the **undo port** command in RPKI server view.
- If you execute the **undo rpki** command, all configurations in RPKI view are removed.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP RPKI view.

   rpki

4. Specify an RPKI server by its IP address and enter RPKI server view.

   server [**vpn-instance** *vpn-instance-name*] **tcp** { *ipv4-address* | *ipv6-address* }

   By default, no RPKI server is specified.

5. Specify the port number of the RPKI server.

   port *port-number*

   By default, the port number of the RPKI server is not specified.

   To enable a router to establish a TCP connection with the RPKI server, you must configure the port number of the RPKI server on the router.

6. (Optional.) Specify the MD5 authentication password.

   passwords { **cipher** | **simple** } *string*

   By default, the RPKI server does not perform MD5 authentication.

   The MD5 authentication password must be the same as the authentication password configured on the RPKI server.

   MD5 authentication ensures the validity of the RPKI server and the security of BGP RPKI packets.

7. (Optional.) Set the RPKI connection check interval.
   refresh-time *refresh-time*
   By default, the RPKI connection check interval is 600 seconds.
8. (Optional.) Set the time to wait for the response from the RPKI server.
   response-time *response-time*
   By default, the time to wait for the response from the RPKI server is 30 seconds.
9. (Optional.) Set the aging time for the ROA information.
   purge-time *purge-time*
   By default, the aging time for the ROA information is 60 seconds.
10. (Optional.) Set the maximum number of ROAs that can be received from the RPKI server.
    rpki-limit *limit* [**alert-only** | **discard** | **reconnect** *reconnect-time*]
    By default, the number of ROAs that can be received from an RPKI server is not limited.

Enabling BGP RPKI validation

About this task

After you configure this feature, BGP validates the prefix and origin AS number of a received route. BGP determines the validation state of the route based on the following requirements:

1. A minimum of one ROA in the ROA database contains the prefix.
2. The mask length of the route does not exceed the longest mask length among all ROAs in the ROA database.
3. The AS number of an ROA in the ROA database is the same as the origin AS number of the route.

If the route meets all requirements, the route is in **Valid** state. If the route cannot meet requirement 1, the route is in **Not-found** state. If the route meets requirement 1, but cannot meet requirement 2 or 3, the route is in **Invalid** state.

Restrictions and guidelines

If you configure this feature, BGP uses the local RPKI validation states. If you do not configure this feature, BGP uses the validation states in the received BGP routes.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP RPKI view.
   rpki
4. Enable BGP RPKI validation.
   check-origin-validation
   By default, BGP RPKI validation is disabled.

Applying the BGP RPKI validation state to optimal route selection

About this task

If multiple routes to the same destination are available, BGP first discards routes with unreachable next hops, and then selects the optimal route according to the following rules:

- Routes with a BGP RPKI validation state of **Valid** takes precedence over routes with a validation state of **Not-found** or **Invalid**.
- Routes with a BGP RPKI validation state of **Not-found** takes precedence over routes with a validation state of **Invalid**.
- Routes without a BGP RPKI validation state have the same priority as routes with a BGP RPKI validation state of **Not-found**.
- For routes that have the same BGP RPKI validation state, BGP selects the optimal route according to the rules in "BGP route selection."

You can configure a routing policy to filter routes based on the BGP RPKI validation state. For more information about routing policies, see "Configuring routing policies."

Procedure (IPv4 unicast)

1. Enter system view.

   system-view
2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
3. Apply the BGP RPKI validation state to optimal route selection.

   bestroute origin-as-validation [**allow-invalid**]

   By default, BGP ignores the BGP RPKI validation state during optimal route selection.

   To allow routes with a validation state of **Invalid** to participate in optimal route selection, you must specify the **allow-invalid** keyword.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view
2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   - Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

address-family ipv6 [**unicast**]

3. Apply the BGP RPKI validation state to optimal route selection.

   bestroute origin-as-validation [**allow-invalid**]

   By default, BGP ignores the BGP RPKI validation state during optimal route selection.

   To allow routes with a validation state of **Invalid** to participate in optimal route selection, you must specify the **allow-invalid** keyword.

Advertising BGP RPKI validation state to a peer or peer group

Restrictions and guidelines

BGP advertises the BGP RPKI validation state to a peer or peer group through the extended community attribute. To enable this feature, you must first enable BGP to advertise the extended community attribute to the peer or peer group and make sure RPKI settings are correct.

In the current software version, BGP can advertise the BGP RPKI validation state only to IBGP peers and peer groups.

Procedure (IPv4 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.

   - Execute the following commands in sequence to enter BGP IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family ipv4 [**unicast**]

   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

3. Advertise the extended community attribute to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise-ext-community**

   By default, BGP does not advertise the extended community attribute.

4. Advertise the BGP RPKI validation state to the specified peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise origin-as-validation**

   interface-peer *interface-type interface-number* **advertise origin-as-validation**

   By default, BGP does not advertise the BGP RPKI validation state.

Procedure (IPv6 unicast)

1. Enter system view.

   system-view

2. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.

- Execute the following commands in sequence to enter BGP IPv6 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  address-family ipv6 [**unicast**]

- Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view:

  bgp *as-number* [**instance** *instance-name*]

  ip vpn-instance *vpn-instance-name*

  address-family ipv6 [**unicast**]

3. Advertise the extended community attribute to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **advertise-ext-community**

   By default, BGP does not advertise the extended community attribute.

4. Advertise the BGP RPKI validation state to the specified peer or peer group. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **advertise origin-as-validation**

   interface-peer *interface-type interface-number* **advertise origin-as-validation**

   By default, BGP does not advertise the BGP RPKI validation state.

Resetting BGP RPKI sessions

Execute the following command in user view.

reset bgp [**instance** *instance-name*] **rpki server** [**vpn-instance** *vpn-instance-name*] **tcp** { *ipv4 address* | *ipv6 address* }

CAUTION:

After you execute this command, the device will delete and re-establish the specified BGP RPKI session and cause temporary session interruption.

Display and maintenance commands for BGP security features

Execute **display** commands in any view.

Displaying BGP (IPv4 unicast address family)

Step	Command
Display information about connections to RPKI servers.	**display bgp** [**instance** *instance-name*] **rpki server** [[**vpn-instance** *vpn-instance-name*] *ipv4-address*]
Display the ROA information obtained from RPKI servers.	**display bgp** [**instance** *instance-name*] **rpki table ipv4** [*ipv4-address* **min** *min-length* **max** *max-length*]

Displaying BGP (IPv6 unicast address family)

Step	Command
Display information about connections to RPKI servers.	**display bgp** [**instance** *instance-name*] **rpki server** [[**vpn-instance** *vpn-instance-name*] *ipv6-address*]
Display the ROA information obtained from RPKI servers.	**display bgp** [**instance** *instance-name*] **rpki table ipv6** [*ipv6-address* **min** *min-length* **max** *max-length*

IPv4 BGP security feature configuration examples

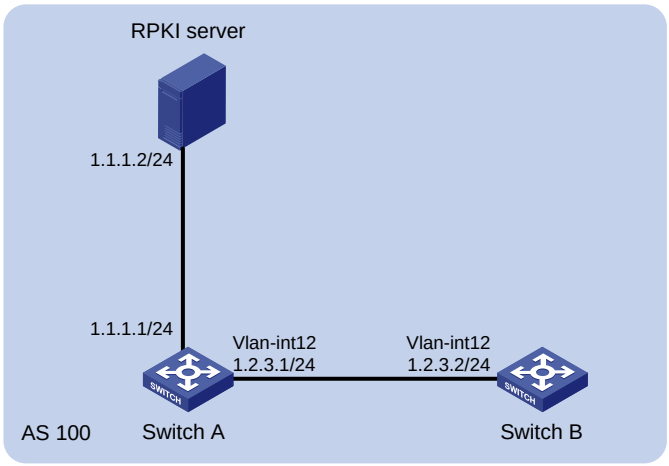
Example: Configuring BGP RPKI

Network configuration

As shown in Figure 61, all switches run BGP. Establish a TCP connection between Switch A and the RPKI server, and establish an IBGP connection between Switch A and Switch B.

Configure Switch A to advertise the BGP RPKI validation state to Switch B, and configure a routing policy to filter routes based on the validation state for Switch B.

Figure 61 Network diagram



Procedure

1. Configure IP addresses for interfaces and establish an IBGP connection between Switch A and Switch B. (Details not shown.)
2. Configure Switch A to establish a TCP connection to the RPKI server.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] rpki
[SwitchA-bgp-default-rpki] server tcp 1.1.1.2
[SwitchA-bgp-default-rpki-server] port 1234
[SwitchA-bgp-default-rpki-server] quit
```

3. Enable BGP RPKI validation on Switch A.

```
[SwitchA-bgp-default-rpki] check-origin-validation
[SwitchA-bgp-default-rpki] quit
```

4. Apply the BGP RPKI validation state to optimal route selection.

```
[SwitchA-bgp-default] address-family ipv4
[SwitchA-bgp-default-ipv4] bestroute origin-as-validation
```

5. Configure Switch A to advertise the BGP RPKI validation state to peer 1.2.3.2.

```
[SwitchA-bgp-default-ipv4] peer 1.2.3.2 advertise-ext-community
[SwitchA-bgp-default-ipv4] peer 1.2.3.2 advertise origin-as-validation
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

6. Configure Switch B to permit routes with a BGP RPKI validation state of **Valid**:

Configure a routing policy.

```
<SwitchB> system-view
[SwitchB] route-policy rpki_policy permit node 0
[SwitchB-route-policy-rpki_policy-0] if-match rpki valid
[SwitchB-route-policy-rpki_policy-0] quit
```

Apply the routing policy.

```
[SwitchB] bgp 100
[SwitchB-bgp-default] address-family ipv4
[SwitchB-bgp-default-ipv4] peer 1.2.3.1 route-policy rpki_policy import
```

Verifying the configuration

Display information about the connection between Switch A and the RPKI server.

```
[SwitchA] display bgp rpki server

  Server          VPN-index  Port        State      Time       ROAs(IPv4/IPv6)
  1.1.1.2         0          1234        Establish  00:04:43   5/4
```

The output shows that Switch A has established a TCP connection to the RPKI server.

Display the ROA information on Switch A.

```
[SwitchA] display bgp rpki table ipv4

 Total number of entries: 5

 Status codes: S - stale, U - used

  Network         Mask-range    Origin-AS    Server       Status
  1.2.3.4         8-24          100          1.1.1.2      U
  2.2.3.6         8-32          100          1.1.1.2      U
  2.2.3.6         10-24         4294967295   1.1.1.2      U
  2.2.3.9         20-24         4294967295   1.1.1.2      U
  3.2.3.5         8-26          200          1.1.1.2      U
```

The output shows that Switch A has obtained the ROA information from the RPKI server.

Display the BGP RPKI validation state on Switch A.

```
[SwitchA] display bgp routing-table ipv4 1.2.3.0

 BGP local router ID: 2.2.2.2
 Local AS number: 100
```

```
 Paths:   1 available, 1 best

 BGP routing table information of 1.2.3.0/24:
 Imported route.
 Original nexthop: 0.0.0.0
 Route age       : 03h30m59s
 OutLabel        : NULL
 RxPathID        : 0x0
 TxPathID        : 0x0
 Org-validation  : Valid
 AS-path         : (null)
 Origin          : incomplete
 Attribute value : MED 0, pref-val 32768
 State           : valid, local, best
 IP precedence   : N/A
 QoS local ID    : N/A
 Traffic index   : N/A
```

The route to 1.2.3.0 matches the ROA with a prefix of 1.2.3.4, a mask length range of 8 to 24, and an origin AS number of 100. Therefore, the BGP RPKI validation state of the route to 1.2.3.0 is **Valid**.

On Switch B, display detailed information about the BGP IPv4 unicast route to 1.2.3.0.

```
[SwitchB] display bgp routing-table ipv4 1.2.3.0
RR-client route.
 From            : 1.2.3.1 (192.168.56.22)
 Rely nexthop    : 1.2.3.1
 Original nexthop: 1.2.3.1
 Route age       : 03h30m59s
 OutLabel        : NULL
 Ext-Community   : <Origin Valid State: Valid >
 RxPathID        : 0x0
 TxPathID        : 0x0
 Org-validation  : Valid
 AS-path         : (null)
 Origin          : incomplete
 Attribute value : MED 0, localpref 100, pref-val 0
 State           : valid, internal, best
 IP precedence   : N/A
 QoS local ID    : N/A
 Traffic index   : N/A
```

The output displays information about the BGP IPv4 unicast route to 1.2.3.0 because Switch B permits routes with a BGP RPKI validation state of **Valid**.

Example: Configuring BGP keychain authentication

Network configuration

Configure keychain authentication to better protect the BGP connection between Switch A and Switch B.

Figure 62 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure EBGP connections.

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] bgp 100
   [SwitchA-bgp-default] peer 10.1.1.2 as-number 200
   [SwitchA-bgp-default] address-family ipv4
   [SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
   [SwitchA-bgp-default-ipv4] quit
   [SwitchA-bgp-default] quit
   ```

   # Configure Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] bgp 200
   [SwitchB-bgp-default] peer 10.1.1.1 as-number 100
   [SwitchB-bgp-default] address-family ipv4
   [SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
   [SwitchB-bgp-default-ipv4] quit
   [SwitchB-bgp-default] quit
   ```

3. Configure basic keychain authentication settings.

   # Configure Switch A.

   ```
   [SwitchA] keychain abc mode absolute
   [SwitchA-keychain-abc] key 1
   [SwitchA-keychain-abc-key-1] authentication-algorithm hmac-md5
   [SwitchA-keychain-abc-key-1] key-string plain pwd123
   [SwitchA-keychain-abc-key-1] send-lifetime utc 11:00:00 2023/01/01 to 12:00:00
   2023/12/30
   [SwitchA-keychain-abc-key-1] accept-lifetime utc 11:00:00 2023/01/01 to 12:00:00
   2023/12/30
   [SwitchA-keychain-abc-key-1] quit
   [SwitchA-keychain-abc] quit
   ```

   # Configure Switch B.

   ```
   [SwitchB] keychain abc mode absolute
   [SwitchB-keychain-abc] key 1
   [SwitchB-keychain-abc-key-1] authentication-algorithm hmac-md5
   [SwitchB-keychain-abc-key-1] key-string plain pwd123
   [SwitchB-keychain-abc-key-1] send-lifetime utc 11:00:00 2023/01/01 to 12:00:00
   2023/12/30
   [SwitchB-keychain-abc-key-1] accept-lifetime utc 11:00:00 2023/01/01 to 12:00:00
   2023/12/30
   [SwitchB-keychain-abc-key-1] quit
   [SwitchB-keychain-abc] quit
   ```

4. Configure BGP keychain authentication settings.

   # Configure Switch A.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] peer 10.1.1.2 keychain abc
[SwitchA-bgp-default] quit
```

   # Configure Switch B.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 10.1.1.1 keychain abc
[SwitchB-bgp-default] quit
```

Verifying the configuration

View IPv4 unicast peer information on Switch A.

```
[SwitchA] display bgp peer ipv4 unicast

 BGP local router ID: 10.1.1.1
 Local AS number: 100
 Total number of peers: 1                 Peers in established state: 1

 * - Dynamically created peer
 Peer                    AS  MsgRcvd  MsgSent OutQ  PrefRcv Up/Down  State

 10.1.1.2                200      13       16    0        0 00:10:34 Established
```

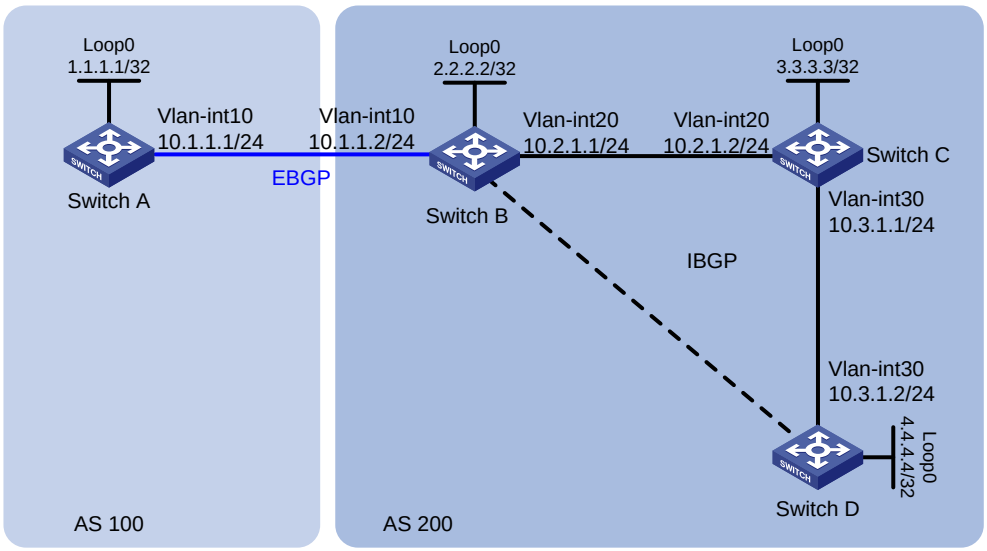
The output shows that Switch A and Switch B have passed keychain authentication and established a BGP session.

Example: Configuring BGP GTSM

Network configuration

As shown in Figure 63, Switch A is located in AS 100, and Switches B, C, and D are in AS 200. They all run BGP. Switch A and Switch B are EBGP peers, and Switches B, C, and D are IBGP peers. Configure BGP GTSM to protect all devices from CPU utilization-type packet attacks.

Figure 63 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure OSPF to ensure the loopback interfaces of Switch B, Switch C, and Switch D are reachable.

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf 1
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[SwitchB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf 1
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[SwitchC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

   # Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ospf 1
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 4.4.4.4 0.0.0.0
[SwitchD-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

3. Configure IBGP connections in a full-mesh manner.

   To avoid route flaps caused by unstable port states, this example uses loopback interfaces for IBGP peer session establishment.

   # Configure Switch B.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 3.3.3.3 as-number 65009
[SwitchB-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchB-bgp-default] peer 4.4.4.4 as-number 65009
[SwitchB-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 3.3.3.3 enable
[SwitchB-bgp-default-ipv4] peer 4.4.4.4 enable
[SwitchB-bgp-default-ipv4] peer 3.3.3.3 next-hop-local
[SwitchB-bgp-default-ipv4] peer 4.4.4.4 next-hop-local
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

   # Configure Switch C.

```
[SwitchC] bgp 200
```

```
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 2.2.2.2 as-number 65009
[SwitchC-bgp-default] peer 2.2.2.2 connect-interface loopback 0
[SwitchC-bgp-default] peer 4.4.4.4 as-number 65009
[SwitchC-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 2.2.2.2 enable
[SwitchC-bgp-default-ipv4] peer 4.4.4.4 enable
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Configure Switch D.

```
[SwitchD] bgp 200
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] peer 3.3.3.3 as-number 65009
[SwitchD-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchD-bgp-default] peer 2.2.2.2 as-number 65009
[SwitchD-bgp-default] peer 2.2.2.2 connect-interface loopback 0
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] peer 3.3.3.3 enable
[SwitchD-bgp-default-ipv4] peer 2.2.2.2 enable
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

4. Configure EBGP connections.

Configure Switch A.

```
<SwitchA>
[SwitchA] bgp 100
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 10.1.1.2 as-number 200
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
```

Configure Switch B.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 10.1.1.1 as-number 100
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
```

5. Configure BGP GTSM.

Configure BGP GTSM between Switch A and Switch B. The maximum number of hops to each other is 1 and the valid TTL range is [255,255], because the two devices are directly connected.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] peer 10.1.1.2 ttl-security hops 1
[SwitchA-bgp-default] quit
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 10.1.1.1 ttl-security hops 1
```

Configure BGP GTSM between Switch B and Switch C. The maximum number of hops to each other is 1 and the valid TTL range is [255,255], because the two devices are directly connected.

```
[SwitchC] bgp 200
```

```
[SwitchC-bgp-default] peer 2.2.2.2 ttl-security hops 1
[SwitchB-bgp-default] peer 3.3.3.3 ttl-security hops 1
```

Configure BGP GTSM between Switch C and Switch D. The maximum number of hops to each other is 1 and the valid TTL range is [255,255], because the two devices are directly connected.

```
[SwitchD] bgp 200
[SwitchD-bgp-default] peer 3.3.3.3 ttl-security hops 1
[SwitchC-bgp-default] peer 4.4.4.4 ttl-security hops 1
[SwitchC-bgp-default] quit
```

Configure BGP GTSM between Switch B and Switch D. The maximum number of hops to each other is 2 and the valid TTL range is [254,255], because the two devices are indirectly connected via an intermediate device.

```
[SwitchB-bgp-default] peer 4.4.4.4 ttl-security hops 2
[SwitchB-bgp-default] quit
[SwitchD-bgp-default] peer 2.2.2.2 ttl-security hops 2
[SwitchD-bgp-default] quit
```

IPv6 BGP security feature configuration examples

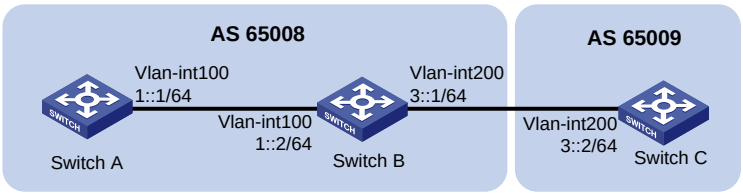
Example: Configuring IPsec for IPv6 BGP packets

Network configuration

As shown in Figure 64, all switches run IPv6 BGP. Establish an IBGP connection between Switch A and Switch B, and establish an EBGP connection between Switch B and Switch C.

To enhance security, configure IPsec to protect IPv6 BGP packets.

Figure 64 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Establish an IBGP connection between Switch A and Switch B:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] group ibgp internal
[SwitchA-bgp-default] peer 1::2 group ibgp
[SwitchA-bgp-default] address-family ipv6 unicast
[SwitchA-bgp-default-ipv6] peer ibgp enable
[SwitchA-bgp-default-ipv6] quit
```

```
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65008
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] group ibgp internal
[SwitchB-bgp-default] peer 1::1 group ibgp
[SwitchB-bgp-default] address-family ipv6 unicast
[SwitchB-bgp-default-ipv6] peer ibgp enable
[SwitchB-bgp-default-ipv6] quit
```

3. Establish an EBGP connection between Switch B and Switch C:

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] group ebgp external
[SwitchC-bgp-default] peer 3::1 as-number 65008
[SwitchC-bgp-default] peer 3::1 group ebgp
[SwitchC-bgp-default] address-family ipv6 unicast
[SwitchC-bgp-default-ipv6] peer ebgp enable
[SwitchC-bgp-default-ipv6] quit
[SwitchC-bgp-default] quit
```

Configure Switch B.

```
[SwitchB-bgp-default] group ebgp external
[SwitchB-bgp-default] peer 3::2 as-number 65009
[SwitchB-bgp-default] peer 3::2 group ebgp
[SwitchB-bgp-default] address-family ipv6 unicast
[SwitchB-bgp-default-ipv6] peer ebgp enable
[SwitchB-bgp-default-ipv6] quit
[SwitchB-bgp-default] quit
```

4. Configure IPsec transform sets and IPsec profiles:

On Switch A, create an IPsec transform set named **tran1**.

```
[SwitchA] ipsec transform-set tran1
```

Set the encapsulation mode to transport mode.

```
[SwitchA-ipsec-transform-set-tran1] encapsulation-mode transport
```

Set the security protocol to ESP, the encryption algorithm to DES, and authentication algorithm to SHA1.

```
[SwitchA-ipsec-transform-set-tran1] esp encryption-algorithm des
[SwitchA-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchA-ipsec-transform-set-tran1] quit
```

Create an IPsec profile named **policy001**, and specify the manual mode for it.

```
[SwitchA] ipsec profile policy001 manual
```

Use IPsec transform set **tran1**.

```
[SwitchA-ipsec-profile-policy001-manual] transform-set tran1
```

Set the SPIs of the inbound and outbound SAs to **12345**.

```
[SwitchA-ipsec-profile-policy001-manual] sa spi outbound esp 12345
[SwitchA-ipsec-profile-policy001-manual] sa spi inbound esp 12345
```

Set the keys for the inbound and outbound SAs using ESP to **abcdefg**.

```
[SwitchA-ipsec-profile-policy001-manual] sa string-key outbound esp simple abcdefg
[SwitchA-ipsec-profile-policy001-manual] sa string-key inbound esp simple abcdefg
[SwitchA-ipsec-profile-policy001-manual] quit
```

On Switch B, create an IPsec transform set named **tran1**.

```
[SwitchB] ipsec transform-set tran1
```

Set the encapsulation mode to transport mode.

```
[SwitchB-ipsec-transform-set-tran1] encapsulation-mode transport
```

Set the security protocol to ESP, the encryption algorithm to DES, and authentication algorithm to SHA1.

```
[SwitchB-ipsec-transform-set-tran1] esp encryption-algorithm des
[SwitchB-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran1] quit
```

Create IPsec profile named **policy001**, and specify the manual mode for it.

```
[SwitchB] ipsec profile policy001 manual
```

Use IPsec transform set **tran1**.

```
[SwitchB-ipsec-profile-policy001-manual] transform-set tran1
```

Set the SPIs of the inbound and outbound SAs to **12345**.

```
[SwitchB-ipsec-profile-policy001-manual] sa spi outbound esp 12345
[SwitchB-ipsec-profile-policy001-manual] sa spi inbound esp 12345
```

Set the keys for the inbound and outbound SAs using ESP to **abcdefg**.

```
[SwitchB-ipsec-profile-policy001-manual] sa string-key outbound esp simple abcdefg
[SwitchB-ipsec-profile-policy001-manual] sa string-key inbound esp simple abcdefg
[SwitchB-ipsec-profile-policy001-manual] quit
```

Create an IPsec transform set named **tran2**.

```
[SwitchB] ipsec transform-set tran2
```

Set the encapsulation mode to transport mode.

```
[SwitchB-ipsec-transform-set-tran2] encapsulation-mode transport
```

Set the security protocol to ESP, the encryption algorithm to DES, and authentication algorithm to SHA1.

```
[SwitchB-ipsec-transform-set-tran2] esp encryption-algorithm des
[SwitchB-ipsec-transform-set-tran2] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran2] quit
```

Create IPsec profile named **policy002**, and specify the manual mode for it.

```
[SwitchB] ipsec profile policy002 manual
```

Use IPsec transform set **tran2**.

```
[SwitchB-ipsec-profile-policy002-manual] transform-set tran2
```

Set the SPIs of the inbound and outbound SAs to **54321**.

```
[SwitchB-ipsec-profile-policy002-manual] sa spi outbound esp 54321
[SwitchB-ipsec-profile-policy002-manual] sa spi inbound esp 54321
```

Set the keys for the inbound and outbound SAs using ESP to **gfedcba**.

```
[SwitchB-ipsec-profile-policy002-manual] sa string-key outbound esp simple gfedcba
[SwitchB-ipsec-profile-policy002-manual] sa string-key inbound esp simple gfedcba
[SwitchB-ipsec-profile-policy002-manual] quit
```

On Switch C, create an IPsec transform set named **tran2**.

```
[SwitchC] ipsec transform-set tran2
```

Set the encapsulation mode to transport mode.

```
[SwitchC-ipsec-transform-set-tran2] encapsulation-mode transport
```

Set the security protocol to ESP, the encryption algorithm to DES, and authentication algorithm to SHA1.

```
[SwitchC-ipsec-transform-set-tran2] esp encryption-algorithm des
[SwitchC-ipsec-transform-set-tran2] esp authentication-algorithm sha1
[SwitchC-ipsec-transform-set-tran2] quit
```

Create IPsec profile named **policy002**, and specify the manual mode for it.

```
[SwitchC] ipsec profile policy002 manual
```

Use IPsec transform set **tran2**.

```
[SwitchC-ipsec-profile-policy002-manual] transform-set tran2
```

Set the SPIs of the inbound and outbound SAs to **54321**.

```
[SwitchC-ipsec-profile-policy002-manual] sa spi outbound esp 54321
[SwitchC-ipsec-profile-policy002-manual] sa spi inbound esp 54321
```

Set the keys for the inbound and outbound SAs using ESP to **gfedcba**.

```
[SwitchC-ipsec-profile-policy002-manual] sa string-key outbound esp simple gfedcba
[SwitchC-ipsec-profile-policy002-manual] sa string-key inbound esp simple gfedcba
[SwitchC-ipsec-profile-policy002-manual] quit
```

5. Configure IPsec to protect IPv6 BGP packets between Switch A and Switch B:

Configure Switch A.

```
[SwitchA] bgp 65008
[SwitchA-bgp-default] peer 1::2 ipsec-profile policy001
[SwitchA-bgp-default] quit
```

Configure Switch B.

```
[SwitchB] bgp 65008
[SwitchB-bgp-default] peer 1::1 ipsec-profile policy001
[SwitchB-bgp-default] quit
```

6. Configure IPsec to protect IPv6 BGP packets between Router B and Switch C:

Configure Switch C.

```
[SwitchC] bgp 65009
[SwitchC-bgp-default] peer ebgp ipsec-profile policy002
[SwitchC-bgp-default] quit
```

Configure Switch B.

```
[SwitchB] bgp 65008
[SwitchB-bgp-default] peer ebgp ipsec-profile policy002
[SwitchB-bgp-default] quit
```

Verifying the configuration

Display detailed information about IPv6 BGP peers on Switch B.

```
[SwitchB] display bgp peer ipv6 verbose

  Peer: 1::1      Local: 2.2.2.2
  Type: IBGP link
  BGP version 4, remote router ID 1.1.1.1
  BGP current state: Established, Up for 00h05m54s
  BGP current event: KATimerExpired
  BGP last state: OpenConfirm
  Port:  Local - 24896    Remote - 179
  Local source address: 1::2
```

```
        Configured:  Min Hold Time: 0 sec
        Configured: Active Hold Time: 180 sec   Keepalive Time: 60 sec
        Received  : Active Hold Time: 180 sec
        Negotiated: Active Hold Time: 180 sec   Keepalive Time: 60 sec
        Peer optional capabilities:
        Peer supports BGP multi-protocol extension
        Peer supports BGP route refresh capability
        Peer supports BGP route AS4 capability
        Address family IPv6 Unicast: advertised and received

InQ updates: 0, OutQ updates: 0
 NLRI statistics:
        Rcvd:   UnReach NLRI          0,       Reach NLRI          0
        Sent:   UnReach NLRI          0,       Reach NLRI          3

 Message statistics:
 Msg type     Last rcvd time/      Current rcvd count/      History rcvd count/
              Last sent time       Current sent count       History sent count
 Open         18:59:15-2019.4.24   1                        1
              18:59:15-2019.4.24   1                        2
 Update       -                    0                        0
              18:59:16-2019.4.24   1                        1
 Notification -                    0                        0
              18:59:15-2019.4.24   0                        1
 Keepalive    18:59:15-2019.4.24   1                        1
              18:59:15-2019.4.24   1                        1
 RouteRefresh -                    0                        0
              -                    0                        0
 Total        -                    2                        2
              -                    3                        5

 Maximum allowed prefix number: 4294967295
 Threshold: 75%
 Minimum time between advertisements is 15 seconds
 Optional capabilities:
  Multi-protocol extended capability has been enabled
  Route refresh capability has been enabled
 Peer preferred value: 0
 IPsec profile name: policy001

 Routing policy configured:
 No routing policy is configured

        Peer: 3::2      Local: 2.2.2.2
        Type: EBGP link
        BGP version 4, remote router ID 3.3.3.3
        BGP current state: Established, Up for 00h05m00s
        BGP current event: KATimerExpired
```

```
        BGP last state: OpenConfirm
        Port:  Local - 24897    Remote - 179
        Local source address: 3::1
        Configured:  Min Hold Time: 0 sec
        Configured: Active Hold Time: 180 sec   Keepalive Time: 60 sec
        Received  : Active Hold Time: 180 sec
        Negotiated: Active Hold Time: 180 sec   Keepalive Time: 60 sec
        Peer optional capabilities:
        Peer supports BGP multi-protocol extension
        Peer supports BGP route refresh capability
        Peer supports BGP route AS4 capability
        Address family IPv6 Unicast: advertised and received

Maximum allowed prefix number: 4294967295
Threshold: 75%
Minimum time between advertisements is 30 seconds
Optional capabilities:
 Multi-protocol extended capability has been enabled
 Route refresh capability has been enabled
Peer preferred value: 0
IPsec profile name: policy002

Routing policy configured:
No routing policy is configured
```

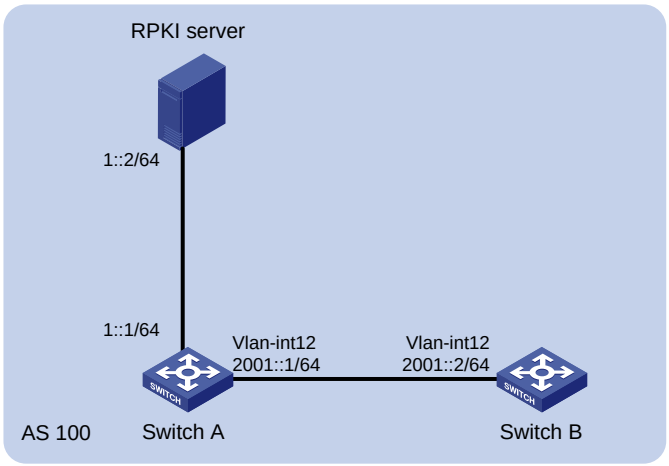
Example: Configuring BGP RPKI

Network configuration

As shown in Figure 65, all switches run IPv6 BGP. Establish a TCP connection between Switch A and the RPKI server, and establish an IBGP connection between Switch A and Switch B.

Configure Switch A to advertise the BGP RPKI validation state to Switch B, and configure a routing policy to filter routes based on the validation state for Switch B.

Figure 65 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces and establish an IBGP connection between Switch A and Switch B. (Details not shown.)
2. Configure Switch A to establish a TCP connection to the RPKI server.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] rpki
[SwitchA-bgp-default-rpki] server tcp 1::2
[RouterA-bgp-default-rpki-server] port 1234
[RouterA-bgp-default-rpki-server] quit
```

3. Enable BGP RPKI validation on Switch A.

```
[SwitchA-bgp-default-rpki] check-origin-validation
[SwitchA-bgp-default-rpki] quit
```

4. Apply the BGP RPKI validation state to optimal route selection.

```
[SwitchA-bgp-default] address-family ipv6
[SwitchA-bgp-default-ipv6] bestroute origin-as-validation
```

5. Configure Switch A to advertise the BGP RPKI validation state to peer 2001::2.

```
[SwitchA-bgp-default-ipv6] peer 2001::2 advertise-ext-community
[SwitchA-bgp-default-ipv6] peer 2001::2 advertise origin-as-validation
[SwitchA-bgp-default-ipv6] quit
[SwitchA-bgp-default] quit
```

6. Configure Switch B to permit routes with a BGP RPKI validation state of **Valid**:

   # Configure a routing policy.

```
<SwitchB> system-view
[SwitchB] route-policy rpki_policy permit node 0
[SwitchB-route-policy-rpki_policy-0] if-match rpki valid
```

   # Apply the routing policy.

```
<SwitchB> system-view
[SwitchB] bgp 100
[SwitchB-bgp-default] address-family ipv6
[SwitchB-bgp-default-ipv6] peer 2001::1 route-policy rpki_policy import
```

Verifying the configuration

Display information about the connection between Switch A and the RPKI server.

```
[SwitchA] display bgp rpki server

  Server          VPN-index  Port        State      Time       ROAs(IPv4/IPv6)
  1::2            0          1234        Establish  00:04:43   5/5
```

The output shows that Switch A has established a TCP connection to the RPKI server.

Display the ROA information on Switch A.

```
[SwitchA] display bgp rpki table ipv6

 Total number of entries: 5

 Status codes: S - stale, U - used

  Network         Mask-range    Origin-AS    Server       Status
  2001:4860::     32-32         100          1::2         U
```

```
  2404:6800::     32-32         100          1::2         U
  2607:F8B0::     28-28         4294967295   1::2         U
  2A03:ACE0::     40-40         4294967295   1::2         U
  2001::1         64-64         200          1::2         U
```

The output shows that Switch A has obtained ROAs from the RPKI server.

Display the BGP RPKI validation state on Switch A.

```
[SwitchA] display bgp routing-table ipv6 2001::1 64

 BGP local router ID: 2.2.2.2
 Local AS number: 100

 Paths:   1 available, 1 best

 BGP routing table information of 2001::1/64:
 Imported route.
 Original nexthop: 0.0.0.0
 Route age       : 03h30m59s
 OutLabel        : NULL
 RxPathID        : 0x0
 TxPathID        : 0x0
 Org-validation  : Valid
 AS-path         : (null)
 Origin          : incomplete
 Attribute value : MED 0, pref-val 32768
 State           : valid, local, best
 IP precedence   : N/A
 QoS local ID    : N/A
```

The route to 2001::1 matches the ROA with a prefix of 2001::1, a prefix length range of 64 to 64, and an origin AS number of 200. Therefore, the BGP RPKI validation state of the route to 2001::1 is **Valid**.

On Switch B, display detailed information about the BGP IPv6 unicast route to network 2001::1.

```
[SwitchB] display bgp routing-table ipv6 2001::1 64

RR-client route.
 From            : 2001::1 64 (192.168.56.22)
 Rely nexthop    : 2001::1
 Original nexthop: 2001::1
 Route age       : 03h30m59s
 OutLabel        : NULL
 Ext-Community   : <Origin Valid State: Valid >
 RxPathID        : 0x0
 TxPathID        : 0x0
 Org-validation  : Invalid
 AS-path         : (null)
 Origin          : incomplete
 Attribute value : MED 0, localpref 100, pref-val 0
 State           : valid, internal, best
 IP precedence   : N/A
```

```
QoS local ID     : N/A
Traffic index    : N/A
```

The output displays information about the BGP IPv6 unicast route to 2001::1 because Switch B permits routes with a BGP RPKI validation state of **Valid**.

Improving BGP network reliability

BGP network reliability improvement tasks at a glance

To improve the BGP network reliability, perform the following tasks:

- Configuring BGP GR
- Configuring BGP NSR
- Configuring BFD for BGP
- Configuring BGP FRR
- Configuring BGP maintenance features
  - Configuring BGP isolation
  - Configuring BGP shutdown

Configuring BGP GR

BGP GR operating mechanism

Graceful Restart (GR) ensures forwarding continuous when a routing protocol restarts or an active/standby switchover occurs. Two routers are required to complete a GR process. The following are router roles in a GR process:

- **GR restarter**—Performs GR upon a BGP restart or active/standby switchover.
- **GR helper**—Helps the GR restarter to complete the GR process.

A device can act as a GR restarter and GR helper at the same time.

BGP GR works as follows:

1. The BGP GR restarter and helper exchange OPEN messages for GR capability negotiation. If both parties have the GR capability, they establish a GR-capable session. The GR restarter sends the GR timer set by the **`graceful-restart timer restart`** command to the GR helper in an OPEN message.
2. When an active/standby switchover occurs or BGP restarts, the GR restarter does not remove existing BGP routes from Routing Information Base (RIB) and Forwarding Information Base (FIB). It still uses these routes for packet forwarding, and it starts the RIB purge timer (set by the **`graceful-restart timer purge-time`** command). The GR helper marks all routes learned from the GR restarter as stale instead of deleting them. It continues to use these routes for packet forwarding. During the GR process, packet forwarding is not interrupted.
3. After the active/standby switchover or BGP restart completes, the GR restarter re-establishes a BGP session to the GR helper. If the BGP session fails to be established after both the GR timer and the extra timer to wait expire, the GR helper removes the stale routes. To set the extra timer to wait after the restart timer expires, execute the **`peer graceful-restart timer restart extra`** command.
4. If the BGP session is established, routing information is exchanged for the GR restarter to retrieve route entries and for the GR helper to recover stale routes.
5. Both the GR restarter and the GR helper start the End-Of-RIB marker waiting timer.

   The End-Of-RIB marker waiting timer is set by the **`graceful-restart timer wait-for-rib`** command.

- If routing information exchange is completed within the waiting timer, the GR restarter and the GR helper act as follows:
  - The GR restarter updates its routing table with learned BGP routes, removes the aged routes from the RIB, and quits the GR process.
  - The GR helper removes the stale routes and quits the GR process.
- If routing information exchange is not completed within the waiting timer, the GR restarter and the GR helper act as follows:
  - The GR restarter updates the RIB with the BGP routes already learned, removes the aged routes from the RIB, and quits the GR process.
  - The GR helper removes the stale routes and quits the GR process.

The End-Of-RIB marker is an update message that does not contain any NLRI or withdrawn NLRI. After the GR restarter and the GR helper complete route exchange, they additionally send an End-Of-RIB marker to each other. The marker indicates the completion of route update.

6. The GR restarter quits the GR process if routing information exchange is not completed within the RIB purge timer. It updates the RIB with the BGP routes already learned, and removes the aged routes.

After BGP completes GR, it advertises updated routes as follows:

- If the routes rely on other protocols, for example, redistributed OSPF routes and routes with MPLS labels, BGP starts a wait timer for these protocols to complete GR. This ensures that incorrect and unreachable routes are not advertised.
  - If the protocols complete GR and notify BGP before the wait timer expires, BGP cancels the wait timer and advertises the routes.
  - If the protocols does not complete GR or notify BGP before the wait timer expires, BGP immediately advertises the routes after the wait timer expires.
- If the routes do not rely on other protocols, BGP immediately advertises the routes.

Enabling the GR capability globally or on a per peer basis

The **graceful-restart** command globally enables the BGP GR capability for a BGP instance. The BGP instance can establishes GR-capable sessions to all peers. However, enabling the BGP GR capability globally will terminate and re-establish all BGP sessions and thus cause unnecessary traffic interruption. To avoid this issue, use the **peer capability-advertise graceful-restart** command to enable the capability of BGP GR negotiation with only desired peers.

After you configure the **peer capability-advertise graceful-restart** command, the local device re-establishes the session to the specified peer and performs GR capability negotiation with the peer through Open messages. If both the local device and the peer have GR capability, they can establish a GR-capable session.

Enabling BGP to gracefully reset peer sessions gracefully

When the TCP connection goes down, the hold timer expires, or the support for a new address family is added, BGP terminates and re-establishes the peer sessions, which will cause traffic interruption. To avoid traffic interruption in these cases, enable BGP to reset peer sessions gracefully.

Setting the F-bit to 0 in OPEN messages used for GR negotiation

When BGP has a GR negotiation with a peer, it sends OPEN messages with the F-bit to that peer. Supported F-bit values include:

- **1**—When BGP acts as the GR helper, it retains forwarding entries and thus can forward traffic to the GR restarter.
- **0**—When BGP acts as the GR helper, it updates forwarding entries and thus cannot forward traffic to the GR restarter.

By default, the GR restarter sets the F-bit to 1 in OPEN messages used for GR negotiation. During the GR process, the GR restarter does not have forwarding capabilities, but it still receives traffic from the GR helper, which causes data loss. To avoid data loss, use this command on the GR helper to set the F-bit to 0 in OPEN messages used for GR negotiation.

Restrictions and guidelines

Follow these guidelines when you configure BGP GR:

- The maximum time to wait for the End-of-RIB marker configured on the local end is not advertised to the peer. It controls the time for the local end to receive updates from the peer. Set a large value for the maximum time to wait for the End-of-RIB marker when a large number of routes need to be exchanged. When you use the **peer graceful-restart timer restart extra** command, you can specify the **no-limit** keyword. In this situation, the GR helper does not start the End-of-RIB marker waiting timer after BGP session re-establishment and keeps waiting for the End-of-RIB marker from the GR restarter.
- As a best practice, perform the BGP GR configuration on both the GR restarter and GR helper.
- When the following conditions exist, BGP might advertise incomplete routes after completing GR:
  - The routes rely on other protocols, for example, redistributed OSPF routes.
  - BGP maintains a large amount of routing information. In this case, BGP and the protocols take a long time to complete GR.

  For BGP to correctly advertise the routes after BGP and the protocols complete GR, set a longer wait timer for BGP.
- The local device can negotiate GR capability with a peer successfully when the GR negotiation capability is enabled globally or only for the peer on the local device.
- If the local device starts BGP GR through the configuration of the **peer capability-advertise graceful-restart** command, enable or disable the global GR capability with caution. This operation will terminate the current BGP GR process and re-establish BGP sessions.
- If the local device starts BGP GR through the configuration of the **graceful-restart** command, use the **peer capability-advertise graceful-restart** command with caution. This command will terminate the BGP GR process associated with the local device and the specified peer and re-establish the session between the two peers.
- If you enable the capability of GR negotiation with a single peer, the time settings used by the GR process that involves the local device and the peer depend on the following commands:
  - **peer graceful-restart timer restart**
  - **peer graceful-restart timer wait-for-rib**
  - **graceful-restart timer purge-time**
  - **peer graceful-restart timer restart extra**

  You can enable the GR negotiation capability globally and for a single peer on the local device at the same time. In this case, the GR time settings configured for the peer take precedence over the global GR time settings.

Enabling the GR capability globally

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable the GR capability for BGP globally.

   graceful-restart

   By default, the GR capability is disabled for BGP.

4. Configure the GR timer.

   graceful-restart timer restart *timer*

   The default setting is 150 seconds.

   The time that a peer waits to re-establish a session must be less than the hold time.

5. (Optional.) Set the extra time to wait after the restart timer expires. Choose one command as needed:

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **graceful-restart timer restart extra** { *time* | **no-limit** }

   interface-peer *interface-type interface-number* **graceful-restart timer restart extra** { *time* | **no-limit** }

   By default, the extra time to wait after the restart timer expires is 0 seconds.

6. Configure the maximum time to wait for the End-of-RIB marker.

   graceful-restart timer wait-for-rib *timer*

   The default setting is 600 seconds.

7. Configure the RIB purge timer.

   graceful-restart timer purge-time *timer*

   The default setting is 480 seconds.

8. (Optional.) Configure the time that BGP must wait for other protocols to complete GR after BGP completes GR.

   bgp update-delay wait-other-protocol *seconds*

   By default, after BGP completes GR, it must wait a maximum of 300 seconds for other protocols to complete GR.

9. (Optional.) Enable BGP to reset peer sessions gracefully.

   graceful-restart peer-reset [**all**]

   By default, BGP does not reset peer sessions gracefully.

Enabling the GR capability on a per peer basis

1. Enter system view.

   system-view

2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

3. Enable the capability of GR negotiation with a peer.

   peer { *group-name* | *ipv4-address* | *ipv6-address* } **capability-advertise graceful-restart**

   By default, the capability of GR negotiation with peers is disabled.

4. Configure the GR timer for a peer.

   peer { *group-name* | *ipv4-address* | *ipv6-address* } **graceful-restart timer restart** *time-value*

   By default, the GR timer is 150 seconds for peers.

   The time that a peer waits to re-establish a session must be less than the hold time.

5. (Optional.) Set the extra time to wait after the GR timer expires.

   peer { *group-name* | *ipv4-address* | *ipv6-address* } **graceful-restart timer restart extra** { *time* | **no-limit** }

   By default, the extra time to wait after the GR timer expires is 0 seconds.

6. Configure the maximum time to wait for the End-of-RIB marker from a peer.

   peer { *group-name* | *ipv4-address* | *ipv6-address* } **graceful-restart timer wait-for-rib** *time-value*

   The default setting is 600 seconds.

7. Configure the RIB purge timer.

   graceful-restart timer purge-time *timer*

   The default setting is 480 seconds.

Configuring BGP NSR

About this task

BGP nonstop routing (NSR) backs up BGP state and data information from the active BGP process to the standby BGP process. The standby BGP process takes over when any of the following events occurs:

- The active BGP process restarts.
- The member device that runs the active BGP process fails.
- An ISSU starts on the member device that runs the active BGP process.

Choose BGP NSR or BGP GR according to the actual situation and make sure data forwarding is not interrupted. The differences between BGP NSR and BGP GR as follows:

- **Device requirements**—The active and standby processes of BGP must run on different IRF member devices. To run BGP NSR, you must set up an IRF fabric that contains a minimum of two member devices. To run BGP GR, IRF fabric setup is not required. If an IRF fabric already exists, you can run BGP GR when the IRF fabric contains only one member device.
- **BGP peer requirements**—When you use BGP NSR, BGP peers cannot detect process restart or active/standby switchover on the local device. BGP NSR can recover BGP route information without requiring the cooperation of BGP peers. BGP GR requires BGP peers to have the GR helper capability to help the GR restarter complete the GR process.

After BGP completes NSR, it advertises updated routes as follows:

- If the routes rely on other protocols, for example, redistributed OSPF routes and routes with MPLS labels, BGP starts a wait timer for these protocols to complete NSR. This ensures that incorrect and unreachable routes are not advertised.
  - If the protocols complete NSR and notify BGP before the wait timer expires, BGP cancels the wait timer and advertises the routes.
  - After the wait timer expires, BGP immediately advertises the routes.
- If the routes do not rely on other protocols, BGP immediately advertises the routes.

Restrictions and guidelines

When both BGP GR and BGP NSR are configured on the device, the following rules apply:

- BGP NSR has a higher priority than BGP GR. When the active process goes down, the device does not act as the GR restarter and performs NSR to ensure nonstop forwarding.
- When the following conditions exist, the GR process might fail even if BGP NSR is enabled on the GR helper:
  - BGP is restarted through GR.
  - An active/standby switchover is triggered on the GR helper before the GR process finishes.

To use BGP NSR in MPLS L3VPN, you must enable RIB NSR. For information about RIB NSR, see IP routing basics configuration in *Layer 3—IP Routing Configuration Guide*.

When the following conditions exist, BGP might advertise incomplete routes after completing NSR:

- The routes rely on other protocols, for example, redistributed OSPF routes.
- BGP maintains a large amount of routing information. In this case, BGP and the protocols take a long time to complete NSR.

To ensure BGP advertises the routes after BGP and the protocols complete NSR, set a longer wait timer for BGP.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP NSR.

   non-stop-routing

   By default, BGP NSR is enabled.
4. (Optional.) Configure the time that BGP must wait for other protocols to complete NSR after BGP completes NSR.

   bgp update-delay wait-other-protocol *seconds*

   By default, after BGP completes NSR, it must wait a maximum of 300 seconds for other protocols to complete NSR.

Configuring BFD for BGP

About this task

BGP maintains neighbor relationships based on the keepalive timer and hold timer in seconds. It requires that the hold time must be at least three times the keepalive interval. This mechanism slows down link failure detection. Once a failure occurs on a high-speed link, a large quantity of packets will be dropped before routing convergence completes. BFD for BGP can solve this problem by fast detecting link failures to reduce convergence time.

Before you enable BFD for a BGP peer or peer group, you must establish a BGP session between the local router and the peer or peer group.

For more information about BFD, see *High Availability Configuration Guide*.

Restrictions and guidelines

Follow these restrictions and guidelines when you configure echo packet mode BFD:

- Echo packet mode BFD is applicable only to directly connected BGP peers.
- Echo packet mode BFD is not applicable to BGP peers established through loopback interfaces.

To detect the link to a BGP peer established through link-local addresses, you must configure single-hop BFD in control packet mode.

To establish a control packet mode BFD session to a BGP peer, you must configure the same BFD detection mode (multi-hop or single-hop) on the local device and the BGP peer.

If you have enabled GR, use BFD with caution because BFD might detect a failure before the system performs GR, which will result in GR failure. If you have enabled both BFD and GR for BGP, do not disable BFD during a GR process to avoid GR failure.

Procedure (IPv4 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enable BFD to detect the link to the specified BGP peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **bfd** [**echo** | **multi-hop** | **single-hop**]
   By default, BFD is disabled.

Procedure (IPv6 peers)

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enable BFD to detect the link to the specified IPv6 BGP peer or peer group. Choose one command as needed:
   peer { *group-name* | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **bfd** [**echo** | **multi-hop** | **single-hop**]
   interface-peer *interface-type interface-number* **bfd** [**multi-hop** | **single-hop**]
   By default, BFD is disabled.
   The *link-local-address* **interface** *interface-type interface-number* option and the **echo** keyword are mutually exclusive.

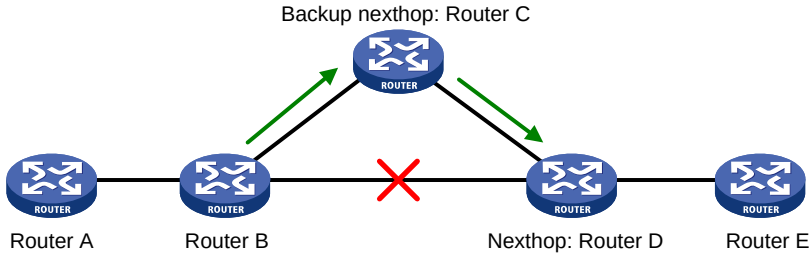
Configuring BGP FRR

About BGP FRR

When a link fails, the packets on the link are discarded, and a routing loop might occur until BGP completes routing convergence based on the new network topology.

You can enable BGP fast reroute (FRR) to resolve this issue.

Figure 66 Network diagram for BGP FRR



After you configure FRR on Router B as shown in Figure 66, BGP generates a backup next hop Router C for the primary route. BGP uses ARP or BFD echo packet mode in an IPv4 network or ND in an IPv6 network to detect the connectivity to Router D. When the link to Router D fails, BGP directs packets to the backup next hop. At the same time, BGP calculates a new optimal route, and forwards packets over the optimal route.

You can use the following methods to configure BGP FRR:

- **Method 1**—Execute the **pic** command in BGP address family view. BGP calculates a backup next hop for each BGP route in the address family if there are two or more unequal-cost routes that reach the destination.
- **Method 2**—Execute the **fast-reroute route-policy** command to use a routing policy in which a backup next hop is specified by using the command **apply** [**ipv6**] **fast-reroute backup-nexthop**. The backup next hop calculated by BGP must be the same as the specified backup next hop. Otherwise, BGP does not generate a backup next hop for the primary route. You can also configure **if-match** clauses in the routing policy to identify the routes protected by FRR.

If both methods are configured, Method 2 takes precedence over Method 1.

BGP supports FRR for IPv4 and IPv6 unicast routes, but not for IPv4 and IPv6 multicast routes.

Configuring BGP FRR by using a routing policy (IPv4 unicast address family)

1. Enter system view.

   system-view
2. Create a routing policy and enter routing policy view.

   route-policy *route-policy-name* **permit node** *node-number*

   For more information about this command, see routing policy commands in *Layer 3—IP Routing Command Reference*.
3. Set the backup next hop for FRR.

   apply fast-reroute backup-nexthop *ipv4-address*

   By default, no backup next hop is set.

   For more information about this command, see routing policy commands in *Layer 3—IP Routing Command Reference*.
4. Return to system view.

   quit
5. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
6. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   - Enter BGP IPv4 unicast address family view.

address-family ipv4 [**unicast**]

- Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

  ip vpn-instance *vpn-instance-name*

  address-family ipv4 [**unicast**]

7. Apply a routing policy to FRR for the address family.

   fast-reroute route-policy *route-policy-name*

   By default, no routing policy is applied.

   The **apply fast-reroute backup-nexthop** command and **apply ipv6 fast-reroute backup-nexthop** command can take effect in the applied routing policy. Other **apply** commands do not take effect.

Configuring BGP FRR by using a routing policy (IPv6 unicast address family)

1. Enter system view.

   system-view

2. Create a routing policy and enter routing policy view.

   route-policy *route-policy-name* **permit node** *node-number*

   For more information about this command, see routing policy commands in *Layer 3—IP Routing Command Reference*.

3. Set the backup next hop for FRR.

   apply ipv6 fast-reroute backup-nexthop *ipv6-address*

   By default, no backup next hop is set.

   For more information about this command, see routing policy commands in *Layer 3—IP Routing Command Reference*.

4. Return to system view.

   quit

5. Enter BGP instance view or BGP-VPN instance view.

   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]

   - Enter BGP-VPN instance view.

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

6. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.

   address-family ipv6 [**unicast**]

7. Apply a routing policy to FRR for the address family.

   fast-reroute route-policy *route-policy-name*

   By default, no routing policy is applied.

   The **apply fast-reroute backup-nexthop** and **apply ipv6 fast-reroute backup-nexthop** commands can take effect in the applied routing policy. Other **apply** commands do not take effect.

Configuring BGP FRR through PIC (IPv4 unicast address family)

Restrictions and guidelines

This feature might result in routing loops. Use it with caution.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enter BGP IPv4 unicast address family view or BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
4. Enable BGP FRR through PIC.
   pic
   By default, BGP FRR is disabled.

Configuring BGP FRR through PIC (IPv6 unicast address family)

Restrictions and guidelines

This feature might result in routing loops. Use it with caution.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.
     bgp *as-number* [**instance** *instance-name*]
   - Enter BGP-VPN instance view.
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
3. Enter BGP IPv6 unicast address family view or BGP-VPN IPv6 unicast address family view.
   address-family ipv6 [**unicast**]
4. Enable BGP FRR through PIC.
   pic
   By default, BGP FRR is disabled.

Configuring BFD-powered nexthop connectivity detection for BGP FRR

About this task

By default, BGP FRR uses ARP to detect next hop connectivity for the primary route. This method is not efficient at detecting the failure of the primary route. As a result, data loss might occur when the primary route fails, because the backup path might not take over to forward traffic in time. To resolve this issue, enable BFD-powered nexthop connectivity detection for BGP FRR. This feature enables BGP to create an IP FRR BFD session that detects next hop connectivity for the primary route. This can speed up primary and backup route switchover when the next hop of the primary route fails.

Restrictions and guidelines

This feature takes effect and creates an IP FRR BFD session only when the primary route has a backup next hop.

BGP does not support backup next hop calculation for the routes used for load balancing. This command cannot take effect when the primary route is an equal-cost route.

BGP can establish a control-mode BFD session with a peer only after negotiation. To use control-mode BFD to detect the next hop of the primary link, perform one of the following tasks:

- Configure the **primary-path-detect bfd** command on the peer device in the primary link.

  This task is applicable when both ends of the primary link are configured with FRR.
- Manually configure a static BFD session that uses the following settings on the peer device in the primary link:
  - Source IP address: Destination IP address of the BFD session automatically created on the local device in the primary link.
  - Destination IP address: Source IP address of the BFD session automatically created on the local device in the primary link.
  - Remote ID: Local ID of the BFD session automatically created on the local device in the primary link.

When you configure echo-mode BFD for nexthop connectivity detection, you only need to configure this feature on the local router.

When another routing protocol (such as OSPF or IS-IS) uses BFD to detect next hop connectivity for the primary route, it also creates a BFD session. If the detected link is the same as the link attached to the next hop of the BGP primary route, BGP reuses the BFD session created by the protocol. In this situation, BGP will not create an additional BFD session.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. (Optional.) Enable BFD-powered nexthop connectivity detection for BGP FRR.

   primary-path-detect bfd { **ctrl** | **echo** }

   By default, BGP FRR uses ARP to detect nexthop connectivity for the primary route.

Configuring BGP maintenance features

About this task

For maintenance purposes, you can use either BGP isolation or BGP shutdown to remove the device from the network. The device will not be used for forwarding traffic from its peers after they reselect an optimal route. Both BGP isolation and BGP shutdown retain the device configuration to facilitate maintenance workload.

After maintenance, you can disable BGP isolation or BGP shutdown to add the device back to the network.

Configuring BGP isolation

About this task

To reduce impact on traffic forwarding, you can isolate the device from the network before upgrading it. BGP isolation works as follows:

1. BGP withdraws all routes advertised by the device except for the directly-connected routes.
2. BGP keeps all routes learned from its peers.

   With BGP isolation enabled, BGP minimizes the preferences of direct routes advertised to peers by maximizing their MED values and minimizing their local preferences.
3. Each peer of the device reselects an optimal route and updates the FIB table. During optimal route selection, the peers can still use the routes advertised by the device to forward traffic.
4. After an optimal route is selected and the FIB table is updated, the peers stop forwarding packets except for those destined for the device to the device. The device is fully isolated from the network and you can upgrade it.
5. After the maintenance, disable BGP isolation on the device to gracefully add it back to the network. After returning to the network, the device advertises and learns routes as follows:
   - Advertises routes to its peers.
   - Learns routes if BGP was restarted during the isolation.

Restrictions and guidelines

To avoid isolation failure, do not use this feature when GR or NSR is enabled for the device.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP isolation to gracefully remove the device from the network.

   isolate enable

   By default, BGP isolation is disabled.
4. Disable BGP isolation to gracefully add the device back to the network.

   undo isolate enable

Configuring BGP shutdown

About this task

For maintenance purposes, you can use this feature to temporarily disconnect BGP sessions from all peers and peer groups. After maintenance, you can disable this feature to restore these sessions.

Restrictions and guidelines

With this feature enabled, the device tears down all sessions and clears all routing information.

The device cannot establish BGP sessions in one of the following conditions:

- BGP shutdown is configured.
- BGP session establishment is disabled. For more information about disabling BGP session establishment, see "Disabling BGP session establishment."

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Shut down BGP sessions to all peers and peer groups.
   shutdown process
   By default, BGP does not shut down sessions to any peers or peer groups.
4. Restore the BGP sessions to add the device back to the network.
   undo shutdown process

Display and maintenance commands for BGP network reliability improvement

Execute **display** commands in any view.

Displaying BGP (IPv4 unicast address family)

Task	Command
Display BGP NSR status information.	**display bgp** [**instance** *instance-name*] **non-stop-routing status**

Displaying BGP (IPv6 unicast address family)

Task	Command
Display BGP NSR status information.	**display bgp** [**instance** *instance-name*] **non-stop-routing status**

Displaying BGP (IPv4 multicast address family)

Task	Command
Display BGP NSR status information.	**display bgp** [**instance** *instance-name*] **non-stop-routing status**

Displaying BGP (IPv6 multicast address family)

Task	Command
Display BGP NSR status information.	**display bgp** [**instance** *instance-name*] **non-stop-routing status**

IPv4 BGP network reliability improvement configuration examples

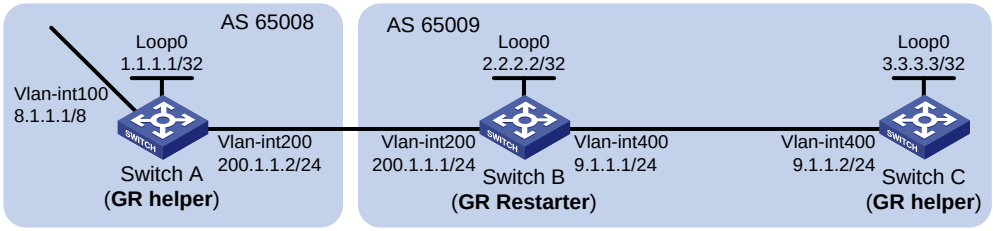
Example: Configuring BGP GR

Network configuration

As shown in Figure 67, all switches run BGP. EBGP runs between Switch A and Switch B. IBGP runs between Switch B and Switch C.

Enable GR capability for BGP so that the communication between Switch A and Switch C is not affected when an active/standby switchover occurs on Switch B.

Figure 67 Network diagram



Procedure

1. Configure Switch A:

   # Configure IP addresses for interfaces. (Details not shown.)

   # Configure the EBGP connection.

   ```
   <SwitchA> system-view
   [SwitchA] bgp 65008
   [SwitchA-bgp-default] router-id 1.1.1.1
   [SwitchA-bgp-default] peer 200.1.1.1 as-number 65009
   ```

   # Enable GR capability for BGP.

   ```
   [SwitchA-bgp-default] graceful-restart
   ```

   # Inject network 8.0.0.0/8 to the BGP routing table.

   ```
   [SwitchA-bgp-default] address-family ipv4
   [SwitchA-bgp-default-ipv4] network 8.0.0.0
   ```

   # Enable Switch A to exchange IPv4 unicast routing information with Switch B.

   ```
   [SwitchA-bgp-default-ipv4] peer 200.1.1.1 enable
   ```

2. Configure Switch B:

   # Configure IP addresses for interfaces. (Details not shown.)

   # Configure the EBGP connection.

   ```
   <SwitchB> system-view
   [SwitchB] bgp 65009
   ```

```
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 200.1.1.2 as-number 65008
```

Configure the IBGP connection.

```
[SwitchB-bgp-default] peer 9.1.1.2 as-number 65009
```

Enable GR capability for BGP.

```
[SwitchB-bgp-default] graceful-restart
```

Inject networks 200.1.1.0/24 and 9.1.1.0/24 to the BGP routing table.

```
[SwitchB-bgp-default] address-family ipv4
[SwitchB-bgp-default-ipv4] network 200.1.1.0 24
[SwitchB-bgp-default-ipv4] network 9.1.1.0 24
```

Enable Switch B to exchange IPv4 unicast routing information with Switch A and Switch C.

```
[SwitchB-bgp-default-ipv4] peer 200.1.1.2 enable
[SwitchB-bgp-default-ipv4] peer 9.1.1.2 enable
```

3. Configure Switch C:

Configure IP addresses for interfaces. (Details not shown.)

Configure the IBGP connection.

```
<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 9.1.1.1 as-number 65009
```

Enable GR capability for BGP.

```
[SwitchC-bgp-default] graceful-restart
```

Enable Switch C to exchange IPv4 unicast routing information with Switch B.

```
[SwitchC-bgp-default-ipv4] peer 9.1.1.1 enable
```

Verifying the configuration

Ping Switch C on Switch A. Meanwhile, perform an active/standby switchover on Switch B. The ping operation is successful during the whole switchover process. (Details not shown.)

Example: Configuring BFD for BGP

Network configuration

As shown in Figure 68, configure OSPF as the IGP in AS 200.

- Establish two IBGP connections between Switch A and Switch C. When both paths operate correctly, Switch C uses the path Switch A<—>Switch B<—>Switch C to exchange packets with network 1.1.1.0/24.
- Configure BFD over the path. When the path fails, BFD can quickly detect the failure and notify it to BGP. Then, the path Switch A<—>Switch D<—>Switch C takes effect immediately.

Figure 68 Network diagram

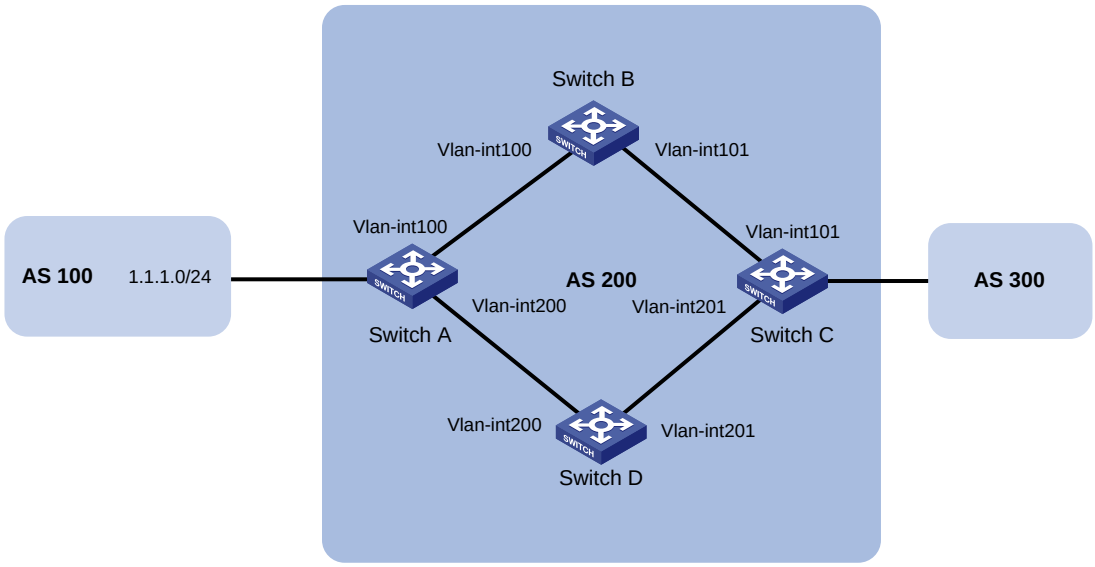


Table 16 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	3.0.1.1/24	Switch C	Vlan-int101	3.0.2.2/24
	Vlan-int200	2.0.1.1/24		Vlan-int201	2.0.2.2/24
Switch B	Vlan-int100	3.0.1.2/24	Switch D	Vlan-int200	2.0.1.2/24
	Vlan-int101	3.0.2.1/24		Vlan-int201	2.0.2.1/24

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure OSPF to ensure that Switch A and Switch C are reachable to each other. (Details not shown.)
3. Configure BGP on Switch A:

   # Establish two IBGP connections to Switch C.

```
<SwitchA> system-view
[SwitchA] bgp 200
[SwitchA-bgp-default] peer 3.0.2.2 as-number 200
[SwitchA-bgp-default] peer 2.0.2.2 as-number 200
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 3.0.2.2 enable
[SwitchA-bgp-default-ipv4] peer 2.0.2.2 enable
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

   # Create IPv4 basic ACL 2000 to permit 1.1.1.0/24 to pass.

```
[SwitchA] acl basic 2000
[SwitchA-acl-ipv4-basic-2000] rule permit source 1.1.1.0 0.0.0.255
[SwitchA-acl-ipv4-basic-2000] quit
```

   # Create two routing policies to set the MED for route 1.1.1.0/24. The policy **apply_med_50** sets the MED to 50, and the policy **apply_med_100** sets the MED to 100.

```
[SwitchA] route-policy apply_med_50 permit node 10
```

```
[SwitchA-route-policy-apply_med_50-10] if-match ip address acl 2000
[SwitchA-route-policy-apply_med_50-10] apply cost 50
[SwitchA-route-policy-apply_med_50-10] quit
[SwitchA] route-policy apply_med_100 permit node 10
[SwitchA-route-policy-apply_med_100-10] if-match ip address acl 2000
[SwitchA-route-policy-apply_med_100-10] apply cost 100
[SwitchA-route-policy-apply_med_100-10] quit
```

Apply routing policy **apply_med_50** to routes outgoing to peer 3.0.2.2, and apply routing policy **apply_med_100** to routes outgoing to peer 2.0.2.2.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 3.0.2.2 route-policy apply_med_50 export
[SwitchA-bgp-default-ipv4] peer 2.0.2.2 route-policy apply_med_100 export
[SwitchA-bgp-default-ipv4] quit
```

Enable BFD for peer 3.0.2.2.

```
[SwitchA-bgp-default] peer 3.0.2.2 bfd
[SwitchA-bgp-default] quit
```

4. Configure BGP on Switch C:

Establish two IBGP connections to Switch A.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 3.0.1.1 as-number 200
[SwitchC-bgp-default] peer 2.0.1.1 as-number 200
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 3.0.1.1 enable
[SwitchC-bgp-default-ipv4] peer 2.0.1.1 enable
[SwitchC-bgp-default-ipv4] quit
```

Enable BFD for peer 3.0.1.1.

```
[SwitchC-bgp-default] peer 3.0.1.1 bfd
[SwitchC-bgp-default] quit
[SwitchC] quit
```

Verifying the configuration

Display detailed BFD session information on Switch C.

```
<SwitchC> display bfd session verbose
 Total Session Num: 1     Up Session Num: 1     Init Mode: Active

 IPv4 Session Working in control packet mode:
      Local Discr: 513                 Remote Discr: 513
        Source IP: 3.0.2.2           Destination IP: 3.0.1.1
    Session State: Up                     Interface: N/A
     Min Tx Inter: 500ms                Act Tx Inter: 500ms
     Min Rx Inter: 500ms                Detect Inter: 2500ms
         Rx Count: 135                      Tx Count: 135
     Connect Type: Indirect           Running Up for: 00:00:58
        Hold Time: 2457ms                 Auth mode: None
      Detect Mode: Async                       Slot: 01
         Protocol: BGP
          Version: 1
```

```
         Diag Info: No Diagnostic
     Hardware mode: Disable
```

The output shows that a BFD session has been established between Switch A and Switch C.

Display BGP peer information on Switch C.

```
<SwitchC> display bgp peer ipv4

 BGP local router ID: 3.3.3.3
 Local AS number: 200
 Total number of peers: 2                 Peers in established state: 2

  * - Dynamically created peer
  ^ - Peer created through link-local address
  Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

  2.0.1.1                200        4        5    0       0 00:01:55 Established
  3.0.1.1                200        4        5    0       0 00:01:52 Established
```

The output shows that Switch C has established two BGP connections with Switch A, and both connections are in Established state.

Display route 1.1.1.0/24 on Switch C.

```
<SwitchC> display ip routing-table 1.1.1.0 24 verbose

Summary count : 1

Destination: 1.1.1.0/24
   Protocol: BGP             Process ID: 0
  SubProtID: 0x1                    Age: 00h00m09s
       Cost: 50              Preference: 255
      IpPre: N/A             QosLocalID: N/A
        Tag: 0                    State: Active Adv
  OrigTblID: 0x1               OrigVrf: default-vrf
    TableID: 0x2                OrigAs: 0
      NibID: 0x15000001         LastAs: 0
     AttrID: 0x1               Neighbor: 3.0.1.1
      Flags: 0x10060        OrigNextHop: 3.0.1.1
      Label: NULL           RealNextHop: 3.0.2.1
    BkLabel: NULL             BkNextHop: N/A
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface101
BkTunnel ID: Invalid        BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

The output shows that Switch C communicates with network 1.1.1.0/24 through the path Switch C<—>Switch B<—>Switch A.

Break down the path Switch C<—>Switch B<—>Switch A and then display route 1.1.1.0/24 on Switch C.

```
<SwitchC> display ip routing-table 1.1.1.0 24 verbose

Summary count : 1

Destination: 1.1.1.0/24
   Protocol: BGP             Process ID: 0
  SubProtID: 0x1                    Age: 00h03m08s
       Cost: 100             Preference: 255
      IpPre: N/A             QosLocalID: N/A
        Tag: 0                    State: Active Adv
  OrigTblID: 0x1                OrigVrf: default-vrf
    TableID: 0x2                 OrigAs: 0
      NibID: 0x15000000          LastAs: 0
     AttrID: 0x0                Neighbor: 2.0.1.1
      Flags: 0x10060        OrigNextHop: 2.0.1.1
      Label: NULL           RealNextHop: 2.0.2.1
    BkLabel: NULL              BkNextHop: N/A
    SRLabel: NULL              BkSRLabel: NULL
   SIDIndex: NULL                InLabel: NULL
  Tunnel ID: Invalid           Interface: Vlan-interface201
BkTunnel ID: Invalid         BkInterface: N/A
   FtnIndex: 0x0            TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0                MicroSegID: 0
RealFIRType: Normal            RealThres: 0
```

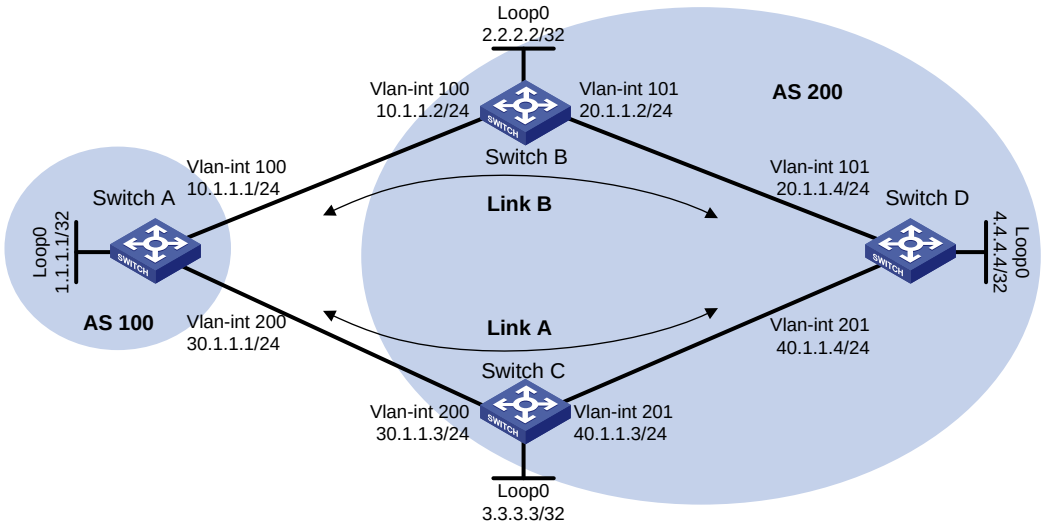
The output shows that Switch C communicates with network 1.1.1.0/24 through the path Switch C<—>Switch D<—>Switch A.

Example: Configuring BGP FRR

Network configuration

As shown in Figure 69, configure BGP FRR so that when Link B fails, BGP uses Link A to forward traffic.

Figure 69 Network diagram



Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure OSPF in AS 200 to ensure connectivity among Switch B, Switch C, and Switch D. (Details not shown.)
3. Configure BGP connections:

   # Configure Switch A to establish EBGP sessions to Switch B and Switch C, and advertise network 1.1.1.1/32.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 10.1.1.2 as-number 200
[SwitchA-bgp-default] peer 30.1.1.3 as-number 200
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 enable
[SwitchA-bgp-default-ipv4] peer 30.1.1.3 enable
[SwitchA-bgp-default-ipv4] network 1.1.1.1 32
```

   # Configure Switch B to establish an EBGP session to Switch A, and an IBGP session to Switch D.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 10.1.1.1 as-number 100
[SwitchB-bgp-default] peer 4.4.4.4 as-number 200
[SwitchB-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 10.1.1.1 enable
[SwitchB-bgp-default-ipv4] peer 4.4.4.4 enable
[SwitchB-bgp-default-ipv4] peer 4.4.4.4 next-hop-local
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

Configure Switch C to establish an EBGP session to Switch A, and an IBGP session to Switch D.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 30.1.1.1 as-number 100
[SwitchC-bgp-default] peer 4.4.4.4 as-number 200
[SwitchC-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 30.1.1.1 enable
[SwitchC-bgp-default-ipv4] peer 4.4.4.4 enable
[SwitchC-bgp-default-ipv4] peer 4.4.4.4 next-hop-local
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

Configure Switch D to establish IBGP sessions to Switch B and Switch C, and advertise network 4.4.4.4/32.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] peer 2.2.2.2 as-number 200
[SwitchD-bgp-default] peer 2.2.2.2 connect-interface loopback 0
[SwitchD-bgp-default] peer 3.3.3.3 as-number 200
[SwitchD-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] peer 2.2.2.2 enable
[SwitchD-bgp-default-ipv4] peer 3.3.3.3 enable
[SwitchD-bgp-default-ipv4] network 4.4.4.4 32
```

4. Configure preferred values so Link B is used to forward traffic between Switch A and Switch D:

Configure Switch A to set the preferred value to 100 for routes received from Switch B.

```
[SwitchA-bgp-default-ipv4] peer 10.1.1.2 preferred-value 100
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

Configure Switch D to set the preferred value to 100 for routes received from Switch B.

```
[SwitchD-bgp-default-ipv4] peer 2.2.2.2 preferred-value 100
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

5. Configure BGP FRR:

On Switch A, create routing policy **frr** to set a backup next hop 30.1.1.3 (Switch C) for the route destined for 4.4.4.4/32.

```
[SwitchA] ip prefix-list abc index 10 permit 4.4.4.4 32
[SwitchA] route-policy frr permit node 10
[SwitchA-route-policy] if-match ip address prefix-list abc
[SwitchA-route-policy] apply fast-reroute backup-nexthop 30.1.1.3
[SwitchA-route-policy] quit
```

Use BFD echo packet mode to detect the connectivity to Switch D.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] primary-path-detect bfdecho
```

Apply the routing policy to BGP FRR for BGP IPv4 unicast address family.

```
[SwitchA-bgp-default] address-family ipv4 unicast
[SwitchA-bgp-default-ipv4] fast-reroute route-policy frr
[SwitchA-bgp-default-ipv4] quit
[SwitchA-bgp-default] quit
```

On Switch D, create routing policy **frr** to set a backup next hop 3.3.3.3 (Switch C) for the route destined for 1.1.1.1/32.

```
[SwitchD] ip prefix-list abc index 10 permit 1.1.1.1 32
[SwitchD] route-policy frr permit node 10
[SwitchD-route-policy] if-match ip address prefix-list abc
[SwitchD-route-policy] apply fast-reroute backup-nexthop 3.3.3.3
[SwitchD-route-policy] quit
```

Use BFD echo packet mode to detect the connectivity to Switch A.

```
[SwitchD] bgp 200
[SwitchD-bgp-default] primary-path-detect bfdecho
```

Apply the routing policy to BGP FRR for BGP IPv4 unicast address family.

```
[SwitchD-bgp-default] address-family ipv4 unicast
[SwitchD-bgp-default-ipv4] fast-reroute route-policy frr
[SwitchD-bgp-default-ipv4] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

Display detailed information about the route to 4.4.4.4/32 on Switch A. The output shows the backup next hop for the route.

```
[SwitchA] display ip routing-table 4.4.4.4 32 verbose

Summary count : 1

Destination: 4.4.4.4/32
   Protocol: BGP             Process ID: 0
  SubProtID: 0x2                    Age: 00h01m52s
       Cost: 0              Preference: 255
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0x2                OrigAs: 200
      NibID: 0x15000003         LastAs: 200
     AttrID: 0x5              Neighbor: 10.1.1.2
      Flags: 0x10060       OrigNextHop: 10.1.1.2
      Label: NULL          RealNextHop: 10.1.1.2
    BkLabel: NULL            BkNextHop: 30.1.1.3
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface 100
BkTunnel ID: Invalid       BkInterface: Vlan-interface 200
   FtnIndex: 0x0          TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal          RealThres: 0
```

Display detailed information about the route to 1.1.1.1/32 on Switch D. The output shows the backup next hop for the route.

```
[SwitchD] display ip routing-table 1.1.1.1 32 verbose

Summary count : 1

Destination: 1.1.1.1/32
   Protocol: BGP             Process ID: 0
  SubProtID: 0x1                    Age: 00h00m36s
       Cost: 0              Preference: 255
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0x2                OrigAs: 100
      NibID: 0x15000003         LastAs: 100
     AttrID: 0x1               Neighbor: 2.2.2.2
      Flags: 0x10060        OrigNextHop: 2.2.2.2
      Label: NULL           RealNextHop: 20.1.1.2
    BkLabel: NULL             BkNextHop: 40.1.1.3
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface 101
BkTunnel ID: Invalid        BkInterface: Vlan-interface 201
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

IPv6 BGP network reliability improvement configuration examples

Example: Configuring BFD for IPv6 BGP

Network configuration

As shown in Figure 70, configure OSPFv3 as the IGP in AS 200.

- Establish two IBGP connections between Switch A and Switch C. When both paths operate correctly, Switch C uses the path Switch A<—>Switch B<—>Switch C to exchange packets with network 1200::0/64.
- Configure BFD over the path. When the path fails, BFD can quickly detect the failure and notify it to IPv6 BGP. Then, the path Switch A<—>Switch D<—>Switch C takes effect immediately.

Figure 70 Network diagram

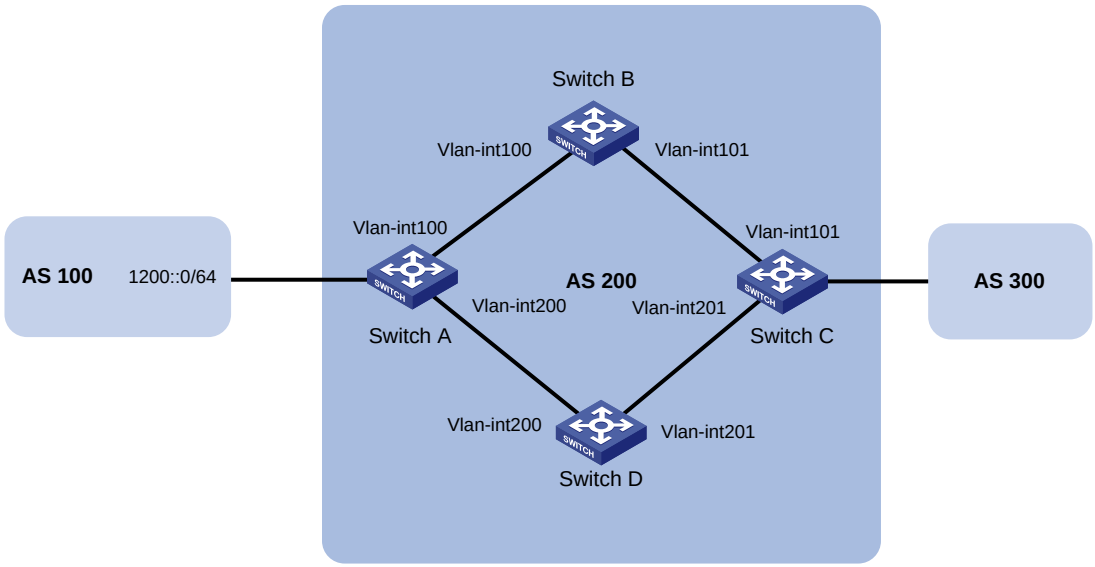


Table 17 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	3000::1/64	Switch C	Vlan-int101	3001::3/64
	Vlan-int200	2000::1/64		Vlan-int201	2001::3/64
Switch B	Vlan-int100	3000::2/64	Switch D	Vlan-int200	2000::2/64
	Vlan-int101	3001::2/64		Vlan-int201	2001::2/64

Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure OSPFv3 so that Switch A and Switch C can reach each other. (Details not shown.)
3. Configure IPv6 BGP on Switch A:

   # Establish two IBGP connections to Switch C.

```
<SwitchA> system-view
[SwitchA] bgp 200
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 3001::3 as-number 200
[SwitchA-bgp-default] peer 2001::3 as-number 200
[SwitchA-bgp-default] address-family ipv6
[SwitchA-bgp-default-ipv6] peer 3001::3 enable
[SwitchA-bgp-default-ipv6] peer 2001::3 enable
[SwitchA-bgp-default-ipv6] quit
```

   # Create IPv6 basic ACL 2000 to permit 1200::0/64 to pass.

```
[SwitchA] acl ipv6 basic 2000
[SwitchA-acl-ipv6-basic-2000] rule permit source 1200:: 64
[SwitchA-acl-ipv6-basic-2000] quit
```

   # Create two routing policies to set the MED for route 1200::0/64. The policy **apply_med_50** sets the MED to 50, and the policy **apply_med_100** sets the MED to 100.

```
[SwitchA] route-policy apply_med_50 permit node 10
```

```
[SwitchA-route-policy-apply_med_50-10] if-match ipv6 address acl 2000
[SwitchA-route-policy-apply_med_50-10] apply cost 50
[SwitchA-route-policy-apply_med_50-10] quit
[SwitchA] route-policy apply_med_100 permit node 10
[SwitchA-route-policy-apply_med_100-10] if-match ipv6 address acl 2000
[SwitchA-route-policy-apply_med_100-10] apply cost 100
[SwitchA-route-policy-apply_med_100-10] quit
```

Apply routing policy **apply_med_50** to routes outgoing to peer 3001::3, and apply routing policy **apply_med_100** to routes outgoing to peer 2001::3.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] address-family ipv6 unicast
[SwitchA-bgp-default-ipv6] peer 3001::3 route-policy apply_med_50 export
[SwitchA-bgp-default-ipv6] peer 2001::3  route-policy apply_med_100 export
[SwitchA-bgp-default-ipv6] quit
```

Enable BFD for peer 3001::3.

```
[SwitchA-bgp-default] peer 3001::3 bfd
[SwitchA-bgp-default] quit
```

4. Configure IPv6 BGP on Switch C:

Establish two IBGP connections to Switch A.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 3000::1 as-number 200
[SwitchC-bgp-default] peer 2000::1 as-number 200
[SwitchC-bgp-default] address-family ipv6
[SwitchC-bgp-default-ipv6] peer 3000::1 enable
[SwitchC-bgp-default-ipv6] peer 2000::1 enable
[SwitchC-bgp-default-ipv6] quit
```

Enable BFD for peer 3001::1.

```
[SwitchC-bgp-default] peer 3000::1 bfd
[SwitchC-bgp-default] quit
[SwitchC] quit
```

Verifying the configuration

Display detailed BFD session information on Switch C.

```
<SwitchC> display bfd session verbose
Total Session Num: 1     Up Session Num: 1     Init Mode: Active

 IPv6 Session Working in control packet mode:
       Local Discr: 513                 Remote Discr: 513
         Source IP: 3001::3
    Destination IP: 3000::1
     Session State: Up                     Interface: N/A
      Min Tx Inter: 500ms                Act Tx Inter: 500ms
      Min Rx Inter: 500ms                Detect Inter: 2500ms
          Rx Count: 13                       Tx Count: 14
      Connect Type: Indirect           Running Up for: 00:00:05
         Hold Time: 2243ms                  Auth mode: None
       Detect Mode: Async                        Slot: 01
```

```
          Protocol: BGP4+
           Version:1
          Diag Info: No Diagnostic
     Hardware mode: Disable
```

The output shows that a BFD session has been established between Switch A and Switch C.

Display BGP peer information on Switch C.

```
<SwitchC> display bgp peer ipv6

 BGP local router ID: 3.3.3.3
 Local AS number: 200
 Total number of peers: 2                 Peers in established state: 2

  * - Dynamically created peer
  ^ - Peer created through link-local address
  Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

  2000::1                200        8        8    0       0 00:04:45 Established
  3000::1                200        5        4    0       0 00:01:53 Established
```

The output shows that Switch C has established two BGP connections with Switch A, and both connections are in Established state.

Display route 1200::0/64 on Switch C.

```
<SwitchC> display ipv6 routing-table 1200::0 64 verbose

Summary count : 1

Destination: 1200::/64
   Protocol: BGP4+           Process ID: 0
  SubProtID: 0x1                    Age: 00h01m07s
       Cost: 50              Preference: 255
      IpPre: N/A             QosLocalID: N/A
        Tag: 0                    State: Active Adv
  OrigTblID: 0x1               OrigVrf: default-vrf
    TableID: 0xa                OrigAs: 0
      NibID: 0x25000001         LastAs: 0
     AttrID: 0x1               Neighbor: 3000::1
      Flags: 0x10060        OrigNextHop: 3000::1
      Label: NULL           RealNextHop: FE80::20C:29FF:FE4A:3873
    BkLabel: NULL             BkNextHop: N/A
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface101
BkTunnel ID: Invalid        BkInterface: N/A
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

The output shows that Switch C communicates with network 1200::0/64 through the path Switch C<—>Switch B<—>Switch A.

Break down the path Switch C<—>Switch B<—>Switch A and then display route 1200::0/64 on Switch C.

```
<SwitchC> display ipv6 routing-table 1200::0 64 verbose

Summary count : 1

Destination: 1200::/64
   Protocol: BGP4+           Process ID: 0
  SubProtID: 0x1                    Age: 00h00m57s
       Cost: 100             Preference: 255
      IpPre: N/A             QosLocalID: N/A
        Tag: 0                    State: Active Adv
  OrigTblID: 0x1                OrigVrf: default-vrf
    TableID: 0xa                 OrigAs: 0
      NibID: 0x25000000          LastAs: 0
     AttrID: 0x0                Neighbor: 2000::1
      Flags: 0x10060         OrigNextHop: 2000::1
      Label: NULL            RealNextHop: FE80::20C:29FF:FE40:715
    BkLabel: NULL              BkNextHop: N/A
    SRLabel: NULL              BkSRLabel: NULL
   SIDIndex: NULL                InLabel: NULL
  Tunnel ID: Invalid           Interface: Vlan-interface201
BkTunnel ID: Invalid         BkInterface: N/A
   FtnIndex: 0x0            TrafficIndex: N/A
  Connector: N/A                  PathID: 0x0
   LinkCost: 0                MicroSegID: 0
RealFIRType: Normal            RealThres: 0
```

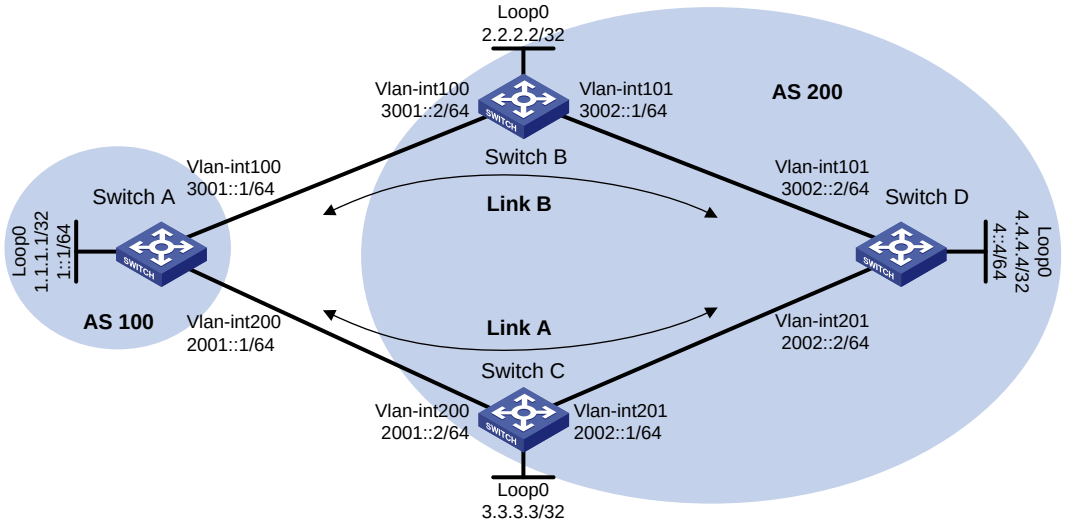
The output shows that Switch C communicates with network 1200::0/64 through the path Switch C<—>Switch D<—>Switch A.

Example: Configuring IPv6 BGP FRR

Network configuration

As shown in Figure 71, configure BGP FRR so that when Link B fails, BGP uses Link A to forward traffic.

Figure 71 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure OSPFv3 in AS 200 to ensure connectivity among Switch B, Switch C, and Switch D. (Details not shown.)
3. Configure BGP connections:

   # Configure Switch A to establish EBGP sessions to Switch B and Switch C, and advertise network 1::/64.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA] router-id 1.1.1.1
[SwitchA-bgp-default] peer 3001::2 as-number 200
[SwitchA-bgp-default] peer 2001::2 as-number 200
[SwitchA-bgp-default] address-family ipv6 unicast
[SwitchA-bgp-default-ipv6] peer 3001::2 enable
[SwitchA-bgp-default-ipv6] peer 2001::2 enable
[SwitchA-bgp-default-ipv6] network 1:: 64
[SwitchA-bgp-default-ipv6] quit
[SwitchA-bgp-default] quit
```

   # Configure Switch B to establish an EBGP session to Switch A, and an IBGP session to Switch D.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB] router-id 2.2.2.2
[SwitchB-bgp-default] peer 3001::1 as-number 100
[SwitchB-bgp-default] peer 3002::2 as-number 200
[SwitchB-bgp-default] address-family ipv6 unicast
[SwitchB-bgp-default-ipv6] peer 3001::1 enable
[SwitchB-bgp-default-ipv6] peer 3002::2 enable
[SwitchB-bgp-default-ipv6] peer 3002::2 next-hop-local
[SwitchB-bgp-default-ipv6] quit
[SwitchB-bgp-default] quit
```

Configure Switch C to establish an EBGP session to Switch A, and an IBGP session to Switch D.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC] router-id 3.3.3.3
[SwitchC-bgp-default] peer 2001::1 as-number 100
[SwitchC-bgp-default] peer 2002::2 as-number 200
[SwitchC-bgp-default] address-family ipv6 unicast
[SwitchC-bgp-default-ipv6] peer 2001::1 enable
[SwitchC-bgp-default-ipv6] peer 2002::2 enable
[SwitchC-bgp-default-ipv6] peer 2002::2 next-hop-local
[SwitchC-bgp-default-ipv6] quit
[SwitchC-bgp-default] quit
```

Configure Switch D to establish IBGP sessions to Switch B and Switch C, and advertise network 4::/64.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] peer 3002::1 as-number 200
[SwitchD-bgp-default] peer 2002::1 as-number 200
[SwitchD-bgp-default] address-family ipv6 unicast
[SwitchD-bgp-default-ipv6] peer 3002::1 enable
[SwitchD-bgp-default-ipv6] peer 2002::1 enable
[SwitchD-bgp-default-ipv6] network 4:: 64
[SwitchD-bgp-default-ipv6] quit
[SwitchD-bgp-default] quit
```

4. Configure preferred values so Link B is used to forward traffic between Switch A and Switch D:

Configure Switch A to set the preferred value to 100 for routes received from Switch B.

```
[SwitchA-bgp-default-ipv6] peer 3001::2 preferred-value 100
[SwitchA-bgp-default-ipv6] quit
[SwitchA-bgp-default] quit
```

Configure Switch D to set the preferred value to 100 for routes received from Switch B.

```
[SwitchD-bgp-default-ipv6] peer 3002::1 preferred-value 100
[SwitchD-bgp-default-ipv6] quit
[SwitchD-bgp-default] quit
```

5. Configure BGP FRR:

On Switch A, create routing policy **frr** to set a backup next hop 2001::2 (Switch C) for the route destined for 4::/64.

```
<SwitchA> system-view
[SwitchA] ipv6 prefix-list abc index 10 permit 4:: 64
[SwitchA] route-policy frr permit node 10
[SwitchA-route-policy] if-match ipv6 address prefix-list abc
[SwitchA-route-policy] apply ipv6 fast-reroute backup-nexthop 2001::2
[SwitchA-route-policy] quit
```

Apply the routing policy to BGP FRR for BGP IPv6 unicast address family.

```
[SwitchA] bgp 100
[SwitchA-bgp-default] address-family ipv6 unicast
[SwitchA-bgp-default-ipv6] fast-reroute route-policy frr
[SwitchA-bgp-default-ipv6] quit
```

```
[SwitchA-bgp-default] quit
```

On Switch D, create routing policy **frr** to set a backup next hop 2002::1 (Switch C) for the route destined for 1::/64.

```
<SwitchD> system-view
[SwitchD] ipv6 prefix-list abc index 10 permit 1:: 64
[SwitchD] route-policy frr permit node 10
[SwitchD-route-policy] if-match ipv6 address prefix-list abc
[SwitchD-route-policy] apply ipv6 fast-reroute backup-nexthop 2002::1
[SwitchD-route-policy] quit
```

Apply the routing policy to BGP FRR for BGP IPv6 unicast address family.

```
[SwitchD] bgp 200
[SwitchD-bgp-default] address-family ipv6 unicast
[SwitchD-bgp-default-ipv6] fast-reroute route-policy frr
[SwitchD-bgp-default-ipv6] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

Display detailed information about the route to 4::/64 on Switch A. The output shows the backup next hop for the route.

```
[SwitchA] display ipv6 routing-table 4:: 64 verbose

Summary count : 1

Destination: 4::/64
   Protocol: BGP4+           Process ID: 0
  SubProtID: 0x2                    Age: 00h00m58s
       Cost: 0              Preference: 255
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0xa                OrigAs: 200
      NibID: 0x25000003         LastAs: 200
     AttrID: 0x3              Neighbor: 3001::2
      Flags: 0x10060        OrigNextHop: 3001::2
      Label: NULL           RealNextHop: 3001::2
    BkLabel: NULL             BkNextHop: 2001::2
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface 100
BkTunnel ID: Invalid        BkInterface: Vlan-interface 200
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Display detailed information about the route to 1::/64 on Switch D. The output shows the backup next hop for the route.

```
[SwitchD] display ipv6 routing-table 1:: 64 verbose

Summary count : 1
```

```
Destination: 1::/64
   Protocol: BGP4+           Process ID: 0
  SubProtID: 0x1                    Age: 00h03m24s
       Cost: 0              Preference: 255
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
    TableID: 0xa                OrigAs: 100
      NibID: 0x25000003         LastAs: 100
     AttrID: 0x4              Neighbor: 3002::1
      Flags: 0x10060       OrigNextHop: 3002::1
      Label: NULL          RealNextHop: 3002::1
    BkLabel: NULL            BkNextHop: 2002::1
    SRLabel: NULL            BkSRLabel: NULL
   SIDIndex: NULL              InLabel: NULL
  Tunnel ID: Invalid         Interface: Vlan-interface 101
BkTunnel ID: Invalid       BkInterface: Vlan-interface 201
   FtnIndex: 0x0          TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal          RealThres: 0
```

Configuring extended BGP features

Extended BGP feature configuration tasks at a glance

To configure extended BGP features, perform the following tasks:

- Configuring BMP
- Configuring BGP LS
  - Configuring basic BGP LS
  - (Optional.) Configuring BGP LS route reflection
  - (Optional.) Specifying an AS number and a router ID for BGP LS messages
  - (Optional.) Performing manual soft-reset for BGP sessions of LS address family
- Configuring BGP dedicated routes
  - Configuring basic BGP dedicated route settings
  - (Optional.) Configuring BGP dedicated route reflection
  - (Optional.) Resetting BGP sessions in dedicated address family
- Configuring 6PE
  - Configuring basic 6PE
  - (Optional.) Controlling path selection for 6PE
  - (Optional.) Controlling 6PE route advertisement and reception
  - (Optional.) Tuning and optimizing 6PE
  - (Optional.) Configuring 6PE route reflection
  - (Optional.) Resetting 6PE connections

Configuring BMP

About this task

BGP monitors only the real-time states of BGP sessions and BGP routes. It cannot directly monitor session state change and route update processes. To resolve this issue, configure the BGP monitoring protocol (BMP) feature. BMP enables a BGP router (BMP client) to send session state information of the specified peers to BMP servers for monitoring. The session status information includes peer relationship establishment and termination as well as routing information.

BMP benefits

- BMP uses TCP as its transmission layer protocol, which ensures the transmission reliability.
- As an extended feature of BGP, BMP collects the required information through a simple work mechanism.
- All BMP settings are configured on the BMP client, which simplifies the configuration and management.
- To implement BMP, the BMP client only needs to send BMP packets to the BMP server. It does not require any BMP packets from the BMP server, which barely affects the current network.

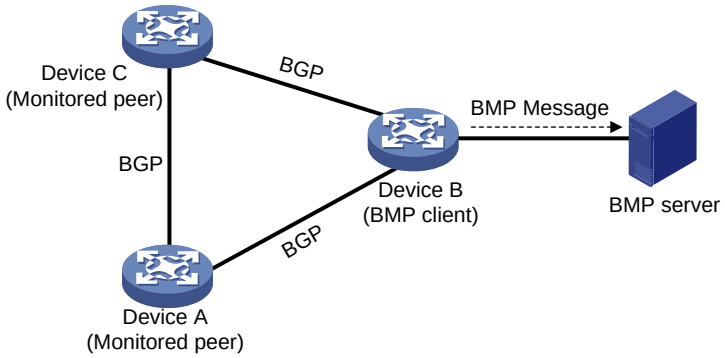
BMP components

BMP contains the following components:

- **Client**—Device configured with the BMP feature. You must specify the IP address and port number of the BMP server on the BMP client to enable them to establish a TCP connection. A BMP client can connect to multiple BMP servers.
- **Server**— Monitoring server.
- **Monitored peer**—Device being monitored. It must be a peer of the BMP client.

As shown in Figure 72, Device A, Device B, and Device C have established a BGP session with one another. You can configure BMP on Device B (BMP client) to enable the BMP server to monitor the running state changes and route updates of Device A and Device C (monitored peers).

Figure 72 BMP components



BMP messages

BMP defines the following messages:

- **Initiation**—After establishing a TCP connection to the BMP server, the BMP client sends an initiation message to the BMP server to report its software version and vendor.
- **Termination**—Reports TCP connection down causes to the BMP server. After sending a termination message to the BMP server, the BMP client terminates the TCP connection and stops sending messages to the BMP server through the TCP connection.
- **Peer Up Notification**—Reports information about BGP sessions between the BMP client and monitored peers to the BMP server. A peer up notification contains information in the open messages exchanged with a monitored peer and TCP connection information. The BMP client will send a peer up notification for each BGP session in established state.
- **Peer Down Notification**—Reports BGP session down events and causes to the BMP server. The BMP client sends a peer down notification when the state of a monitored peer changes from Established to another value.
- **Stats Reports**—Reports BGP running statistics information to the BMP server. The statistics information includes number of routes discarded because of duplicated AS path attributes, number of routes received from peers, and number of duplicated up messages.
- **Route Monitoring**—Reports BGP routes (including initial, updated, and withdrawal routes) exchanged between the BMP client and monitored peers to the BMP server.

BMP mechanism

BMP works as follows:

1. The BMP client establishes a TCP connection with the BMP server.
2. The BMP client sends an initiation message to the BMP server to advertise its information.
3. The BMP client sends a peer up notification and a route monitoring message to the BMP server to advertise the initial state and routing information of the monitored peer.
4. The BMP client sends peer up notifications, peer down notifications, route monitoring messages, and stats reports to the BMP server to report state changes and route updates of the monitored peers.

5. The BMP client sends a termination message to the BMP server and tears down the TCP connection to terminate the monitoring.

BMP configuration tasks at a glance

To configure BMP, perform the following tasks on the BMP client:

- Establishing a TCP connection between the BMP client and BMP server
- Enabling a BMP server to monitor a peer or peer group
- Sending routes to the BMP server
- Configuring the BMP client to send notifications with mode flags to the BMP server
- Sending statistics information to the BMP server

Establishing a TCP connection between the BMP client and BMP server

1. Enter system view.

   system-view
2. Create a BMP server and enter BMP server view.

   bmp server *server-number*
3. Configure an IP address and port number for the BMP server.

   server address { *ipv4-address* | *ipv6-address* } **port** *port-number*

   By default, no IP address and port number are configured for the BMP server.
4. (Optional.) Specify a VPN instance for the BMP server.

   server vpn-instance *vpn-instance-name*

   By default, no VPN instance is specified for a BMP server. A BMP server belongs to the public network.
5. (Optional.) Specify the source interface of TCP connections to the BMP server.

   server connect-interface *interface-type interface-number*

   By default, BGP uses the primary IPv4 or IPv6 address of the output interface in the optimal route to the BMP server as the source address of TCP connections to the BMP server.

   For a BMP server, this command does not take effect if the VPN instance of the specified interface is different from that specified by the **server vpn-instance** command.
6. (Optional.) Specify the source address of TCP connections to the BMP server.

   server source-address { *ipv4-address* | *ipv6-address* }

   By default, BGP uses the primary IPv4 or IPv6 address of the output interface in the optimal route to the BMP server as the source address of TCP connections to the BMP server.

   If you execute both this command and the **server connect-interface** command for a BMP server, the configuration of this command takes effect.
7. (Optional.) Specify the authentication mode and key for the local device to establish TCP connections to the BMP server.

   server password { **keychain** *keychain-name* | **md5** { **cipher** | **simple** } *string* }

   By default, the local device establishes TCP connections to the BMP server without authentication.

Enabling a BMP server to monitor a peer or peer group

About this task

You can execute one of the following commands to enable a BMP server to monitor BGP peers:

- Execute the **peer bmp server** command to enable the specified BMP server to monitor the specified peer or peer group.
- Execute the **bmp server monitor current-instance** command to enable the specified BMP server to monitor all peers in the current instance.
- Execute the **bmp server monitor all-vpn-instance** command to enable the specified BMP server to monitor all peers in all BGP-VPN instances.

For a BGP peer, the following rules apply:

- The BMP server specified by the **peer bmp server** command takes precedence over that specified by the **bmp server monitor current-instance** command.
- The BMP server specified by the **bmp server monitor current-instance** command takes precedence over that specified by the **bmp server monitor all-vpn-instance** command.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*
3. Enable a BMP server to monitor a peer or peer group. Choose the options to configure as needed:
   - Enable the specified BMP server to monitor the specified peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | *link-local-address* **interface** *interface-type interface-number* } **bmp server** *server-number-list*

     interface-peer *interface-type interface-number* **bmp server** *server-number-list*

     By default, a BMP server cannot monitor a peer or peer group.
   - Enable the specified BMP server to monitor all peers in the current instance.

     bmp server *server-number* **monitor current-instance** [**route-mode** { **adj-rib-in** { **pre-policy** | **post-policy** | **both** } | **adj-rib-out** { **pre-policy** | **post-policy** | **both** } } *]

     By default, a BMP server cannot monitor the peers in the current instance.
   - Enable the specified BMP server to monitor all peers in all BGP-VPN instances.

     bmp server *server-number* **monitor all-vpn-instance** [**route-mode** { **adj-rib-in** { **pre-policy** | **post-policy** | **both** } | **adj-rib-out** { **pre-policy** | **post-policy** | **both** } } *]

     By default, a BMP server cannot monitor the peers in a BGP-VPN instance.

     This command is not supported in BGP-VPN instance view.

Sending routes to the BMP server

About this task

For a peer or peer group, the following rules apply when BGP determines the type of routes to send to the BMP server:

- The route type specified by the **peer route-mode** command takes precedence over that specified by the **bmp server monitor current-instance route-mode** command.
- The route type specified by the **bmp server monitor current-instance route-mode** command takes precedence over that specified by the **bmp server monitor all-vpn-instance route-mode** command.
- The route type specified by the **bmp server monitor all-vpn-instance route-mode** command takes precedence over that specified by the **route-mode adj-rib-in** or **route-mode adj-rib-out** command.

You can execute the **route-mode loc-rib** command to configure BGP to send the optimal routes in the routing table to the BMP server.

Procedure

1. Enter system view.

   system-view
2. Enter BMP server view.

   bmp server *server-number*
3. (Optional.) Enable BGP to send routes to the BMP server globally. Choose the following options to configure as needed:
   - Enable BGP to send routes advertised to all the monitored peers and peer groups to the BMP server.

     route-mode adj-rib-out [**pre-policy** | **post-policy** | **both**]

     By default, BGP does not send routes advertised to a monitored peer or peer group to the BMP server.
   - Enable BGP to send routes received from all the monitored peers and peer groups to the BMP server.

     route-mode adj-rib-in [**pre-policy** | **post-policy** | **both**]

     By default, BGP does not send routes received from a monitored peer or peer group to the BMP server.
   - Configure BGP to send the optimal routes in the routing table to the BMP server.

     route-mode loc-rib

     By default, BGP does not send the optimal routes in the routing table to the BMP server.

     With this feature configured, BGP sends all routes to the BMP server first. If routes update later, BGP sends only the optimal routes to the server.
4. Return to system view.

   quit
5. Enter BGP instance view or BGP-VPN instance view.
   - Enter BGP instance view.

     bgp *as-number* [**instance** *instance-name*]
   - Execute the following commands in sequence to enter BGP-VPN instance view:

     bgp *as-number* [**instance** *instance-name*]

ip vpn-instance *vpn-instance-name*

6. Enable BGP to send routes exchanged with a monitored peer or peer group to the BMP server. Choose the following options to configure as needed:
   - Enable BGP to send routes exchanged with the specified monitored peer or peer group to the BMP server.

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-mode** { **adj-rib-in** { **pre-policy** | **post-policy** | **both** } | **adj-rib-out** { **pre-policy** | **post-policy** | **both** } } *
   - Enable BGP to send routes exchanged with all the monitored peers and peer groups in the current instance to the BMP server.

     bmp server *server-number* **monitor current-instance route-mode** { **adj-rib-in** { **pre-policy** | **post-policy** | **both** } | **adj-rib-out** { **pre-policy** | **post-policy** | **both** } } *
   - Enable BGP to send routes exchanged with all the monitored peers and peer groups in all BGP-VPN instances to the BMP server.

     bmp server *server-number* **monitor all-vpn-instance route-mode** { **adj-rib-in** { **pre-policy** | **post-policy** | **both** } | **adj-rib-out** { **pre-policy** | **post-policy** | **both** } } *

     This command is not supported in BGP-VPN instance view.

   By default, BGP determines whether to send routes exchanged with a peer or peer group to the BMP server based on the following configurations:
   - Configuration of the **route-mode adj-rib-in** command in BMP server view.
   - Configuration of the **route-mode adj-rib-out** command in BMP server view.

Configuring the BMP client to send notifications with mode flags to the BMP server

About this task

Perform this task to configure the BMP client to send peer up/down notifications with the following flags:

- **Adj-RIB-In**—The BMP client sends routes received from monitored peers to the BMP server.
- **Adj-RIB-Out**—The BMP client sends routes advertised to monitored peers to the BMP server.
- **pre-policy**—The BMP client sends routes to the BMP server without filtering.
- **post-policy**—The BMP client sends routes to the BMP server after filtering.

By default, the BMP client sends peer up/down notifications with the Adj-RIB-In and pre-policy flags to the BMP server after you configure this feature.

- After you add a route sending mode by executing the **route-mode adj-rib-in**, **route-mode adj-rib-out**, or **peer route-mode** command, the BMP client will send peer up notifications with the new mode flag to the BMP server.
- After you delete the route sending mode by executing the **route-mode adj-rib-in**, **route-mode adj-rib-out**, or **peer route-mode** command, the BMP client will send peer down notifications with the new mode flag to the BMP server.

For example, if you execute the **route-mode adj-rib-out post-policy** command after executing the **route-mode adj-rib-out pre-policy** command, the BMP client sends routes to the BMP server after filtering. The BMP client will send both peer up and peer down notifications with the new mode flags to the BMP server.

Procedure

1. Enter system view.
   system-view
2. Enter BMP server view.
   bmp server *server-number*
3. Configure the BMP client to send notifications with mode flags to the BMP server. Choose a minimum of one task.
   - Configure the BMP client to send peer up notifications with mode flags to the BMP server.
     pu-monitor-mode enable
   - Configure the BMP client to send peer down notifications with mode flags to the BMP server.
     pd-monitor-mode enable

   By default, the BMP client sends peer up/down notifications without mode flags to the BMP server.

Sending statistics information to the BMP server

About this task

The BMP client uses stats reports to report statistics information to the BMP server. Perform this task to set the interval at which the BMP client sends stats reports to the BMP server.

Procedure

1. Enter system view.
   system-view
2. Create a BMP server and enter BMP server view.
   bmp server *server-number*
3. Set the interval at which the BMP client sends statistics information to the BMP server.
   statistics-interval *value*

   By default, the BMP client does not send statistics information to the BMP server.

Configuring BGP LS

About BGP LS

Traditional BGP focuses on the propagation of routing information and lacks detailed understanding of network topology and link conditions. BGP Link State (LS) is a feature that collects and reports network topology information and link state information to the controller. This feature helps network administrators better understand and manage network links. Although IGPs can report network topology information and link state information to the controller, they have the following disadvantages:

- The controllers need to support those IGPs and their algorithms, which requires higher compute performance.
- IGPs cannot collect information across different domains.
- Different IGPs report information to the controller separately. It is complex for the controller to process the reported information.

With BGP LS configured, BGP can report the link state and topology information collected by IGPs to the controller, regardless of the IGP capabilities of the controller. In addition, BGP supports

cross-domain deployment. All BGP LS-enabled devices can maintain a complete link state database with the route advertisement mechanism of BGP. The topology information reported to the controller will also be complete.

The link state and topology information collected through BGP LS helps achieve the following purposes:

- **Topology maintenance and calculation**: This information can be used for calculating the shortest paths and building network maps.
- **Path selection optimization**: Network administrators can optimize path selection results based on this information.
- **Link failure detection**: Network administrators can quickly detect link failures by using this information.
- **Support for advanced applications**: This information provides support for multiple tasks, such as traffic engineering, network planning, and resource allocation.

BGP LS extensions

BGP places network topology and link state information in a series of NLRIs, and adds those NLRIs into the MP_REACH_NLRI and MP_UNREACH_NLRI attributes of BGP UPDATE messages. This approach allows BGP LS peers to exchange topology and link state information.

Configuring basic BGP LS

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Specify an AS number for an LS peer or peer group.

   peer { *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **as-number** *as-number*

   By default, no AS number is specified.

4. Create the BGP LS address family and enter its view.

   address-family link-state

5. Enable the device to exchange LS information with the peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **enable**

   By default, the device cannot exchange LS information with the peer or peer group.

6. (Optional.) Set the optimal route selection delay timer for BGP LS routes. Choose one of the following option:

   - Enable optimal route selection delay for all routes in the address family.

     route-select delay *delay-value*

     With this command configured, the device delays optimal route selection as long as the address family has a route change.

   - Enable optimal route selection delay based on DOWN-to-UP peer state changes.

     route-select suppress on-peer-up *milliseconds*

     With this command configured, when the state of a peer changes from DOWN to UP, the device performs optimal route selection for BGP routes received from that peer only after the delay timer expires.

   By default, the device does not delay optimal route selection.

Configuring BGP LS route reflection

About this task

Perform this task to configure a BGP route reflector and its clients. The route reflector and its clients automatically form a cluster identified by the router ID of the route reflector. The route reflector forwards route updates among its clients.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP LS address family view.
   address-family link-state
4. Configure BGP LS route reflection.
   - Configure the device as a route reflector and specify a peer or peer group as its client.
     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **reflect-client**
     By default, no route reflector or client is configured.
   - (Optional.) Enable route reflection between clients.
     reflect between-clients
     By default, route reflection between clients is enabled.
     This command can reduce the number of IBGP connections in an AS.
   - (Optional.) Configure the cluster ID of the route reflector.
     reflector cluster-id { *cluster-id* | *ipv4-address* }
     By default, a route reflector uses its own router ID as the cluster ID.

Specifying an AS number and a router ID for BGP LS messages

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP LS address family view.
   address-family link-state
4. Specify an AS number and a router ID for BGP LS messages.
   domain-distinguisher *as-number:router-id*
   By default, the AS number and router ID of the current BGP process are used.
   Configure this command to ensure that LS messages sent by devices in the same AS have the same AS number and router ID.

Performing manual soft-reset for BGP sessions of LS address family

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP route refresh.
   - Enable BGP route refresh for a peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **capability-advertise route-refresh**
   - Enable the BGP route refresh, multi-protocol extension, and 4-byte AS number features for a peer or peer group.

     undo peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **capability-advertise conventional**

   By default, the BGP route refresh, multi-protocol extension, and 4-byte AS number features are enabled.
4. Perform manual soft-reset for BGP sessions of LS address family:
   a. Return to system view.

      quit

   b. Return to user view.

      quit

   c. Perform manual soft-reset for BGP sessions of LS address family.

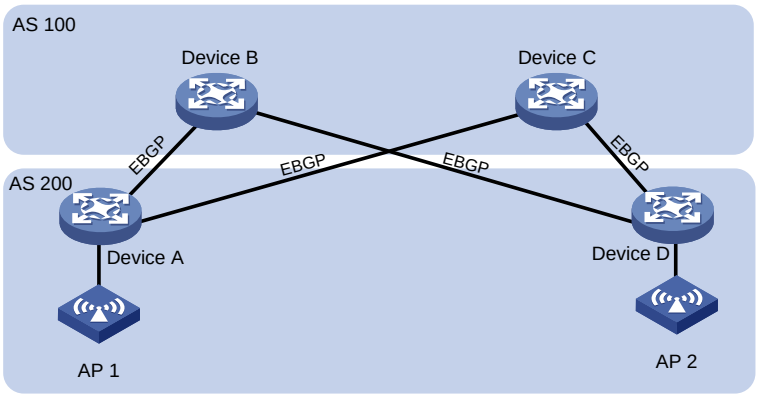
      refresh bgp [**instance** *instance-name*] { *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **link-state**

Configuring BGP dedicated routes

About BGP dedicated routes

As shown in Figure 73, Device B and Device C are EBGP peers of both Device A and Device D, and the EBGP peers can exchange dedicated routes. Both Device A and Device D are enabled with IPSG. When a client accesses Device A through AP 1, Device A generates an IPSG binding for the client based on the DHCP, DHCPv6, or ND snooping entry. Then, Device B and Device C synchronize the IPSG binding of the client through BGP dedicated routes, and they synchronize the IPSG binding to Device D. When the client migrates from Device A to Device D, the client can quickly come online based on the synchronized IPSG binding. For more information about IPSG and IPSG bindings, see IP source guard configuration in *Security Configuration Guide*.

Figure 73 Network diagram



Configuring basic BGP dedicated route settings

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Specify an AS number for a peer or peer group.

   peer { *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **as-number** *as-number*

   By default, no AS number is specified for a peer or peer group.

4. Create the BGP dedicated address family and enter its view.

   address-family dedicated

5. Enable BGP to exchange dedicated routes with a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **enable**

   By default, BGP cannot exchange dedicated routes with a peer or peer group.

6. (Optional.) Specify the device as the next hop for routes sent to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **next-hop-local**

   By default, the device sets itself as the next hop for routes sent to an EBGP peer or peer group. It does not set itself as the next hop for routes sent to an IBGP peer or peer group.

7. (Optional.) Permit the local AS number to appear in routes from a peer or peer group and set the appearance times.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **allow-as-loop** [*number*]

   By default, the local AS number is not allowed in routes from a peer or peer group.

Configuring BGP dedicated route reflection

About this task

Perform this task to configure a BGP route reflector and its clients. The route reflector and its clients automatically form a cluster identified by the router ID of the route reflector. The route reflector forwards route updates among its clients.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP dedicated address family view.
   address-family dedicated
4. Configure the device as a route reflector and specify a peer or peer group as its client.
   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **reflect-client**
   By default, no route reflector or client is configured.
5. (Optional.) Enable route reflection between clients.
   reflect between-clients
   By default, route reflection between clients is enabled.
   This command can reduce the number of IBGP connections in an AS.
6. (Optional.) Configure the cluster ID of the route reflector.
   reflector cluster-id { *cluster-id* | *ipv4-address* }
   By default, a route reflector uses its own router ID as the cluster ID.

Resetting BGP sessions in dedicated address family

Execute one command in user view.

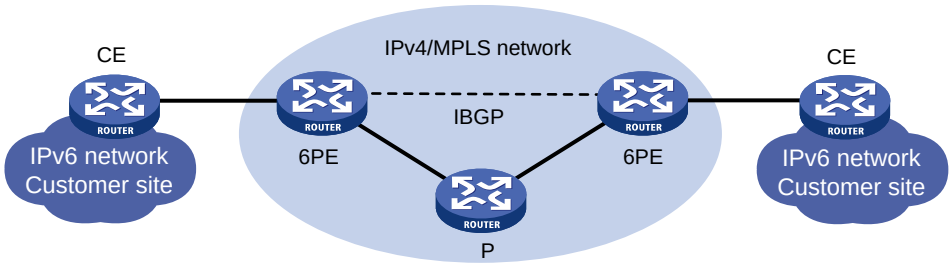
- Perform manual soft-reset on BGP sessions of the dedicated address family.
  refresh bgp [**instance** *instance-name*] { *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **dedicated**
- Reset BGP sessions of the dedicated address family.
  reset bgp [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } **dedicated**

Configuring 6PE

About 6PE

IPv6 provider edge (6PE) is a transition technology that uses MPLS to connect sparsely populated IPv6 networks through an existing IPv4 backbone network. It is an efficient solution for ISP IPv4/MPLS networks to provide IPv6 traffic switching capability.

Figure 74 Network diagram for 6PE



6PE mainly performs the following operations:

- 6PE assigns a label to IPv6 routing information received from a CE router, and sends the labeled IPv6 routing information to the peer 6PE device through an MP-BGP session. The peer 6PE device then forwards the IPv6 routing information to the attached customer site.
- 6PE provides tunnels over the IPv4 backbone so the IPv4 backbone can forward packets for IPv6 networks. The tunnels can be GRE tunnels or MPLS LSPs.
- Upon receiving an IPv6 packet, 6PE adds an inner tag (corresponding to the IPv6 packet) and then an outer tag (corresponding to the public network tunnel) to the IPv6 packet. Devices in the IPv4 backbone network forwards the packet based on the outer tag. When the peer 6PE device receives the packet, it removes the outer and inner tags and forwards the original IPv6 packet to the attached customer site.

To implement exchange of IPv6 routing information, you can configure IPv6 static routing, an IPv6 IGP protocol, or IPv6 BGP between CE and 6PE devices.

For more information about MPLS, CE, and P, see *MPLS Configuration Guide*. For more information about GRE, see *Layer 3—IP Services Configuration Guide*.

Prerequisites

Before you configure 6PE, perform the following tasks:

- Establish tunnels in the IPv4 backbone network (see *Layer 3—IP Services Configuration Guide*).
- Configure basic MPLS on 6PE devices (see *MPLS Configuration Guide*).
- Configure BGP on 6PE devices so that they can advertise tagged IPv6 routing information through BGP sessions. The following tasks describe only BGP configurations on 6PE devices.

Configuring basic 6PE

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Specify a 6PE peer or peer group and its AS number.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
   By default, no 6PE peer is specified.
4. Enter BGP IPv6 unicast address family view.
   address-family ipv6 [**unicast**]
5. Enable BGP to exchange IPv6 unicast routing information with the 6PE peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**

By default, BGP cannot exchange IPv6 unicast routing information with a 6PE peer or peer group.

6. Enable BGP to exchange labeled IPv6 routes with the 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **label-route-capability**

   By default, BGP cannot exchange labeled IPv6 routes with a 6PE peer or peer group.

Controlling path selection for 6PE

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP IPv6 unicast address family view.

   address-family ipv6 [**unicast**]

4. Advertise COMMUNITY attribute to a 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise-community**

   By default, the COMMUNITY attribute is not advertised.

5. Advertise extended community attribute to a 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise-ext-community**

   By default, the extended community attribute is not advertised.

6. Remove private AS numbers in BGP updates sent to a 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **public-as-only** [{ **force** | **limited** } [**replace**] [**include-peer-as**]]

   By default, BGP updates sent to a 6PE peer or peer group can carry both public and private AS numbers.

7. Specify a preferred value for routes received from a 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **preferred-value** *value*

   By default, the preferred value for routes received from a 6PE peer or peer group is 0.

Controlling 6PE route advertisement and reception

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP IPv6 unicast address family view.

   address-family ipv6 [**unicast**]

4. Advertise a default route to a 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **default-route-advertise** [**route-policy** *route-policy-name*]

   By default, no default route is advertised to a 6PE peer or peer group.

5. Configure route filtering policies.

- Specify an AS path list to filter routes advertised to or received from a 6PE peer or peer group.

  peer { *group-name* | *ipv4-address* [*mask-length*] } **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } { **export** | **import** }

  By default, no AS path list is specified.
- Specify an IPv6 ACL to filter routes advertised to or received from a 6PE peer or peer group.

  peer { *group-name* | *ipv4-address* [*mask-length*] } **filter-policy** { *ipv6-acl-number* | **name** *ipv6-acl-name* } { **export** | **import** }

  By default, no IPv6 ACL is specified.
- Specify an IPv6 prefix list to filter routes advertised to or received from a 6PE peer or peer group.

  peer { *group-name* | *ipv4-address* [*mask-length*] } **prefix-list** *ipv6-prefix-list-name* { **export** | **import** }

  By default, no IPv6 prefix list is specified.
- Specify a routing policy to filter routes advertised to or received from a 6PE peer or peer group.

  peer { *group-name* | *ipv4-address* [*mask-length*] } **route-policy** *route-policy-name* { **export** | **import** }

  By default, no routing policy is specified.

6. Save all routes from a 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **keep-all-routes**

   By default, routes from a 6PE peer or peer group are not saved.
7. Set the maximum number of routes that BGP can receive from a 6PE peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **route-limit** *prefix-number* [{ **alert-only** | **discard** | **reconnect** *reconnect-time* } | *percentage-value*] *

   By default, the number of routes that BGP can receive from a 6PE peer or peer group is not limited.
8. Configure the SoO attribute for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **soo** *site-of-origin*

   By default, no SoO attribute is configured for a peer or peer group.

Tuning and optimizing 6PE

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP IPv6 unicast address family view.

   address-family ipv6 [**unicast**]
4. Allow the local AS number to appear in routes from a 6PE peer or peer group and specify the repeat times.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **allow-as-loop** [*number*]

   By default, the local AS number is not allowed to appear in routes from a 6PE peer or peer group.

Configuring 6PE route reflection

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP IPv6 unicast address family view.
   address-family ipv6 [**unicast**]
4. Configure the device as a route reflector and a 6PE peer or peer group as a client.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **reflect-client**
   By default, no route reflector or client is configured.

Resetting 6PE connections

To reset 6PE connections, execute one of the following commands in user view:

- Soft-reset a BGP 6PE connection.
  refresh bgp [**instance** *instance-name*] *ipv4-address* [*mask-length*] { **export** | **import** } **ipv6** [**unicast**]
- Reset a BGP 6PE connection.
  reset bgp [**instance** *instance-name*] *ipv4-address* [*mask-length*] **ipv6** [**unicast**]

CAUTION:
A reset operation tears down BGP sessions for a short period of time.

Display and maintenance commands for extended BGP features

Displaying BGP

Execute **display** commands in any view.

Displaying BGP (IPv4 unicast address family)

Task	Command
Display BMP server information.	**display bgp** [**instance** *instance-name*] **bmp server** *server-number*
Display information about BGP peers monitored by the specified BMP server for the specified BGP instance.	**display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer all** **display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer** [**vpn-instance** *vpn-instance-name*] **ipv4** { *ipv4-address* | *ipv6-address* }

Displaying BGP (IPv6 unicast address family)

Task	Command
Display information about BGP peers monitored by the specified BMP server for the specified BGP instance.	**display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer all** **display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer** [**vpn-instance** *vpn-instance-name*] **ipv6** { *ipv4-address* | *ipv6-address* }

Displaying BGP (IPv4 multicast address family)

Task	Command
Display BMP server information.	**display bgp** [**instance** *instance-name*] **bmp server** *server-number*

Displaying BGP (LS address family)

Task	Command
Display BGP LS peer group information.	**display bgp** [**instance** *instance-name*] **group link-state** [**group-name** *group-name*]
Display BGP LS information.	**display bgp** [**instance** *instance-name*] **link-state** [*ls-prefix* [**advertise-info** | **as-path** | **cluster-list**] | **peer** { *ipv4-address* | *ipv6-address* } { **advertised** | **received** } [**statistics**] | **statistics**]
Display BGP LS peer or peer group information.	**display bgp** [**instance** *instance-name*] **peer link-state** [*ipv4-address mask-length* | *ipv6-address prefix-length* | { *ipv4-address* | *ipv6-address* | **group-name** *group-name* } **log-info** | [*ipv4-address* | *ipv6-address*] **verbose**]
Display BGP LS address family update group information.	**display bgp** [**instance** *instance-name*] **update-group link-state** [*ipv4-address* | *ipv6-address*]

Displaying BGP (dedicated address family)

Task	Command
Display BGP dedicated peer group information.	**display bgp** [**instance** *instance-name*] **group dedicated** [**group-name** *group-name*]
Display BGP dedicated peer or peer group information.	**display bgp** [**instance** *instance-name*] **peer dedicated** [{ *ipv4-address mask-length* | *ipv6-address prefix-length* } | { *ipv4-address* | *ipv6-address* | **group-name** *group-name* } **log-info** | [*ipv4-address* | *ipv6-address*] **verbose**]

Display BGP dedicated address family update group information.	**display bgp** [**instance** *instance-name*] **update-group dedicated** [*ipv4-address* | *ipv6-address*]

Displaying BGP (VPNv4 address family)

Task	Command
Display information about BGP peers monitored by the specified BMP server for the specified BGP instance.	**display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer all** **display bgp** [**instance** *instance-name*] **server** *server-number* **monitor-peer vpnv4** { *ipv4-address* | *ipv6-address* } **display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer vpn-instance** *vpn-instance-name* **vpnv4** *ipv4-address*

Displaying BGP (EVPN address family)

Task	Command
Display information about BGP peers monitored by the specified BMP server for the specified BGP instance.	**display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer all** **display bgp** [**instance** *instance-name*] **bmp server** *server-number* **monitor-peer l2vpn evpn** { *ipv4-address* | *ipv6-address* }

Resetting BGP sessions

CAUTION:
A reset operation tears down BGP sessions for a short period of time.

Execute **reset** commands in user view.

Task	Command
Reset BGP sessions for LS address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } **link-state**
Reset BGP sessions for dedicated address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } **dedicated**
Perform manual soft-reset on BGP sessions of the dedicated address family.	**refresh bgp** [**instance** *instance-name*] { *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **dedicated**

Clearing BGP information

Execute **reset** commands in user view.

Task	Command
Reset the connection to a BMP server and clear statistics information.	**reset bgp** [**instance** *instance-name*] **bmp server** { *server-number* | **all** }
Clear BMP server statistics.	**reset bgp** [**instance** *instance-name*] **bmp server** *server-number* **statistics**

Extended BGP feature configuration examples

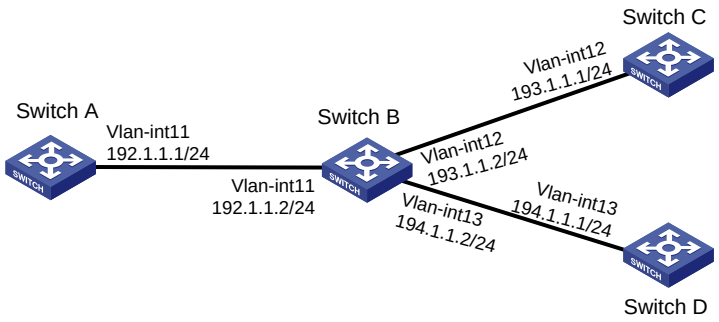
Example: Configuring BGP LS

Network configuration

As shown in Figure 75, all switches run BGP. Run IBGP between Switch A and Switch B, between Switch B and Switch C, and between Switch B and Switch D.

Configure Switch B as a route reflector with client Switch A to allow Switch A to learn LS information advertised by Switch C and Switch D.

Figure 75 Network diagram



Procedure

1. Configure IP addresses for interfaces and configure OSPF on Switch C and Switch D. (Details not shown.)
2. Configure BGP connections:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] bgp 100
   [SwitchA-bgp-default] peer 192.1.1.2 as-number 100
   [SwitchA-bgp-default] address-family link-state
   [SwitchA-bgp-default-ls] peer 192.1.1.2 enable
   [SwitchA-bgp-default-ls] quit
   [SwitchA-bgp-default] quit
   ```

   # Configure Switch B.

   ```
   <SwitchB> system-view
   ```

```
[SwitchB] bgp 100
[SwitchB-bgp-default] peer 192.1.1.1 as-number 100
[SwitchB-bgp-default] peer 193.1.1.1 as-number 100
[SwitchB-bgp-default] peer 194.1.1.1 as-number 100
[SwitchB-bgp-default] address-family link-state
[SwitchB-bgp-default-ls] peer 192.1.1.1 enable
[SwitchB-bgp-default-ls] peer 193.1.1.1 enable
[SwitchB-bgp-default-ls] peer 194.1.1.1 enable
[SwitchB-bgp-default-ls] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 100
[SwitchC-bgp-default] peer 193.1.1.2 as-number 100
[SwitchC-bgp-default] address-family link-state
[SwitchC-bgp-default-ls] peer 193.1.1.2 enable
[SwitchC-bgp-default-ls] quit
[SwitchC-bgp-default] quit
[SwitchC] ospf
[SwitchC-ospf-1] distribute bgp-ls
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 0.0.0.0 0.0.0.0
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 100
[SwitchD-bgp-default] peer 194.1.1.2 as-number 100
[SwitchD-bgp-default] address-family link-state
[SwitchD-bgp-default-ls] peer 194.1.1.2 enable
[SwitchD-bgp-default-ls] quit
[SwitchD-bgp-default] quit
[SwitchD] ospf
[SwitchD-ospf-1] distribute bgp-ls
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 0.0.0.0 0.0.0.0
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

3. Configure Switch B as the route reflector.

```
[SwitchB] bgp 100
[SwitchB-bgp-default] address-family link-state
[SwitchB-bgp-default-ls] peer 192.1.1.1 reflect-client
[SwitchB-bgp-default-ls] quit
[SwitchB-bgp-default] quit
```

Verifying the configuration

Verify that Switch A has learned LS information advertised by Switch C and Switch D.

```
[SwitchA] display bgp link-state
```

```
 Total number of routes: 4

 BGP local Switch ID is 192.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete
 Prefix codes: E link, V node, T IP reachable route, u/U unknown,
               I Identifier, N local node, R remote node, L link, P prefix,
               L1/L2 ISIS level-1/level-2, O OSPF, D direct, S static, B BGP
               a area-ID, , l link-ID, t topology-ID, s ISO-ID,
               c confed-ID/ASN, b bgp-identifier, r Switch-ID,
               i if-address, n peer-address, o OSPF Route-type, p IP-prefix
               d designated Switch address
i Network : [V][O][I0x0][N[c100][b193.1.1.1][a0.0.0.0][r193.1.1.1]]/376
    NextHop : 193.1.1.1                               LocPrf    : 100
    PrefVal : 0                                       OutLabel  : NULL
    MED     :
    Path/Ogn: i

i Network : [V][O][I0x0][N[c100][b194.1.1.1][a0.0.0.0][r194.1.1.1]]/376
    NextHop : 194.1.1.1                               LocPrf    : 100
    PrefVal : 0                                       OutLabel  : NULL
    MED     :
    Path/Ogn: i

i Network :
[T][O][I0x0][N[c100][b193.1.1.1][a0.0.0.0][r193.1.1.1]][P[o0x1][p193.1.1.0/24]]/
480
    NextHop : 193.1.1.1                               LocPrf    : 100
    PrefVal : 0                                       OutLabel  : NULL
    MED     :
    Path/Ogn: i

i Network :
[T][O][I0x0][N[c100][b194.1.1.1][a0.0.0.0][r194.1.1.1]][P[o0x1][p194.1.1.0/24]]/
480
    NextHop : 194.1.1.1                               LocPrf    : 100
    PrefVal : 0                                       OutLabel  : NULL
    MED     :
    Path/Ogn: i
```

Example: Configuring BGP dedicated routes

Network configuration

As shown in Figure 76, all switches run BGP. Configure Switch B and Switch C as EBGP peers of both Switch A and Switch D. Switch A and Switch D provide access services to clients through AP 1 and AP 2. To ensure that a client can quickly come online on the access devices after migration, perform the following tasks:

- Configure IPSG on Switch A and Switch D to enable Switch A and Switch D to generate an IPSG binding for the client after the client comes online.
- Enable BGP to exchange dedicated routes with its peers and peer groups. Then, Switch B and Switch C can synchronize IPSG bindings from Switch A and Switch D, and Switch A and Switch D can withdraw expired IPSG bindings.

Figure 76 Network diagram

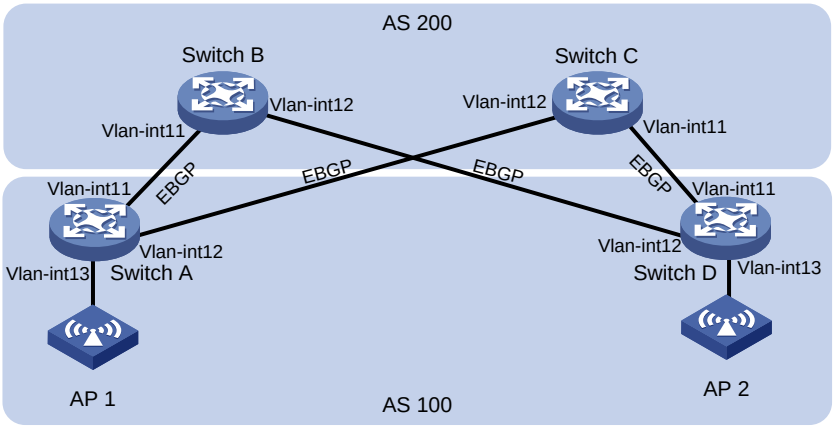


Table 18 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int11	11.11.11.11/24	Switch C	Vlan-int12	22.22.22.22/24
	Vlan-int12	22.22.22.11/24		Vlan-int11	44.44.44.11/24
Switch B	Vlan-int11	11.11.11.22/24	Switch D	Vlan-int11	44.44.44.22/24
	Vlan-int12	33.33.33.11/24		Vlan-int12	33.33.33.22/24

Procedure

1. Configure IP addresses for interfaces. (Details not shown.)
2. Configure BGP connections:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] peer 11.11.11.22 as-number 200
[SwitchA-bgp-default] peer 22.22.22.22 as-number 200
[SwitchA-bgp-default] address-family dedicated
[SwitchA-bgp-default-dedicated] peer 11.11.11.22 enable
[SwitchA-bgp-default-dedicated] peer 22.22.22.22 enable
[SwitchA-bgp-default-dedicated] quit
[SwitchA-bgp-default] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 11.11.11.11 as-number 100
[SwitchB-bgp-default] peer 33.33.33.22 as-number 100
[SwitchB-bgp-default] address-family dedicated
[SwitchB-bgp-default-dedicated] peer 11.11.11.11 enable
[SwitchB-bgp-default-dedicated] peer 33.33.33.22 enable
```

```
[SwitchB-bgp-default-dedicated] quit
[SwitchB-bgp-default] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 22.22.22.11 as-number 100
[SwitchC-bgp-default] peer 44.44.44.22 as-number 100
[SwitchC-bgp-default] address-family dedicated
[SwitchC-bgp-default-dedicated] peer 22.22.22.11 enable
[SwitchC-bgp-default-dedicated] peer 44.44.44.22 enable
[SwitchC-bgp-default-dedicated] quit
[SwitchC-bgp-default] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 100
[SwitchD-bgp-default] peer 33.33.33.11 as-number 200
[SwitchD-bgp-default] peer 44.44.44.11 as-number 200
[SwitchD-bgp-default] address-family dedicated
[SwitchD-bgp-default-dedicated] peer 33.33.33.11 enable
[SwitchD-bgp-default-dedicated] peer 44.44.44.11 enable
```

On Switch D, enable BGP to permit the local AS number to appear in routes from Switch A.

```
[SwitchD-bgp-default-dedicated] peer 33.33.33.11 allow-as-loop 1
[SwitchD-bgp-default-dedicated] peer 44.44.44.11 allow-as-loop 1
[SwitchD-bgp-default-dedicated] quit
[SwitchD-bgp-default] quit
```

3. Configure IPv4SG on Switch A and Switch D.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 13
[SwitchA-Vlan-interface13] ip verify source ip-address mac-address
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] interface vlan-interface 13
[SwitchD-Vlan-interface13] ip verify source ip-address mac-address
```

Verifying the configuration

View IPv4SG bindings on Switch A.

```
[SwitchA] display ip source binding
Total entries found: 3
IP Address      MAC Address    Interface               VLAN Type
192.168.100.4   1234-5678-9abc Vlan-interface14         N/A  DHCP relay
192.168.100.5   1234-5678-9abd Vlan-interface14         N/A  DHCP relay
192.168.100.6   1234-5678-9abe Vlan-interface14         N/A  DHCP relay
```

View BGP dedicated routes on Switch A.

```
[SwitchA] display bgp routing-table dedicated

 Total number of routes: 3
```

```
 BGP local Switch ID is 1.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

* >  Network : [1][48][1234-5678-9abc][32][192.168.100.1][3]/120
     NextHop : 0.0.0.0                                 LocPrf    : 100
     PrefVal : 32768                                   OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >  Network : [1][48][1234-5678-9abd][32][192.168.100.2][3]/120
     NextHop : 0.0.0.0                                 LocPrf    : 100
     PrefVal : 32768                                   OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >  Network : [1][48][1234-5678-9abe][32][192.168.100.3][3]/120
     NextHop : 0.0.0.0                                 LocPrf    : 100
     PrefVal : 32768                                   OutLabel  : NULL
     MED     : 0
     Path/Ogn: i
```

View BGP dedicated routes on Switch D.

```
[SwitchD] display bgp routing-table dedicated

 Total number of routes: 6

 BGP local Switch ID is 4.4.4.4
 Status codes: * - valid, > - best, d - dampened, h - history
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

* >e Network : [1][48][1234-5678-9abc][32][192.168.100.1][3]/120
     NextHop : 33.33.33.11                             LocPrf    :
     PrefVal : 0                                       OutLabel  : NULL
     MED     :
     Path/Ogn: 200 100i

*  e Network : [1][48][1234-5678-9abc][32][192.168.100.1][3]/120
     NextHop : 44.44.44.11                             LocPrf    :
     PrefVal : 0                                       OutLabel  : NULL
     MED     :
     Path/Ogn: 200 100i

* >e Network : [1][48][1234-5678-9abd][32][192.168.100.2][3]/120
     NextHop : 33.33.33.11                             LocPrf    :
```

```
     PrefVal : 0                                     OutLabel  : NULL
     MED     :
     Path/Ogn: 200 100i

*  e Network : [1][48][1234-5678-9abd][32][192.168.100.2][3]/120
     NextHop : 44.44.44.11                           LocPrf    :
     PrefVal : 0                                     OutLabel  : NULL
     MED     :
     Path/Ogn: 200 100i

* >e Network : [1][48][1234-5678-9abe][32][192.168.100.3][3]/120
     NextHop : 33.33.33.11                           LocPrf    :
     PrefVal : 0                                     OutLabel  : NULL
     MED     :
     Path/Ogn: 200 100i

*  e Network : [1][48][1234-5678-9abe][32][192.168.100.3][3]/120
     NextHop : 44.44.44.11                           LocPrf    :
     PrefVal : 0                                     OutLabel  : NULL
     MED     :
     Path/Ogn: 200 100i
```

View IPv4SG bindings on Switch D.

```
[SwitchD] display ip source binding remote
Total entries found: 3
IP Address      MAC Address    Router ID               VLAN Type
192.168.100.1   1234-5678-9abc 1.1.1.1                 N/A  DHCP relay
192.168.100.2   1234-5678-9abd 1.1.1.1                 N/A  DHCP relay
192.168.100.3   1234-5678-9abe 1.1.1.1                 N/A  DHCP relay
```

The output shows that Switch D has synchronized IPv4SG bindings from Switch A.

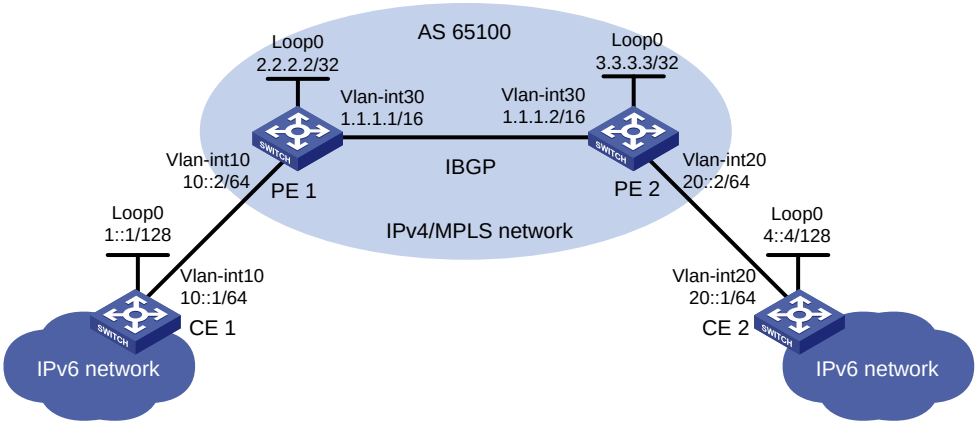
Example: Configuring 6PE

Network configuration

As shown in Figure 77, use 6PE to connect two isolated IPv6 networks over an IPv4/MPLS network.

- The ISP uses OSPF as the IGP.
- PE 1 and PE 2 are edge devices of the ISP, and establish an IPv4 IBGP connection between them.
- CE 1 and CE 2 are edge devices of the IPv6 networks, and they connect the IPv6 networks to the ISP.
- A CE and a PE exchange IPv6 packets through IPv6 static routing.

Figure 77 Network diagram



Procedure

1. Configure IPv6 addresses and IPv4 addresses for interfaces. (Details not shown.)
2. Configure PE 1:

   # Enable LDP globally, and configure the LSP generation policy.

   ```
   <PE1> system-view
   [PE1] mpls lsr-id 2.2.2.2
   [PE1] mpls ldp
   [PE1-ldp] lsp-trigger all
   [PE1-ldp] quit
   ```

   # Enable MPLS and LDP on VLAN-interface 30.

   ```
   [PE1] interface vlan-interface 30
   [PE1-Vlan-interface30] mpls enable
   [PE1-Vlan-interface30] mpls ldp enable
   [PE1-Vlan-interface30] quit
   ```

   # Configure IBGP, enable the peer's 6PE capabilities, and redistribute IPv6 direct and static routes.

   ```
   [PE1] bgp 65100
   [PE1-bgp-default] router-id 2.2.2.2
   [PE1-bgp-default] peer 3.3.3.3 as-number 65100
   [PE1-bgp-default] peer 3.3.3.3 connect-interface loopback 0
   [PE1-bgp-default] address-family ipv6
   [PE1-bgp-default-ipv6] import-route direct
   [PE1-bgp-default-ipv6] import-route static
   [PE1-bgp-default-ipv6] peer 3.3.3.3 enable
   [PE1-bgp-default-ipv6] peer 3.3.3.3 label-route-capability
   [PE1-bgp-default-ipv6] quit
   [PE1-bgp-default] quit
   ```

   # Configure a static route to CE 1.

   ```
   [PE1] ipv6 route-static 1::1 128 10::1
   ```

   # Configure OSPF for the ISP.

   ```
   [PE1] ospf
   [PE1-ospf-1] area 0
   [PE1-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
   [PE1-ospf-1-area-0.0.0.0] network 1.1.0.0 0.0.255.255
   ```

```
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

3. Configure PE 2:

   # Enable LDP globally, and configure the LSP generation policy.

```
<PE2> system-view
[PE2] mpls lsr-id 3.3.3.3
[PE2] mpls ldp
[PE2-mpls-ldp] lsp-trigger all
[PE2-mpls-ldp] quit
```

   # Enable MPLS and LDP on VLAN-interface 30.

```
[PE2] interface vlan-interface 30
[PE2-Vlan-interface30] mpls enable
[PE2-Vlan-interface30] mpls ldp enable
[PE2-Vlan-interface30] quit
```

   # Configure IBGP, enable the peer's 6PE capabilities, and redistribute IPv6 direct and static routes.

```
[PE2] bgp 65100
[PE2-bgp-default] router-id 3.3.3.3
[PE2-bgp-default] peer 2.2.2.2 as-number 65100
[PE2-bgp-default] peer 2.2.2.2 connect-interface loopback 0
[PE2-bgp-default] address-family ipv6
[PE2-bgp-default-ipv6] import-route direct
[PE2-bgp-default-ipv6] import-route static
[PE2-bgp-default-ipv6] peer 2.2.2.2 enable
[PE2-bgp-default-ipv6] peer 2.2.2.2 label-route-capability
[PE2-bgp-default-ipv6] quit
[PE2-bgp-default] quit
```

   # Configure the static route to CE 2.

```
[PE2] ipv6 route-static 4::4 128 20::1
```

   # Configure OSPF for the ISP.

```
[PE2] ospf
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 1.1.0.0 0.0.255.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

4. Configure a static route on CE 1, with PE 1 as the default next hop.

```
<CE1> system-view
[CE1] ipv6 route-static :: 0 10::2
```

5. Configure a static route on CE 2, with PE 2 as the default next hop.

```
<CE2> system-view
[CE2] ipv6 route-static :: 0 20::2
```

Verifying the configuration

Display the IPv6 BGP routing tables on PE 1 and PE 2. The output shows that each of them has two IPv6 network routes. The following shows the output on PE 1:

```
[PE1] display bgp routing-table ipv6
```

```
 Total number of routes: 5

 BGP local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a – additional-path
       Origin: i - IGP, e - EGP, ? - incomplete

* >  Network : 1::1                                    PrefixLen : 128
     NextHop : 10::1                                   LocPrf    :
     PrefVal : 32768                                   OutLabel  : NULL
     MED     : 0
     Path/Ogn: ?

* >i Network : 4::4                                    PrefixLen : 128
     NextHop : ::FFFF:3.3.3.3                          LocPrf    : 100
     PrefVal : 0                                       OutLabel  : 1279
     MED     : 0
     Path/Ogn: ?

* >  Network : 10::                                    PrefixLen : 64
     NextHop : ::                                      LocPrf    :
     PrefVal : 32768                                   OutLabel  : NULL
     MED     : 0
     Path/Ogn: ?

* >  Network : 10::2                                   PrefixLen : 128
     NextHop : ::1                                     LocPrf    :
     PrefVal : 32768                                   OutLabel  : NULL
     MED     : 0
     Path/Ogn: ?

* >i Network : 20::                                    PrefixLen : 64
     NextHop : ::FFFF:3.3.3.3                          LocPrf    : 100
     PrefVal : 0                                       OutLabel  : 1278
     MED     : 0
     Path/Ogn: ?
```

Verify that CE 1 can ping the IPv6 address 4::4 (loopback interface address) of CE 2. (Details not shown.)

Contents

Configuring PBR 1

About PBR 1
Packet forwarding process 1
PBR types 1
Policy 1
PBR and Track 2
Restrictions and guidelines: PBR configuration 3
PBR tasks at a glance 3
Configuring a policy 3
Creating a node 3
Setting match criteria for a node 3
Configuring actions for a node 4
Specifying a policy for PBR 6
Specifying a policy for local PBR 6
Specifying a policy for interface PBR 7
Specifying a policy for global PBR 7
Display and maintenance commands for PBR 8
PBR configuration examples 8
Example: Configuring packet type-based local PBR 8
Example: Configuring packet type-based interface PBR 10
Example: Configuring packet type-based global PBR 11

Configuring PBR

About PBR

Policy-based routing (PBR) uses user-defined policies to route packets. A policy can specify parameters for packets that match specific criteria such as ACLs. The parameters include the next hop and default next hop .

Packet forwarding process

When the device receives a packet, the device searches the PBR policy for a matching node to forward that packet.

- If a matching node is found and its match mode is **permit**, the device performs the following operations:
  a. Uses the next hops specified on the node to forward the packet.
  b. Searches the routing table for a route to forward the packet if one of the following conditions exists:
    - No next hops are specified on the node.
    - Forwarding failed based on the next hops.

  c. Uses the default next hops specified on the node to forward the packet if one of the following conditions exists:
    - No matching route was found in the routing table.
    - The routing table-based forwarding failed.

  d. Uses the default route to forward the packet if one of the following conditions exists:
    - No default next hops are specified on the node.
    - The forwarding failed based on the default next hops.
- The device perfoms routing table lookup to forward the packet in either of the following conditions:
  - No matching node is found.
  - A matching node is found, but its match mode is **deny**.

PBR types

PBR includes the following types:

- **Local PBR**—Guides the forwarding of locally generated packets, such as ICMP packets generated by using the **`ping`** command.
- **Interface PBR**—Guides the forwarding of packets received on an interface.

Policy

A policy includes match criteria and actions to be taken on the matching packets. A policy can have one or multiple nodes as follows:

- Each node is identified by a node number. A smaller node number has a higher priority.
- A node contains **`if-match`** and **`apply`** clauses. An **`if-match`** clause specifies a match criterion, and an **`apply`** clause specifies an action.

- A node has a match mode of **permit** or **deny**.

A policy compares packets with nodes in priority order. If a packet matches the criteria on a node, it is processed by the action on the node. If the packet does not match any criteria on the node, it goes to the next node for a match. If the packet does not match the criteria on any node, the device performs a routing table lookup.

Relationship between if-match clauses

PBR supports only the **`if-match acl`** clause to set an ACL match criterion. On a node, you can specify only one **`if-match`** clause.

Relationship between apply clauses

You can specify multiple **`apply`** clauses for a node, but some of them might not be executed. For more information about relationship between **`apply`** clauses, see "Configuring actions for a node."

Relationship between the match mode and clauses on the node

Does a packet match all the if-match **clauses on the node?**	**Match mode**	
	Permit	**Deny**
Yes.	• If the node contains **`apply`** clauses, PBR executes the **`apply`** clauses on the node. ○ If PBR-based forwarding succeeds, PBR does not compare the packet with the next node. ○ If PBR-based forwarding fails, PBR does not compare the packet with the next node. • If the node does not contain **`apply`** clauses, the device performs a routing table lookup for the packet.	The device does not execute any **`apply`** clauses except the **`apply mirror-to-destination`** clause, and performs a routing table lookup for the packet.
No.	PBR compares the packet with the next node.	PBR compares the packet with the next node.

NOTE:

A node that has no **`if-match`** clauses matches any packet.

PBR and Track

PBR can work with the Track feature to dynamically adapt the availability status of an **`apply`** clause to the link status of a tracked object. The tracked object can be a next hop or default next hop.

- When the track entry associated with an object changes to **Negative**, the **`apply`** clause is invalid.
- When the track entry changes to **Positive** or **NotReady**, the **`apply`** clause is valid.

For more information about Track and PBR collaboration, see *High Availability Configuration Guide*.

Restrictions and guidelines: PBR configuration

If a packet destined for the local device matches a PBR policy, PBR will execute the apply clauses in the policy, including the clause for forwarding. When you configure a PBR policy, be careful to avoid this situation.

PBR tasks at a glance

To configure PBR, perform the following tasks:

1. Configuring a policy
   a. Creating a node
   b. Setting match criteria for a node
   c. Configuring actions for a node
2. Specifying a policy for PBR

   Choose the following tasks as needed:
   - Specifying a policy for local PBR
   - Specifying a policy for interface PBR
   - Specifying a policy for global PBR

Configuring a policy

Creating a node

1. Enter system view.

   system-view
2. Create a node for a policy, and enter its view.

   policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. (Optional.) Configure a description for the policy node.

   description *text*

   By default, no description is configured for a policy node.

Setting match criteria for a node

Procedure

1. Enter system view.

   system-view
2. Enter policy node view.

   policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Set match criteria.
   - Set an ACL match criterion.

     if-match acl [**user-defined**] { *acl-number* | **name** *acl-name* }

     By default, no ACL match criterion is set.

     The ACL match criterion cannot match Layer 2 information.

When using the ACL to match packets, PBR ignores the action (**permit** or **deny**) and time range settings in the ACL. If the specified ACL does not exist, no packets will match this criterion.

- Set a local QoS ID match criterion.
  if-match qos-local-id *local-id-value*
  By default, no local QoS ID match criterion is set.

Configuring actions for a node

About this task

The **apply** clauses allow you to specify the actions to be taken on matching packets on a node.

The following **apply** clauses determine the packet forwarding paths in a descending order:

- **apply next-hop**
- **apply output-interface**
- **apply default-next-hop**

PBR supports the **apply** clauses in Table 1.

Table 1 Apply clauses supported in PBR

Clause	Meaning	Remarks
apply precedence	Sets an IP precedence.	This clause is always executed.
apply loadshare next-hop	Enables load sharing among multiple next hops.	Multiple next hop options operate in either primary/backup or load sharing mode. • **Primary/backup mode**—One option is selected from all options in configuration order for packet forwarding, with all remaining options as backups. For example, if multiple next hops are configured, the first configured next hop is selected. When the selected next hop fails, the next available next hop takes over. • **Load sharing mode**—Matching traffic is distributed across the available options in round robin manner, starting from the first configured option. They perform per-packet load sharing for traffic that does not match any fast forwarding entry, and perform per-flow load sharing for traffic that matches a fast forwarding entry. After you configure the **apply loadshare-mode** command, the load sharing mode specified in the command applies. By default, the primary/backup mode applies. For the load sharing mode to take effect, make sure multiple next hops are set in the policy.
apply loadshare-mode next-hop robin	Specifies the load sharing mode for multiple next	Specify one of the following load sharing modes for multiple next hops:

	hops.	• **robin**: Selects load-shared links randomly based on packets.
apply next-hop and **apply output-interface**	Sets next hops and sets output interfaces.	If both clauses are configured, only the **apply next-hop** clause is executed.
apply default-next-hop	Sets default next hops.	The configuration takes effect only in the following cases: • No next hops are set or the next hops are invalid. • The packet does not match any route in the routing table.
apply statistics	Counts successful matches and matching bytes on the policy node.	To include the number of successful matches and matching bytes on a policy node in the statistics displayed by a **display** command, execute this command. This clause is always executed for matching packets.

Restrictions and guidelines

If you specify a next hop, PBR periodically performs FIB table lookup to determine its availability. Temporary service interruption might occur if PBR does not update the route immediately after its availability status changes.

If you configure the **apply loadshare-mode** command, you cannot specify next hop addresses as VXLAN tunnel addresses.

In a policy node, if you configure per-packet forwarding with the **apply loadshare-mode** command, and apply the policy locally with the **ip local policy-based-route** command, per-packet forwarding does not take effect. Instead, per-flow forwarding still takes effect. As a best practice to have PBR take effect correctly, configure other PBR types with the policy.

Configuring actions to modify packet fields

1. Enter system view.
   system-view
2. Enter policy node view.
   policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Set an IP precedence.
   apply precedence { *type* | *value* }
   By default, no IP precedence is specified.

Configuring actions to direct packet forwarding

1. Enter system view.
   system-view
2. Enter policy node view.
   policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Configure actions.
   - Set next hops.
     apply next-hop [**vpn-instance** *vpn-instance-name*] { *ip-address* [**direct**] [**track** *track-entry-number*] }&<1-8>
     By default, no next hops are specified.

On a node, you can specify a maximum of eight next hops for backup or load sharing in one command line or by executing this command multiple times.

If multiple next hops on the same subnet are specified for backup, the device first uses the subnet route for the next hops to forward packets when the primary next hop fails. If the subnet route is not available, the device selects a backup next hop.

- Enable load sharing among multiple next hops.

  apply loadshare next-hop

  By default, the next hops operate in primary/backup mode.

- Specify the load sharing mode for multiple next hops.

  apply loadshare-mode next-hop robin

  By default, multiple next hops load share traffic in round robin manner, starting from the first configured next hop. They perform per-packet load sharing for traffic that does not match any fast forwarding entry, and perform per-flow load sharing for traffic that matches a fast forwarding entry.

  .

- Set output interfaces.

  apply output-interface NULL 0

  By default, no output interfaces are specified.

- Set default next hops.

  apply default-next-hop [**vpn-instance** *vpn-instance-name*] { *ip-address* [**direct**] [**track** *track-entry-number*] }&<1-4>

  By default, no default next hops are specified.

  On a node, you can specify a maximum of eight default next hops for backup in one command line or by executing this command multiple times.

Collecting policy node statistics

1. Enter system view.

   system-view

2. Enter policy node view.

   policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*

3. Count successful matches and matching bytes on the policy node.

   apply statistics

   By default, counting successful matches is disabled on policy nodes.

Specifying a policy for PBR

Specifying a policy for local PBR

About this task

Perform this task to specify a policy for local PBR to guide the forwarding of locally generated packets.

Restrictions and guidelines

You can specify only one policy for local PBR and must make sure the specified policy already exists. Before you apply a new policy, you must first remove the current policy.

Local PBR might affect local services such as ping and Telnet. When you use local PBR, make sure you fully understand its impact on local services of the device.

Procedure

1. Enter system view.
   system-view
2. Specify a policy for local PBR.
   ip local policy-based-route *policy-name*
   By default, local PBR is not enabled.

Specifying a policy for interface PBR

About this task

Perform this task to apply a policy to an interface to guide the forwarding of packets received on the interface.

Restrictions and guidelines

You can apply only one policy to an interface and must make sure the specified policy already exists. Before you can apply a new interface PBR policy to an interface, you must first remove the current policy from the interface.

You can apply a policy to multiple interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify a policy for interface PBR.
   ip policy-based-route *policy-name*
   By default, no interface policy is applied to an interface.

Specifying a policy for global PBR

About this task

Perform this task to apply a policy to all interfaces on the device to guide the forwarding of packets received on the interfaces.

Restrictions and guidelines

You can apply only one policy for global PBR and the specified policy must already exist. Before you can apply a new policy, you must first remove the current policy.

Interface PBR takes precedence over global PBR on an interface. When they are both configured and packets fail to match the interface PBR policy, global PBR applies.

Procedure

1. Enter system view.
   system-view
2. Specify a policy for global PBR.
   ip global policy-based-route *policy-name*
   By default, no policy is specified for global PBR.

Display and maintenance commands for PBR

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display PBR policy information.	**display ip policy-based-route** [**policy** *policy-name*]
Display global PBR configuration and statistics.	**display ip policy-based-route global** [**slot** *slot-number*]
Display interface PBR configuration and statistics.	**display ip policy-based-route interface** *interface-type interface-number* [**slot** *slot-number*]
Display local PBR configuration and statistics.	**display ip policy-based-route local** [**slot** *slot-number*]
Display PBR configuration.	**display ip policy-based-route setup**
Clear PBR statistics.	**reset ip policy-based-route statistics** [**policy** *policy-name*]

PBR configuration examples

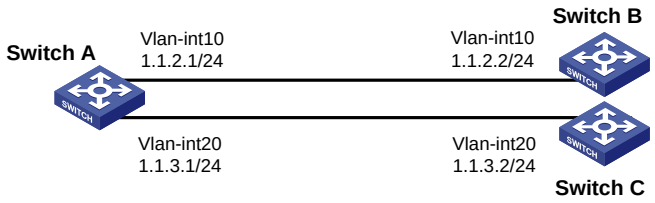
Example: Configuring packet type-based local PBR

Network configuration

As shown in Figure 1, Switch B and Switch C do not have a route to reach each other.

Configure PBR on Switch A to forward all TCP packets to the next hop 1.1.2.2 (Switch B).

Figure 1 Network diagram



Procedure

1. Configure Switch A:

Create VLAN 10 and VLAN 20.

```
<SwitchA> system-view
[SwitchA] vlan 10
[SwitchA-vlan10] quit
[SwitchA] vlan 20
[SwitchA-vlan20] quit
```

Configure the IP addresses of VLAN-interface 10 and VLAN-interface 20.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ip address 1.1.2.1 24
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] ip address 1.1.3.1 24
[SwitchA-Vlan-interface20] quit
```

Configure ACL 3101 to match TCP packets.

```
[SwitchA] acl advanced 3101
[SwitchA-acl-ipv4-adv-3101] rule permit tcp
[SwitchA-acl-ipv4-adv-3101] quit
```

Configure Node 5 for the policy **aaa** to forward TCP packets to next hop 1.1.2.2.

```
[SwitchA] policy-based-route aaa permit node 5
[SwitchA-pbr-aaa-5] if-match acl 3101
[SwitchA-pbr-aaa-5] apply next-hop 1.1.2.2
[SwitchA-pbr-aaa-5] quit
```

Configure local PBR by applying the policy **aaa** to Switch A.

```
[SwitchA] ip local policy-based-route aaa
```

2. Configure Switch B:

Create VLAN 10.

```
<SwitchB> system-view
[SwitchB] vlan 10
[SwitchB-vlan10] quit
```

Configure the IP address of VLAN-interface 10.

```
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ip address 1.1.2.2 24
[SwitchB-Vlan-interface10] quit
```

3. Configure Switch C:

Create VLAN 20.

```
<SwitchC> system-view
```

```
[SwitchC] vlan 20
[SwitchC-vlan20] quit
```

Configure the IP address of VLAN-interface 20.

```
[SwitchC] interface vlan-interface 20
[SwitchC-Vlan-interface20] ip address 1.1.3.2 24
[SwitchC-Vlan-interface20] quit
```

Verifying the configuration

1. Perform telnet operations to verify that local PBR on Switch A operates as configured to forward the matching TCP packets to the next hop 1.1.2.2 (Switch B), as follows:

   # Verify that you can telnet to Switch B from Switch A successfully. (Details not shown.)

   # Verify that you cannot telnet to Switch C from Switch A. (Details not shown.)
2. Verify that Switch A forwards packets other than TCP packets through VLAN-interface 20. For example, verify that you can ping Switch C from Switch A. (Details not shown.)

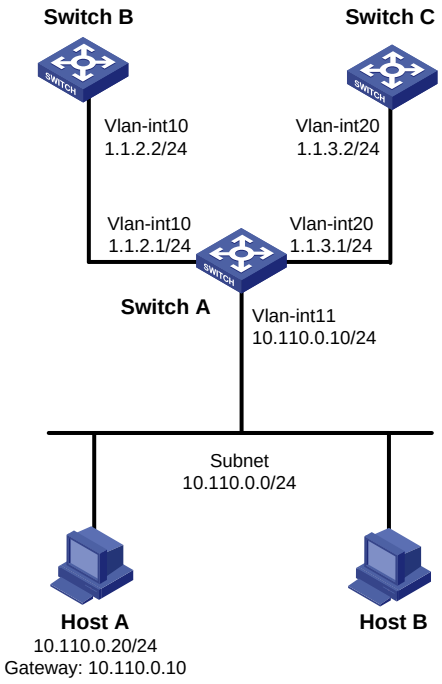
Example: Configuring packet type-based interface PBR

Network configuration

As shown in Figure 2, Switch B and Switch C do not have a route to reach each other.

Configure PBR on Switch A to forward all TCP packets received on VLAN-interface 11 to the next hop 1.1.2.2 (Switch B).

Figure 2 Network diagram



Procedure

1. Configure IP addresses and unicast routing protocol settings to make sure Switch B and Switch C can reach Host A. (Details not shown.)
2. Configure Switch A:

   # Configure ACL 3101 to match TCP packets.

   ```
   <SwitchA> system-view
   ```

```
[SwitchA] acl advanced 3101
[SwitchA-acl-ipv4-adv-3101] rule permit tcp
[SwitchA-acl-ipv4-adv-3101] quit
```

Configure Node 5 for the policy **aaa** to forward TCP packets to next hop 1.1.2.2.

```
[SwitchA] policy-based-route aaa permit node 5
[SwitchA-pbr-aaa-5] if-match acl 3101
[SwitchA-pbr-aaa-5] apply next-hop 1.1.2.2
[SwitchA-pbr-aaa-5] quit
```

Configure interface PBR by applying the policy **aaa** to VLAN-interface 11.

```
[SwitchA] interface vlan-interface 11
[SwitchA-Vlan-interface11] ip policy-based-route aaa
[SwitchA-Vlan-interface11] quit
```

Verifying the configuration

1. Perform telnet operations to verify that interface PBR on Switch A operates as configured to forward the matching TCP packets to the next hop 1.1.2.2 (Switch B), as follows:

   # Verify that you can telnet to Switch B from Host A successfully. (Details not shown.)

   # Verify that you cannot telnet to Switch C from Host A. (Details not shown.)
2. Verify that Switch A forwards packets other than TCP packets through VLAN-interface 20. For example, verify that you can ping Switch C from Host A. (Details not shown.)

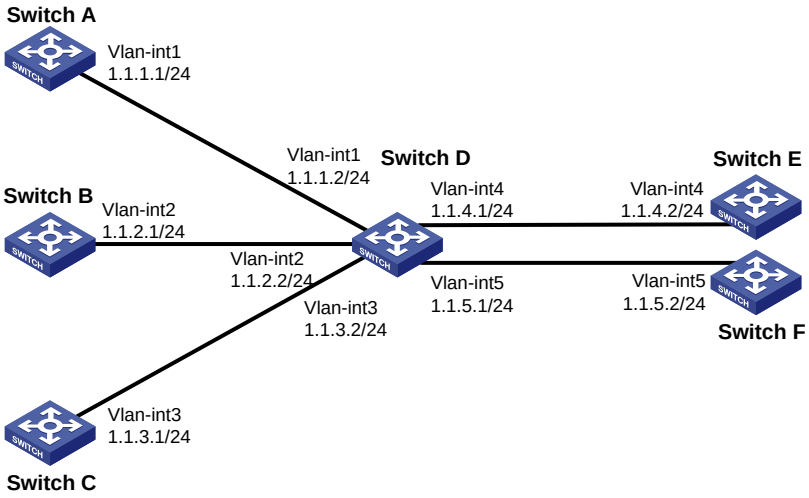
Example: Configuring packet type-based global PBR

Network configuration

As shown in Figure 3, Switch E and Switch F do not have a route to reach each other.

Configure global PBR on Switch D to forward TCP packets to the next hop 1.1.4.2 (Switch E).

Figure 3 Network diagram



Procedure

1. Assign IP addresses to the interfaces, as shown in Figure 3. (Details not shown.)
2. Configure unicast routing protocol settings to make sure Switch A, B and C can communicate with Switch E and Switch F, respectively. (Details not shown.)
3. Configure Switch D:

Configure ACL 3101 to match TCP packets sourced from networks 1.1.1.0/24, 1.1.2.0/24, and 1.1.3.0/24.

```
<SwitchD> system-view
[SwitchD] acl advanced 3101
[SwitchD-acl-ipv4-adv-3101] rule permit tcp source 1.1.1.0 0.0.0.255
[SwitchD-acl-ipv4-adv-3101] rule permit tcp source 1.1.2.0 0.0.0.255
[SwitchD-acl-ipv4-adv-3101] rule permit tcp source 1.1.3.0 0.0.0.255
[SwitchD-acl-ipv4-adv-3101] quit
```

Configure node 5 in PBR policy **aaa** to forward TCP packets that match ACL 3101 to next hop 1.1.4.2.

```
[SwitchD] policy-based-route aaa permit node 5
[SwitchD-pbr-aaa-5] if-match acl 3101
[SwitchD-pbr-aaa-5] apply next-hop 1.1.4.2
[SwitchD-pbr-aaa-5] quit
```

Specify PBR policy **aaa** as the global PBR policy.

```
[SwitchD] ip global policy-based-route aaa
```

Verifying the configuration

1. Perform telnet operations to verify that global PBR on Switch D operates as configured to forward the matching TCP packets to the next hop 1.1.4.2 (Switch E), as follows:

   # Verify that you can telnet to Switch E from Switch A, Switch B, and Switch C successfully. (Details not shown.)

   # Verify that you cannot telnet to Switch F from Switch A, Switch B, or Switch C. (Details not shown.)

2. Verify that Switch D forwards packets other than TCP packets as long as a route is available. For example, verify that you can ping Switch F from Switch A, Switch B, and Switch C. (Details not shown.)

Contents

Configuring IPv6 static routing 1
About IPv6 static routing 1
Configuring an IPv6 static route 1
Configuring IPv6 floating static routes 2
Deleting IPv6 static routes 2
Configuring BFD for IPv6 static routes 3
About BFD for IPv6 static routes 3
Restrictions and guidelines for BFD 3
Configuring BFD control packet mode 3
Configuring BFD echo packet mode 4
Enabling periodic sending of ND requests to the next hops of IPv6 static routes 4
Display and maintenance commands for IPv6 static routing 5
IPv6 static routing configuration examples 6
Example: Configuring basic IPv6 static route 6
Example: Configuring BFD for IPv6 static routes (direct next hop) 7
Example: Configuring BFD for IPv6 static routes (indirect next hop) 10
Configuring an IPv6 default route 13

Configuring IPv6 static routing

About IPv6 static routing

Static routes are manually configured and cannot adapt to network topology changes. If a fault or a topological change occurs in the network, the network administrator must modify the static routes manually. IPv6 static routing works well in a simple IPv6 network.

Configuring an IPv6 static route

1. Enter system view.

   system-view

2. Configure an IPv6 static route.

   Public network:

   ipv6 route-static *ipv6-address prefix-length interface-type interface-number* [*next-hop-address*] [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ipv6 route-static *ipv6-address prefix-length* [**vpn-instance** *d-vpn-instance-name*] *next-hop-address* [**recursive-lookup** { **host-route** | **longest-match** }] [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ipv6 route-static *ipv6-address prefix-length* **vpn-instance** *d-vpn-instance-name* [**track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   ipv6 route-static *ipv6-address prefix-length interface-type interface-number next-hop-address* **track-nd** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, no IPv6 static route is configured.

   The **recursive-lookup host-route**, **recursive-lookup longest-match**, and **recursive-lookup** keywords are mutually exclusive. You cannot specify them at the same time.

   VPN:

   ipv6 route-static vpn-instance *s-vpn-instance-name ipv6-address prefix-length interface-type interface-number* [*next-hop-address*] [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ipv6 route-static vpn-instance *s-vpn-instance-name ipv6-address prefix-length* { *nexthop-address* [**recursive-lookup** { **host-route** | **longest-match** }] [**public**] | **vpn-instance** *d-vpn-instance-name nexthop-address* [**recursive-lookup** { **host-route** | **longest-match** }] } [**permanent** | **track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**recursive-lookup**] [**description** *text*]

   ipv6 route-static vpn-instance *s-vpn-instance-name ipv6-address prefix-length* { **public** | **vpn-instance** *d-vpn-instance-name* } [**track** *track-entry-number*] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

ipv6 route-static vpn-instance *s-vpn-instance-name ipv6-address prefix-length interface-type interface-number next-hop-address* **track-nd** [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

By default, no IPv6 static route is configured.

The **recursive-lookup host-route**, **recursive-lookup longest-match**, and **recursive-lookup** keywords are mutually exclusive. You cannot specify them at the same time.

If you specify a destination VPN instance without specifying a next hop address, the destination VPN instance cannot be the same as the source VPN instance.

3. (Optional.) Set the default preference for IPv6 static routes.

   ipv6 route-static default-preference *default-preference*

   The default setting is 60.

Configuring IPv6 floating static routes

Perform this task to configure route backup to improve network reliability.

When an IPv6 static or dynamic route to a destination address already exists on the device, you configure another IPv6 static route with a lower priority as the backup route to improve the network reliability. This backup IPv6 static route is called an IPv6 floating static route and is activated only when the primary route fails. After the primary route recovers from failure, the IPv6 floating static route becomes inactive and data forwarding switches back to the primary route.

An IPv6 floating static route can be configured in either of the following ways

- Configure different priorities for multiple IPv6 static routes to the same destination address. The route with the lower priority automatically becomes the IPv6 floating static route.
- When an IPv6 route to a destination address already exists on the device, configure an IPv6 static route with a lower priority to the same destination address.

When you configure an IPv6 floating static route, the priority value of the route must be larger than then priority value of the primary route. For more information, see "Configuring an IPv6 static route."

Deleting IPv6 static routes

About this task

To delete an IPv6 static route, use the **undo ipv6 route-static** command. To delete all IPv6 static routes including the default route, use the **delete ipv6 static-routes all** command.

Procedure

1. Enter system view.

   system-view

2. Delete all IPv6 static routes, including the default route.

   delete ipv6 [**vpn-instance** *vpn-instance-name*] **static-routes all**

CAUTION:

Deleting all IPv6 static routes with caution. This operation might cause network connectivity failure and packet forwarding failure.

Configuring BFD for IPv6 static routes

About BFD for IPv6 static routes

BFD provides a general purpose, standard, and medium- and protocol-independent fast failure detection mechanism. It can uniformly and quickly detect the failures of the bidirectional forwarding paths between two routers for protocols, such as routing protocols. BFD for IPv6 static routes tests the reachability of the next hop for each IPv6 static route. If a next hop is unreachable, BFD deletes the associated IPv6 static route.

For more information about BFD, see *High Availability Configuration Guide*.

Restrictions and guidelines for BFD

When you configure BFD for IPv6 static routes, follow these restrictions and guidelines:

- If you specify a source IPv6 address for BFD packets on the local device, you must specify that IPv6 address as the next hop IPv6 address on the peer device.
- If you specify a non-P2P output interface and a direct next hop, specify the **bfd-source** *ipv6-address* option as a best practice. Make sure the source IPv6 address of BFD packets meets the following requirements:
  - The address is the same as the IPv6 address of the output interface.
  - The address is on the same network segment as the next hop IPv6 address of the same type.

    For example, if the next hop IPv6 address is a link-local address, the source IPv6 address of BFD packets must also be a link-local address.
- Enabling BFD for a flapping route could worsen the situation.

Configuring BFD control packet mode

About this task

This mode uses BFD control packets to detect the status of a link bidirectionally at a millisecond level.

BFD control packet mode can be applied to IPv6 static routes with a direct next hop or with an indirect next hop.

Restrictions and guidelines for BFD control packet mode

If you configure BFD control packet mode at the local end, you must also configure this mode at the peer end.

Configuring BFD control packet mode for an IPv6 static route (direct next hop)

1. Enter system view.

   system-view
2. Configure BFD control packet mode for an IPv6 static route.

   ipv6 route-static [**vpn-instance** *s-vpn-instance-name*] *ipv6-address prefix-length interface-type interface-number next-hop-address* **bfd control-packet** [**bfd-source** *ipv6-address*] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, BFD control packet mode for an IPv6 static route is not configured.

Configuring BFD control packet mode for an IPv6 static route (indirect next hop)

1. Enter system view.

   system-view

2. Configure BFD control packet mode for an IPv6 static route.

   ipv6 route-static [**vpn-instance** *s-vpn-instance-name*] *ipv6-address prefix-length* [**vpn-instance** *d-vpn-instance-name*] { *next-hop-address* [**recursive-lookup** { **host-route** | **longest-match** }] **bfd control-packet bfd-source** *ipv6-address* } [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, BFD control packet mode for an IPv6 static route is not configured.

Configuring BFD echo packet mode

About this task

With BFD echo packet mode enabled for a static route, the output interface sends BFD echo packets to the destination device, which loops the packets back to test the link reachability.

Restrictions and guidelines

You do not need to configure BFD echo packet mode at the peer end.

Do not use BFD for a static route with the output interface in spoofing state.

Procedure

1. Enter system view.

   system-view

2. Configure the source address of echo packets.

   bfd echo-source-ipv6 *ipv6-address*

   By default, the source address of echo packets is not configured.

   As a best practice to avoid network congestion caused by excessive ICMPv6 redirect packets from the peer, use this command. Make sure the source IPv6 address is not on the subnet of any interfaces on the device.

   The source IPv6 address of echo packets can only be the global unicast address.

   For more information about this command, see *High Availability Command Reference*.

3. Configure BFD echo packet mode for an IPv6 static route.

   ipv6 route-static [**vpn-instance** *s-vpn-instance-name*] *ipv6-address prefix-length interface-type interface-number next-hop-address* **bfd echo-packet** [**bfd-source** *ipv6-address*] [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, BFD echo packet mode for an IPv6 static route is not configured.

   The next hop IPv6 address must be a global unicast address.

Enabling periodic sending of ND requests to the next hops of IPv6 static routes

About this task

As a best practice, use this feature when the following types of IPv6 static routes exist on the device:

- IPv6 static routes that recurse to host routes, which include the following IPv6 static routes:

- IPv6 static routes with the **recursive-lookup host-route** keywords specified to recurse only to host routes.
- IPv6 static routes that recurse to host routes according to routing policies (configured by using the **protocol nexthop recursive-lookup** command).

The device might delete the ND entry for the next hop host address of an IPv6 static route after that entry ages out because the device cannot update that entry in time. As a result, the device does not have the host route of the directly connected device in the IPv6 routing table for that IPv6 static route. In this case, route recursion fails and that IPv6 static route cannot be activated.

- IPv6 static routes with the **track-nd** keyword specified to associate their next hops with the ND entries for these next hops.

  The device might delete the ND entry for the next hop address of an IPv6 static route after that entry ages out because the device cannot update that entry in time. As a result, that IPv6 static route cannot be activated.

For IPv6 static routes that recurse to host routes, the **ipv6 route-static nd-request** command enables the device to periodically send ND requests to the next hops of the IPv6 static routes that meet the following requirements:

- The IPv6 static routes have no output interfaces specified.
- The IPv6 static routes fail the next-hop recursion.

When the device receives an ND response from the next hop host of an IPv6 static route, it automatically performs the following operations:

1. Adds the host route to the routing table.
2. Activates the IPv6 static route that recurses to the host route.
3. Stops sending ND requests to the host.

For IPv6 static routes with the **track-nd** keyword specified, the **ipv6 route-static nd-request** command enables the device to periodically send ND requests to the next hops of the IPv6 static routes associated with next hop ND entries. When the device receives an ND response from the next hop of an IPv6 static route, it automatically activates that IPv6 static route and stops sending ND requests to the next hop of that IPv6 static route.

Procedure

1. Enter system view.

   system-view
2. Enable periodic sending of ND requests to the next hops of IPv6 static routes.

   ipv6 route-static nd-request [**interval** *interval*]

   By default, periodic sending of ND requests to the next hops of IPv6 static routes is disabled.

Display and maintenance commands for IPv6 static routing

Execute **display** commands in any view.

Task	Command
Display IPv6 static route next hop information.	**display ipv6 route-static nib** [*nib-id*] [**verbose**]
Display IPv6 static routing table information.	**display ipv6 route-static routing-table** [**vpn-instance** *vpn-instance-name*]

Task	Command
	[*ipv6-address prefix-length*]
Display IPv6 static route information.	**display ipv6 routing-table protocol static** [**inactive** | **verbose**]

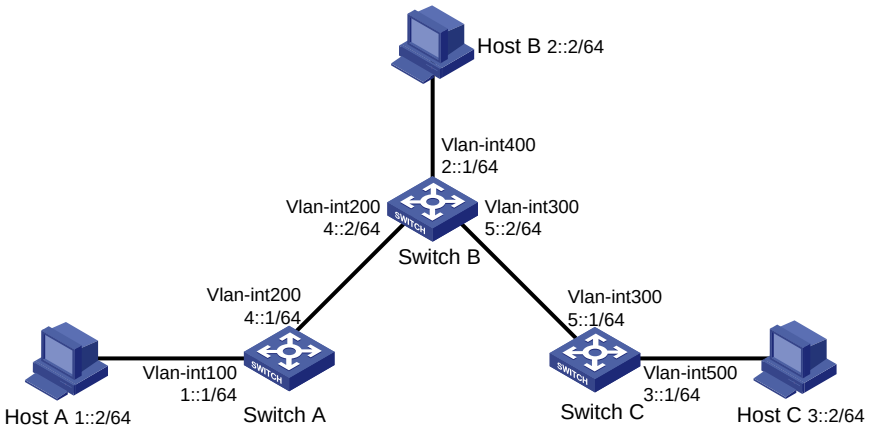
IPv6 static routing configuration examples

Example: Configuring basic IPv6 static route

Network configuration

As shown in Figure 1, configure IPv6 static routes so that hosts can reach one another.

Figure 1 Network diagram



Procedure

1. Configure the IPv6 addresses for all VLAN interfaces. (Details not shown.)
2. Configure IPv6 static routes:

   # Configure a default IPv6 static route on Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 route-static :: 0 4::2
   ```

   # Configure two IPv6 static routes on Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] ipv6 route-static 1:: 64 4::1
   [SwitchB] ipv6 route-static 3:: 64 5::1
   ```

   # Configure a default IPv6 static route on Switch C.

   ```
   <SwitchC> system-view
   [SwitchC] ipv6 route-static :: 0 5::2
   ```
3. Configure the IPv6 addresses for all the hosts and configure the default gateway of Host A, Host B, and Host C as 1::1, 2::1, and 3::1.

Verifying the configuration

Display the IPv6 static route information on Switch A.

```
[SwitchA] display ipv6 routing-table protocol static
```

```
Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination: ::/0                                       Protocol  : Static
NextHop    : 4::2                                       Preference: 60
Interface  : Vlan200                                    Cost      : 0

Static Routing table status : <Inactive>
Summary count : 0
```

Display the IPv6 static route information on Switch B.

```
[SwitchB] display ipv6 routing-table protocol static

Summary count : 2

Static Routing table status : <Active>
Summary count : 2

Destination: 1::/64                                     Protocol  : Static
NextHop    : 4::1                                       Preference: 60
Interface  : Vlan200                                    Cost      : 0

Destination: 3::/64                                     Protocol  : Static
NextHop    : 5::1                                       Preference: 60
Interface  : Vlan300                                    Cost      : 0

Static Routing table status : <Inactive>
Summary count : 0
```

Use the **ping** command to test the reachability.

```
[SwitchA] ping ipv6 3::1
Ping6(56 data bytes) 4::1 --> 3::1, press CTRL+C to break
56 bytes from 3::1, icmp_seq=0 hlim=62 time=0.700 ms
56 bytes from 3::1, icmp_seq=1 hlim=62 time=0.351 ms
56 bytes from 3::1, icmp_seq=2 hlim=62 time=0.338 ms
56 bytes from 3::1, icmp_seq=3 hlim=62 time=0.373 ms
56 bytes from 3::1, icmp_seq=4 hlim=62 time=0.316 ms

--- Ping6 statistics for 3::1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.316/0.416/0.700/0.143 ms
```

Example: Configuring BFD for IPv6 static routes (direct next hop)

Network configuration

As shown in Figure 2:

- Configure an IPv6 static route to subnet 120::/64 on Switch A.
- Configure an IPv6 static route to subnet 121::/64 on Switch B.
- Enable BFD for both routes.
- Configure an IPv6 static route to subnet 120::/64 and an IPv6 static route to subnet 121::/64 on Switch C.

When the link between Switch A and Switch B through the Layer 2 switch fails, BFD can detect the failure immediately, and Switch A and Switch B can communicate through Switch C.

Figure 2 Network diagram

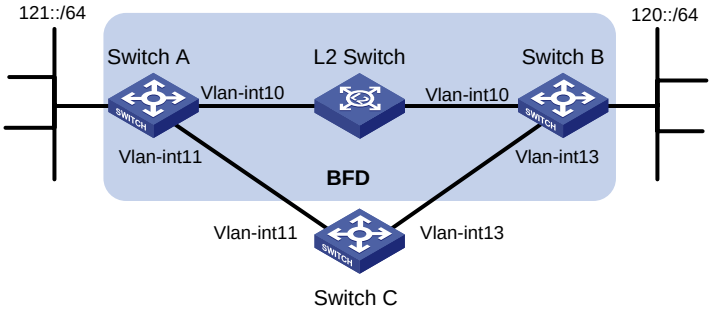


Table 1 Interface and IP address assignment

Device	Interface	IPv6 address
Switch A	Vlan-int10	12::1/64
Switch A	Vlan-int11	10::102/64
Switch B	Vlan-int10	12::2/64
Switch B	Vlan-int13	13::1/64
Switch C	Vlan-int11	10::100/64
Switch C	Vlan-int13	13::2/64

Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure IPv6 static routes and BFD:

   # Configure IPv6 static routes on Switch A and enable BFD control packet mode for the static route that traverses the Layer 2 switch.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 10
[SwitchA-vlan-interface10] bfd min-control-interval 500
[SwitchA-vlan-interface10] bfd detect-multiplier 9
[SwitchA-vlan-interface10] quit
[SwitchA] ipv6 route-static 120:: 64 vlan-interface 10 12::2 bfd control-packet
[SwitchA] ipv6 route-static 120:: 64 10::100 preference 65
[SwitchA] quit
```

   # Configure IPv6 static routes on Switch B and enable BFD control packet mode for the static route that traverses the Layer 2 switch.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 10
[SwitchB-vlan-interface10] bfd min-control-interval 500
```

```
[SwitchB-vlan-interface10] bfd detect-multiplier 9
[SwitchB-vlan-interface10] quit
[SwitchB] ipv6 route-static 121:: 64 vlan-interface 10 12::1 bfd control-packet
[SwitchB] ipv6 route-static 121:: 64 vlan-interface 13 13::2 preference 65
[SwitchB] quit
```

Configure IPv6 static routes on Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6 route-static 120:: 64 13::1
[SwitchC] ipv6 route-static 121:: 64 10::102
```

Verifying the configuration

Display the BFD sessions on Switch A.

```
<SwitchA> display bfd session
 Total sessions: 1        Up sessions: 1        Init mode: Active

 IPv6 session working in control packet mode:

       Local Discr: 513                Remote Discr: 33
         Source IP: 12::1
    Destination IP: 12::2
     Session State: Up                     Interface: Vlan10
         Hold Time: 2012ms
```

The output shows that the BFD session has been created.

Display IPv6 static routes on Switch A.

```
<SwitchA> display ipv6 routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination: 120::/64                                 Protocol  : Static
NextHop    : 12::2                                    Preference: 60
Interface  : Vlan10                                   Cost      : 0

Direct Routing table status : <Inactive>
Summary count : 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 10. The link over VLAN-interface 10 fails.

Display IPv6 static routes on Switch A again.

```
<SwitchA> display ipv6 routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination: 120::/64                                 Protocol  : Static
```

```
NextHop    : 10::100                                Preference: 65
Interface  : Vlan11                                 Cost      : 0

Static Routing table status : < Inactive>
Summary count : 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 11.

Example: Configuring BFD for IPv6 static routes (indirect next hop)

Network configuration

As shown in Figure 3:

- Switch A has a route to interface Loopback 1 (2::9/128) on Switch B, and the output interface is VLAN-interface 10.
- Switch B has a route to interface Loopback 1 (1::9/128) on Switch A, and the output interface is VLAN-interface 12.
- Switch D has a route to 1::9/128, and the output interface is VLAN-interface 10. It also has a route to 2::9/128, and the output interface is VLAN-interface 12.

Configure the following:

- Configure an IPv6 static route to subnet 120::/64 on Switch A.
- Configure an IPv6 static route to subnet 121::/64 on Switch B.
- Enable BFD for both routes.
- Configure an IPv6 static route to subnet 120::/64 and an IPv6 static route to subnet 121::/64 on both Switch C and Switch D.

When the link between Switch A and Switch B through Switch D fails, BFD can detect the failure immediately and Switch A and Switch B can communicate through Switch C.

Figure 3 Network diagram

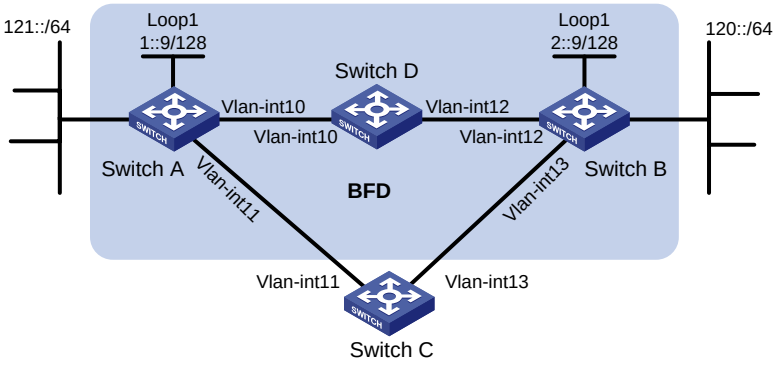


Table 2 Interface and IP address assignment

Device	Interface	IPv6 address
Switch A	Vlan-int10	12::1/64
Switch A	Vlan-int11	10::102/64
Switch A	Loop1	1::9/128
Switch B	Vlan-int12	11::2/64

Device	Interface	IPv6 address
Switch B	Vlan-int13	13::1/64
Switch B	Loop1	2::9/128
Switch C	Vlan-int11	10::100/64
Switch C	Vlan-int13	13::2/64
Switch D	Vlan-int10	12::2/64
Switch D	Vlan-int12	11::1/64

Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure IPv6 static routes and BFD:

   # Configure IPv6 static routes on Switch A and enable BFD control packet mode for the IPv6 static route that traverses Switch D.

```
<SwitchA> system-view
[SwitchA] bfd multi-hop min-control-interval 500
[SwitchA] bfd multi-hop detect-multiplier 9
[SwitchA] ipv6 route-static 120:: 64 2::9 bfd control-packet bfd-source 1::9
[SwitchA] ipv6 route-static 120:: 64 10::100 preference 65
[SwitchA] ipv6 route-static 2::9 128 12::2
[SwitchA] quit
```

   # Configure IPv6 static routes on Switch B and enable BFD control packet mode for the static route that traverses Switch D.

```
<SwitchB> system-view
[SwitchB] bfd multi-hop min-control-interval 500
[SwitchB] bfd multi-hop detect-multiplier 9
[SwitchB] ipv6 route-static 121:: 64 1::9 bfd control-packet bfd-source 2::9
[SwitchB] ipv6 route-static 121:: 64 13::2 preference 65
[SwitchB] ipv6 route-static 1::9 128 11::1
[SwitchB] quit
```

   # Configure IPv6 static routes on Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6 route-static 120:: 64 13::1
[SwitchC] ipv6 route-static 121:: 64 10::102
```

   # Configure IPv6 static routes on Switch D.

```
<SwitchD> system-view
[SwitchD] ipv6 route-static 120:: 64 11::2
[SwitchD] ipv6 route-static 121:: 64 12::1
[SwitchD] ipv6 route-static 2::9 128 11::2
[SwitchD] ipv6 route-static 1::9 128 12::1
```

Verifying the configuration

Display the BFD sessions on Switch A.

```
<SwitchA> display bfd session
 Total sessions: 1        Up sessions: 1        Init mode: Active

 IPv6 session working in control packet mode:
```

```
       Local Discr: 513                Remote Discr: 33
         Source IP: 1::9
    Destination IP: 2::9
     Session State: Up                      Interface: N/A
         Hold Time: 2012ms
```

The output shows that the BFD session has been created.

Display the IPv6 static routes on Switch A.

```
<SwitchA> display ipv6 routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination: 120::/64                                  Protocol  : Static
NextHop    : 2::9                                      Preference: 60
Interface  : Vlan10                                    Cost      : 0

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that Switch A communicates Switch B through VLAN-interface 10. The link over VLAN-interface 10 fails.

Display IPv6 static routes on Switch A again.

```
<SwitchA> display ipv6 routing-table protocol static

Summary count : 1

Static Routing table status : <Active>
Summary count : 1

Destination: 120::/64                                  Protocol  : Static
NextHop    : 10::100                                   Preference: 65
Interface  : Vlan11                                    Cost      : 0

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that Switch A communicates with Switch B through VLAN-interface 11.

Configuring an IPv6 default route

A default IPv6 route is used to forward packets that match no entry in the routing table.

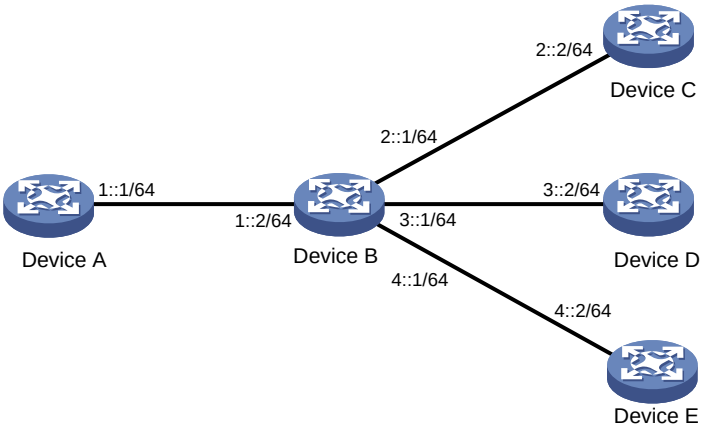
A default IPv6 route can be configured in either of the following ways:

- The network administrator can configure a default route with a destination prefix of **::/0**. For more information, see "Configuring IPv6 static routing."
- Some dynamic routing protocols (such as OSPFv3, IPv6 IS-IS, and RIPng) can generate a default IPv6 route. For example, an upstream router running OSPFv3 can generate a default IPv6 route and advertise it to other routers. These routers install the default IPv6 route with the next hop being the upstream router. For more information, see the respective chapters on those routing protocols in this configuration guide.

As shown in Figure 4, Device B is the next hop for packets from Device A to Device C, Device D, and Device E. You can configure a default route on Device A to replace the three IPv6 static routes from Device A to Device C, Device D, and Device E, respectively.

The next hop address, destination address, and subnet mask of the IPv6 default route configured on Device A are 1::2, ::, and 0, respectively.

Figure 4 Configuring a default route



Contents

Configuring RIPng 1

About RIPng 1
RIPng routing metrics 1
RIPng route entries 1
RIPng packets and advertisement 1
Protocols and standards 1
RIPng tasks at a glance 2
Configuring basic RIPng 2
Enabling RIPng 2
Suppressing an interface 3
Configuring RIPng route control 3
Configuring an additional routing metric 3
Configuring RIPng route summarization 3
Advertising a default route 4
Configuring received/redistributed route filtering 4
Setting a preference for RIPng 5
Configuring RIPng route redistribution 5
Tuning and optimizing the RIPng network 6
Setting RIPng timers 6
Configuring split horizon and poison reverse 6
Setting the maximum number of ECMP routes 7
Configuring the RIPng packet sending rate 7
Setting the interval for sending triggered updates 8
Configuring RIPng GR 8
Configuring RIPng NSR 9
Configuring BFD for RIPng 9
Configuring BFD for RIPng 9
Configuring RIPng FRR 10
About RIPng FRR 10
Restrictions and guidelines for RIPng FRR 10
Enabling RIPng FRR 10
Enabling BFD for RIPng FRR 11
Enhancing RIPng security 11
Configuring zero field check for RIPng packets 11
Applying an IPsec profile 12
Display and maintenance commands for RIPng 12
RIPng configuration examples 13
Example: Configuring basic RIPng 13
Example: Configuring RIPng route redistribution 16
Example: Configuring RIPng GR 18
Example: Configuring RIPng NSR 19
Example: Configuring RIPng FRR 21
Example: Configuring RIPng IPsec profile 24

Configuring RIPng

About RIPng

RIP next generation (RIPng), as an extension of RIP-2 for support of IPv6, is a distance vector routing protocol. It employs UDP to exchange route information through port 521. Most RIP concepts are applicable to RIPng.

RIPng routing metrics

RIPng uses a hop count to measure the distance to a destination. The hop count is the metric or cost. The hop count from a router to a directly connected network is 0. The hop count between two directly connected routers is 1. When the hop count is greater than or equal to 16, the destination network or host is unreachable.

RIPng route entries

RIPng stores route entries in a database. Each route entry contains the following elements:

- **Destination address**—IPv6 address of a destination host or a network.
- **Next hop address**—IPv6 address of the next hop.
- **Egress interface**—Egress interface of the route.
- **Metric**—Cost from the local router to the destination.
- **Route time**—Time elapsed since the most recent update. The time is reset to 0 every time the route entry is updated.
- **Route tag**—Used for route control. For more information, see "Configuring routing policies."

RIPng packets and advertisement

RIPng multicasts request and response packets to exchange routing information. It uses FF02::9 as the destination address and link-local address FE80::/10 as the source address. RIPng exchanges routing information as follows:

1. When RIPng starts or needs to update some route entries, it sends a multicast request packet to neighbors.
2. When a RIPng neighbor receives the request packet, it sends back a response packet that contains the local routing table. RIPng can also advertise route updates in response packets periodically or advertise a triggered update caused by a route change.
3. After RIPng receives the response, it checks the validity of the response before adding routes to its routing table, including the following details:
   - Whether the source IPv6 address is the link-local address.
   - Whether the port number is correct.
4. A response packet that fails the check is discarded.

Protocols and standards

- RFC 2080, *RIPng for IPv6*
- RFC 2081, *RIPng Protocol Applicability Statement*

RIPng tasks at a glance

To configure RIPng, perform the following tasks:

1. Configuring basic RIPng
   a. Enabling RIPng
   b. (Optional.) Suppressing an interface
2. (Optional.) Configuring RIPng route control
   - Configuring an additional routing metric
   - Configuring RIPng route summarization
   - Advertising a default route
   - Configuring received/redistributed route filtering
   - Setting a preference for RIPng
   - Configuring RIPng route redistribution
3. (Optional.) Tuning and optimizing the RIPng network
   - Setting RIPng timers
   - Configuring split horizon and poison reverse
   - Setting the maximum number of ECMP routes
   - Configuring the RIPng packet sending rate
   - Setting the interval for sending triggered updates
4. (Optional.) Enhancing RIPng availability
   - Configuring RIPng GRConfiguring RIPng NSR
   - Enabling BFD for RIPng FRR
   - Configuring RIPng FRR
5. (Optional.) Enhancing RIPng security
   - Configuring zero field check for RIPng packetsApplying an IPsec profile

Configuring basic RIPng

Enabling RIPng

1. Enter system view.
   system-view
2. Enable RIPng and enter its view.
   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]
   By default, RIPng is disabled.
3. Return to system view.
   quit
4. Enter interface view.
   interface *interface-type interface-number*
5. Enable RIPng on the interface.
   ripng *process-id* **enable**
   By default, RIPng is disabled on the interface.
   If RIPng is not enabled on an interface, the interface does not send or receive any RIPng route.

Suppressing an interface

About this task

Perform this task to disable interfaces from sending RIPng messages. The interfaces can still receive RIPng messages.

Procedure

1. Enter system view.

   system-view
2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Suppress an interface.

   silent-interface { *interface-type interface-number* | **all** }

   By default, all RIPng-enabled interfaces can send RIP messages.

   The suppressed interface can still receive RIP messages and respond to unicast requests containing unknown ports.

Configuring RIPng route control

Configuring an additional routing metric

About this task

An additional routing metric (hop count) can be added to the metric of an inbound or outbound RIPng route.

- An outbound additional metric is added to the metric of a sent route, and it does not change the route's metric in the routing table.
- An inbound additional metric is added to the metric of a received route before the route is added into the routing table, and the route's metric is changed.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify an inbound additional routing metric.

   ripng metricin *value*

   The default additional metric of an inbound route is 0.
4. Specify an outbound additional routing metric.

   ripng metricout *value*

   The default additional metric of an outbound route is 1.

Configuring RIPng route summarization

About this task

RIPng route summarization is interface-based. RIPng advertises a summary route based on the longest match.

RIPng route summarization improves network scalability, reduces routing table size, and increases routing table lookup efficiency.

RIPng advertises a summary route with the smallest metric of all the specific routes.

For example, RIPng has two specific routes to be advertised through an interface: 1:11:11::24 with a metric of a 2 and 1:11:12::34 with a metric of 3. Configure route summarization on the interface, so RIPng advertises a single route 11::0/16 with a metric of 2.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Advertise a summary IPv6 prefix.

   ripng summary-address *ipv6-address prefix-length*

   By default, no summary IPv6 prefix is configured on the interface.

Advertising a default route

About this task

You can configure RIPng to advertise a default route with the specified cost to its neighbors.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure RIPng to advertise a default route.

   ripng default-route { **only** | **originate** } [**cost** *cost-value* | **route-policy** *route-policy-name*] *

   By default, RIPng does not advertise a default route.

   This command advertises a default route on the current interface regardless of whether the default route exists in the local IPv6 routing table.

Configuring received/redistributed route filtering

About this task

Perform this task to filter received or redistributed routes by using an IPv6 ACL, IPv6 prefix list, or routing policy. You can also configure RIPng to filter routes redistributed from other routing protocols and routes from a specified neighbor.

Procedure

1. Enter system view.

   system-view
2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Configure a filter policy to filter received routes.

   filter-policy { *ipv6-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } **import**

By default, RIPng does not filter received routes.

4. Configure a filter policy to filter redistributed routes.

   filter-policy { *ipv6-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } **export** [*protocol* [*process-id*]]

   By default, RIPng does not filter redistributed routes.

Setting a preference for RIPng

About this task

Routing protocols each have a preference. When they find routes to the same destination, the route found by the routing protocol with the highest preference is selected as the optimal route. You can manually set a preference for RIPng. The smaller the value, the higher the preference.

Procedure

1. Enter system view.

   system-view

2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Set a preference for RIPng.

   preference { *preference* | **route-policy** *route-policy-name* } *

   By default, the preference of RIPng is 100.

Configuring RIPng route redistribution

1. Enter system view.

   system-view

2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Configure route redistribution.

   - Redistribute routes from BGP4+.

     import-route bgp4+ [*as-number*] [**allow-ibgp**] [**cost** *cost-value* | **route-policy** *route-policy-name*] *

   - Redistribute direct or static routes.

     import-route { **direct** | **static** } [**cost** *cost-value* | **route-policy** *route-policy-name*] *

   - Redistribute routes from OSPFv3, IPv6 IS-IS, or other RIPng processes.

     import-route { **isisv6** | **ospfv3** | **ripng** } [*process-id*] [**allow-direct** | **cost** *cost-value* | **route-policy** *route-policy-name*] *

   By default, RIPng does not redistribute routes.

4. (Optional.) Set a default routing metric for redistributed routes.

   default cost *cost-value*

   The default metric of redistributed routes is 0.

Tuning and optimizing the RIPng network

Setting RIPng timers

About this task

You can adjust RIPng timers to optimize the performance of the RIPng network.

Restrictions and guidelines

When you adjust RIPng timers, consider the network performance, and perform unified configurations on routers running RIPng to avoid unnecessary network traffic or route oscillation.

Procedure

1. Enter system view.

   system-view

2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Set RIPng timers.

   timers { **garbage-collect** *garbage-collect-value* | **suppress** *suppress-value* | **timeout** *timeout-value* | **update** *update-value* } *

   The default settings are as follows:

   - The update timer is 30 seconds.
   - The timeout timer is 180 seconds.
   - The suppress timer is 120 seconds.
   - The garbage-collect timer is 120 seconds.

Configuring split horizon and poison reverse

Restrictions and guidelines for split horizon and poison reverse

When you configure split horizon and poison reverse, following these restrictions and guidelines:

- If both split horizon and poison reverse are configured, only the poison reverse feature takes effect.
- Split horizon disables RIPng from sending routes through the interface where the routes were learned to prevent routing loops between neighbors. As a best practice, enable split horizon to prevent routing loops in normal cases.
- Poison reverse enables a route learned from an interface to be advertised through the interface. However, the metric of the route is set to 16, which means the route is unreachable.

Configuring split horizon

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable split horizon.

   ripng split-horizon

   By default, split horizon is enabled.

Configuring poison reverse

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable poison reverse.
   ripng poison-reverse
   By default, poison reverse is disabled.

Setting the maximum number of ECMP routes

1. Enter system view.
   system-view
2. Enter RIPng view.
   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Set the maximum number of RIPng ECMP routes.
   maximum load-balancing *number*
   By default, the maximum number of RIPng ECMP routes equals the maximum number of ECMP routes, which is configurable by using the **max-ecmp-num** command.

Configuring the RIPng packet sending rate

About this task

Perform this task to specify the interval for sending RIPng packets and the maximum number of RIPng packets that can be sent at each interval. This feature can avoid excessive RIPng packets from affecting system performance and consuming too much bandwidth.

Procedure

1. Enter system view.
   system-view
2. Enter RIPng view.
   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]
3. Configuring the RIPng packet sending rate.
   - Execute the following commands in sequence to configure the RIPng packet sending rate in RIPng view:
     ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]
     output-delay *time* **count** *count*
     By default, an interface that runs the RIPng process sends a maximum of three RIPng packets every 20 milliseconds.
   - Execute the following commands in sequence to configure the RIPng packet sending rate in interface view:
     interface *interface-type interface-number*
     ripng output-delay *time* **count** *count*
     By default, an interface uses the RIPng packet sending rate of the RIPng process that it runs.

Setting the interval for sending triggered updates

About this task

Perform this task to avoid network overhead and reduce system resource consumption caused by frequent RIPng triggered updates.

You can use the **timer triggered** command to set the maximum interval, minimum interval, and incremental interval for sending RIPng triggered updates.

For a stable network, the minimum interval is used. If network changes become frequent, the triggered update sending interval is incremented by the incremental interval × 2^{n-2} for each triggered update until the maximum interval is reached. The value *n* is the number of triggered update times.

Procedure

1. Enter system view.

   system-view

2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Set the interval for sending triggered updates.

   timer triggered *maximum-interval* [*minimum-interval* [*incremental-interval*]]

   The default maximum interval is 5 seconds, the default minimum interval is 50 milliseconds, and the default incremental interval is 200 milliseconds.

Configuring RIPng GR

About this task

GR ensures forwarding continuity when a routing protocol restarts or an active/standby switchover occurs.

Two routers are required to complete a GR process. The following are router roles in a GR process:

- **GR restarter**—Graceful restarting router. It must have GR capability.
- **GR helper**—A neighbor of the GR restarter. It helps the GR restarter to complete the GR process.

After RIPng restarts on a router, the router must learn RIPng routes again and updates its FIB table, which causes network disconnections and route reconvergence.

With the GR feature, the restarting router (known as the GR restarter) can notify the event to its GR capable neighbors. GR capable neighbors (known as GR helpers) maintain their adjacencies with the router within a configurable GR interval. During this process, the FIB table of the router does not change. After the restart, the router contacts its neighbors to retrieve its FIB.

By default, a RIPng-enabled device acts as the GR helper. Perform this task on the GR restarter.

Procedure

1. Enter system view.

   system-view

2. Enable RIPng and enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enable the GR capability for RIPng.

   graceful-restart

   By default, RIPng GR is disabled.

4. (Optional.) Set the GR interval.

   graceful-restart interval *interval*

   The default GR interval is 60 seconds.

Configuring RIPng NSR

About this task

Nonstop routing (NSR) backs up RIPng routing information from the active process to the standby process. After an active/standby switchover, NSR can complete route regeneration without tearing down adjacencies or impacting forwarding services.

To use RIPng NSR, the system must have a minimum of two IRF member devices. NSR ensures nonstop services when RIPng has redundant processes on multiple IRF member devices.

In contrast to GR, NSR does not require a neighbor device to recover routing information.

Restrictions and guidelines

When both GR and NSR are configured for RIPng, the following rules apply:

- NSR has a higher priority than GR. When the active RIPng process goes down, the device does not act as the GR restarter and performs RIPng NSR to ensure nonstop forwarding.
- The device can participate in a GR process only when it acts as the GR helper. If it performs an active/standby RIPng process switchover before the GR process finishes, the GR process might fail even if RIPng NSR is enabled on the device.

RIPng NSR enabled for a RIPng process takes effect only on that process. As a best practice, enable RIPng NSR for each process if multiple RIPng processes exist.

Procedure

1. Enter system view.

   system-view

2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enable RIPng NSR.

   non-stop-routing

   By default, RIPng NSR is disabled.

Configuring BFD for RIPng

About this task

Configuring BFD for RIPng

About this task

RIPng detects route failures by periodically sending requests. If it receives no response for a route within a certain time, RIPng considers the route unreachable. To speed up convergence, perform this task to enable BFD for RIPng. For more information about BFD, see *High Availability Configuration Guide*.

RIPng supports BFD in control packet mode. In this mode, a BFD session is established only when both ends have routes to send and BFD is enabled on the receiving interface.

Restrictions and guidelines

You need to perform this task to configure control-mode BFD detection on both the local and peer ends.

Procedure

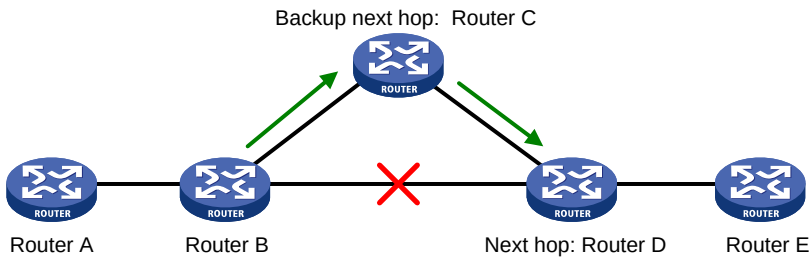
1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable BFD for RIPng.
   ripng bfd enable
   By default, BFD for RIPng is disabled.

Configuring RIPng FRR

About RIPng FRR

A link or router failure on a path can cause packet loss and even routing loop until RIPng completes routing convergence based on the new network topology. FRR enables fast rerouting to minimize the impact of link or node failures.

Figure 1 Network diagram for RIPng FRR



As shown in Figure 1, configure FRR on Router B by using a routing policy to specify a backup next hop. When the primary link fails, RIPng directs packets to the backup next hop. At the same time, RIPng calculates the shortest path based on the new network topology. Then, the device forwards packets over that path after network convergence.

Restrictions and guidelines for RIPng FRR

RIPng FRR is available only when the state of the primary link (with Layer 3 interfaces staying up) changes from bidirectional to unidirectional or down.

RIPng FRR is only effective for RIPng routes that are learned from directly connected neighbors.

Equal-cost routes do not support RIPng FRR.

Enabling RIPng FRR

1. Enter system view.
   system-view
2. Configure a routing policy.

You must specify a next hop by using the **apply ipv6 fast-reroute backup-interface** command in a routing policy and specify the routing policy for FRR.

For more information about routing policy configuration, see "Configuring routing policies."

3. Enter RIPng view.
   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]
4. Enable RIPng FRR.
   fast-reroute route-policy *route-policy-name*

   By default, RIPng FRR is disabled.

Enabling BFD for RIPng FRR

About this task

By default, RIPng FRR does not use BFD to detect primary link failures. For quicker RIPng FRR, use BFD on the primary link of redundant links to detect link failure.

Procedure

1. Enter system view.
   system-view
2. Configure the source IP address of BFD echo packets.
   bfd echo-source-ipv6 *ipv6-address*

   By default, the source IP address of BFD echo packets is not configured.

   As a best practice, do not configure the source IP address on the same network segment as any local interfaces.

   Execute this command only when configuring echo-packet-mode BFD sessions.

   For more information about this command, see *High Availability Command Reference*.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable BFD for RIPng FRR.
   ripng primary-path-detect bfd { **ctrl** | **echo** }

   By default, BFD is disabled for RIPng FRR.

Enhancing RIPng security

Configuring zero field check for RIPng packets

About this task

Some fields in the RIPng packet header must be zero. These fields are called zero fields. You can enable zero field check for incoming RIPng packets. If a zero field of a packet contains a non-zero value, RIPng does not process the packets. If you are certain that all packets are trustworthy, disable the zero field check to save CPU resources.

Procedure

1. Enter system view.
   system-view
2. Enter RIPng view.
   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Enable the zero field check for incoming RIPng packets.

   checkzero

   By default, zero field check for incoming RIPng packets is enabled.

Applying an IPsec profile

About this task

To protect routing information and prevent attacks, you can configure RIPng to authenticate protocol packets by using an IPsec profile.

An IPsec profile contains inbound and outbound security parameter indexes (SPIs). RIPng compares the inbound SPI defined in the IPsec profile with the outbound SPI in the received packets. Two RIPng devices accept the packets from each other and establish a neighbor relationship only if the SPIs are the same and the relevant IPsec profiles match.

For more information about IPsec profiles, see *Security Configuration Guide*.

Restrictions and guidelines

You can apply an IPsec profile to a RIPng process or to an interface. If an interface and its process each have an IPsec profile, the IPsec profile applied to the interface takes effect.

Applying an IPsec profile to a process

1. Enter system view.

   system-view

2. Enter RIPng view.

   ripng [*process-id*] [**vpn-instance** *vpn-instance-name*]

3. Apply an IPsec profile to the process.

   enable ipsec-profile *profile-name*

   By default, no IPsec profile is applied.

Applying an IPsec profile to an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Apply an IPsec profile to the interface.

   ripng ipsec-profile *profile-name*

   By default, no IPsec profile is applied.

Display and maintenance commands for RIPng

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display configuration information for a RIPng process.	**display ripng** [*process-id*]
Display RIPng GR information.	**display ripng** [*process-id*] **graceful-restart**
Display RIPng NSR information.	**display ripng** [*process-id*]

Task	Command
	non-stop-routing
Display routes in the RIPng database.	**display ripng** *process-id* **database** [*ipv6-address prefix-length*]
Display interface information for a RIPng process.	**display ripng** *process-id* **interface** [*interface-type interface-number*]
Display neighbor information for a RIPng process.	**display ripng** *process-id* **neighbor** [*interface-type interface-number*]
Display the routing information for a RIPng process.	**display ripng** *process-id* **route** [*ipv6-address prefix-length* [**verbose**] | **peer** *ipv6-address* | **statistics**]
Restart a RIPng process.	**reset ripng** *process-id* **process**
Clear statistics for a RIPng process.	**reset ripng** *process-id* **statistics**

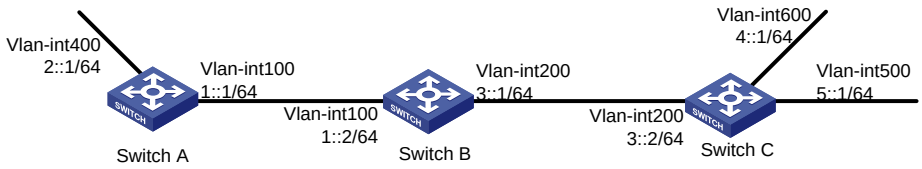
RIPng configuration examples

Example: Configuring basic RIPng

Network configuration

As shown in Figure 2, Switch A, Switch B, and Switch C run RIPng. Configure route filtering on Switch B to accept all received routes except the route 2::/64 and to advertise only the route 4::/64.

Figure 2 Network diagram



Procedure

1. Configure IPv6 addresses for the interfaces. (Details not shown.)
2. Configure basic RIPng settings:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] ripng 1
   [SwitchA-ripng-1] quit
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] ripng 1 enable
   [SwitchA-Vlan-interface100] quit
   [SwitchA] interface vlan-interface 400
   [SwitchA-Vlan-interface400] ripng 1 enable
   [SwitchA-Vlan-interface400] quit
   ```

   # Configure Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] ripng 1
   ```

```
[SwitchB-ripng-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ripng 1 enable
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ripng 1
[SwitchC-ripng-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] ripng 1 enable
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 500
[SwitchC-Vlan-interface500] ripng 1 enable
[SwitchC-Vlan-interface500] quit
[SwitchC] interface vlan-interface 600
[SwitchC-Vlan-interface600] ripng 1 enable
[SwitchC-Vlan-interface600] quit
```

Display the RIPng routing table on Switch B.

```
[SwitchB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect, D – Direct
                O - Optimal, F - Flush to RIB
 ------------------------------------------------------------------

 Peer FE80::20F:E2FF:FE23:82F5 on Vlan-interface100
 Destination 2::/64,
     via FE80::20F:E2FF:FE23:82F5, cost 1, tag 0, AOF, 6 secs
 Peer FE80::20F:E2FF:FE00:100  on Vlan-interface200
 Destination 4::/64,
     via FE80::20F:E2FF:FE00:100, cost 1, tag 0, AOF, 11 secs
 Destination 5::/64,
     via FE80::20F:E2FF:FE00:100, cost 1, tag 0, AOF, 11
 Local route
 Destination 1::/64,
     via ::, cost 0, tag 0, DOF
 Destination 3::/64,
     via ::, cost 0, tag 0, DOF
```

Display the RIPng routing table on Switch A.

```
[SwitchA] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect, D – Direct
                O - Optimal, F - Flush to RIB
 ------------------------------------------------------------------

 Peer FE80::200:2FF:FE64:8904 on Vlan-interface100
 Destination 3::/64,
     via FE80::200:2FF:FE64:8904, cost 1, tag 0, AOF, 31 secs
```

```
 Destination 4::/64,
     via FE80::200:2FF:FE64:8904, cost 2, tag 0, AOF, 31 secs
 Destination 5::/64,
     via FE80::200:2FF:FE64:8904, cost 2, tag 0, AOF, 31 secs
 Local route
 Destination 2::/64,
     via ::, cost 0, tag 0, DOF
 Destination 1::/64,
     via ::, cost 0, tag 0, DOF
```

3. Configure route filtering:

Use IPv6 prefix lists on Switch B to filter received and redistributed routes.

```
[SwitchB] ipv6 prefix-list aaa permit 4:: 64
[SwitchB] ipv6 prefix-list bbb deny 2:: 64
[SwitchB] ipv6 prefix-list bbb permit :: 0 less-equal 128
[SwitchB] ripng 1
[SwitchB-ripng-1] filter-policy prefix-list aaa export
[SwitchB-ripng-1] filter-policy prefix-list bbb import
[SwitchB-ripng-1] quit
```

Display RIPng routing tables on Switch B and Switch A.

```
[SwitchB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect, D – Direct
                O - Optimal, F - Flush to RIB
 ------------------------------------------------------------

 Peer FE80::1:100 on Vlan-interface100

 Peer FE80::3:200 on Vlan-interface200
 Destination 4::/64,
     via FE80::2:200, cost 1, tag 0, AOF, 11 secs
 Destination 5::/64,
     via FE80::2:200, cost 1, tag 0, AOF, 11 secs
 Local route
 Destination 1::/64,
     via ::, cost 0, tag 0, DOF
 Destination 3::/64,
     via ::, cost 0, tag 0, DOF
[SwitchA] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect, D – Direct
                O - Optimal, F - Flush to RIB
 ------------------------------------------------------------

 Peer FE80::2:100 on Vlan-interface100
 Destination 4::/64,
     via FE80::1:100, cost 2, tag 0, AOF, 2 secs
 Local route
 Destination 1::/64,
     via ::, cost 0, tag 0, DOF
 Destination 2::/64,
```

```
via ::, cost 0, tag 0, DOF
```

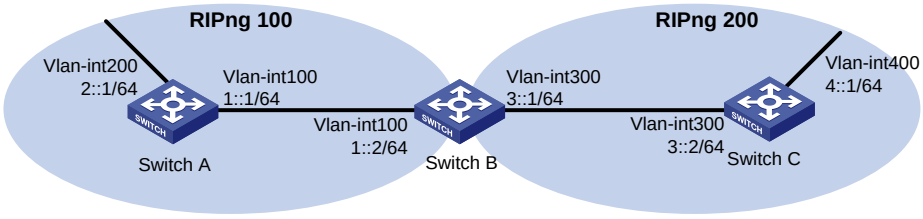
Example: Configuring RIPng route redistribution

Network configuration

As shown in Figure 3, Switch B communicates with Switch A through RIPng 100 and with Switch C through RIPng 200.

Configure route redistribution on Switch B, so the two RIPng processes can redistribute routes from each other.

Figure 3 Network diagram



Procedure

1. Configure IPv6 addresses for the interfaces. (Details not shown.)
2. Configure basic RIPng settings:

   # Enable RIPng 100 on Switch A.

```
<SwitchA> system-view
[SwitchA] ripng 100
[SwitchA-ripng-100] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ripng 100 enable
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ripng 100 enable
[SwitchA-Vlan-interface200] quit
```

   # Enable RIPng 100 and RIPng 200 on Switch B.

```
<SwitchB> system-view
[SwitchB] ripng 100
[SwitchB-ripng-100] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ripng 100 enable
[SwitchB-Vlan-interface100] quit
[SwitchB] ripng 200
[SwitchB-ripng-200] quit
[SwitchB] interface vlan-interface 300
[SwitchB-Vlan-interface300] ripng 200 enable
[SwitchB-Vlan-interface300] quit
```

   # Enable RIPng 200 on Switch C.

```
<SwitchC> system-view
[SwitchC] ripng 200
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] ripng 200 enable
```

```
[SwitchC-Vlan-interface300] quit
[SwitchC] interface vlan-interface 400
[SwitchC-Vlan-interface400] ripng 200 enable
[SwitchC-Vlan-interface400] quit
```

Display the routing table on Switch A.

```
[SwitchA] display ipv6 routing-table

Destinations : 7 Routes : 7

Destination: ::1/128                                Protocol  : Direct
NextHop    : ::1                                    Preference: 0
Interface  : InLoop0                                Cost      : 0

Destination: 1::/64                                 Protocol  : Direct
NextHop    : ::                                     Preference: 0
Interface  : Vlan100                                Cost      : 0

Destination: 1::1/128                               Protocol  : Direct
NextHop    : ::1                                    Preference: 0
Interface  : InLoop0                                Cost      : 0

Destination: 2::/64                                 Protocol  : Direct
NextHop    : ::                                     Preference: 0
Interface  : Vlan200                                Cost      : 0

Destination: 2::1/128                               Protocol  : Direct
NextHop    : ::1                                    Preference: 0
Interface  : InLoop0                                Cost      : 0

Destination: FE80::/10                              Protocol  : Direct
NextHop    : ::                                     Preference: 0
Interface  : NULL0                                  Cost      : 0

Destination: FF00::/8                               Protocol  : Direct
NextHop    : ::                                     Preference: 0
Interface  : NULL0                                  Cost      : 0
```

3. Configure RIPng route redistribution:

Configure route redistribution between the two RIPng processes on Switch B.

```
[SwitchB] ripng 100
[SwitchB-ripng-100] import-route ripng 200
[SwitchB-ripng-100] quit
[SwitchB] ripng 200
[SwitchB-ripng-200] import-route ripng 100
[SwitchB-ripng-200] quit
```

Display the routing table on Switch A.

```
[SwitchA] display ipv6 routing-table

Destinations : 8 Routes : 8
```

```
Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 1::/64                                     Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan100                                    Cost      : 0

Destination: 1::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 2::/64                                     Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan200                                    Cost      : 0

Destination: 2::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 4::/64                                     Protocol  : RIPng
NextHop    : FE80::200:BFF:FE01:1C02                    Preference: 100
Interface  : Vlan100                                    Cost      : 1

Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0

Destination: FF00::/8                                   Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0
```

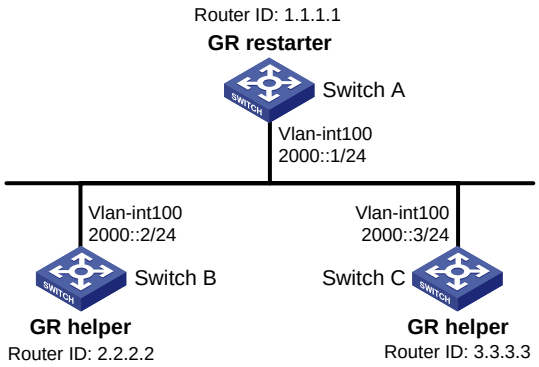
Example: Configuring RIPng GR

Network configuration

As shown in Figure 4, Switch A, Switch B, and Switch C learn IPv6 routing information through RIPng.

Configure Switch A as the GR restarter. Configure Switch B and Switch C as the GR helpers to synchronize their routing tables with Switch A by using GR.

Figure 4 Network diagram



Procedure

1. Configure IPv6 addresses for the interfaces. (Details not shown.)
2. Configure RIPng on the switches to ensure the following: (Details not shown.)
   - Switch A, Switch B, and Switch C can communicate with each other at Layer 3.
   - Dynamic route update can be implemented among them with RIPng.
3. Enable RIPng GR on Switch A.

```
<SwitchA> system-view
[SwitchA] ripng 1
[SwitchA-ripng-1] graceful-restart
```

Verifying the configuration

Restart RIPng or trigger an active/standby switchover, and then display GR status on Switch A.

```
<SwitchA> display ripng 1 graceful-restart
RIPng process: 1
 Graceful Restart capability     : Enabled
 Current GR state                : Normal
 Graceful Restart period         : 60  seconds
 Graceful Restart remaining time: 0   seconds
```

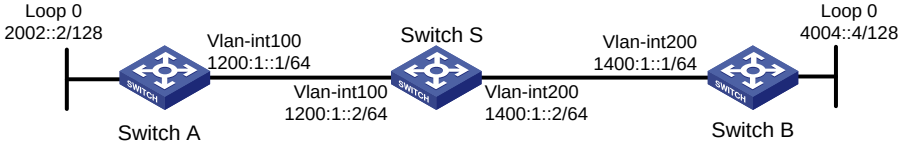
Example: Configuring RIPng NSR

Network configuration

As shown in Figure 5, Switch S, Switch A, and Switch B learn IPv6 routing information through RIPng.

Enable RIPng NSR on Switch S to ensure correct routing when an active/standby switchover occurs on Switch S.

Figure 5 Network diagram



Procedure

1. Configure IPv6 addresses for the interfaces. (Details not shown.)

2. Configure RIPng on the switches to ensure the following: (Details not shown.)
   - Switch S, Switch A, and Switch B can communicate with each other at Layer 3.
   - Dynamic route update can be implemented among them with RIPng.
3. Enable RIPng NSR.

   # Enable RIPng NSR on Switch S.

```
<SwitchS> system-view
[SwitchS] ripng 1
[SwitchS-ripng-1] non-stop-routing
[SwitchS-ripng-1] quit
```

Verifying the configuration

Perform an active/standby switchover on Switch S.

```
[SwitchS] placement reoptimize
Predicted changes to the placement
Service Group(instance name)                        Cur location  New location
--------------------------------------------------------------------------
lb                                                  1/0           1/0
lsm                                                 1/0           1/0
slsp                                                1/0           1/0
rib6                                                1/0           1/0
routepolicy                                         1/0           1/0
rib                                                 1/0           1/0
staticroute6                                        1/0           1/0
staticroute                                         1/0           1/0
eviisis                                             1/0           1/0
ripng                                               1/0           1/0
Continue? [y/n]:y
Re-optimization of the placement start. You will be notified on completion
Re-optimization of the placement complete. Use 'display placement' to view the new
placement
```

During the switchover period, display RIPng neighbors on Switch A to verify the neighbor relationship between Switch A and Switch S.

```
[SwitchA] display ripng 1 neighbor
 Neighbor Address: FE80::AE45:5CE7:422E:2867
     Interface  : Vlan-interface100
     Version    : RIPng version 1     Last update: 00h00m23s
     BFD session: None
     Bad packets: 0                   Bad routes : 0
```

Display RIPng routes on Switch A to verify if Switch A has a route to the loopback interface on Switch B.

```
[SwitchA] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect, D - Direct
                O - Optimal, F - Flush to RIB
 ----------------------------------------------------------------
 Peer FE80::AE45:5CE7:422E:2867 on Vlan-interface100
 Destination 1400:1::/64,
     via FE80::AE45:5CE7:422E:2867, cost 1, tag 0, AOF, 1 secs
 Destination 4004::4/128,
```

```
    via FE80::AE45:5CE7:422E:2867, cost 2, tag 0, AOF, 1 secs
 Local route
 Destination 2002::2/128,
    via ::, cost 0, tag 0, DOF
 Destination 1200:1::/64,
    via ::, cost 0, tag 0, DOF
```

Display RIPng neighbors on Switch B to verify the neighbor relationship between Switch B and Switch S.

```
[SwitchB] display ripng 1 neighbor
 Neighbor Address: FE80::20C:29FF:FECE:6277
    Interface  : Vlan-interface200
    Version    : RIPng version 1     Last update: 00h00m18s
    BFD session: None
    Bad packets: 0                   Bad routes : 0
```

Display RIPng routes on Switch B to verify if Switch B has a route to the loopback interface on Switch A.

```
[SwitchB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect, D - Direct
                O - Optimal, F - Flush to RIB
 ----------------------------------------------------------------
 Peer FE80::20C:29FF:FECE:6277 on Vlan-interface200
 Destination 2002::2/128,
    via FE80::20C:29FF:FECE:6277, cost 2, tag 0, AOF, 24 secs
 Destination 1200:1::/64,
    via FE80::20C:29FF:FECE:6277, cost 1, tag 0, AOF, 24 secs
 Local route
 Destination 4004::4/128,
    via ::, cost 0, tag 0, DOF
 Destination 1400:1::/64,
    via ::, cost 0, tag 0, DOF
```

The output shows the following when an active/standby switchover occurs on Switch S:

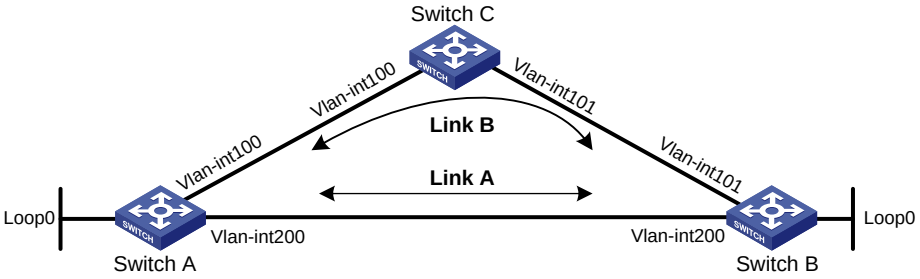
- The neighbor relationships and routing information on Switch A and Switch B have not changed.
- The traffic from Switch A to Switch B has not been impacted.

Example: Configuring RIPng FRR

Network configuration

As shown in Figure 6, Switch A, Switch B, and Switch C run RIPng. Configure RIPng FRR so that when Link A becomes unidirectional, traffic can be switched to Link B immediately.

Figure 6 Network diagram



Device	Interface	IP address
Switch A	VLAN-interface 100	1::1/64
Switch A	VLAN-interface 200	2::1/64
Switch A	Loopback 0	10::1/128
Switch B	VLAN-interface 101	3::1/64
Switch B	VLAN-interface 200	2::2/64
Switch B	Loopback 0	20::1/128
Switch C	VLAN-interface 100	1::2/64
Switch C	VLAN-interface 101	3::2/64

Procedure

1. Configure IPv6 addresses for the interfaces on the switches. (Details not shown.)
2. Configure RIPng on the switches to make sure Switch A, Switch B, and Switch C can communicate with each other at Layer 3. (Details not shown.)
3. Configure RIPng FRR:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6 prefix-list abc index 10 permit 20::1 128
[SwitchA] route-policy frr permit node 10
[SwitchA-route-policy-frr-10] if-match ipv6 address prefix-list abc
[SwitchA-route-policy-frr-10] apply ipv6 fast-reroute backup-interface
vlan-interface 100 backup-nexthop fe80::6eab:8bff:fed4:205
[SwitchA-route-policy-frr-10] quit
[SwitchA] ripng 1
[SwitchA-ripng-1] fast-reroute route-policy frr
[SwitchA-ripng-1] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6 prefix-list abc index 10 permit 10::1 128
[SwitchB] route-policy frr permit node 10
[SwitchB-route-policy-frr-10] if-match ipv6 address prefix-list abc
[SwitchB-route-policy-frr-10] apply ipv6 fast-reroute backup-interface
vlan-interface 101 backup-nexthop fe80::6eab:8bff:fed4:206
[SwitchB-route-policy-frr-10] quit
[SwitchB] ripng 1
[SwitchB-ripng-1] fast-reroute route-policy frr
```

```
[SwitchB-ripng-1] quit
```

Verifying the configuration

Display the route 20::1/128 on Switch A to view the backup next hop information.

```
[SwitchA] display ipv6 routing-table 20::1 128 verbose

Summary count : 1

Destination: 20::1/128
   Protocol: RIPng
 Process ID: 1
   SubProtID: 0x0                  Age: 00h17m42s
        Cost: 1              Preference: 100
       IpPre: N/A            QosLocalID: N/A
         Tag: 0                   State: Inactive Adv
   OrigTblID: 0x0               OrigVrf: default-vrf
     TableID: 0xa                OrigAs: 0
       NibID: 0x22000003         LastAs: 0
      AttrID: 0xffffffff       Neighbor: FE80::6EAB:90FF:FEF3:305
       Flags: 0x41          OrigNextHop: FE80::6EAB:90FF:FEF3:305
       Label: NULL          RealNextHop: FE80::6EAB:90FF:FEF3:305
    BkLabel: NULL             BkNextHop: FE80::6EAB:8BFF:FED4:205
    SRLabel: NULL             BkSRLabel: NULL
   SIDIndex: NULL               InLabel: NULL
  Tunnel ID: Invalid          Interface: Vlan-interface200
BkTunnel ID: Invalid        BkInterface: Vlan-interface100
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                PathID: 0x0
   LinkCost: 0               MicroSegID: 0
RealFIRType: Normal           RealThres: 0
```

Display the route 10::1/128 on Switch B to view the backup next hop information.

```
[SwitchB] display ipv6 routing-table 10::1 128 verbose

Summary count : 1

Destination: 10::1/128
   Protocol: RIPng
 Process ID: 1
  SubProtID: 0x0                  Age: 00h22m34s
        Cost: 1              Preference: 100
       IpPre: N/A            QosLocalID: N/A
         Tag: 0                   State: Inactive Adv
  OrigTblID: 0x0               OrigVrf: default-vrf
     TableID: 0xa                OrigAs: 0
       NibID: 0x22000001         LastAs: 0
      AttrID: 0xffffffff       Neighbor: FE80::6EAB:81FF:FE67:106
       Flags: 0x41          OrigNextHop: FE80::6EAB:81FF:FE67:106
       Label: NULL          RealNextHop: FE80::6EAB:81FF:FE67:106
    BkLabel: NULL             BkNextHop: FE80::6EAB:8BFF:FED4:206
```

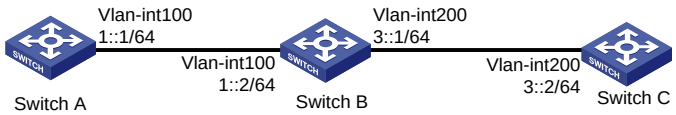
```
    SRLabel: NULL              BkSRLabel: NULL
   SIDIndex: NULL                InLabel: NULL
  Tunnel ID: Invalid           Interface: Vlan-interface200
BkTunnel ID: Invalid         BkInterface: Vlan-interface101
   FtnIndex: 0x0           TrafficIndex: N/A
  Connector: N/A                 PathID: 0x0
   LinkCost: 0                MicroSegID: 0
RealFIRType: Normal            RealThres: 0
```

Example: Configuring RIPng IPsec profile

Network configuration

As shown in Figure 7, configure RIPng on the switches, and configure IPsec profiles on the switches to authenticate and encrypt protocol packets.

Figure 7 Network diagram



Procedure

1. Configure IPv6 addresses for the interfaces. (Details not shown.)
2. Configure basic RIPng settings:

   # Configure Switch A.

   ```
   <SwitchA> system-view
   [SwitchA] ripng 1
   [SwitchA-ripng-1] quit
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] ripng 1 enable
   [SwitchA-Vlan-interface100] quit
   ```

   # Configure Switch B.

   ```
   <SwitchB> system-view
   [SwitchB] ripng 1
   [SwitchB-ripng-1] quit
   [SwitchB] interface vlan-interface 200
   [SwitchB-Vlan-interface200] ripng 1 enable
   [SwitchB-Vlan-interface200] quit
   [SwitchB] interface vlan-interface 100
   [SwitchB-Vlan-interface100] ripng 1 enable
   [SwitchB-Vlan-interface100] quit
   ```

   # Configure Switch C.

   ```
   <SwitchC> system-view
   [SwitchC] ripng 1
   [SwitchC-ripng-1] quit
   [SwitchC] interface vlan-interface 200
   [SwitchC-Vlan-interface200] ripng 1 enable
   [SwitchC-Vlan-interface200] quit
   ```

3. Configure RIPng IPsec profiles:

- On Switch A:

  # Create an IPsec transform set named **protrf1**.

  ```
  [SwitchA] ipsec transform-set protrf1
  ```

  # Specify the ESP encryption and authentication algorithms.

  ```
  [SwitchA-ipsec-transform-set-protrf1] esp encryption-algorithm 3des-cbc
  [SwitchA-ipsec-transform-set-protrf1] esp authentication-algorithm md5
  ```

  # Specify transport mode for encapsulation.

  ```
   [SwitchA-ipsec-transform-set-protrf1] encapsulation-mode transport
  [SwitchA-ipsec-transform-set-protrf1] quit
  ```

  # Create a manual IPsec profile named **profile001**.

  ```
  [SwitchA] ipsec profile profile001 manual
  ```

  # Reference IPsec transform set **protrf1**.

  ```
  [SwitchA-ipsec-profile-profile001-manual] transform-set protrf1
  ```

  # Configure the inbound and outbound SPIs for ESP.

  ```
  [SwitchA-ipsec-profile-profile001-manual] sa spi inbound esp 256
  [SwitchA-ipsec-profile-profile001-manual] sa spi outbound esp 256
  ```

  # Configure the inbound and outbound SA keys for ESP.

  ```
  [SwitchA-ipsec-profile-profile001-manual] sa string-key inbound esp simple abc
  [SwitchA-ipsec-profile-profile001-manual] sa string-key outbound esp simple abc
  [SwitchA-ipsec-profile-profile001-manual] quit
  ```

- On Switch B:

  # Create an IPsec transform set named **protrf1**.

  ```
  [SwitchB] ipsec transform-set protrf1
  ```

  # Specify the ESP encryption and authentication algorithms.

  ```
  [SwitchB-ipsec-transform-set-protrf1] esp encryption-algorithm 3des-cbc
  [SwitchB-ipsec-transform-set-protrf1] esp authentication-algorithm md5
  ```

  # Specify transport mode for encapsulation.

  ```
   [SwitchB-ipsec-transform-set-protrf1] encapsulation-mode transport
  [SwitchB-ipsec-transform-set-protrf1] quit
  ```

  # Create a manual IPsec profile named **profile001**.

  ```
  [SwitchB] ipsec profile profile001 manual
  ```

  # Reference IPsec transform set **protrf1**.

  ```
  [SwitchB-ipsec-profile-profile001-manual] transform-set protrf1
  ```

  # Configure the inbound and outbound SPIs for ESP.

  ```
  [SwitchB-ipsec-profile-profile001-manual] sa spi inbound esp 256
  [SwitchB-ipsec-profile-profile001-manual] sa spi outbound esp 256
  ```

  # Configure the inbound and outbound SA keys for ESP.

  ```
  [SwitchB-ipsec-profile-profile001-manual] sa string-key inbound esp simple abc
  [SwitchB-ipsec-profile-profile001-manual] sa string-key outbound esp simple abc
  [SwitchB-ipsec-profile-profile001-manual] quit
  ```

- On Switch C:

  # Create an IPsec transform set named **protrf1**.

  ```
  [SwitchC] ipsec transform-set protrf1
  ```

  # Specify the ESP encryption and authentication algorithms.

  ```
  [SwitchC-ipsec-transform-set-protrf1] esp encryption-algorithm 3des-cbc
  ```

```
[SwitchC-ipsec-transform-set-protrf1] esp authentication-algorithm md5
```

Specify transport mode for encapsulation.

```
 [SwitchC-ipsec-transform-set-protrf1] encapsulation-mode transport
[SwitchC-ipsec-transform-set-protrf1] quit
```

Create a manual IPsec profile named **profile001**.

```
[SwitchC] ipsec profile profile001 manual
```

Reference IPsec transform set **protrf1**.

```
[SwitchC-ipsec-profile-profile001-manual] transform-set protrf1
```

Configure the inbound and outbound SPIs for ESP.

```
[SwitchC-ipsec-profile-profile001-manual] sa spi inbound esp 256
[SwitchC-ipsec-profile-profile001-manual] sa spi outbound esp 256
```

Configure the inbound and outbound SA keys for ESP.

```
[SwitchC-ipsec-profile-profile001-manual] sa string-key inbound esp simple abc
[SwitchC-ipsec-profile-profile001-manual] sa string-key outbound esp simple abc
[SwitchC-ipsec-profile-profile001-manual] quit
```

4. Apply the IPsec profiles to the RIPng process on each device:
   - On Switch A:

     ```
     [SwitchA] ripng 1
     [SwitchA-ripng-1] enable ipsec-profile profile001
     [SwitchA-ripng-1] quit
     ```

   - On Switch B:

     ```
     [SwitchB] ripng 1
     [SwitchB-ripng-1] enable ipsec-profile profile001
     [SwitchB-ripng-1] quit
     ```

   - On Switch C:

     ```
     [SwitchC] ripng 1
     [SwitchC-ripng-1] enable ipsec-profile profile001
     [SwitchC-ripng-1] quit
     ```

Verifying the configuration

Verify that the RIPng packets between Switches A, B and C are protected by IPsec. (Details not shown.)

Contents

Configuring OSPFv3 1
About OSPFv3 1
Comparison of OSPFv3 with OSPFv2 1
OSPFv3 packets 1
OSPFv3 LSA types 1
Protocols and standards 2
OSPFv3 tasks at a glance 2
Enabling OSPFv3 4
Establishing OSPFv3 neighbor relationships on an M-LAG network 4
Configuring OSPFv3 area parameters 5
About OSPFv3 areas 5
Configuring a stub area 6
Configuring an NSSA area 6
Configuring an OSPFv3 virtual link 7
Configuring OSPFv3 network types 7
Restrictions and guidelines for OSPFv3 network type configuration 7
Setting the broadcast network type for an OSPFv3 interface 7
Setting the NBMA network type for an OSPFv3 interface 8
Setting the P2MP network type for an OSPFv3 interface 8
Setting the P2P network type for an OSPFv3 interface 9
Configuring OSPFv3 route control 9
Configuring OSPFv3 inter-area route summarization 9
Configuring redistributed route summarization 9
Configuring OSPFv3 received route filtering 10
Configuring Inter-Area-Prefix LSA filtering 10
Advertising the maximum link cost to neighbors within the specified period 11
Setting an OSPFv3 cost for an interface 11
Associating OSPFv3 with Track 12
Setting the maximum number of OSPFv3 ECMP routes 14
Setting a preference for OSPFv3 15
Configuring OSPFv3 route redistribution 15
Configuring default route redistribution 16
Setting OSPFv3 timers 16
Setting OSPFv3 packet timers 16
Setting the LSA transmission delay 17
Setting the OSPFv3 route calculation interval 17
Setting the LSA generation interval 18
Setting the LSU transmit rate 19
Setting a DR priority for an interface 19
Configuring OSPFv3 packet parameters 19
Ignoring MTU check for DD packets 19
Disabling interfaces from receiving and sending OSPFv3 packets 20
Configuring prefix suppression 20
About prefix suppression 20
Restrictions and guidelines for prefix suppression 21
Configuring prefix suppression for an OSPFv3 process 21
Configuring prefix suppression for an interface 21
Configuring OSPFv3 advanced features 21
Configuring a stub router 21
Configuring OSPFv3 isolation 22
Configuring OSPFv3 shutdown 23
Configuring OSPFv3 GR 24
About OSPFv3 GR 24
Restrictions and guidelines for OSPFv3 GR 24
Configuring GR restarter 24
Configuring GR helper 24
Triggering OSPFv3 GR 25

Configuring OSPFv3 NSR 25
Configuring BFD for OSPFv3 26
Configuring OSPFv3 FRR 26
About OSPFv3 FRR 26
Configuring OSPFv3 FRR to use the LFA algorithm to calculate a backup next hop 27
Configuring OSPFv3 FRR to use a backup next hop in a routing policy 27
Configuring BFD for OSPFv3 FRR 28
Configuring BFD echo packet mode for OSPFv3 FRR 28
Enhancing OSPFv3 security 29
Configuring OSPFv3 authentication 29
Applying an IPsec profile for authenticating OSPFv3 packets 30
Configuring OSPFv3 logging and SNMP notifications 31
Enabling logging for neighbor state changes 31
Setting the maximum number of OSPFv3 logs 31
Configuring OSPFv3 network management 32
Display and maintenance commands for OSPFv3 33
OSPFv3 configuration examples 34
Example: Configuring OSPFv3 stub area 34
Example: Configuring OSPFv3 NSSA area 40
Example: Configuring OSPFv3 DR election 43
Example: Configuring OSPFv3 route redistribution 46
Example: Configuring OSPFv3 route summarization 49
Example: Configuring OSPFv3 GR 52
Example: Configuring OSPFv3 NSR 54
Example: Configuring BFD for OSPFv3 55
Example: Configuring OSPFv3 FRR 57
Example: Configuring OSPFv3 IPsec profile 61

Configuring OSPFv3

About OSPFv3

This chapter describes how to configure Open Shortest Path First version 3 (OSPFv3) for an IPv6 network.

Comparison of OSPFv3 with OSPFv2

OSPFv3 and OSPFv2 have the following in common:

- 32-bit router ID and area ID.
- Hello, Database Description (DD), Link State Request (LSR), Link State Update (LSU), Link State Acknowledgment (LSAck).
- Mechanisms for finding neighbors and establishing adjacencies.
- Mechanisms for advertising and aging LSAs.

OSPFv3 and OSPFv2 have the following differences:

- OSPFv3 runs on a per-link basis and uses link-local addresses for neighbor relationship establishment and LSA database synchronization. All OSPFv3 interfaces use their link-local addresses to forward packets except for virtual links.
- OSPFv3 supports running multiple processes on an interface, but OSPFv2 does not support.
- OSPFv3 identifies neighbors by router ID. OSPFv2 identifies neighbors by IP address.

For more information about OSPFv2, see "Configuring OSPF."

OSPFv3 packets

OSPFv3 uses the following packet types:

- **Hello**—Periodically sent to find and maintain neighbors, containing timer values, information about the DR, BDR, and known neighbors.
- **DD**—Describes the digest of each LSA in the LSDB, exchanged between two routers for data synchronization.
- **LSR**—Requests needed LSAs from the neighbor. After exchanging the DD packets, the two routers know which LSAs of the neighbor are missing from their LSDBs. They then send an LSR packet to each other, requesting the missing LSAs. The LSA packet contains the digest of the missing LSAs.
- **LSU**—Transmits the requested LSAs to the neighbor.
- **LSAck**—Acknowledges received LSU packets.

OSPFv3 LSA types

OSPFv3 sends routing information in LSAs. The following LSAs are commonly used:

- **Router LSA**—Type-1 LSA, originated by all routers. This LSA describes the collected states of the router's interfaces to an area, and is flooded throughout a single area only.
- **Network LSA**—Type-2 LSA, originated for broadcast and NBMA networks by the DR. This LSA contains the list of routers connected to the network, and is flooded throughout a single area only.

- **Inter-Area-Prefix LSA**—Type-3 LSA, originated by ABRs and flooded throughout the LSA's associated area. Each Inter-Area-Prefix LSA describes a route with IPv6 address prefix to a destination outside the area, yet still inside the AS.
- **Inter-Area-Router LSA**—Type-4 LSA, originated by ABRs and flooded throughout the LSA's associated area. Each Inter-Area-Router LSA describes a route to ASBR.
- **AS External LSA**—Type-5 LSA, originated by ASBRs, and flooded throughout the AS, except stub areas and Not-So-Stubby Areas (NSSAs). Each AS External LSA describes a route to another AS. A default route can be described by an AS External LSA.
- **NSSA LSA**—Type-7 LSA, originated by ASBRs in NSSAs and flooded throughout a single NSSA. NSSA LSAs describe routes to other ASs.
- **Link LSA**—Type-8 LSA. A router originates a separate Link LSA for each attached link. Link LSAs have link-local flooding scope. Each Link LSA describes the IPv6 address prefix of the link and Link-local address of the router.
- **Intra-Area-Prefix LSA**—Type-9 LSA. Each Intra-Area-Prefix LSA contains IPv6 prefix information on a router, stub area, or transit area information, and has area flooding scope. It was introduced because Router LSAs and Network LSAs contain no address information.
- **Grace LSA**—Type-11 LSA, generated by a GR restarter at reboot and transmitted on the local link. The GR restarter describes the cause and interval of the reboot in the Grace LSA to notify its neighbors that it performs a GR operation.

Protocols and standards

- RFC 2328, *OSPF Version 2*
- RFC 3101, *OSPF Not-So-Stubby Area (NSSA) Option*
- RFC 4552, *Authentication/Confidentiality for OSPFv3*
- RFC 5187, *OSPFv3 Graceful Restart*
- RFC 5286, *Basic Specification for IP Fast Reroute: Loop-Free Alternates*
- RFC 5329, *Traffic Engineering Extensions to OSPF Version 3*
- RFC 5340, *OSPF for IPv6*
- RFC 5523, *OSPFv3-Based Layer 1 VPN Auto-Discovery*
- RFC 5643, *Management Information Base for OSPFv3*
- RFC 6506, *Supporting Authentication Trailer for OSPFv3*
- RFC 6565, *OSPFv3 as a Provider Edge to Customer Edge (PE-CE) Routing Protocol*
- RFC 6969, *OSPFv3 Instance ID Registry Update*
- RFC 7166, *Supporting Authentication Trailer for OSPFv3*

OSPFv3 tasks at a glance

To configure OSPFv3, perform the following tasks:

1. Enabling OSPFv3
2. (Optional.) Establishing OSPFv3 neighbor relationships on an M-LAG network
3. (Optional.) Configuring OSPFv3 area parameters
   - Configuring a stub area
   - Configuring an NSSA area
   - Configuring an OSPFv3 virtual link

     Perform this task on an ABR to create a virtual link when connectivity cannot be maintained between a non-backbone area and the backbone, or within the backbone.

4. (Optional.) Configuring OSPFv3 network types
   - Setting the broadcast network type for an OSPFv3 interface
   - Setting the NBMA network type for an OSPFv3 interface
   - Setting the P2MP network type for an OSPFv3 interface
   - Setting the P2P network type for an OSPFv3 interface
5. (Optional.) Configuring OSPFv3 route control
   - Configuring OSPFv3 inter-area route summarization
   - Configuring redistributed route summarization
   - Configuring OSPFv3 received route filtering
   - Configuring Inter-Area-Prefix LSA filtering
   - Setting an OSPFv3 cost for an interface
   - Advertising the maximum link cost to neighbors within the specified period
   - Associating OSPFv3 with Track
   - Setting the maximum number of OSPFv3 ECMP routes
   - Setting a preference for OSPFv3
   - Configuring OSPFv3 route redistribution
6. (Optional.) Setting OSPFv3 timers
   - Setting OSPFv3 packet timers
   - Setting the LSA transmission delay
   - Setting the OSPFv3 route calculation interval
   - Setting the LSA generation interval
   - Setting the LSU transmit rate
7. (Optional.) Setting a DR priority for an interface
8. (Optional.) Configuring OSPFv3 packet parameters
   - Ignoring MTU check for DD packets
   - Disabling interfaces from receiving and sending OSPFv3 packets
9. (Optional.) Configuring prefix suppression
10. (Optional.) Configuring OSPFv3 advanced features
    - Configuring a stub router
    - Configuring OSPFv3 isolation
    - Configuring OSPFv3 shutdown
11. (Optional.) Enhancing OSPFv3 availability
    - Configuring OSPFv3 GR
    - Configuring OSPFv3 NSR
    - Configuring BFD for OSPFv3
    - Configuring OSPFv3 FRR
12. (Optional.) Enhancing OSPFv3 security
    - Configuring OSPFv3 authentication
    - Applying an IPsec profile for authenticating OSPFv3 packets
13. (Optional.) Configuring OSPFv3 logging and SNMP notifications
    - Enabling logging for neighbor state changes
    - Setting the maximum number of OSPFv3 logs
    - Configuring OSPFv3 network management

Enabling OSPFv3

About this task

To enable an OSPFv3 process on a router:

1. Enable the OSPFv3 process globally.
2. Assign the OSPFv3 process a router ID.
3. Enable the OSPFv3 process on related interfaces.

An OSPFv3 process ID has only local significance. Process 1 on a router can exchange packets with process 2 on another router.

OSPFv3 requires you to manually specify a router ID for each router in an AS. Make sure all assigned router IDs in the AS are unique.

Restrictions and guidelines

If a router runs multiple OSPFv3 processes, you must specify a unique router ID for each process.

Procedure

1. Enter system view.
   system-view
2. Enable an OSPFv3 process and enter its view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
   By default, no OSPFv3 processes are enabled.
3. Specify a router ID.
   router-id *router-id*
   By default, no router ID is configured.
4. (Optional.) Configure a description for the OSPFv3 process.
   description *text*
   By default, no description is configured for an OSPFv3 process.
   To identify and manage different OSPFv3 processes, configure a unique description for each process.
5. Return to system view.
   quit
6. Enter interface view.
   interface *interface-type interface-number*
7. Enable an OSPFv3 process on the interface.
   ospfv3 *process-id* **area** *area-id* [**instance** *instance-id*]
   By default, no OSPFv3 processes are enabled on an interface.

Establishing OSPFv3 neighbor relationships on an M-LAG network

About this task

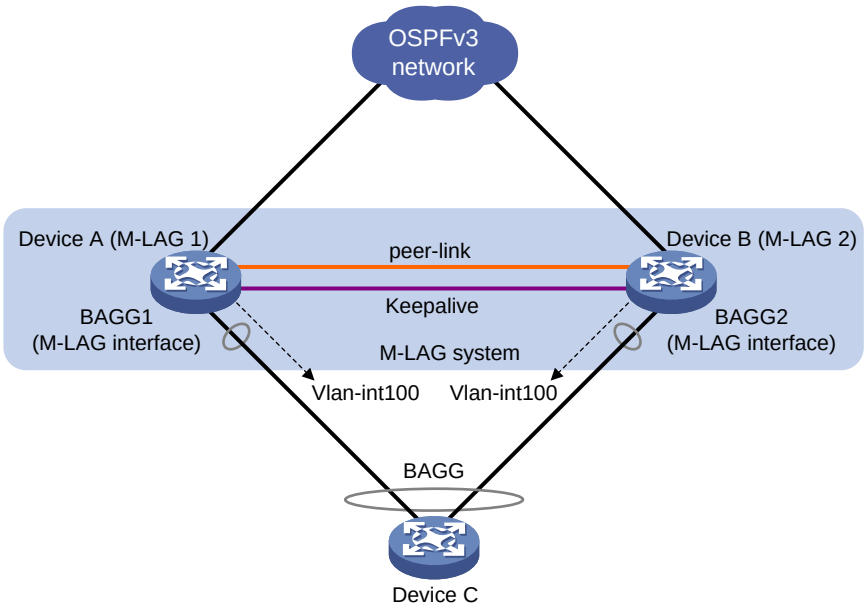
When a customer-side device accesses the OSPFv3 network through both M-LAG member devices in an M-LAG system, the M-LAG member devices provide Layer 3 gateway services. The gateway interfaces (such as VLAN interfaces) on different M-LAG member devices might generate the same link-local address because they have the same IP address and MAC address. In this case, the

M-LAG member devices cannot establish OSPFv3 neighbor relationships with the customer-side device.

To resolve this issue, configure the gateway interfaces to use M-LAG virtual IPv6 addresses for OSPFv3 neighbor relationship establishment. The M-LAG virtual IPv6 address configured for a gateway interface must be a link-local address because OSPFv3 uses link-local addresses for OSPFv3 neighbor relationship establishment. For more information about M-LAG, see M-LAG configuration in *Layer 2—LAN Switching Configuration Guide*.

As shown in Figure 1, Device C accesses the OSPFv3 network through the M-LAG system that consists of Device A and Device B. Device A and Device B provide Layer 3 gateway services. The gateway interface VLAN-interface 100 on Device A and the gateway interface VLAN-interface 100 on Device B have the same IP address and the same MAC address. To ensure that both Device A and Device B can establish an OSPFv3 neighbor relationship with Device C, configure different M-LAG virtual IPv6 addresses for the VLAN interfaces on Device A and Device B.

Figure 1 OSPFv3 neighbor relationship establishment on an M-LAG network



Procedure

1. Enter system view.
   `system-view`
2. Enter interface view.
   `interface` *`interface-type interface-number`*
3. Specify a link-local address as the M-LAG virtual IPv6 address to be assigned to the interface.
   `port m-lag ipv6 virtual-ip` *`ipv6-address`* **`link-local`**
   By default, no M-LAG virtual IPv6 address is assigned to an interface.
   For more information about this command, see M-LAG commands in *Layer 2—LAN Switching Ethernet Command Reference*.

Configuring OSPFv3 area parameters

About OSPFv3 areas

OSPFv3 has the same stub area, NSSA area, and virtual link features as OSPFv2.

After you split an OSPFv3 AS into multiple areas, the LSA number is reduced and OSPFv3 applications are extended. To further reduce the size of routing tables and the number of LSAs, configure the non-backbone areas at an AS edge as stub areas.

A stub area cannot import external routes, but an NSSA area can import external routes into the OSPFv3 routing domain while retaining other stub area characteristics.

Non-backbone areas exchange routing information through the backbone area, so the backbone and non-backbone areas (including the backbone itself) must be fully meshed. If no connectivity can be achieved, configure virtual links.

Configuring a stub area

Restrictions and guidelines

To configure a stub area, you must perform this task on all routers attached to the area.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter OSPFv3 area view.
   area *area-id*
4. Configure the area as a stub area.
   stub [**default-route-advertise-always** | **no-summary**] *
   By default, no area is configured as a stub area.
   The **no-summary** keyword is only available on the ABR of a stub area. If you specify the **no-summary** keyword, the ABR only advertises a default route in an Inter-Area-Prefix LSA into the stub area.
5. (Optional.) Set a cost for the default route advertised to the stub area.
   default-cost *cost-value*
   By default, the cost for the default route advertised to the stub area is 1.

Configuring an NSSA area

Restrictions and guidelines

To configure an NSSA area, you must perform this task on all routers attached to the area.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter OSPFv3 area view.
   area *area-id*
4. Configure the area as an NSSA area.
   nssa [**default-route-advertise** [**cost** *cost-value* | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] * |

no-import-route | **no-summary** | [**translate-always** | **translate-never**] | **suppress-fa** | **translator-stability-interval** *value*] *

By default, no area is configured as an NSSA area.

To configure a totally NSSA area, execute the **nssa no-summary** command on the ABR. The ABR of a totally NSSA area does not advertise inter-area routes into the area.

5. (Optional.) Set a cost for the default route advertised to the NSSA area.

   default-cost *cost-value*

   By default, the cost for the default route advertised to the NSSA area is 1.

   This command takes effect only on the ABR/ASBR of an NSSA or totally NSSA area.

Configuring an OSPFv3 virtual link

About this task

You can configure a virtual link to maintain connectivity between a non-backbone area and the backbone, or in the backbone itself.

Restrictions and guidelines

Both ends of a virtual link are ABRs that must be configured with the **vlink-peer** command.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter OSPFv3 area view.

   area *area-id*

4. Configure a virtual link.

   vlink-peer *router-id* [**dead** *seconds* | **hello** *seconds* | **instance** *instance-id* | **ipsec-profile** *profile-name* | **keychain** *keychain-name* | **retransmit** *seconds* | **trans-delay** *seconds*] *

Configuring OSPFv3 network types

Restrictions and guidelines for OSPFv3 network type configuration

Based on the link layer protocol, OSPFv3 classifies networks into different types, including broadcast, NBMA, P2MP, and P2P.

- If any routers in a broadcast network do not support multicasting, you can change the network type to NBMA.
- If only two routers running OSPFv3 exist on a network segment, you can change the network type to P2P to save costs.

Setting the broadcast network type for an OSPFv3 interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the network type to broadcast for the OSPFv3 interface.

   ospfv3 network-type broadcast [**instance** *instance-id*]

   By default, the network type of an OSPFv3 interface depends on the encapsulated link layer protocol.

Setting the NBMA network type for an OSPFv3 interface

Restrictions and guidelines

For NBMA interfaces, you must specify the link-local IP addresses and DR priorities for their neighbors because these interfaces cannot find neighbors by broadcasting hello packets.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the network type to NBMA for the OSPFv3 interface.

   ospfv3 network-type nbma [**instance** *instance-id*]

   By default, the network type of an OSPFv3 interface depends on the encapsulated link layer protocol.

4. (Optional.) Set the router priority for the interface

   ospfv3 dr-priority *priority*

   By default, an interface has a router priority of 1.

   An interface's router priority determines its privilege in DR/BDR selection.

5. Specify an NBMA neighbor.

   ospfv3 peer *ipv6-address* [**cost** *cost-value* | **dr-priority** *priority*] [**instance** *instance-id*]

   By default, no link-local address is specified for the neighbor interface.

Setting the P2MP network type for an OSPFv3 interface

Restrictions and guidelines

For P2MP interfaces (only when in unicast mode), you must specify the link-local IP addresses of their neighbors because these interfaces cannot find neighbors by broadcasting hello packets.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the network type to P2MP for the OSPFv3 interface.

   ospfv3 network-type p2mp [**unicast**] [**instance** *instance-id*]

   By default, the network type of an OSPFv3 interface depends on the encapsulated link layer protocol.

4. Specify a P2MP unicast neighbor.

ospfv3 peer *ipv6-address* [**cost** *cost-value* | **dr-priority** *priority*] [**instance** *instance-id*]

By default, no link-local address is specified for the neighbor interface.

Setting the P2P network type for an OSPFv3 interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the network type to P2P for the OSPFv3 interface.
   ospfv3 network-type p2p [**instance** *instance-id*]
   By default, the network type of an OSPFv3 interface depends on the encapsulated link layer protocol.

Configuring OSPFv3 route control

Configuring OSPFv3 inter-area route summarization

About this task

If contiguous network segments exist in an area, you can summarize them into one network segment on the ABR. The ABR will advertise only the summary route. Any LSA on the specified network segment will not be advertised, reducing the LSDB size in other areas.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter OSPFv3 area view.
   area *area-id*
4. Configure route summarization on the ABR.
   abr-summary *ipv6-address prefix-length* [**not-advertise**] [**cost** *cost-value*]
   By default, route summarization is not configured on an ABR.

Configuring redistributed route summarization

About this task

Perform this task to enable an ASBR to summarize external routes within the specified address range into a single route.

An ASBR can summarize routes in the following LSAs:

- Type-5 LSAs.
- Type-7 LSAs in an NSSA area.

- Type-5 LSAs translated from Type-7 LSAs in an NSSA area if the ASBR (also an ABR) is a translator. If the ASBR is not a translator, it cannot summarize routes in Type-5 LSAs translated from Type-7 LSAs.

Procedure

1. Enter system view.

   system-view
2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Configure route summarization on an ASBR.

   asbr-summary *ipv6-address prefix-length* [**cost** *cost-value* | **not-advertise** | **nssa-only** | **tag** *tag*] *

   By default, route summarization is not configured on an ASBR.

Configuring OSPFv3 received route filtering

About this task

This task allows you to filter routes calculated by using received LSAs.

Procedure

1. Enter system view.

   system-view
2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Configure OSPFv3 to filter routes calculated by using received LSAs.

   filter-policy { *ipv6-acl-number* [**gateway** *prefix-list-name*] | **prefix-list** *prefix-list-name* [**gateway** *prefix-list-name*] | **gateway** *prefix-list-name* | **route-policy** *route-policy-name* } **import**

   By default, OSPFv3 accepts all routes calculated by using received LSAs.

   This command can only filter routes computed by OSPFv3. Only routes not filtered out can be added into the local routing table.

Configuring Inter-Area-Prefix LSA filtering

Restrictions and guidelines

The **filter** command takes effect only on ABRs.

Procedure

1. Enter system view.

   system-view
2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter OSPFv3 area view.

   area *area-id*
4. Configure OSPFv3 to filter Inter-Area-Prefix LSAs.

   filter { *ipv6-acl-number* | **prefix-list** *prefix-list-name* | **route-policy** *route-policy-name* } { **export** | **import** }

By default, OSPFv3 accepts all Inter-Area-Prefix LSAs.

Advertising the maximum link cost to neighbors within the specified period

About this task

On an OSPFv3 network, when a link recovers from failures or the state of an interface changes, OSPFv3 will re-establish neighbor relationships and perform route convergence. During the route convergence process, routing loops and traffic loss might occur because the convergence speeds of the nodes are different. To address this issue, enable OSPFv3 to advertise the maximum link cost (65535) to neighbors within the specified period, so the traffic forwarding path remains unchanged. After the specified period, OSPFv3 advertises the original link cost to neighbors and performs optimal route selection again.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set the duration of the maximum link cost advertisement.

   ospfv3 peer hold-max-cost duration *timer* [**instance** *instance-id*]

   By default, OSPFv3 advertises the original link cost to neighbors during a route convergence.

Setting an OSPFv3 cost for an interface

About this task

You can set an OSPFv3 cost for an interface with one of the following methods:

- Set the cost value in interface view.
- Set a bandwidth reference value for the interface, and OSPFv3 computes the cost automatically based on the bandwidth reference value by using the following formula:

  Interface OSPFv3 cost = Bandwidth reference value (100 Mbps) / Interface bandwidth (Mbps)
  - If the calculated cost is greater than 65535, the value of 65535 is used.
  - If the calculated cost is smaller than 1, the value of 1 is used.
- If no cost is set for an interface, OSPFv3 automatically computes the cost for the interface.

Setting a cost in interface view

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set an OSPFv3 cost for the interface.

   ospfv3 cost *cost-value* [**instance** *instance-id*]

   By default, the OSPFv3 cost is 1 for a VLAN interface, is 0 for a loopback interface. The OSPFv3 cost is automatically computed according to the interface bandwidth for other interfaces.

Setting a bandwidth reference value

1. Enter system view.

system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Set a bandwidth reference value.

   bandwidth-reference *value*

   The default bandwidth reference value is 100 Mbps.

Associating OSPFv3 with Track

About this task

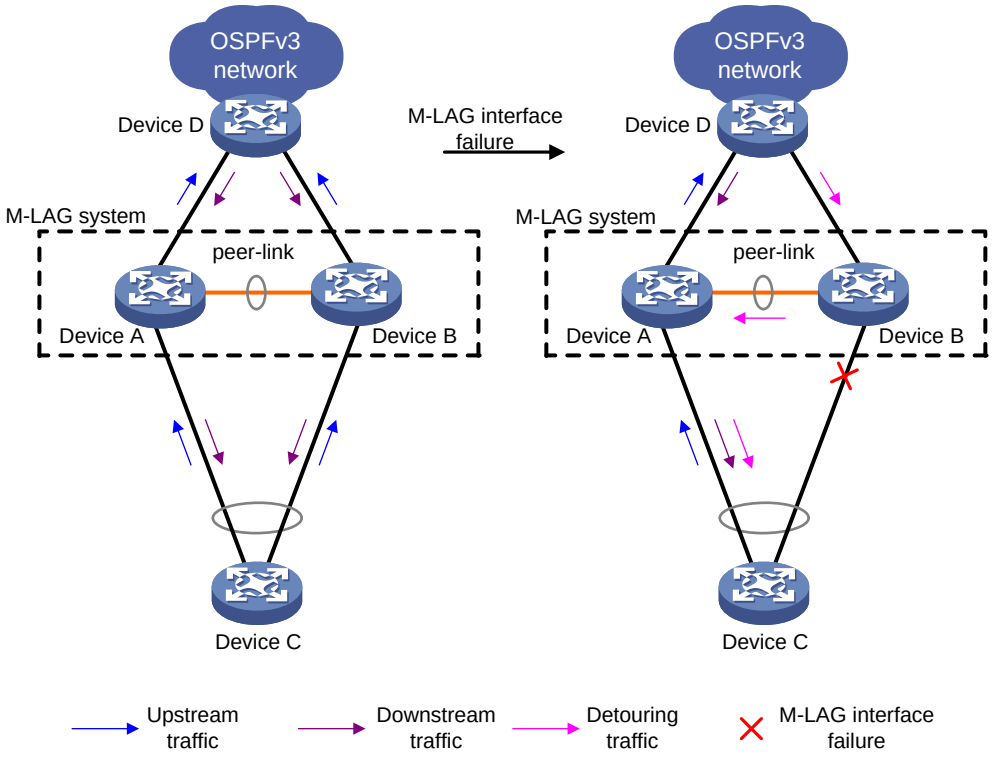
Perform this task to associate an OSPFv3 interface with a track entry to adjust the cost of the interface as follows:

- When the track entry state becomes negative, OSPFv3 increases the cost of the interface.
- When the track entry state becomes positive, OSPFv3 restores the cost of the interface to the original value.

When a customer side device is dual-homed to an OSPFv3 network through M-LAG, you can configure this feature to prevent the traffic from detouring through the M-LAG peer link. As shown in Figure 2, customer-side device Device C is dual-homed to the OSPFv3 network through an M-LAG system formed by Device A and Device B. Device A and Device B share the traffic load between Device D and Device C. When the M-LAG interface that connects Device B to Device C fails, the upstream traffic and downstream traffic are forwarded as follows:

- Device C detects the M-LAG interface failure, and traffic from Device C to Device D is switched from path Device C->Device B->Device D to path Device C->Device A->Device D.
- Device D cannot detect the M-LAG interface failure and still forwards traffic through Device B. Because the link between Device B and Device C fails, traffic from Device D to Device C is switched from path Device D->Device B->Device C to path Device D->Device B->Device A->Device C.

Figure 2 Traffic forwarding path after M-LAG interface failure



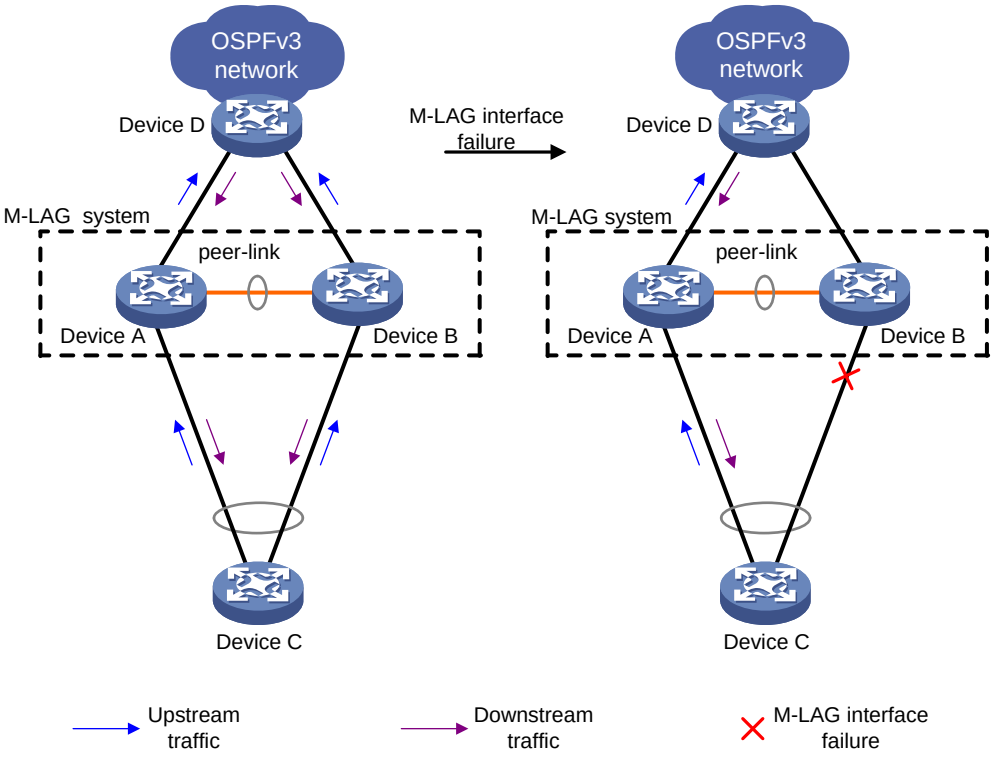
To prevent the downstream traffic from detouring through the M-LAG peer link, perform the following tasks on Device B:

1. Execute the **track interface** command to create a track entry and associate it with the link state of the M-LAG interface. For more information about the **track interface** command, see track commands in *High Availability Command Reference*.
2. Execute the **ospfv3 track adjust-cost** command to associate the gateway interface with the track entry.

Then, Device B adjusts the cost of the gateway interface based on the state of the associated track entry:

- When the M-LAG interface fails, the track entry state becomes negative, and OSPFv3 increases the cost of the gateway interface. Then, traffic from Device D to Device C is switched to path Device D->Device A->Device C.
- When the M-LAG interface recovers, the track entry state becomes positive, and OSPFv3 restores the cost of the gateway interface to the original value. Then, traffic from Device D to Device C is switched back to path Device D->Device B->Device C.

Figure 3 Traffic forwarding path after associating OSPFv3 with Track



Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Associate the OSPFv3 interface with a track entry to adjust the cost of the interface based on the track entry state.

   ospfv3 track *track-entry-number* **adjust-cost** { *cost-offset* | **max** }

   By default, an OSPFv3 interface is not associated with any track entry.

Setting the maximum number of OSPFv3 ECMP routes

About this task

OSPFv3 might find multiple equal-cost routes to the same destination, which can be used to share the traffic load. This task allows you to set the maximum number of ECMP routes.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the maximum number of ECMP routes.

   maximum load-balancing *number*

By default, the maximum number of OSPFv3 ECMP routes equals the maximum number of ECMP routes supported by the system.

Setting a preference for OSPFv3

About this task

A router can run multiple routing protocols. The system assigns a priority for each protocol. When these routing protocols find the same route, the route found by the protocol with the highest priority is selected.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Set a preference for OSPFv3.

   preference [**ase**] { *preference* | **route-policy** *route-policy-name* } *

   By default, the preference of OSPFv3 internal routes is 10, and the preference of OSPFv3 external routes is 150.

Configuring OSPFv3 route redistribution

Restrictions and guidelines

Because OSPFv3 is a link state routing protocol, it cannot directly filter LSAs to be advertised. OSPFv3 filters only redistributed routes. Only routes that are not filtered out can be advertised in LSAs.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure route redistribution.

   import-route bgp4+ [*as-number*] [**allow-ibgp**] [[**cost** *cost-value* | **inherit-cost**] | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   import-route { **direct** | **static** } [[**cost** *cost-value* | **inherit-cost**] | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   import-route { **isisv6** | **ospfv3** | **ripng** } [*process-id* | **all-processes**] [**allow-direct** | [**cost** *cost-value* | **inherit-cost**] | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   By default, OSPFv3 does not redistribute routes.

   The **import-route bgp4+** command redistributes only EBGP routes. The **import-route bgp4+ allow-ibgp** command redistributes both EBGP and IBGP routes, which might cause routing loops. Therefore, use the **import-route bgp4+ allow-ibgp** command with caution.

4. (Optional.) Configure OSPFv3 to filter redistributed routes.

filter-policy { *ipv6-acl-number* | **prefix-list** *prefix-list-name* } **export** [**bgp4+** | **direct** | { **isisv6** | **ospfv3** | **ripng** } [*process-id*] | **static**]

By default, OSPFv3 accepts all redistributed routes.

This command filters only routes redistributed by the **import-route** command. If no routes are redistributed by the **import-route** command, this command does not take effect.

5. Set a tag for redistributed routes.

   default tag *tag*

   By default, the tag of redistributed routes is 1.

Configuring default route redistribution

About this task

The **import-route** command cannot redistribute a default external route. To redistribute a default route, perform this task.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Redistribute a default route.

   default-route-advertise [[**always** | **permit-calculate-other**] | **cost** *cost-value* | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   By default, no default route is redistributed.

4. Set a tag for redistributed routes.

   default tag *tag*

   By default, the tag of redistributed routes is 1.

Setting OSPFv3 timers

Setting OSPFv3 packet timers

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the hello interval.

   ospfv3 timer hello *seconds* [**instance** *instance-id*]

   The default hello interval on P2P and broadcast interfaces is 10 seconds. The default hello interval on P2MP and NBMA interfaces is 30 seconds.

4. Set the dead interval.

   ospfv3 timer dead *seconds* [**instance** *instance-id*]

   The default dead interval on P2P and broadcast interfaces is 40 seconds. The default dead interval on P2MP and NBMA interfaces is 120 seconds.

The dead interval set on neighboring interfaces cannot be too short. If the interval is too short, a neighbor is easily down.

5. Set the poll interval.

   ospfv3 timer poll *seconds* [**instance** *instance-id*]

   By default, the poll interval is 120 seconds.

6. Set the LSA retransmission interval.

   ospfv3 timer retransmit *interval* [**instance** *instance-id*]

   The default LSA retransmission interval is 5 seconds.

   The LSA retransmission interval cannot be too short. If the interval is too short, unnecessary retransmissions will occur.

Setting the LSA transmission delay

About this task

Each LSA in the LSDB has an age that increases by 1 every second, but the age does not change during transmission. Therefore, it is necessary to add a transmission delay into the age time, especially for low-speed links.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the LSA transmission delay.

   ospfv3 trans-delay *seconds* [**instance** *instance-id*]

   By default, the LSA transmission delay is 1 second.

Setting the OSPFv3 route calculation interval

About this task

LSDB changes result in SPF calculations. When the topology changes frequently, a large amount of network and router resources are occupied by SPF calculation. You can adjust the route calculation interval to reduce the impact.

The default route calculation interval settings are as follows:

- The interval for the first route calculation is 50 milliseconds.
- The interval for the nth (n ≥ 2) route calculation is 50 + (200 x 2^{n-2}) milliseconds, and cannot exceed 5 seconds.

The **spf-schedule-interval** command functions as follows:

- If the **spf-schedule-interval** *maximum-interval* command is executed:
  - The values for the *minimum-interval* and *incremental-interval* arguments are both 0.
  - The route calculation interval is fixed at *maximum-interval*.
- If the **spf-schedule-interval** *maximum-interval minimum-interval* command is executed:
  - The value for the *incremental-interval* argument is 0.
  - The interval for the first route calculation is *minimum-interval*.

- The interval for subsequent route calculations is *maximum-interval*.

- If the **spf-schedule-interval** *maximum-interval minimum-interval incremental-interval* command is executed:
  - The interval for the first route calculation is *minimum-interval*.
  - The interval for the nth (n ≥ 2) route calculation is *minimum-interval* + *incremental-interval* x 2^{n-2}, and cannot exceed *maximum-interval*.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Set the route calculation interval for OSPFv3.
   spf-schedule-interval *maximum-interval* [*minimum-interval* [*incremental-interval*]]
   By default, the maximum interval is 5 seconds, the minimum interval is 50 milliseconds, and the incremental interval is 200 milliseconds.

Setting the LSA generation interval

About this task

If network changes become frequent, bandwidth resources and router resources might be overconsumed. To avoid this issue, use the **lsa-generation-interval** command to adjust the LSA generation interval settings for OSPFv3.

By default, the LSA generation interval is fixed at 5 seconds for OSPFv3.

The **lsa-generation-interval** command functions as follows:

- If the **lsa-generation**-*interval maximum-interval* command is executed:
  - The values for the *minimum-interval* and *incremental-interval* arguments are both 0.
  - The LSA generation interval is fixed at *maximum-interval*.
- If the **lsa-generation-interval** *maximum-interval minimum-interval* command is executed:
  - The value for the *incremental-interval* argument is 0.
  - The interval for the first LSA generation is *minimum-interval*.
  - The interval for subsequent LSA generations is *maximum-interval*.
- If the **lsa-generation-interval** *maximum-interval minimum-interval incremental-interval* command is executed:
  - The interval for the first LSA generation is *minimum-interval*.
  - The interval for the nth (n ≥ 2) LSA generation is *minimum-interval* + *incremental-interval* x 2^{n-2}, and cannot exceed *maximum-interval*.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the LSA generation interval.

   lsa-generation-interval *maximum-interval* [*minimum-interval* [*incremental-interval*]]

   By default, the maximum interval is 5 seconds, the minimum interval is 0 milliseconds, and the incremental interval is 0 milliseconds.

Setting the LSU transmit rate

About this task

Sending large numbers of LSU packets affects router performance and consumes a large amount of network bandwidth. You can configure the router to send LSU packets at an interval and to limit the maximum number of LSU packets sent out of an OSPFv3 interface at each interval.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the LSU transmit rate.

   transmit-pacing interval *interval* **count** *count*

   By default, an OSPFv3 interface sends a maximum of three LSU packets every 20 milliseconds.

Setting a DR priority for an interface

About this task

The router priority is used for DR election. Interfaces having the priority 0 cannot become a DR or BDR.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set a router priority.

   ospfv3 dr-priority *priority* [**instance** *instance-id*]

   The default router priority is 1.

Configuring OSPFv3 packet parameters

Ignoring MTU check for DD packets

About this task

When LSAs are few in DD packets, it is unnecessary to check the MTU in DD packets to improve efficiency.

Restrictions and guidelines

A neighbor relationship can be established only if the interface's MTU is the same as that of the peer.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Ignore MTU check for DD packets.
   ospfv3 mtu-ignore [**instance** *instance-id*]
   By default, OSPFv3 does not ignore MTU check for DD packets.

Disabling interfaces from receiving and sending OSPFv3 packets

About this task

After an OSPFv3 interface is set to **silent**, direct routes of the interface can still be advertised in Intra-Area-Prefix LSAs through other interfaces, but other OSPFv3 packets cannot be advertised. No neighboring relationship can be established on the interface. This feature can enhance the adaptability of OSPFv3 networking.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Disable interfaces from receiving and sending OSPFv3 packets.
   silent-interface { *interface-type interface-number* | **all** }
   By default, the interfaces can receive and send OSPFv3 packets.
   This command disables only the interfaces that run the current process. However, multiple OSPFv3 processes can disable the same interface from receiving and sending OSPFv3 packets.

Configuring prefix suppression

About prefix suppression

By default, an OSPFv3 interface advertises all of its prefixes in LSAs. To speed up OSPFv3 convergence, you can suppress interfaces from advertising all of their prefixes. This feature helps improve network security by preventing IP routing to the suppressed networks.

When prefix suppression is enabled:

- OSPFv3 does not advertise the prefixes of suppressed interfaces in Type-8 LSAs.
- On broadcast and NBMA networks, the DR does not advertise the prefixes of suppressed interfaces in Type-9 LSAs that reference Type-2 LSAs.
- On P2P and P2MP networks, OSPFv3 does not advertise the prefixes of suppressed interfaces in Type-9 LSAs that reference Type-1 LSAs.

Restrictions and guidelines for prefix suppression

As a best practice, configure prefix suppression on all OSPFv3 routers if you want to use prefix suppression.

Configuring prefix suppression for an OSPFv3 process

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable prefix suppression for the OSPFv3 process.
   prefix-suppression
   By default, prefix suppression is disabled for an OSPFv3 process.
   Enabling prefix suppression for an OSPFv3 process does not suppress the prefixes of loopback interfaces and passive interfaces.

Configuring prefix suppression for an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable prefix suppression for the interface.
   ospfv3 prefix-suppression [**disable**] [**instance** *instance-id*]
   By default, prefix suppression is disabled for an interface.

Configuring OSPFv3 advanced features

Configuring a stub router

About this task

A stub router is used for traffic control. It reports its status as a stub router to neighboring OSPFv3 routers. The neighboring routers can have a route to the stub router, but they do not use the stub router to forward data.

Use either of the following methods to configure a router as a stub router:

- Clear the R-bit of the Option field in Type-1 LSAs. When the R-bit is clear, the OSPFv3 router can participate in OSPFv3 topology distribution without forwarding traffic.
- Use the OSPFv3 max-metric router LSA feature. This feature enables OSPFv3 to advertise its locally generated Type-1 LSAs with a maximum cost of 65535. Neighbors do not send packets to the stub router as long as they have a route with a smaller cost.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.

ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Configure the router as a stub router.
   - Configure the router as a stub router and clear the R-bit of the Option field in Type-1 LSAs.

     stub-router r-bit [**include-stub** | **on-startup** { *seconds* | **wait-for-bgp** [*seconds*] }] *
   - Configure the router as a stub router and advertise the locally generated Type-1 LSAs with the maximum cost of 65535.

     stub-router max-metric [**external-lsa** [*max-metric-value*] | **summary-lsa** [*max-metric-value*] | **include-stub** | **on-startup** { *seconds* | **wait-for-bgp** [*seconds*] }] *

   By default, the router is not configured as a stub router.

   A stub router is not related to a stub area.

Configuring OSPFv3 isolation

About this task

Isolation is a method used for network device maintenance. It gracefully removes a device from the packet forwarding path for maintenance and gracefully adds the device to the network after maintenance.

To reduce impact on traffic forwarding, you can isolate a device before upgrading it. OSPFv3 isolation works as follows:

1. After OSPFv3 isolation is enabled for a device, OSPFv3 increases the link cost in LSAs advertised by the device based on the following rules:
   - The link cost in Type-1 LSAs (Router LSAs) is set to 65535.
   - The link cost in Type-9 LSAs (Intra-area-prefix LSAs) that reference Type-1 LSAs is set to 65535.
   - The link cost in the following LSAs is set to 16711680:
     - Type-3 LSAs (Inter-area-prefix LSAs).
     - Type-4 LSAs (Inter-area-router LSAs).
     - Type-5 LSAs (AS external LSAs).
     - Type-7 LSAs (NSSA LSAs).
2. Each neighbor of the device reselects an optimal route based on the LSAs and stops forwarding traffic to the device. The device is fully isolated from the network and you can upgrade the device.
3. After the maintenance, disable OSPFv3 isolation on the device to restore its link cost and gracefully add it back to the network.

Restrictions and guidelines

Both the **isolate enable** and **stub-router max-metric external-lsa 16711680 summary-lsa 16711680 include-stub** commands can isolate the device from the network.

When you execute both the **isolate enable** and **stub-router max-metric** commands, follow these restrictions and guidelines:

- If you specify the **external-lsa** and **summary-lsa** keywords in the **stub-router max-metric** command, the higher one of the two link costs provided by the isolation feature and the stub router feature takes effect.
- If the **on-startup** keyword is specified in the **stub-router max-metric** command, traffic forwarding path selection is affected only by the isolation feature when the stub router feature does not take effect.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Configure OSPFv3 isolation.
   - Enable OSPFv3 isolation to gracefully isolate the device from the network.
     isolate enable
   - Disable OSPFv3 isolation to gracefully add the device back to the network.
     undo isolate enable

   By default, OSPFv3 isolation is disabled.

Configuring OSPFv3 shutdown

About this task

For maintenance purposes, you can use this feature to shut down processes on the device with small impact on the network. If you shut down an OSPFv3 process, it will perform the following operations:

- Send 1-way hello packets to its neighbors.

  On receipt of the packets, the neighbors disconnect from the process and use the backup path to forward traffic.
- Stop receiving and sending OSPFv3 packets.
- Clear its neighbor, LSDB, and OSPFv3 route information.

After maintenance, you can use the **undo shutdown process** command to restart the process for neighbor relationship re-establishment.

This feature shuts down an OSPFv3 process while retaining the process-associated settings to facilitate your maintenance.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Configure OSPFv3 shutdown for OSPFv3 maintenance.
   a. Shut down an OSPFv3 process to isolate it from the network.
      shutdown process
   b. Restart the OSPFv3 process to add it back to the network.
      undo shutdown process

   By default, the OSPFv3 process is not shut down.

Configuring OSPFv3 GR

About OSPFv3 GR

GR ensures forwarding continuity when a routing protocol restarts or an active/standby switchover occurs.

Two routers are required to complete a GR process. The following are router roles in a GR process:

- **GR restarter**—Graceful restarting router. It must be Graceful Restart capable.
- **GR helper**—The neighbor of the GR restarter. It helps the GR restarter to complete the GR process.

To prevent service interruption after a master/backup switchover, a GR restarter running OSPFv3 must perform the following tasks:

- Keep the GR restarter forwarding entries stable during reboot.
- Establish all adjacencies and obtain complete topology information after reboot.

After the active/standby switchover, the GR restarter sends a Grace LSA to tell its neighbors that it performs a GR. Upon receiving the Grace LSA, the neighbors with the GR helper capability enter the helper mode (and are called GR helpers). Then, the GR restarter retrieves its adjacencies and LSDB with the help of the GR helpers.

Restrictions and guidelines for OSPFv3 GR

You cannot enable OSPFv3 NSR on a device that acts as GR restarter.

Configuring GR restarter

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable the GR capability.
   graceful-restart enable [**global** | **planned-only**] *
   By default, OSPFv3 GR restarter capability is disabled.
4. (Optional.) Set the GR interval.
   graceful-restart interval *interval*
   By default, the GR interval is 120 seconds.

Configuring GR helper

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable the GR helper capability.
   graceful-restart helper enable [**planned-only**]
   By default, the GR helper capability is enabled.

4. Enable strict LSA checking.

   graceful-restart helper strict-lsa-checking

   By default, strict LSA checking is disabled.

Triggering OSPFv3 GR

About this task

OSPFv3 GR is triggered by an active/standby switchover or when this task is performed.

Procedure

To trigger OSPFv3 GR, execute the **reset ospfv3** [*process-id*] **process graceful-restart** command in user view.

Configuring OSPFv3 NSR

About this task

Nonstop routing (NSR) backs up OSPFv3 link state information from the active process to the standby process. After an active/standby switchover, NSR can complete link state recovery and route regeneration without tearing down adjacencies or impacting forwarding services.

Choose OSPFv3 NSR or OSPFv3 GR according to the actual situation and make sure data forwarding is not interrupted. The differences between OSPFv3 NSR and OSPFv3 GR as follows:

- **Device requirements**—The active and standby processes of OSPFv3 must run on different IRF member devices. To run OSPFv3 NSR, you must set up an IRF fabric that contains a minimum of two member devices. To run OSPFv3 GR, IRF fabric setup is not required. If an IRF fabric already exists, you can run OSPFv3 GR when the IRF fabric contains only one member device.
- **OSPFv3 neighbor requirements**—When you use OSPFv3 NSR, OSPFv3 neighbors cannot detect process restart or active/standby switchover on the local device. OSPFv3 NSR can recover OSPFv3 route information without requiring the cooperation of OSPFv3 neighbors. OSPFv3 GR requires OSPFv3 neighbors to have the GR helper capability to help the GR restarter complete the GR process.

Restrictions and guidelines

When both GR and NSR are configured for OSPFv3 on a device, the following rules apply:

- OSPFv3 NSR has a higher priority than OSPFv3 GR. When the active process goes down, the device does not act as the GR restarter and performs NSR to ensure nonstop forwarding.
- When the following conditions exist, the GR process might fail even if NSR is enabled on the GR helper:
  - OSPFv3 is restarted through GR.
  - An active/standby switchover is triggered on the GR helper before the GR process finishes.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Enable OSPFv3 NSR.

   non-stop-routing

   By default, OSPFv3 NSR is enabled.

This command takes effect only for the current process. As a best practice, enable OSPFv3 NSR for each process if multiple OSPFv3 processes exist.

Configuring BFD for OSPFv3

About this task

Bidirectional forwarding detection (BFD) provides a mechanism to quickly detect the connectivity of links between OSPFv3 neighbors, improving the convergence speed of OSPFv3. For more information about BFD, see *High Availability Configuration Guide*.

After discovering neighbors by sending hello packets, OSPFv3 notifies BFD of the neighbor addresses, and BFD uses these addresses to establish sessions. Before a BFD session is established, it is in the down state. In this state, BFD control packets are sent at an interval of no less than 1 second to reduce BFD control packet traffic. After the BFD session is established, BFD control packets are sent at the negotiated interval, thereby implementing fast fault detection.

To configure BFD for OSPFv3, you need to configure OSPFv3 first.

Restrictions and guidelines

In the current software version, this feature is not supported on VSI interfaces.

Procedure

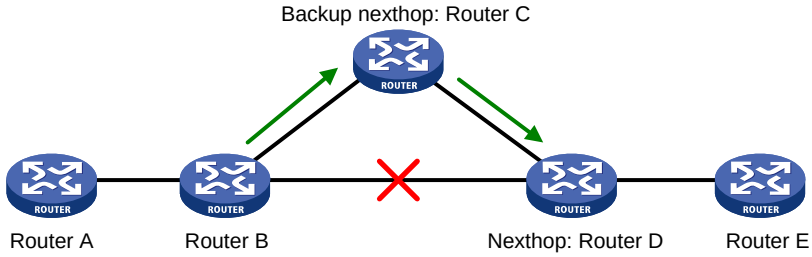
1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Specify a router ID.
   router-id *router-id*
4. Quit the OSPFv3 view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Enable an OSPFv3 process on the interface.
   ospfv3 *process-id* **area** *area-id* [**instance** *instance-id*]
7. Enable BFD on the interface.
   ospfv3 bfd enable [**instance** *instance-id*]
   By default, BFD is disabled on the OSPFv3 interface.

Configuring OSPFv3 FRR

About OSPFv3 FRR

A primary link failure can cause packet loss and even a routing loop until OSPFv3 completes routing convergence based on the new network topology. OSPFv3 FRR enables fast rerouting to minimize the failover time.

Figure 4 Network diagram for OSPFv3 FRR



As shown in Figure 4, configure FRR on Router B. OSPFv3 FRR automatically calculates a backup next hop or specifies a backup next hop by using a routing policy. When the primary link fails, OSPFv3 directs packets to the backup next hop. At the same time, OSPFv3 calculates the shortest path based on the new network topology. It forwards packets over the path after network convergence.

You can configure OSPFv3 FRR to calculate a backup next hop by using the loop free alternate (LFA) algorithm, or specify a backup next hop by using a routing policy.

Configuring OSPFv3 FRR to use the LFA algorithm to calculate a backup next hop

Restrictions and guidelines

Do not use the **fast-reroute lfa** command together with the **vlink-peer** command.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. (Optional.) Disable LFA on an interface.
   ospfv3 fast-reroute lfa-backup exclude
   By default, the interface on which LFA is enabled can be selected as a backup interface.
4. Return to system view.
   quit
5. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
6. Enable OSPFv3 FRR to use the LFA algorithm to calculate a backup next hop.
   fast-reroute lfa [**abr-only**]
   By default, OSPFv3 FRR is disabled.
   If **abr-only** is specified, the route to the ABR is selected as the backup path.

Configuring OSPFv3 FRR to use a backup next hop in a routing policy

About this task

Before you perform this task, use the **apply ipv6 fast-reroute backup-interface** command to specify a backup next hop in the routing policy to be used. For more information about

the **apply ipv6 fast-reroute backup-interface** command and routing policy configuration, see "Configuring routing policies."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. (Optional.) Disable LFA on an interface.
   ospfv3 fast-reroute lfa-backup exclude
   By default, the interface is enabled with LFA and it can be selected as a backup interface.
4. Return to system view.
   quit
5. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
6. Configure OSPFv3 FRR to use a backup next hop in a routing policy.
   fast-reroute route-policy *route-policy-name*
   By default, OSPFv3 FRR is disabled.

Configuring BFD for OSPFv3 FRR

About this task

By default, OSPFv3 FRR does not use BFD to detect primary link failures. To speed up OSPFv3 convergence, enable BFD control packet mode for OSPFv3 FRR to detect primary link failures. This mode requires BFD configuration on both OSPFv3 routers on the link.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable BFD control packet mode for OSPFv3 FRR.
   ospfv3 primary-path-detect bfd ctrl [**instance** *instance-id*]
   By default, BFD control packet mode is disabled for OSPFv3 FRR.

Configuring BFD echo packet mode for OSPFv3 FRR

About this task

By default, OSPFv3 FRR does not use BFD to detect primary link failures. To speed up OSPFv3 convergence, enable BFD echo packet mode for OSPFv3 FRR to detect primary link failures. This mode requires BFD configuration on one OSPFv3 router on the link.

Procedure

1. Enter system view.
   system-view
2. Configure the source IPv6 address of BFD echo packets.
   bfd echo-source-ipv6 *ipv6-address*

By default, the source IPv6 address of BFD echo packets is not configured.

The source IPv6 address cannot be on the same network segment as any local interface's IP address.

For more information about this command, see *High Availability Command Reference*.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable BFD echo packet mode for OSPFv3 FRR.

   ospfv3 primary-path-detect bfd echo [**instance** *instance-id*]

   By default, BFD echo packet mode is disabled for OSPFv3 FRR.

Enhancing OSPFv3 security

Configuring OSPFv3 authentication

About this task

OSPFv3 uses keychain authentication to prevent routing information from being leaked and routers from being attacked.

OSPFv3 adds the Authentication Trailer option into outgoing packets, and uses the authentication information in the option to authenticate incoming packets. Only packets that pass the authentication can be received. If a packet fails the authentication, the OSPFv3 neighbor relationship cannot be established.

Restrictions and guidelines

The authentication mode specified for an OSPFv3 interface has a higher priority than the mode specified for an OSPFv3 area.

OSPFv3 supports only the HMAC-SHA-256 and HMAC-SM3 authentication algorithms.

The ID of keys used for authentication can only be in the range of 0 to 65535.

Configuring OSPFv3 area authentication

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter OSPFv3 area view.

   area *area-id*

4. Specify an authentication mode for the area.

   For HMAC-SHA-256/HMAC-SM3 authentication:

   authentication-mode { **hmac-sha-256** | **hmac-sm3** } *key-id* { **cipher** | **plain** } *string*

   For keychain authentication:

   authentication-mode keychain *keychain-name*

   By default, no authentication is performed for the area.

   For more information about keychains, see *Security Configuration Guide*.

Configuring OSPFv3 interface authentication

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify an authentication mode for the interface.

   For HMAC-SHA-256/HMAC-SM3 authentication:

   ospfv3 authentication-mode { **hmac-sha-256** | **hmac-sm3** } *key-id* { **cipher** | **plain** } *string* [**instance** *instance-id*]

   For keychain authentication:

   ospfv3 authentication-mode keychain *keychain-name* [**instance** *instance-id*]

   By default, no authentication is performed for the interface.

   For more information about keychains, see *Security Configuration Guide*.

Applying an IPsec profile for authenticating OSPFv3 packets

About this task

To protect routing information and prevent attacks, OSPFv3 can authenticate protocol packets by using an IPsec profile. For more information about IPsec profiles, see *Security Configuration Guide*.

Outbound OSPFv3 packets carry the Security Parameter Index (SPI) defined in the relevant IPsec profile. A device compares the SPI carried in a received packet with the configured IPsec profile. If they match, the device accepts the packet. Otherwise, the device discards the packet and will not establish a neighbor relationship with the sending device.

Restrictions and guidelines for applying an IPsec profile

You can configure an IPsec profile for an area, an interface, or a virtual link .

- To implement area-based IPsec protection, configure the same IPsec profile on the routers in the target area.
- To implement interface-based IPsec protection, configure the same IPsec profile on the interfaces between two neighboring routers.
- To implement virtual link-based IPsec protection, configure the same IPsec profile on the two routers connected over the virtual link.
- If an interface and its area each have an IPsec profile configured, the interface uses its own IPsec profile.
- If a virtual link and area 0 each have an IPsec profile configured, the virtual link uses its own IPsec profile.

Applying an IPsec profile to an area

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Enter OSPFv3 area view.

   area *area-id*

4. Apply an IPsec profile to the area.

   enable ipsec-profile *profile-name*

   By default, no IPsec profile is applied.

Applying an IPsec profile to an interface

1. Enter system view.

system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Apply an IPsec profile to the interface.
   ospfv3 ipsec-profile *profile-name* [**instance** *instance-id*]
   By default, no IPsec profile is applied.

Applying an IPsec profile to a virtual link

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enter OSPFv3 area view.
   area *area-id*
4. Apply an IPsec profile to a virtual link.
   vlink-peer *router-id* [**dead** *seconds* | **hello** *seconds* | **instance** *instance-id* | **ipsec-profile** *profile-name* | { { **hmac-sha-256** | **hmac-sm3** } *key-id* { **cipher** | **plain** } *string* | **keychain** *keychain-name* } | **retransmit** *seconds* | **trans-delay** *seconds*] *
   By default, no IPsec profile is applied.

Configuring OSPFv3 logging and SNMP notifications

Enabling logging for neighbor state changes

About this task

With this feature enabled, the router delivers logs about neighbor state changes to its information center. The information center processes logs according to user-defined output rules (whether to output logs and where to output). For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter OSPFv3 view.
   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *
3. Enable logging for neighbor state changes.
   log-peer-change
   By default, this feature is enabled.

Setting the maximum number of OSPFv3 logs

About this task

OSPFv3 logs include route calculation logs, neighbor logs, and LSA aging logs.

Procedure

1. Enter system view.

   system-view

2. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

3. Set the maximum number of OSPFv3 logs.

   event-log { **lsa-flush** | **peer** | **spf** } **size** *count*

   By default, the maximum number of LSA aging logs, neighbor logs, or route calculation logs is 10.

Configuring OSPFv3 network management

About this task

This task involves the following configurations:

- Bind an OSPFv3 process to MIB so that you can use network management software to manage the specified OSPFv3 process.
- Enable SNMP notifications for OSPFv3 to report important events.
- Set the SNMP notification output interval and the maximum number of SNMP notifications that can be output at each interval.

To report critical OSPFv3 events to an NMS, enable SNMP notifications for OSPFv3. For SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

The standard OSPFv3 MIB provides only single-instance MIB objects. To identify multiple OSPFv3 processes in the standard OSPFv3 MIB, you must assign a unique context name to each OSPFv3 process.

Context is a method introduced to SNMPv3 for multiple-instance management. For SNMPv1/v2c, you must specify a community name as a context name for protocol identification.

Procedure

1. Enter system view.

   system-view

2. Bind MIB to an OSPFv3 process.

   ospfv3 mib-binding *process-id*

   By default, MIB is bound to the process with the smallest process ID.

3. Enable SNMP notifications for OSPFv3.

   snmp-agent trap enable ospfv3 [**grhelper-status-change** | **grrestarter-status-change** | **if-bad-pkt** | **if-cfg-error** | **if-networktype-mismatch** | **if-state-change** | **inter-area-router-id-conflict** | **intra-area-router-id-conflict** | **last-auth-key-expiry** | **neighbor-state-change** | **nssatranslator-status-change** | **virtgrhelper-status-change** | **virtif-bad-pkt** | **virtif-cfg-error** | **virtif-state-change** | **virtneighbor-state-change**]*

   By default, SNMP notifications for OSPFv3 are enabled.

4. Enter OSPFv3 view.

   ospfv3 [*process-id* | **vpn-instance** *vpn-instance-name*] *

5. Configure an SNMP context for the OSPFv3 process.

snmp context-name *context-name*

By default, no SNMP context is configured for the OSPFv3 process.

6. (Optional.) Set the SNMP notification output interval and the maximum number of SNMP notifications that can be output at each interval.

   snmp trap rate-limit interval *trap-interval* **count** *trap-number*

   By default, OSPFv3 outputs a maximum of seven SNMP notifications within 10 seconds.

Display and maintenance commands for OSPFv3

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display summary route information on the OSPFv3 ABR.	**display ospfv3** [*process-id*] [**area** *area-id*] **abr-summary** [*ipv6-address prefix-length*] [**verbose**]
Display OSPFv3 neighbor information.	**display ospfv3** [*process-id*] [**area** *area-id*] **peer** [[*interface-type interface-number*] [**verbose**] | *peer-router-id* | **statistics**]
Display OSPFv3 request list information.	**display ospfv3** [*process-id*] [**area** *area-id*] **request-queue** [*interface-type interface-number*] [*neighbor-id*]
Display OSPFv3 retransmission list information.	**display ospfv3** [*process-id*] [**area** *area-id*] **retrans-queue** [*interface-type interface-number*] [*neighbor-id*]
Display OSPFv3 topology information.	**display ospfv3** [*process-id*] [**area** *area-id*] **spf-tree** [**verbose**]
Display OSPFv3 process information.	**display ospfv3** [*process-id*] [**verbose**]
Display information about the routes to OSPFv3 ABR and ASBR.	**display ospfv3** [*process-id*] **abr-asbr**
Display summary route information on the OSPFv3 ASBR.	**display ospfv3** [*process-id*] **asbr-summary** [*ipv6-address prefix-length*] [**verbose**]
Display OSPFv3 log information.	**display ospfv3** [*process-id*] **event-log** { **lsa-flush** | **peer** | **spf** }
Display OSPFv3 GR information.	**display ospfv3** [*process-id*] **graceful-restart** [**verbose**]
Display OSPFv3 interface information.	**display ospfv3** [*process-id*] **interface** [*interface-type interface-number* | **verbose**]
Display OSPFv3 LSDB information.	**display ospfv3** [*process-id*] **lsdb** [{ **e-inter-prefix** | **e-inter-router** | **e-router** | **external** | **grace** | **inter-prefix** | **inter-router** | **intra-prefix** | **intra-te** | **link** | **locator** | **network** | **nssa** | **router** | **router-information** | **unknown** [*type*] } [*link-state-id*] [**originate-router** *router-id* | **self-originate**] | **statistics** | **total** | **verbose**]

Task	Command
Display OSPFv3 next hop information.	**display ospfv3** [*process-id*] **nexthop**
Display OSPFv3 NSR information.	**display ospfv3** [*process-id*] **non-stop-routing**
Display OSPFv3 routing information.	**display ospfv3** [*process-id*] **routing** [*ipv6-address prefix-length*]
Display OSPFv3 statistics.	**display ospfv3** [*process-id*] **statistics** [**error**]
Display OSPFv3 virtual link information.	**display ospfv3** [*process-id*] **vlink**
Clear OSPFv3 log information.	**reset ospfv3** [*process-id*] **event-log** [**lsa-flush** | **peer** | **spf**]
Restart an OSPFv3 process.	**reset ospfv3** [*process-id*] **process** [**graceful-restart**]
Restart OSPFv3 route redistribution.	**reset ospfv3** [*process-id*] **redistribution**
Clear OSPFv3 statistics.	**reset ospfv3** [*process-id*] **statistics**

OSPFv3 configuration examples

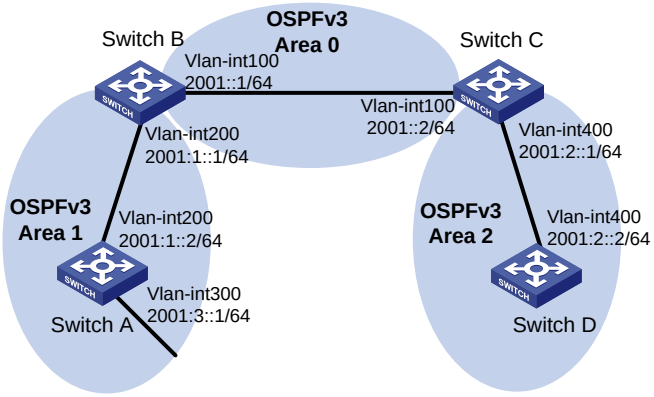
Example: Configuring OSPFv3 stub area

Network configuration

As shown in Figure 5:

- Enable OSPFv3 on all switches.
- Split the AS into three areas.
- Configure Switch B and Switch C as ABRs to forward routing information between areas.
- Configure Area 2 as a stub area to reduce LSAs in the area without affecting route reachability.

Figure 5 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)

2. Configure basic OSPFv3:

On Switch A, enable OSPFv3 and specify the router ID as 1.1.1.1.

```
<SwitchA> system-view
[SwitchA] ospfv3
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 300
[SwitchA-Vlan-interface300] ospfv3 1 area 1
[SwitchA-Vlan-interface300] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ospfv3 1 area 1
[SwitchA-Vlan-interface200] quit
```

On Switch B, enable OSPFv3 and specify the router ID as 2.2.2.2.

```
<SwitchB> system-view
[SwitchB] ospfv3
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ospfv3 1 area 0
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 1 area 1
[SwitchB-Vlan-interface200] quit
```

On Switch C, enable OSPFv3 and specify the router ID as 3.3.3.3.

```
<SwitchC> system-view
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 1 area 0
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 400
[SwitchC-Vlan-interface400] ospfv3 1 area 2
[SwitchC-Vlan-interface400] quit
```

On Switch D, enable OSPFv3 and specify the router ID as 4.4.4.4.

```
<SwitchD> system-view
[SwitchD] ospfv3
[SwitchD-ospfv3-1] router-id 4.4.4.4
[SwitchD-ospfv3-1] quit
[SwitchD] interface vlan-interface 400
[SwitchD-Vlan-interface400] ospfv3 1 area 2
[SwitchD-Vlan-interface400] quit
```

Display OSPFv3 neighbors on Switch B.

```
[SwitchB] display ospfv3 peer

              OSPFv3 Process 1 with Router ID 2.2.2.2

 Area: 0.0.0.0
```

```
---------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 3.3.3.3         1   Full/BDR          00:00:40  0      Vlan100

 Area: 0.0.0.1
---------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 1.1.1.1         1   Full/DR           00:00:40  0      Vlan200
```

Display OSPFv3 neighbors on Switch C.

```
[SwitchC] display ospfv3 peer

              OSPFv3 Process 1 with Router ID 3.3.3.3

 Area: 0.0.0.0
---------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 2.2.2.2         1   Full/DR           00:00:40  0      Vlan100

 Area: 0.0.0.2
---------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 4.4.4.4         1   Full/BDR          00:00:40  0      Vlan400
```

Display OSPFv3 routing table information on Switch D.

```
[SwitchD] display ospfv3 routing

              OSPFv3 Process 1 with Router ID 4.4.4.4
---------------------------------------------------------------------------
 I  - Intra area route,  E1 - Type 1 external route,  N1 - Type 1 NSSA route
 IA - Inter area route,  E2 - Type 2 external route,  N2 - Type 2 NSSA route
 *  - Selected route

*Destination: 2001::/64
  Type       : IA                       Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                  Preference : 10
  NibID      : 0x23000004               Cost       : 2
  Interface  : Vlan400                  BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely
  Delay Flag : N/A

 *Destination: 2001:1::/64
  Type       : IA                       Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                  Preference : 10
  NibID      : 0x23000004               Cost       : 3
  Interface  : Vlan400                  BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
```

```
 Status     : Rely
 Delay Flag : N/A

*Destination: 2001:2::/64
 Type       : I                          Area       : 0.0.0.2
 AdvRouter  : 4.4.4.4                    Preference : 10
 NibID      : 0x23000002                 Cost       : 1
 Interface  : Vlan400                    BkInterface: N/A
 Nexthop    : ::
 BkNexthop  : N/A
 Status     : Direct
 Delay Flag : N/A

*Destination: 2001:3::1/128
 Type       : IA                         Area       : 0.0.0.2
 AdvRouter  : 3.3.3.3                    Preference : 10
 NibID      : 0x23000004                 Cost       : 3
 Interface  : Vlan400                    BkInterface: N/A
 Nexthop    : FE80::48C0:26FF:FEDA:305
 BkNexthop  : N/A
 Status     : Rely
 Delay Flag : N/A

 Total: 4
 Intra area: 1         Inter area: 3         ASE: 0         NSSA: 0
```

3. Configure Area 2 as a stub area:

Configure Switch D.

```
[SwitchD] ospfv3
[SwitchD-ospfv3-1] area 2
[SwitchD-ospfv3-1-area-0.0.0.2] stub
[SwitchD-ospfv3-1-area-0.0.0.2] quit
[SwitchD-ospfv3-1] quit
```

Configure Switch C, and specify the cost of the default route sent to the stub area as 10.

```
[SwitchC] ospfv3
[SwitchC-ospfv3-1] area 2
[SwitchC-ospfv3-1-area-0.0.0.2] stub
[SwitchC-ospfv3-1-area-0.0.0.2] default-cost 10
[SwitchC-ospfv3-1-area-0.0.0.2] quit
[SwitchC-ospfv3-1] quit
```

Display OSPFv3 routing table information on Switch D.

```
[SwitchD] display ospfv3 routing

              OSPFv3 Process 1 with Router ID 4.4.4.4
------------------------------------------------------------------------
 I  - Intra area route,  E1 - Type 1 external route,  N1 - Type 1 NSSA route
 IA - Inter area route,  E2 - Type 2 external route,  N2 - Type 2 NSSA route
 *  - Selected route
```

```
*Destination: ::/0
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000003                Cost       : 11
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely
  Delay Flag : N/A

 *Destination: 2001::/64
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000003                Cost       : 2
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely
  Delay Flag : N/A

 *Destination: 2001:1::/64
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000003                Cost       : 3
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely
  Delay Flag : N/A

 *Destination: 2001:2::/64
  Type       : I                         Area       : 0.0.0.2
  AdvRouter  : 4.4.4.4                   Preference : 10
  NibID      : 0x23000001                Cost       : 1
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : ::
  BkNexthop  : N/A
  Status     : Direct
  Delay Flag : N/A

 *Destination: 2001:3::1/128
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000003                Cost       : 3
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely
```

```
  Delay Flag : N/A

 Total: 5
 Intra area: 1         Inter area: 4         ASE: 0         NSSA: 0
```

The output shows that a default route is added, and its cost is the cost of a direct route plus the configured cost.

4. Configure Area 2 as a totally stub area:

Configure Area 2 as a totally stub area on Switch C.

```
[SwitchC] ospfv3
[SwitchC-ospfv3-1] area 2
[SwitchC-ospfv3-1-area-0.0.0.2] stub no-summary
[SwitchC-ospfv3-1-area-0.0.0.2] quit
[SwitchC-ospfv3-1] quit
```

Display OSPFv3 routing table information on Switch D.

```
[SwitchD] display ospfv3 routing

               OSPFv3 Process 1 with Router ID 4.4.4.4
------------------------------------------------------------------------------
 I  - Intra area route,  E1 - Type 1 external route,  N1 - Type 1 NSSA route
 IA - Inter area route,  E2 - Type 2 external route,  N2 - Type 2 NSSA route
 *  - Selected route

*Destination: ::/0
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000003                Cost       : 11
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely
  Delay Flag : N/A

 *Destination: 2001:2::/64
  Type       : I                         Area       : 0.0.0.2
  AdvRouter  : 4.4.4.4                   Preference : 10
  NibID      : 0x23000001                Cost       : 1
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : ::
  BkNexthop  : N/A
  Status     : Direct
  Delay Flag : N/A

 Total: 2
 Intra area: 1         Inter area: 1         ASE: 0         NSSA: 0
```

The output shows that route entries are reduced. All indirect routes are removed, except the default route.

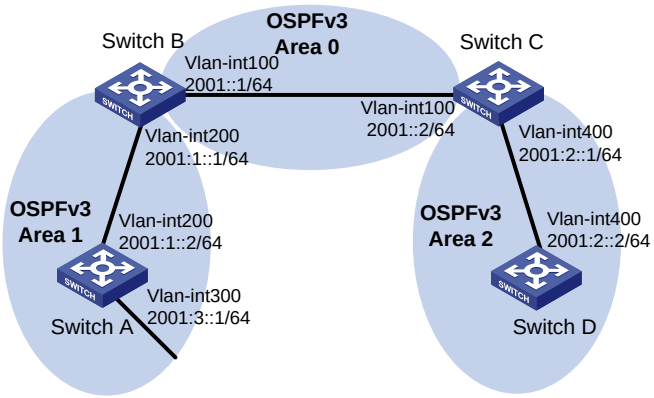
Example: Configuring OSPFv3 NSSA area

Network configuration

As shown in Figure 6:

- Configure OSPFv3 on all switches and split the AS into three areas.
- Configure Switch B and Switch C as ABRs to forward routing information between areas.
- Configure Area 1 as an NSSA area and configure Switch A as an ASBR to redistribute static routes into the AS.

Figure 6 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure basic OSPFv3 (see "Example: Configuring OSPFv3 stub area").
3. Configure Area 1 as an NSSA area:

   # Configure Switch A.

```
[SwitchA] ospfv3
[SwitchA-ospfv3-1] area 1
[SwitchA-ospfv3-1-area-0.0.0.1] nssa
[SwitchA-ospfv3-1-area-0.0.0.1] quit
[SwitchA-ospfv3-1] quit
```

   # Configure Switch B.

```
[SwitchB] ospfv3
[SwitchB-ospfv3-1] area 1
[SwitchB-ospfv3-1-area-0.0.0.1] nssa
[SwitchB-ospfv3-1-area-0.0.0.1] quit
[SwitchB-ospfv3-1] quit
```

   # Display OSPFv3 routing information on Switch D.

```
[SwitchD] display ospfv3 1 routing

                OSPFv3 Process 1 with Router ID 4.4.4.4
------------------------------------------------------------------------
 I   - Intra area route,  E1 - Type 1 external route,  N1 - Type 1 NSSA route
 IA  - Inter area route,  E2 - Type 2 external route,  N2 - Type 2 NSSA route
 *   - Selected route
```

```
*Destination: 2001::/64
  Type       : IA                         Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                    Preference : 10
  NibID      : 0x23000003                 Cost       : 2
  Interface  : Vlan200                    BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely

 *Destination: 2001:1::/64
  Type       : IA                         Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                    Preference : 10
  NibID      : 0x23000003                 Cost       : 3
  Interface  : Vlan200                    BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely

 *Destination: 2001:2::/64
  Type       : I                          Area       : 0.0.0.2
  AdvRouter  : 4.4.4.4                    Preference : 10
  NibID      : 0x23000001                 Cost       : 1
  Interface  : Vlan200                    BkInterface: N/A
  Nexthop    : ::
  BkNexthop  : N/A
  Status     : Direct

 *Destination: 2001:3::/64
  Type       : IA                         Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                    Preference : 10
  NibID      : 0x23000003                 Cost       : 4
  Interface  : Vlan200                    BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely

 Total: 4
 Intra area: 1         Inter area: 3         ASE: 0         NSSA: 0
```

4. Configure route redistribution:

Configure an IPv6 static route, and configure OSPFv3 to redistribute the static route on Switch A.

```
[SwitchA] ipv6 route-static 1234:: 64 null 0
[SwitchA] ospfv3 1
[SwitchA-ospfv3-1] import-route static
[SwitchA-ospfv3-1] quit
```

Display OSPFv3 routing information on Switch D.

```
[SwitchD] display ospfv3 1 routing
```

```
                OSPFv3 Process 1 with Router ID 4.4.4.4
------------------------------------------------------------------------------
 I  - Intra area route,  E1 - Type 1 external route,  N1 - Type 1 NSSA route
 IA - Inter area route,  E2 - Type 2 external route,  N2 - Type 2 NSSA route
 *  - Selected route

*Destination: 2001::/64
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000002                Cost       : 2
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely

 *Destination: 2001:1::/64
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000002                Cost       : 3
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely

 *Destination: 2001:2::/64
  Type       : I                         Area       : 0.0.0.2
  AdvRouter  : 4.4.4.4                   Preference : 10
  NibID      : 0x23000004                Cost       : 1
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : ::
  BkNexthop  : N/A
  Status     : Direct

 *Destination: 2001:3::/64
  Type       : IA                        Area       : 0.0.0.2
  AdvRouter  : 3.3.3.3                   Preference : 10
  NibID      : 0x23000002                Cost       : 4
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
  BkNexthop  : N/A
  Status     : Rely

 *Destination: 1234::/64
  Type       : E2                        Tag        : 1
  AdvRouter  : 2.2.2.2                   Preference : 150
  NibID      : 0x23000001                Cost       : 1
  Interface  : Vlan400                   BkInterface: N/A
  Nexthop    : FE80::48C0:26FF:FEDA:305
```

```
 BkNexthop  : N/A
 Status     : Normal

Total: 5
Intra area: 1         Inter area: 3         ASE: 1         NSSA: 0
```

The output shows an AS external route imported from the NSSA area exists on Switch D.

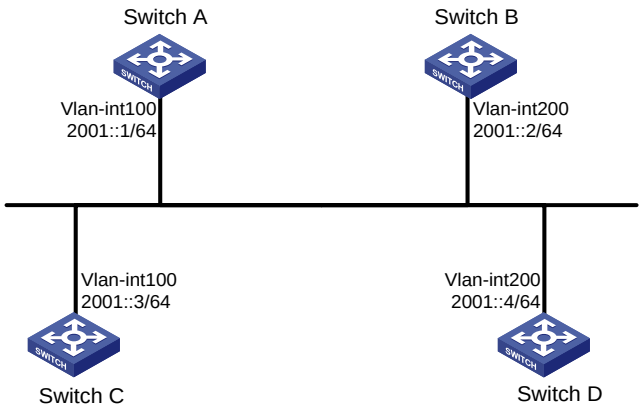
Example: Configuring OSPFv3 DR election

Network configuration

As shown in Figure 7:

- Configure router priority 100 for Switch A, the highest priority on the network, so it will become the DR.
- Configure router priority 2 for Switch C, the second highest priority on the network, so it will become the BDR.
- Configure router priority 0 for Switch B, so it cannot become a DR or BDR.
- Switch D uses the default router priority 1.

Figure 7 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure basic OSPFv3:

   # On Switch A, enable OSPFv3 and specify the router ID as 1.1.1.1.

```
<SwitchA> system-view
[SwitchA] ospfv3
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 1 area 0
[SwitchA-Vlan-interface100] quit
```

   # On Switch B, enable OSPFv3 and specify the router ID as 2.2.2.2.

```
<SwitchB> system-view
[SwitchB] ospfv3
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
```

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 1 area 0
[SwitchB-Vlan-interface200] quit
```

On Switch C, enable OSPFv3 and specify the router ID as 3.3.3.3.

```
<SwitchC> system-view
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 1 area 0
[SwitchC-Vlan-interface100] quit
```

On Switch D, enable OSPFv3 and specify the router ID as 4.4.4.4.

```
<SwitchD> system-view
[SwitchD] ospfv3
[SwitchD-ospfv3-1] router-id 4.4.4.4
[SwitchD-ospfv3-1] quit
[SwitchD] interface vlan-interface 200
[SwitchD-Vlan-interface200] ospfv3 1 area 0
[SwitchD-Vlan-interface200] quit
```

Display neighbor information on Switch A. The switches have the same default DR priority 1, so Switch D (the switch with the highest router ID) is elected as the DR, and Switch C is the BDR.

```
[SwitchA] display ospfv3 peer

              OSPFv3 Process 1 with Router ID 1.1.1.1

 Area: 0.0.0.0
--------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 2.2.2.2         1   2-Way/DROther     00:00:36  0      Vlan200
 3.3.3.3         1   Full/BDR          00:00:35  0      Vlan100
 4.4.4.4         1   Full/DR           00:00:33  0      Vlan200
```

Display neighbor information on Switch D. The neighbor states are all full.

```
[SwitchD] display ospfv3 peer

              OSPFv3 Process 1 with Router ID 4.4.4.4

 Area: 0.0.0.0
--------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 1.1.1.1         1   Full/DROther      00:00:30  0      Vlan100
 2.2.2.2         1   Full/DROther      00:00:37  0      Vlan200
 3.3.3.3         1   Full/BDR          00:00:31  0      Vlan100
```

3. Configure router priorities for interfaces:

Set the router priority of VLAN-interface 100 to 100 on Switch A.

```
[SwitchA] interface Vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 dr-priority 100
[SwitchA-Vlan-interface100] quit
```

Set the router priority of VLAN-interface 200 to 0 on Switch B.

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 dr-priority 0
[SwitchB-Vlan-interface200] quit
```

Set the router priority of VLAN-interface 100 to 2 on Switch C.

```
[SwitchC] interface Vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 dr-priority 2
[SwitchC-Vlan-interface100] quit
```

Display neighbor information on Switch A. Router priorities have been updated, but the DR and BDR are not changed.

```
[SwitchA] display ospfv3 peer

              OSPFv3 Process 1 with Router ID 1.1.1.1

 Area: 0.0.0.0
--------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 2.2.2.2         0   2-Way/DROther     00:00:36  0      Vlan200
 3.3.3.3         2   Full/BDR          00:00:35  0      Vlan200
 4.4.4.4         1   Full/DR           00:00:33  0      Vlan200
```

Display neighbor information on Switch D. Switch D is still the DR.

```
[SwitchD] display ospfv3 peer

              OSPFv3 Process 1 with Router ID 4.4.4.4

 Area: 0.0.0.0
--------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 1.1.1.1         100 Full/DROther      00:00:30  0      Vlan100
 2.2.2.2         0   Full/DROther      00:00:37  0      Vlan200
 3.3.3.3         2   Full/BDR          00:00:31  0      Vlan100
```

4. Restart DR and BDR election:

Use the **shutdown** and **undo shutdown** commands on interfaces to restart DR and BDR election. (Details not shown.)

Display neighbor information on Switch A. The output shows that Switch C becomes the BDR.

```
[SwitchA] display ospfv3 peer

              OSPFv3 Process 1 with Router ID 1.1.1.1

 Area: 0.0.0.0
--------------------------------------------------------------------------
 Router ID       Pri State             Dead-Time InstID Interface
 2.2.2.2         0   Full/DROther      00:00:36  0      Vlan200
 3.3.3.3         2   Full/BDR          00:00:35  0      Vlan100
 4.4.4.4         1   Full/DROther      00:00:33  0      Vlan200
```

Display neighbor information on Switch D.

```
[SwitchD] display ospfv3 peer
```

```
                OSPFv3 Process 1 with Router ID 4.4.4.4

 Area: 0.0.0.0
------------------------------------------------------------------------
 Router ID       Pri State            Dead-Time InstID Interface
 1.1.1.1         100 Full/DR          00:00:30  0      Vlan100
 2.2.2.2         0   2-Way/DROther    00:00:37  0      Vlan200
 3.3.3.3         2   Full/BDR         00:00:31  0      Vlan100
```

The output shows that Switch A becomes the DR.

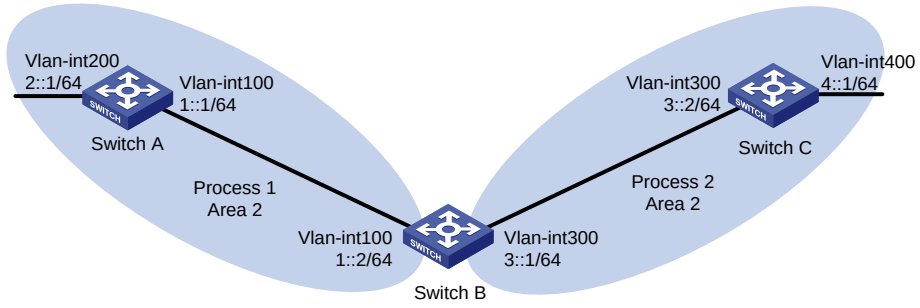
Example: Configuring OSPFv3 route redistribution

Network configuration

As shown in Figure 8:

- Switch A, Switch B, and Switch C are in Area 2.
- OSPFv3 process 1 and OSPFv3 process 2 run on Switch B. Switch B communicates with Switch A and Switch C through OSPFv3 process 1 and OSPFv3 process 2.
- Configure OSPFv3 process 2 to redistribute direct routes and the routes from OSPFv3 process 1 on Switch B, and set the metric for redistributed routes to 3. Switch C can then learn the routes destined for 1::0/64 and 2::0/64, and Switch A cannot learn the routes destined for 3::0/64 or 4::0/64.

Figure 8 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure basic OSPFv3:

   # Enable OSPFv3 process 1 on Switch A.

```
<SwitchA> system-view
[SwitchA] ospfv3 1
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 1 area 2
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ospfv3 1 area 2
[SwitchA-Vlan-interface200] quit
```

   # Enable OSPFv3 process 1 and OSPFv3 process 2 on Switch B.

```
<SwitchB> system-view
```

```
[SwitchB] ospfv3 1
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ospfv3 1 area 2
[SwitchB-Vlan-interface100] quit
[SwitchB] ospfv3 2
[SwitchB-ospfv3-2] router-id 3.3.3.3
[SwitchB-ospfv3-2] quit
[SwitchB] interface vlan-interface 300
[SwitchB-Vlan-interface300] ospfv3 2 area 2
[SwitchB-Vlan-interface300] quit
```

Enable OSPFv3 process 2 on Switch C.

```
<SwitchC> system-view
[SwitchC] ospfv3 2
[SwitchC-ospfv3-2] router-id 4.4.4.4
[SwitchC-ospfv3-2] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] ospfv3 2 area 2
[SwitchC-Vlan-interface300] quit
[SwitchC] interface vlan-interface 400
[SwitchC-Vlan-interface400] ospfv3 2 area 2
[SwitchC-Vlan-interface400] quit
```

Display the routing table on Switch C.

```
[SwitchC] display ipv6 routing-table

Destinations : 7 Routes : 7

Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 3::/64                                     Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan300                                    Cost      : 0

Destination: 3::2/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 4::/64                                     Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan400                                    Cost      : 0

Destination: 4::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0
```

```
Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0

Destination: FF00::/8                                   Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0
```

3. Configure OSPFv3 route redistribution:

Configure OSPFv3 process 2 to redistribute direct routes and the routes from OSPFv3 process 1 on Switch B, and set the metric for redistributed routes to 3.

```
[SwitchB] ospfv3 2
[SwitchB-ospfv3-2] import-route ospfv3 1 cost 3
[SwitchB-ospfv3-2] import-route direct cost 3
[SwitchB-ospfv3-2] quit
```

Display the routing table on Switch C.

```
[SwitchC] display ipv6 routing-table

Destinations : 9 Routes : 9

Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 1::/64                                     Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                    Preference: 150
Interface  : Vlan300                                    Cost      : 3

Destination: 2::/64                                     Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                    Preference: 150
Interface  : Vlan300                                    Cost      : 3

Destination: 3::/64                                     Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan300                                    Cost      : 0

Destination: 3::2/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 4::/64                                     Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan400                                    Cost      : 0

Destination: 4::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: FE80::/10                                  Protocol  : Direct
```

```
NextHop    : ::                                    Preference: 0
Interface  : NULL0                                 Cost      : 0

Destination: FF00::/8                              Protocol  : Direct
NextHop    : ::                                    Preference: 0
Interface  : NULL0                                 Cost      : 0
```

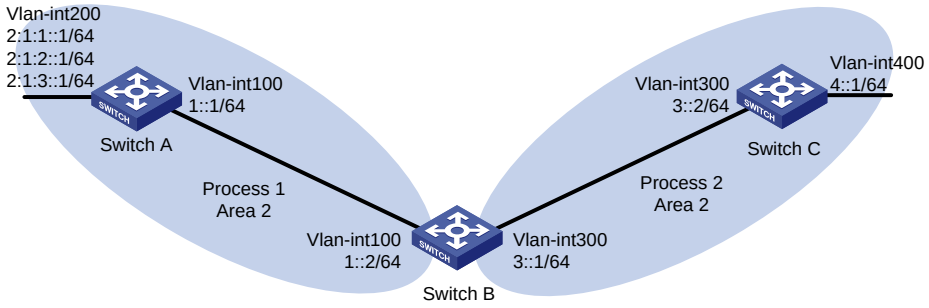
Example: Configuring OSPFv3 route summarization

Network configuration

As shown in Figure 9:

- Switch A, Switch B, and Switch C are in Area 2.
- OSPFv3 process 1 and OSPFv3 process 2 run on Switch B. Switch B communicates with Switch A and Switch C through OSPFv3 process 1 and OSPFv3 process 2, respectively.
- On Switch A, configure IPv6 addresses 2:1:1::1/64, 2:1:2::1/64, and 2:1:3::1/64 for VLAN-interface 200.
- On Switch B, configure OSPFv3 process 2 to redistribute direct routes and the routes from OSPFv3 process 1. Switch C can then learn the routes destined for 2::/64, 2:1:1::/64, 2:1:2::/64, and 2:1:3::/64.
- On Switch B, configure route summarization to advertise only summary route 2::/16 to Switch C.

Figure 9 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure OSPFv3:

   # Enable OSPFv3 process 1 on Switch A.

```
<SwitchA> system-view
[SwitchA] ospfv3 1
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 1 area 2
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ipv6 address 2:1:1::1 64
[SwitchA-Vlan-interface200] ipv6 address 2:1:2::1 64
[SwitchA-Vlan-interface200] ipv6 address 2:1:3::1 64
```

```
[SwitchA-Vlan-interface200] ospfv3 1 area 2
[SwitchA-Vlan-interface200] quit
```

Enable OSPFv3 process 1 and OSPFv3 process 2 on Switch B.

```
<SwitchB> system-view
[SwitchB] ospfv3 1
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ospfv3 1 area 2
[SwitchB-Vlan-interface100] quit
[SwitchB] ospfv3 2
[SwitchB-ospfv3-2] router-id 3.3.3.3
[SwitchB-ospfv3-2] quit
[SwitchB] interface vlan-interface 300
[SwitchB-Vlan-interface300] ospfv3 2 area 2
[SwitchB-Vlan-interface300] quit
```

Enable OSPFv3 process 2 on Switch C.

```
<SwitchC> system-view
[SwitchC] ospfv3 2
[SwitchC-ospfv3-2] router-id 4.4.4.4
[SwitchC-ospfv3-2] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] ospfv3 2 area 2
[SwitchC-Vlan-interface300] quit
[SwitchC] interface vlan-interface 400
[SwitchC-Vlan-interface400] ospfv3 2 area 2
[SwitchC-Vlan-interface400] quit
```

3. Configure OSPFv3 route redistribution:

Configure OSPFv3 process 2 to redistribute direct routes and the routes from OSPFv3 process 1 on Switch B.

```
[SwitchB] ospfv3 2
[SwitchB-ospfv3-2] import-route ospfv3 1
[SwitchB-ospfv3-2] import-route direct
[SwitchB-ospfv3-2] quit
```

Display the routing table on Switch C.

```
[SwitchC] display ipv6 routing-table

Destinations : 12 Routes : 12

Destination: ::1/128                                Protocol  : Direct
NextHop    : ::1                                    Preference: 0
Interface  : InLoop0                                Cost      : 0

Destination: 1::/64                                 Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                Preference: 150
Interface  : Vlan300                                Cost      : 1

Destination: 2::/64                                 Protocol  : O_ASE2
```

```
NextHop    : FE80::200:CFF:FE01:1C03                Preference: 150
Interface  : Vlan300                                Cost      : 1

Destination: 2:1:1::/64                             Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                Preference: 150
Interface  : Vlan300                                Cost      : 1

Destination: 2:1:2::/64                             Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                Preference: 150
Interface  : Vlan300                                Cost      : 1

Destination: 2:1:3::/64                             Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                Preference: 150
Interface  : Vlan300                                Cost      : 1

Destination: 3::/64                                 Protocol  : Direct
NextHop    : 3::2                                   Preference: 0
Interface  : Vlan300                                Cost      : 0

Destination: 3::2/128                               Protocol  : Direct
NextHop    : ::1                                    Preference: 0
Interface  : InLoop0                                Cost      : 0

Destination: 4::/64                                 Protocol  : Direct
NextHop    : 4::1                                   Preference: 0
Interface  : Vlan400                                Cost      : 0

Destination: 4::1/128                               Protocol  : Direct
NextHop    : ::1                                    Preference: 0
Interface  : InLoop0                                Cost      : 0

Destination: FE80::/10                              Protocol  : Direct
NextHop    : ::                                     Preference: 0
Interface  : NULL0                                  Cost      : 0

Destination: FF00::/8                               Protocol  : Direct
NextHop    : ::                                     Preference: 0
Interface  : NULL0                                  Cost      : 0
```

4. Configure ASBR route summarization:

On Switch B, configure OSPFv3 process 2 to advertise a single route 2::/16.

```
[SwitchB] ospfv3 2
[SwitchB-ospfv3-2] asbr-summary 2:: 16
[SwitchB-ospfv3-2] quit
```

Display the routing table on Switch C.

```
[SwitchC] display ipv6 routing-table

Destinations : 9 Routes : 9
```

```
Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 1::/64                                     Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                    Preference: 150
Interface  : Vlan300                                    Cost      : 1

Destination: 2::/16                                     Protocol  : O_ASE2
NextHop    : FE80::200:CFF:FE01:1C03                    Preference: 150
Interface  : Vlan300                                    Cost      : 1

Destination: 3::/64                                     Protocol  : Direct
NextHop    : 3::2                                       Preference: 0
Interface  : Vlan300                                    Cost      : 0

Destination: 3::2/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 4::/64                                     Protocol  : Direct
NextHop    : 4::1                                       Preference: 0
Interface  : Vlan400                                    Cost      : 0

Destination: 4::1/128                                   Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0

Destination: FF00::/8                                   Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0
```

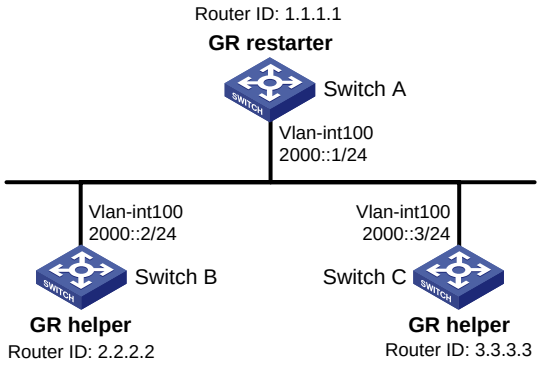
Example: Configuring OSPFv3 GR

Network configuration

As shown in Figure 10:

- Switch A, Switch B, and Switch C that reside in the same AS and the same OSPFv3 routing domain are GR capable.
- Switch A acts as the GR restarter. Switch B and Switch C act as the GR helpers, and synchronize their LSDBs with Switch A through GR.

Figure 10 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure basic OSPFv3:

   # On Switch A, enable OSPFv3 process 1, enable GR, and set the router ID to 1.1.1.1.

   ```
   <SwitchA> system-view
   [SwitchA] ospfv3 1
   [SwitchA-ospfv3-1] router-id 1.1.1.1
   [SwitchA-ospfv3-1] graceful-restart enable
   [SwitchA-ospfv3-1] quit
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] ospfv3 1 area 1
   [SwitchA-Vlan-interface100] quit
   ```

   # On Switch B, enable OSPFv3 and set the router ID to 2.2.2.2. (By default, GR helper is enabled on Switch B.)

   ```
   <SwitchB> system-view
   [SwitchB] ospfv3 1
   [SwitchB-ospfv3-1] router-id 2.2.2.2
   [SwitchB-ospfv3-1] quit
   [SwitchB] interface vlan-interface 100
   [SwitchB-Vlan-interface100] ospfv3 1 area 1
   [SwitchB-Vlan-interface100] quit
   ```

   # On Switch C, enable OSPFv3 and set the router ID to 3.3.3.3. (By default, GR helper is enabled on Switch C.)

   ```
   <SwitchC> system-view
   [SwitchC] ospfv3 1
   [SwitchC-ospfv3-1] router-id 3.3.3.3
   [SwitchC-ospfv3-1] quit
   [SwitchC] interface vlan-interface 100
   [SwitchC-Vlan-interface100] ospfv3 1 area 1
   [SwitchC-Vlan-interface100] quit
   ```

Verifying the configuration

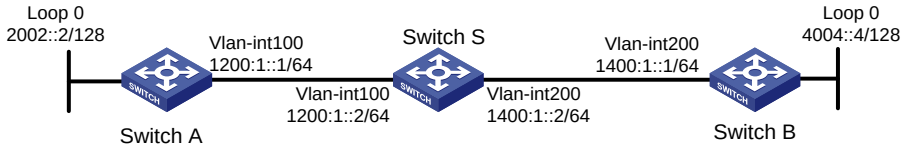
Perform a master/backup switchover on Switch A to trigger an OSPFv3 GR operation. (Details not shown.)

Example: Configuring OSPFv3 NSR

Network configuration

As shown in Figure 11, Switch S, Switch A, and Switch B belong to the same OSPFv3 routing domain. Enable OSPFv3 NSR on Switch S to ensure correct routing when an active/standby switchover occurs on Switch S.

Figure 11 Network diagram



Procedure

1. Configure IP addresses and subnet masks for interfaces on the switches. (Details not shown.)
2. Configure OSPFv3 on the switches to ensure that Switch S, Switch A, and Switch B can communicate with each other at Layer 3. (Details not shown.)
3. Configure OSPFv3:

   # On Switch A, enable OSPFv3, and set the router ID to 1.1.1.1.

   ```
   <SwitchA> system-view
   [SwitchA] ospfv3 1
   [SwitchA-ospfv3-1] router-id 1.1.1.1
   [SwitchA-ospfv3-1] quit
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] ospfv3 1 area 1
   [SwitchA-Vlan-interface100] quit
   ```

   # On Switch B, enable OSPFv3, and set the router ID to 2.2.2.2.

   ```
   <SwitchB> system-view
   [SwitchB] ospfv3 1
   [SwitchB-ospfv3-1] router-id 2.2.2.2
   [SwitchB-ospfv3-1] quit
   [SwitchB] interface vlan-interface 200
   [SwitchB-Vlan-interface200] ospfv3 1 area 1
   [SwitchB-Vlan-interface200] quit
   ```

   # On Switch S, enable OSPFv3, and set the router ID to 3.3.3.3.

   ```
   <SwitchS> system-view
   [SwitchS] ospfv3 1
   [SwitchS-ospfv3-1] router-id 3.3.3.3
   [SwitchS-ospfv3-1] quit
   [SwitchS] interface vlan-interface 100
   [SwitchS-Vlan-interface100] ospfv3 1 area 1
   [SwitchS-Vlan-interface100] quit
   [SwitchS] interface vlan-interface 200
   [SwitchS-Vlan-interface200] ospfv3 1 area 1
   [SwitchS-Vlan-interface200] quit
   ```

4. Configure OSPFv3 NSR:

   # Enable OSPFv3 NSR on Switch S.

```
[SwitchS] ospfv3 1
[SwitchS-ospfv3-1] non-stop-routing
[SwitchS-ospfv3-1] quit
```

Verifying the configuration

Verify the following:

- When an active/standby switchover occurs on Switch S, the neighbor relationships and routing information on Switch A and Switch B have not changed. (Details not shown.)
- The traffic from Switch A to Switch B has not been impacted. (Details not shown.)

Example: Configuring BFD for OSPFv3

Network configuration

As shown in Figure 12:

- Configure OSPFv3 on Switch A, Switch B and Switch C and configure BFD over the link Switch A<—>L2 Switch<—>Switch B.
- After the link Switch A<—>L2 Switch<—>Switch B fails, BFD can quickly detect the failure and notify OSPFv3 of the failure. Then Switch A and Switch B communicate through Switch C.

Figure 12 Network diagram

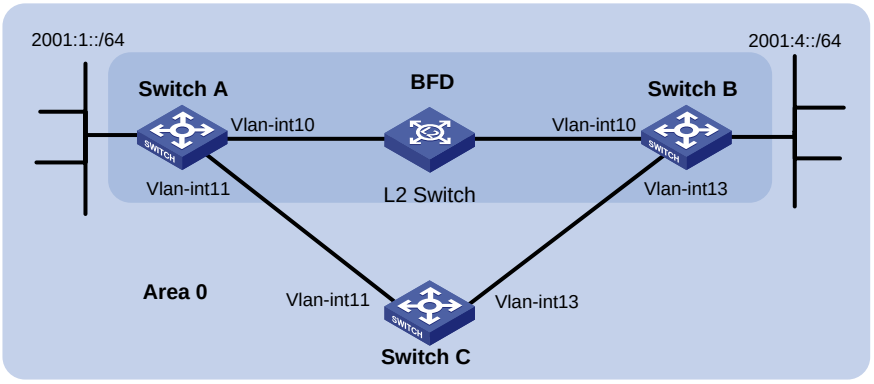


Table 1 Interface and IP address assignment

Device	Interface	IPv6 address
Switch A	Vlan-int10	2001::1/64
Switch A	Vlan-int11	2001:2::1/64
Switch B	Vlan-int10	2001::2/64
Switch B	Vlan-int13	2001:3::2/64
Switch C	Vlan-int11	2001:2::2/64
Switch C	Vlan-int13	2001:3::1/64

Procedure

1. Configure IPv6 addresses for the interfaces. (Details not shown.)
2. Configure basic OSPFv3:

   # On Switch A, enable OSPFv3 and specify the router ID as 1.1.1.1.

   ```
   <SwitchA> system-view
   [SwitchA] ospfv3
   ```

```
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ospfv3 1 area 0
[SwitchA-Vlan-interface10] quit
[SwitchA] interface vlan-interface 11
[SwitchA-Vlan-interface11] ospfv3 1 area 0
[SwitchA-Vlan-interface11] quit
```

On Switch B, enable OSPFv3 and specify the router ID as 2.2.2.2.

```
<SwitchB> system-view
[SwitchB] ospfv3
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ospfv3 1 area 0
[SwitchB-Vlan-interface10] quit
[SwitchB] interface vlan-interface 13
[SwitchB-Vlan-interface13] ospfv3 1 area 0
[SwitchB-Vlan-interface13] quit
```

On Switch C, enable OSPFv3 and specify the router ID as 3.3.3.3.

```
<SwitchC> system-view
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 11
[SwitchC-Vlan-interface11] ospfv3 1 area 0
[SwitchC-Vlan-interface11] quit
[SwitchC] interface vlan-interface 13
[SwitchC-Vlan-interface13] ospfv3 1 area 0
[SwitchC-Vlan-interface13] quit
```

3. Configure BFD:

Enable BFD and configure BFD parameters on Switch A.

```
[SwitchA] bfd session init-mode active
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ospfv3 bfd enable
[SwitchA-Vlan-interface10] bfd min-transmit-interval 500
[SwitchA-Vlan-interface10] bfd min-receive-interval 500
[SwitchA-Vlan-interface10] bfd detect-multiplier 7
[SwitchA-Vlan-interface10] quit
```

Enable BFD and configure BFD parameters on Switch B.

```
[SwitchB] bfd session init-mode active
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ospfv3 bfd enable
[SwitchB-Vlan-interface10] bfd min-transmit-interval 500
[SwitchB-Vlan-interface10] bfd min-receive-interval 500
[SwitchB-Vlan-interface10] bfd detect-multiplier 6
[SwitchB-Vlan-interface10] quit
```

Verifying the configuration

Display the BFD information on Switch A.

```
[SwitchA] display bfd session

 Total sessions: 1     Up sessions: 1     Init mode: Active

 IPv6 session working in control packet mode:

          Local discr: 1441                      Remote discr: 1450
            Source IP: FE80::20F:FF:FE00:1202 (link-local address of VLAN-interface 10
on Switch A)
       Destination IP: FE80::20F:FF:FE00:1200 (link-local address of VLAN-interface 10
on Switch B)
         Session state: Up                           Interface: Vlan10
             Hold time: 2319ms
```

Display routes destined for 2001:4::0/64 on Switch A.

```
[SwitchA] display ipv6 routing-table 2001:4::0 64

Summary count : 1

Destination: 2001:4::/64                                Protocol  : O_INTRA
NextHop    : FE80::20F:FF:FE00:1200                     Preference: 10
Interface  : Vlan10                                     Cost      : 1
```

The output information shows that Switch A communicates with Switch B through VLAN-interface 10. The link over VLAN-interface 10 fails.

Display routes to 2001:4::0/64 on Switch A.

```
[SwitchA] display ipv6 routing-table 2001:4::0 64

Summary count : 1

Destination: 2001:4::/64                                Protocol  : O_INTRA
NextHop    : FE80::BAAF:67FF:FE27:DCD0                  Preference: 10
Interface  : Vlan11                                     Cost      : 2
```

The output shows that Switch A communicates with Switch B through VLAN-interface 11.

Example: Configuring OSPFv3 FRR

Network configuration

As shown in Figure 13, Switch A, Switch B, and Switch C reside in the same OSPFv3 domain. Configure OSPFv3 FRR so that when Link A fails, traffic is immediately switched to Link B.

Figure 13 Network diagram

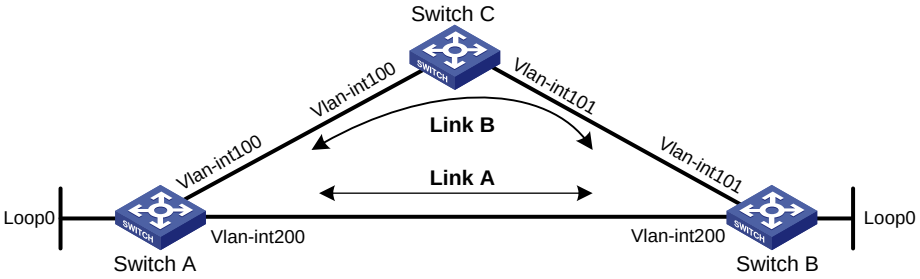


Table 2 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	1::1/64	Switch B	Vlan-int101	3::1/64
	Vlan-int200	2::1/64		Vlan-int200	2::2/64
	Loop0	10::1/128		Loop0	20::1/128
Switch C	Vlan-int100	1::2/64			
	Vlan-int101	3::2/64			

Procedure

1. Configure IPv6 addresses and subnet masks for interfaces on the switches. (Details not shown.)
2. Configure OSPFv3 on the switches to ensure that Switch A, Switch B, and Switch C can communicate with each other at the network layer. (Details not shown.)
3. Configure OSPFv3 FRR to automatically calculate the backup next hop:

   You can enable OSPFv3 FRR to either calculate a backup next hop by using the LFA algorithm, or specify a backup next hop by using a routing policy.

   - (Method 1.) Enable OSPFv3 FRR to calculate the backup next hop by using the LFA algorithm:

     # Configure Switch A.

     ```
     <SwitchA> system-view
     [SwitchA] ospfv3 1
     [SwitchA-ospfv3-1] fast-reroute lfa
     [SwitchA-ospfv3-1] quit
     ```

     # Configure Switch B.

     ```
     <SwitchB> system-view
     [SwitchB] ospfv3 1
     [SwitchB-ospfv3-1] fast-reroute lfa
     [SwitchB-ospfv3-1] quit
     ```

   - (Method 2.) Enable OSPFv3 FRR to designate a backup next hop by using a routing policy:

     # Configure Switch A.

     ```
     <SwitchA> system-view
     [SwitchA] ipv6 prefix-list abc index 10 permit 20::1 128
     [SwitchA] route-policy frr permit node 10
     [SwitchA-route-policy-frr-10] if-match ipv6 address prefix-list abc
     [SwitchA-route-policy-frr-10] apply ipv6 fast-reroute backup-interface
     vlan-interface 100 backup-nexthop fe80::6eab:8bff:fed4:205
     ```

```
[SwitchA-route-policy-frr-10] quit
[SwitchA] ospfv3 1
[SwitchA-ospfv3-1] fast-reroute route-policy frr
[SwitchA-ospfv3-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6 prefix-list abc index 10 permit 10::1 128
[SwitchB] route-policy frr permit node 10
[SwitchB-route-policy-frr-10] if-match ipv6 address prefix-list abc
[SwitchB-route-policy-frr-10] apply ipv6 fast-reroute backup-interface
vlan-interface 101 backup-nexthop fe80::6eab:8bff:fed4:206
[SwitchB-route-policy-frr-10] quit
[SwitchB] ospfv3 1
[SwitchB-ospfv3-1] fast-reroute route-policy frr
[SwitchB-ospfv3-1] quit
```

Verifying the configuration

Display the route 20::1/128 on Switch A to view the backup next hop information.

```
[SwitchA] display ipv6 routing-table 20::1 128 verbose

Summary count : 1

Destination: 20::1/128
    Protocol: O_INTRA
  Process ID: 1
   SubProtID: 0x1                    Age: 00h03m45s
        Cost: 6                Preference: 10
       IpPre: N/A              QosLocalID: N/A
         Tag: 0                    State: Active Adv
   OrigTblID: 0x0                OrigVrf: default-vrf
     TableID: 0xa                 OrigAs: 0
       NibID: 0x23000005          LastAs: 0
      AttrID: 0xffffffff
    BkAttrID: 0x1000000          Neighbor: ::
       Flags: 0x10004          OrigNextHop: FE80::6EAB:90FF:FEF3:305
       Label: NULL             RealNextHop: FE80::6EAB:90FF:FEF3:305
     BkLabel: NULL               BkNextHop: FE80::6EAB:8BFF:FED4:205
     SRLabel: NULL               Interface: Vlan-interface200
   BkSRLabel: NULL             BkInterface: N/A
   Tunnel ID: Invalid          IPInterface: Vlan-interface200
 BkTunnel ID: Invalid        BkIPInterface: Vlan-interface100
    LinkCost: 0                 MicroSegID: 0
 RealFIRType: Normal             RealThres: 0
     InLabel: 0               ColorInterface: N/A
    SIDIndex: NULL          BkColorInterface: N/A
    FtnIndex: 0x0            TunnelInterface: N/A
TrafficIndex: N/A          BkTunnelInterface: N/A
   Connector: N/A                    PathID: 0x0
    SID Type: N/A                       NID: Invalid
```

```
    FlushNID: Invalid                BkNID: Invalid
  BkFlushNID: Invalid             StatFlags: 0x0
         SID: N/A
       BkSID: N/A
CommBlockLen: 0                SRTunnelID: Invalid
  MemberPort: N/A                  ExtFlags: 0x0
     SliceID: 0                   BkSliceID: 0
```

Display the route 10::1/128 on Switch B to view the backup next hop information.

```
[SwitchB] display ipv6 routing-table 10::1 128 verbose

Summary count : 1

Destination: 10::1/128
    Protocol: O_INTRA
  Process ID: 1
   SubProtID: 0x1                     Age: 00h03m10s
        Cost: 1                  Preference: 10
       IpPre: N/A                QosLocalID: N/A
         Tag: 0                       State: Active Adv
   OrigTblID: 0x0                   OrigVrf: default-vrf
     TableID: 0xa                    OrigAs: 0
       NibID: 0x23000006             LastAs: 0
      AttrID: 0xffffffff
    BkAttrID: 0x1000000            Neighbor: ::
       Flags: 0x10004           OrigNextHop: FE80::6EAB:81FF:FE67:106
       Label: NULL              RealNextHop: FE80::6EAB:81FF:FE67:106
     BkLabel: NULL                BkNextHop: FE80::6EAB:8BFF:FED4:206
     SRLabel: NULL                Interface: Vlan-interface200
   BkSRLabel: NULL              BkInterface: N/A
   Tunnel ID: Invalid           IPInterface: Vlan-interface200
 BkTunnel ID: Invalid         BkIPInterface: Vlan-interface101
    LinkCost: 0                  MicroSegID: 0
 RealFIRType: Normal              RealThres: 0
     InLabel: 0               ColorInterface: N/A
    SIDIndex: NULL          BkColorInterface: N/A
    FtnIndex: 0x0            TunnelInterface: N/A
TrafficIndex: N/A          BkTunnelInterface: N/A
   Connector: N/A                     PathID: 0x0
    SID Type: N/A                        NID: Invalid
    FlushNID: Invalid                BkNID: Invalid
  BkFlushNID: Invalid             StatFlags: 0x0
         SID: N/A
       BkSID: N/A
CommBlockLen: 0                SRTunnelID: Invalid
  MemberPort: N/A                  ExtFlags: 0x0
     SliceID: 0                   BkSliceID: 0
```

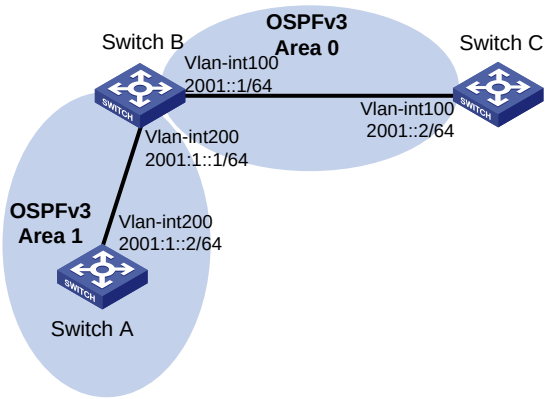
Example: Configuring OSPFv3 IPsec profile

Network configuration

As shown in Figure 14, all switches run OSPFv3, and the AS is divided into two areas.

Configure IPsec profiles on the switches to authenticate and encrypt protocol packets.

Figure 14 Network diagram



Procedure

1. Configure IPv6 addresses for interfaces. (Details not shown.)
2. Configure OSPFv3 basic features:

   # On Switch A, enable OSPFv3 and specify the router ID as 1.1.1.1.

```
<SwitchA> system-view
[SwitchA] ospfv3 1
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ospfv3 1 area 1
[SwitchA-Vlan-interface200] quit
```

   # On Switch B, enable OSPFv3 and specify the router ID as 2.2.2.2.

```
<SwitchB> system-view
[SwitchB] ospfv3 1
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ospfv3 1 area 0
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 1 area 1
[SwitchB-Vlan-interface200] quit
```

   # On Switch C, enable OSPFv3 and specify the router ID as 3.3.3.3.

```
<SwitchC> system-view
[SwitchC] ospfv3 1
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
```

```
[SwitchC-Vlan-interface100] ospfv3 1 area 0
[SwitchC-Vlan-interface100] quit
```

3. Configure OSPFv3 IPsec profiles:

- On Switch A:

  # Create an IPsec transform set named **trans**.

  ```
  [SwitchA] ipsec transform-set trans
  ```

  # Specify the encapsulation mode as **transport**.

  ```
  [SwitchA-ipsec-transform-set-trans] encapsulation-mode transport
  ```

  # Specify the ESP encryption and authentication algorithms.

  ```
  [SwitchA-ipsec-transform-set-trans] protocol esp
  [SwitchA-ipsec-transform-set-trans] esp encryption-algorithm aes-cbc-128
  [SwitchA-ipsec-transform-set-trans] esp authentication-algorithm sha1
  [SwitchA-ipsec-transform-set-trans] quit
  ```

  # Create a manual IPsec profile named **profile001**.

  ```
  [SwitchA] ipsec profile profile001 manual
  ```

  # Use IPsec transform set **trans**.

  ```
  [SwitchA-ipsec-profile-manual-profile001] transform-set trans
  ```

  # Configure the inbound and outbound SPIs for ESP.

  ```
  [SwitchA-ipsec-profile-manual-profile001] sa spi outbound esp 123456
  [SwitchA-ipsec-profile-manual-profile001] sa spi inbound esp 123456
  ```

  # Configure the inbound and outbound SA keys for ESP.

  ```
  [SwitchA-ipsec-profile-manual-profile001] sa string-key outbound esp simple
  abcdefg
  [SwitchA-ipsec-profile-manual-profile001] sa string-key inbound esp simple
  abcdefg
  [SwitchA-ipsec-profile-manual-profile001] quit
  ```

- On Switch B:

  # Create an IPsec transform set named **trans**.

  ```
  [SwitchB] ipsec transform-set trans
  ```

  # Specify the encapsulation mode as **transport**.

  ```
  [SwitchB-ipsec-transform-set-trans] encapsulation-mode transport
  ```

  # Specify the ESP encryption and authentication algorithms.

  ```
  [SwitchB-ipsec-transform-set-trans] protocol esp
  [SwitchB-ipsec-transform-set-trans] esp encryption-algorithm aes-cbc-128
  [SwitchB-ipsec-transform-set-trans] esp authentication-algorithm sha1
  [SwitchB-ipsec-transform-set-trans] quit
  ```

  # Create a manual IPsec profile named **profile001**.

  ```
  [SwitchB] ipsec profile profile001 manual
  ```

  # Use IPsec transform set **trans**.

  ```
  [SwitchB-ipsec-profile-manual-profile001] transform-set trans
  ```

  # Configure the inbound and outbound SPIs for ESP.

  ```
  [SwitchB-ipsec-profile-manual-profile001] sa spi outbound esp 123456
  [SwitchB-ipsec-profile-manual-profile001] sa spi inbound esp 123456
  ```

  # Configure the inbound and outbound SA keys for ESP.

  ```
  [SwitchB-ipsec-profile-manual-profile001] sa string-key outbound esp simple
  abcdefg
  ```

```
[SwitchB-ipsec-profile-manual-profile001] sa string-key inbound esp simple
abcdefg
[SwitchB-ipsec-profile-manual-profile001] quit
```

Create a manual IPsec profile named **profile002**.

```
[SwitchB] ipsec profile profile002 manual
```

Use IPsec transform set **trans**.

```
[SwitchB-ipsec-profile-manual-profile002] transform-set trans
```

Configure the inbound and outbound SPIs for ESP.

```
[SwitchB-ipsec-profile-manual-profile002] sa spi outbound esp 256
[SwitchB-ipsec-profile-manual-profile002] sa spi inbound esp 256
```

Configure the inbound and outbound SA keys for ESP.

```
[SwitchB-ipsec-profile-manual-profile002] sa string-key outbound esp simple
byebye
[SwitchB-ipsec-profile-manual-profile002] sa string-key inbound esp simple byebye
[SwitchB-ipsec-profile-manual-profile002] quit
```

- On Switch C:

Create an IPsec transform set named **trans**.

```
[SwitchC] ipsec transform-set trans
```

Specify the encapsulation mode as **transport**.

```
[SwitchC-ipsec-transform-set-trans] encapsulation-mode transport
```

Specify the ESP encryption and authentication algorithms.

```
[SwitchC-ipsec-transform-set-trans] protocol esp
[SwitchC-ipsec-transform-set-trans] esp encryption-algorithm aes-cbc-128
[SwitchC-ipsec-transform-set-trans] esp authentication-algorithm sha1
[SwitchC-ipsec-transform-set-trans] quit
```

Create a manual IPsec profile named **profile002**.

```
[SwitchC] ipsec profile profile002 manual
```

Use IPsec transform set **trans**.

```
[SwitchC-ipsec-profile-manual-profile002] transform-set trans
```

Configure the inbound and outbound SPIs for ESP.

```
[SwitchC-ipsec-profile-manual-profile002] sa spi outbound esp 256
[SwitchC-ipsec-profile-manual-profile002] sa spi inbound esp 256
```

Configure the inbound and outbound SA keys for ESP.

```
[SwitchC-ipsec-profile-manual-profile002] sa string-key outbound esp simple
byebye
[SwitchC-ipsec-profile-manual-profile002] sa string-key inbound esp simple byebye
[SwitchC-ipsec-profile-manual-profile002] quit
```

4. Apply the IPsec profiles to areas:

Configure Switch A.

```
[SwitchA] ospfv3 1
[SwitchA-ospfv3-1] area 1
[SwitchA-ospfv3-1-area-0.0.0.1] enable ipsec-profile profile001
[SwitchA-ospfv3-1-area-0.0.0.1] quit
[SwitchA-ospfv3-1] quit
```

Configure Switch B.

```
[SwitchB] ospfv3 1
[SwitchB-ospfv3-1] area 0
```

```
[SwitchB-ospfv3-1-area-0.0.0.0] enable ipsec-profile profile002
[SwitchB-ospfv3-1-area-0.0.0.0] quit
[SwitchB-ospfv3-1] area 1
[SwitchB-ospfv3-1-area-0.0.0.1] enable ipsec-profile profile001
[SwitchB-ospfv3-1-area-0.0.0.1] quit
[SwitchB-ospfv3-1] quit
```

Configure Switch C.

```
[SwitchC] ospfv3 1
[SwitchC-ospfv3-1] area 0
[SwitchC-ospfv3-1-area-0.0.0.0] enable ipsec-profile profile002
[SwitchC-ospfv3-1-area-0.0.0.0] quit
[SwitchC-ospfv3-1] quit
```

Verifying the configuration

Verify that OSPFv3 packets between Switches A, B, and C are protected by IPsec. (Details not shown.)

Contents

Configuring IPv6 PBR 1

About IPv6 PBR 1
IPv6 packet forwarding process 1
IPv6 PBR types 1
Policy 1
IPv6 PBR and Track 2
Restrictions and guidelines: IPv6 PBR configuration 3
IPv6 PBR tasks at a glance 3
Configuring an IPv6 policy 3
Creating an IPv6 node 3
Setting match criteria for an IPv6 node 3
Configuring actions for an IPv6 node 4
Specifying a policy for IPv6 PBR 6
Specifying an IPv6 policy for IPv6 local PBR 6
Specifying an IPv6 policy for IPv6 interface PBR 7
Specifying an IPv6 policy for IPv6 global PBR 7
Display and maintenance commands for IPv6 PBR 8
IPv6 PBR configuration examples 9
Example: Configuring packet type-based IPv6 local PBR 9
Example: Configuring packet type-based IPv6 interface PBR 10
Example: Configuring packet type-based IPv6 global PBR 12

Configuring IPv6 PBR

About IPv6 PBR

IPv6 policy-based routing (PBR) uses user-defined policies to route IPv6 packets. A policy can specify parameters for packets that match specific criteria such as ACLs. The parameters include the next hop and default next hop.

IPv6 packet forwarding process

When the device receives an IPv6 packet, the device searches the IPv6 PBR policy for a matching node to forward that packet.

- If a matching node is found and its match mode is **permit**, the device performs the following operations:
  a. Uses the next hops specified on the node to forward the packet.
  b. Searches the routing table for a route (except the default route) to forward the packet if one of the following conditions exists:
    - No next hops are specified on the node.
    - Forwarding failed based on the next hops.

  c. Uses the default next hops specified on the node to forward the packet if one of the following conditions exists:
    - No matching route was found in the routing table.
    - The routing table-based forwarding failed.

  d. Uses the default route to forward the packet if one of the following conditions exists:
    - No default next hops are specified on the node.
    - The forwarding failed based on the default next hops.
- The device perfoms routing table lookup to forward the packet in either of the following conditions:
  - No matching node is found.
  - A matching node is found, but its match mode is **deny**.

IPv6 PBR types

IPv6 PBR includes the following types:

- **Local PBR**—Guides the forwarding of locally generated packets, such as the ICMP packets generated by using the **`ping`** command.
- **Interface PBR**—Guides the forwarding of packets received on an interface only.

Policy

An IPv6 policy includes match criteria and actions to be taken on the matching packets. A policy can have one or multiple nodes as follows:

- Each node is identified by a node number. A smaller node number has a higher priority.
- A node contains **`if-match`** and **`apply`** clauses. An **`if-match`** clause specifies a match criterion, and an **`apply`** clause specifies an action.

- A node has a match mode of **permit** or **deny**.

An IPv6 policy compares packets with nodes in priority order. If a packet matches the criteria on a node, it is processed by the action on the node. If the packet does not match any criteria on the node, it goes to the next node for a match. If the packet does not match the criteria on any node, the device performs a routing table lookup for the packet.

Relationship between if-match clauses

IPv6 PBR supports only the **`if-match acl`** clause to set an ACL match criterion. On a node, you can specify only one **`if-match`** clause.

Relationship between apply clauses

You can specify multiple **`apply`** clauses for a node, but some of them might not be executed. For more information about the relationship between the **`apply`** clauses, see "Configuring actions for an IPv6 node."

Relationship between the match mode and clauses on the node

Does a packet match all the if-match **clauses on the node?**	**Match mode**	
	In permit mode	**In deny mode**
Yes	• If the node contains **`apply`** clauses, IPv6 PBR executes the **`apply`** clauses on the node. ○ If IPv6 PBR-based forwarding succeeds, IPv6 PBR does not compare the packet with the next node. ○ If IPv6 PBR-based forwarding fails, IPv6 PBR compares the packet with the next node. • If the node does not contain **`apply`** clauses, the device performs a routing table lookup for the packet.	The device performs a routing table lookup for the packet.
No	IPv6 PBR compares the packet with the next node.	IPv6 PBR compares the packet with the next node.

NOTE:

A node that has no **`if-match`** clauses matches any packet.

IPv6 PBR and Track

IPv6 PBR can work with the Track feature to dynamically adapt the availability status of an **`apply`** clause to the link status of a tracked object. The tracked object can be a next hop or default next hop.

- When the track entry associated with an object changes to **Negative**, the **`apply`** clause is invalid.
- When the track entry changes to **Positive** or **NotReady**, the **`apply`** clause is valid.

For more information about Track and IPv6 PBR collaboration, see *High Availability Configuration Guide*.

Restrictions and guidelines: IPv6 PBR configuration

If a packet destined for the local device matches an IPv6 PBR policy, IPv6 PBR will execute the apply clauses in the policy, including the clause for forwarding. When you configure an IPv6 PBR policy, be careful to avoid this situation.

IPv6 PBR tasks at a glance

To configure IPv6 PBR, perform the following tasks:

1. Configuring an IPv6 policy
   a. Creating an IPv6 node
   b. Setting match criteria for an IPv6 node
   c. Configuring actions for an IPv6 node
2. Specifying a policy for IPv6 PBR

   Choose the following tasks as needed:
   - Specifying an IPv6 policy for IPv6 local PBR
   - Specifying an IPv6 policy for IPv6 interface PBR
   - Specifying an IPv6 policy for IPv6 global PBR

Configuring an IPv6 policy

Creating an IPv6 node

1. Enter system view.

   system-view
2. Create an IPv6 policy or policy node and enter its view.

   ipv6 policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. (Optional.) Configure a description for the IPv6 policy node.

   description *text*

   By default, no description is configured for an IPv6 policy node.

Setting match criteria for an IPv6 node

1. Enter system view.

   system-view
2. Enter IPv6 policy node view.

   ipv6 policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Set match criteria.
   - Set an ACL match criterion.

     if-match acl [**user-defined**] { *ipv6-acl-number* | **name** *ipv6-acl-name* }

     By default, no ACL match criterion is set.

     The ACL match criterion cannot match Layer 2 information.

When using the ACL to match packets, IPv6 PBR ignores the action (**permit** or **deny**) and time range settings in the ACL. If the specified ACL does not exist, no packets will match this criterion.

- Set a local QoS ID match criterion.
  if-match qos-local-id *local-id-value*
  By default, no local QoS ID match criterion is set.

Configuring actions for an IPv6 node

About this task

The **apply** clauses allow you to specify actions to take on matching packets on a node.

The following **apply** clauses determine the packet forwarding paths in a descending order:

- **apply next-hop**
- **apply output-interface**
- **apply default-next-hop**

IPv6 PBR supports the **apply** clauses in Table 1.

Table 1 Apply clauses supported in IPv6 PBR

Clause	Meaning	Remarks
apply precedence	Sets an IP precedence.	This clause is always executed.
apply loadshare next-hop	Enables load sharing among multiple next hops.	Multiple next hop options operate in either primary/backup or load sharing mode. • **Primary/backup mode**—One option is selected from all options in configuration order for packet forwarding, with all remaining options as backups. For example, if multiple next hops are configured, the first configured next hop is selected. When the selected next hop fails, the next available next hop takes over. • **Load sharing mode**—Matching traffic is distributed across the available options. They load share traffic in round robin manner, starting from the first configured option. They perform per-packet load sharing for traffic that does not match any fast forwarding entry, and perform per-flow load sharing for traffic that matches a fast forwarding entry. After you configure the **apply loadshare-mode** command, the load sharing mode specified in the command applies. By default, the primary/backup mode applies. For the load sharing mode to take effect, make sure multiple next hops are set in the policy.
apply	Specifies the load sharing mode for	Specify one of the following load sharing

loadshare-modenext-hop robin	multiple next hops.	modes for multiple next hops: • **robin**: Selects load-shared links randomly based on packets.
apply next-hop and **apply output-interface**	Sets next hops and sets output interfaces.	If both clauses are configured, only the **apply next-hop** clause is executed.
apply default-next-hop	Sets default next hops.	The configuration take effect only in the following cases: • No next hops or output interfaces are set or the next hops and output interfaces are invalid. • The IPv6 packet does not match any route in the routing table.
apply statistics	Counts successful matches and matching bytes on the IPv6 policy node.	To include the number of successful matches and matching bytes on an IPv6 policy node in the statistics displayed by a **display** command, execute this command. This clause is always executed for matching packets.

Restrictions and guidelines for action configuration

If you specify a next hop, IPv6 PBR periodically performs FIB table lookup to determine its availability. Temporary service interruption might occur if IPv6 PBR does not update the route immediately after its availability status changes.

If you configure the **apply loadshare-mode** command, you cannot specify next hop addresses as VXLAN tunnel addresses.

In an IPv6 policy node, if you configure per-packet forwarding with the **apply loadshare-mode** command, and apply the policy locally with the **ip local policy-based-route** command, per-packet forwarding does not take effect. Instead, per-flow forwarding still takes effect. As a best practice to have IPv6 PBR take effect correctly, configure other IPv6 PBR types with the policy.

Setting an IP preference

1. Enter system view.
   system-view
2. Enter IPv6 policy node view.
   ipv6 policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Set an IP precedence.
   apply precedence { *type* | *value* }
   By default, no IP precedence is specified.

Configuring actions for a node

1. Enter system view.
   system-view
2. Enter IPv6 policy node view.
   ipv6 policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Configure actions for a node.
   - Set next hops for permitted IPv6 packets.
     apply next-hop [**vpn-instance** *vpn-instance-name*] { *ipv6-address* [**direct**] [**track** *track-entry-number*] } &<1-8>

By default, no next hops are specified.

You can specify multiple next hops for backup or load sharing in one command line or by executing this command multiple times. You can specify a maximum of eight next hops for a node.

If multiple next hops on the same subnet are specified for backup, the device first uses the subnet route for the next hops to forward packets when the primary next hop fails. If the subnet route is not available, the device selects a backup next hop.

- Enable load sharing among multiple next hops.

  apply loadshare next-hop

  By default, the next hops operate in primary/backup mode.
- Specify the load sharing mode for multiple next hops.

  apply loadshare-mode next-hop robin

  By default, multiple next hops load share traffic in round robin manner, starting from the first configured next hop. They perform per-packet load sharing for traffic that does not match any fast forwarding entry, and perform per-flow load sharing for traffic that matches a fast forwarding entry.
- Set output interfaces.

  apply output-interface NULL 0

  By default, no output interfaces are specified.
- Set default next hops.

  apply default-next-hop [**vpn-instance** *vpn-instance-name*] { *ipv6-address* [**direct**] [**track** *track-entry-number*] }&<1-4>

  By default, no default next hops are specified.

  You can specify multiple default next hops for backup in one command line or by executing this command multiple times. You can specify a maximum of eight default next hops for a node.

Collecting statistics

1. Enter system view.

   system-view
2. Enter IPv6 policy node view.

   ipv6 policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Count successful matches and matching bytes on the IPv6 policy node.

   apply statistics

   By default, counting successful matches is disabled on IPv6 policy nodes.

Specifying a policy for IPv6 PBR

Specifying an IPv6 policy for IPv6 local PBR

About this task

Perform this task to specify an IPv6 policy for IPv6 local PBR to guide the forwarding of locally generated packets.

Restrictions and guidelines

You can specify only one policy for IPv6 local PBR and must make sure the specified policy already exists. Before you apply a new policy, you must first remove the current policy.

IPv6 local PBR might affect local services, such as ping and Telnet. When you use IPv6 local PBR, make sure you fully understand its impact on local services of the device.

Procedure

1. Enter system view.
   system-view
2. Specify an IPv6 policy for IPv6 local PBR.
   ipv6 local policy-based-route *policy-name*
   By default, IPv6 local PBR is not enabled.

Specifying an IPv6 policy for IPv6 interface PBR

About this task

Perform this task to apply an IPv6 policy to an interface to guide the forwarding of packets received on the interface only.

Restrictions and guidelines

You can apply only one policy to an interface and must make sure the specified policy already exists. Before you apply a new policy, you must first remove the current policy from the interface.

You can apply a policy to multiple interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify an IPv6 policy for IPv6 interface PBR.
   ipv6 policy-based-route *policy-name*
   By default, no IPv6 policy is applied to the interface.

Specifying an IPv6 policy for IPv6 global PBR

About this task

Perform this task to apply an IPv6 policy to all interfaces on the device to guide the forwarding of packets received on the interfaces.

Restrictions and guidelines

You can apply only one IPv6 policy for IPv6 global PBR and the specified IPv6 policy must already exist. Before you can apply a new IPv6 policy, you must first remove the current IPv6 policy.

IPv6 interface PBR takes precedence over IPv6 global PBR on an interface. When they are both configured and packets fail to match the IPv6 interface PBR policy, IPv6 global PBR applies.

Procedure

1. Enter system view.
   system-view
2. Specify an IPv6 policy for IPv6 global PBR.
   ipv6 global policy-based-route *policy-name*
   By default, no IPv6 policy is specified for IPv6 global PBR.

Display and maintenance commands for IPv6 PBR

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display IPv6 PBR policy information.	**display ipv6 policy-based-route** [**policy** *policy-name*]
Display IPv6 global PBR configuration and statistics.	**display ipv6 policy-based-route global** [**slot** *slot-number*]
Display IPv6 interface PBR configuration and statistics.	**display ipv6 policy-based-route interface** *interface-type interface-number* [**slot** *slot-number*]
Display IPv6 local PBR configuration and statistics.	**display ipv6 policy-based-route local** [**slot** *slot-number*]
Display IPv6 PBR configuration.	**display ipv6 policy-based-route setup**
Clear IPv6 PBR statistics.	**reset ipv6 policy-based-route statistics** [**policy** *policy-name*]

IPv6 PBR configuration examples

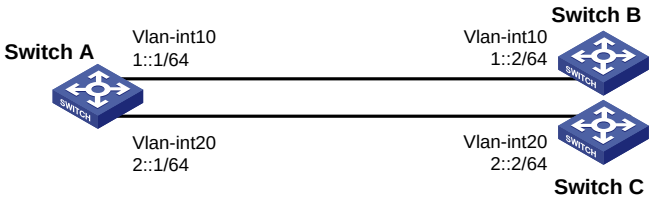
Example: Configuring packet type-based IPv6 local PBR

Network configuration

As shown in Figure 1, Switch B and Switch C are connected through Switch A. Switch B and Switch C do not have a route to reach each other.

Configure IPv6 PBR on Switch A to forward all TCP packets to the next hop 1::2 (Switch B).

Figure 1 Network diagram



Procedure

1. Configure Switch A:

   # Create VLAN 10 and VLAN 20.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 10
   [SwitchA-vlan10] quit
   [SwitchA] vlan 20
   [SwitchA-vlan20] quit
   ```

   # Configure the IPv6 addresses of VLAN-interface 10 and VLAN-interface 20.

   ```
   [SwitchA] interface vlan-interface 10
   [SwitchA-Vlan-interface10] ipv6 address 1::1 64
   [SwitchA-Vlan-interface10] quit
   [SwitchA] interface vlan-interface 20
   [SwitchA-Vlan-interface20] ipv6 address 2::1 64
   [SwitchA-Vlan-interface20] quit
   ```

   # Configure ACL 3001 to match TCP packets.

   ```
   [SwitchA] acl ipv6 advanced 3001
   [SwitchA-acl-ipv6-adv-3001] rule permit tcp
   [SwitchA-acl-ipv6-adv-3001] quit
   ```

   # Configure Node 5 for policy **aaa** to forward TCP packets to next hop 1::2.

   ```
   [SwitchA] ipv6 policy-based-route aaa permit node 5
   [SwitchA-pbr6-aaa-5] if-match acl 3001
   [SwitchA-pbr6-aaa-5] apply next-hop 1::2
   [SwitchA-pbr6-aaa-5] quit
   ```

   # Configure IPv6 local PBR by applying policy **aaa** to Switch A.

   ```
   [SwitchA] ipv6 local policy-based-route aaa
   ```

2. Configure Switch B:

   # Create VLAN 10.

   ```
   <SwitchB> system-view
   [SwitchB] vlan 10
   ```

```
[SwitchB-vlan10] quit
```

Configure the IPv6 address of VLAN-interface 10.

```
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ipv6 address 1::2 64
[SwitchB-Vlan-interface10] quit
```

3. Configure Switch C:

Create VLAN 20.

```
<SwitchC> system-view
[SwitchC] vlan 20
[SwitchC-vlan20] quit
```

Configure the IPv6 address of VLAN-interface 20.

```
[SwitchC] interface vlan-interface 20
[SwitchC-Vlan-interface20] ipv6 address 2::2 64
[SwitchC-Vlan-interface20] quit
```

Verifying the configuration

1. Perform telnet operations to verify that IPv6 local PBR on Switch A operates as configured to forward the matching TCP packets to the next hop 1::2 (Switch B), as follows:

   # Verify that you can telnet to Switch B from Switch A successfully. (Details not shown.)

   # Verify that you cannot telnet to Switch C from Switch A. (Details not shown.)

2. Verify that Switch A forwards packets other than TCP packets through VLAN-interface 20. For example, verify that you can ping Switch C from Switch A. (Details not shown.)

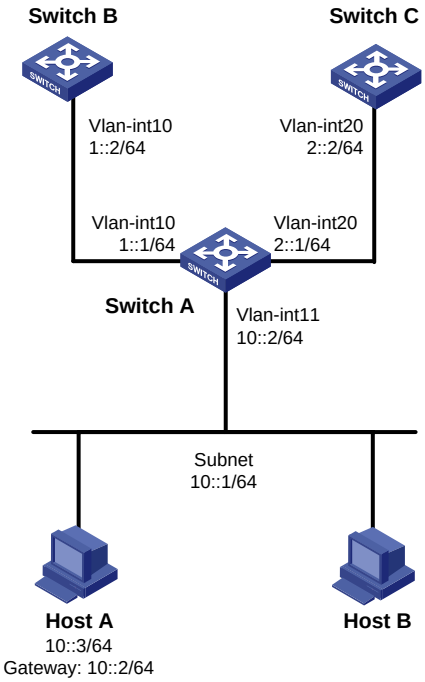
Example: Configuring packet type-based IPv6 interface PBR

Network configuration

As shown in Figure 2, Switch B and Switch C do not have a route to reach each other.

Configure IPv6 PBR on Switch A to forward all TCP packets received on VLAN-interface 11 to the next hop 1::2 (Switch B).

Figure 2 Network diagram



Procedure

1. Configure IPv6 addresses and unicast routing protocol settings to make sure that Switch B and Switch C each have a route to reach Host A. (Details not shown.)
2. Configure Switch A:

   # Configure ACL 3001 to match TCP packets.

   ```
   <SwitchA> system-view
   [SwitchA] acl ipv6 advanced 3001
   [SwitchA-acl-ipv6-adv-3001] rule permit tcp
   [SwitchA-acl-ipv6-adv-3001] quit
   ```

   # Configure Node 5 for policy **aaa** to forward TCP packets to next hop 1::2.

   ```
   [SwitchA] ipv6 policy-based-route aaa permit node 5
   [SwitchA-pbr6-aaa-5] if-match acl 3001
   [SwitchA-pbr6-aaa-5] apply next-hop 1::2
   [SwitchA-pbr6-aaa-5] quit
   ```

   # Configure IPv6 interface PBR by applying policy **aaa** to VLAN-interface 11.

   ```
   [SwitchA] interface vlan-interface 11
   [SwitchA-Vlan-interface11] ipv6 policy-based-route aaa
   [SwitchA-Vlan-interface11] quit
   ```

Verifying the configuration

1. Enable IPv6 and configure the IPv6 address 10::3 for Host A.

   ```
   C:\>ipv6 install
   Installing...
   Succeeded.
   C:\>ipv6 adu 4/10::3
   ```

2. Perform telnet operations to verify that IPv6 interface PBR on Switch A operates as configured to forward the matching TCP packets to the next hop 1::2 (Switch B), as follows:

Verify that you can telnet to Switch B from Host A successfully. (Details not shown.)

Verify that you cannot telnet to Switch C from Host A. (Details not shown.)

3. Verify that Switch A forwards packets other than TCP packets through VLAN-interface 20. For example, verify that you can ping Switch C from Host A. (Details not shown.)

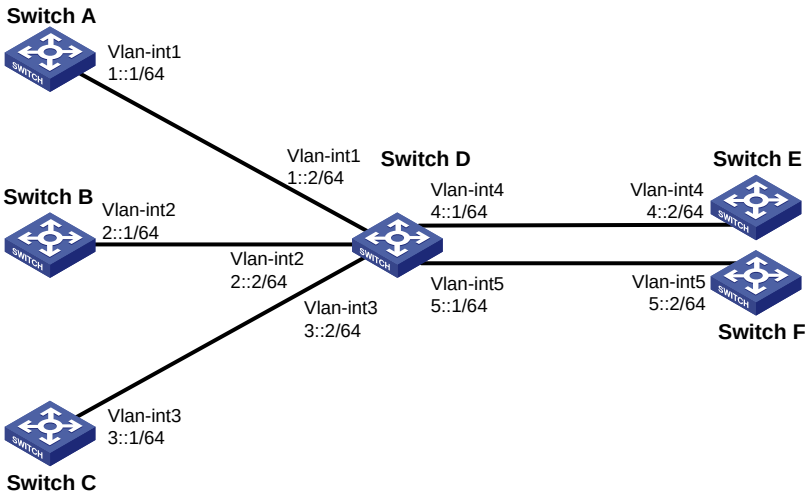
Example: Configuring packet type-based IPv6 global PBR

Network configuration

As shown in Figure 3, Switch E and Switch F do not have a route to reach each other.

Configure IPv6 global PBR on Switch D to forward TCP packets to the next hop 4::2 (Switch E).

Figure 3 Network diagram



Procedure

1. Assign IPv6 addresses to the interfaces, as shown in Figure 3. (Details not shown.)
2. Configure unicast routing protocol settings to make sure Switch A, B and C can communicate with Switch E and Switch F, respectively. (Details not shown.)
3. Configure Switch D:

   # Configure IPv6 ACL 3101 to match TCP packets sourced from networks 1::0/64, 2::0/64, and 3::0/64.

   ```
   <SwitchD> system-view
   [SwitchD] acl ipv6 advanced 3101
   [SwitchD-acl-ipv6-adv-3101] rule permit tcp source 1::0 64
   [SwitchD-acl-ipv6-adv-3101] rule permit tcp source 2::0 64
   [SwitchD-acl-ipv6-adv-3101] rule permit tcp source 3::0 64
   [SwitchD-acl-ipv6-adv-3101] quit
   ```

   # Configure node 5 in IPv6 PBR policy **aaa** to forward TCP packets that match ACL 3101 to next hop 4::2.

   ```
   [SwitchD] ipv6 policy-based-route aaa permit node 5
   [SwitchD-pbr6-aaa-5] if-match acl 3101
   [SwitchD-pbr6-aaa-5] apply next-hop 4::2
   [SwitchD-pbr6-aaa-5] quit
   ```

   # Specify IPv6 PBR policy **aaa** as the IPv6 global PBR policy.

   ```
   [SwitchD] ipv6 global policy-based-route aaa
   ```

Verifying the configuration

1. Perform telnet operations to verify that IPv6 global PBR on Switch D operates as configured to forward the matching TCP packets to the next hop 4::2 (Switch E), as follows:

   # Verify that you can telnet to Switch E from Switch A, Switch B, and Switch C successfully. (Details not shown.)

   # Verify that you cannot telnet to Switch F from Switch A, Switch B, or Switch C. (Details not shown.)

2. Verify that Switch D forwards packets other than TCP packets as long as a route is available. For example, verify that you can ping Switch F from Switch A, Switch B, and Switch C. (Details not shown.)

Contents

Configuring routing policies 1

About routing policies 1
Implementation of a routing policy 1
Filters 1
Routing policy tasks at a glance 2
Configuring an IPv4 prefix list 3
Configuring an IPv6 prefix list 3
Configuring an AS path list 4
Configuring a community list 7
Configuring an extended community list 7
Configuring an RD list 8
Configuring a routing policy 8
Creating a routing policy 8
Configuring if-match clauses 9
Configuring apply clauses 10
Configuring the continue clause 12
Configuring the routing policy change delay timer 13
Display and maintenance commands for routing policies 14
Routing policy configuration examples 14
Example: Configuring a routing policy for redistributing static routes to RIP 14
Example: Configuring a routing policy for redistributing IS-IS routes to OSPF 16
Example: Configuring a routing policy for IPv6 route redistribution 19

Configuring routing policies

About routing policies

Routing policies control routing paths by filtering and modifying routing information.

Routing policies can filter advertised, received, and redistributed routes, and modify attributes for specific routes.

Implementation of a routing policy

To configure a routing policy:

1. **Define route filters**—Configure filters based on route attributes.
2. **Apply route filters**—Create a routing policy and apply filters to the routing policy.

Filters

Routing policies can use the following filters to match routes.

ACL

An Access Control List (ACL) can match the destination or next hop of routes.

For more information about ACLs, see ACL configuration in *ACL and QoS Configuration Guide*.

IP prefix list

An IP prefix list matches the destination address of routes.

An IP prefix list can contain multiple items that specify prefix ranges. Each destination IP address prefix of a route is compared with these items in ascending order of their index numbers. A prefix matches the IP prefix list if it matches one item in the list.

AS path list

An AS path list matches the AS_PATH attribute of BGP routes. The AS_PATH attribute identifies the ASs through which a route has passed.

For more information about the AS_PATH attribute, see BGP configuration in *Layer 3—IP Routing Configuration Guide*.

Community list

A community list matches the COMMUNITY attribute of BGP routes. The COMMUNITY attribute identifies the community of BGP routes.

A route can carry one or more COMMUNITY attribute values (each of which is represented by a 4-byte integer). A router uses the COMMUNITY attribute to determine whether to advertise the route and the advertising scope without using complex filters such as ACLs. This mechanism simplifies routing policy configuration and network maintenance.

For more information about community lists, see BGP configuration in *Layer 3—IP Routing Configuration Guide*.

Extended community list

An extended community list matches the extended community attribute of BGP routes.

The device supports the link bandwidth extended community attribute, route target (also called VPN target) extended community attribute, and Site of Origin (SoO) extended community attribute. The extended community list is actually a route filter based on these attributes.

For more information about extended community lists, see *MCE Configuration Guide.*

RD list

A route distinguisher (RD) list matches the RD of routes. After you add an RD as a prefix to an IPv4 address prefix, the address prefix becomes a unique VPN-IPv4 address prefix.

An RD list is identified by an RD list number and can contain multiple items that specify RD ranges. Each item is identified by an index number. The RD of a route is compared with these items in ascending order of their index numbers. An RD matches the RD list if it matches one item in the list.

For information about RD, see MCE configuration in *MCE Configuration Guide*.

Routing policy

A routing policy can contain multiple nodes, which are in a logical OR relationship. A node with a smaller number is matched first. A route matches the routing policy if it matches one node (except the node configured with the **continue** clause) in the routing policy.

Each node has a match mode of **permit** or **deny**.

- **permit**—Specifies the **permit** match mode for a routing policy node. If a route meets all the **if-match** clauses of the node, it is handled by the **apply** clauses of the node. The route is not compared with the next node unless the **continue** clause is configured. If a route does not meet all the **if-match** clauses of the node, it is compared with the next node.
- **deny**—Specifies the **deny** match mode for a routing policy node. The **apply** and **continue** clauses of a deny node are never executed. If a route meets all the **if-match** clauses of the node, it is denied without being compared with the next node. If a route does not meet all the **if-match** clauses of the node, it is compared with the next node.

A node can contain a set of **if-match**, **apply**, and **continue** clauses.

- **if-match** clauses—Specify the match criteria that match the attributes of routes. The **if-match** clauses of different types are in a logical AND relationship and the **if-match** clauses of the same type are in a logical OR relationship. A route must meet **if-match** clauses of all types to match the node.
- **apply** clauses—Specify the actions to be taken on permitted routes, such as modifying a route attribute.
- **continue** clause—Specifies the next node. A route that matches the current node (permit node) must match the specified next node in the same routing policy. The **continue** clause combines the **if-match** and **apply** clauses of the two nodes to improve flexibility of the routing policy. After you configure a **continue** clause, a route can pass the routing policy even if it does not match the specified next node. To reject such a route, add a **deny** node without clauses.

Follow these guidelines when you configure **if-match**, **apply**, and **continue** clauses:

- If you only want to filter routes, do not configure **apply** clauses.
- If you do not configure any **if-match** clauses for a permit node, the node will permit all routes.
- Configure a permit node containing no **if-match** or **apply** clauses following multiple deny nodes to allow unmatched routes to pass.

Routing policy tasks at a glance

To configure a routing policy, perform the following tasks:

1. (Optional.) Configure filters:
   - Configuring an IPv4 prefix list
   - Configuring an IPv6 prefix list
   - Configuring an AS path list
   - Configuring a community list

- Configuring an extended community list
- Configuring an RD list

2. Configuring a routing policy:
   a. Creating a routing policy
   b. Configuring if-match clauses
   c. Configuring apply clauses
   d. Configuring the continue clause
3. (Optional.) Configuring the routing policy change delay timer

Configuring an IPv4 prefix list

About this task

An IPv4 prefix list is used to filter IPv4 addresses. It can contain multiple items, each of which specifies a range of IPv4 prefixes. The relation between the items is logical OR. A route can pass an IPv4 prefix list as long as it successfully matches an item. If the route cannot match any item, the IPv4 prefix list cannot be passed.

Restrictions and guidelines

If the *ip-address mask-length* argument is specified as **0.0.0.0 0**, only the default route will be matched. To match all routes, use **0.0.0.0 0 less-equal 32**.

If all the items are set to deny mode, no routes can pass the IPv4 prefix list. To permit IPv4 routes that do not match any of the items, you must configure the **permit 0.0.0.0 0 less-equal 32** item following multiple deny items.

Procedure

1. Enter system view.

   system-view
2. Configure an IPv4 prefix list.

   ip prefix-list *prefix-list-name* [**index** *index-number*] { **deny** | **permit** } *ip-address mask-length* [**greater-equal** *min-mask-length*] [**less-equal** *max-mask-length*]

Configuring an IPv6 prefix list

About this task

An IPv6 prefix list is used to filter IPv6 addresses. It can contain multiple items, each of which specifies a range of IPv6 prefixes. The relation between the items is logical OR. A route can pass an IPv6 prefix list as long as it successfully matches an item. If the route cannot match any item, the IPv6 prefix list cannot be passed.

Restrictions and guidelines

If the *ipv6-address prefix-length* argument is specified as :: 0, only the default route matches. To match all routes, configure :: 0 **less-equal** 128.

If all items are set to **deny** mode, no routes can pass the IPv6 prefix list. To permit IPv6 routes that do not match any of the items, you must configure the **permit** :: 0 **less-equal** 128 item following multiple **deny** items.

Procedure

1. Enter system view.

system-view

2. Configure an IPv6 prefix list.

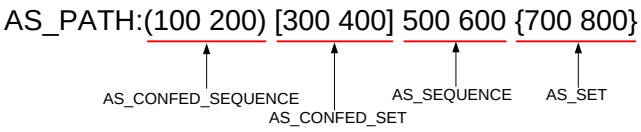
   ipv6 prefix-list *prefix-list-name* [**index** *index-number*] { **deny** | **permit** } *ipv6-address* { **inverse** *inverse-prefix-length* | *prefix-length* [**greater-equal** *min-prefix-length*] [**less-equal** *max-prefix-length*] }

Configuring an AS path list

About this task

The AS_PATH attribute identifies the ASs through which a BGP route has passed. Figure 1 shows the AS_PATH attribute format in the BGP routing table.

Figure 1 AS_PATH attribute



The AS_PATH attribute value is a character string containing digits, parentheses, brackets, braces, and spaces. It is a sequence of the following AS path segments:

- **AS_CONFED_SEQUENCE**—Ordered set of member AS numbers in the local confederation that the UPDATE message has traversed.
- **AS_CONFED_SET**—Unordered set of member AS numbers in the local confederation that the UPDATE message has traversed.
- **AS_SEQUENCE**—Ordered set of ASs a route in the UPDATE message has traversed.
- **AS_SET**—Unordered set of ASs a route in the UPDATE message has traversed.

NOTE:

The AS_PATH attribute of a BGP route might not contain all of the four AS path segments.

An AS path list consists of regular expressions used for filtering BGP routes by their AS_PATH attributes. A regular expression can contain the special characters described in Table 1.

Table 1 Special characters supported in a regular expression

Character	Description	Example
^	Matches the beginning of a line.	Match a local route if its AS_PATH attribute is null: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **^$**
$	Matches the end of a line.	Match a route if it is originated from AS 100: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **100$**
.	Matches any single character.	Match a route if it has passed AS 100 but is not originated or received from AS 100: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **. 100 .**

Character	Description	Example
*	Matches the preceding character or string zero, one, or multiple times.	Match a route if it has an AS_PATH attribute: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **.***
+	Matches the preceding character or string one or multiple times.	Match a route if its AS_PATH attribute value contains **5**: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **5+**
x|y	Matches the preceding or succeeding string.	Match a route if it is originated from AS 100 or AS 200: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **100$\|200$**
( )	Matches the string in the parentheses, usually used together with the plus sign (+) or asterisk sign (*).	Match a route if its AS_PATH attribute value contains **123**: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **(123)+**
[xyz]	Matches a single character in the brackets.	Match a route if it is received from AS 10 through AS 19: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **^1[0-9]**$
[^xyz]	Matches a single character that is not in the brackets.	Match a route if its originating AS number does not end with **2** or **4**: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **[^24]$**
{n}	Matches the preceding character *n* times. The number *n* must be a nonnegative integer.	Match a route if its AS_PATH attribute value contains two **5**: **ip as-path** { *as-path-number* | *as-path-name* } { **deny** | **permit** } 5{2}
{n,}	Matches the preceding character *n* times or more. The number *n* must be a nonnegative integer.	Match a route if its AS_PATH attribute value contains two or more consecutive **5**: **ip as-path** { *as-path-number* | *as-path-name* } { **deny** | **permit** } 5{2,}
{n,m}	Matches the preceding character *n* to *m* times or more. The numbers *n* and *m* must be nonnegative integers and *n* cannot be greater than *m*.	Match a route if its AS_PATH attribute value contains one or more consecutive **5**: **ip as-path** { *as-path-number* | *as-path-name* } { **deny** | **permit** } 5{1,2}
[a-z]	Matches a single character within the specified range for only once.	Match a route if its AS_PATH attribute value contains **0**, **1**, or **2**: **ip as-path** { *as-path-number* | *as-path-name* } { **deny** | **permit** }[0-2]

Character	Description	Example
[^a-z]	Matches a single character that is not in the specified range.	Match a route if its AS_PATH attribute value contains digits other than **0**, **1**, or **2**: **ip as-path** { *as-path-number* | *as-path-name* } { **deny** | **permit** }^ [^0-2]$
_	Matches a punctuation. The expression can begin or end with _.	Match a route if it has passed AS101 100: **ip as-path** { *as-path-number* | *as-path-name* } { **deny** | **permit** } 101_100_
\b	Matches a word that starts with the pattern following \b or ends with the pattern preceding \b.	Match a route if it is originated from AS 100: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **\b100$**
\B	Matches a word that contains the pattern but does not start or end with the pattern.	Match a route if its AS_PATH attribute value starts with **1**: **ip as-path** { *as-path-number*| *as-path-name* } { **deny** | **permit** } **^1\B**
\	Escape character. If a special character listed in this table follows \, the specific meaning of the character is removed.	N/A
\w	Same as [A-Za-z0-9_], matches a digit, letter, or underscore.	Match a route if its AS_PATH attribute value contains double digits: **ip as-path** { *as-path-number* | *as-path-name* } \w{2}
\W	Same as [^A-Za-z0-9_], matches a character that is not a digit, letter, or underscore.	N/A
\index	Matches the specified string in the parentheses followed by \ twice. Each string in the parentheses is numbered from 1 in order. The *index* argument represents the sequence number of the string you want to match. If the parentheses contain *n* strings, you can specify the sequence number of the string in the range of 1 to *n*.	Match a route if its AS_PATH attribute value contains two consecutive **1**: **ip as-path** { *as-path-number* | *as-path-name* } (1)\1

NOTE:

The regular expressions shown in Table 1 are used for illustration only. Other regular expressions might be available to achieve the same effect.

Restrictions and guidelines

You can configure multiple items for an AS path list that is identified by a number. The relationship between the items is logical OR. A route matches the AS path list if it matches one item in the list.

Procedure

1. Enter system view.

   system-view

2. Configure an AS path list.

   ip as-path { *as-path-number*| *as-path-name* } { **deny** | **permit** } *regular-expression*

Configuring a community list

About this task

You can configure multiple items for a community list that is identified by a number. The relationship between the items is logical OR. A route matches the community list if it matches one item in the list.

An advanced community list matches the community attributes of BGP routes based on regular expressions. For more information about regular expressions, see the table that describes the special characters supported in a regular expression in "Configuring an AS path list."

Procedure

1. Enter system view.

   system-view

2. Configure a community list.
   - Configure a basic community list.

     ip community-list { *basic-comm-list-num* | **basic** *basic-comm-list-name* } { **deny** | **permit** } [*community-number*&<1-32> | *aa:nn*&<1-32>] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] *
   - Configure an advanced community list.

     ip community-list { *adv-comm-list-num* | **advanced** *adv-comm-list-name* } { **deny** | **permit** } *regular-expression*

Configuring an extended community list

About this task

You can configure multiple items for an extended community list that is identified by a number. The relationship between the items is logical OR. A route matches the extended community list if it matches one item in the list.

Procedure

1. Enter system view.

   system-view

2. Configure an extended community list.

   ip extcommunity-list *ext-comm-list-number* [**index** *index-number*] { **deny** | **permit** } { **bandwidth** *link-bandwidth-value* | **rt** *route-target* | **soo** *site-of-origin* }&<1-32>

Configuring an RD list

About this task

RD lists are used to filter VPN routing information. You can configure multiple items for an RD list that is identified by a number. The relationship between the items is logical OR. A route matches the RD list if it matches one item in the list.

Procedure

1. Enter system view.

   system-view

2. Configure an RD list.

   ip rd-list *rd-list-number* [**index** *index-number*] { **deny** | **permit** } *route-distinguisher*&<1-10>

Configuring a routing policy

Creating a routing policy

About this task

A routing policy must have a minimum of one permit node. If all the nodes are in **deny** mode, no routes can pass the routing policy.

Procedure

1. Enter system view.

   system-view

2. Create a routing policy and a node, and enter routing policy node view.

   route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*

3. (Optional.) Enable consistency check between the IP version in routing information and the IP version of the prefix list specified by an **if-match** clause. Routing information with a different IP version does not match the clause.

   a. Return to system view.

      quit

   b. Enable consistency check between the IP version in routing information and the IP version of the prefix list specified by an **if-match** clause. Routing information with a different IP version does not match the clause.

      route-policy prefix-protocol-check

      By default, routing policies do not check the consistency between the IP version in routing information and the IP version of the prefix list specified by an if-match clause:

      - If the if-match clause uses an IPv6 address prefix list, all IPv4 routes can match the clause.
      - If the if-match clause uses an IPv4 address prefix list, all IPv6 routes can match the clause.

Configuring if-match clauses

About this task

You can either specify no **if-match** clauses or multiple **if-match** clauses for a routing policy node. If no **if-match** clause is specified for a permit node, all routes can pass the node. If no **if-match** clause is specified for a deny node, no routes can pass the node.

Restrictions and guidelines

When you configure **if-match** clauses, follow these restrictions and guidelines:

- The **if-match** clauses of a routing policy node have a logical AND relationship. A route must meet all **if-match** clauses before it can be executed by the **apply** clauses of the node. If an **if-match** command exceeds the maximum length, multiple **if-match** clauses of the same type are generated. These clauses have a logical OR relationship. A route only needs to meet one of them.
- All IPv4 routes match a node if the **if-match** clauses of the node use only IPv6 ACLs. All IPv6 routes match a node if the **if-match** clauses of the node use only IPv4 ACLs.
- If the ACL used by an **if-match** clause does not exist or the ACL is not configured with any rules, the clause is always matched. When the ACL of an **if-match** clause is configured with rules, a route cannot match the **if-match** clause in one of the following scenarios:
  - The route cannot match any rule of the specified ACL.
  - The route matches a deny-mode rule of the specified ACL.
- If the prefix list, AS path list, community list, extended community list, MAC list, or RD list used by an **if-match** clause does not exist, the clause is always matched. The clause is not matched if one of the following conditions exists:
  - No rules of the specified prefix list, AS path list, community list, extended community list, MAC list, or RD list are matched.
  - The specified prefix list, AS path list, community list, extended community list, MAC list, or RD list is in **deny** mode.
- The **rule** [*rule-id*] **permit ip source** *sour-addr sour-wildcard* **destination** *dest-addr dest-wildcard* command can create a rule for an advanced ACL to match specific routes. The **source** *sour-addr sour-wildcard* option represents the destination network address and the **destination** *dest-addr dest-wildcard* option represents the mask in dotted decimal notation. For example, to match route 3.3.3.0/24, the rule configuration should be **rule** [*rule-id*] **permit ip source** 3.3.3.0 0.0.0.255 **destination** 255.255.255.0 0.0.0.255.

Procedure

1. Enter system view.

   system-view

2. Enter routing policy node view.

   route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*

3. Match routes whose destination, next hop, or source address matches an ACL or prefix list.

   IPv4:

   if-match ip { **address** | **next-hop** | **route-source** } { **acl** *ipv4-acl-number* | **prefix-list** *prefix-list-name* }

   IPv6:

   if-match ipv6 { **address** | **next-hop** | **route-source** } { **acl** *ipv6-acl-number* | **prefix-list** *prefix-list-name* }

   By default, no ACL or prefix list match criterion is configured.

   The ACL specified in an **if-match** clause must be a non-VPN ACL.

4. Configure BGP route match criteria.
   - Match BGP routes whose AS_PATH attribute matches a specified AS path list.
     if-match as-path { *as-path-number*&<1-32> | *as-path-name* }
   - Match BGP routes whose COMMUNITY attribute matches a specified community list.
     if-match community { { *basic-community-list-number* | **name** *comm-list-name* } [**whole-match**] | *adv-community-list-number* }&<1-32>
   - Match BGP routes whose extended community attribute matches a specified extended community list.
     if-match extcommunity *ext-comm-list-number*&<1-32>
   - Match BGP routes having the specified local preference.
     if-match local-preference *preference*
   - Match BGP routes with the specified RPKI validation state.
     if-match rpki { **invalid** | **not-found** | **valid** }

   By default, no BGP route match criteria are configured.
5. Configure route match criteria.
   - Match routes having the specified cost.
     if-match cost *cost-value*
   - Match routes having the specified output interface.
     if-match interface { *interface-type interface-number* }&<1-16>

     This command is not supported by BGP.
   - Match routes having the specified route type.
     if-match route-type { **external-type1** | **external-type1or2** | **external-type2** | **internal** | **is-is-level-1** | **is-is-level-2** | **nssa-external-type1** | **nssa-external-type1or2** | **nssa-external-type2** } *
   - Match IGP routes having the specified tag value.
     if-match tag *tag-value*

   By default, no route match criteria are configured.
6. Match routes whose RD matches the specified RD list.
   if-match rd-list *rd-list-number*

   By default, no RD match criterion is configured.
7. Match IPv6 route prefixes of the specified length whose last bit is an odd or even number.
   - Match IPv6 route prefixes of the specified length whose last bit is an odd number.
     if-match ipv6 odd-prefix-length *prefix-length-number*

     By default, no match criterion is configured to match the length and last bit of IPv6 route prefixes.
   - Match IPv6 route prefixes of the specified length whose last bit is an even number.
     if-match ipv6 even-prefix-length *prefix-length-number*

     By default, no match criterion is configured to match the length and last bit of IPv6 route prefixes.

Configuring apply clauses

1. Enter system view.

system-view

2. Enter routing policy node view.

   route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*

3. Configure BGP route attributes.

   - Set the AS_PATH attribute for BGP routes.

     apply as-path { *as-number*&<1-32> [**delete** | **replace**] | **clear** }

   - Delete the specified COMMUNITY attribute for BGP routes.

     apply comm-list { *comm-list-number* | *comm-list-name* } **delete**

     By default, no COMMUNITY attribute is deleted for BGP routes.

   - Set the specified COMMUNITY attribute for BGP routes.

     apply community { **none** | **additive** | { *community-number*&<1-32> | *aa:nn*&<1-32> | **internet** | **no-advertise** | **no-export** | **no-export-subconfed** } * [**additive**] }

   - Set the link bandwidth extended community attribute for BGP routes.

     apply extcommunity bandwidth *link-bandwidth-value* [**additive**]

   - Set the RT extended community attribute for BGP routes.

     apply extcommunity { **rt** *route-target* }&<1-32> [**additive**]

   - Set the SoO extended community attribute for BGP routes.

     apply extcommunity soo *site-of-origin*&<1-32> [**additive**]

   - Set a local preference for BGP routes.

     apply local-preference *preference*

   - Set the ORIGIN attribute for BGP routes.

     apply origin { **egp** *as-number* | **igp** | **incomplete** }

   - Set a preferred value for BGP routes.

     apply preferred-value *preferred-value*

   - Set a traffic index for BGP routes.

     apply traffic-index { *value* | **clear** }

   By default, no BGP route attributes are configured.

4. Configure the route cost and cost type.

   - Set a cost for routes.

     apply cost [**+** | **-**] *cost-value*

     By default, no cost is set for routes.

   - Set a cost type for routes.

     apply cost-type { **external** | **inherit-link-cost** | **internal** | **internal-inc-ibgp** | **med-plus-igp** | **type-1** | **type-2** }

     By default, no cost type is set for routes.

5. Set the next hop for routes.

   IPv4:

   apply ip-address next-hop *ip-address* [**public** | **vpn-instance** *vpn-instance-name*]

   IPv6:

   apply ipv6 next-hop *ipv6-address*

   By default, no next hop is set for routes.

   The configuration does not apply to redistributed routes.

6. Configure route priorities.
   - Set an IP precedence for matching routes.
     apply ip-precedence { *value* | **clear** }
     By default, no IP precedence is set.
   - Set a preference.
     apply preference *preference*
     By default, no preference is set.
   - Set a prefix priority.
     apply prefix-priority { **critical** | **high** | **medium** }
     By default, the prefix priority is low.
7. Associate a Track entry with BGP routes.
   apply track *track-entry-number*
   By default, no track entry is associated with BGP routes.
8. Redistribute routes to the specified IS-IS level.
   apply isis { **level-1** | **level-1-2** | **level-2** }
   By default, routes are not redistributed into the specified IS-IS level.
9. Set a tag value for IGP routes.
   apply tag *tag-value*
   By default, no tag value is set for IGP routes.
10. Set a backup link for fast reroute (FRR).
    IPv4:
    apply fast-reroute { **backup-interface** *interface-type interface-number* [**backup-nexthop** *ip-address*] | **backup-nexthop** *ip-address* }
    IPv6:
    apply ipv6 fast-reroute { **backup-interface** *interface-type interface-number* [**backup-nexthop** *ipv6-address*] | **backup-nexthop** *ipv6-address* }
    By default, no backup link is set for FRR.

Configuring the continue clause

Restrictions and guidelines

When you configure the **continue** clause to combine multiple nodes, follow these restrictions and guidelines:

- If you configure an **apply** clause that sets different attribute values on all the nodes, the **apply** clause of the node configured most recently takes effect.
- If you configure the following **apply** clauses on all the nodes, the **apply** clause of each node takes effect:
  - **apply as-path** without the **replace** or **clear** keyword.
  - **apply cost** with the **+** or **–** keyword.
  - **apply community** with the **additive** keyword.
  - **apply extcommunity** with the **additive** keyword.
- The **apply comm-list delete** clause configured on the current node cannot delete the community attributes set by the **apply community** clauses of the preceding nodes.

- The **apply as-path delete** or **apply as-path clear** clause configured on the current node will delete or clear the AS numbers added or replaced by the **apply as-path** clauses of the preceding nodes.

Procedure

1. Enter system view.

   system-view
2. Enter routing policy node view.

   route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*
3. Specify the next node to be matched.

   continue [*node-number*]

   By default, no continue clause is configured.

   The specified next node must have a larger number than the current node.

Configuring the routing policy change delay timer

About this task

This feature makes a routing policy take effect after a delayed time interval, which prevents incomplete routing policy configuration from being issued to cause incorrect route advertisement. The system automatically starts the timer when a routing policy changes. The changes will not take effect on the policy until the change delay timer expires.

A routing policy changes when one of the following events occurs:

- A routing policy is created.
- A routing policy node, **if-match** clause, or **apply** clause is added, modified, or deleted for a routing policy.
- You add, modify, or delete one of the following lists:
  - IPv4 prefix list.
  - IPv6 prefix list.
  - AS path list.
  - Community list.
  - Extended community list.
- The ACL used by an **if-match** clause changes.

Procedure

1. Enter system view.

   system-view
2. Set the routing policy change delay timer.

   route-policy-change delay-time { *time-value* | **unlimited** }

   By default, routing policy changes immediately take effect, but the routing protocol waits five seconds before processing routes from the new routing policy.

   When the delay timer expires, the routing protocol waits five seconds before processing routes from the new routing policy.

Display and maintenance commands for routing policies

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display BGP AS path list information.	**display ip as-path** [*as-path-number* | *as-path-name*]
Display BGP community list information.	**display ip community-list** [*basic-community-list-number* | *adv-community-list-number* | **name** *comm-list-name*]
Display BGP extended community list information.	**display ip extcommunity-list** [*ext-comm-list-number*]
Display IPv4 prefix list statistics.	**display ip prefix-list** [**name** *prefix-list-name*]
Display RD list information.	**display ip rd-list** [*rd-list-number*]
Display IPv6 prefix list statistics.	**display ipv6 prefix-list** [**name** *prefix-list-name*]
Display routing policy information.	**display route-policy** [**name** *route-policy-name*]
Clear BGP AS path list statistics.	**reset ip as-path** [*as-path-number* | *as-path-name*]
Clear IPv4 prefix list statistics.	**reset ip prefix-list** [*prefix-list-name*]
Clear IPv6 prefix list statistics.	**reset ipv6 prefix-list** [*prefix-list-name*]

Routing policy configuration examples

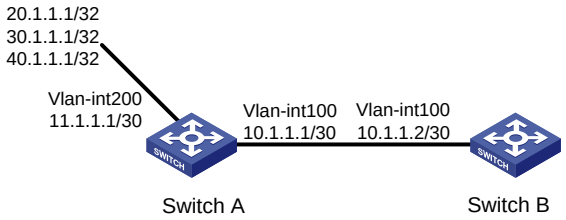
Example: Configuring a routing policy for redistributing static routes to RIP

Network configuration

As shown in Figure 2, Switch A exchanges routing information with Switch B by using RIP.

On Switch A, configure three static routes. Use a routing policy to configure Switch B to redistribute networks 20.1.1.1/32 and 40.1.1.1/32 and block network 30.1.1.1/32.

Figure 2 Network diagram



Procedure

1. Configure Switch A:

Configure IP addresses for interfaces VLAN-interface 100 and VLAN-interface 200.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 100
[SwitchA-vlan-interface100] ip address 10.1.1.1 30
[SwitchA-vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-vlan-interface200] ip address 11.1.1.1 30
[SwitchA-vlan-interface200] quit
```

Enable RIP on interface VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-vlan-interface100] rip 1 enable
[SwitchA-vlan-interface100] quit
```

Configure three static routes and set the next hop of the three routes to 11.1.1.2.

```
[SwitchA] ip route-static 20.1.1.1 32 11.1.1.2
[SwitchA] ip route-static 30.1.1.1 32 11.1.1.2
[SwitchA] ip route-static 40.1.1.1 32 11.1.1.2
```

Configure a routing policy.

```
[SwitchA] ip prefix-list a index 10 permit 30.1.1.1 32
[SwitchA] route-policy static2rip deny node 0
[SwitchA-route-policy-static2rip-0] if-match ip address prefix-list a
[SwitchA-route-policy-static2rip-0] quit
[SwitchA] route-policy static2rip permit node 10
[SwitchA-route-policy-static2rip-10] quit
```

Enable RIP and apply routing policy **static2rip** to filter redistributed static routes.

```
[SwitchA] rip
[SwitchA-rip-1] import-route static route-policy static2rip
```

2. Configure Switch B:

Configure an IP address for interface VLAN-interface 100.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 100
[SwitchB-vlan-interface100] ip address 10.1.1.2 30
```

Enable RIP.

```
[SwitchB] rip
[SwitchB-rip-1] quit
```

Enable RIP on the interface.

```
[SwitchB] interface vlan-interface 100
```

```
[SwitchB-vlan-interface100] rip 1 enable
[SwitchB-vlan-interface100] quit
```

Verifying the configuration

Display the routing table information on Switch B.

```
<SwitchB> display ip routing-table

Destinations : 14       Routes : 14

Destination/Mask    Proto   Pre Cost        NextHop         Interface
0.0.0.0/32          Direct  0   0           127.0.0.1       InLoop0
10.1.1.0/30         Direct  0   0           10.1.1.2        Vlan100
10.1.1.0/32         Direct  0   0           10.1.1.2        Vlan100
10.1.1.2/32         Direct  0   0           127.0.0.1       InLoop0
10.1.1.3/32         Direct  0   0           10.1.1.2        Vlan100
20.0.0.0/8          RIP     100 1           10.1.1.1        Vlan100
40.0.0.0/8          RIP     100 1           10.1.1.1        Vlan100
127.0.0.0/8         Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32        Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4         Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24        Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32  Direct  0   0           127.0.0.1       InLoop
```

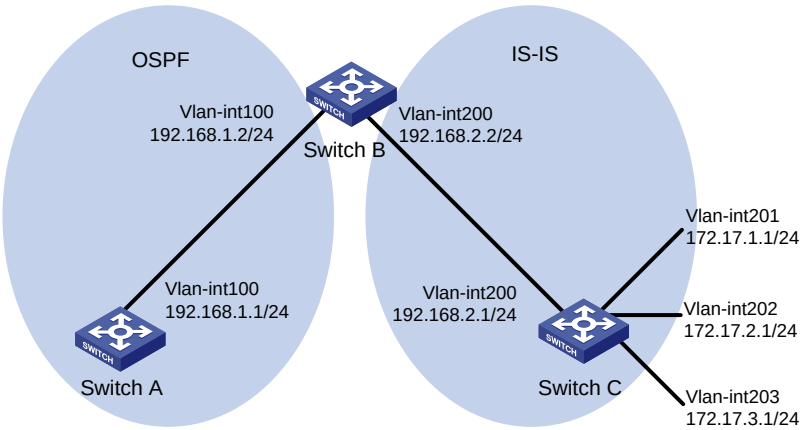
Example: Configuring a routing policy for redistributing IS-IS routes to OSPF

Network configuration

As shown in Figure 3, Switch B exchanges routing information with Switch A by using OSPF and with Switch C by using IS-IS.

On Switch B, enable route redistribution from IS-IS to OSPF. Use a routing policy to set the cost of route 172.17.1.0/24 to 100 and the tag of route 172.17.2.0/24 to 20.

Figure 3 Network diagram



Procedure

1. Specify IP addresses for interfaces. (Details not shown.)
2. Configure IS-IS:

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis
[SwitchC-isis-1] is-level level-2
[SwitchC-isis-1] network-entity 10.0000.0000.0001.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 201
[SwitchC-Vlan-interface201] isis enable
[SwitchC-Vlan-interface201] quit
[SwitchC] interface vlan-interface 202
[SwitchC-Vlan-interface202] isis enable
[SwitchC-Vlan-interface202] quit
[SwitchC] interface vlan-interface 203
[SwitchC-Vlan-interface203] isis enable
[SwitchC-Vlan-interface203] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis
[SwitchB-isis-1] is-level level-2
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable
[SwitchB-Vlan-interface200] quit
```

3. Configure OSPF and route redistribution:

   # Configure OSPF on Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

   # On Switch B, configure OSPF and enable route redistribution from IS-IS to OSPF.

```
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] import-route isis 1
[SwitchB-ospf-1] quit
```

   # Display the OSPF routing table on Switch A to view redistributed routes.

```
[SwitchA] display ospf routing
```

```
          OSPF Process 1 with Router ID 192.168.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop        AdvRouter     Area
 192.168.1.0/24     1        Stub    192.168.1.1    192.168.1.1   0.0.0.0

 Routing for ASEs
 Destination        Cost     Type    Tag        NextHop        AdvRouter
 172.17.1.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 172.17.2.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 172.17.3.0/24      1        Type2   1          192.168.1.2    192.168.2.2

 Total Nets: 4
 Intra Area: 1  Inter Area: 0  ASE: 3  NSSA: 0
```

4. Configure filtering lists:

Configure IPv4 basic ACL 2002 to permit route 172.17.2.0/24.

```
[SwitchB] acl basic 2002
[SwitchB-acl-ipv4-basic-2002] rule permit source 172.17.2.0 0.0.0.255
[SwitchB-acl-ipv4-basic-2002] quit
```

Configure IP prefix list **prefix-a** to permit route 172.17.1.0/24.

```
[SwitchB] ip prefix-list prefix-a index 10 permit 172.17.1.0 24
```

5. Configure a routing policy.

```
[SwitchB] route-policy isis2ospf permit node 10
[SwitchB-route-policy-isis2ospf-10] if-match ip address prefix-list prefix-a
[SwitchB-route-policy-isis2ospf-10] apply cost 100
[SwitchB-route-policy-isis2ospf-10] quit
[SwitchB] route-policy isis2ospf permit node 20
[SwitchB-route-policy-isis2ospf-20] if-match ip address acl 2002
[SwitchB-route-policy-isis2ospf-20] apply tag 20
[SwitchB-route-policy-isis2ospf-20] quit
[SwitchB] route-policy isis2ospf permit node 30
[SwitchB-route-policy-isis2ospf-30] quit
```

6. Apply the routing policy to route redistribution:

On Switch B, enable route redistribution from IS-IS to OSPF and apply the routing policy.

```
[SwitchB] ospf
[SwitchB-ospf-1] import-route isis 1 route-policy isis2ospf
[SwitchB-ospf-1] quit
```

Display the OSPF routing table on Switch A.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter     Area
 192.168.1.0/24     1        Transit 192.168.1.1     192.168.1.1   0.0.0.0
```

```
Routing for ASEs
Destination        Cost     Type    Tag         NextHop       AdvRouter
172.17.1.0/24      100      Type2   1           192.168.1.2   192.168.2.2
172.17.2.0/24      1        Type2   20          192.168.1.2   192.168.2.2
172.17.3.0/24      1        Type2   1           192.168.1.2   192.168.2.2

Total Nets: 4
Intra Area: 1  Inter Area: 0  ASE: 3  NSSA: 0
```

The output shows that the cost of route 172.17.1.0/24 is 100 and the tag of route 172.17.2.0/24 is 20.

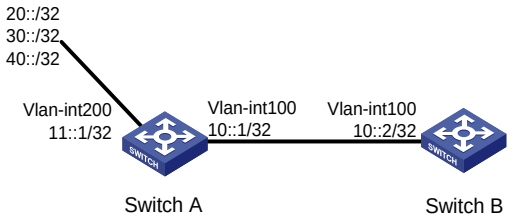
Example: Configuring a routing policy for IPv6 route redistribution

Network configuration

As shown in Figure 4:

- Run RIPng on Switch A and Switch B.
- Configure three static routes on Switch A.
- On Switch A, apply a routing policy to redistribute static routes 20::/32 and 40::/32 and deny route 30::/32.

Figure 4 Network diagram



Procedure

1. Configure Switch A:

   # Configure IPv6 addresses for VLAN-interface 100 and VLAN-interface 200.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipv6 address 10::1 32
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ipv6 address 11::1 32
[SwitchA-Vlan-interface200] quit
```

   # Enable RIPng on VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ripng 1 enable
[SwitchA-Vlan-interface100] quit
```

   # Configure three static routes with next hop 11::2, and make sure the static routes are active.

```
[SwitchA] ipv6 route-static 20:: 32 11::2
[SwitchA] ipv6 route-static 30:: 32 11::2
```

```
[SwitchA] ipv6 route-static 40:: 32 11::2
```

Configure a routing policy.

```
[SwitchA] ipv6 prefix-list a index 10 permit 30:: 32
[SwitchA] route-policy static2ripng deny node 0
[SwitchA-route-policy-static2ripng-0] if-match ipv6 address prefix-list a
[SwitchA-route-policy-static2ripng-0] quit
[SwitchA] route-policy static2ripng permit node 10
[SwitchA-route-policy-static2ripng-10] quit
```

Enable RIPng and apply the routing policy to static route redistribution.

```
[SwitchA] ripng
[SwitchA-ripng-1] import-route static route-policy static2ripng
```

2. Configure Switch B:

Configure the IPv6 address for VLAN-interface 100.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ipv6 address 10::2 32
```

Enable RIPng.

```
[SwitchB] ripng
[SwitchB-ripng-1] quit
```

Enable RIPng on VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit
```

Verifying the configuration

Display the RIPng routing table on Switch B.

```
[SwitchB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect, D - Direct
                O - Optimal, F - Flush to RIB
 ------------------------------------------------------------------

 Peer FE80::7D58:0:CA03:1 on Vlan-interface 100
 Destination 20::/32,
     via FE80::7D58:0:CA03:1, cost 1, tag 0, A, 8 secs
 Destination 40::/32,
     via FE80::7D58:0:CA03:1, cost 1, tag 0, A, 3 secs
 Local route
 Destination 10::/32,
     via ::, cost 0, tag 0, DOF
```

Contents

Configuring DCN 1

About DCN 1
Basic concepts 1
GNE 1
NE search 1
Automatic report 1
DCN tasks at a glance 2
Enabling DCN 2
Configuring the NE ID and NE IP 2
Configuring DCN VPN 3
Enabling the automatic report feature 3
Configuring the LLDP features 4
Configuring the source MAC address of LLDP frames 4
Advertising the LLDP management address 4
Enabling the device to generate ARP entries for received management address LLDP TLVs 5
Display and maintenance commands for DCN 5
DCN configuration examples 6
Example: Configuring DCN 6

Configuring DCN

About DCN

Data communication network (DCN) is built for the network management system (NMS) to implement operation, administration, and maintenance (OAM) on the network elements (NEs).

On large-scaled networks with DCN configured, the NMS remotely manages and controls all NEs through the gateway network element (GNE), which reduces operation and maintenance costs.

On the DCN network, all NEs must operate in Area 0 of OSPF process 65535.

Basic concepts

NE

A network device managed by the NMS on the DCN network.

GNE

The NE that is directly connected to the NMS. The NSM communicates with the GNE at the network layer or application layer and manages the NEs through the GNE.

NE table

Stores the NE ID-to-NE IP mappings for each NE in the DCN domain. After DCN is enabled, an NE encapsulates its NE ID and NE IP in a Type-10 Opaque LSA. It floods the LSA to synchronize the mappings with other NEs in the DCN domain.

The NE IP in the Type-10 Opaque LSA might be inconsistent with the NE's router ID when DCN is enabled on an OSPF network. Devices from other vendors will fail to identify the NE. To avoid this problem, make sure the following requirements are met:

- Change the router ID by modifying the NE IP. Do not change the default IP address of the DCN's loopback interface.
- Do not associate other loopback interfaces with the DCN VPN instance.

GNE

The GNE allows the NMS to access NEs on the DCN network. The GNE provides the NE ID-to-NE IP mappings of all NEs for the Telnet users to access the target NE remotely.

NE search

The NE search feature allows the NMS to search for all NEs that are connected to the GNE. With this feature, the GNE can report information about all NEs in its NE table and the GNE itself to the NMS.

Automatic report

The automatic report feature enables the GNE to automatically report online or offline events of NEs to the NMS.

DCN tasks at a glance

To configure DCN, perform the following tasks:

1. Enabling DCN
2. (Optional.) Configuring the NE ID and NE IP
3. Configuring DCN VPN
4. (Optional.) Enabling the automatic report feature
5. Configuring the LLDP features
   a. (Optional.) Configuring the source MAC address of LLDP frames
   b. Advertising the LLDP management address
   c. Enabling the device to generate ARP entries for received management address LLDP TLVs

Enabling DCN

About this task

After DCN is enabled, the device assigns an NE IP to the loopback interface with the largest interface number and uses the interface for communication. The NE IP is specified as the management address and advertised in an LLDP frame. For more information about LLDP, see LLDP configuration in *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Enable DCN and enter DCN view.
   `dcn`
   By default, DCN is disabled.

Configuring the NE ID and NE IP

About this task

Typically an NE automatically generates an NE IP according to its NE ID. The automatically generated NE IP changes when the NE ID changes. When you replan your network or an NE ID conflict occurs, you can manually configure the NE ID and NE IP. The manually configured NE IP does not change when the NE ID changes.

Procedure

1. Enter system view.
   `system-view`
2. Enter DCN view.
   `dcn`
3. Configure the NE ID.
   `ne-id` *`id-number`*
   By default, the NE ID is automatically generated by using the 24 low-order bits of the bridge MAC address.
4. Configure the NE IP.
   `ne-ip` *`ip-address`* { *`mask-length`* | *`mask`* }

By default, the NE IP is automatically generated by using 129 as the eight high-order bits and the NE ID as the rest 16 bits. The mask length is 32.

Configuring DCN VPN

About this task

Perform this task to ensure the independence of DCN services. For more information about VPN instances, see MCE configuration in *MCE Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Create a VPN instance and enter its view, or enter the view of an existing VPN instance.
   ip vpn-instance *vpn-instance-name*
   For more information about this command, see MCE commands in *MCE Command Reference*.
3. Quit VPN instance view.
   quit
4. Create a loopback interface and enter its view, or enter the view of an existing loopback interface.
   interface loopback *interface-number*
5. Associate the loopback interface with a VPN instance.
   ip binding vpn-instance *vpn-instance-name*
   By default, the loopback interface is not associated with a VPN instance.
   For more information about this command, see MCE commands in *MCE Command Reference*.

Enabling the automatic report feature

Restrictions and guidelines

Perform this task only on GNE to enable it to automatically report online or offline events of NEs to the NMS.

Procedure

1. Enter system view.
   system-view
2. Enter DCN view.
   dcn
3. Enable the automatic report feature.
   auto-report
   By default, the automatic report feature is disabled.

Configuring the LLDP features

Configuring the source MAC address of LLDP frames

About this task

Perform this task to configure the MAC address of the specified Layer 3 Ethernet interface as the source MAC address of LLDP frames. For more information about LLDP, see LLDP configuration in *Layer 2—LAN Switching Configuration Guide*.

Restrictions and guidelines

This feature is supported only on Layer 3 physical interfaces.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 3 Ethernet interface view.

   interface *interface-type interface-number*

3. Configure the MAC address of the Dot1q termination-enabled specified Layer 3 Ethernet interface as the source MAC address of LLDP frames.

   lldp source-mac vlan *vlan-id*

   By default, the MAC address of the current Ethernet interface is specified as the source MAC address of LLDP frames.

   For more information about this command, see LLDP commands in *Layer 2—LAN Switching Command Reference*.

Advertising the LLDP management address

About this task

An NE on the DCN network learns the MAC address of its neighbor through LLDP. Perform this task to enable an NE to advertise the IP address of the specified loopback interface (management address) for ARP entry learning. For more information about LLDP, see LLDP configuration in *Layer 2—LAN Switching Configuration Guide*.

Restrictions and guidelines

This feature is supported only on Layer 3 physical interfaces.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Advertise the LLDP management address.

   lldp tlv-enable basic-tlv management-address-tlv [*ip-address* | **interface loopback** *interface-number*]

   By default, the LLDP management address is not advertised.

   For more information about this command, see LLDP commands in *Layer 2—LAN Switching Command Reference*.

Enabling the device to generate ARP entries for received management address LLDP TLVs

About this task

This feature enables the device to generate an ARP entry for a received LLDP frame that meets the following requirements:

- The LLDP frame carries a management address TLV.
- The management address TLV contains an IPv4 address.

The generated ARP entry contains the management address and the source MAC address of the frame. For more information about LLDP, see LLDP configuration in *Layer 2—LAN Switching Configuration Guide*.

Restrictions and guidelines

This feature is supported only on Layer 3 physical interfaces.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable the device to generate an ARP entry when receiving a management address LLDP TLV.

   lldp management-address arp-learning [**vlan** *vlan-id*]

   By default, the device does not generate an ARP entry when receiving a management address LLDP TLV.

   The device uses the current Layer 3 Ethernet subinterface as the output interface for generated ARP entries.

   For more information about this command, see LLDP commands in *Layer 2—LAN Switching Configuration Guide*.

Display and maintenance commands for DCN

Execute the **display** commands in any view.

Task	Command
Display the brief DCN information.	**display dcn**
Display information about all online NEs on the DCN network.	**display dcn ne-info**

DCN configuration examples

Example: Configuring DCN

Network configuration

As shown in Figure 1, the GNE, Device A, and Device B run DCN in the same VPN instance. The NMS uses SNMP to manage the GNE, and the GNE automatically sends notifications to the NMS to report online or offline events of NEs.

Figure 1 Network diagram

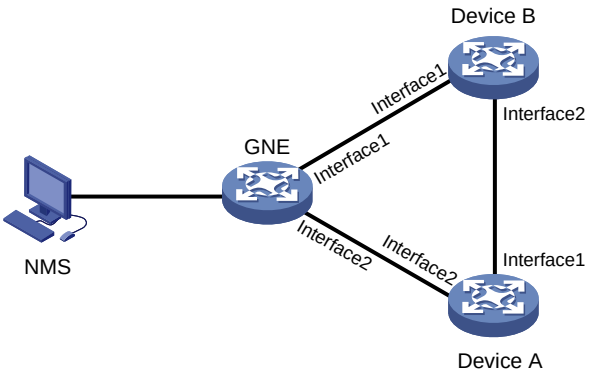


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure the GNE:

   # Enable SNMP on the GNE. (Details not shown. For more information, see *Network Management and Monitoring Configuration Guide*.)

   # Enable DCN, configure the NE ID as 100001 and NE IP as 11.1.1.1/32, and enable the automatic report feature.

   ```
   <GNE> system-view
   [GNE] dcn
   [GNE-dcn] ne-id 100001
   [GNE-dcn] ne-ip 11.1.1.1 32
   [GNE-dcn] auto-report
   [GNE-dcn] quit
   ```

   # Create a VPN instance named **dcn_vpn**.

   ```
   [GNE] ip vpn-instance dcn_vpn
   [GNE-vpn-instance-dcn_vpn] quit
   ```

   # Create interface Loopback 1023, and associate it with VPN instance **dcn_vpn**.

   ```
   [GNE] interface loopback 1023
   [GNE-LoopBack1023] ip binding vpn-instance dcn_vpn
   [GNE-LoopBack1023] quit
   ```

   # Enable LLDP globally.

```
[GNE] lldp global enable
```

Enable the nearest bridge agents on GigabitEthernet 1/0/1 to advertise basic LLDP TLVs and management address TLVs. The IP address of interface Loopback 1023 is specified as the management address.

```
[GNE] interface gigabitethernet 1/0/1
[GNE-GigabitEthernet1/0/1] port access vlan 3001
[GNE-GigabitEthernet1/0/1] lldp source-mac vlan 3001
[GNE-GigabitEthernet1/0/1] lldp tlv-enable basic-tlv management-address-tlv
interface loopback 1023
```

Configure the system to issue the generated ARP entry to the Layer 3 Ethernet subinterface associated with VLAN 3001 in Dot1q termination after GigabitEthernet 1/0/1 receives an LLDP frame.

```
[GNE-GigabitEthernet1/0/1] lldp management-address arp-learning vlan 3001
[GNE-GigabitEthernet1/0/1] quit
```

Enable the nearest bridge agents on GigabitEthernet 1/0/2 to advertise basic LLDP TLVs and management address TLVs. The IP address of interface Loopback 1023 is specified as the management address.

```
[GNE] interface gigabitethernet 1/0/2
[GNE-GigabitEthernet1/0/2] port access vlan 3002
[GNE-GigabitEthernet1/0/2] lldp source-mac vlan 3002
[GNE-GigabitEthernet1/0/2] lldp tlv-enable basic-tlv management-address-tlv
interface loopback 1023
```

Configure the system to issue the generated ARP entry to the Layer 3 Ethernet subinterface associated with VLAN 3002 in Dot1q termination after GigabitEthernet 1/0/2 receives an LLDP frame.

```
[GNE-GigabitEthernet1/0/2] lldp management-address arp-learning vlan 3002
[GNE-GigabitEthernet1/0/2] quit
```

Create VLAN-interface 3001 that borrows the IP address of Loopback 1023.

```
[GNE] interface Vlan-interface 3001
[GNE-Vlan-interface3001] ip binding vpn-instance dcn_vpn
[GNE-Vlan-interface3001] ip address unnumbered interface loopback 1023
[GNE-Vlan-interface3001] quit
```

Create VLAN-interface 3002 that borrows the IP address of Loopback 1023.

```
[GNE] interface Vlan-interface 3002
[GNE-Vlan-interface3002] ip binding vpn-instance dcn_vpn
[GNE-Vlan-interface3002] ip address unnumbered interface loopback 1023
[GNE-Vlan-interface3002] quit
```

Enable OSPF process 65535 and create area 0.

```
[GNE] ospf 65535 vpn-instance dcn_vpn
[GNE-ospf-65535] area 0
[GNE-ospf-65535-area-0.0.0.0] network 0.0.0.0 255.255.255.255
[GNE-ospf-65535-area-0.0.0.0] quit
[GNE-ospf-65535] quit
```

Set the OSPF network type for VLAN-interface 3001 to P2P.

```
[GNE] interface Vlan-interface 3001
[GNE-Vlan-interface3001] ospf network-type p2p
[GNE-Vlan-interface3001] quit
```

Set the OSPF network type for VLAN-interface 3002 to P2P.

```
[GNE] interface Vlan-interface 3002
```

```
[GNE-Vlan-interface3002] ospf network-type p2p
[GNE-Vlan-interface3002] quit
```

2. Configure Device A:

Enable DCN, configure the NE ID as 200002 and NE IP as 22.2.2.2/32.

```
<DeviceA> system-view
[DeviceA] dcn
[DeviceA-dcn] ne-id 200002
[DeviceA-dcn] ne-ip 22.2.2.2 32
[DeviceA-dcn] quit
```

Create a VPN instance named **dcn_vpn**.

```
[DeviceA] ip vpn-instance dcn_vpn
[DeviceA-vpn-instance-dcn_vpn] quit
```

Create interface Loopback 1023, and associate it with VPN instance **dcn_vpn**.

```
[DeviceA] interface loopback 1023
[DeviceA-LoopBack1023] ip binding vpn-instance dcn_vpn
[DeviceA-LoopBack1023] quit
```

Enable LLDP globally.

```
[DeviceA] lldp global enable
```

Enable the nearest bridge agents on GigabitEthernet 1/0/2 to advertise basic LLDP TLVs and management address TLVs. The IP address of interface Loopback 1023 is specified as the management address.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port access vlan 3002
[DeviceA-GigabitEthernet1/0/2] lldp source-mac vlan 3002
[DeviceA-GigabitEthernet1/0/2] lldp tlv-enable basic-tlv management-address-tlv
interface loopback 1023
```

Configure the system to issue the generated ARP entry to the Layer 3 Ethernet subinterface associated with VLAN 3002 in Dot1q termination after GigabitEthernet 1/0/2 receives an LLDP frame.

```
[DeviceA-GigabitEthernet1/0/2] lldp management-address arp-learning vlan 3002
[DeviceA-GigabitEthernet1/0/2] quit
```

Enable the nearest bridge agents on GigabitEthernet 1/0/1 to advertise basic LLDP TLVs and management address TLVs. The IP address of interface Loopback 1023 is specified as the management address.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port access vlan 3003
[DeviceA-GigabitEthernet1/0/1] lldp source-mac vlan 3003
[DeviceA-GigabitEthernet1/0/1] lldp tlv-enable basic-tlv management-address-tlv
interface loopback 1023
```

Configure the system to issue the generated ARP entry to the Layer 3 Ethernet subinterface associated with VLAN 3003 in Dot1q termination after GigabitEthernet 1/0/1 receives an LLDP frame.

```
[DeviceA-GigabitEthernet1/0/1] lldp management-address arp-learning vlan 3003
[DeviceA-GigabitEthernet1/0/1] quit
```

Create VLAN-interface 3002 that borrows the IP address of Loopback 1023.

```
[DeviceA] interface Vlan-interface 3002
[DeviceA-Vlan-interface3002] ip binding vpn-instance dcn_vpn
[DeviceA-Vlan-interface3002] ip address unnumbered interface loopback 1023
[DeviceA-Vlan-interface3002] quit
```

Create VLAN-interface 3003 that borrows the IP address of Loopback 1023.

```
[DeviceA] interface Vlan-interface 3003
[DeviceA-Vlan-interface3003] ip binding vpn-instance dcn_vpn
[DeviceA-Vlan-interface3003] ip address unnumbered interface loopback 1023
[DeviceA-Vlan-interface3003] quit
```

Enable OSPF process 65535 and create area 0.

```
[DeviceA] ospf 65535 vpn-instance dcn_vpn
[DeviceA-ospf-65535] area 0
[DeviceA-ospf-65535-area-0.0.0.0] network 0.0.0.0 255.255.255.255
[DeviceA-ospf-65535-area-0.0.0.0] quit
[DeviceA-ospf-65535] quit
```

Set the OSPF network type for VLAN-interface 3002 to P2P.

```
[DeviceA] interface Vlan-interface 3002
[DeviceA-Vlan-interface3002] ospf network-type p2p
[DeviceA-Vlan-interface3002] quit
```

Set the OSPF network type for VLAN-interface 3003 to P2P.

```
[DeviceA] interface Vlan-interface 3003
[DeviceA-Vlan-interface3003] ospf network-type p2p
[DeviceA-Vlan-interface3003] quit
```

3. Configure Device B:

Enable DCN, configure the NE ID as 300003 and NE IP as 33.3.3.3/32.

```
<DeviceB> system-view
[DeviceB] dcn
[DeviceB-dcn] ne-id 300003
[DeviceB-dcn] ne-ip 33.3.3.3 32
[DeviceB-dcn] quit
```

Create a VPN instance named **dcn_vpn**.

```
[DeviceB] ip vpn-instance dcn_vpn
[DeviceB-vpn-instance-dcn_vpn] quit
```

Create interface Loopback 1023, and associate it with VPN instance **dcn_vpn**.

```
[DeviceB] interface loopback 1023
[DeviceB-LoopBack1023] ip binding vpn-instance dcn_vpn
[DeviceB-LoopBack1023] quit
```

Enable LLDP globally.

```
[DeviceB] lldp global enable
```

Enable the nearest bridge agents on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to advertise basic LLDP TLVs and management address TLVs. The IP address of interface Loopback 1023 is specified as the management address. Configure the system to issue the generated ARP entry after GigabitEthernet 1/0/1 or GigabitEthernet 1/0/2 receives an LLDP frame.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-mode route
[DeviceB-GigabitEthernet1/0/1] lldp tlv-enable basic-tlv management-address-tlv
interface loopback 1023
[DeviceB-GigabitEthernet1/0/1] lldp management-address arp-learning vlan 3001
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-mode route
```

```
[DeviceB-GigabitEthernet1/0/2] lldp tlv-enable basic-tlv management-address-tlv
interface loopback 1023
[DeviceB-GigabitEthernet1/0/2] lldp management-address arp-learning vlan 3003
[DeviceB-GigabitEthernet1/0/2] quit
```

Create VLAN-interface 3001 that borrows the IP address of Loopback 1023.

```
[DeviceB] interface Vlan-interface 3001
[DeviceB-Vlan-interface3001] ip binding vpn-instance dcn_vpn
[DeviceB-Vlan-interface3001] ip address unnumbered interface loopback 1023
[DeviceB-Vlan-interface3001] quit
```

Create VLAN-interface 3003 that borrows the IP address of Loopback 1023.

```
[DeviceB] interface Vlan-interface 3003
[DeviceB-Vlan-interface3003] ip binding vpn-instance dcn_vpn
[DeviceB-Vlan-interface3003] ip address unnumbered interface loopback 1023
[DeviceB-Vlan-interface3003] quit
```

Enable OSPF process 65535 and create area 0.

```
[DeviceB] ospf 65535 vpn-instance dcn_vpn
[DeviceB-ospf-65535] area 0
[DeviceB-ospf-65535-area-0.0.0.0] network 0.0.0.0 255.255.255.255
[DeviceB-ospf-65535-area-0.0.0.0] quit
[DeviceB-ospf-65535] quit
```

Set the OSPF network type for VLAN-interface 3001 to P2P.

```
[DeviceB] interface Vlan-interface 3001
[DeviceB-Vlan-interface3001] ospf network-type p2p
[DeviceB-Vlan-interface3001] quit
```

Set the OSPF network type for VLAN-interface 3003 to P2P.

```
[DeviceB] interface Vlan-interface 3003
[DeviceB-Vlan-interface3003] ospf network-type p2p
[DeviceB-Vlan-interface3003] quit
```

Verifying the configuration

Display brief DCN information on the GNE.

```
[GNE] display dcn

              DCN Brief Information

 NE ID        : 0x100001
 NE IP        : 11.1.1.1
 Mask         : 255.255.255.255
 DCN interface: LoopBack1023
 Auto report  : Enabled
```

Display all DCN NE information on the GNE.

```
[GNE] display dcn ne-info

              DCN Network Elements Information

 NE ID          NE IP            Metric   Device Type
 0x100001       11.1.1.1         0        H3C S5590-EI
 0x200002       22.2.2.2         1        H3C S5590-EI
```

```
0x300003       33.3.3.3         1        H3C S5590-EI

Total number: 3
```

The output shows that GNE, Device A, and Device B are online. The GNE notifies the NMS of the online event. You can successfully ping the NE IP addresses of Device A and Device B from the GNE.

Remove Device B from the DCN network and display all DCN NE information for the GNE.

```
[GNE] display dcn ne-info

              DCN Network Elements Information

NE ID          NE IP            Metric   Device Type
0x100001       11.1.1.1         0        H3C S5590-EI
0x200002       22.2.2.2         1        H3C S5590-EI

Total number: 2
```

The output shows that GNE and Device A are online. The GNE notifies the NMS of the offline event of Device B. You can still successfully ping the NE IP address of Device A from the GNE.

Contents

Configuring dual-stack PBR 1

About dual-stack PBR 1
Dual-stack policy 1
Packet forwarding process 2
Dual-stack PBR and Track 2
Restrictions and guidelines: Dual-stack PBR configuration 2
Dual-stack PBR tasks at a glance 3
Configuring a dual-stack PBR policy 3
Creating a dual-stack policy node 3
Setting match criteria for a dual-stack policy node 3
Configuring actions for a dual-stack policy node 4
Specifying a policy for dual-stack PBR 5
Specifying a policy for interface dual-stack PBR 5
Specifying a policy for global dual-stack PBR 6
Display and maintenance commands for dual-stack PBR 6
Dual-stack PBR configuration examples 7
Example: Configuring packet type-based interface dual-stack PBR 7
Example: Configuring packet type-based global dual-stack PBR 9

Configuring dual-stack PBR

About dual-stack PBR

Dual-stack policy-based routing (dual-stack PBR) uses user-defined policies to route packets. A policy can specify parameters for IPv4 or IPv6 packets that match specific criteria such as ACLs. The parameters include the next hop , output interface, and default next hop.

Dual-stack policy

A dual-stack PBR policy includes match criteria and actions to be taken on the matching packets. A dual-stack PBR policy can have one or multiple nodes as follows:

- Each node is identified by a node number. A smaller node number has a higher priority.
- A node contains **`if-match`** and **`apply`** clauses. An **`if-match`** clause specifies a match criterion, and an **`apply`** clause specifies an action.
- A node has a match mode of **permit** or **deny**.

A dual-stack PBR policy compares packets with dual stack PBR policy nodes in priority order. If a packet matches the criteria on a node, it is processed by the action on the node. If the packet does not match any criteria on the node, it goes to the next node for a match. If the packet does not match the criteria on any node, the device performs a routing table lookup.

Relationship between if-match clauses

On a node, you can specify only one **`if-match acl`** clause.

Relationship between apply clauses

You can specify multiple **`apply`** clauses for a node, but some of them might not be executed. For more information about relationship between **`apply`** clauses, see "Configuring actions for a dual-stack policy node."

Relationship between the match mode and clauses on the node

Does a packet match all the if-match **clauses on the node?**	**Match mode**	
	Permit	**Deny**
Yes.	• If the node contains **`apply`** clauses, dual-stack PBR executes the **`apply`** clauses on the node. ○ If dual-stack PBR-based forwarding succeeds, dual-stack PBR does not compare the packet with the next node. ○ If dual-stack PBR-based forwarding fails, dual-stack PBR does not compare the packet with the next node. • If the node does not contain **`apply`** clauses, the device performs a routing table lookup for the packet.	The device performs a routing table lookup for the packet.
No.	Dual-stack PBR compares the packet	Dual-stack PBR compares the

Does a packet match all the if-match clauses on the node?	Match mode	
	Permit	**Deny**
	with the next node.	packet with the next node.

NOTE:

A node that has no **`if-match`** clauses matches any packet.

Packet forwarding process

You can apply a dual-stack PBR policy to an interface or to all interfaces on the device. Once the policy is applied, the device searches the policy for a matching node to forward that packet upon receiving a packet.

- If a matching node is found and its match mode is **permit**, the device performs the following operations:
  a. Uses the next hops specified on the node to forward the packet.
  b. Searches the routing table for a route (except the default route) to forward the packet if one of the following conditions exists:
    – No next hops are specified on the node.
    – Forwarding failed based on the next hops.
  c. Uses the default route to forward the packet if no matching route was found in the routing table.
- The device performs routing table lookup to forward the packet in either of the following conditions:
  - No matching node is found.
  - A matching node is found, but its match mode is **deny**.

Dual-stack PBR and Track

Dual-stack PBR can work with the Track feature to dynamically adapt the availability status of an **`apply`** clause to the link status of a tracked object. The tracked object can be a next hop or default next hop.

- When the track entry associated with an object changes to **Negative**, the **`apply`** clause is invalid.
- When the track entry changes to **Positive** or **NotReady**, the **`apply`** clause is valid.

For more information about Track and dual-stack PBR collaboration, see *High Availability Configuration Guide*.

Restrictions and guidelines: Dual-stack PBR configuration

If a packet destined for the local device matches a dual-stack PBR policy, dual-stack PBR will execute the apply clauses in the policy, including the clause for forwarding. When you configure a dual-stack PBR policy, be careful to avoid this situation.

Dual-stack PBR tasks at a glance

To configure dual-stack PBR, perform the following tasks:

1. Configuring a dual-stack PBR policy
   d. Creating a dual-stack policy node
   e. Setting match criteria for a dual-stack policy node
   f. Configuring actions for a dual-stack policy node
2. Specifying a policy for dual-stack PBR

   Choose the following tasks as needed:
   - Specifying a policy for interface dual-stack PBR
   - Specifying a policy for global dual-stack PBR

     A global dual-stack PBR policy applies to all interfaces on the device.

Configuring a dual-stack PBR policy

Creating a dual-stack policy node

1. Enter system view.

   system-view
2. Create a node for a dual-stack policy, and enter its view.

   dual-stack policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. (Optional.) Configure a description for the policy node.

   description *text*

   By default, no description is configured for a dual-stack policy node.

Setting match criteria for a dual-stack policy node

1. Enter system view.

   system-view
2. Enter dual-stack policy node view.

   dual-stack policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Set match criteria.
   - Set an ACL match criterion.

     if-match acl { **ipv4** | **ipv6** } { *acl-number* | **name** *acl-name* }

     By default, no ACL match criterion is set.

     When using the ACL to match packets, dual-stack PBR ignores the action (**permit** or **deny**) and time range settings in the ACL.
   - Set a local QoS ID match criterion.

     if-match qos-local-id *local-id-value*

     By default, no local QoS ID match criterion is set.

Configuring actions for a dual-stack policy node

About this task

The **apply** clauses allow you to specify the actions to be taken on matching packets on a dual-stack policy node.

The following **apply** clauses determine the packet forwarding paths in a descending order:

- **apply next-hop**
- **apply output-interface**
- **apply default-next-hop**

Dual-stack PBR supports the **apply** clauses in Table 1.

Table 1 Apply clauses supported in PBR

Clause	Meaning	Remarks
apply precedence	Sets an IP precedence.	This clause is always executed.
apply next-hop and **apply output-interface**	Sets next hops and sets output interfaces.	If both clauses are configured, only the **apply next-hop** clause is executed.
apply default-next-hop	Sets default next hops	The clause takes effect only in the following cases: • No next hops are set or the next hops are invalid. • The packet does not match any route in the routing table.
apply statistics	Counts successful matches and matching bytes on the policy node.	To include the number of successful matches and matching bytes on a policy node in the statistics displayed by a **display** command, execute this command. This clause is always executed for matching packets.

Restrictions and guidelines

If you specify a next hop, dual-stack PBR periodically performs FIB table lookup to determine its availability. Temporary service interruption might occur if dual-stack PBR does not update the route immediately after its availability status changes.

Configuring actions to set the IP precedence for packets

1. Enter system view.

   system-view

2. Enter dual-stack policy node view.

   dual-stack policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*

3. Set an IP precedence for packets.

   apply precedence { *type* | *value* }

   By default, no IP precedence is specified for packets.

Configuring actions to direct packet forwarding

1. Enter system view.

   system-view

2. Enter dual-stack policy node view.

   dual-stack policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*

3. Configure actions.

   - Set next hops.

     apply next-hop [**vpn-instance** *vpn-instance-name*] { { *ipv4-address* | *ipv6-address* } [**direct**] [**track** *track-entry-number*] }&<1-4>

     By default, no next hops are specified.

     On a node, you can specify a maximum of four next hops for backup in one command line or by executing this command multiple times.

     If multiple next hops on the same subnet are specified for backup, the device first uses the subnet route for the next hops to forward packets when the primary next hop fails. If the subnet route is not available, the device selects a backup next hop.

   - Set output interfaces.

     apply output-interface NULL 0

     By default, no output interfaces are specified.

   - Set default next hops.

     apply default-next-hop [**vpn-instance** *vpn-instance-name*] { { *ipv4-address* | *ipv6-address* } [**direct**] [**track** *track-entry-number*] }&<1-4>

     By default, no default next hops are specified.

     On a node, you can specify a maximum of four default next hops for backup in one command line or by executing this command multiple times.

Collecting policy node statistics

1. Enter system view.

   system-view

2. Enter dual-stack policy node view.

   dual-stack policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*

3. Count successful matches and matching bytes on the policy node.

   apply statistics

   By default, matching counting is disabled on policy nodes.

Specifying a policy for dual-stack PBR

Specifying a policy for interface dual-stack PBR

About this task

Perform this task to apply a dual-stack policy to an interface to guide the forwarding of packets received on the interface.

Restrictions and guidelines

You can apply only one dual-stack policy to an interface and make sure the specified policy already exists. Before you can apply a new dual-stack PBR policy to an interface, you must first remove the current dual-stack policy from the interface.

You can apply a dual-stack policy to multiple interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify a dual-stack policy for interface PBR.
   dual-stack policy-based-route *policy-name*
   By default, no interface dual-stack policy is applied to an interface.

Specifying a policy for global dual-stack PBR

About this task

Perform this task to apply a dual-stack policy to all interfaces on the device to guide the forwarding of packets received on the interfaces.

Restrictions and guidelines

You can apply only one policy for global dual-stack PBR and the specified policy must already exist. Before you can apply a new policy, you must first remove the current policy.

Interface dual-stack PBR takes precedence over global dual-stack PBR on an interface. When they are both configured and packets fail to match the interface dual-stack PBR policy, global dual-stack PBR applies.

Procedure

1. Enter system view.
   system-view
2. Specify a policy for global dual-stack PBR.
   dual-stack global policy-based-route *policy-name*
   By default, no policy is specified for global dual-stack PBR.

Display and maintenance commands for dual-stack PBR

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display dual-stack PBR policy information.	**display dual-stack policy-based-route** [**policy** *policy-name*]

Task	Command
Display global dual-stack PBR configuration and statistics.	**display dual-stack policy-based-route global** [**slot** *slot-number*]
Display interface dual-stack PBR configuration and statistics.	**display dual-stack policy-based-route interface** *interface-type interface-number* [**slot** *slot-number*]
Display dual-stack PBR configuration.	**display dual-stack policy-based-route setup**
Clear dual-stack PBR statistics.	**reset dual-stack policy-based-route statistics** [**policy** *policy-name*]

Dual-stack PBR configuration examples

Example: Configuring packet type-based interface dual-stack PBR

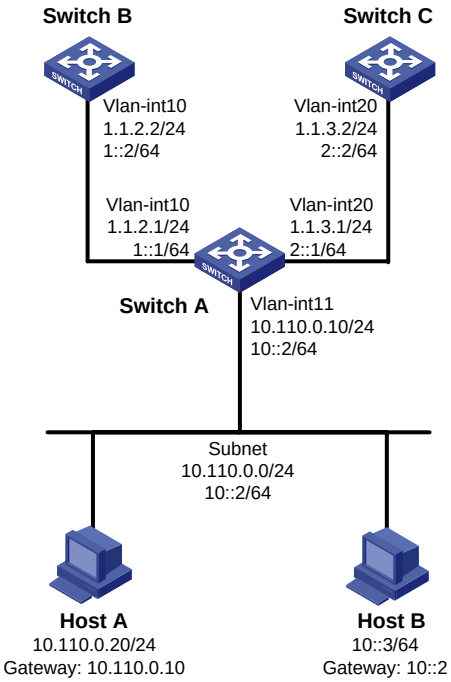
Network configuration

As shown in Figure 1, Switch B and Switch C do not have a route to reach each other.

Configure dual-stack PBR on Switch A to implement the following requirements:

- Forward all IPv4 TCP packets received on VLAN-interface 11 to next hop 1.1.2.2 (Switch B).
- Forward all IPv6 TCP packets received on VLAN-interface 11 to next hop 2::2 (Switch C).

Figure 1 Network diagram



Procedure

1. Assign IP addresses to the interfaces as shown in Figure 1. (Details not shown.)
2. Configure static or dynamic routing protocol settings to make sure Host A and B can communicate with Switch B and Switch C, respectively. (Details not shown.)
3. Configure dual-stack PBR on Switch A:

   # Configure ACL 3101 to match IPv4 TCP packets.

   ```
   <SwitchA> system-view
   [SwitchA] acl advanced 3101
   [SwitchA-acl-ipv4-adv-3101] rule permit tcp
   [SwitchA-acl-ipv4-adv-3101] quit
   ```

   # Configure ACL 3102 to match IPv6 TCP packets.

   ```
   [SwitchA] acl ipv6 advanced 3102
   [SwitchA-acl-ipv6-adv-3102] rule permit tcp
   [SwitchA-acl-ipv6-adv-3102] quit
   ```

   # Configure Node 5 for the dual-stack policy **aaa** to forward IPv4 TCP packets to next hop 1.1.2.2.

   ```
   [SwitchA] dual-stack policy-based-route aaa permit node 5
   [SwitchA-pbrdual-aaa-5] if-match acl ipv4 3101
   [SwitchA-pbrdual-aaa-5] apply next-hop 1.1.2.2
   [SwitchA-pbrdual-aaa-5] quit
   ```

   # Configure Node 10 for the dual-stack policy **aaa** to forward IPv6 TCP packets to next hop 2::2.

   ```
   [SwitchA] dual-stack policy-based-route aaa permit node 10
   [SwitchA-pbrdual-aaa-10] if-match acl ipv6 3102
   [SwitchA-pbrdual-aaa-10] apply next-hop 2::2
   [SwitchA-pbrdual-aaa-10] quit
   ```

   # Configure interface dual-stack PBR by applying policy **aaa** to VLAN-interface 11.

   ```
   [SwitchA] interface vlan-interface 11
   ```

```
[SwitchA-Vlan-interface11] dual-stack policy-based-route aaa
[SwitchA-Vlan-interface11] quit
```

Verifying the configuration

Perform telnet and ping operations to verify that interface dual-stack PBR on Switch A operates as configured to perform packet forwarding as follows:

- Forwards the matching IPv4 TCP packets to the next hop 1.1.2.2 (Switch B).
- Forwards the matching IPv6 TCP packets to the next hop 2::2 (Switch C).

Verify that you can telnet to Switch B (1.1.2.2) from Host A successfully. (Details not shown.)

Verify that you cannot telnet to Switch C (1.1.3.2) from Host A. (Details not shown.)

Verify that you can ping Switch C (1.1.3.2) from Host A successfully. (Details not shown.)

Verify that you can telnet to Switch C (2::2) from Host B successfully. (Details not shown.)

Verify that you cannot telnet to Switch B (1::2) from Host B. (Details not shown.)

Verify that you can ping Switch B (1::2) from Host B successfully. (Details not shown.)

Example: Configuring packet type-based global dual-stack PBR

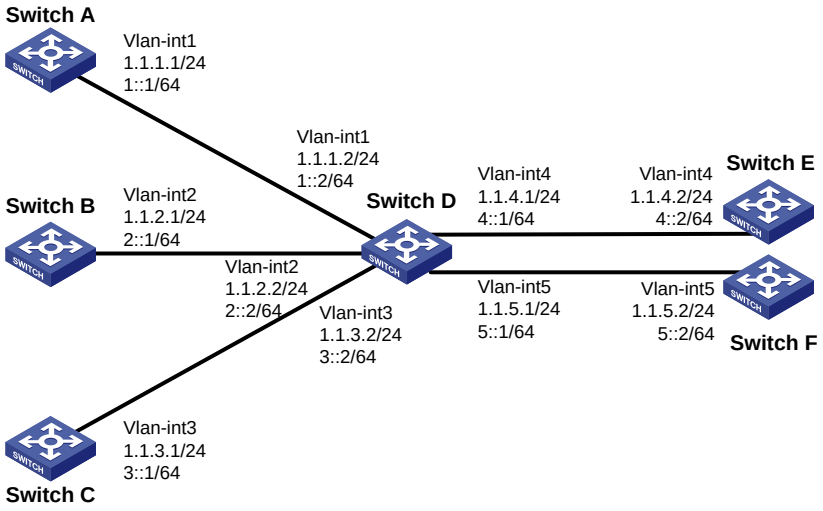
Network configuration

As shown in Figure 2, Switch E and Switch F do not have a route to reach each other.

Configure global dual-stack PBR on Switch D to implement the following requirements:

- Forward IPv4 TCP packets to next hop 1.1.4.2 (Switch E).
- Forward IPv6 TCP packets to next hop 5::2 (Switch F).

Figure 2 Network diagram



Procedure

1. Assign IP addresses to the interfaces, as shown in Figure 2. (Details not shown.)
2. Configure static or dynamic routing protocol settings to make sure Switch A, B and C can communicate with Switch E and Switch F, respectively. (Details not shown.)
3. Configure dual-stack PBR on Switch D:

Configure ACL 3101 to match IPv4 TCP packets sourced from networks 1.1.1.0/24, 1.1.2.0/24, and 1.1.3.0/24.

```
<SwitchD> system-view
[SwitchD] acl advanced 3101
[SwitchD-acl-ipv4-adv-3101] rule permit tcp source 1.1.1.0 0.0.0.255
[SwitchD-acl-ipv4-adv-3101] rule permit tcp source 1.1.2.0 0.0.0.255
[SwitchD-acl-ipv4-adv-3101] rule permit tcp source 1.1.3.0 0.0.0.255
[SwitchD-acl-ipv4-adv-3101] quit
```

Configure ACL 3102 to match IPv6 TCP packets sourced from networks 1::0/64, 2::0/64, and 3::0/64.

```
<RouterD> system-view
[RouterD] acl ipv6 advanced 3102
[RouterD-acl-ipv6-adv-3102] rule permit tcp source 1::0 64
[RouterD-acl-ipv6-adv-3102] rule permit tcp source 2::0 64
[RouterD-acl-ipv6-adv-3102] rule permit tcp source 3::0 64
[RouterD-acl-ipv6-adv-3102] quit
```

Configure node 5 in dual-stack PBR policy **aaa** to forward IPv4 TCP packets that match ACL 3101 to next hop 1.1.4.2.

```
[SwitchD] dual-stack policy-based-route aaa permit node 5
[SwitchD-pbrdual-aaa-5] if-match acl ipv4 3101
[SwitchD-pbrdual-aaa-5] apply next-hop 1.1.4.2
[SwitchD-pbrdual-aaa-5] quit
```

Configure node 10 in dual-stack PBR policy **aaa** to forward IPv6 TCP packets that match ACL 3102 to next hop 5::2.

```
[SwitchD] dual-stack policy-based-route aaa permit node 10
[SwitchD-pbrdual-aaa-10] if-match acl ipv6 3102
[SwitchD-pbrdual-aaa-10] apply next-hop 5::2
[SwitchD-pbrdual-aaa-10] quit
```

Specify dual-stack PBR policy **aaa** as the global PBR policy.

```
[SwitchD] dual-stack global policy-based-route aaa
```

Verifying the configuration

Perform telnet and ping operations to verify that global dual-stack PBR on Switch D operates as configured to perform packet forwarding as follows:

- Forwards the matching IPv4 TCP packets to the next hop 1.1.4.2 (Switch E).
- Forwards the matching IPv6 TCP packets to the next hop 5::2 (Switch F).

Verify that you can telnet to Switch E (1.1.4.2) successfully and cannot telnet to Switch F (1.1.5.2), from Switch A, B, and C. (Details not shown.)

Verify that you can ping Switch F (1.1.5.2) from Switch A, B and C successfully. (Details not shown.)

Verify that you can telnet to Switch F (5::2) successfully and cannot telnet to Switch E (4::2/64), from Switch A, B, and C. (Details not shown.)

Verify that you can ping Switch F (4::2) from Switch A, B and C successfully. (Details not shown.)

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

IP Multicast Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes IP multicast fundamentals and configuration procedures. It covers the IPv4 multicast configuration and IPv6 multicast configuration. With multicast, you can implement efficient point-to-multipoint data transmission in your network.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** >

	Folder.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Multicast overview ····· 1

Introduction to multicast ····· 1
Information transmission techniques ····· 1
Multicast features ····· 3
Multicast benefits and applications ····· 4
Multicast models ····· 4
ASM model ····· 4
SFM model ····· 4
SSM model ····· 5
Multicast addresses ····· 5
IP multicast addresses ····· 5
Ethernet multicast MAC addresses ····· 7
Multicast protocols ····· 8
Layer 3 multicast protocols ····· 9
Layer 2 multicast protocols ····· 9
Multicast packet forwarding mechanism ····· 10
IP multicast architecture ····· 11
Multicast support for VPNs ····· 11
Introduction to VPN instances ····· 11
Multicast application in VPNs ····· 12
Common notations in multicast ····· 12

Multicast overview

Introduction to multicast

As a technique that coexists with unicast and broadcast, the multicast technique effectively addresses the issue of point-to-multipoint data transmission. By enabling high-efficiency point-to-multipoint data transmission over a network, multicast greatly saves network bandwidth and reduces network load.

By using multicast technology, a network operator can easily provide bandwidth-critical and time-critical information services. These services include live webcasting, Web TV, distance learning, telemedicine, Web radio, and real-time video conferencing.

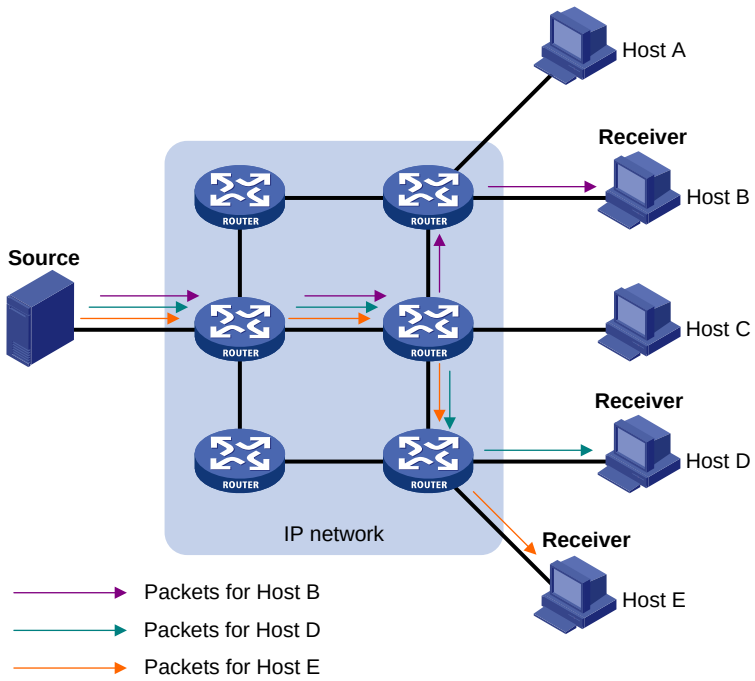
Information transmission techniques

The information transmission techniques include unicast, broadcast, and multicast.

Unicast

In unicast transmission, the information source must send a separate copy of information to each host that needs the information.

Figure 1 Unicast transmission



In Figure 1, Host B, Host D, and Host E need the information. A separate transmission channel must be established from the information source to each of these hosts.

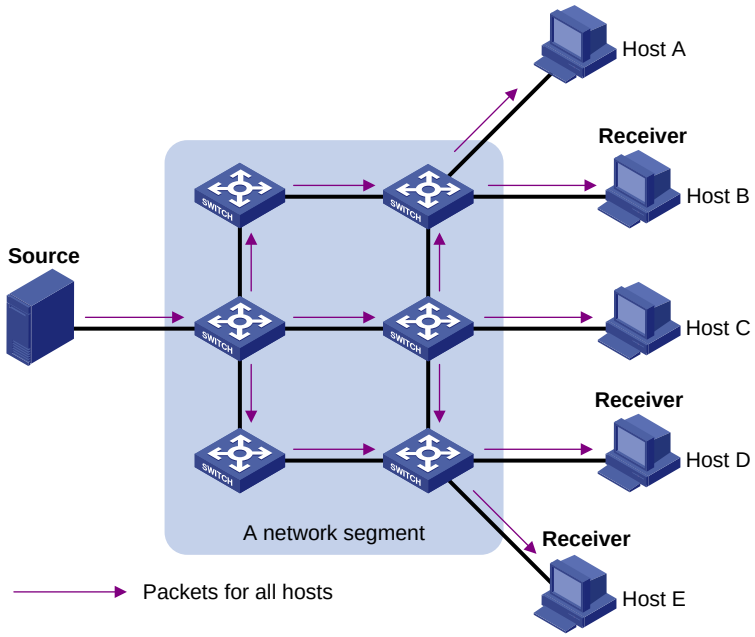
In unicast transmission, the traffic transmitted over the network is proportional to the number of hosts that need the information. If a large number of hosts need the information, the information source must send a separate copy of the same information to each of these hosts. Sending many copies can place a tremendous pressure on the information source and the network bandwidth.

Unicast is not suitable for batch transmission of information.

Broadcast

In broadcast transmission, the information source sends information to all hosts on the subnet, even if some hosts do not need the information.

Figure 2 Broadcast transmission



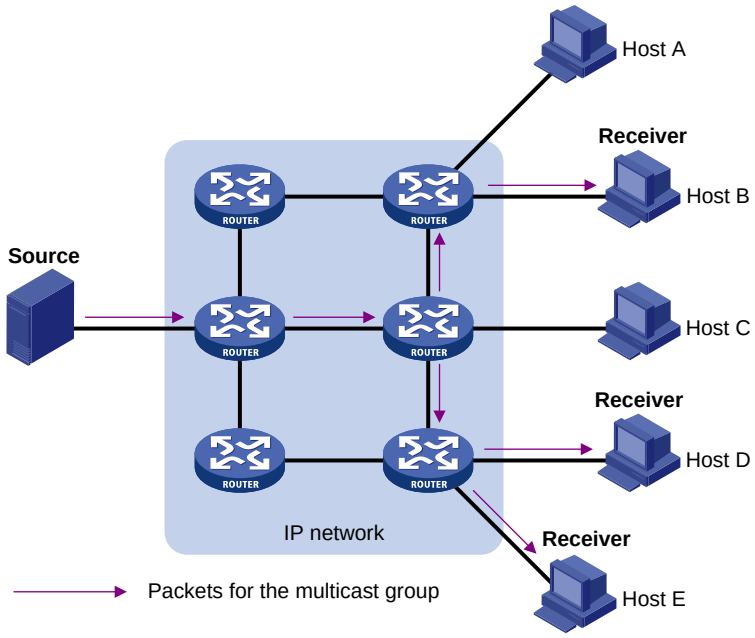
In Figure 2, only Host B, Host D, and Host E need the information. If the information is broadcast to the subnet, Host A and Host C also receive it. In addition to information security issues, broadcasting to hosts that do not need the information also causes traffic flooding on the same subnet.

Broadcast is disadvantageous in transmitting data to specific hosts. Moreover, broadcast transmission is a significant waste of network resources.

Multicast

Multicast provides point-to-multipoint data transmissions with the minimum network consumption. When some hosts on the network need multicast information, the information sender, or multicast source, sends only one copy of the information. Multicast distribution trees are built through multicast routing protocols, and the packets are replicated only on nodes where the trees branch.

Figure 3 Multicast transmission



In Figure 3, the multicast source sends only one copy of the information to a multicast group. Host B, Host D, and Host E, which are information receivers, must join the multicast group. The routers on the network duplicate and forward the information based on the distribution of the group members. Finally, the information is correctly delivered to Host B, Host D, and Host E.

To summarize, multicast has the following advantages:

- **Advantages over unicast**—Multicast data is replicated and distributed until it flows to the farthest-possible node from the source. The increase of receiver hosts will not remarkably increase the load of the source or the usage of network resources.
- **Advantages over broadcast**—Multicast data is sent only to the receivers that need it. This saves network bandwidth and enhances network security. In addition, multicast data is not confined to the same subnet.

Multicast features

- A multicast group is a multicast receiver set identified by an IP multicast address. Hosts must join a multicast group to become members of the multicast group before they receive the multicast data addressed to that multicast group. Typically, a multicast source does not need to join a multicast group.
- A multicast source is an information sender. It can send data to multiple multicast groups at the same time. Multiple multicast sources can send data to the same multicast group at the same time.
- The group memberships are dynamic. Hosts can join or leave multicast groups at any time. Multicast groups are not subject to geographic restrictions.
- Multicast routers or Layer 3 multicast devices are routers or Layer 3 switches that support Layer 3 multicast. They provide multicast routing and manage multicast group memberships on stub subnets with attached group members. A multicast router itself can be a multicast group member.

For a better understanding of the multicast concept, you can compare multicast transmission to the transmission of TV programs.

Table 1 Comparing TV program transmission and multicast transmission

TV program transmission	Multicast transmission
A TV station transmits a TV program through a channel.	A multicast source sends multicast data to a multicast group.
A user tunes the TV set to the channel.	A receiver joins the multicast group.
The user starts to watch the TV program transmitted by the TV station on the channel.	The receiver starts to receive the multicast data sent by the source to the multicast group.
The user turns off the TV set or tunes to another channel.	The receiver leaves the multicast group or joins another group.

Multicast benefits and applications

Multicast benefits

- **Enhanced efficiency**—Reduces the processor load of information source servers and network devices.
- **Optimal performance**—Reduces redundant traffic.
- **Distributed application**—Enables point-to-multipoint applications at the price of minimum network resources.

Multicast applications

- Multimedia and streaming applications, such as Web TV, Web radio, and real-time video/audio conferencing
- Communication for training and cooperative operations, such as distance learning and telemedicine
- Data warehouse and financial applications (stock quotes)
- Any other point-to-multipoint application for data distribution

Multicast models

Based on how the receivers treat the multicast sources, the multicast models include any-source multicast (ASM), source-filtered multicast (SFM), and source-specific multicast (SSM).

ASM model

In the ASM model, any multicast sources can send information to a multicast group. Receivers can join a multicast group and get multicast information addressed to that multicast group from any multicast sources. In this model, receivers do not know the positions of the multicast sources in advance.

SFM model

The SFM model is derived from the ASM model. To a multicast source, the two models appear to have the same multicast membership architecture.

The SFM model functionally extends the ASM model. The upper-layer software checks the source address of received multicast packets and permits or denies multicast traffic from specific sources.

The receivers obtain the multicast data from only part of the multicast sources. To a receiver, multicast sources are not all valid, but are filtered.

SSM model

The SSM model provides a transmission service that enables multicast receivers to specify the multicast sources in which they are interested.

In the SSM model, receivers have already determined the locations of the multicast sources. This is the main difference between the SSM model and the ASM model. In addition, the SSM model uses a different multicast address range than the ASM/SFM model. Dedicated multicast forwarding paths are established between receivers and the specified multicast sources.

Multicast addresses

IP multicast addresses

IPv4 multicast addresses

IANA assigned the Class D address block (224.0.0.0 to 239.255.255.255) to IPv4 multicast.

Table 2 Class D IP address blocks and description

Address block	Description
224.0.0.0 to 224.0.0.255	Reserved permanent group addresses. The IP address 224.0.0.0 is reserved. Other IP addresses can be used by routing protocols and for topology searching, protocol maintenance, and so on. Table 3 lists common permanent group addresses. A packet destined for an address in this block will not be forwarded beyond the local subnet regardless of the TTL value in the IP header.
224.0.1.0 to 238.255.255.255	Globally scoped group addresses. This block includes the following types of designated group addresses: • **232.0.0.0/8**—SSM group addresses. • **233.0.0.0/8**—Glop group addresses.
239.0.0.0 to 239.255.255.255	Administratively scoped multicast addresses. These addresses are considered locally unique rather than globally unique. You can reuse them in domains administered by different organizations without causing conflicts. For more information, see RFC 2365.

NOTE:

Glop is a mechanism for assigning multicast addresses between different ASs. By filling an AS number into the middle two bytes of 233.0.0.0, you get 255 multicast addresses for that AS. For more information, see RFC 2770.

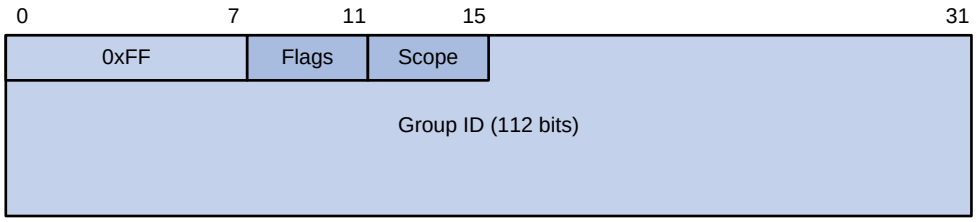
Table 3 Common permanent multicast group addresses

Address	Description
224.0.0.1	All systems on this subnet, including hosts and routers.
224.0.0.2	All multicast routers on this subnet.
224.0.0.3	Unassigned.
224.0.0.4	DVMRP routers.
224.0.0.5	OSPF routers.

Address	Description
224.0.0.6	OSPF designated routers and backup designated routers.
224.0.0.7	Shared Tree (ST) routers.
224.0.0.8	ST hosts.
224.0.0.9	RIPv2 routers.
224.0.0.11	Mobile agents.
224.0.0.12	DHCP server/relay agent.
224.0.0.13	All Protocol Independent Multicast (PIM) routers.
224.0.0.14	This protocol is not supported in the current software version. RSVP encapsulation.
224.0.0.15	All Core-Based Tree (CBT) routers.
224.0.0.16	Designated SBM.
224.0.0.17	All SBMs.
224.0.0.18	VRRP.

IPv6 multicast addresses

Figure 4 IPv6 multicast format



The following describes the fields of an IPv6 multicast address:

- **0xFF**—The most significant eight bits are 11111111.
- **Flags**—The Flags field contains four bits.

  Figure 5 Flags field format

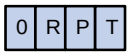
  

  Table 4 Flags field description

Bit	Description
0	Reserved, set to 0.
R	• When set to 0, this address is an IPv6 multicast address without an embedded RP address. • When set to 1, this address is an IPv6 multicast address with an embedded RP address. (The P and T bits must also be set to 1.)
P	• When set to 0, this address is an IPv6 multicast address not based on a unicast prefix. • When set to 1, this address is an IPv6 multicast address based on a unicast prefix. (The T bit must also be set to

Bit	Description
	1.)
T	• When set to 0, this address is an IPv6 multicast address permanently-assigned by IANA. • When set to 1, this address is a transient or dynamically assigned IPv6 multicast address.

- **Scope**—The Scope field contains four bits, which represent the scope of the IPv6 internetwork for which the multicast traffic is intended.

  Table 5 Values of the Scope field

Value	Meaning
0, F	Reserved.
1	Interface-local scope.
2	Link-local scope.
3	Subnet-local scope.
4	Admin-local scope.
5	Site-local scope.
6, 7, 9 through D	Unassigned.
8	Organization-local scope.
E	Global scope.

- **Group ID**—The Group ID field contains 112 bits. It uniquely identifies an IPv6 multicast group in the scope that the Scope field defines.

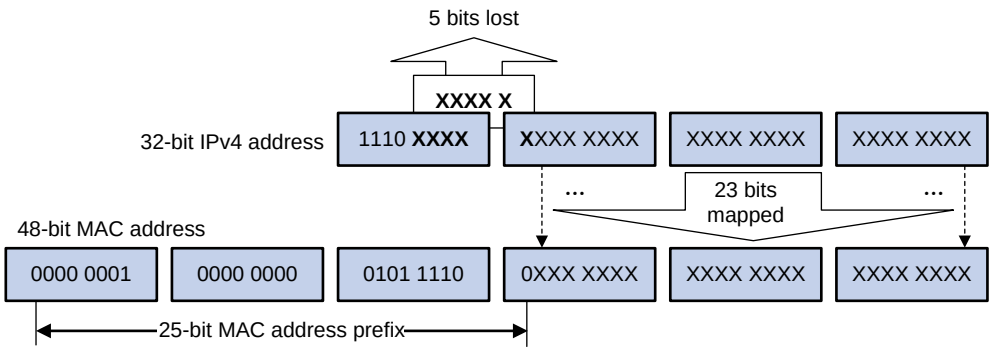
Ethernet multicast MAC addresses

An Ethernet multicast MAC address identifies receivers that belong to the same multicast group at the data link layer.

IPv4 multicast MAC addresses

As defined by IANA, the most significant 24 bits of an IPv4 multicast MAC address are 0x01005E. Bit 25 is 0, and the other 23 bits are the least significant 23 bits of an IPv4 multicast address.

Figure 6 IPv4-to-MAC address mapping

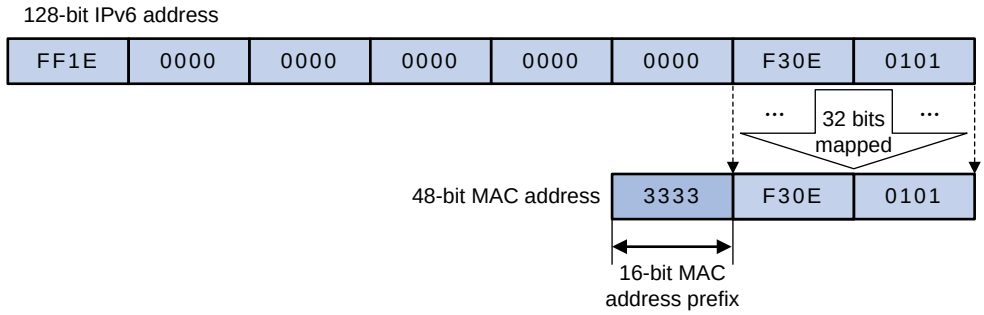


The most significant four bits of an IPv4 multicast address are fixed at 1110. In an IPv4-to-MAC address mapping, five bits of the IPv4 multicast address are lost. As a result, 32 IPv4 multicast addresses are mapped to the same IPv4 multicast MAC address. A device might receive unwanted multicast data at Layer 2 processing, which needs to be filtered by the upper layer.

IPv6 multicast MAC addresses

As defined by IANA, the most significant 16 bits of an IPv6 multicast MAC address are 0x3333. The least significant 32 bits are mapped from the least significant 32 bits of an IPv6 multicast address. Therefore, the problem of duplicate IPv6-to-MAC address mapping also arises like IPv4-to-MAC address mapping.

Figure 7 IPv6-to-MAC address mapping



IMPORTANT:

Because of the duplicate mapping from multicast IP address to multicast MAC address, the device might inadvertently send multicast protocol packets as multicast data in Layer 2 forwarding. To avoid this, do not use the IP multicast addresses that are mapped to multicast MAC addresses 0100-5E00-00xx and 3333-0000-00xx (where "x" represents any hexadecimal number from 0 to F).

Multicast protocols

Multicast protocols include the following categories:

- Layer 3 and Layer 2 multicast protocols:
  - Layer 3 multicast refers to IP multicast operating at the network layer.

    Layer 3 multicast protocols—IGMP, MLD, PIM, IPv6 PIM, MSDP, MBGP, and IPv6 MBGP.
  - Layer 2 multicast refers to IP multicast operating at the data link layer.

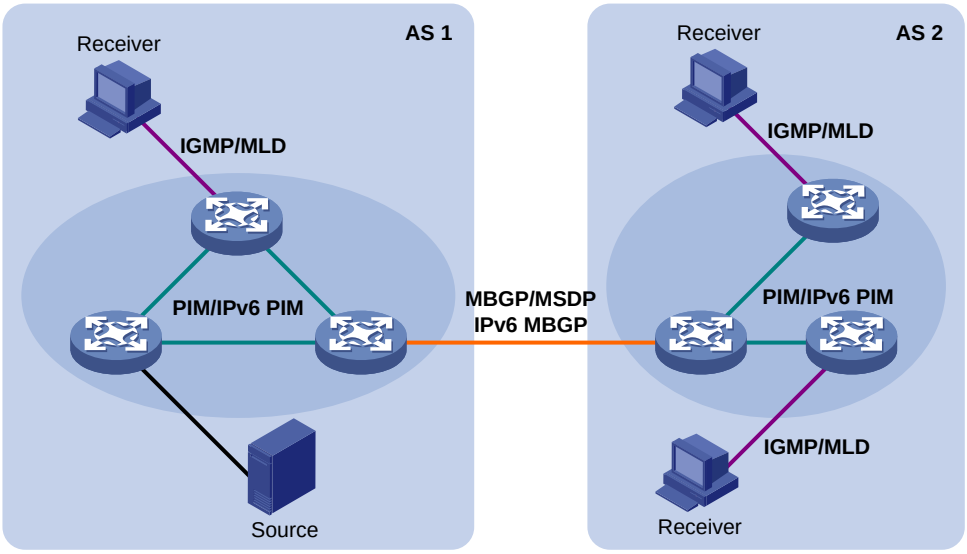
    Layer 2 multicast protocols—IGMP snooping, MLD snooping, PIM snooping, IPv6 PIM snooping, multicast VLAN, and IPv6 multicast VLAN.
- IPv4 and IPv6 multicast protocols:
  - **For IPv4 networks**—IGMP snooping, PIM snooping, multicast VLAN, IGMP, PIM, MSDP, and MBGP.
  - **For IPv6 networks**—MLD snooping, IPv6 PIM snooping, IPv6 multicast VLAN, MLD, IPv6 PIM, and IPv6 MBGP.

This section provides only general descriptions about applications and functions of the Layer 2 and Layer 3 multicast protocols in a network. For more information about these protocols, see the related chapters.

Layer 3 multicast protocols

In Figure 8, Layer 3 multicast protocols include multicast group management protocols and multicast routing protocols.

Figure 8 Positions of Layer 3 multicast protocols



- Multicast group management protocols:

  Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) protocol are multicast group management protocols. Typically, they run between hosts and Layer 3 multicast devices that directly connect to the hosts to establish and maintain multicast group memberships.

- Multicast routing protocols:

  A multicast routing protocol runs on Layer 3 multicast devices to establish and maintain multicast routes and correctly and efficiently forward multicast packets. Multicast routes constitute loop-free data transmission paths (also known as multicast distribution trees) from a data source to multiple receivers.

  In the ASM model, multicast routes include intra-domain routes and inter-domain routes.

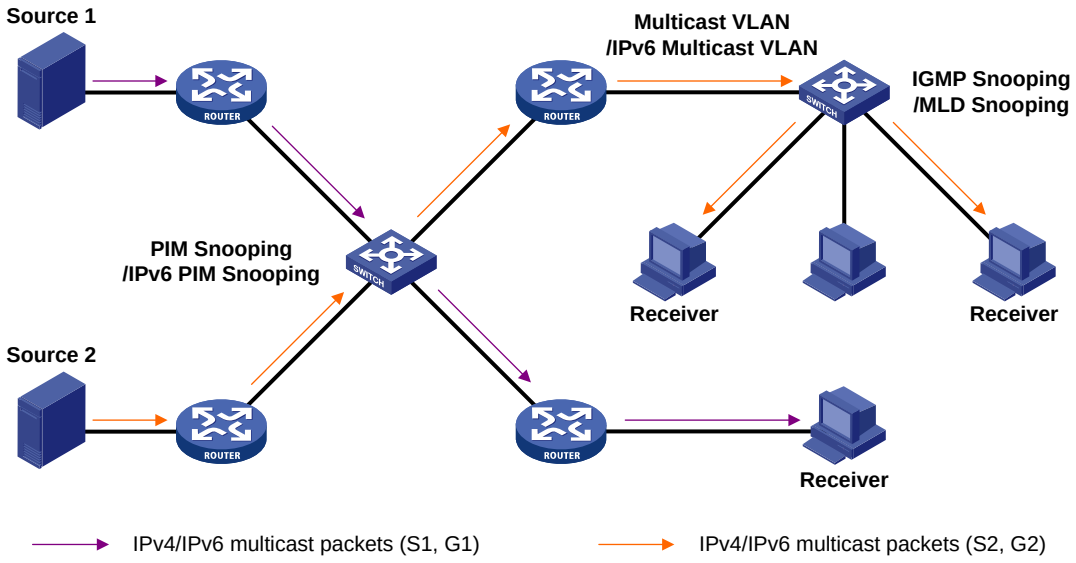
  - An intra-domain multicast routing protocol discovers multicast sources and builds multicast distribution trees within an AS to deliver multicast data to receivers. Among a variety of mature intra-domain multicast routing protocols, PIM is most widely used. Based on the forwarding mechanism, PIM has dense mode (often referred to as PIM-DM) and sparse mode (often referred to as PIM-SM).
  - An inter-domain multicast routing protocol is used for delivering multicast information between two ASs. So far, mature solutions include Multicast Source Discovery Protocol (MSDP) and MBGP. MSDP propagates multicast source information among different ASs. MBGP is an extension of the MP-BGP for exchanging multicast routing information among different ASs.

  For the SSM model, multicast routes are not divided into intra-domain routes and inter-domain routes. Because receivers know the positions of the multicast sources, channels established through PIM-SM are sufficient for the transport of multicast information.

Layer 2 multicast protocols

In Figure 9, Layer 2 multicast protocols include IGMP snooping, MLD snooping, PIM snooping, IPv6 PIM snooping, multicast VLAN, and IPv6 multicast VLAN.

Figure 9 Positions of Layer 2 multicast protocols



- IGMP snooping and MLD snooping:

  IGMP snooping and MLD snooping are multicast constraining mechanisms that run on Layer 2 devices. They manage and control multicast groups by monitoring and analyzing IGMP or MLD messages exchanged between the hosts and Layer 3 multicast devices. This effectively controls the flooding of multicast data in Layer 2 networks.

- PIM snooping and IPv6 PIM snooping:

  PIM snooping and IPv6 PIM snooping run on Layer 2 devices. They work with IGMP snooping or MLD snooping to analyze received PIM messages. Then, they add the ports that are interested in specific multicast data to a PIM snooping routing entry or IPv6 PIM snooping routing entry. In this way, multicast data can be forwarded to only the ports that are interested in the data.

- Multicast VLAN and IPv6 multicast VLAN:

  Multicast VLAN or IPv6 multicast VLAN runs on a Layer 2 device in a multicast network where multicast receivers for the same group exist in different VLANs. With these protocols, the Layer 3 multicast device sends only one copy of multicast to the multicast VLAN or IPv6 multicast VLAN on the Layer 2 device. This method avoids waste of network bandwidth and extra burden on the Layer 3 device.

Multicast packet forwarding mechanism

In a multicast model, receiver hosts of a multicast group are usually located at different areas on the network. They are identified by the same multicast group address. To deliver multicast packets to these receivers, a multicast source encapsulates the multicast data in an IP packet with the multicast group address as the destination address. Multicast routers on the forwarding paths forward multicast packets that an incoming interface receives through multiple outgoing interfaces. Compared to a unicast model, a multicast model is more complex in the following aspects:

- To ensure multicast packet transmission on the network, different routing tables are used to guide multicast forwarding. These routing tables include unicast routing tables, routing tables for multicast (for example, the MBGP routing table), and static multicast routing tables.
- To process the same multicast information from different peers received on different interfaces, the multicast device performs an RPF check on each multicast packet. The RPF check result determines whether the packet will be forwarded or discarded. The RPF check mechanism is the basis for most multicast routing protocols to implement multicast forwarding.

For more information about the RPF mechanism, see "Configuring multicast routing and forwarding" and "Configuring IPv6 multicast routing and forwarding."

IP multicast architecture

IP multicast addresses the following issues:

- Where should the multicast source transmit information to? (Multicast addressing.)
- What receivers exist on the network? (Host registration.)
- Where is the multicast source that will provide data to the receivers? (Multicast source discovery.)
- How is the information transmitted to the receivers? (Multicast routing.)

IP multicast is an end-to-end service. The multicast architecture involves the following parts:

- **Addressing mechanism**—A multicast source sends information to a group of receivers through a multicast address.
- **Host registration**—Receiver hosts can join and leave multicast groups dynamically. This mechanism is the basis for management of group memberships.
- **Multicast routing**—A multicast distribution tree (a forwarding path tree for multicast data on the network) is constructed for delivering multicast data from a multicast source to receivers.
- **Multicast applications**—A software system that supports multicast applications, such as video conferencing, must be installed on multicast sources and receiver hosts. The TCP/IP stack must support reception and transmission of multicast data.

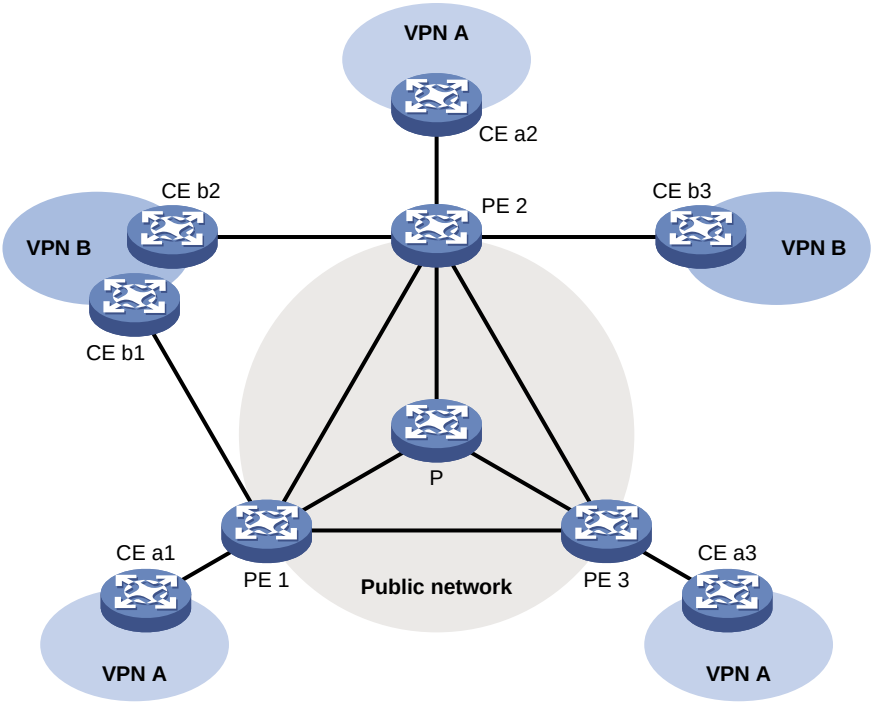
Multicast support for VPNs

Multicast support for VPNs refers to multicast applied in VPNs.

Introduction to VPN instances

VPNs are isolated from one another and from the public network. As shown in Figure 10, VPN A and VPN B separately access the public network through PE devices.

Figure 10 VPN networking diagram



- The P device belongs to the public network. The CE devices belong to their respective VPNs. Each CE device serves its own VPN and maintains only one set of forwarding mechanisms.
- The PE devices connect to the public network and the VPNs. Each PE device must strictly distinguish the information for different networks, and maintain a separate forwarding mechanism for each network. On a PE device, a set of software and hardware that serve the same network forms an instance. Multiple instances can exist on the same PE device, and an instance can reside on different PE devices. On a PE device, the instance for the public network is called the public network instance, and those for VPNs are called VPN instances.

Multicast application in VPNs

A PE device that supports multicast for VPNs performs the following operations:

- Maintains an independent set of multicast forwarding mechanisms for each VPN, including the multicast protocols, PIM neighbor information, and multicast routing table. In a VPN, the device forwards multicast data based on the forwarding table or routing table for that VPN.
- Implements the isolation between different VPNs.
- Implements information exchange and data conversion between the public network and VPN instances.

For example, as shown in Figure 10, a multicast source in VPN A sends multicast data to a multicast group. Only receivers that belong to both the multicast group and VPN A can receive the multicast data. The multicast data is multicast both in VPN A and on the public network. If you configure MVPN extranet, multicast data can be transmitted across different VPNs.

Common notations in multicast

The following notations are commonly used in multicast transmission:

- **(*, G)**—Rendezvous point tree (RPT), or a multicast packet that any multicast source sends to multicast group G. The asterisk (*) represents any multicast source, and "G" represents a specific multicast group.

- **(S, G)**—Shortest path tree (SPT), or a multicast packet that multicast source "S" sends to multicast group "G." "S" represents a specific multicast source, and "G" represents a specific multicast group.

For more information about the concepts RPT and SPT, see "PIM overview" and "IPv6 PIM overview."

Contents

Configuring IGMP snooping 1

About IGMP snooping 1
Fundamentals of IGMP snooping 1
IGMP snooping ports 1
How IGMP snooping works 3
IGMP snooping proxying 4
Protocols and standards 5
Restrictions and guidelines: IGMP snooping configuration 6
VLAN-based IGMP snooping tasks at a glance 6
VSI-based IGMP snooping tasks at a glance 7
Enabling the IGMP snooping feature 8
Enabling IGMP snooping 8
Enabling IGMP snooping globally 8
Enabling IGMP snooping for VLANs 9
Enabling IGMP snooping for a VSI 9
Configuring basic IGMP snooping features 10
Specifying an IGMP snooping version 10
Setting the maximum number of IGMP snooping forwarding entries 11
Configuring static multicast MAC address entries 11
Setting the IGMP last member query interval 12
Configuring IGMP snooping port features 13
Setting aging timers for dynamic ports 13
Configuring a static member port 14
Configuring a static router port 14
Configuring a port as a simulated member host 15
Enabling fast-leave processing 15
Disabling a port from becoming a dynamic router port 16
Preventing the device from forwarding multicast data to router ports in a VLAN 17
Specifying the Layer 2 multicast forwarding mode 17
About Layer 2 multicast forwarding modes 17
Restrictions and guidelines for specifying the Layer 2 multicast forwarding mode 17
Specifying the Layer 2 multicast forwarding mode for a VLAN 18
Configuring the IGMP snooping querier 18
Enabling the IGMP snooping querier 18
Enabling IGMP snooping querier election 19
Configuring parameters for IGMP general queries and responses 19
Enabling the device to send IGMP general queries upon a path change 20
Enabling the device to send IGMP general queries upon a port state change 21
Enabling IGMP snooping proxying 22
About enabling IGMP snooping proxying 22
Restrictions and guidelines for enabling IGMP snooping proxying 22
Enabling IGMP snooping proxying for a VLAN 23
Enabling IGMP snooping proxying for a VSI 23
Configuring parameters for IGMP messages 23
Configuring source IP addresses for IGMP messages 23
Setting the 802.1p priority for IGMP messages 25
Configuring IGMP snooping policies 25
Configuring a multicast group policy 25
Enabling dropping unknown multicast data 26
Enabling IGMP report suppression 27
Setting the maximum number of multicast groups on a port 27
Enabling multicast group replacement 27
Enabling host tracking 28
Setting the DSCP value for outgoing IGMP protocol packets 29
Enabling SNMP notifications for IGMP snooping 29
Configuring IGMP snooping logging 30
Enabling IGMP snooping logging 30

Enabling IGMP snooping logging suppression ····30
Display and maintenance commands for IGMP snooping ····30
IGMP snooping configuration examples ····32
Example: Configuring VLAN-based IGMP snooping group polices and simulated joining ····32
Example: Configuring VLAN-based IGMP snooping static ports ····35
Example: Configuring the VLAN-based IGMP snooping querier ····41
Example: Configuring VLAN-based IGMP snooping proxying ····46
Troubleshooting IGMP snooping ····49
Layer 2 multicast forwarding cannot function ····49
Multicast group policy does not work ····49

Configuring IGMP snooping

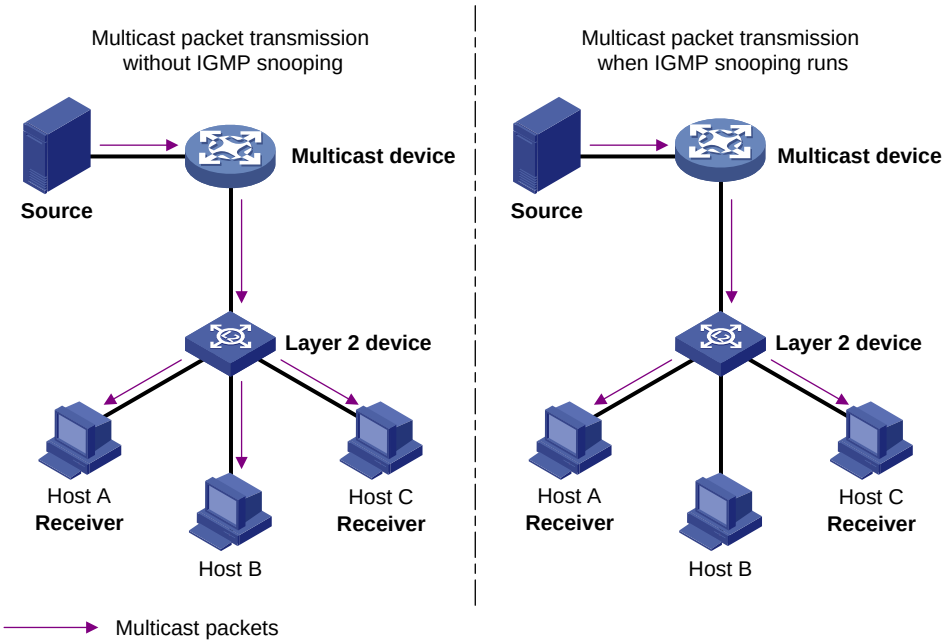
About IGMP snooping

IGMP snooping runs on a Layer 2 device as a multicast constraining mechanism to improve multicast forwarding efficiency. It creates Layer 2 multicast forwarding entries from IGMP packets that are exchanged between the hosts and the Layer 3 device.

Fundamentals of IGMP snooping

As shown in Figure 1, when IGMP snooping is not enabled, the Layer 2 switch floods multicast packets to all hosts in a VLAN or VSI. When IGMP snooping is enabled, the Layer 2 switch forwards multicast packets of known multicast groups to only the receivers.

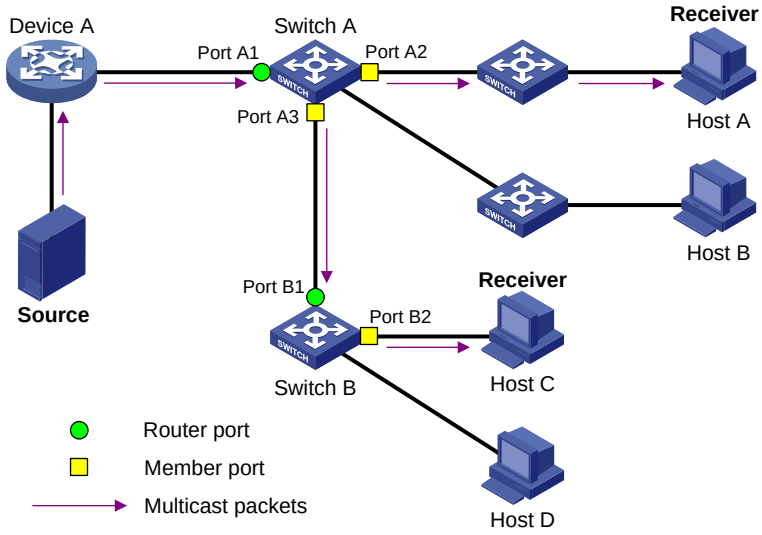
Figure 1 Multicast packet transmission without and with IGMP snooping



IGMP snooping ports

As shown in Figure 2, IGMP snooping runs on Switch A and Switch B, and Host A and Host C are receivers in a multicast group. IGMP snooping ports are divided into member ports and router ports.

Figure 2 IGMP snooping ports



Router ports

On an IGMP snooping Layer 2 device, the ports toward Layer 3 multicast devices are called router ports. In Figure 2, Port A1 of Switch A and Port B1 of Switch B are router ports.

Router ports contain the following types:

- **Dynamic router port**—When a port receives an IGMP general query whose source address is not 0.0.0.0 or receives a PIM hello message, the port is added into the dynamic router port list. At the same time, an aging timer is started for the port. If the port receives either of the messages before the timer expires, the timer is reset. If the port does not receive an IGMP general query whose source address is not 0.0.0.0 or a PIM hello message when the timer expires, the port is removed from the dynamic router port list.
- **Static router port**—When a port is statically configured as a router port, it is added into the static router port list. The static router port does not age out, and it can be deleted only manually.

Do not confuse the "router port" in IGMP snooping with the "routed interface" commonly known as the "Layer 3 interface." The router port in IGMP snooping is a Layer 2 interface.

Member ports

On an IGMP snooping Layer 2 device, the ports toward receiver hosts are called member ports. In Figure 2, Port A2 and Port A3 of Switch A and Port B2 of Switch B are member ports.

Member ports contain the following types:

- **Dynamic member port**—When a port receives an IGMP report, it is added to the associated dynamic IGMP snooping forwarding entry as an outgoing interface. At the same time, an aging timer is started for the port. If the port receives an IGMP report before the timer expires, the timer is reset. If the port does not receive an IGMP report when the timer expires, the port is removed from the associated dynamic forwarding entry.
- **Static member port**—When a port is statically configured as a member port, it is added to the associated static IGMP snooping forwarding entry as an outgoing interface. The static member port does not age out, and it can be deleted only manually.

Unless otherwise specified, router ports and member ports in this document include both static and dynamic router ports and member ports.

How IGMP snooping works

The ports in this section are dynamic ports. For information about how to configure and remove static ports, see "Configuring a static member port" and "Configuring a static router port."

IGMP messages types include general query, IGMP report, and leave message. An IGMP snooping-enabled Layer 2 device performs differently depending on the message types.

General query

The IGMP querier periodically sends IGMP general queries to all hosts and devices on the local subnet to check for the existence of multicast group members.

After receiving an IGMP general query, the Layer 2 device forwards the query to all ports in the VLAN or VSI except the receiving port. The Layer 2 device also performs one of the following actions:

- If the receiving port is a dynamic router port in the dynamic router port list, the Layer 2 device restarts the aging timer for the port.
- If the receiving port does not exist in the dynamic router port list, the Layer 2 device adds the port to the dynamic router port list. It also starts an aging timer for the port.

IGMP report

A host sends an IGMP report to the IGMP querier for the following purposes:

- Responds to queries if the host is a multicast group member.
- Applies for a multicast group membership.

After receiving an IGMP report from a host, the Layer 2 device forwards the report through all the router ports in the VLAN or VSI. It also resolves the address of the reported multicast group, and looks up the forwarding table for a matching entry as follows:

- If no match is found, the Layer 2 device creates a forwarding entry with the receiving port as an outgoing interface. It also marks the receiving port as a dynamic member port and starts an aging timer for the port.
- If a match is found but the matching forwarding entry does not contain the receiving port, the Layer 2 device adds the receiving port to the outgoing interface list. It also marks the receiving port as a dynamic member port and starts an aging timer for the port.
- If a match is found and the matching forwarding entry contains the receiving port, the Layer 2 device restarts the aging timer for the port.

NOTE:

A Layer 2 device does not forward an IGMP report through a non-router port because of the host IGMP report suppression mechanism. If a non-router port has member host attached, the member hosts suppress their IGMP reports upon receiving IGMP reports forwarded by the non-router port. The Layer 2 device cannot know the existence of the member hosts attached to the non-router port. For more information about the IGMP report suppression mechanism, see "Configuring IGMP."

Leave message

An IGMPv1 receiver host does not send any leave messages when it leaves a multicast group. The Layer 2 device cannot immediately update the status of the port that connects to the receiver host. The Layer 2 device does not remove the port from the outgoing interface list in the associated forwarding entry until the aging time for the group expires.

An IGMPv2 or IGMPv3 host sends an IGMP leave message when it leaves a multicast group.

When the Layer 2 device receives an IGMP leave message on a dynamic member port, the Layer 2 device first examines whether a forwarding entry matches the group address in the message.

- If no match is found, the Layer 2 device discards the IGMP leave message.

- If a match is found but the receiving port is not an outgoing interface in the forwarding entry, the Layer 2 device discards the IGMP leave message.
- If a match is found and the receiving port is not the only outgoing interface in the forwarding entry, the Layer 2 device performs the following actions:
  - Discards the IGMP leave message.
  - Sends an IGMP group-specific query to identify whether the group has active receivers attached to the receiving port.
  - Sets the aging timer for the receiving port to twice the IGMP last member query interval.
- If a match is found and the receiving port is the only outgoing interface in the forwarding entry, the Layer 2 device performs the following actions:
  - Forwards the IGMP leave message to all router ports in the VLAN or VSI.
  - Sends an IGMP group-specific query to identify whether the group has active receivers attached to the receiving port.
  - Sets the aging timer for the receiving port to twice the IGMP last member query interval.

After receiving the IGMP leave message on a port, the IGMP querier resolves the multicast group address in the message. Then, it sends an IGMP group-specific query to the multicast group through the receiving port.

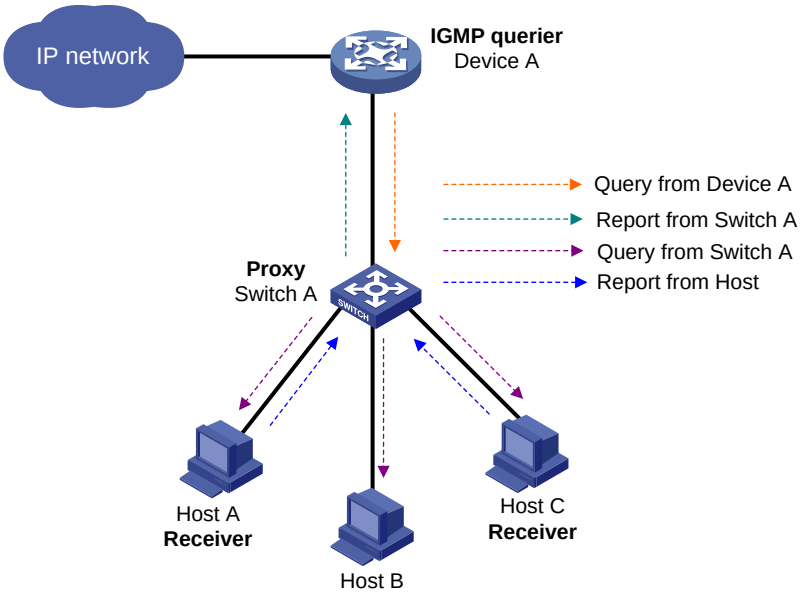
After receiving the IGMP group-specific query, the Layer 2 device forwards the query through all router ports and member ports of the group in the VLAN or VSI. Then, it waits for the responding IGMP report from the directly connected hosts. For the dynamic member port that received the leave message, the Layer 2 device also performs one of the following actions:

- If the port receives an IGMP report before the aging timer expires, the Layer 2 device resets the aging timer.
- If the port does not receive an IGMP report when the aging timer expires, the Layer 2 device removes the port from the forwarding entry for the multicast group.

IGMP snooping proxying

As shown in Figure 3, to reduce the number of IGMP report and leave messages received by the upstream device, you can enable IGMP snooping proxying on the edge device. With IGMP snooping proxying enabled, the edge device acts as a host for the upstream IGMP snooping querier to send IGMP report and leave messages to Device A. The host IGMP report suppression mechanism on the edge device does not take effect. For more information about the IGMP report suppression mechanism, see "Configuring IGMP."

Figure 3 IGMP snooping proxying



The IGMP snooping proxy device processes different IGMP messages as follows:

- General query.

  After receiving an IGMP general query, the device forwards the query to all ports in the VLAN or VSI except the receiving port. The device also generates an IGMP report based on the local membership information and sends the report to all router ports.

- Group-specific query or group-and-source-specific query.

  After receiving an IGMP group-specific query or group-and-source-specific query, the device forwards the query to all ports in the VLAN or VSI except the receiving port. If the forwarding entry has a member port, the device sends a report to all router ports in the VLAN or VSI.

- Report.

  After receiving an IGMP report from a host, the device looks up the forwarding table for a matching entry as follows:

  - If a match is found and the matching forwarding entry contains the receiving port, the device resets the aging timer for the port.
  - If a match is found but the matching forwarding entry does not contain the receiving port, the device adds the receiving port to the outgoing interface list. It also marks the receiving port as a dynamic member port and starts an aging timer for the port.
  - If no match is found, the device creates a forwarding entry with the receiving port as an outgoing interface. It also marks the receiving port as a dynamic member port and starts an aging timer for the port. Then it sends the report to all router ports.

- Leave message.

  After receiving the IGMP leave message on a port, the device sends an IGMP group-specific query through the receiving port. The device sends the IGMP leave message to all router ports only when the last member port is removed from the forwarding entry.

Protocols and standards

RFC 4541, *Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches*

Restrictions and guidelines: IGMP snooping configuration

The IGMP snooping configurations made on Layer 2 aggregate interfaces do not interfere with the configurations made on member ports. In addition, the configurations made on Layer 2 aggregate interfaces do not take part in aggregation calculations. The configuration made on a member port of the aggregate group takes effect after the port leaves the aggregate group.

Some features can be configured for a VLAN in VLAN view or for multiple VLANs in IGMP-snooping view. The configuration made in VLAN view and the configuration made in IGMP-snooping view have the same priority, and the most recent configuration takes effect.

Some features can be configured for a VLAN in VLAN view or globally for all VLANs in IGMP-snooping view. The VLAN-specific configuration takes priority over the global configuration.

Some features can be configured for a VSI in VSI view or globally for all VSIs in IGMP-snooping view. The VSI-specific configuration takes priority over the global configuration.

Some features can be configured for an interface in interface view or for all interfaces of the specified VLANs in IGMP-snooping view. The interface-specific configuration takes priority over the configuration made in IGMP-snooping view.

VLAN-based IGMP snooping tasks at a glance

To configure IGMP snooping for VLANs, perform the following tasks:

1. Enabling the IGMP snooping feature
2. Enabling IGMP snooping

   Choose the following tasks as needed:
   - Enabling IGMP snooping globally
   - Enabling IGMP snooping for VLANs
3. (Optional.) Configuring basic IGMP snooping features
   - Specifying an IGMP snooping version
   - Setting the maximum number of IGMP snooping forwarding entries
   - Configuring static multicast MAC address entries
   - Setting the IGMP last member query interval
4. (Optional.) Configuring IGMP snooping port features
   - Setting aging timers for dynamic ports
   - Configuring a static member port
   - Configuring a static router port
   - Configuring a port as a simulated member host
   - Enabling fast-leave processing
   - Disabling a port from becoming a dynamic router port
   - Preventing the device from forwarding multicast data to router ports in a VLAN
5. (Optional.) Specifying the Layer 2 multicast forwarding mode
6. (Optional.) Configuring the IGMP snooping querier
   - Enabling the IGMP snooping querier
   - Enabling IGMP snooping querier election
   - Configuring parameters for IGMP general queries and responses
   - Enabling the device to send IGMP general queries upon a path change

- Enabling the device to send IGMP general queries upon a port state change

7. (Optional.) Enabling IGMP snooping proxying
8. (Optional.) Configuring parameters for IGMP messages
   - Configuring source IP addresses for IGMP messages
   - Setting the 802.1p priority for IGMP messages
9. (Optional.) Configuring IGMP snooping policies
   - Configuring a multicast group policy
   - Enabling dropping unknown multicast data
   - Enabling IGMP report suppression
   - Setting the maximum number of multicast groups on a port
   - Enabling multicast group replacement
   - Enabling host tracking
10. (Optional.) Setting the DSCP value for outgoing IGMP protocol packets
11. (Optional.) Enabling SNMP notifications for IGMP snooping
12. (Optional.) Configuring IGMP snooping logging
    - Enabling IGMP snooping logging
    - Enabling IGMP snooping logging suppression

VSI-based IGMP snooping tasks at a glance

To configure IGMP snooping for VSIs, perform the following tasks:

1. Enabling the IGMP snooping feature
2. Enabling IGMP snooping

   Choose the following tasks as needed:
   - Enabling IGMP snooping globally
   - Enabling IGMP snooping for a VSI
3. (Optional.) Configuring basic IGMP snooping features
   - Specifying an IGMP snooping version
   - Setting the maximum number of IGMP snooping forwarding entries
   - Setting the IGMP last member query interval
4. (Optional.) Setting aging timers for dynamic ports
5. (Optional.) Configuring the IGMP snooping querier
   - Enabling the IGMP snooping querier
   - Enabling IGMP snooping querier election
   - Configuring parameters for IGMP general queries and responses
   - Enabling the device to send IGMP general queries upon a path change
6. (Optional.) Enabling IGMP snooping proxying
7. (Optional.) Configuring parameters for IGMP messages
   - Configuring source IP addresses for IGMP messages
   - Setting the 802.1p priority for IGMP messages
8. (Optional.) Configuring IGMP snooping policies
   - Enabling dropping unknown multicast data
   - Enabling IGMP report suppression
9. (Optional.) Setting the DSCP value for outgoing IGMP protocol packets

10. (Optional.) Enabling SNMP notifications for IGMP snooping
11. (Optional.) Configuring IGMP snooping logging
    - Enabling IGMP snooping logging
    - Enabling IGMP snooping logging suppression

Enabling the IGMP snooping feature

About this task

You must enable the IGMP snooping feature before you configure other IGMP snooping features.

Procedure

1. Enter system view.
   `system-view`
2. Enable the IGMP snooping feature and enter IGMP-snooping view.
   `igmp-snooping`
   By default, the IGMP snooping feature is disabled.

Enabling IGMP snooping

Enabling IGMP snooping globally

About this task

After you enable IGMP snooping globally, IGMP snooping is enabled for all VLANs or VSIs. You can disable IGMP snooping for a VLAN or VSI when IGMP snooping is globally enabled.

Restrictions and guidelines

To configure other IGMP snooping features for VLANs or VSIs, you must enable IGMP snooping for the specific VLANs or VSIs even though IGMP snooping is enabled globally.

The VLAN-specific or VSI-specific IGMP snooping configuration takes priority over the global IGMP snooping configuration. For example, if you enable IGMP snooping globally and then use the **`igmp-snooping disable`** command to disable IGMP snooping for a VLAN, IGMP snooping is disabled in the VLAN.

Procedure

1. Enter system view.
   `system-view`
2. Enter IGMP-snooping view.
   `igmp-snooping`
3. Enable IGMP snooping globally.
4. **`global-enable`**
   By default, IGMP snooping is disabled globally.
5. (Optional.) Disable IGMP snooping for a VLAN or VSI.
   a. Return to system view.
      `quit`
   b. Enter VLAN view.
      `vlan` *`vlan-id`*

c. Disable IGMP snooping for the VLAN or VSI.

igmp-snooping disable

By default, the IGMP snooping status in a VLAN or VSI is consistent with the global IGMP snooping status.

Enabling IGMP snooping for VLANs

Restrictions and guidelines

You can enable IGMP snooping for multiple VLANs by using the **enable vlan** command in IGMP-snooping view or for a VLAN by using the **igmp-snooping enable** command in VLAN view. The configuration in VLAN view has the same priority as the configuration in IGMP-snooping view.

IGMP snooping configuration in a VLAN takes effect only on the member ports in the VLAN.

Enabling IGMP snooping for multiple VLANs

1. Enter system view.

   system-view
2. Enter IGMP-snooping view.

   igmp-snooping
3. Enable IGMP snooping for the specified VLANs.

   enable vlan *vlan-list*

   By default, the IGMP snooping status in a VLAN is consistent with the global IGMP snooping status.

Enabling IGMP snooping for a VLAN

1. Enter system view.

   system-view
2. Enter VLAN view.

   vlan *vlan-id*
3. Enable IGMP snooping for the VLAN.

   igmp-snooping enable

   By default, the IGMP snooping status in a VLAN is consistent with the global IGMP snooping status.

Enabling IGMP snooping for a VSI

Restrictions and guidelines

For VSIs, you can only enable IGMP snooping for a VSI by using the **igmp-snooping enable** command in VSI view.

IGMP snooping configuration in a VSI takes effect only on the member ports in the VSI.

On an EVPN M-LAG network, you must enable both IGMP snooping and IGMP snooping proxying for VSIs. For more information about EVPN M-LAG, see EVPN VXLAN configuration in *EVPN Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter VSI view.

vsi *vsi-name*

3. Enable IGMP snooping for the VSI.

   igmp-snooping enable

   By default, the IGMP snooping status in a VSI is consistent with the global IGMP snooping status.

Configuring basic IGMP snooping features

Specifying an IGMP snooping version

About this task

Different IGMP snooping versions process different versions of IGMP messages.

- IGMPv2 snooping processes IGMPv1 messages, IGMPv2 messages, and IGMPv3 queries, but it floods IGMPv3 reports in the VLAN or VSI instead of processing them.
- IGMPv3 snooping processes IGMPv1, IGMPv2, and IGMPv3 messages.

Restrictions and guidelines

If you change the IGMP snooping version from 2 to 3, the device performs the following actions:

- Clears all IGMP snooping forwarding entries that are dynamically created.
- Keeps static IGMPv3 snooping forwarding entries (*, G).
- Clears static IGMPv3 snooping forwarding entries (S, G), which will be restored when the IGMP snooping version is switched back to 3.

For more information about static IGMP snooping forwarding entries, see "Configuring a static member port."

Specifying an IGMP snooping version for multiple VLANs

1. Enter system view.

   system-view

2. Enter IGMP-snooping view.

   igmp-snooping

3. Specify an IGMP snooping version for multiple VLANs.

   version *version-number* **vlan** *vlan-list*

   By default, the IGMP snooping version for a VLAN is 2.

Specifying an IGMP snooping version for a VLAN

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Specify an IGMP snooping version for the VLAN.

   igmp-snooping version *version-number*

   By default, the IGMP snooping version for a VLAN is 2.

Specifying an IGMP snooping version for a VSI

1. Enter system view.

   system-view

2. Enter VSI view.

vsi *vsi-name*

3. Specify an IGMP snooping version for the VSI.
   igmp-snooping version *version-number*
   By default, the IGMP snooping version for a VSI is 2.

Setting the maximum number of IGMP snooping forwarding entries

About this task

You can modify the maximum number of IGMP snooping forwarding entries, including dynamic entries and static entries. When the number of forwarding entries on the device reaches the upper limit, the device does not automatically remove any existing entries. To allow new entries to be created, remove some entries manually.

Procedure

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Set the maximum number of IGMP snooping forwarding entries.
   entry-limit *limit*
   By default, the maximum number of IGMP snooping forwarding entries is 4294967295.

Configuring static multicast MAC address entries

About this task

In Layer 2 multicast, multicast MAC address entries can be dynamically created through Layer 2 multicast protocols (such as IGMP snooping). You can also manually configure static multicast MAC address entries by binding multicast MAC addresses and ports to control the destination ports of the multicast data.

Restrictions and guidelines

You must specify an unused multicast MAC address when configuring a static multicast MAC address entry. A multicast MAC address is the MAC address in which the least significant bit of the most significant octet is 1.

Configuring a static multicast MAC address entry in system view

1. Enter system view.
   system-view
2. Configure a static multicast MAC address entry.
   mac-address multicast *mac-address* **interface** *interface-list* **vlan** *vlan-id*

Configuring a static multicast MAC address entry in interface view

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*

- Enter Layer 2 aggregate interface view.
  interface bridge-aggregation *interface-number*

3. Configure a static multicast MAC address entry.
   mac-address multicast *mac-address* **vlan** *vlan-id*

Setting the IGMP last member query interval

About this task

A receiver host starts a report delay timer for a multicast group when it receives an IGMP group-specific query for the group. This timer is set to a random value in the range of 0 to the maximum response time advertised in the query. When the timer value decreases to 0, the host sends an IGMP report to the group.

The IGMP last member query interval defines the maximum response time advertised in IGMP group-specific queries. Set an appropriate value for the IGMP last member query interval to speed up hosts' responses to IGMP group-specific queries and avoid IGMP report traffic bursts.

Setting the IGMP last member query interval globally

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Set the IGMP last member query interval globally.
   last-member-query-interval *interval*

   By default, the global IGMP last member query interval is 1 second.

Setting the IGMP last member query interval for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Set the IGMP last member query interval for the VLAN.
   igmp-snooping last-member-query-interval *interval*

   By default, the IGMP last member query interval is 1 second for a VLAN.

Setting the IGMP last member query interval for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Set the IGMP last member query interval for the VSI.
   igmp-snooping last-member-query-interval *interval*

   By default, the IGMP last member query interval is 1 second for a VSI.

Configuring IGMP snooping port features

Setting aging timers for dynamic ports

About this task

A dynamic router port is removed from the dynamic router port list if it does not receive an IGMP general query or PIM hello message when its aging timer expires.

A dynamic member port is removed from the dynamic member port if it does not receive an IGMP report when its aging timer expires.

Restrictions and guidelines

Set an appropriate value for the aging timers of dynamic ports. For example, if the memberships of multicast groups frequently change, set a relatively small value for the aging timer of the dynamic member ports. If the memberships of multicast groups rarely change, set a relatively large value.

If a dynamic router port receives a PIMv2 hello message, the aging timer for the port is specified by the hello message. In this case, the **router-aging-time** or **igmp-snooping router-aging-time** command does not take effect on the port.

IGMP group-specific queries originated by the Layer 2 device trigger the adjustment of aging timers for dynamic member ports. If a dynamic member port receives such a query, its aging timer is set to twice the IGMP last member query interval. For more information about setting the IGMP last member query interval on the Layer 2 device, see "Setting the IGMP last member query interval."

Setting the aging timers for dynamic ports globally

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Set the aging timer for dynamic router ports globally.
   router-aging-time *seconds*
   By default, the aging timer for dynamic router ports is 260 seconds.
4. Set the global aging timer for dynamic member ports globally.
   host-aging-time *seconds*
   By default, the aging timer for dynamic member ports is 260 seconds.

Setting the aging timers for dynamic ports in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Set the aging timer for dynamic router ports in the VLAN.
   igmp-snooping router-aging-time *seconds*
   By default, a VLAN uses the global aging timer for dynamic router ports.
4. Set the aging timer for dynamic member ports in the VLAN.
   igmp-snooping host-aging-time *seconds*
   By default, a VLAN uses the global aging timer for dynamic member ports.

Setting the aging timers for dynamic ports in a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Set the aging timer for dynamic router ports in the VSI.
   igmp-snooping router-aging-time *seconds*
   By default, a VSI uses the global aging timer for dynamic router ports.
4. Set the aging timer for dynamic member ports in the VSI.
   igmp-snooping host-aging-time *seconds*
   By default, a VSI uses the global aging timer for dynamic member ports.

Configuring a static member port

About this task

You can configure a port as a static member port for a multicast group so that hosts attached to the port can always receive multicast for the group. The static member port does not respond to IGMP queries. When you complete or cancel this configuration on a port, the port does not send an unsolicited IGMP report or leave message.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Configure the port as a static member port.
   igmp-snooping static-group *group-address* [**source-ip** *source-address*] { **vlan** *vlan-id* | **service-instance** *instance-id* }
   By default, a port is not a static member port.

Configuring a static router port

About this task

You can configure a port as a static router port for a multicast group so that all multicast data for the group received on the port will be forwarded.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.

interface bridge-aggregation *interface-number*

3. Configure the port as a static router port.

   igmp-snooping static-router-port vlan *vlan-id*

   By default, a port is not a static router port.

Configuring a port as a simulated member host

About this task

When a port is configured as a simulated member host, it is equivalent to an independent host in the following ways:

- It sends an unsolicited IGMP report when you complete the configuration.
- It responds to IGMP general queries with IGMP reports.
- It sends an IGMP leave message when you cancel the configuration.

The version of IGMP running on the simulated member host is the same as the version of IGMP snooping running on the port. The port ages out in the same way as a dynamic member port.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*

3. Configure the port as a simulated member host.

   igmp-snooping host-join *group-address* [**source-ip** *source-address*] **vlan** *vlan-id*

   By default, the port is not a simulated member host.

Enabling fast-leave processing

About this task

This feature enables the Layer 2 device to immediately remove a port from the forwarding entry for a multicast group when the port receives a leave massage. The Layer 2 device no longer sends or forwards IGMP group-specific queries for the group to the port.

Restrictions and guidelines

Do not enable fast-leave processing on a port that has multiple receiver hosts in a VLAN. If you do so, the remaining receivers cannot receive multicast data for a group after a receiver leaves the group.

On a network where both Layer 2 multicast and Layer 3 multicast are enabled, when multicast group members on a port need to leave, you only need to enable fast-leave processing on the Layer 2 port. To avoid multicast service interruption on the other ports in the same VLAN, do not enable fast-leave processing for Layer 3 multicast group members at the same time (by using the **igmp fast leave** command). For more information about the **igmp fast leave** command, see IGMP commands in *IP Multicast Command Reference*.

Enabling fast-leave processing globally

1. Enter system view.

system-view

2. Enter IGMP-snooping view.
   igmp-snooping
3. Enable fast-leave processing globally.
   fast-leave [**vlan** *vlan-list*]
   By default, fast-leave processing is disabled globally.

Enabling fast-leave processing on a port

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Enable fast-leave processing on the port.
   igmp-snooping fast-leave [**vlan** *vlan-list*]
   By default, fast-leave processing is disabled on a port.

Disabling a port from becoming a dynamic router port

About this task

A receiver host might send IGMP general queries or PIM hello messages for testing purposes. On the Layer 2 device, the port that receives either of the messages becomes a dynamic router port. Before the aging timer for the port expires, the following problems might occur:

- All multicast data for the VLAN to which the port belongs flows to the port. Then, the port forwards the data to attached receiver hosts. The receiver hosts will receive multicast data that it does not want to receive.
- The port forwards the IGMP general queries or PIM hello messages to its upstream Layer 3 devices. These messages might affect the multicast routing protocol state (such as the IGMP querier or M-LAG election) on the Layer 3 devices. This might further cause network interruption.

To solve these problems, you can disable a port from becoming a dynamic router port. This also improves network security and the control over receiver hosts.

Restrictions and guidelines

This configuration and the static router port configuration do not interfere with each other.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Disable the port from becoming a dynamic router port.

igmp-snooping router-port-deny [**vlan** *vlan-list*]

By default, a port is allowed to become a dynamic router port.

Preventing the device from forwarding multicast data to router ports in a VLAN

About this task

If all ports on an IGMP snooping Layer 2 device are statically configured as router ports, the Layer 2 device will forwards multicast data from Layer 3 multicast devices to all its router ports. This wastes bandwidth resources. Perform this task to not forward multicast data to router ports in a VLAN if they are not interested in receiving the data.

Prerequisites

Before performing this task, you must enable IGMP snooping in the VLAN.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Prevent the device from forwarding multicast data to static router ports in the VLAN.
   igmp-snooping router-port-discard
   By default, the device forwards multicast data to static router ports in a VLAN.

Specifying the Layer 2 multicast forwarding mode

About Layer 2 multicast forwarding modes

The device supports the following Layer 2 multicast forwarding modes:

- **MAC-based mode**—The device forwards Layer 2 multicast packets based on the multicast MAC addresses in the packets.
- **IP-based mode**—The device forwards Layer 2 multicast packets based on the IP addresses (multicast source addresses and multicast group addresses) in the packets.

To display Layer 2 multicast IP forwarding entries, use the **display l2-multicast ip forwarding** command.

To display Layer 2 multicast MAC forwarding entries, use the **display l2-multicast mac forwarding** command.

Restrictions and guidelines for specifying the Layer 2 multicast forwarding mode

Multiple multicast IP addresses might be mapped to the same multicast MAC address. This causes that non-members of a multicast group receive the multicast data of this group if the device uses the MAC-based forwarding mode. As a best practice, specify the IP-based Layer 2 multicast forwarding mode on Layer 3 multicast devices.

As a best practice, specify the IP-based Layer 2 multicast forwarding mode on the device if a device running IGMPv3 exists in the network.

Specifying the Layer 2 multicast forwarding mode for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Specify the Layer 2 multicast forwarding mode.
   igmp-snooping forwarding-mode { **ip** | **mac** }
   By default, the device forwards Layer 2 multicast packets based on the MAC addresses in the packets.

Configuring the IGMP snooping querier

Enabling the IGMP snooping querier

About this task

This feature enables the Layer 2 device to periodically send IGMP general queries to establish and maintain multicast forwarding entries at the data link Layer. You can configure an IGMP snooping querier on a network without Layer 3 multicast devices.

Restrictions and guidelines

Do not enable the IGMP snooping querier on a multicast network that runs IGMP. An IGMP snooping querier does not take part in IGMP querier elections. However, it might affect IGMP querier elections if it sends IGMP general queries with a low source IP address. For more information about IGMP querier, see "Configuring IGMP."

On a VXLAN network, the IGMP snooping querier in a VSI does not include VLAN tags in IGMP general queries. As a best practice, do not enable the IGMP snooping querier in a VSI if the VSI uses the Ethernet access mode. For more information about the Ethernet access mode, see *VXLAN Configuration Guide*.

Enabling the IGMP snooping querier for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable the IGMP snooping querier for the VLAN.
   igmp-snooping querier
   By default, the IGMP snooping querier is disabled for a VLAN.

Enabling the IGMP snooping querier for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable the IGMP snooping querier for the VSI.
   igmp-snooping querier

By default, the IGMP snooping querier is disabled for a VSI.

Enabling IGMP snooping querier election

About this task

To avoid traffic interruption caused by the failure of a single querier in a VLAN or VSI, configure multiple queriers in the VLAN or VSI and enable querier election. When the elected querier fails, the device starts a new querier election to ensure multicast forwarding. The mechanism for IGMP snooping querier election is the same as that for IGMP querier election. For more information about IGMP querier election, see "Configuring IGMP."

Prerequisites for enabling IGMP snooping querier election

Before you enable IGMP snooping querier election, you must complete the following tasks:

- Enable the IGMP snooping querier for a VLAN or VSI. For more information about enabling the IGMP snooping querier, see "Enabling the IGMP snooping querier."
- Configure the source IP address for IGMP general queries as an IP address different from 0.0.0.0 and the local querier IP address. An IGMP snooping querier performs querier election only if the source IP address of a received IGMP general query is not 0.0.0.0 or its own IP address.
- Make sure the candidate IGMP snooping queriers run the same IGMP snooping version. To specify the IGMP snooping version, use the **igmp-snooping version** command.

Enabling IGMP snooping querier election for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable IGMP snooping querier election for the VLAN.
   igmp-snooping querier-election
   By default, IGMP snooping querier election is disabled for a VLAN.

Enabling IGMP snooping querier election for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable IGMP snooping querier election for the VSI.
   igmp-snooping querier-election
   By default, IGMP snooping querier election is disabled for a VSI.

Configuring parameters for IGMP general queries and responses

About this task

You can modify the IGMP general query interval based on the actual condition of the network.

A receiver host starts a report delay timer for each multicast group that it has joined when it receives an IGMP general query. This timer is set to a random value in the range of 0 to the maximum response time advertised in the query. When the timer value decreases to 0, the host sends an IGMP report to the corresponding multicast group.

Set an appropriate value for the maximum response time for IGMP general queries to speed up hosts' responses to IGMP general queries and avoid IGMP report traffic bursts.

Restrictions and guidelines

To avoid mistakenly deleting multicast group members, make sure the IGMP general query interval is greater than the maximum response time for IGMP general queries.

Configuring parameters for IGMP general queries and responses globally

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Set the maximum response time for IGMP general queries.
   max-response-time *seconds*
   By default, the maximum response time for IGMP general queries is 10 seconds.

Configuring parameters for IGMP general queries and responses in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Set the IGMP general query interval in the VLAN.
   igmp-snooping query-interval *interval*
   By default, the IGMP general query interval is 125 seconds for a VLAN.
4. Set the maximum response time for IGMP general queries in the VLAN.
   igmp-snooping max-response-time *seconds*
   By default, the maximum response time for IGMP general queries is 10 seconds for a VLAN.

Configuring parameters for IGMP general queries and responses in a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Set the IGMP general query interval in the VSI.
   igmp-snooping query-interval *interval*
   By default, the IGMP general query interval is 125 seconds for a VSI.
4. Set the maximum response time for IGMP general queries in the VSI.
   igmp-snooping max-response-time *seconds*
   By default, the maximum response time for IGMP general queries is 10 seconds for a VSI.

Enabling the device to send IGMP general queries upon a path change

About this task

Perform this task to enable the device to immediately send an IGMP general query to all ports in the affected VLAN or VSI when a path changes. This feature ensures that the device can update the

member port information in multicast forwarding entries and switch multicast traffic to the new path promptly.

Restrictions and guidelines

Perform this task on the IGMP snooping querier in a network of which the topology frequently changes.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enter IGMP-snooping view.
   ```
   igmp-snooping
   ```
3. Enable the device to send IGMP general queries upon a path change.
   ```
   send-query enable
   ```
   By default, the device does not send IGMP general queries upon a path change.

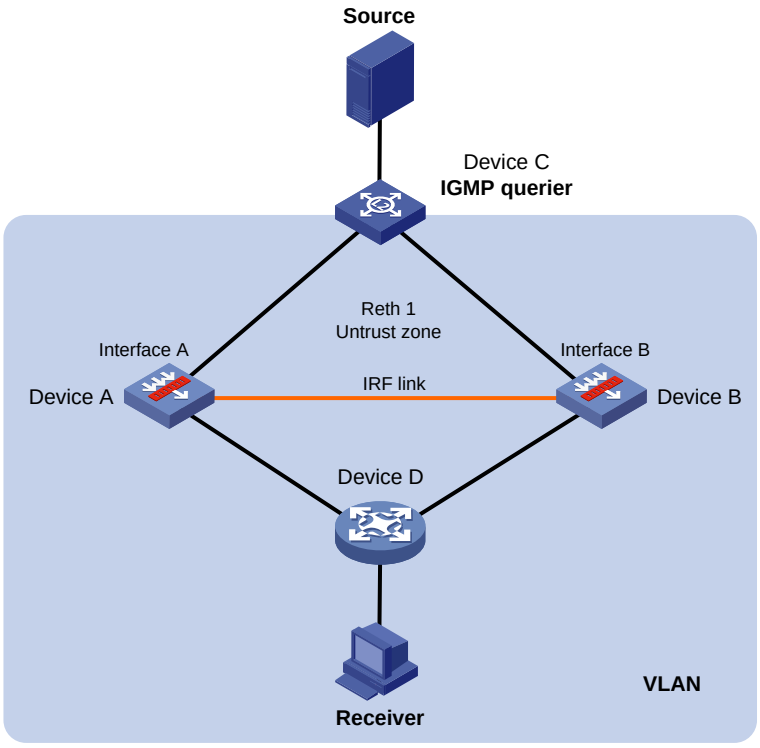
Enabling the device to send IGMP general queries upon a port state change

About this task

As shown in Figure 4, the receiver side's devices connect to the Layer 2 multicast device through a Reth interface for high availability. Interface A and interface B are member interfaces of the Reth interface, and interface A is active. Multicast data is sent to the receivers through Device A. When interface A fails, Device C does not send an IGMP general query immediately and cannot update interface B to multicast forwarding entries.

Perform this task to enable the device to immediately send an IGMP general query to all ports in the affected VLAN when the state of a port changes. This feature ensures that the device can update the member port information in multicast forwarding entries and switch multicast traffic to the new link promptly.

Figure 4 Network diagram



Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable the device to send IGMP general queries upon a port state change.
   igmp-snooping triggered-query enable { **port-down** | **port-up** } *
   By default, the device does not send IGMP general queries upon a port state change.

Enabling IGMP snooping proxying

About enabling IGMP snooping proxying

The device enabled with IGMP snooping proxying is called an IGMP snooping proxy. The IGMP snooping proxy acts as a host to the upstream device. Enabled with IGMP snooping querier, the IGMP snooping proxy acts as the router to downstream devices and receives report and leave messages on behalf of the upstream device. As a best practice, enable IGMP snooping proxy on the edge device to alleviate the effect caused by excessive packets.

Restrictions and guidelines for enabling IGMP snooping proxying

Before you enable IGMP snooping proxying for a VLAN or VSI, you must first enable IGMP snooping globally and enable IGMP snooping for the VLAN or VSI. IGMP snooping proxying does not take effect on sub VLANs of a multicast VLAN.

Use this feature with IGMP snooping querier. For more information about enabling IGMP snooping querier, see "Enabling the IGMP snooping querier."

Enabling IGMP snooping proxying for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable IGMP snooping proxying for the VLAN.
   igmp-snooping proxy enable
   By default, IGMP snooping proxying is disabled for a VLAN.

Enabling IGMP snooping proxying for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable IGMP snooping proxying for the VSI.
   igmp-snooping proxy enable
   By default, IGMP snooping proxying is disabled for a VSI.

Configuring parameters for IGMP messages

Configuring source IP addresses for IGMP messages

About this task

The IGMP snooping querier might send IGMP general queries with the source IP address 0.0.0.0. The port that receives such queries will not be maintained as a dynamic router port. This might prevent the associated dynamic IGMP snooping forwarding entry from being correctly created at the data link layer and eventually cause multicast traffic forwarding failures. To avoid this problem, you can configure a non-all-zero IP address as the source IP address of the IGMP queries on the IGMP snooping querier. This configuration might affect the IGMP querier election within the subnet.

You can also change the source IP address of IGMP reports or leave messages sent by a simulated member host or an IGMP snooping proxy.

Configuring the source IP addresses for IGMP messages in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Configure the source IP address for IGMP general queries.
   igmp-snooping general-query source-ip *ip-address*
   By default, the source IP address of IGMP general queries is the IP address of the current VLAN interface. If the current VLAN interface does not have an IP address, the source IP address is 0.0.0.0.

4. Configure the source IP address for IGMP group-specific queries.

   igmp-snooping special-query source-ip *ip-address*

   By default, the source IP address of IGMP group-specific queries is one of the following:

   - The source address of IGMP group-specific queries if the IGMP snooping querier of the VLAN has received IGMP general queries.
   - The IP address of the current VLAN interface if the IGMP snooping querier does not receive an IGMP general query.
   - 0.0.0.0 if the IGMP snooping querier does not receive an IGMP general query and the current VLAN interface does not have an IP address.

5. Configure the source IP address for IGMP reports.

   igmp-snooping report source-ip *ip-address*

   By default, the source IP address of IGMP reports is the IP address of the current VLAN interface. If the current VLAN interface does not have an IP address, the source IP address is 0.0.0.0.

6. Configure the source IP address for IGMP leave messages.

   igmp-snooping leave source-ip *ip-address*

   By default, the source IP address of IGMP leave messages is the IP address of the current VLAN interface. If the current VLAN interface does not have an IP address, the source IP address is 0.0.0.0.

Configuring the source IP addresses for IGMP messages in a VSI

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Configure the source IP address for IGMP general queries.

   igmp-snooping general-query source-ip *ip-address*

   By default, the source IP address of IGMP general queries is the IP address of the gateway interface for a VSI. If the gateway interface does not have an IP address, the source IP address is 0.0.0.0.

4. Configure the source IP address for IGMP group-specific queries.

   igmp-snooping special-query source-ip *ip-address*

   By default, the source IP address of IGMP group-specific queries is one of the following:

   - The source IP address of IGMP general queries if the IGMP snooping querier of a VSI has received IGMP general queries.
   - The IP address of the gateway interface for the VSI if the IGMP snooping querier does not receive an IGMP general query.
   - 0.0.0.0 if the gateway interface of the VSI does not have an IP address.

5. Configure the source IP address for IGMP reports.

   igmp-snooping report source-ip *ip-address*

   By default, the source IP address of IGMP reports is the IP address of the gateway interface for a VSI. If the gateway interface does not have an IP address, the source IP address is 0.0.0.0.

6. Configure the source IP address for IGMP leave messages.

   igmp-snooping leave source-ip *ip-address*

   By default, the source IP address of IGMP leave messages is the IP address of the gateway interface for a VSI. If the gateway interface does not have an IP address, the source IP address is 0.0.0.0.

Setting the 802.1p priority for IGMP messages

About this task

When congestion occurs on outgoing ports of the Layer 2 device, it forwards IGMP messages in their 802.1p priority order, from highest to lowest. You can assign a higher 802.1p priority to IGMP messages that are created or forwarded by the device.

Restrictions and guidelines

The configuration in IGMP-snooping view takes effect only for all VLANs. For VSIs, you can set the 802.1p priority for IGMP messages only in VSI view.

Setting the 802.1p priority for IGMP messages globally

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Set the 802.1p priority for IGMP messages.
   dot1p-priority *priority*
   By default, the global 802.1p priority is not set.

Setting the 802.1p priority for IGMP messages in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Set the 802.1p priority for IGMP messages in the VLAN.
   igmp-snooping dot1p-priority *priority*
   By default, the 802.1p priority is not set.

Configuring IGMP snooping policies

Configuring a multicast group policy

About this task

This feature enables the Layer 2 device to filter IGMP reports by using an ACL that specifies the multicast groups and the optional sources. It is used to control the multicast groups that hosts can join. This configuration takes effect only on the multicast groups that ports join dynamically.

In a multicast application, a host sends an unsolicited IGMP report when a user requests a multicast program. The Layer 2 device uses the multicast group policy to filter the IGMP report. The host can join the multicast group only if the IGMP report is permitted by the multicast group policy.

Configuring a multicast group policy globally

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Configure a multicast group policy globally.

group-policy *ipv4-acl-number* [**vlan** *vlan-list*]

By default, no multicast group policies are configured, and hosts can join any multicast groups.

Configuring a multicast group policy on a port

1. Enter system view.

   system-view

2. **Enter Layer 2 interface view.**
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
3. Configure a multicast group policy on the port.

   igmp-snooping group-policy *ipv4-acl-number* [**vlan** *vlan-list*]

   By default, no multicast group policies are configured on a port, and hosts attached to the port can join any multicast groups.

Enabling dropping unknown multicast data

About this task

Unknown multicast data refers to multicast data for which no forwarding entries exist in the IGMP snooping forwarding table. This feature enables the device to forward unknown multicast data only to the router port. If the device does not have a router port, unknown multicast data will be dropped.

If you do not enable this feature, the unknown multicast data is flooded in the VLAN or VSI to which the data belongs.

Restrictions and guidelines

When dropping unknown IPv4 multicast data is enabled, the device also drops unknown IPv6 multicast data.

When this feature is enabled for a VLAN or VSI, the device still forwards unknown multicast data out of router ports except the receiving router port in this VLAN or VSI.

Enabling dropping unknown multicast data for a VLAN

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Enable dropping unknown multicast data for the VLAN.

   igmp-snooping drop-unknown

   By default, dropping unknown multicast data is disabled for a VLAN. Unknown multicast data is flooded in the VLAN.

Enabling dropping unknown multicast data for a VSI

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Enable dropping unknown multicast data for the VSI.

   igmp-snooping drop-unknown

By default, dropping unknown multicast data is disabled for a VSI. Unknown multicast data is flooded in the VSI.

Enabling IGMP report suppression

About this task

This feature enables the Layer 2 device to forward only the first IGMP report for a multicast group to its directly connected Layer 3 device. Other reports for the same group in the same query interval are discarded. Use this feature to reduce multicast traffic.

Procedure

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Enable IGMP report suppression.
   report-aggregation
   By default, IGMP report suppression is enabled.

Setting the maximum number of multicast groups on a port

About this task

You can set the maximum number of multicast groups on a port to regulate the port traffic. This feature takes effect only on the multicast groups that a port joins dynamically.

If the number of multicast groups on a port exceeds the limit, the system removes all the forwarding entries associated with the port. The receiver hosts attached to that port can join multicast groups again before the number of multicast groups on the port reaches the limit. When the number of multicast groups on the port reaches the limit, the port automatically drops IGMP reports for new groups.

Procedure

1. Enter system view.
   system-view
2. **Enter Layer 2 interface view.**
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Set the maximum number of multicast groups on the port.
   igmp-snooping group-limit *limit* [**vlan** *vlan-list*]
   By default, no limit is placed on the maximum number of multicast groups on a port.

Enabling multicast group replacement

About this task

When multicast group replacement is enabled, the port does not drop IGMP reports for new groups if the number of multicast groups on the port reaches the upper limit. Instead, the port leaves the

multicast group that has the lowest IP address and joins the new group contained in the IGMP report. The multicast group replacement feature is typically used in the channel switching application.

Restrictions and guidelines

This feature takes effect only on the multicast groups that a port joins dynamically.

This feature does not take effect if the following conditions exist:

- The number of the IGMP snooping forwarding entries on the device reaches the upper limit.
- The multicast group that the port newly joins is not included in the multicast group list maintained by the device.

Enabling multicast group replacement globally

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Enable multicast group replacement globally.
   overflow-replace [**vlan** *vlan-list*]
   By default, multicast group replacement is disabled globally.

Enabling multicast group replacement on a port

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Enable multicast group replacement on the port.
   igmp-snooping overflow-replace [**vlan** *vlan-list*]
   By default, multicast group replacement is disabled on a port.

Enabling host tracking

About this task

This feature enables the Layer 2 device to record information about member hosts that are receiving multicast data. The information includes IP addresses of the hosts, length of time elapsed since the hosts joined multicast groups, and remaining timeout time for the hosts. This feature facilitates monitoring and managing member hosts.

Restrictions and guidelines

The global configuration does not take effect on VSIs.

Enabling host tracking globally

1. Enter system view.
   system-view
2. Enter IGMP-snooping view.
   igmp-snooping
3. Enable host tracking globally.

host-tracking

By default, host tracking is disabled globally.

Enabling host tracking for a VLAN

1. Enter system view.

   system-view
2. Enter VLAN view.

   vlan *vlan-id*
3. Enable host tracking for the VLAN.

   igmp-snooping host-tracking

   By default, host tracking is disabled for a VLAN.

Setting the DSCP value for outgoing IGMP protocol packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.

   system-view
2. Enter IGMP snooping view.

   igmp-snooping
3. Set the DSCP value for outgoing IGMP protocol packets.

   dscp *dscp-value*

   By default, the DSCP value is 48 for outgoing IGMP protocol packets.

Enabling SNMP notifications for IGMP snooping

About this task

To report critical IGMP snooping events to an SNMP NMS, enable SNMP notifications for IGMP snooping.

For IGMP snooping SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about configuring SNMP, see Network Management and Monitoring Configuration Guide.

Procedure

1. Enter system view.

   system-view
2. Enable SNMP notifications for IGMP snooping.

   snmp-agent trap enable igmp-snooping [**entry-refresh-failed**]

   By default, SNMP notifications are enabled for IGMP snooping.

Configuring IGMP snooping logging

Enabling IGMP snooping logging

About this task

This feature enables the IGMP snooping module to send log messages to the information center for troubleshooting. With the information center, you can set log message filtering and output rules, including output destinations. For more information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Excessive number of logs affects the device performance. You can disable IGMP snooping logging or use the **igmp-snooping log suppress** command to limit the rate of log messages sent.

Procedure

1. Enter system view.
   system-view
2. Enter IGMP snooping view.
   igmp-snooping
3. Enable IGMP snooping logging.
   igmp-snooping log enable [**group**]
   By default, IGMP snooping logging is enabled.

Enabling IGMP snooping logging suppression

About this task

Excessive number of logs affects the device performance. This feature limits the rate of log messages sent from the IGMP snooping module to the information center.

Procedure

1. Enter system view.
   system-view
2. Enter IGMP snooping view.
   igmp-snooping
3. Enable IGMP snooping logging suppression.
   igmp-snooping log suppress [**group**]
   By default, IGMP snooping logging suppression is enabled.

Display and maintenance commands for IGMP snooping

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display IGMP snooping status.	**display igmp-snooping** [**global** | **vlan** *vlan-id* | **vsi** *vsi-name*]

Task	Command
Display dynamic IGMP snooping group entries.	**display igmp-snooping group** [*group-address* | *source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**interface** *interface-type interface-number* | [**verbose**] [**slot** *slot-number*]]
Display host tracking information.	**display igmp-snooping host-tracking vlan** *vlan-id* **group** *group-address* [**source** *source-address*] [**slot** *slot-number*]
Display dynamic router port information.	**display igmp-snooping router-port** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**verbose**] [**slot** *slot-number*]
Display static IGMP snooping group entries.	**display igmp-snooping static-group** [*group-address* | *source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**verbose**] [**slot** *slot-number*]
Display static router port information.	**display igmp-snooping static-router-port** [**vlan** *vlan-id*] [**verbose**] [**slot** *slot-number*]
Display statistics for the IGMP messages and PIMv2 hello messages learned by IGMP snooping.	**display igmp-snooping statistics**
Display Layer 2 multicast fast forwarding entries.	**display l2-multicast fast-forwarding cache** [**vlan** *vlan-id*] [*source-address* | *group-address*] * [**slot** *slot-number*]
Display information about Layer 2 IP multicast groups.	**display l2-multicast ip** [**group** *group-address* | **source** *source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display Layer 2 IP multicast group entries.	**display l2-multicast ip forwarding** [**group** *group-address* | **source** *source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display information about Layer 2 MAC multicast groups.	**display l2-multicast mac** [*mac-address*] [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display Layer 2 MAC multicast group entries.	**display l2-multicast mac forwarding** [*mac-address*] [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display static multicast MAC address entries.	**display mac-address** [*mac-address* [**vlan** *vlan-id*] | [**multicast**] [**vlan** *vlan-id*] [**count**]]
Clear dynamic IGMP snooping group entries.	**reset igmp-snooping group** { *group-address* [*source-address*] | **all** } [**vlan** *vlan-id* | **vsi** *vsi-name*]
Clear dynamic router port information.	**reset igmp-snooping router-port** { **all** | **vlan** *vlan-id* | **vsi** *vsi-name* }

Task	Command
Clear statistics for IGMP messages and PIMv2 hello messages learned through IGMP snooping.	**reset igmp-snooping statistics**
Clear Layer 2 multicast fast forwarding entries.	**reset l2-multicast fast-forwarding cache** [**vlan** *vlan-id*] { { *source-address* | *group-address* } * | **all** } [**slot** *slot-number*]

IGMP snooping configuration examples

Example: Configuring VLAN-based IGMP snooping group polices and simulated joining

Network configuration

As shown in Figure 5, Router A runs IGMPv2 and acts as the IGMP querier. Switch A runs IGMPv2 snooping.

Configure a multicast group policy and simulated joining to meet the following requirements:

- Host A and Host B receive only the multicast data addressed to multicast group 224.1.1.1. Multicast data can be forwarded through GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A uninterruptedly, even though Host A and Host B fail to receive the multicast data.
- Switch A will drop unknown multicast data instead of flooding it in VLAN 100.

Figure 5 Network diagram

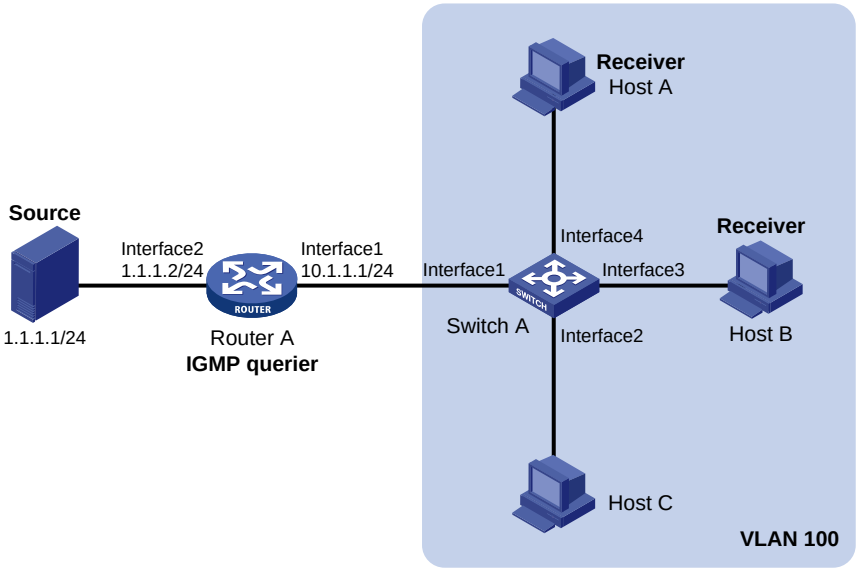


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Router A:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname RouterA
[RouterA] multicast routing
[RouterA-mrib] quit
```

Enable IGMP on GigabitEthernet 1/0/1.

```
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] ip address 10.1.1.1 24
[RouterA-GigabitEthernet1/0/1] igmp enable
[RouterA-GigabitEthernet1/0/1] quit
```

Enable PIM-DM on GigabitEthernet 1/0/2.

```
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/1] ip address 10.1.1.2 24
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] quit
```

2. Configure Switch A:

Enable the IGMP snooping feature.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable IGMP snooping, and enable dropping unknown multicast data for VLAN 100.

```
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping drop-unknown
[SwitchA-vlan100] quit
```

Configure a multicast group policy so that hosts in VLAN 100 can join only multicast group 224.1.1.1.

```
[SwitchA] acl basic 2001
[SwitchA-acl-ipv4-basic-2001] rule permit source 224.1.1.1 0
[SwitchA-acl-ipv4-basic-2001] quit
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] group-policy 2001 vlan 100
[SwitchA-igmp-snooping] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 as simulated member hosts of multicast group 224.1.1.1.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] igmp-snooping host-join 224.1.1.1 vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
```

```
[SwitchA] interface gigabitethernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] igmp-snooping host-join 224.1.1.1 vlan 100
[SwitchA-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Send IGMP reports from Host A and Host B to join multicast groups 224.1.1.1 and 224.2.2.2. (Details not shown.)

Display dynamic IGMP snooping group entries for VLAN 100 on Switch A.

```
[SwitchA] display igmp-snooping group vlan 100
Total 1 entries.

VLAN 100: Total 1 entries.
  (0.0.0.0, 224.1.1.1)
    Host ports (2 in total):
      GE1/0/3                              (00:03:23)
      GE1/0/4                              (00:04:10)
```

The output shows the following information:

- Host A and Host B have joined multicast group 224.1.1.1 through the member ports GigabitEthernet 1/0/4 and GigabitEthernet 1/0/3 on Switch A, respectively.
- Host A and Host B have failed to join multicast group 224.2.2.2.

Configuration files

- Router A:

```
#
 sysname RouterA
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 10.1.1.1 255.255.255.0
 igmp enable
#
interface GigabitEthernet1/0/2
 port link-mode route
 ip address 1.1.1.2 255.255.255.0
 pim dm
#
multicast routing
#
return
```

- Switch A:

```
#
 sysname SwitchA
#
igmp-snooping
 group-policy 2001 vlan 100
#
vlan 100
 igmp-snooping enable
```

```
 igmp-snooping drop-unknown
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
 igmp-snooping host-join 224.1.1.1 vlan 100
#
interface GigabitEthernet1/0/4
 port link-mode bridge
 port access vlan 100
 igmp-snooping host-join 224.1.1.1 vlan 100
#
acl basic 2001
 rule 5 permit source 224.1.1.1 0
#
return
```

Example: Configuring VLAN-based IGMP snooping static ports

Network configuration

As shown in Figure 6:

- Router A runs IGMPv2 and acts as the IGMP querier. Switch A, Switch B, and Switch C run IGMPv2 snooping.
- Host A and host C are permanent receivers of multicast group 224.1.1.1.

Configure static ports to meet the following requirements:

- To enhance the reliability of multicast traffic transmission, configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C as static member ports for multicast group 224.1.1.1.
- Suppose the STP runs on the network. To avoid data loops, the forwarding path from Switch A to Switch C is blocked. Multicast data flows to the receivers attached to Switch C only along the path of Switch A—Switch B—Switch C. When this path is blocked, a minimum of one IGMP query-response cycle must be completed before multicast data flows to the receivers along the path of Switch A—Switch C. In this case, the multicast delivery is interrupted during the process. For more information about the STP, see *Layer 2—LAN Switching Configuration Guide*.

  Configure GigabitEthernet 1/0/3 on Switch A as a static router port. Then, multicast data can flow to the receivers nearly uninterruptedly along the path of Switch A—Switch C when the path of Switch A—Switch B—Switch C is blocked.

Figure 6 Network diagram

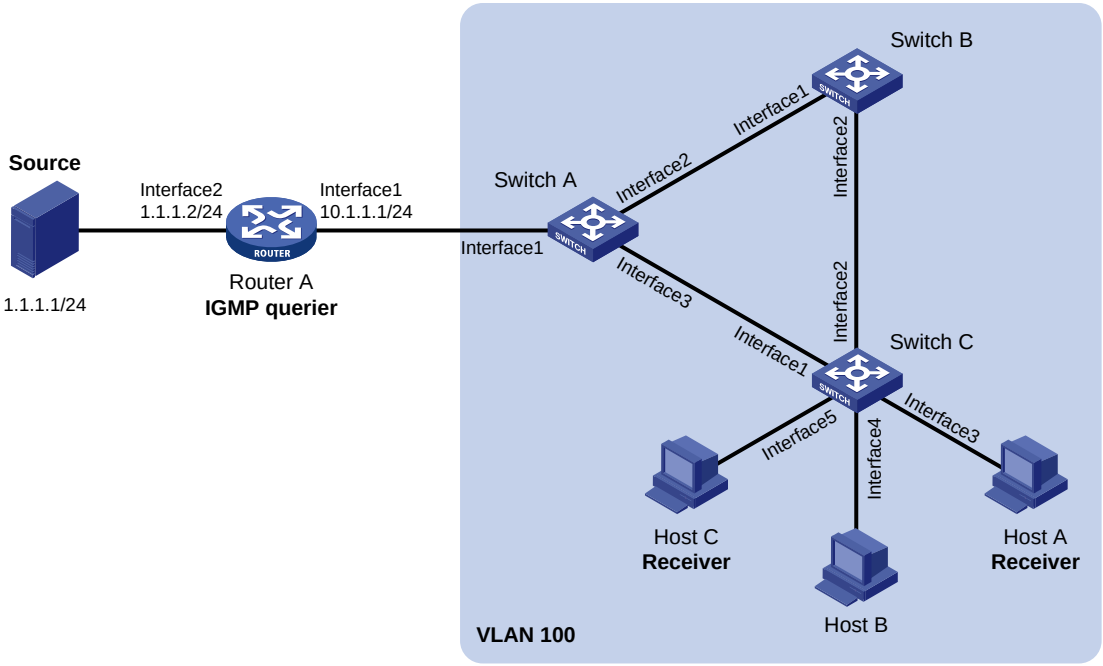


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

1. Configure Router A:

   # Enable IP multicast routing.

   ```
   <Sysname> system-view
   [Sysname] sysname RouterA
   [RouterA] multicast routing
   [RouterA-mrib] quit
   ```

   # Enable IGMP on GigabitEthernet 1/0/1.

   ```
   [RouterA] interface gigabitethernet 1/0/1
   [RouterA-GigabitEthernet1/0/1] ip address 10.1.1.1 24
   [RouterA-GigabitEthernet1/0/1] igmp enable
   [RouterA-GigabitEthernet1/0/1] quit
   ```

   # Enable PIM-DM on GigabitEthernet 1/0/2.

   ```
   [RouterA] interface gigabitethernet 1/0/2
   [RouterA-GigabitEthernet1/0/1] ip address 10.1.1.2 24
   [RouterA-GigabitEthernet1/0/2] pim dm
   [RouterA-GigabitEthernet1/0/2] quit
   ```

2. Configure Switch A:

Enable the IGMP snooping feature.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
```

Enable IGMP snooping for VLAN 100.

```
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] quit
```

Configure GigabitEthernet 1/0/3 as a static router port.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] igmp-snooping static-router-port vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
```

Configure STP.

```
[SwitchA] stp region-configuration
[SwitchA-mst-region] instance 1 vlan 100
[SwitchA-mst-region] active region-configuration
[SwitchA-mst-region] quit
[SwitchA] stp global enable
```

3. Configure Switch B:

Enable the IGMP snooping feature.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 gigabitethernet 1/0/2
```

Enable IGMP snooping for VLAN 100.

```
[SwitchB-vlan100] igmp-snooping enable
[SwitchB-vlan100] quit
```

Configure STP.

```
[SwitchB] stp region-configuration
[SwitchB-mst-region] instance 1 vlan 100
[SwitchB-mst-region] active region-configuration
[SwitchB-mst-region] quit
[SwitchB] stp global enable
```

4. Configure Switch C:

Enable the IGMP snooping feature.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] igmp-snooping
[SwitchC-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/5 to the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/5
```

Enable IGMP snooping for VLAN 100.

```
[SwitchC-vlan100] igmp-snooping enable
[SwitchC-vlan100] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 as static member ports for multicast group 224.1.1.1.

```
[SwitchC] interface gigabitethernet 1/0/3
[SwitchC-GigabitEthernet1/0/3] igmp-snooping static-group 224.1.1.1 vlan 100
[SwitchC-GigabitEthernet1/0/3] quit
[SwitchC] interface gigabitethernet 1/0/5
[SwitchC-GigabitEthernet1/0/5] igmp-snooping static-group 224.1.1.1 vlan 100
[SwitchC-GigabitEthernet1/0/5] quit
```

Configure STP.

```
[SwitchC] stp region-configuration
[SwitchC-mst-region] instance 1 vlan 100
[SwitchC-mst-region] active region-configuration
[SwitchC-mst-region] quit
[SwitchC] stp global enable
```

Verifying the configuration

Display static router port information for VLAN 100 on Switch A.

```
[SwitchA] display igmp-snooping static-router-port vlan 100
VLAN 100:
  Router ports (1 in total):
    GE1/0/3
```

The output shows that GigabitEthernet 1/0/3 on Switch A has become a static router port.

Display static IGMP snooping group entries for VLAN 100 on Switch C.

```
[SwitchC] display igmp-snooping static-group vlan 100
Total 1 entries.

VLAN 100: Total 1 entries.
  (0.0.0.0, 224.1.1.1)
    Host ports (2 in total):
      GE1/0/3
      GE1/0/5
```

The output shows that GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C have become static member ports of multicast group 224.1.1.1.

Configuration files

- Router A:

```
#
 sysname RouterA
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 10.1.1.1 255.255.255.0
```

```
 igmp enable
#
interface GigabitEthernet1/0/2
 port link-mode route
 ip address 1.1.1.2 255.255.255.0
 pim dm
#
multicast routing
#
return
```

- Switch A:

```
#
 sysname SwitchA
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
#
stp region-configuration
 instance 1 vlan 100
 active region-configuration
#
 stp global enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
 igmp-snooping static-router-port vlan 100
#
return
```

- Switch B:

```
#
 sysname SwitchB
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
#
```

```
stp region-configuration
 instance 1 vlan 100
 active region-configuration
#
 stp instance 1 root primary
 stp global enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
return
```

- Switch C:

```
#
 sysname SwitchC
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
#
stp region-configuration
 instance 1 vlan 100
 active region-configuration
#
 stp global enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
 igmp-snooping static-group 224.1.1.1 vlan 100
#
interface GigabitEthernet1/0/4
 port link-mode bridge
 port access vlan 100
#
```

```
interface GigabitEthernet1/0/5
 port link-mode bridge
 port access vlan 100
 igmp-snooping static-group 224.1.1.1 vlan 100
#
return
```

Example: Configuring the VLAN-based IGMP snooping querier

Network configuration

As shown in Figure 7:

- The network is a Layer 2-only network.
- Source 1 and Source 2 send multicast data to multicast groups 224.1.1.1 and 225.1.1.1, respectively.
- Host A and Host C are receivers of multicast group 224.1.1.1, and Host B and Host D are receivers of multicast group 225.1.1.1.
- All host receivers run IGMPv2, and all switches run IGMPv2 snooping. Switch A (which is close to the multicast sources) acts as the IGMP snooping querier.

Configure the switches to meet the following requirements:

- To prevent the switches from flooding unknown data in the VLAN, enable all the switches to drop unknown multicast data.
- A switch does not mark a port that receives an IGMP query with source IP address 0.0.0.0 as a dynamic router port. This adversely affects the establishment of Layer 2 forwarding entries and multicast traffic forwarding. To avoid this, configure the source IP address of IGMP queries as a non-zero IP address.

Figure 7 Network diagram

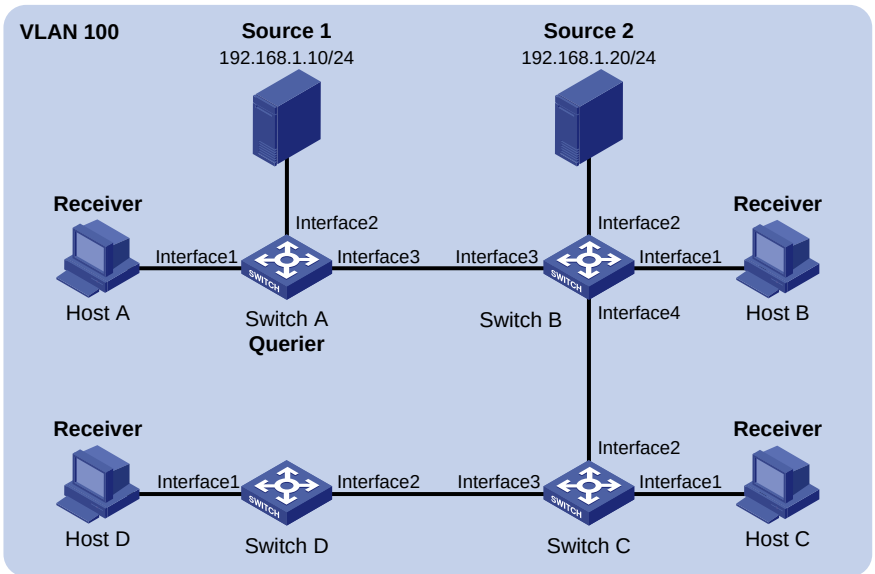


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Switch A:

   # Enable the IGMP snooping feature.

   ```
   <Sysname> system-view
   [Sysname] sysname SwitchA
   [SwitchA] igmp-snooping
   [SwitchA-igmp-snooping] quit
   ```

   # Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the VLAN.

   ```
   [SwitchA] vlan 100
   [SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
   ```

   # Enable IGMP snooping, and enable dropping unknown multicast data for VLAN 100.

   ```
   [SwitchA-vlan100] igmp-snooping enable
   [SwitchA-vlan100] igmp-snooping drop-unknown
   ```

   # Configure Switch A as the IGMP snooping querier.

   ```
   [SwitchA-vlan100] igmp-snooping querier
   [SwitchA-vlan100] quit
   ```

   # In VLAN 100, specify 192.168.1.1 as the source IP address of IGMP general queries.

   ```
   [SwitchA-vlan100] igmp-snooping general-query source-ip 192.168.1.1
   ```

   # In VLAN 100, specify 192.168.1.1 as the source IP address of IGMP group-specific queries.

   ```
   [SwitchA-vlan100] igmp-snooping special-query source-ip 192.168.1.1
   [SwitchA-vlan100] quit
   ```

2. Configure Switch B:

   # Enable the IGMP snooping feature.

   ```
   <Sysname> system-view
   [Sysname] sysname SwitchB
   [SwitchB] igmp-snooping
   [SwitchB-igmp-snooping] quit
   ```

   # Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

   ```
   [SwitchB] vlan 100
   [SwitchB-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
   ```

   # Enable IGMP snooping, and enable dropping unknown multicast data for VLAN 100.

   ```
   [SwitchB-vlan100] igmp-snooping enable
   [SwitchB-vlan100] igmp-snooping drop-unknown
   [SwitchB-vlan100] quit
   ```

3. Configure Switch C:

   # Enable the IGMP snooping feature.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] igmp-snooping
[SwitchC-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
```

Enable IGMP snooping, and enable dropping unknown multicast data for VLAN 100.

```
[SwitchC-vlan100] igmp-snooping enable
[SwitchC-vlan100] igmp-snooping drop-unknown
[SwitchC-vlan100] quit
```

4. Configure Switch D:

Enable the IGMP snooping feature.

```
<Sysname> system-view
[Sysname] sysname SwitchD
[SwitchD] igmp-snooping
[SwitchD-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the VLAN.

```
[SwitchD] vlan 100
[SwitchD-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/2
```

Enable IGMP snooping, and enable dropping unknown multicast data for VLAN 100.

```
[SwitchD-vlan100] igmp-snooping enable
[SwitchD-vlan100] igmp-snooping drop-unknown
[SwitchD-vlan100] quit
```

Verifying the configuration

Display statistics for IGMP messages and PIMv2 hello messages learned through IGMP snooping on Switch B.

```
[SwitchB] display igmp-snooping statistics
Received IGMP messages:
  IGMP general queries: 19
    M-LAG: 19
  IGMPv2 specific queries: 10
    M-LAG: 0
  IGMPv3 specific queries: 0
    M-LAG: 0
  IGMPv3 specific sg queries: 0
    M-LAG: 0
  IGMPv1 reports: 0
    M-LAG: 0
  IGMPv2 reports: 19
    M-LAG: 19
  IGMPv3 reports: 0
    M-LAG: 0
  IGMPv3 reports with right and wrong records: 0
    M-LAG: 0
  IGMP leaves: 5
    M-LAG: 5
```

```
  Error IGMP messages: 1
    M-LAG: 1
Sent IGMP messages:
  IGMPv2 specific queries: 0
  IGMPv3 specific queries: 0
  IGMPv3 specific sg queries: 0
Received PIMv2 hello : 0
```

The output shows that all switches except Switch A can receive the IGMP general queries after Switch A acts as the IGMP snooping querier.

Configuration files

- Switch A:

```
#
 sysname SwitchA
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
 igmp-snooping drop-unknown
 igmp-snooping querier
 igmp-snooping general-query source-ip 192.168.1.1
 igmp-snooping special-query source-ip 192.168.1.1
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
#
return
```

- Switch B:

```
#
 sysname SwitchB
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
 igmp-snooping drop-unknown
#
interface GigabitEthernet1/0/1
```

```
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/4
 port link-mode bridge
 port access vlan 100
#
return
```

- Switch C:

```
#
 sysname SwitchC
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
 igmp-snooping drop-unknown
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
#
return
```

- Switch D:

```
#
 sysname SwitchD
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
 igmp-snooping drop-unknown
```

```
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
return
```

Example: Configuring VLAN-based IGMP snooping proxying

Network configuration

As shown in Figure 8, Router A runs IGMPv2 and acts as the IGMP querier. Switch A runs IGMPv2 snooping. Configure IGMP snooping proxying so that Switch A can perform the following actions:

- Forward IGMP report and leave messages to Router A.
- Respond to IGMP queries sent by Router A and forward the queries to downstream hosts.

Figure 8 Network diagram

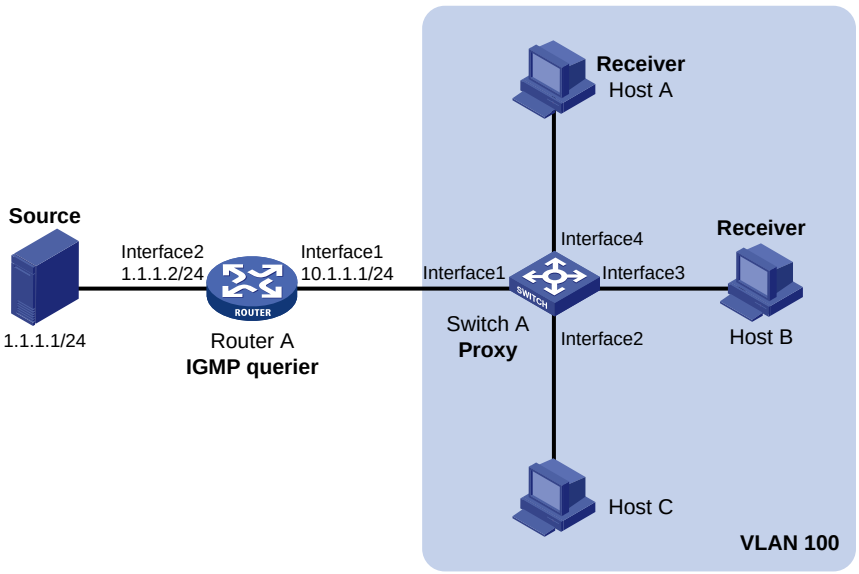


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Router A:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname RouterA
[RouterA] multicast routing
[RouterA-mrib] quit
```

Enable IGMP and PIM-DM on GigabitEthernet 1/0/1.

```
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] ip address 10.1.1.1 24
[RouterA-GigabitEthernet1/0/1] igmp enable
[RouterA-GigabitEthernet1/0/1] pim dm
[RouterA-GigabitEthernet1/0/1] quit
```

Enable PIM-DM on GigabitEthernet 1/0/2.

```
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/1] ip address 10.1.1.2 24
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] quit
```

2. Configure Switch A:

Enable the IGMP snooping feature.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable IGMP snooping and IGMP snooping proxying for the VLAN.

```
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping proxy enable
[SwitchA-vlan100] quit
```

Verifying the configuration

Send IGMP reports from Host A and Host B to join multicast groups 224.1.1.1 and 224.1.1.1. (Details not shown.)

Display brief information about IGMP snooping group entries on Switch A.

```
[SwitchA] display igmp-snooping group
 Total 1 entries.

VLAN 100: Total 1 entries.
  (0.0.0.0, 224.1.1.1)
    Host ports (2 in total):
      GE1/0/3                                  (00:04:00)
      GE1/0/4                                  (00:04:04)
```

The output shows that GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 are member ports of multicast group 224.1.1.1. Host A and Host B become receivers of the group.

Display IGMP group membership information on Router A.

```
[RouterA] display igmp group
IGMP groups in total: 1
```

```
 GigabitEthernet1/0/1(10.1.1.1):
  IGMP groups reported in total: 1
   Group address   Last reporter   Uptime      Expires
   224.1.1.1       0.0.0.0         00:00:31    00:02:03
```

Send an IGMP leave message from Host A to leave multicast group 224.1.1.1. (Details not shown.)

Display brief information about IGMP snooping group entries on Switch A.

```
[SwitchA] display igmp-snooping group
Total 1 entries.

VLAN 100: Total 1 entries.
  (0.0.0.0, 224.1.1.1)
    Host ports (1 in total):
      GE1/0/3                                  ( 00:01:23 )
```

The output shows that GigabitEthernet 1/0/3 is the only member port of multicast group 224.1.1.1. Only Host B remains as the receiver of the group.

Configuration files

- Router A:

```
#
 sysname RouterA
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 10.1.1.1 255.255.255.0
 pim dm
 igmp enable
#
interface GigabitEthernet1/0/2
 port link-mode route
 ip address 1.1.1.2 255.255.255.0
 pim dm
#
multicast routing
#
return
```

- Switch A:

```
#
 sysname SwitchA
#
igmp-snooping
#
vlan 100
 igmp-snooping enable
 igmp-snooping proxy enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 100
```

```
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
#
interface GigabitEthernet1/0/4
 port link-mode bridge
 port access vlan 100
#
return
```

Troubleshooting IGMP snooping

Layer 2 multicast forwarding cannot function

Symptom

Layer 2 multicast forwarding cannot function on the Layer 2 device.

Solution

To resolve the problem:

1. Use the **display igmp-snooping** command to display IGMP snooping status.
2. If IGMP snooping is not enabled, use the **igmp-snooping** command in system view to enable the IGMP snooping feature. Then, use the **igmp-snooping enable** command in VLAN view or VSI view to enable IGMP snooping for the VLAN or VSI.
3. If IGMP snooping is enabled globally but not enabled for the VLAN or VSI, use the **igmp-snooping enable** command in VLAN view or VSI view to enable IGMP snooping for the VLAN or VSI.
4. If the problem persists, contact H3C Support.

Multicast group policy does not work

Symptom

Hosts can receive multicast data for multicast groups that are not permitted by the multicast group policy.

Solution

To resolve the problem:

1. Use the **display acl** command to verify that the configured ACL meets the multicast group policy requirements.
2. Use the **display this** command in IGMP-snooping view or in a corresponding interface view to verify that the correct multicast group policy has been applied. If the applied policy is not correct, use the **group-policy** or **igmp-snooping group-policy** command to apply the correct multicast group policy.

3. Use the **display igmp-snooping** command to verify that dropping unknown multicast data is enabled. If it is not, use the **igmp-snooping drop-unknown** command to enable dropping unknown multicast data.
4. If the problem persists, contact H3C Support.

Contents

Configuring PIM snooping 1

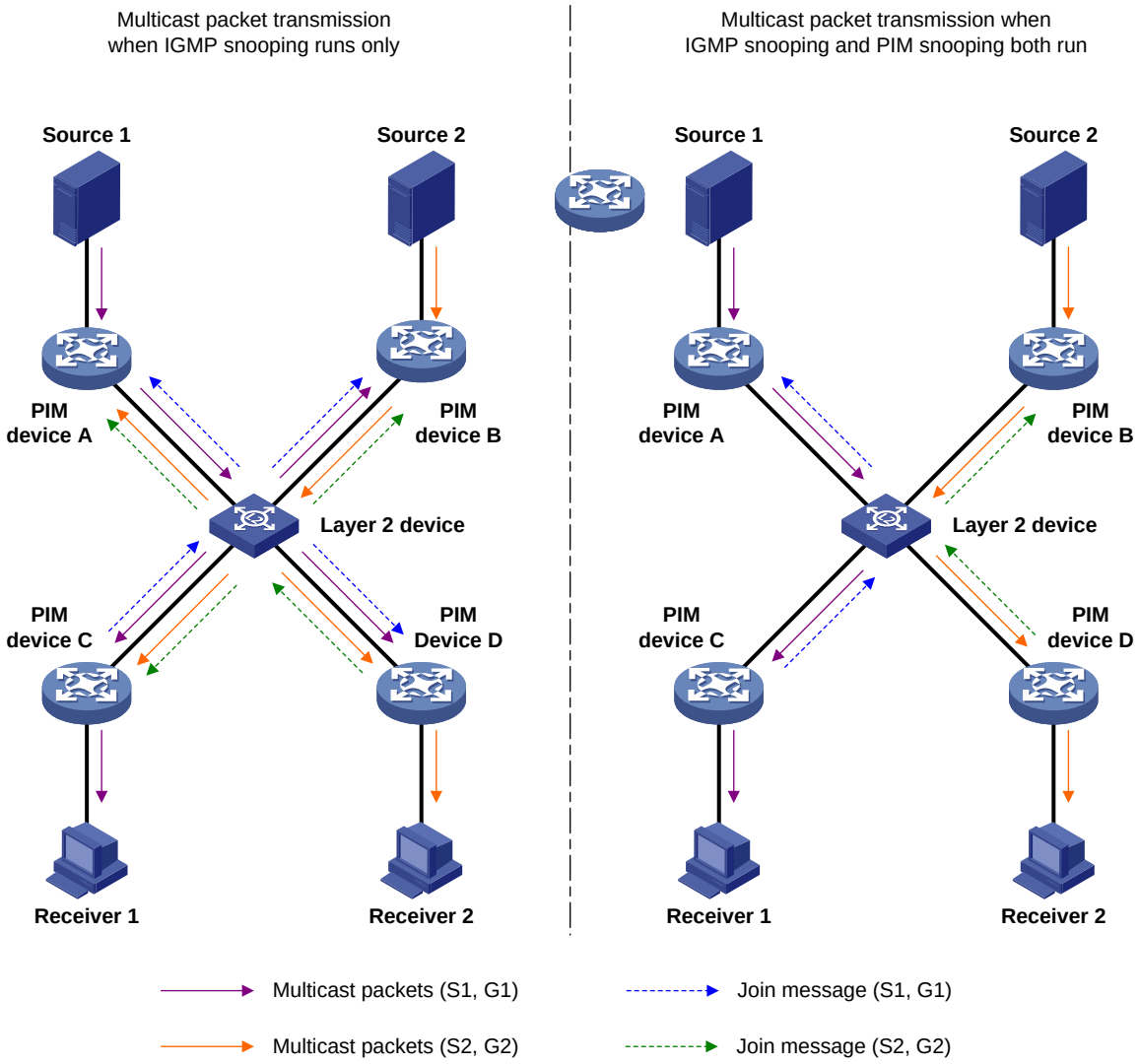
About PIM snooping 1
Restrictions and guidelines: PIM snooping configuration 2
PIM snooping tasks at a glance 2
Enabling PIM snooping 2
Setting the aging time for global ports after a master/subordinate switchover 3
About global ports 3
Restrictions and guidelines 3
Setting the aging time for global neighbor ports 3
Setting the aging time for global downstream ports and global router ports 4
Display and maintenance commands for PIM snooping 4
PIM snooping configuration examples 5
Example: Configuring PIM snooping 5
Troubleshooting PIM snooping 8
PIM snooping does not work on a Layer 2 device 8

Configuring PIM snooping

About PIM snooping

PIM snooping runs on Layer 2 devices. It works with IGMP snooping to analyze received PIM messages, and adds the ports that are interested in specific multicast data to a PIM snooping routing entry. In this way, the multicast data can be forwarded to only the ports that are interested in the data.

Figure 1 Multicast packet transmission without or with PIM snooping



As shown in Figure 1, Source 1 sends multicast data to multicast group G1, and Source 2 sends multicast data to multicast group G2. Receiver 1 belongs to G1, and Receiver 2 belongs to G2. The Layer 2 switch's interfaces that connect to the PIM routers are in the same VLAN or VSI.

- When the Layer 2 switch runs only IGMP snooping, it performs the following actions:
  - **a.** Maintains the router ports according to the received PIM hello messages that PIM routers send.
  - **b.** Floods all other types of received PIM messages except PIM hello messages in the VLAN or VSI.

c. Forwards all multicast data to all router ports in the VLAN or VSI.

Each PIM router in the VLAN, whether interested in the multicast data or not, can receive all multicast data and all PIM messages except PIM hello messages.

- When the Layer 2 switch runs both IGMP snooping and PIM snooping, it performs the following actions:

  d. Examines whether a PIM router is interested in the multicast data addressed to a multicast group according to the received PIM messages that the router sends.

  e. Adds only the ports that connect to the router and are interested in the data to a PIM snooping routing entry.

  f. Forwards PIM messages and multicast data to only the routers that are interested in the data, which saves network bandwidth.

For more information about IGMP snooping and the router port, see "Configuring IGMP snooping."

Restrictions and guidelines: PIM snooping configuration

Make sure the maximum size of a PIM join or prune message is smaller than the path MTU of devices connected to PIM snooping devices. Otherwise, fragmented PIM join or prune messages prevent PIM snooping from correctly forwarding multicast data. For more information about setting the maximum size of a join or prune message, see "Configuring PIM."

PIM snooping is supported only for PIM-SM and PIM-SSM. As a best practice, do not configure PIM snooping for PIM-DM and BIDIR-PIM. For more information about PIM, see "PIM overview."

After you configure PIM snooping for a VLAN or VSI, PIM snooping takes effect only on ports that belong to the VLAN or VSI.

PIM snooping tasks at a glance

To configure PIM snooping, perform the following tasks:

1. Enabling PIM snooping
2. (Optional.) Setting the aging time for global ports after a master/subordinate switchover
   - Setting the aging time for global neighbor ports
   - Setting the aging time for global downstream ports and global router ports

Enabling PIM snooping

Restrictions and guidelines

You must enable the IGMP snooping feature and enable IGMP snooping for a VLAN or VSI before you enable PIM snooping for the VLAN or VSI.

PIM snooping does not take effect on sub-VLANs of a multicast VLAN.

Procedure

1. Enter system view.

   `system-view`

2. Enable the IGMP snooping feature and enter IGMP-snooping view.

   `igmp-snooping`

   By default, IGMP snooping is disabled.

For more information about this command, see IGMP snooping commands in *IP Multicast Command Reference*.

3. Return to system view.

   quit

4. Enter VLAN view or VSI view.
   - Enter VLAN view.

     vlan *vlan-id*
   - Enter VSI view.

     vsi *vsi-name*

5. Enable IGMP snooping for the VLAN or VSI.

   igmp-snooping enable

   By default, IGMP snooping is disabled in a VLAN or VSI.

   For more information about this command, see *IP Multicast Command Reference*.

6. Enable PIM snooping for the VLAN or VSI.

   pim-snooping enable

   By default, PIM snooping is disabled in a VLAN or VSI.

Setting the aging time for global ports after a master/subordinate switchover

About global ports

A global port is a virtual port on the master device, such as a Layer 2 aggregate interface, AC, N-PW, and U-PW. A global port that acts as a neighbor port, downstream port, or router port is called a global neighbor port, global downstream port, and global router port, respectively.

Perform this task to decrease Layer 2 multicast data interruption caused by the aging of PIM snooping entries after a master/subordinate switchover.

Restrictions and guidelines

For a global neighbor port, the set aging time does not take effect when the port receives a PIM hello message after a master/subordinate switchover. The aging time for the port is determined by the aging time in the PIM hello message.

For a global router port or global downstream port, the set aging time does not take effect when the port receives a PIM join message after a master/subordinate switchover. The aging time for the port is determined by the aging time in the PIM join message.

Setting the aging time for global neighbor ports

1. Enter system view.

   system-view

2. Enter VLAN view or VSI view.
   - Enter VLAN view.

     vlan *vlan-id*
   - Enter VSI view.

     vsi *vsi-name*

3. Set the aging time for global neighbor ports after a master/subordinate switchover.

   pim-snooping graceful-restart neighbor-aging-time *seconds*

   By default, the aging time for global neighbor ports after a master/subordinate switchover is 105 seconds.

Setting the aging time for global downstream ports and global router ports

1. Enter system view.

   system-view

2. Enter VLAN view or VSI view.
   - Enter VLAN view.

     vlan *vlan-id*
   - Enter VSI view.

     vsi *vsi-name*

3. Set the aging time for global downstream ports and global router ports after a master/subordinate switchover.

   pim-snooping graceful-restart join-aging-time *seconds*

   By default, the aging time for downstream ports and global router ports after a master/subordinate switchover is 210 seconds.

Display and maintenance commands for PIM snooping

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display PIM snooping neighbor information.	**display pim-snooping neighbor** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*] [**verbose**]
Display PIM snooping router port information.	**display pim-snooping router-port** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*] [**verbose**]
Display PIM snooping routing entries.	**display pim-snooping routing-table** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*] [**verbose**]
Display statistics for the PIM messages learned through PIM snooping.	**display pim-snooping statistics**
Clear statistics for the PIM messages learned through PIM snooping.	**reset pim-snooping statistics**

PIM snooping configuration examples

Example: Configuring PIM snooping

Network configuration

As shown in Figure 2:

- OSPF runs on the network.
- Source 1 and Source 2 send multicast data to multicast groups 224.1.1.1 and 225.1.1.1, respectively.
- Receiver 1 and Receiver 2 belong to multicast groups 224.1.1.1 and 225.1.1.1, respectively.
- Router C and Router D run IGMP on GigabitEthernet 1/0/1. Router A, Router B, Router C, and Router D run PIM-SM.
- GigabitEthernet 1/0/2 on Router A acts as a C-BSR and a C-RP.

Configure IGMP snooping and PIM snooping on Switch A. Then, Switch A forwards PIM protocol packets and multicast data packets only to the routers that are connected to receivers.

Figure 2 Network diagram

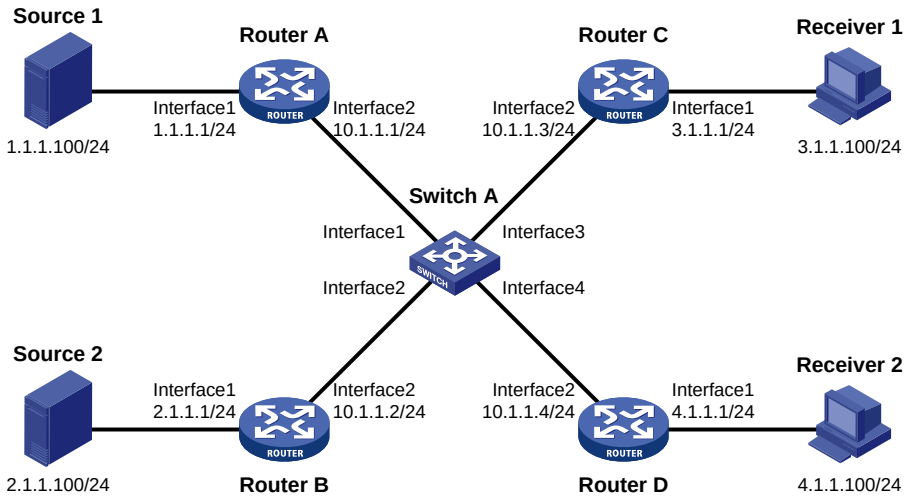


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 2, and configure OSPF on the routers.

Procedure

1. Configure Router A:

   # Enable IP multicast routing.

```
<RouterA> system-view
[RouterA] multicast routing
[RouterA-mrib] quit
```

Enable PIM-SM on each interface.

```
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] pim sm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim sm
[RouterA-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/2 as a C-BSR and a C-RP.

```
[RouterA] pim
[RouterA-pim] c-bsr 10.1.1.1
[RouterA-pim] c-rp 10.1.1.1
[RouterA-pim] quit
```

2. Configure Router B:

Enable IP multicast routing.

```
<RouterB> system-view
[RouterB] multicast routing
[RouterB-mrib] quit
```

Enable PIM-SM on each interface.

```
[RouterB] interface gigabitethernet 1/0/1
[RouterB-GigabitEthernet1/0/1] pim sm
[RouterB-GigabitEthernet1/0/1] quit
[RouterB] interface gigabitethernet 1/0/2
[RouterB-GigabitEthernet1/0/2] pim sm
[RouterB-GigabitEthernet1/0/2] quit
```

3. Configure Router C:

Enable IP multicast routing.

```
<RouterC> system-view
[RouterC] multicast routing
[RouterC-mrib] quit
```

Enable IGMP on GigabitEthernet 1/0/1.

```
[RouterC] interface gigabitethernet 1/0/1
[RouterC-GigabitEthernet1/0/1] igmp enable
[RouterC-GigabitEthernet1/0/1] quit
```

Enable PIM-SM on GigabitEthernet 1/0/2.

```
[RouterC] interface gigabitethernet 1/0/2
[RouterC-GigabitEthernet1/0/2] pim sm
[RouterC-GigabitEthernet1/0/2] quit
```

4. Configure Router D:

Enable IP multicast routing.

```
<RouterD> system-view
[RouterD] multicast routing
[RouterD-mrib] quit
```

Enable IGMP on GigabitEthernet 1/0/1.

```
[RouterD] interface gigabitethernet 1/0/1
```

```
[RouterD-GigabitEthernet1/0/1] igmp enable
[RouterD-GigabitEthernet1/0/1] quit
```

Enable PIM-SM on GigabitEthernet 1/0/2.

```
[RouterD] interface gigabitethernet 1/0/2
[RouterD-GigabitEthernet1/0/2] pim sm
[RouterD-GigabitEthernet1/0/2] quit
```

5. Configure Switch A:

Enable the IGMP snooping feature.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable IGMP snooping and PIM snooping for VLAN 100.

```
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] pim-snooping enable
[SwitchA-vlan100] quit
```

Verifying the configuration

On Switch A, display PIM snooping neighbor information for VLAN 100.

```
[SwitchA] display pim-snooping neighbor vlan 100
Total 4 neighbors.

VLAN 100: Total 4 neighbors.
  10.1.1.1
    Ports (1 in total):
      GE1/0/1                 (00:32:43)
  10.1.1.2
    Ports (1 in total):
      GE1/0/2                 (00:32:43)
  10.1.1.3
    Ports (1 in total):
      GE1/0/3                 (00:32:43)
  10.1.1.4
    Ports (1 in total):
      GE1/0/4                 (00:32:43)
```

The output shows that Router A, Router B, Router C, and Router D are PIM snooping neighbors.

On Switch A, display PIM snooping routing entries for VLAN 100.

```
[SwitchA] display pim-snooping routing-table vlan 100
Total 4 entries.
FSM Flag: NI-no info, J-join, PP-prune pending

VLAN 100: Total 4 entries.
  (*, 224.1.1.1)
    Upstream neighbor: 10.1.1.1
      Upstream Ports (1 in total):
```

```
      GE1/0/1
    Downstream Ports (1 in total):
      GE1/0/3
        Expires: 00:03:01, FSM: J
(1.1.1.100, 224.1.1.1)
  Upstream neighbor: 10.1.1.1
    Upstream Slots (0 in total):
    Upstream Ports (1 in total):
      GE1/0/1
    Downstream Slots (0 in total):
    Downstream Ports (1 in total):
      GE1/0/3
        Expires: 00:02:33, FSM: J
(*, 225.1.1.1)
  Upstream neighbor: 10.1.1.1
    Upstream Slots (0 in total):
    Upstream Ports (1 in total):
      GE1/0/1
    Downstream Slots (0 in total):
    Downstream Ports (1 in total):
      GE1/0/4
        Expires: 00:03:11, FSM: J
(2.1.1.100, 225.1.1.1)
  Upstream neighbor: 10.1.1.2
    Upstream Slots (0 in total):
    Upstream Ports (1 in total):
      GE1/0/2
    Downstream Slots (0 in total):
    Downstream Ports (1 in total):
      GE1/0/4
        Expires: 00:02:34, FSM: J
```

The output shows the following information:

- Switch A will forward the multicast data intended for multicast group 224.1.1.1 only to Router C.
- Switch A will forward the multicast data intended for multicast group 225.1.1.1 only to Router D.

Troubleshooting PIM snooping

PIM snooping does not work on a Layer 2 device

Symptom

PIM snooping does not work on a Layer 2 device.

Solution

To resolve the problem:

1. Use the **`display current-configuration`** command to display information about IGMP snooping and PIM snooping.

2. If IGMP snooping is not enabled, enable the IGMP snooping feature, and then enable IGMP snooping and PIM snooping for the VLAN or VSI.
3. If PIM snooping is not enabled, enable PIM snooping for the VLAN or VSI.
4. If the problem persists, contact H3C Support.

Contents

Configuring multicast VLANs 1

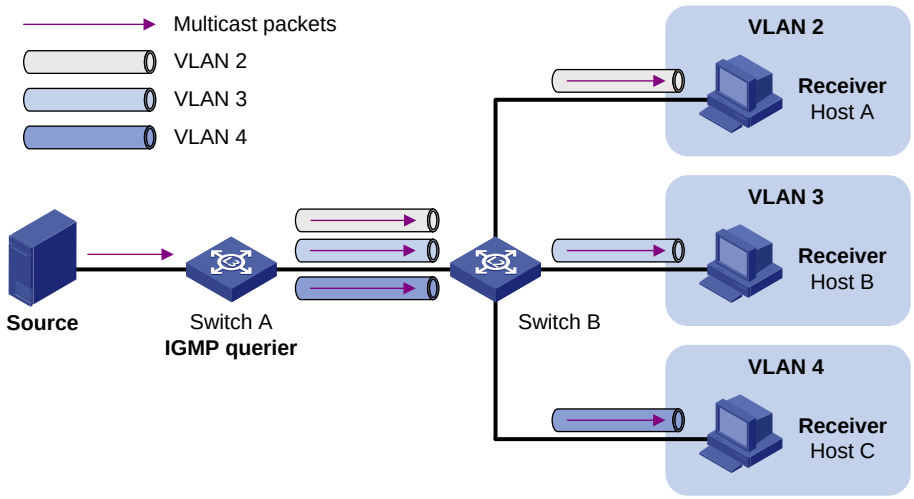
Multicast VLAN feature 1
Multicast VLAN implementations 1
Sub-VLAN-based multicast VLAN 1
Port-based multicast VLAN 2
Restrictions and guidelines: Multicast VLAN configuration 3
Configuring a sub-VLAN-based multicast VLAN 3
Configuring a port-based multicast VLAN 3
Setting the maximum number of multicast VLAN forwarding entries 4
Display and maintenance commands for multicast VLAN 5
Multicast VLAN configuration examples 5
Example: Configuring sub-VLAN-based multicast VLAN 5
Example: Configuring port-based multicast VLAN 8

Configuring multicast VLANs

Multicast VLAN feature

As shown in Figure 1, Host A, Host B, and Host C are in three different VLANs and the same multicast group. When Switch A (Layer 3 device) receives multicast data for that group, it sends three copies of the multicast data to Switch B (Layer 2 device). This occupies a large amount of bandwidth and increases the burden on the Layer 3 device.

Figure 1 Multicast transmission without the multicast VLAN feature



After a multicast VLAN is configured on Switch B, Switch A sends only one copy of the multicast data to the multicast VLAN on Switch B. This method saves network bandwidth and lessens the burden on the Layer 3 device.

Multicast VLAN implementations

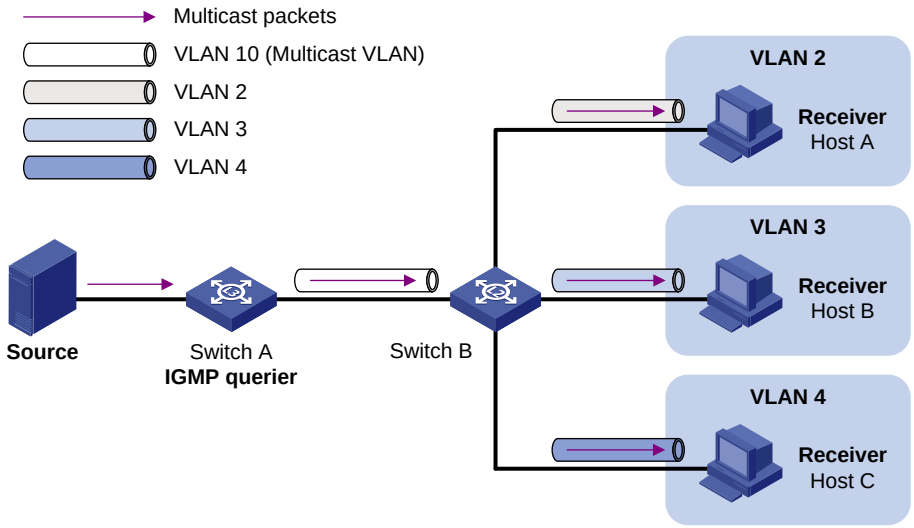
Multicast VLANs include sub-VLAN-based multicast VLANs and port-based multicast VLANs.

Sub-VLAN-based multicast VLAN

As shown in Figure 2:

- Host A, Host B, and Host C are in VLAN 2 through VLAN 4, respectively.
- On Switch B, VLAN 10 is a multicast VLAN. VLAN 2 through VLAN 4 are sub-VLANs of VLAN 10.
- IGMP snooping is enabled for the multicast VLAN and its sub-VLANs.

Figure 2 Sub-VLAN-based multicast VLAN



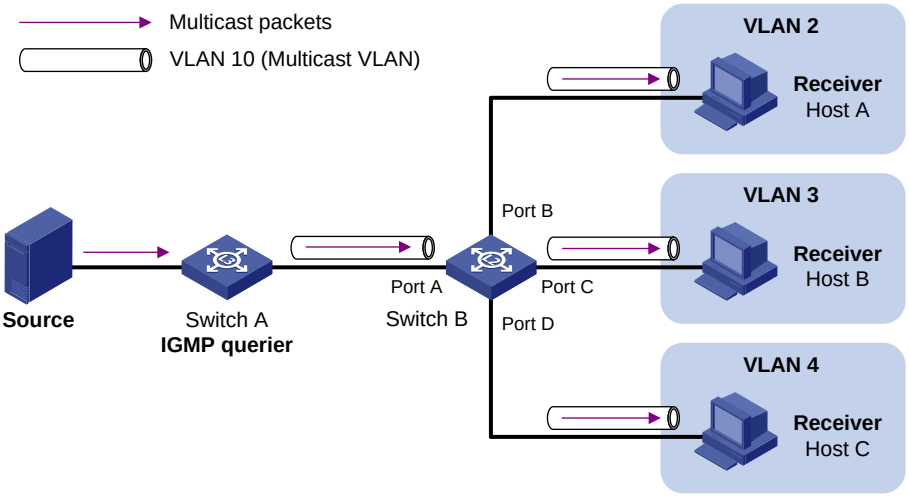
IGMP snooping manages router ports in the multicast VLAN and member ports in each sub-VLAN. When Switch A receives multicast data from the multicast source, it sends only one copy of the multicast data to the multicast VLAN on Switch B. Then, Switch B sends a separate copy to each sub-VLAN in the multicast VLAN.

Port-based multicast VLAN

As shown in Figure 3:

- Host A, Host B, and Host C are in VLAN 2 through VLAN 4, respectively.
- On Switch B, VLAN 10 is a multicast VLAN. All the user ports are hybrid ports and are assigned to VLAN 10.
- IGMP snooping is enabled for the multicast VLAN and VLAN 2 through VLAN 4.

Figure 3 Port-based multicast VLAN



IGMP snooping manages the router ports and member ports in the multicast VLAN. When Switch A receives multicast data from the multicast source, it sends only one copy of the multicast data to the multicast VLAN on Switch B. Switch B sends a separate copy to each user port in the multicast VLAN.

Restrictions and guidelines: Multicast VLAN configuration

The VLAN to be configured as a multicast VLAN must exist.

If you have configured both a sub-VLAN-based multicast VLAN and a port-based multicast VLAN on a device, the port-based multicast VLAN configuration takes effect.

Do not configure a multicast VLAN on a device that is enabled with IP multicast routing. Otherwise, the system displays errors. Do not enable IP multicast routing on a device that is configured with multicast VLANs. Otherwise, multicast forwarding exceptions occur.

Do not configure a VLAN as a multicast VLAN or a multicast sub-VLAN if the VLAN interface is enabled with PIM or IGMP. Otherwise, the system displays errors. Do not enable PIM or IGMP on a VLAN interface if the VLAN interface belongs to a multicast VLAN or a multicast sub-VLAN. Otherwise, multicast forwarding exceptions occur.

Configuring a sub-VLAN-based multicast VLAN

Restrictions and guidelines

The VLANs to be configured as sub-VLANs of a multicast VLAN must exist and cannot be multicast VLANs or sub-VLANs of any other multicast VLAN.

Prerequisites

Before you configure a sub-VLAN-based multicast VLAN, you must complete the following tasks:

- Create VLANs as required.
- Enable IGMP snooping for the VLAN to be configured as the multicast VLAN and for the VLANs to be configured as sub-VLANs.

Procedure

1. Enter system view.
   system-view
2. Configure a VLAN as a multicast VLAN and enter multicast VLAN view.
   multicast-vlan *vlan-id*
   By default, a VLAN is not a multicast VLAN.
3. Assign VLANs to the multicast VLAN as sub-VLANs.
   subvlan *vlan-list*

Configuring a port-based multicast VLAN

Restrictions and guidelines

You can assign user ports to a multicast VLAN in multicast VLAN view or assign a user port to a multicast VLAN in interface view. These configurations have the same priority.

A user port can belong to only one multicast VLAN.

Prerequisites

Before you configure a port-based multicast VLAN, you must complete the following tasks:

- Create VLANs as required.
- Enable IGMP snooping for the VLAN to be configured as the multicast VLAN.

- Enable IGMP snooping for all the VLANs that contain multicast receivers.
- Configure the attributes of user ports. Make sure the ports can forward packets from the VLAN to be configured as the multicast VLAN and send the packets with the VLAN tag removed. For more information about configuring port attributes, see VLAN configuration in *Layer 2—LAN Switching Configuration Guide*.

Assigning user ports to a multicast VLAN in multicast VLAN view

1. Enter system view.
   system-view
2. Configure a VLAN as a multicast VLAN and enter multicast VLAN view.
   multicast-vlan *vlan-id*
   By default, a VLAN is not a multicast VLAN.
3. Assign ports to the multicast VLAN.
   port *interface-list*

Assigning user ports to a multicast VLAN in interface view

1. Enter system view.
   system-view
2. Configure a VLAN as a multicast VLAN and enter multicast VLAN view.
   multicast-vlan *vlan-id*
   By default, a VLAN is not a multicast VLAN.
3. Return to system view.
   quit
4. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
5. Assign the port to the multicast VLAN.
   port multicast-vlan *vlan-id*
   By default, a port does not belong to any multicast VLAN.

Setting the maximum number of multicast VLAN forwarding entries

About this task

You can set the maximum number of multicast VLAN forwarding entries on the device. When the upper limit is reached, the device does not create multicast VLAN forwarding entries until some entries age out or are manually removed.

Procedure

1. Enter system view.
   system-view
2. Set the maximum number of multicast VLAN forwarding entries.
   multicast-vlan entry-limit *limit*
   By default, the maximum number of multicast VLAN forwarding entries is 8192.

Display and maintenance commands for multicast VLAN

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about multicast VLANs.	**display multicast-vlan** [*vlan-id*]
Display information about multicast VLAN forwarding entries.	**display multicast-vlan forwarding-table** [*group-address* [**mask** { *mask-length* | *mask* }] | *source-address* [**mask** { *mask-length* | *mask* }] | **slot** *slot-number* | **subvlan** *vlan-id* | **vlan** *vlan-id*] *
Display information about multicast groups in multicast VLANs.	**display multicast-vlan group** [*source-address* | *group-address* | **slot** *slot-number* | **verbose** | **vlan** *vlan-id*] *
Clear multicast groups in multicast VLANs.	**reset multicast-vlan group** [*source-address* [**mask** { *mask-length* | *mask* }] | *group-address* [**mask** { *mask-length* | *mask* }] | **vlan** *vlan-id*] *

Multicast VLAN configuration examples

Example: Configuring sub-VLAN-based multicast VLAN

Network configuration

As shown in Figure 4:

- Layer 3 device Switch A runs IGMPv2 and acts as the IGMP querier. Layer 2 device Switch B runs IGMPv2 snooping.
- The multicast source sends multicast data to multicast group 224.1.1.1. Receivers Host A, Host B, and Host C belong to VLAN 2, VLAN 3, and VLAN 4, respectively.

Configure a sub-VLAN-based multicast VLAN on Switch B to meet the following requirements:

- Switch A sends the multicast data to Switch B through the multicast VLAN.
- Switch B forwards the multicast data to the receivers in different user VLANs.

Figure 4 Network diagram

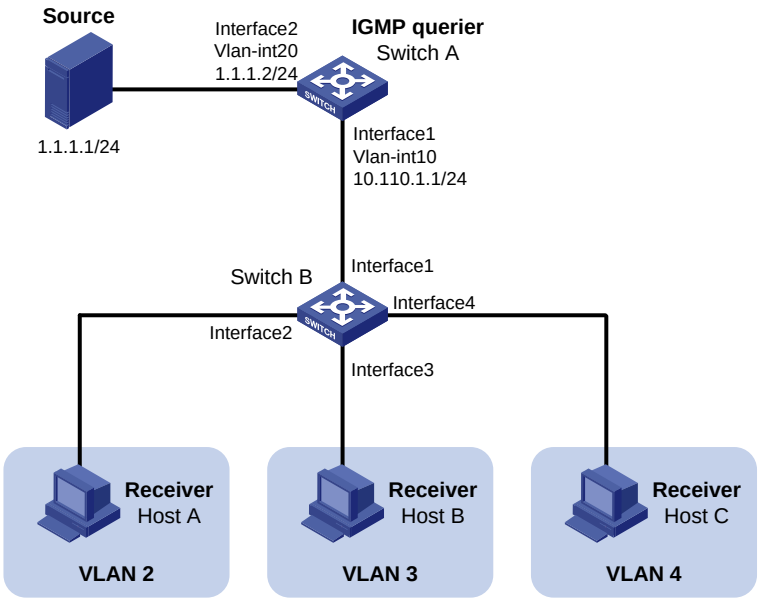


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Switch A:

Enable IP multicast routing.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Create VLAN 20, and assign GigabitEthernet 1/0/2 to the VLAN.

```
[SwitchA] vlan 20
[SwitchA-vlan20] port gigabitethernet 1/0/2
[SwitchA-vlan20] quit
```

Assign an IP address to VLAN-interface 20, and enable PIM-DM on the interface.

```
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] ip address 1.1.1.2 24
[SwitchA-Vlan-interface20] pim dm
[SwitchA-Vlan-interface20] quit
```

Create VLAN 10.

```
[SwitchA] vlan 10
[SwitchA-vlan10] quit
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign the port to VLAN 10 as a tagged VLAN member.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-type hybrid
[SwitchA-GigabitEthernet1/0/1] port hybrid vlan 10 tagged
[SwitchA-GigabitEthernet1/0/1] quit
```

Assign an IP address to VLAN-interface 10, and enable IGMP on the interface.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ip address 10.110.1.1 24
[SwitchA-Vlan-interface10] igmp enable
[SwitchA-Vlan-interface10] quit
```

2. Configure Switch B:

Enable the IGMP snooping feature.

```
<SwitchB> system-view
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

Create VLAN 2, assign GigabitEthernet 1/0/2 to the VLAN, and enable IGMP snooping for the VLAN.

```
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/2
[SwitchB-vlan2] igmp-snooping enable
[SwitchB-vlan2] quit
```

Create VLAN 3, assign GigabitEthernet 1/0/3 to the VLAN, and enable IGMP snooping in the VLAN.

```
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/3
[SwitchB-vlan3] igmp-snooping enable
[SwitchB-vlan3] quit
```

Create VLAN 4, assign GigabitEthernet 1/0/4 to the VLAN, and enable IGMP snooping in the VLAN.

```
[SwitchB] vlan 4
[SwitchB-vlan4] port gigabitethernet 1/0/4
[SwitchB-vlan4] igmp-snooping enable
[SwitchB-vlan4] quit
```

Create VLAN 10, and enable IGMP snooping for the VLAN.

```
[SwitchB] vlan 10
[SwitchB-vlan10] igmp-snooping enable
[SwitchB-vlan10] quit
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign the port to VLAN 10 as a tagged VLAN member.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-type hybrid
[SwitchB-GigabitEthernet1/0/1] port hybrid vlan 10 tagged
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure VLAN 10 as a multicast VLAN, and assign VLAN 2 through VLAN 4 as sub-VLANs to multicast VLAN 10.

```
[SwitchB] multicast-vlan 10
[SwitchB-mvlan-10] subvlan 2 to 4
[SwitchB-mvlan-10] quit
```

Verifying the configuration

Display information about all multicast VLANs on Switch B.

```
[SwitchB] display multicast-vlan
Total 1 multicast VLANs.

Multicast VLAN 10:
  Sub-VLAN list(3 in total):
     2-4
  Port list(0 in total):
```

Display information about multicast groups in multicast VLANs on Switch B.

```
[SwitchB] display multicast-vlan group
Total 1 entries.

Multicast VLAN 10: Total 1 entries.
  (0.0.0.0, 224.1.1.1)
    Sub-VLANs (3 in total):
      VLAN 2
      VLAN 3
      VLAN 4
```

The output shows that multicast group 224.1.1.1 belongs to multicast VLAN 10. Multicast VLAN 10 contains sub-VLANs VLAN 2 through VLAN 4. Switch B will replicate the multicast data of VLAN 10 to VLAN 2 through VLAN 4.

Example: Configuring port-based multicast VLAN

Network configuration

As shown in Figure 5:

- Layer 3 device Switch A runs IGMPv2 and acts as the IGMP querier. Layer 2 device Switch B runs IGMPv2 snooping.
- The multicast source sends multicast data to multicast group 224.1.1.1. Receivers Host A, Host B, and Host C belong to VLAN 2, VLAN 3, and VLAN 4, respectively.

Configure a port-based multicast VLAN on Switch B to meet the following requirements:

- Switch A sends multicast data to Switch B through the multicast VLAN.
- Switch B forwards the multicast data to the receivers in different user VLANs.

Figure 5 Network diagram

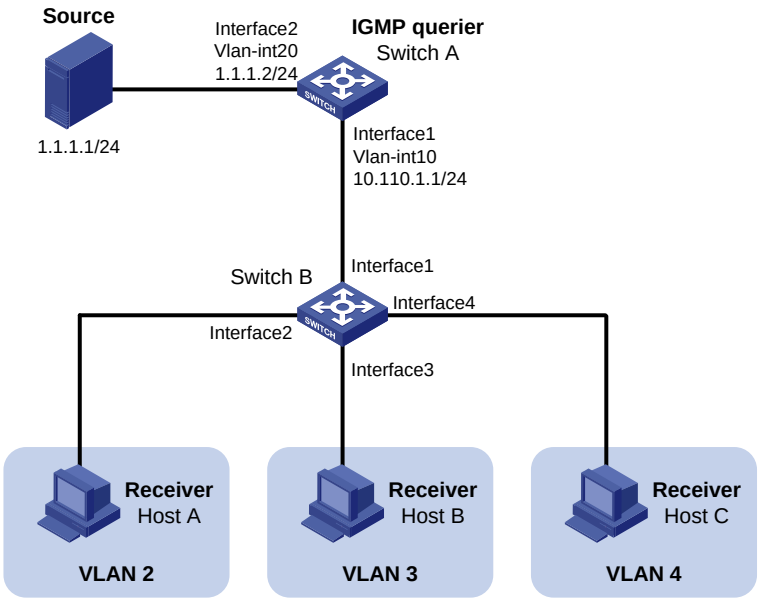


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Switch A:

   # Enable IP multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

   # Create VLAN 20, and assign GigabitEthernet 1/0/2 to the VLAN.

   ```
   [SwitchA] vlan 20
   [SwitchA-vlan20] port gigabitethernet 1/0/2
   [SwitchA-vlan20] quit
   ```

   # Assign an IP address to VLAN-interface 20, and enable PIM-DM on the interface.

   ```
   [SwitchA] interface vlan-interface 20
   [SwitchA-Vlan-interface20] ip address 1.1.1.2 24
   [SwitchA-Vlan-interface20] pim dm
   [SwitchA-Vlan-interface20] quit
   ```

   # Create VLAN 10, and assign GigabitEthernet 1/0/1 to the VLAN.

   ```
   [SwitchA] vlan 10
   [SwitchA-vlan10] port gigabitethernet 1/0/1
   [SwitchA-vlan10] quit
   ```

   # Assign an IP address to VLAN-interface 10, and enable IGMP on the interface.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ip address 10.110.1.1 24
[SwitchA-Vlan-interface10] igmp enable
[SwitchA-Vlan-interface10] quit
```

2. Configure Switch B:

Enable the IGMP snooping feature.

```
<SwitchB> system-view
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

Create VLAN 10, assign GigabitEthernet 1/0/1 to the VLAN, and enable IGMP snooping for the VLAN.

```
[SwitchB] vlan 10
[SwitchB-vlan10] port gigabitethernet 1/0/1
[SwitchB-vlan10] igmp-snooping enable
[SwitchB-vlan10] quit
```

Create VLAN 2, and enable IGMP snooping for the VLAN.

```
[SwitchB] vlan 2
[SwitchB-vlan2] igmp-snooping enable
[SwitchB-vlan2] quit
```

Create VLAN 3, and enable IGMP snooping for the VLAN.

```
[SwitchB] vlan 3
[SwitchB-vlan3] igmp-snooping enable
[SwitchB-vlan3] quit
```

Create VLAN 4, and enable IGMP snooping for the VLAN.

```
[SwitchB] vlan 4
[SwitchB-vlan4] igmp-snooping enable
[SwitchB-vlan4] quit
```

Configure GigabitEthernet 1/0/2 as a hybrid port, and configure VLAN 2 as the PVID of the hybrid port.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-type hybrid
[SwitchB-GigabitEthernet1/0/2] port hybrid pvid vlan 2
```

Assign GigabitEthernet 1/0/2 to VLAN 2 and VLAN 10 as an untagged VLAN member.

```
[SwitchB-GigabitEthernet1/0/2] port hybrid vlan 2 untagged
[SwitchB-GigabitEthernet1/0/2] port hybrid vlan 10 untagged
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a hybrid port, and configure VLAN 3 as the PVID of the hybrid port.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port link-type hybrid
[SwitchB-GigabitEthernet1/0/3] port hybrid pvid vlan 3
```

Assign GigabitEthernet 1/0/3 to VLAN 3 and VLAN 10 as an untagged VLAN member.

```
[SwitchB-GigabitEthernet1/0/3] port hybrid vlan 3 untagged
[SwitchB-GigabitEthernet1/0/3] port hybrid vlan 10 untagged
[SwitchB-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet 1/0/4 as a hybrid port, and configure VLAN 4 as the PVID of the hybrid port.

```
[SwitchB] interface gigabitethernet 1/0/4
```

```
[SwitchB-GigabitEthernet1/0/4] port link-type hybrid
[SwitchB-GigabitEthernet1/0/4] port hybrid pvid vlan 4
```

Assign GigabitEthernet 1/0/4 to VLAN 4 and VLAN 10 as an untagged VLAN member.

```
[SwitchB-GigabitEthernet1/0/4] port hybrid vlan 4 untagged
[SwitchB-GigabitEthernet1/0/4] port hybrid vlan 10 untagged
[SwitchB-GigabitEthernet1/0/4] quit
```

Configure VLAN 10 as a multicast VLAN.

```
[SwitchB] multicast-vlan 10
```

Assign GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 to VLAN 10.

```
[SwitchB-mvlan-10] port gigabitethernet 1/0/2 to gigabitethernet 1/0/3
[SwitchB-mvlan-10] quit
```

Assign GigabitEthernet 1/0/4 to VLAN 10.

```
[SwitchB] interface gigabitethernet 1/0/4
[SwitchB-GigabitEthernet1/0/4] port multicast-vlan 10
[SwitchB-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Display information about multicast VLANs on Switch B.

```
[SwitchB] display multicast-vlan
Total 1 multicast VLANs.

Multicast VLAN 10:
  Sub-VLAN list(0 in total):
  Port list(3 in total):
    GE1/0/2
    GE1/0/3
    GE1/0/4
```

Display dynamic IGMP snooping forwarding entries on Switch B.

```
[SwitchB] display igmp-snooping group
Total 1 entries.

VLAN 10: Total 1 entries.
  (0.0.0.0, 224.1.1.1)
    Host ports (3 in total):
      GE1/0/2        (00:03:23)
      GE1/0/3        (00:04:07)
      GE1/0/4        (00:04:16)
```

The output shows that IGMP snooping maintains the user ports in the multicast VLAN (VLAN 10). Switch B will forward the multicast data of VLAN 10 through these user ports.

Contents

PIM overview 1

About PIM 1
PIM modes 1
PIM-DM 1
Neighbor discovery 1
SPT building 1
Graft 2
Assert 3
PIM-SM 3
Neighbor discovery 4
DR election 4
RP discovery 4
Anycast RP 6
RPT building 7
Multicast source registration 8
Switchover to SPT 9
Assert 9
BIDIR-PIM 9
Neighbor discovery 10
RP discovery 10
DF election 10
Bidirectional RPT building 11
Administrative scoping 12
Administrative scoping mechanism 12
Relationship between admin-scoped zones and the global-scoped zone 13
PIM-SSM 14
Neighbor discovery 14
DR election 14
SPT building 15
Relationship among PIM protocols 15
PIM support for VPNs 16
Protocols and standards 16

Configuring PIM 1

Restrictions and guidelines: PIM configuration 1
Configuring PIM-DM 1
PIM-DM tasks at a glance 1
Prerequisites for PIM-DM 1
Enabling PIM-DM 1
Configuring the state refresh feature 2
Setting the PIM-DM graft retry timer 3
Configuring PIM-SM 3
PIM-SM tasks at a glance 3
Prerequisites for PIM-SM 4
Enabling PIM-SM 4
Configuring static RPs 4
Configuring C-RPs 5
Configuring C-BSRs 5
Configuring the interval for C-BSRs to send BSMs 6
Configuring the timer for C-BSRs to wait BSMs from the BSR 7
Configuring a PIM domain border 7
Disabling BSM semantic fragmentation 8
Disabling the device from forwarding BSMs out of their incoming interfaces 8
Enabling Auto-RP listening 9
Configuring Anycast RP 9
Configuring multicast source registration 10
Configuring the switchover to SPT 11

Configuring the delay time for DR election 11
Configuring BIDIR-PIM 11
BIDIR-PIM tasks at a glance 11
Prerequisites for BIDIR-PIM 12
Enabling BIDIR-PIM 12
Configuring static RPs 13
Configuring C-RPs 13
Configuring C-BSRs 14
Configuring the interval for C-BSRs to send BSMs 15
Configuring the timer for C-BSRs to wait BSMs from the BSR 15
Configuring a PIM domain border 16
Disabling BSM semantic fragmentation 16
Disabling the device from forwarding BSMs out of their incoming interfaces 17
Setting the maximum number of BIDIR-PIM RPs 17
Enabling Auto-RP listening 17
Configuring PIM-SSM 18
PIM-SSM tasks at a glance 18
Prerequisites for PIM-SSM 18
Enabling PIM-SM 18
Configuring the SSM group range 19
Configuring the delay time for DR election 19
Configuring common PIM features 20
Common PIM feature tasks at a glance 20
Configuring a multicast source policy 20
Configuring a PIM hello policy 21
Configuring PIM hello message options 21
Dropping hello messages without the Generation ID option 23
Enabling the device to send hello messages with a different Generation ID upon a port state change 23
Configuring common PIM timers 24
Setting the maximum size of a join or prune message 26
Configuring a PIM join policy 26
Setting the DSCP value for outgoing PIM protocol packets 26
Enabling BFD for PIM 27
Enabling PIM passive mode 27
Enabling PIM NSR 27
Enabling SNMP notifications for PIM 28
Enabling PIM logging 28
Enabling log message suppression for PIM 29
Display and maintenance commands for PIM 29
PIM configuration examples 30
Example: Configuring PIM-DM 30
Example: Configuring non-scoped PIM-SM 34
Example: Configuring admin-scoped PIM-SM 37
Example: Configuring BIDIR-PIM 41
Example: Configuring PIM-SSM 45
Example: Configuring a multicast source policy 48
Troubleshooting PIM 52
A multicast distribution tree cannot be correctly built 52
Multicast data is abnormally terminated on an intermediate device 52
An RP cannot join an SPT in PIM-SM 53
An RPT cannot be built or multicast source registration fails in PIM-SM 53

PIM overview

About PIM

Protocol Independent Multicast (PIM) provides IP multicast forwarding by leveraging unicast static routes or unicast routing tables generated by any unicast routing protocol. PIM uses the underlying unicast routing to generate a multicast routing table without relying on any particular unicast routing protocol. PIM uses the RPF mechanism to implement multicast forwarding. For more information about RPF, see "Configuring multicast routing and forwarding."

PIM modes

Based on the implementation mechanism, PIM includes the following modes:

- Protocol Independent Multicast–Dense Mode (PIM-DM).
- Protocol Independent Multicast–Sparse Mode (PIM-SM).
- Bidirectional Protocol Independent Multicast (BIDIR-PIM).
- Protocol Independent Multicast Source-Specific Multicast (PIM-SSM).

In this document, a PIM domain refers to a network that contains PIM devices.

PIM-DM

PIM-DM uses the push mode for multicast forwarding, and is suitable for small-sized networks with densely distributed multicast members.

PIM-DM assumes that all downstream nodes want to receive multicast data from a source, so multicast data is flooded to all downstream nodes on the network. Branches without downstream receivers are pruned from the forwarding trees. When a pruned branch has new receivers, the graft mechanism turns the pruned branch into a forwarding branch.

In PIM-DM, the multicast forwarding paths for a multicast group constitutes a forwarding tree. The forwarding tree is rooted at the multicast source and has multicast group members as its "leaves." Because the forwarding tree consists of the shortest paths from the multicast source to the receivers, it is also called a "shortest path tree (SPT)."

PIM-DM mechanisms include neighbor discovery, SPT building, graft, and assert.

Neighbor discovery

In a PIM domain, each PIM interface on a device periodically multicasts PIM hello messages to all other PIM devices (identified by the address 224.0.0.13) on the local subnet. Through the exchanging of hello messages, all PIM devices on the subnet determine their PIM neighbors, maintain PIM neighboring relationship with other devices, and build and maintain SPTs.

SPT building

The process of building an SPT is the flood-and-prune process:

1. In a PIM-DM domain, the multicast data from the multicast source S to the multicast group G is flooded throughout the domain. A device performs an RPF check on the multicast data. If the RPF check succeeds, the device creates an (S, G) entry and forwards the data to all

downstream nodes on the network. In the flooding process, all the devices in the PIM-DM domain create the (S, G) entry.

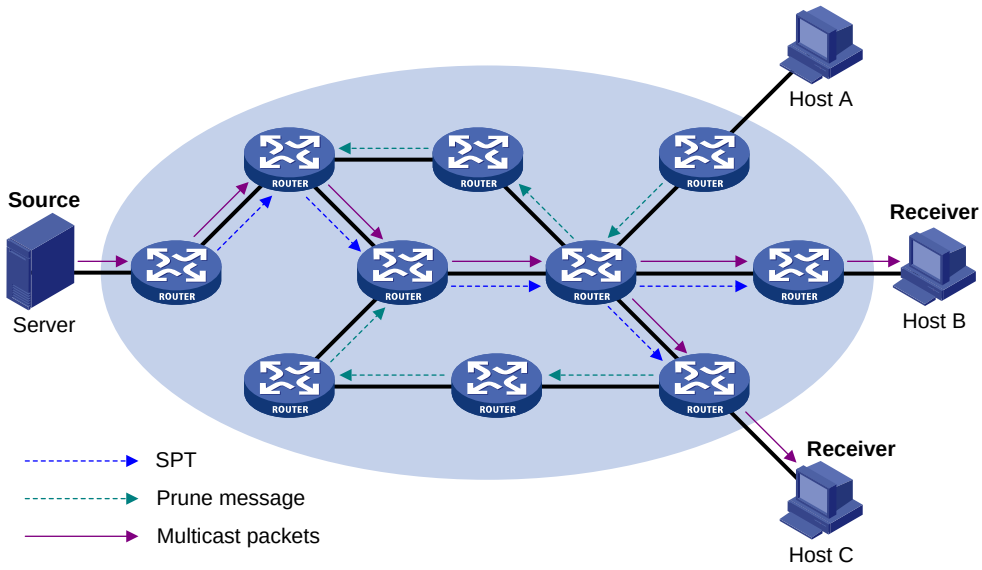
2. The nodes without downstream receivers are pruned. A device that has no downstream receivers multicasts a prune message to all PIM devices on the subnet. When an upstream node receives the prune message, it removes the receiving interface from the (S, G) entry. In this way, the upstream stream node stops forwarding subsequent packets addressed to that multicast group down to this node.

NOTE:

An (S, G) entry contains a multicast source address S, a multicast group address G, an outgoing interface list, and an incoming interface.

A prune process is initiated by a leaf device. As shown in Figure 1, the device interface that does not have any downstream receivers initiates a prune process by sending a prune message toward the multicast source. This prune process goes on until only necessary branches are left in the PIM-DM domain, and these necessary branches constitute an SPT.

Figure 1 SPT building



The pruned state of a branch has a finite holdtime timer. When the timer expires, multicast data is again forwarded to the pruned branch. The flood-and-prune cycle takes place periodically to maintain the forwarding branches.

Graft

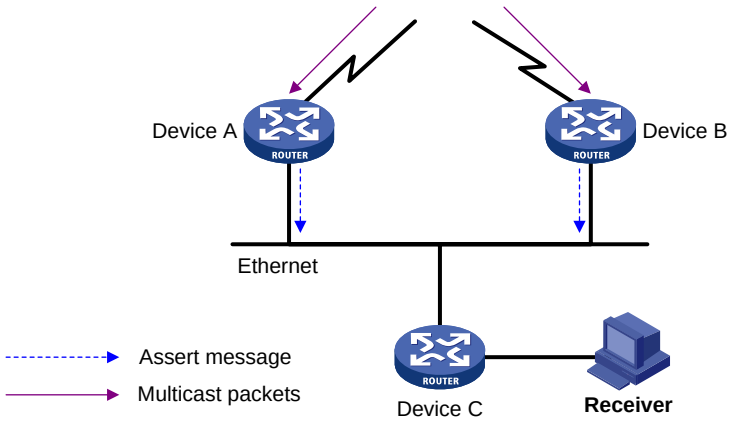
A previously pruned branch might have new downstream receivers. To reduce the latency for resuming the forwarding capability of this branch, a graft mechanism is used as follows:

1. The node that needs to receive the multicast data sends a graft message to its upstream node, telling it to rejoin the SPT.
2. After receiving this graft message on an interface, the upstream node adds the receiving interface to the outgoing interface list of the (S, G) entry. It also sends a graft-ack message to the graft sender.
3. If the graft sender receives a graft-ack message, the graft process finishes. Otherwise, the graft sender continues to send graft messages at a graft retry interval until it receives an acknowledgment from its upstream node.

Assert

On a subnet with more than one multicast device, the assert mechanism shuts off duplicate multicast flows to the network. It does this by electing a unique multicast forwarder for the subnet.

Figure 2 Assert mechanism



As shown in Figure 2, after Device A and Device B receive an (S, G) packet from the upstream node, they both forward the packet to the local subnet. As a result, the downstream node Device C receives two identical multicast packets. In addition, both Device A and Device B, on their downstream interfaces, receive a duplicate packet forwarded by the other. After detecting this condition, both devices send an assert message to all PIM devices (224.0.0.13) on the local subnet through the interface that received the packet. The assert message contains the multicast source address (S), the multicast group address (G), and the metric preference and metric of the unicast route/MBGP route/static multicast route to the multicast source. By comparing these parameters, either Device A or Device B becomes the unique forwarder of the subsequent (S, G) packets on the shared-media LAN. The comparison process is as follows:

1. The device with a higher metric preference to the multicast source wins.
2. If both devices have the same metric preference, the device with a smaller metric wins.
3. If both devices have the same metric, the device with a higher IP address on the downstream interface wins.

PIM-SM

PIM-DM uses the flood-and-prune cycles to build SPTs for multicast data forwarding. Although an SPT has the shortest paths from the multicast source to the receivers, it is built with a low efficiency. Therefore, PIM-DM is not suitable for large and medium-sized networks. PIM-SM uses the pull mode for multicast forwarding, and it is suitable for large- and medium-sized networks with sparsely and widely distributed multicast group members.

PIM-SM assumes that no hosts need multicast data. A multicast receiver must express its interest in the multicast data for a multicast group before the data is forwarded to it. A rendezvous point (RP) is the core of a PIM-SM domain. Relying on the RP, SPTs and rendezvous point trees (RPTs) are established and maintained to implement multicast data forwarding. An SPT is rooted at the multicast source and has the RPs as its leaves. An RPT is rooted at the RP and has the receiver hosts as its leaves.

PIM-SM mechanisms include neighbor discovery, DR election, RP discovery, Anycast RP, RPT building, multicast source registration, switchover to SPT, and assert.

Neighbor discovery

PIM-SM uses the same neighbor discovery mechanism as PIM-DM does. For more information, see "Neighbor discovery."

DR election

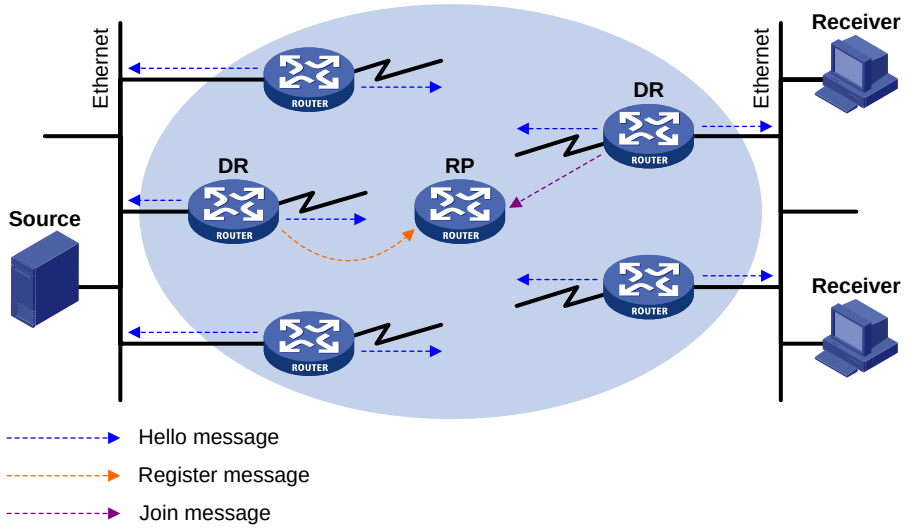
A designated router (DR) is required on both the source-side network and receiver-side network. A source-side DR acts on behalf of the multicast source to send register messages to the RP. The receiver-side DR acts on behalf of the multicast receivers to send join messages to the RP.

PIM-DM does not require a DR. However, if IGMPv1 runs on any shared-media LAN in a PIM-DM domain, a DR must be elected to act as the IGMPv1 querier for the LAN. For more information about IGMP, see "Configuring IGMP."

IMPORTANT:

IGMP must be enabled on the device that acts as the receiver-side DR. Otherwise, the receiver hosts attached to the DR cannot join any multicast groups.

Figure 3 DR election



As shown in Figure 3, the DR election process is as follows:

1. The devices on the shared-media LAN send hello messages to one another. The hello messages contain the DR priority for DR election. The device with the highest DR priority is elected as the DR.
2. The device with the highest IP address wins the DR election under one of following conditions:
   - All the devices have the same DR election priority.
   - A device does not support carrying the DR priority in hello messages.

If the DR fails, its PIM neighbor lifetime expires and the other devices will initiate to elect a new DR.

RP discovery

An RP is the core of a PIM-SM domain. A multicast group can have only one RP for multicast forwarding, and an RP can be designated to multiple multicast groups.

RP selection

An RP can be statically configured or dynamically elected. The priorities of the static RP and the dynamic RP depend on whether the static RP is preferred when the static RP is configured.

- If the static RP is preferred, the static RP takes precedence over the dynamic RP. The dynamic RP takes services over only when the static RP fails.
- If the static RP is not preferred, the dynamic RP takes precedence over the static RP. The static RP takes services over only when the dynamic RP fails.

Static RPs

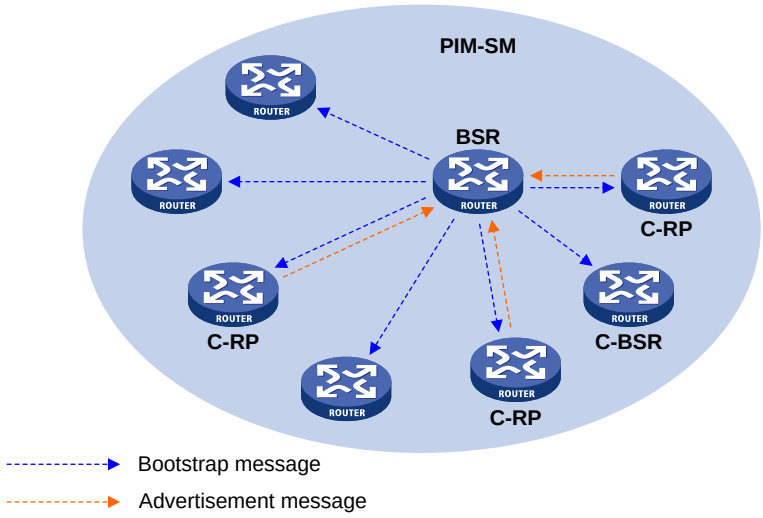
A static RP can avoid a single point of failure and save network bandwidth that is consumed by frequent communications between C-RPs and the BSR.

Dynamic RP election

Dynamic RP election is implemented through the BSR mechanism. BSR mechanism includes the following roles:

- **Candidate-RPs (C-RPs)**—An RP is dynamically elected from C-RPs to provide services to a multicast group.
- **BSR**—A BSR is the core of the administrative core of the PIM-SM domain. It is responsible for collecting and advertising RP information in the whole domain. A PIM-SM domain has only one BSR, and the BSR is elected from C-BSRs.
- **Candidate-BSRs (C-BSRs)**—Any devices in the PIM-SM domain can act as C-BSRs and the BSR is elected from the C-BSRs. Once the BSR fails, a new BSR is elected from the C-BSRs to avoid multicast traffic interruption.

Figure 4 Information exchange between C-RPs and BSR



As shown in Figure 4, an RP is elected as follows:

1. Each C-BSR sends a bootstrap message (BSM) to other devices in the PIM-SM domain.
2. When a C-BSR receives a BSM from another C-BSR, it compares its own priority with the priority carried in the message. The C-BSR with a higher priority wins the BSR election. If a tie exists in the priority, the C-BSR with a higher IP address wins. The loser uses the winner's BSR address to replace its own BSR address and no longer regards itself as the BSR. The winner retains its own BSR address and continues to regard itself as the BSR.
3. Each C-RP periodically unicasts an advertisement message to the BSR. An advertisement message contains the address of the advertising C-RP and the multicast group range to which it is designated.

4. The BSR collects these advertisement messages and organizes the C-RP information into an RP-set, which is a database of mappings between multicast groups and RPs. The BSR encapsulates the RP-set information in the bootstrap messages (BSMs) and floods the BSMs to the entire PIM-SM domain.
5. All devices in the PIM-SM domain select an RP for a multicast group based on the following rules:
   a. The C-RP that is designated to the smallest multicast group range wins.
   b. If the C-RPs are designated to the same group ranges, the C-RP with the highest priority wins.
   c. If the C-RPs have the same priority, the C-RP with the largest hash value wins. The hash value is calculated through the hash algorithm.
   d. If the C-RPs have the same hash value, the C-RP with the highest IP address wins.

Anycast RP

PIM-SM requires only one active RP to serve each multicast group. If the active RP fails, the multicast traffic might be interrupted. The Anycast RP mechanism enables redundancy backup among RPs by configuring multiple RPs with the same IP address. A multicast source registers with the closest RP or a receiver joins the closest RP to implement source information synchronization.

Anycast RP has the following benefits:

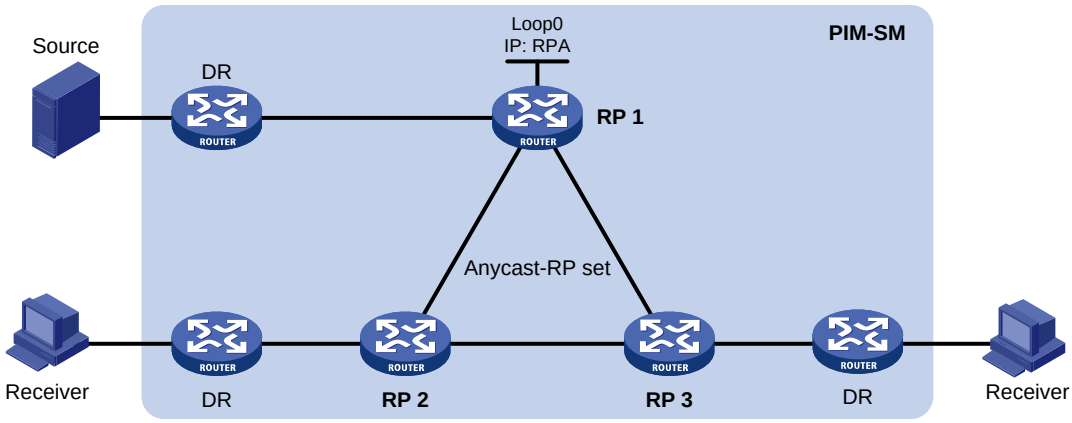
- **Optimal RP path**—A multicast source registers with the closest RP to build an optimal SPT. A receiver joins the closest RP to build an optimal RPT.
- **Redundancy backup among RPs**—When an RP fails, the RP-related sources will register with the closest available RPs and the receiver-side DRs will join the closest available RPs. This provides redundancy backup among RPs.

Anycast RP is implemented in either of the following methods:

- **Anycast RP through MSDP**—In this method, you can configure multiple RPs with the same IP address for one multicast group and configure MSDP peering relationships between them. For more information about Anycast RP through MSDP, see "Configuring MSDP."
- **Anycast RP through PIM-SM**—In this method, you can configure multiple RPs for one multicast group and add them to an Anycast RP set. This method introduces the following concepts:
  - **Anycast RP set**—A set of RPs that are designated to the same multicast group.
  - **Anycast RP member**—Each RP in the Anycast RP set.
  - **Anycast RP member address**—IP address of each Anycast RP member for communication among the RP members.
  - **Anycast RP address**—IP address of the Anycast RP set for communication within the PIM-SM domain. It is also known as RPA.

  As shown in Figure 5, RP 1, RP 2, and RP 3 are members of an Anycast RP set.

Figure 5 Anycast RP through PIM-SM



The following describes how Anycast RP through PIM-SM is implemented:

a. RP 1 receives a register message destined to RPA. Because the message is not from other Anycast RP members (RP 2 or RP 3), RP 1 considers that the register message is from the DR. RP 1 changes the source IP address of the register message to its own address and sends the message to the other members (RP 2 and RP 3).

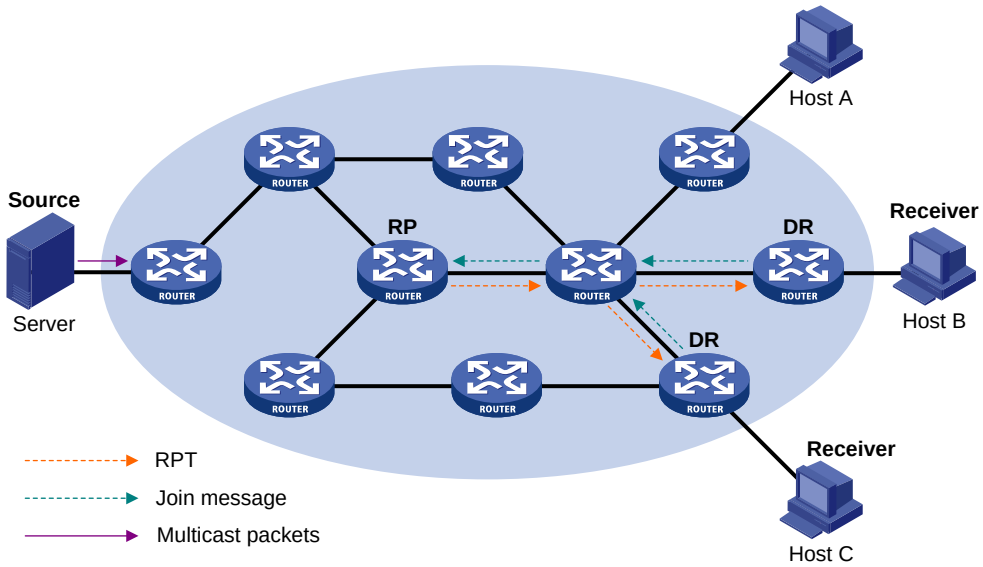
If a device acts as both a DR and an RP, it creates a register message, and then forwards the message to the other RP members.

b. After receiving the register message, RP 2 and RP 3 find out that the source address of the register message is an Anycast RP member address. They stop forwarding the message to other devices.

In Anycast RP implementation, an RP must forward the register message from the DR to other Anycast RP members to synchronize multicast source information.

RPT building

Figure 6 RPT building in a PIM-SM domain



As shown in Figure 6, the process of building an RPT is as follows:

1. When a receiver wants to join the multicast group G, it uses an IGMP message to inform the receiver-side DR.
2. After getting the receiver information, the DR sends a join message, which travels hop by hop to the RP for the multicast group.
3. The devices along the path from the DR to the RP form an RPT branch. Each device on this branch adds to its forwarding table a (*, G) entry, where the asterisk (*) represents any multicast source. The RP is the root of the RPT, and the DR is a leaf of the RPT.

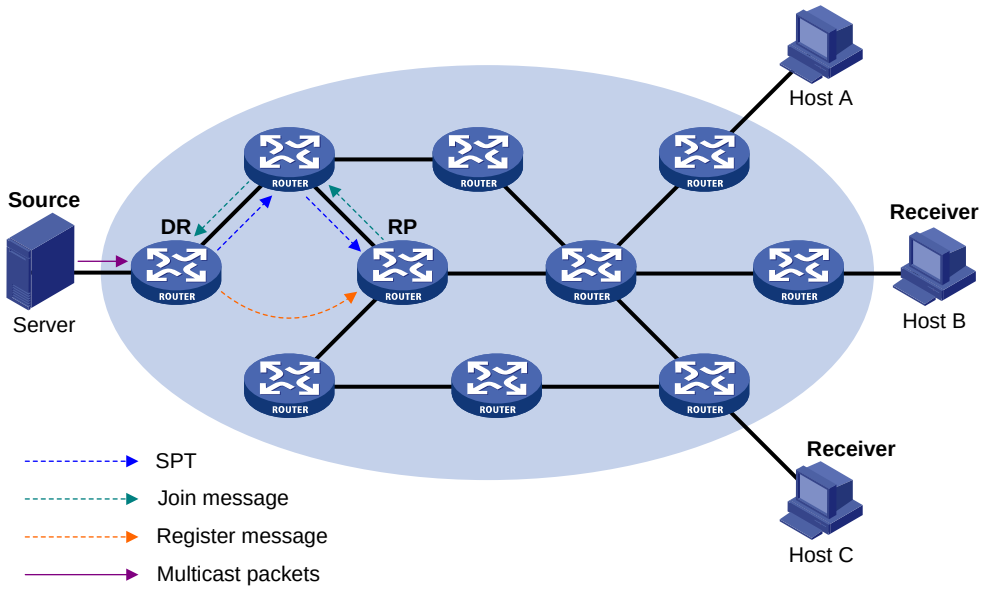
When the multicast data addressed to the multicast group G reaches the RP, the RP forwards the data to the DR along the established RPT, and finally to the receiver.

When a receiver is no longer interested in the multicast data addressed to the multicast group G, the receiver-side DR sends a prune message. The prune message goes hop by hop along the RPT to the RP. After receiving the prune message, the upstream node deletes the interface that connects to this downstream node from the outgoing interface list. At the same time, the upstream device checks for the existence of receivers for that multicast group. If no receivers for the multicast group exist, the device continues to forward the prune message to its upstream device.

Multicast source registration

The multicast source uses the registration process to inform an RP of its presence.

Figure 7 Multicast source registration



As shown in Figure 7, the multicast source registers with the RP as follows:

1. The multicast source S sends the first multicast packet to the multicast group G. When receiving the multicast packet, the source-side DR encapsulates the packet into a PIM register message and unicasts the message to the RP.
2. After the RP receives the register message, it decapsulates the register message and forwards the register message down to the RPT. Meanwhile, it sends an (S, G) source-specific join message toward the multicast source. The devices along the path from the RP to the multicast source constitute an SPT branch. Each device on this branch creates an (S, G) entry in its forwarding table.
3. The subsequent multicast data from the multicast source are forwarded to the RP along the established SPT. When the multicast data reaches the RP along the SPT, the RP forwards the data to the receivers along the RPT. Meanwhile, it unicasts a register-stop message to the source-side DR to prevent the DR from unnecessarily encapsulating the data.

Switchover to SPT

This section takes a router as an example. As compared with the switchover to SPT on a router, switchover to SPT is implemented in a simpler way on a switch.

In a PIM-SM domain, only one RP and one RPT provide services for a specific multicast group. Before the switchover to SPT occurs, the source-side DR encapsulates all multicast data in register messages and sends them to the RP. After receiving these register messages, the RP decapsulates them and forwards them to the receiver-side DR along the RPT.

Multicast forwarding along the RPT has the following weaknesses:

- Encapsulation and decapsulation are complex on the source-side DR and the RP.
- The path for a multicast packet might not be the shortest one.
- The RP might be overloaded by multicast traffic bursts.

To eliminate these weaknesses, PIM-SM allows an RP or the receiver-side DR to initiate the switchover to SPT when the RP or DR receives multicast traffic.

- The RP initiates the switchover to SPT:

  If the RP receives multicast traffic, it sends an (S, G) source-specific join message toward the multicast source. The routers along the path from the RP to the multicast source constitute an SPT. The subsequent multicast data is forwarded to the RP along the SPT without being encapsulated into register messages.

  For more information about the switchover to SPT initiated by the RP, see "Multicast source registration."

- The receiver-side DR initiates the switchover to SPT:

  If the receiver-side DR receives multicast traffic, it initiates the switchover to SPT as follows:

  a. The receiver-side DR sends an (S, G) source-specific join message toward the multicast source. The routers along the path create an (S, G) entry in their forwarding table to constitute an SPT branch.

  b. When the multicast packets reach the router where the RPT and the SPT branches, the router drops the multicast packets that travel along the RPT. It then sends a prune message with the RP bit toward the RP.

  c. After receiving the prune message, the RP forwards it toward the multicast source (supposed only one receiver exists). Thus, the switchover to SPT is completed. The subsequent multicast packets travel along the SPT from the multicast source to the receiver hosts.

With the switchover to SPT, PIM-SM builds SPTs more economically than PIM-DM does.

Assert

PIM-SM uses a similar assert mechanism as PIM-DM does. For more information, see "Assert."

BIDIR-PIM

In some many-to-many applications, such as a multi-side video conference, multiple receivers of a multicast group might be interested in the multicast data from multiple multicast sources. With PIM-DM or PIM-SM, each device along the SPT must create an (S, G) entry for each multicast source, consuming a lot of system resources.

BIDIR-PIM addresses the problem. Derived from PIM-SM, BIDIR-PIM builds and maintains a bidirectional RPT, which is rooted at the RP and connects the multicast sources and the receivers. Along the bidirectional RPT, the multicast sources send multicast data to the RP, and the RP

forwards the data to the receivers. Each device along the bidirectional RPT needs to maintain only one (*, G) entry, saving system resources.

BIDIR-PIM is suitable for a network with dense multicast sources and receivers.

BIDIR-PIM mechanisms include neighbor discovery, RP discovery, DF election, and bidirectional RPT building.

Neighbor discovery

BIDIR-PIM uses the same neighbor discovery mechanism as PIM-SM does. For more information, see "Neighbor discovery."

RP discovery

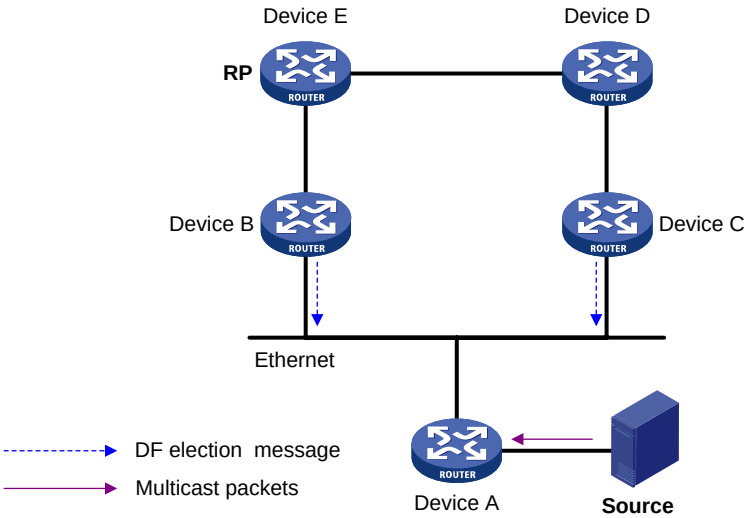
BIDIR-PIM uses the same RP discovery mechanism as PIM-SM does. For more information, see "RP discovery." In BIDIR-PIM, an RPF interface is the interface toward an RP, and an RPF neighbor is the address of the next hop to the RP.

DF election

On a subnet with multiple multicast devices, duplicate multicast packets might be forwarded to the RP. To address this issue, BIDIR-PIM uses a designated forwarder (DF) election mechanism to elect a unique DF for each RP on a subnet. Only the DFs can forward multicast data to the RP.

DF election is not necessary for an RPL.

Figure 8 DF election



As shown in Figure 8, without the DF election mechanism, both Device B and Device C can receive multicast packets from Route A. They also can forward the packets to downstream devices on the local subnet. As a result, the RP (Device E) receives duplicate multicast packets.

With the DF election mechanism, once receiving the RP information, Device B and Device C multicast a DF election message to all PIM devices (224.0.0.13) to initiate a DF election process. The election message carries the RP's address, and the route preference and the metric of the unicast route, MBGP route, or static multicast route to the RP. A DF is elected as follows:

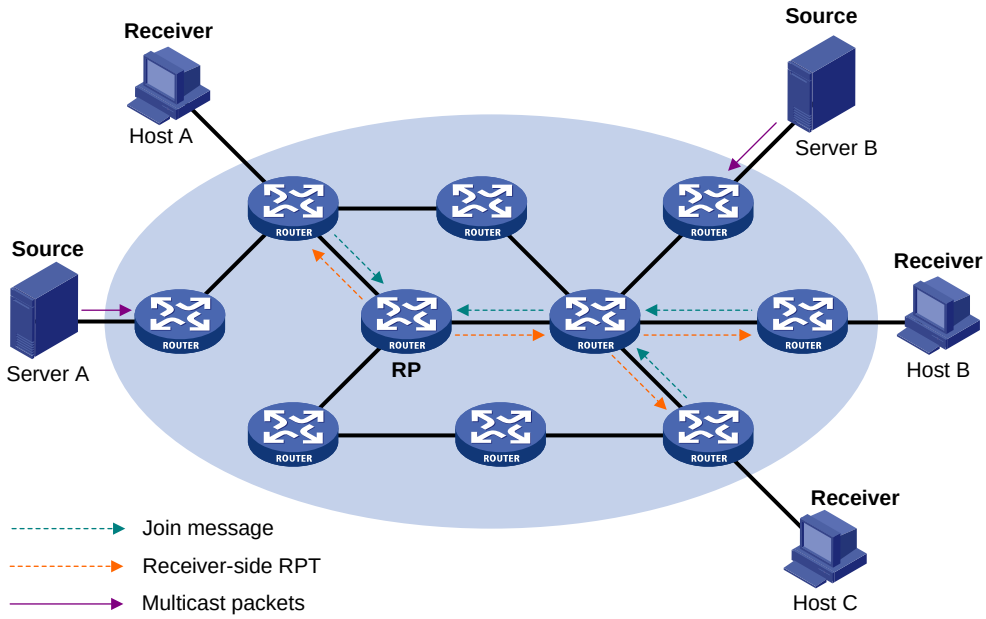
1. The device with a higher route preference becomes the DF.
2. If the devices have the same route preference, the device with a lower metric becomes the DF.

3. If the devices have the same metric, the device with a higher IP address becomes the DF.

Bidirectional RPT building

A bidirectional RPT comprises a receiver-side RPT and a source-side RPT. The receiver-side RPT is rooted at the RP and takes the devices that directly connect to the receivers as leaves. The source-side RPT is also rooted at the RP but takes the devices that directly connect to the sources as leaves. The processes for building these two RPTs are different.

Figure 9 RPT building at the receiver side

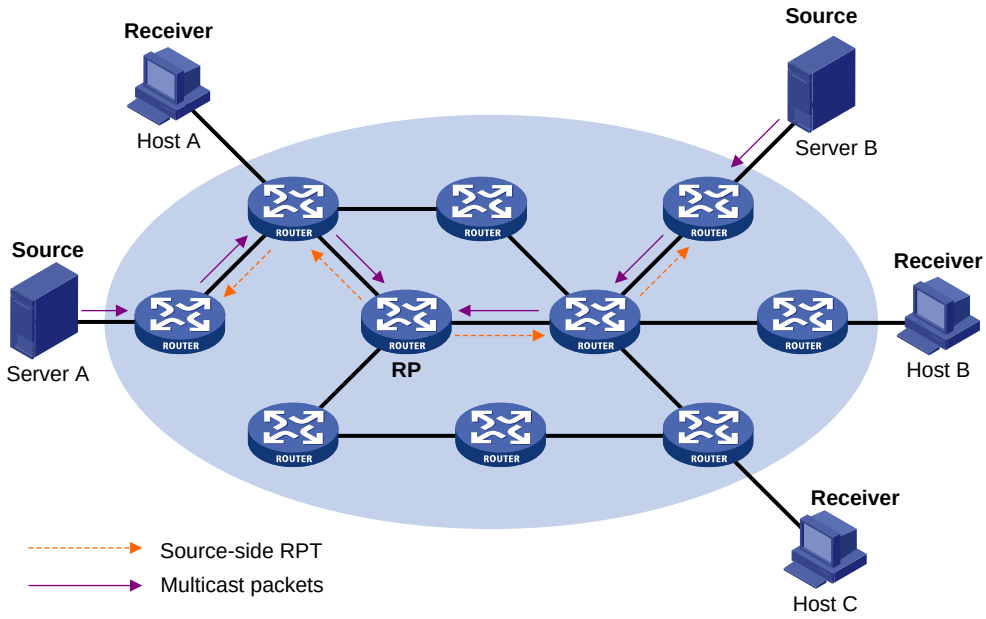


As shown in Figure 9, the process for building a receiver-side RPT is the same as the process for building an RPT in PIM-SM:

1. When a receiver wants to join the multicast group G, it uses an IGMP message to inform the directly connected device.
2. After receiving the message, the device sends a join message, which is forwarded hop by hop to the RP for the multicast group.
3. The devices along the path from the receiver's directly connected device to the RP form an RPT branch. Each device on this branch adds a (*, G) entry to its forwarding table.

After a receiver host leaves the multicast group G, the directly connected device multicasts a prune message to all PIM devices on the subnet. The prune message goes hop by hop along the reverse direction of the RPT to the RP. After receiving the prune message, an upstream node removes the interface that connects to the downstream node from the outgoing interface list. At the same time, the upstream device checks the existence of receivers for that multicast group. If no receivers for the multicast group exist, the device continues to forward the prune message to its upstream device.

Figure 10 RPT building at the multicast source side



As shown in Figure 10, the process for building a source-side RPT is relatively simple:

4. When a multicast source sends multicast packets to the multicast group G, the DF in each subnet unconditionally forwards the packets to the RP.
5. The devices along the path from the source's directly connected device to the RP constitute an RPT branch. Each device on this branch adds to its forwarding table a (*, G) entry.

After a bidirectional RPT is built, the multicast sources send multicast traffic to the RP along the source-side RPT. Then, the RP forwards the traffic to the receivers along the receiver-side RPT.

IMPORTANT:

If a receiver and a multicast source are at the same side of the RP, the source-side RPT and the receiver-side RPT might meet at a node before reaching the RP. In this case, the multicast packets from the multicast source to the receiver are directly forwarded by the node, instead of by the RP.

Administrative scoping

Administrative scoping mechanism

Administrative scoping mechanisms for PIM-SM/BIDIR-PIM include non-scoped BSR mechanism and administrative scoping mechanism.

- **Non-scoped BSR mechanism**—A domain contains only one BSR, which is responsible for advertising RP-set information within the entire domain. The information about all multicast groups is forwarded within the network that the BSR administers.
- **Administrative scoping mechanism**—To implement refined management, you can divide a domain into a global-scoped zone and multiple administratively-scoped zones (admin-scoped zones). The administrative scoping mechanism effectively releases stress on the management in a single-BSR domain and enables provision of zone-specific services through private group addresses.

Admin-scoped zones are divided for multicast groups. Zone border devices (ZBRs) form the boundary of an admin-scoped zone. Each admin-scoped zone maintains one BSR for multicast groups within a specific range. Multicast protocol packets, such as assert messages and BSMs, for a

specific group range cannot cross the boundary of the admin-scoped zone for the group range. Multicast group ranges that are associated with different admin-scoped zones can have intersections. However, the multicast groups in an admin-scoped zone are valid only within the local zone, and theses multicast groups are regarded as private group addresses.

The global-scoped zone maintains a BSR for the multicast groups that do not belong to any admin-scoped zones.

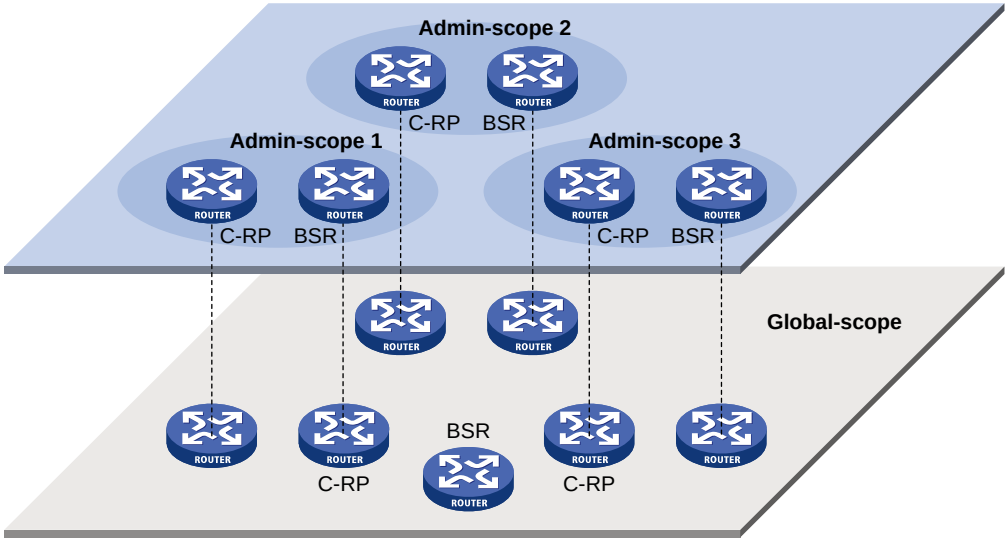
Relationship between admin-scoped zones and the global-scoped zone

The global-scoped zone and each admin-scoped zone have their own C-RPs and BSRs. These devices are effective only on their respective zones, and the BSR election and the RP election are implemented independently. Each admin-scoped zone has its own boundary. The multicast information within a zone cannot cross this boundary in either direction. You can have a better understanding of the global-scoped zone and admin-scoped zones based on geographical locations and multicast group address ranges.

- In view of geographical locations:

  An admin-scoped zone is a logical zone for particular multicast groups. The multicast packets for such multicast groups are confined within the local admin-scoped zone and cannot cross the boundary of the zone.

  Figure 11 Relationship in view of geographical locations

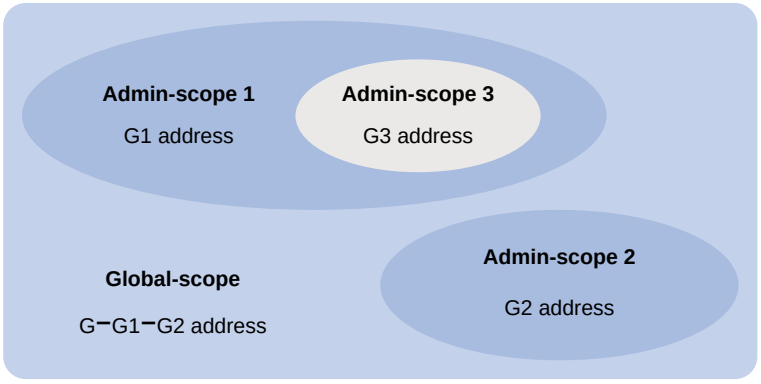


  As shown in Figure 11, for the multicast groups in a specific group address range, the admin-scoped zones must be geographically separated and isolated. A device cannot belong to multiple admin-scoped zones. An admin-scoped zone cannot contain a device that belongs to any other admin-scoped zone. However, the global-scoped zone includes all devices in the PIM-SM domain or BIDIR-PIM domain. Multicast packets that do not belong to any admin-scoped zones are forwarded in the entire PIM-SM domain or BIDIR-PIM domain.

- In view of multicast group address ranges:

  Each admin-scoped zone is designated to specific multicast groups, of which the multicast group addresses are valid only within the local zone. The multicast groups of different admin-scoped zones might have intersections. All the multicast groups other than those of the admin-scoped zones use the global-scoped zone.

Figure 12 Relationship in view of multicast group address ranges



As shown in Figure 12, the admin-scoped zones 1 and 2 have no intersection, but the admin-scoped zone 3 is a subset of the admin-scoped zone 1. The global-scoped zone provides services for all the multicast groups that are not covered by the admin-scoped zones 1 and 2, G—G1—G2 in this case.

PIM-SSM

The ASM model includes PIM-DM and PIM-SM. The SSM model can be implemented by leveraging part of the PIM-SM technique. It is also called "PIM-SSM."

The SSM model provides a solution for source-specific multicast. It maintains the relationship between hosts and devices through IGMPv3.

In actual applications, part of IGMPv3 and PIM-SM techniques are adopted to implement the SSM model. In the SSM model, because receivers have located a multicast source, no RP or RPT is required. Multicast sources do not register with the RP, and the MSDP is not needed for discovering multicast sources in other PIM domains.

PIM-SSM mechanisms include neighbor discovery, DR election, and SPT building.

Neighbor discovery

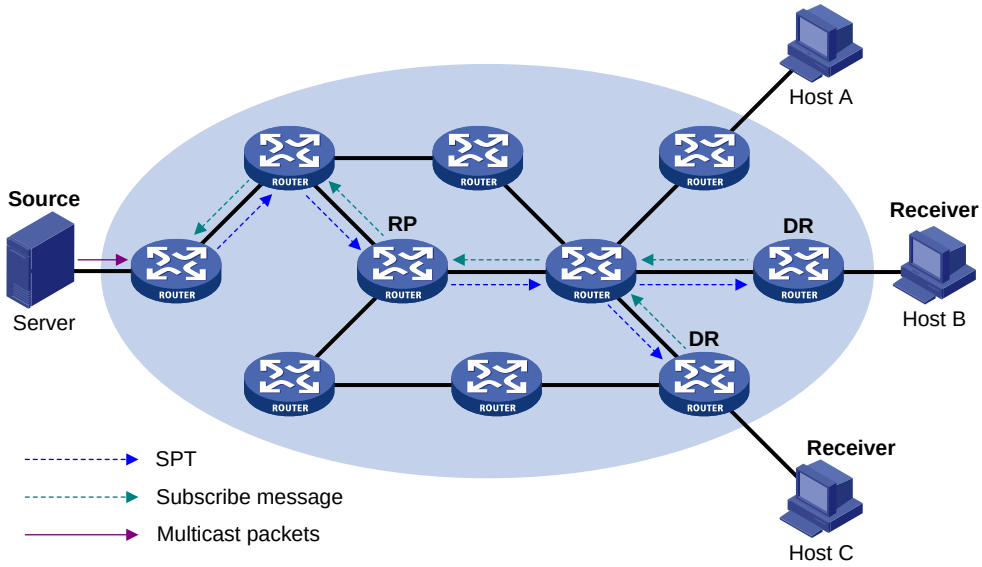
PIM-SSM uses the same neighbor discovery mechanism as PIM-SM. For more information, see "Neighbor discovery."

DR election

PIM-SSM uses the same DR election mechanism as PIM-SM. For more information, see "DR election."

SPT building

Figure 13 SPT building in PIM-SSM



As shown in Figure 13, an SPT is built as follows:

1. Host B and Host C are receivers. They send IGMPv3 report messages to their DRs to express their interest in the multicast information that the multicast source S sends to the multicast group G.
2. After receiving a report message, the DR sends a subscribe message hop by hop toward the multicast source S. All devices along the path from the DR to the source create an (S, G) entry to build an SPT. The SPT is rooted at the multicast source S and has the receivers as its leaves. This SPT is the transmission channel in PIM-SSM.

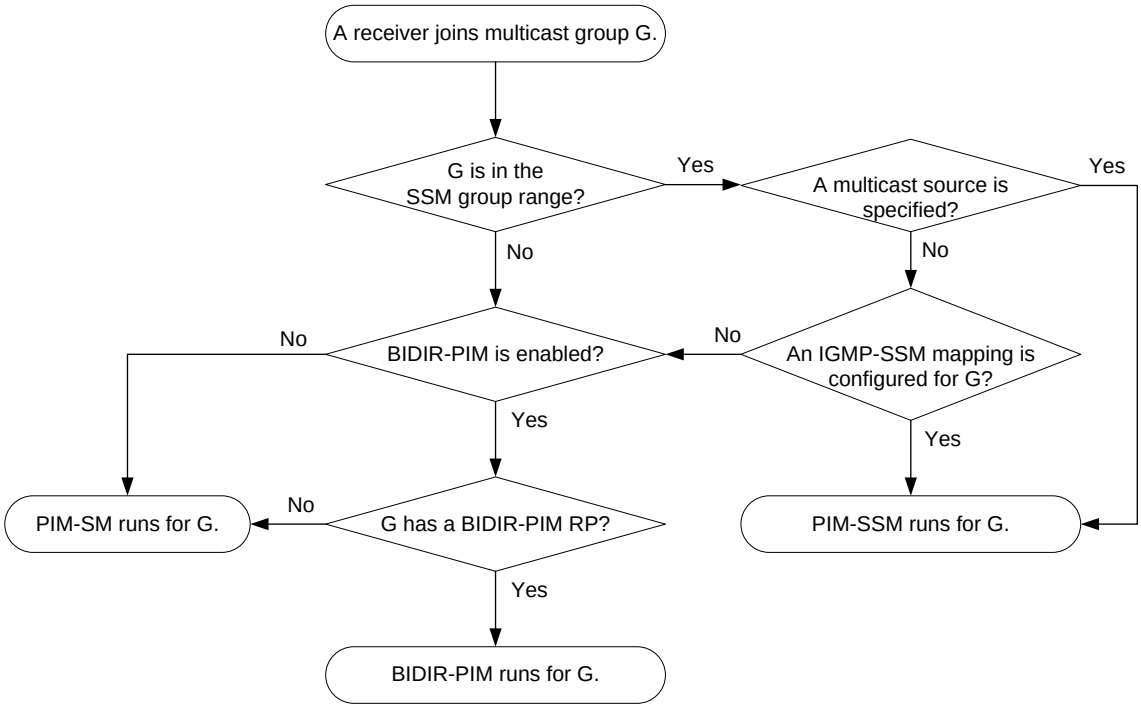
In PIM-SSM, the term "subscribe message" refers to a join message.

Relationship among PIM protocols

In a PIM network, PIM-DM cannot run together with PIM-SM, BIDIR-PIM, or PIM-SSM. However, PIM-SM, BIDIR-PIM, and PIM-SSM can run together. Figure 14 shows how the device selects one protocol from among them for a receiver trying to join a group.

For more information about IGMP SSM mapping, see "Configuring IGMP."

Figure 14 Relationship among PIM protocols



PIM support for VPNs

To support PIM for VPNs, a multicast device that runs PIM maintains an independent set of PIM neighbor table, multicast routing table, BSR information, and RP-set information for each VPN.

After receiving a multicast data packet, the multicast device checks which VPN the data packet belongs to. Then, the device forwards the packet according to the multicast routing table for that VPN or creates a multicast routing entry for that VPN.

Protocols and standards

- RFC 3973, *Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification(Revised)*
- RFC 4601, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised)*
- RFC 4610, *Anycast-RP Using Protocol Independent Multicast (PIM)*
- RFC 5015, *Bidirectional Protocol Independent Multicast (BIDIR-PIM)*
- RFC 5059, *Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)*
- RFC 4607, *Source-Specific Multicast for IP*
- Draft-ietf-ssm-overview-05, *An Overview of Source-Specific Multicast (SSM)*

Configuring PIM

Restrictions and guidelines: PIM configuration

All the interfaces on a device must operate in the same PIM mode on the public network or the same VPN instance.

When both PIM-SM and BIDIR-PIM run on the PIM network, do not use the same RP to provide services for PIM-SM and BIDIR-PIM. Otherwise, exceptions might occur to the PIM routing table.

Configuring PIM-DM

PIM-DM tasks at a glance

To configure PIM-DM, perform the following tasks:

1. Enabling PIM-DM
2. (Optional.) Configuring the state refresh feature
3. (Optional.) Setting the PIM-DM graft retry timer
4. (Optional.) Configuring common PIM timers

Prerequisites for PIM-DM

Before you configure PIM-DM, complete the following tasks:

- Configure a unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable IP multicast routing.

Enabling PIM-DM

About this task

With PIM-DM enabled on interfaces, devices can establish PIM neighbor relationship and process PIM messages from their PIM neighbors.

Restrictions and guidelines

As a best practice, enable PIM-DM on all non-border interfaces of the devices when you deploy a PIM-DM domain.

For PIM-DM to take effect on an interface, you must configure the primary IP address of the interface.

Procedure

1. Enter system view.

   system-view
2. Enable IP multicast routing and enter MRIB view.

   multicast routing [**vpn-instance** *vpn-instance-name*]

   By default, IP multicast routing is disabled.

   For more information about this command, see *IP Multicast Command Reference*.

3. Return to system view.
   quit
4. Enter interface view.
   interface *interface-type interface-number*
5. Enable PIM-DM.
   pim dm
   By default, PIM-DM is disabled.

Configuring the state refresh feature

About this task

- **State refresh capability**—Enables the PIM device directly connected to the source to periodically send state refresh messages. It also enables other PIM devices to refresh pruned state timers after receiving the state refresh messages. It prevents the pruned interfaces from resuming multicast forwarding.
- **State refresh interval**—Determines the interval at which a device sends state refresh messages.
- **Wait time before accepting a new state refresh message**—A device might receive duplicate state refresh messages within a short time. To prevent this situation, you can configure the time that the device must wait to accept a new state refresh message. If the device receives a new state refresh message before the timer expires, it discards the message. If the device receives a new state refresh message after the timer expires, it accepts the message, refreshes its own PIM-DM state, and resets the waiting timer.
- **TTL value of state refresh messages**—The TTL value of a state refresh message decrements by 1 whenever it passes a device before it is forwarded to the downstream node. The state refresh message is not forwarded when the TTL value comes down to 0. A state refresh message with a large TTL value might cycle on a small network.

  To effectively control the propagation scope of state refresh messages, configure an appropriate TTL value based on the network size on the device directly connected with the multicast source.

Restrictions and guidelines

Perform this task on all devices in the PIM-DM domain.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the state refresh feature.
   pim state-refresh-capable
   By default, the state refresh feature is enabled.
4. Return to system view.
   quit
5. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
6. Set the state refresh interval.
   state-refresh-interval *interval*
   The default setting is 60 seconds.

Perform this task on the device directly connected to the multicast source.

7. Set the amount of time that the device must wait to accept a new state refresh message.
   state-refresh-rate-limit *time*
   The default setting is 30 seconds.
8. Set the TTL value of state refresh messages.
   state-refresh-ttl *ttl-value*
   The default setting is 255.

Setting the PIM-DM graft retry timer

About this task

Perform this task to adjust the interval at which the device retransmits a graft message if it does not receive a graft-ack message from the upstream device.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the graft retry timer.
   pim timer graft-retry *interval*
   The default setting is 3 seconds.

Configuring PIM-SM

PIM-SM tasks at a glance

To configure PIM-SM, perform the following tasks:

1. Enabling PIM-SM
2. Configuring static RPs
   As a best practice, configure a static RP when only one dynamic RP exists in the network.
3. Configuring dynamic RPs
   - Configuring C-RPs
   - Configuring C-BSRs
   - (Optional.) Configuring the interval for C-BSRs to send BSMs
   - (Optional.) Configuring the timer for C-BSRs to wait BSMs from the BSR
   - (Optional.) Configuring a PIM domain border
   - (Optional.) Disabling BSM semantic fragmentation
   - (Optional.) Disabling the device from forwarding BSMs out of their incoming interfaces

   As a best practice, configure dynamic RPs when multiple PIM devices exist in the network.
   You can configure static RPs, dynamic RPs, or both.
4. (Optional.) Enabling Auto-RP listening
5. (Optional.) Configuring Anycast RP
6. (Optional.) Configuring multicast source registration
7. (Optional.) Configuring the switchover to SPT

8. (Optional.) Configuring the delay time for DR election
9. (Optional.) Configuring common PIM features

Prerequisites for PIM-SM

Before you configure PIM-SM, complete the following tasks:

- Configure a unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable IP multicast routing.

Enabling PIM-SM

About this task

With PIM-SM enabled on interfaces, devices can establish PIM neighbor relationship and process PIM messages from their PIM neighbors.

Restrictions and guidelines

As a best practice, enable PIM-SM on all non-border interfaces of devices when you deploy a PIM-SM domain.

For PIM-SM to take effect on an interface, you must configure the primary IP address of the interface.

Procedure

1. Enter system view.
   `system-view`
2. Enable IP multicast routing and enter MRIB view.
   `multicast routing` [**`vpn-instance`** *`vpn-instance-name`*]
   By default, IP multicast routing is disabled.
   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.
   `quit`
4. Enter interface view.
   `interface` *`interface-type interface-number`*
5. Enable PIM-SM.
   `pim sm`
   By default, PIM-SM is disabled.

Configuring static RPs

About this task

If only one dynamic RP exists on a network, you can configure a static RP to avoid communication interruption caused by a single point of failure. The static RP can also avoid waste of bandwidth because of frequent message exchanges between C-RPs and the BSR.

Restrictions and guidelines

In a PIM-SM domain, you can configure the same static RP for different multicast groups by using the same RP address but different ACLs.

You do not need to enable PIM for an interface to be configured as a static RP.

If you configure multiple static RPs for a multicast group, only the static RP with the highest IP address takes effect.

The static RP configuration must be the same on all devices in the PIM-SM domain.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure a static RP.

   static-rp *rp-address* [{ *ipv4-acl-number* | **name** *ipv4-acl-number* } | **preferred**] *

   When both a static RP and a dynamically elected RP exist on the network, you can specify the **preferred** keyword in the **static-rp** command to give priority to the static RP. The dynamic RP is used only when the static RP fails. If you do not specify the **preferred** keyword, the dynamic RP takes priority and the static RP is used only when the dynamic RP fails.

Configuring C-RPs

About this task

To avoid C-RP spoofing, configure a C-RP policy to filter C-RP advertisement messages by using an ACL that specifies the packet source address range and multicast group addresses.

Restrictions and guidelines

Configure C-RPs on devices that reside in the backbone network.

Because the RP and other devices exchange a large amount of information in the PIM-SM domain, reserve a large bandwidth between C-RPs and other devices.

You must configure the same C-RP policy on all C-BSRs in the PIM-SM domain because every C-BSR might become the BSR.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-RP.

   c-rp *ip-address* [**advertisement-interval** *adv-interval* | **group-policy** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **holdtime** *hold-time* | **priority** *priority*] *

4. (Optional.) Configure a C-RP policy.

   crp-policy *ipv4-acl-number*

   By default, no C-RP policies are configured. All C-RP advertisement messages are regarded as legal.

Configuring C-BSRs

About this task

You must configure C-BSRs when you configure dynamic RP election.

To prevent a legal BSR from being replaced by a malicious host, configure a BSR policy to filter BSR messages by using an ACL that specifies the legal BSR addresses. It is used to prevent the legal BSR from being replaced by a malicious host.

Restrictions and guidelines

Configure C-BSRs on devices that reside in the backbone network.

Because the BSR and other devices exchange a large amount of information in the PIM-SM domain, reserve a large bandwidth between C-BSRs and other devices.

The C-BSR configuration on the devices in the PIM-SM domain must be the same.

For a successful RPF check, configure static multicast routes to ensure that the next hop to a C-BSR is a tunnel interface when C-BSRs connect to other PIM devices through tunnels. For more information about static multicast routes, see "Configuring multicast routing and forwarding."

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-BSR.

   c-bsr *ip-address* [**scope** *group-address* { *mask-length* | *mask* }] [**hash-length** *hash-length* | **priority** *priority*] *

4. (Optional.) Configure a BSR policy.

   bsr-policy *ipv4-acl-number*

   By default, no BSR policies are configured. All bootstrap messages are regarded as legal.

Configuring the interval for C-BSRs to send BSMs

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires (configured by using the **c-bsr holdtime** command). During this period of time, the BSR periodically sends BSMs to all other devices in the PIM domain. The BSMs carry the IP address of the BSR, RP-set information, and RP aging time. Other devices in the PIM domain select an RP for a multicast group based on the RP-set information. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.

system-view

2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Configure the interval for C-BSRs to send BSMs.
   c-bsr interval *interval*
   By default, a C-BSR sends BSMs at 60-second intervals.

Configuring the timer for C-BSRs to wait BSMs from the BSR

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Configure the timer for C-BSRs to wait BSMs from the BSR.
   c-bsr holdtime *hold-time*
   By default, a C-BSR waits BSMs from the BSR for 130 seconds before sending BSMs.

Configuring a PIM domain border

About this task

A PIM domain border determines the transmission boundary of bootstrap messages. Bootstrap messages cannot cross the domain border in either direction. A number of PIM domain border interfaces partition a network into different PIM-SM domains.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure a PIM domain border.

pim bsr-boundary

By default, an interface is not a PIM domain border.

Disabling BSM semantic fragmentation

About this task

BSM semantic fragmentation enables a BSR to split a BSM into multiple BSM fragments (BSMFs) if the BSM exceeds the MTU. In this way, a non-BSR device can update the RP-set information for a group range after receiving all BSMFs for the range. The loss of one BSMF only affects the RP-set information of the group ranges that the fragment contains.

Restrictions and guidelines

If a device does not support BSM semantic fragmentation, it regards a BSMF as a BSM and updates the RP-set information each time it receives a BSMF. It learns only part of the RP-set information, which further affects the RP election. Therefore, if such a device exists in the PIM-SM domain, you must disable BSM semantic fragmentation on all C-BSRs.

Procedure

1. Enter system view.

   system-view
2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]
3. Disable BSM semantic fragmentation.

   undo bsm-fragment enable

   By default, BSM semantic fragmentation is enabled.

Disabling the device from forwarding BSMs out of their incoming interfaces

About this task

By default, the device forwards BSMs out of their incoming interfaces. This default setting ensures that all devices on the subnet can receive BSMs even when they have inconsistent routing information. However, this default setting results in duplicated multicast traffic. If you are sure that all the devices on the subnet have consistent routing information, you can disable the device from forwarding BSMs out of their incoming interfaces.

Procedure

1. Enter system view.

   system-view
2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]
3. Disable the device from forwarding BSMs out of their incoming interfaces.

   undo bsm-reflection enable

   By default, the device forwards BSMs out of their incoming interfaces.

Enabling Auto-RP listening

About this task

This feature enables the device to receive Auto-RP announcement and discovery messages and learn RP information. The destination IP addresses for Auto-RP announcement and discovery messages are 224.0.1.39 and 224.0.1.40, respectively.

Restrictions and guidelines

After this feature is enabled, the device can receive and forward Auto-RP announcement and discovery messages, but it cannot send these messages unsolicitedly.

Procedure

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Enable Auto-RP listening.
   auto-rp enable
   By default, Auto-RP listening is disabled.

Configuring Anycast RP

Restrictions and guidelines

In an Anycast RP set, RP members must have an interface (for example, loopback interface) with the same IP address. The IP address must be configured as both a C-RP and the Anycast RP address.

To prevent the other Anycast RP member devices from discarding the BSM sent by the BSR, make sure the Anycast RP address is different from the BSR address.

As a best practice to ensure network performance, configure a maximum of 16 Anycast RP members for an Anycast RP set.

As a best practice, configure the IP address of a loopback interface as the RP member address.

If you configure IP addresses of multiple interfaces on the same device as RP member addresses, the lowest IP address takes effect. The rest of the interface addresses become backup RP member addresses.

An Anycast RP address must be a host address with subnet mask 255.255.255.255.

You must add the device where the Anycast RP resides as an RP member to the Anycast RP set.

Prerequisites

You must configure a static RP or C-RPs in the PIM-SM domain before you configure the Anycast RP. Use the address of the static RP or the dynamically elected RP as the Anycast RP address.

Procedure

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Configure Anycast RP.
   anycast-rp *anycast-rp-address member-rp-address*

By default, Anycast RP is not configured.

You can repeat this command to add multiple RP member addresses to the Anycast RP set.

Configuring multicast source registration

About this task

- **PIM register policy**—A PIM register policy enables an RP to filter register messages by using an ACL that specifies the multicast sources and groups. The policy limits the multicast groups to which the RP is designated. If a register message is denied by the ACL or does not match the ACL, the RP discards the register message and sends a register-stop message to the source-side DR. The registration process stops.
- **Checksum computing method for register messages**—For information integrity of a register message, you can configure the device to calculate the checksum based on the entire register message. If a device cannot calculate the checksum based on the entire register message, you can configure the device to calculate the checksum based on the register message header.
- **Register suppression time—**The source-side DR stops sending register messages encapsulated with multicast data and starts a register-stop timer upon receiving a register-stop message from the RP. Before the register-stop timer expires, the DR sends a null register message (a register message without encapsulated multicast data) to the RP and starts a register probe timer. If the DR receives a register-stop message before the register probe timer expires, it resets its register-stop timer. Otherwise, the DR starts sending register messages with encapsulated data again.

  The register-stop timer is set to a random value chosen uniformly from (0.5 × register_suppression_time minus register_probe_time) to (1.5 × register_suppression_time minus register_probe_time). The register_probe_time is fixed to 5 seconds. You can adjust the register suppression time.

Restrictions and guidelines

On all C-RP devices, configure a PIM register policy and the checksum computing method for register messages.

On all devices that might become the source-side DR, set the register suppression time.

As a best practice, set the suppression time to be longer than or equal to 11 seconds.

Procedure

1. Enter system view.

   system-view
2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]
3. Configure a PIM register policy.

   register-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* }

   By default, no PIM register policy is configured. All PIM register messages are regarded as legal.
4. Configure the device to calculate the checksum based on the entire register message.

   register-whole-checksum

   By default, the device calculates the checksum based on the header of a register message.
5. Set the register suppression time.

   register-suppression-timeout *interval*

   The default setting is 60 seconds.

Configuring the switchover to SPT

Restrictions and guidelines

Some devices cannot encapsulate multicast data in register messages destined to the RP. As a best practice to avoid multicast data forwarding failures, do not disable the switchover to SPT on C-RPs that might become the RP.

Procedure

1. Enter system view.

   system-view
2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]
3. Configure RPT to SPT switchover.

   spt-switch-threshold { **immediacy** | **infinity** } [**group-policy** *ipv4-acl-number*]

   By default, the first multicast data packet triggers the RPT to SPT switchover.

Configuring the delay time for DR election

About this task

After a failed DR recovers, it will send PIM hello messages and will be elected as the new DR immediately. In this case, transient multicast traffic loss will occur because the new DR has not generated a complete multicast forwarding table.

Perform this task to delay a recovered DR from being elected as the new DR. Before the delay timer expires, the current DR continues to forward multicast traffic, minimizing multicast traffic loss.

Restrictions and guidelines

As a best practice, perform this task on the user-side interface (interface enabled with IGMP) only on the DR. If you perform this task on both the DR and a non-DR device and both of them fail, traffic loss will occur before the delay timer expires.

As a best practice to avoid sending repeated packets after the failed DR recovers, set the delay time to be longer than the interval for sending PIM hello messages.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the delay time for DR election.

   pim timer dr-elect-delay *delay-time*

   By default, a failed DR is elected as the new DR immediately after it recovers.

Configuring BIDIR-PIM

BIDIR-PIM tasks at a glance

To configure BIDIR-PIM, perform the following tasks:

1. Enabling BIDIR-PIM

2. Configuring static RPs

   As a best practice, configure a static RP when only one dynamic RP exists in the network.
3. Configuring dynamic RPs
   - Configuring C-RPs
   - Configuring C-BSRs
   - (Optional.) Configuring the interval for C-BSRs to send BSMs
   - (Optional.) Configuring the timer for C-BSRs to wait BSMs from the BSR
   - (Optional.) Configuring a PIM domain border
   - (Optional.) Disabling BSM semantic fragmentation
   - (Optional.) Disabling the device from forwarding BSMs out of their incoming interfaces

   As a best practice, configure dynamic RPs when multiple PIM devices exist in the network.

   You can configure static RPs, dynamic RPs, or both.
4. (Optional.) Setting the maximum number of BIDIR-PIM RPs
5. (Optional.) Enabling Auto-RP listening
6. (Optional.) Configuring common PIM features

Prerequisites for BIDIR-PIM

Before you configure BIDIR-PIM, complete the following tasks:

- Configure a unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable PIM-SM.

Enabling BIDIR-PIM

Restrictions and guidelines

As a best practice, enable PIM-SM on all non-border interfaces of devices when you deploy a BIDIR-PIM domain.

Procedure

1. Enter system view.

   `system-view`
2. Enable IP multicast routing and enter MRIB view.

   `multicast routing` [**`vpn-instance`** *`vpn-instance-name`*]

   By default, IP multicast routing is disabled.

   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.

   `quit`
4. Enter interface view.

   `interface` *`interface-type interface-number`*
5. Enable PIM-SM.

   `pim sm`

   By default, PIM-SM is disabled.
6. Return to system view.

   `quit`

7. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

8. Enable BIDIR-PIM.

   bidir-pim enable

   By default, BIDIR-PIM is disabled.

Configuring static RPs

About this task

If only one dynamic RP exists on a network, you can configure a static RP to avoid communication interruption caused by a single point of failure. The static RP can also avoid bandwidth waste caused by frequent message exchanges between C-RPs and the BSR.

Restrictions and guidelines

The static RP configuration must be the same on all devices in the BIDIR-PIM domain.

In a BIDIR-PIM domain, you can configure the same static RP for different multicast groups by using the same RP address but different ACLs.

You must enable PIM for an interface to be configured as a static RP.

If you configure multiple static RPs for a multicast group, only the static RP with the highest IP address takes effect.

You can specify an unused IP address for a static RP. This address must be on the same subnet with the link on which the static RP is configured. For example, if the IP addresses of the interfaces at the two ends of a link are 10.1.1.1/24 and 10.1.1.2/24, you can specify the interface with IP address 10.1.1.100/24 as a static RP. As a result, the link becomes an RPL.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure a static RP for BIDIR-PIM.

   static-rp *rp-address* **bidir** [{ *ipv4-acl-number* | **name** *ipv4-acl-number* } | **preferred**] *

   By default, no static RPs exist.

   When both a static RP and a dynamically elected RP exist on the network, you can specify the **preferred** keyword in the **static-rp** command to give priority to the static RP. The dynamic RP is used only when the static RP fails. If you do not specify the **preferred** keyword, the dynamic RP takes priority and the static RP is used only when the dynamic RP fails.

Configuring C-RPs

About this task

To guard against C-RP spoofing, configure a C-RP policy to filter C-RP advertisement messages by using an ACL that specifies the packet source address range and multicast groups.

Restrictions and guidelines

Configure C-RPs on devices that reside in the backbone network.

Because the RP and other devices exchange a large amount of information in the BIDIR-PIM domain, reserve a large bandwidth between C-RPs and other devices.

You must configure the same C-RP policy on all C-BSRs in the BIDIR-PIM domain because every C-BSR might become the BSR.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-RP to provide services for BIDIR-PIM.

   c-rp *ip-address* [**advertisement-interval** *adv-interval* | **group-policy** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **holdtime** *hold-time* | **priority** *priority*] * **bidir**

   By default, no C-RPs exist.

Configuring C-BSRs

About this task

You must configure C-BSRs when you configure dynamic RP election.

To prevent a legal BSR from being replaced by a malicious host, configure a BSR policy to filter BSR messages by using an ACL that specifies the legal BSR addresses.

Restrictions and guidelines

Configure C-BSRs on devices that reside in the backbone device.

Because the BSR and other devices exchange a large amount of information in the BIDIR-PIM domain, reserve a large bandwidth between the C-BSRs and other devices.

The C-BSR configuration on the devices in the BIDIR-PIM domain must be the same.

For a successful RPF check, configure static multicast routes to ensure that the next hop to a C-BSR is a tunnel interface when C-BSRs connect to other PIM devices through tunnels. For more information about static multicast routes, see "Configuring multicast routing and forwarding."

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-BSR.

   c-bsr *ip-address* [**scope** *group-address* { *mask-length* | *mask* }] [**hash-length** *hash-length* | **priority** *priority*] *

4. (Optional.) Configure a BSR policy.

   bsr-policy *ipv4-acl-number*

   By default, no BSR policy is configured. All bootstrap messages are regarded as legal.

Configuring the interval for C-BSRs to send BSMs

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires (configured by using the **c-bsr holdtime** command). During this period of time, the BSR periodically sends BSMs to all other devices in the PIM domain. The BSMs carry the IP address of the BSR, RP-set information, and RP aging time. Other devices in the PIM domain select an RP for a multicast group based on the RP-set information. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

In a BIDIR-PIM network with active and backup RPs, when the link for the active RP fails, the backup RP must wait the active RP to age out before taking over. The aging time of the active RP is long because it must be longer than the sending interval of bootstrap messages, and the sending interval of BSMs is not configurable. You can use this command to set a smaller sending interval.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.

   system-view
2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]
3. Configure the interval for C-BSRs to send BSMs.

   c-bsr interval *interval*

   By default, a C-BSR sends BSMs at 60-second intervals.

Configuring the timer for C-BSRs to wait BSMs from the BSR

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Configure the timer for C-BSRs to wait BSMs from the BSR.
   c-bsr holdtime *hold-time*
   By default, a C-BSR waits for BSMs from the BSR for 130 seconds before sending BSMs.

Configuring a PIM domain border

About this task

A PIM domain border determines the transmission boundary of bootstrap messages. Bootstrap messages cannot cross the domain border in either direction. A number of PIM domain border interfaces partition a network into different BIDIR-PIM domains.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure a PIM domain border.
   pim bsr-boundary
   By default, an interface is not a PIM domain border.

Disabling BSM semantic fragmentation

About this task

BSM semantic fragmentation enables a BSR to split a BSM into multiple BSM fragments (BSMFs) if the BSM exceeds the MTU. In this way, a non-BSR device can update the RP-set information for a group range after receiving all BSMFs for the group range. The loss of one BSMF only affects the RP-set information of the group ranges that the fragment contains.

Restrictions and guidelines

If a device does not support BSM semantic fragmentation, it regards a BSMF as a BSM and updates the RP-set information each time it receives a BSMF. It learns only part of the RP-set information, which further affects the RP election. Therefore, if such a device presents in the BIDIR-PIM domain, you must disable BSM semantic fragmentation on all C-BSRs.

Procedure

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]

3. Disable BSM semantic fragmentation.

   undo bsm-fragment enable

   By default, BSM semantic fragmentation is enabled.

Disabling the device from forwarding BSMs out of their incoming interfaces

About this task

By default, the device forwards BSMs out of their incoming interfaces. This default setting ensures that all devices on the subnet can receive BSMs even when they have inconsistent routing information. However, this default setting results in duplicated multicast traffic. If you are sure that all the devices on the subnet have consistent routing information, you can disable the device from forwarding BSMs out of their incoming interfaces.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Disable the device from forwarding BSMs out of their incoming interfaces.

   undo bsm-reflection enable

   By default, the device forwards BSMs out of their incoming interfaces.

Setting the maximum number of BIDIR-PIM RPs

About this task

In a BIDIR-PIM domain, one DF election per RP is implemented on all PIM-enabled interfaces. To avoid unnecessary DF elections, do not configure multiple RPs for BIDIR-PIM.

This configuration sets a limit on the number of BIDIR-PIM RPs. If the number of RPs exceeds the limit, excess RPs do not take effect and can be used only for DF election rather than multicast data forwarding. The system does not delete the excess RPs. They must be deleted manually.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Set the maximum number of BIDIR-PIM RPs.

   bidir-rp-limit *limit*

   By default, the maximum number of BIDIR-PIM RPs is 6.

Enabling Auto-RP listening

About this task

This feature enables the device to receive Auto-RP announcement and discovery messages and learn RP information. The destination IP addresses for Auto-RP announcement and discovery messages are 224.0.1.39 and 224.0.1.40, respectively.

Restrictions and guidelines

After this feature is enabled, the device can receive and forward Auto-RP announcement and discovery messages, but it cannot send these messages unsolicitedly.

Procedure

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Enable Auto-RP listening.
   auto-rp enable
   By default, Auto-RP listening is disabled.

Configuring PIM-SSM

PIM-SSM tasks at a glance

To configure PIM-SSM, perform the following tasks:

1. Enabling PIM-SM
2. (Optional.) Configuring the SSM group range
3. (Optional.) Configuring the delay time for DR election
4. (Optional.) Configuring common PIM features

Prerequisites for PIM-SSM

Before you configure PIM-SSM, complete the following tasks:

- Configure a unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable PIM-SM.
- Enable IGMPv3 on PIM devices that connect to multicast receivers.

Enabling PIM-SM

Restrictions and guidelines

As a best practice, enable PIM-SM on all non-border interfaces of devices when you deploy a PIM-SSM domain.

Procedure

1. Enter system view.
   system-view
2. Enable multicast routing and enter MRIB view.
   multicast routing [**vpn-instance** *vpn-instance-name*]
   By default, multicast routing is disabled.
   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.
   quit

4. Enter interface view.

   interface *interface-type interface-number*

5. Enable PIM-SM.

   pim sm

   By default, PIM-SM is disabled.

Configuring the SSM group range

About this task

When a PIM-SM enabled interface receives a multicast packet, it checks whether the multicast group address of the packet is in the SSM group range. If the multicast group address is in this range, the PIM mode for this packet is PIM-SSM. If the multicast group address is not in this range, the PIM mode is PIM-SM.

Restrictions and guidelines

Configure the same SSM group range on all devices in the entire PIM-SSM domain. Otherwise, multicast information cannot be delivered through the SSM model.

When a member of a multicast group in the SSM group range sends an IGMPv1 or IGMPv2 report message, the device does not trigger a (*, G) join.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure the SSM group range.

   ssm-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* }

   By default, the SSM group range is 232.0.0.0/8.

Configuring the delay time for DR election

About this task

After a failed DR recovers, it will send PIM hello messages and will be elected as the new DR immediately. In this case, transient multicast traffic loss will occur because the new DR has not generated a complete multicast forwarding table.

Perform this task to delay a recovered DR from being elected as the new DR. Before the delay timer expires, the current DR continues to forward multicast traffic, minimizing multicast traffic loss.

Restrictions and guidelines

As a best practice, perform this task on the user-side interface (interface enabled with IGMP) only on the DR. If you perform this task on both the DR and a non-DR device and both of them fail, traffic loss will occur before the delay timer expires.

As a best practice to avoid sending repeated packets after the failed DR recovers, set the delay time to be longer than the interval for sending PIM hello messages.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

interface *interface-type interface-number*

3. Configure the delay time for DR election.

   pim timer dr-elect-delay *delay-time*

   By default, a failed DR is elected as the new DR immediately after it recovers.

Configuring common PIM features

Common PIM feature tasks at a glance

All the following tasks are optional.

- Configuring a multicast source policy
- Configuring a PIM hello policy
- Configuring PIM hello message options
- Dropping hello messages without the Generation ID option
- Enabling the device to send hello messages with a different Generation ID upon a port state change
- Configuring common PIM timers
- Setting the maximum size of a join or prune message
- Configuring a PIM join policy
- Setting the DSCP value for outgoing PIM protocol packets
- Enabling BFD for PIM
- Enabling PIM passive mode
- Enabling PIM NSR
- Enabling SNMP notifications for PIM
- Enabling PIM logging
- Enabling log message suppression for PIM

Configuring a multicast source policy

About this task

This feature enables the device to filter multicast data by using an ACL that specifies the multicast sources and the optional groups. It filters not only data packets but also register messages with multicast data encapsulated. It controls the information available to downstream receivers.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Configure a multicast source policy.

   source-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* }

   By default, no multicast source policy is configured. The device does not filter multicast data packets.

Configuring a PIM hello policy

About this task

This feature enables the device to filter PIM hello messages by using an ACL that specifies the packet source addresses. It is used to guard against PIM message attacks and to establish correct PIM neighboring relationships.

If hello messages of an existing PIM neighbor are filtered out by the policy, the neighbor is automatically removed when its aging timer expires.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure a PIM hello policy.

   pim neighbor-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* }

   By default, no PIM hello policy is configured on an interface. All PIM hello messages are regarded as legal.

Configuring PIM hello message options

About this task

In either a PIM-DM domain or a PIM-SM domain, hello messages exchanged among devices contain the following configurable options:

- **DR_Priority** (for PIM-SM only)—Priority for DR election. The device with the highest priority wins the DR election. You can configure this option for all the devices in a shared-media LAN that directly connects to the multicast source or the receivers.
- **Holdtime**—PIM neighbor lifetime. If a device does not receive a hello message from a neighbor when the neighbor lifetime expires, it regards the neighbor failed or unreachable.
- **LAN_Prune_Delay**—Delay of pruning a downstream interface on a shared-media LAN. This option has the LAN delay field, the override interval field, and the T bit.

  The LAN delay defines the PIM message propagation delay. The override interval defines a period for a device to override a prune message. The T bit specifies the capability of tracking downstream device status on upstream devices.

  On the shared-media LAN, the propagation delay and override interval are used as follows:

  - If a device receives a prune message on its upstream interface, it means that there are downstream devices on the shared-media LAN. If this device still needs to receive multicast data, it must send a join message to override the prune message within the override interval.
  - When a device receives a prune message from its downstream interface, it does not immediately prune this interface. Instead, it starts a timer (the propagation delay plus the override interval). If interface receives a join message before the timer expires, the device does not prune the interface. Otherwise, the device prunes the interface.
- **Neighbor tracking**—If you enable neighbor tracking on an upstream device, this device can track the states of the downstream nodes for which the joined state holdtime timer has not expired. All join messages from downstream devices are accepted.

Restrictions and guidelines

If the propagation delay or override interval on different PIM devices on a shared-media LAN are different, the largest ones apply.

If you want to enable neighbor tracking, you must enable it on all PIM devices on a shared-media LAN. Otherwise, the upstream device cannot track join messages from every downstream devices.

You can configure hello message options for all interfaces in PIM view or for the current interface in interface view. The configuration made in interface view takes priority over the configuration made in PIM view.

Configuring hello message options globally

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Set the DR priority.
   hello-option dr-priority *priority*
   The default setting is 1.
4. Set the neighbor lifetime.
   hello-option holdtime *time*
   The default setting is 105 seconds.
5. Set the PIM message propagation delay for a shared-media LAN.
   hello-option lan-delay *delay*
   The default setting is 500 milliseconds.
6. Set the override interval.
   hello-option override-interval *interval*
   The default setting is 2500 milliseconds.
7. Enable neighbor tracking.
   hello-option neighbor-tracking
   By default, neighbor tracking is disabled.

Configuring hello message options on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the DR priority.
   pim hello-option dr-priority *priority*
   The default setting is 1.
4. Set the neighbor lifetime.
   pim hello-option holdtime *time*
   The default setting is 105 seconds.
5. Set the PIM message propagation delay.
   pim hello-option lan-delay *delay*
   The default setting is 500 milliseconds.
6. Set the override interval.
   pim hello-option override-interval *interval*
   The default setting is 2500 milliseconds.
7. Enable neighbor tracking.

pim hello-option neighbor-tracking

By default, neighbor tracking is disabled.

Dropping hello messages without the Generation ID option

About this task

A device generates a generation ID for hello messages when an interface is enabled with PIM. The generation ID is a random value, but it changes only when the status of the device changes. If a PIM device finds that the generation ID in a hello message from the upstream device has changed, it assumes that the status of the upstream device has changed. In this case, it sends a join message to the upstream device for status update. You can configure an interface to drop hello messages without the generation ID options to promptly know the status of an upstream device.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable the interface to drop hello messages without the Generation ID option.

   pim require-genid

   By default, an interface accepts hello messages without the Generation ID option.

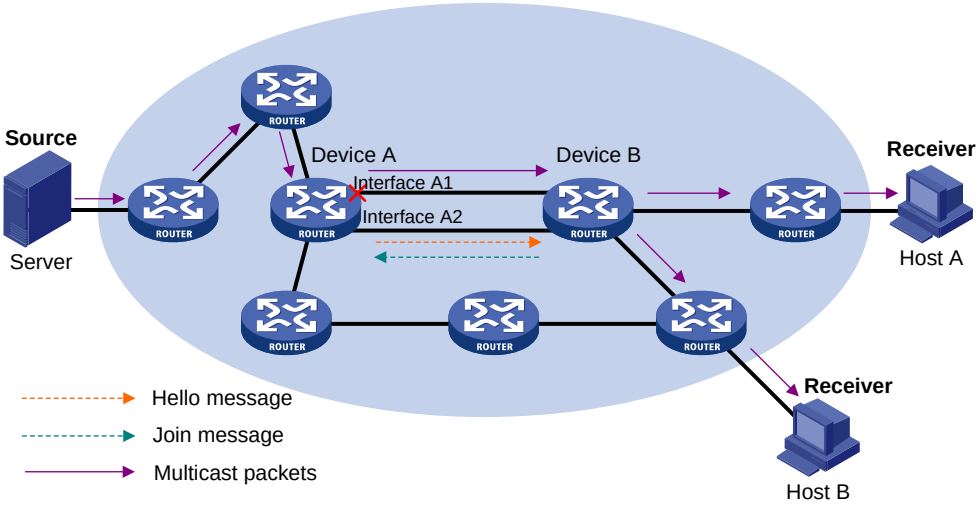
Enabling the device to send hello messages with a different Generation ID upon a port state change

About this task

As shown in Figure 15, Device B connects to Device A through two interfaces in the same VLAN (interface A1 and interface A2) for high availability. Device A sends multicast data to Device B through interface A. When interface A1 fails or recovers from failure, Device A sends multicast data to Device B through interface A2. When interface A1 recovers from failure, Device A sends multicast data to Device B through interface A1 again. In both cases, Device A does not immediately send hello messages with a different Generation ID and multicast data cannot be switched to the new path immediately.

After you perform this task, Device A will immediately send hello messages with a different Generation ID when the port state in the VLAN changes. Then, Device B will send a PIM join message and update the multicast forwarding entry. Multicast data can be switched to the new path immediately.

Figure 15 Network diagram



Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enter VLAN interface view.
   ```
   interface vlan-interface interface-number
   ```
3. Enable the interface to send hello messages with a different Generation ID upon a port state change.
   ```
   pim triggered-hello enable { port-down | port-up } *
   ```
   By default, the device does not send hello messages with a different Generation ID upon a port state change.

Configuring common PIM timers

About this task

The following are common PIM timers:

- **Hello interval**—Interval at which a PIM device sends hello messages to discover PIM neighbors, and maintain PIM neighbor relationship.
- **Triggered hello delay**—Maximum delay for sending a hello message to avoid collisions caused by simultaneous hello messages. After receiving a hello message, a PIM device waits for a random time before sending a hello message. This random time is in the range of 0 to the triggered hello delay.
- **Join/Prune interval**—Interval at which a PIM device sends join or prune messages to its upstream devices for state update.
- **Joined/Pruned state holdtime**—Time for which a PIM device keeps the joined/pruned state for the downstream interfaces. This joined/pruned state holdtime is specified in a join or prune message.
- **Multicast source lifetime**—Lifetime that a PIM device maintains for a multicast source. If a device does not receive subsequent multicast data from the multicast source S when the timer expires, it deletes the (S, G) entry for the multicast source.

Restrictions and guidelines

To prevent upstream neighbors from aging out, set the join or prune interval to be less than the join or pruned state holdtime.

You can configure common PIM timers for all interfaces in PIM view or for the current interface in interface view. The configuration made in interface view takes priority over the configuration made in PIM view.

As a best practice, use the defaults for a network without special requirements.

Configuring common PIM timers globally

1. Enter system view.
   system-view
2. Enter PIM view.
   pim [**vpn-instance** *vpn-instance-name*]
3. Set the hello interval.
   timer hello *interval*
   The default setting is 30 seconds.
4. Set the join or prune interval.
   timer join-prune *interval*
   The default setting is 60 seconds.
   This configuration takes effect after the current interval ends.
5. Set the joined/pruned state holdtime.
   holdtime join-prune *time*
   The default setting is 210 seconds.
6. Set the multicast source lifetime.
   source-lifetime *time*
   The default setting is 210 seconds.

Configuring common PIM timers on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the hello interval.
   pim timer hello *interval*
   The default setting is 30 seconds.
4. Set the triggered hello delay.
   pim triggered-hello-delay *delay*
   The default setting is 5 seconds.
5. Set the join or prune interval.
   pim timer join-prune *interval*
   The default setting is 60 seconds.
   This configuration takes effect after the current interval ends.
6. Set the joined/pruned state holdtime.
   pim holdtime join-prune *time*
   The default setting is 210 seconds.

Setting the maximum size of a join or prune message

About this task

The loss of an oversized join or prune message might result in loss of massive information. You can set a small value for the size of a join or prune message to reduce the impact.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Set the maximum size of a join or prune message.

   jp-pkt-size *size*

   By default, the maximum size of a join or prune message is 1200 bytes.

Configuring a PIM join policy

About this task

The PIM join policy filters joined multicast sources and groups in PIM join or prune messages to protect the device against invalid PIM join or prune message attacks. The policy identifies valid joined multicast source and groups in PIM join or prune messages and filters out invalid joined multicast source and groups. The device creates (*, G) or (S, G) entries only for valid joined multicast sources and groups.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure a PIM join policy to filter joined multicast sources and groups in PIM join or prune messages.

   pim join-policy { { *ipv4-acl-number* | **name** *ipv4-acl-name* } | [**asm** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **ssm** { *ipv4-acl-number* | **name** *ipv4-acl-number* }] * }

   By default, no PIM join policy is configured and joined multicast sources and groups in PIM join or prune messages are not filtered.

Setting the DSCP value for outgoing PIM protocol packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.

   system-view

2. Enter PIM view.

   pim [**vpn-instance** *vpn-instance-name*]

3. Set the DSCP value for outgoing PIM protocol packets.

   dscp *dscp-value*

   By d efault, the DSCP value is 48 for outgoing PIM protocol packets.

Enabling BFD for PIM

About this task

If a DR on a shared-media network fails, a new DR election process does not start until the DR ages out. In addition, it might take a long period of time before other devices detect the link failures and trigger a new DR election. To start a new DR election process immediately after the original DR fails, enable BFD for PIM to detect link failures among PIM neighbors.

You must enable BFD for PIM on all PIM devices on a shared-media network. For more information about BFD, see *High Availability Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable BFD for PIM.

   pim bfd enable

   By default, BFD is disabled for PIM.

Enabling PIM passive mode

About this task

To guard against PIM hello spoofing, you can enable PIM passive mode on a receiver-side interface. The PIM passive interface cannot receive or forward PIM protocol messages (excluding register, register-stop, and C-RP-Adv messages), and it acts as the DR on the subnet.

Restrictions and guidelines

To avoid duplicate multicast data transmission and flow loop, do not enable this feature on a shared-media LAN with multiple PIM devices.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable PIM passive mode on the interface.

   pim passive

   By default, PIM passive mode is disabled on an interface.

Enabling PIM NSR

About this task

This feature enables PIM to back up protocol state information and data, including PIM neighbor information and routes, from the active process to the standby process. The standby process

immediately takes over when the active process fails. Use this feature to avoid route flapping and forwarding interruption for PIM when an active/standby switchover occurs.

Procedure

1. Enter system view.

   system-view

2. Enable PIM NSR.

   pim non-stop-routing

   By default, PIM NSR is disabled.

3. (Optional.) Specify a reserved VLAN for the peer link in an M-LAG system.

   m-lag peer-link reserved

   By default, no reserved VLAN is specified for the peer link in an M-LAG system.

   Execute this command in the view of the VLAN to be specified as the reserved VLAN. For more information about this command, see *Layer 2—LAN Switching Command Reference*.

Enabling SNMP notifications for PIM

About this task

To report critical PIM events to an NMS, enable SNMP notifications for PIM. For PIM event notifications to be sent correctly, you must also configure SNMP as described in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for PIM.

   snmp-agent trap enable pim [**candidate-bsr-win-election** | **elected-bsr-lost-election** | **interface-election** | **invalid-join-prune** | **invalid-register** | **neighbor-loss** | **private-interface-election** | **private-neighbor-loss** | **private-new-neighbor** | **private-rp-mapping-change** | **rp-mapping-change**] *

   By default, SNMP notifications for PIM are enabled.

Enabling PIM logging

About this task

To report critical PIM events to the information center, enable PIM logging. For PIM log messages to be sent correctly, you must also configure the information center on the device. For more information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Outputting a large amount of log messages degrades the device performance. To resolve this issue, disable PIM logging or use the **pim log suppress** command to enable log message suppression for PIM.

Procedure

1. Enter system view.

   system-view

2. Enable PIM logging.

pim log enable [**dr** | **error-packet** | **neighbor** | **rp**] *

PIM logging is enabled by default.

Enabling log message suppression for PIM

About this task

Outputting a large amount of log messages degrades the device performance. To resolve this issue, perform this task to enable log message suppression for PIM. With this feature configured, the speed of sending PIM log messages to the information center will be decreased to reduce the load on the device.

Restrictions and guidelines

Before you configure this feature, make sure PIM logging has been enabled using the **pim log enable** command.

Procedure

1. Enter system view.
   system-view
2. Enable log message suppression for PIM.
   pim log suppress [**dr** | **error-packet** | **neighbor** | **rp**] *
   By default, log message suppression is enabled for PIM.

Display and maintenance commands for PIM

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display register-tunnel interface information.	**display interface** [**register-tunnel** [*interface-number*]] [**brief** [**description** | **down**]]
Display BSR information in the PIM-SM domain.	**display pim** [**vpn-instance** *vpn-instance-name*] **bsr-info**
Display information about the routes used by PIM.	**display pim** [**vpn-instance** *vpn-instance-name*] **claimed-route** [*source-address*]
Display C-RP information in the PIM-SM domain.	**display pim** [**vpn-instance** *vpn-instance-name*] **c-rp** [**local**]
Display DF information in the BIDIR-PIM domain.	**display pim** [**vpn-instance** *vpn-instance-name*] **df-info** [*rp-address*]
Display PIM information on an interface.	**display pim** [**vpn-instance** *vpn-instance-name*] **interface** [*interface-type interface-number*] [**verbose**]
Display PIM neighbor information.	**display pim** [**vpn-instance** *vpn-instance-name*] **neighbor** [*neighbor-address* | **interface** *interface-type interface-number* | **verbose**] *
Display PIM routing entries.	**display pim** [**vpn-instance** *vpn-instance-name*] **routing-table** [*group-address* [**mask** { *mask-length* | *mask* }] | *source-address* [**mask**

Task	Command
	{ *mask-length* | *mask* }] | **flags** *flag-value* | **fsm** | **incoming-interface** *interface-type interface-number* | **mode** *mode-type* | **outgoing-interface** { **exclude** | **include** | **match** } *interface-type interface-number* | **proxy** | **extranet** { **source-vpn-instance** *source-vpn-instance-name* | **source-public-instance** | **receive-vpn-instance** *receive-vpn-instance-name* | **receive-public-instance** }] *
Display RP information in the PIM-SM domain.	**display pim** [**vpn-instance** *vpn-instance-name*] **rp-info** [*group-address*]
Display statistics for PIM packets.	**display pim statistics**

PIM configuration examples

Example: Configuring PIM-DM

Network configuration

As shown in Figure 16:

- OSPF runs on the network.
- VOD streams are sent to receiver hosts in multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist on each stub network.
- The entire PIM domain operates in the dense mode.
- Host A and Host C are multicast receivers on two stub networks.
- IGMPv2 runs between Switch A and N1, and between Switch B, Switch C, and N2.

Figure 16 Network diagram

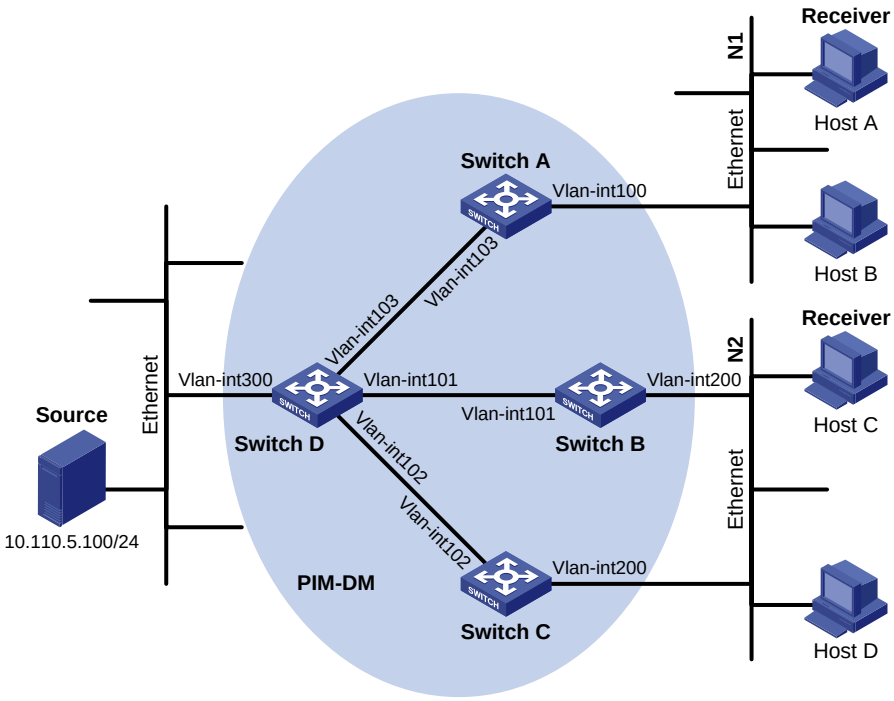


Table 1 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch C	Vlan-int102	192.168.3.1/24
Switch A	Vlan-int103	192.168.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
Switch B	Vlan-int200	10.110.2.1/24	Switch D	Vlan-int103	192.168.1.2/24
Switch B	Vlan-int101	192.168.2.1/24	Switch D	Vlan-int101	192.168.2.2/24
Switch C	Vlan-int200	10.110.2.2/24	Switch D	Vlan-int102	192.168.3.2/24

Prerequisites

Assign an IP address and subnet mask for each interface as shown in Figure 16, and configure OSPF on the switches in the PIM-DM domain.

Procedure

1. Enable IP multicast routing, IGMP, and PIM-DM:

On Switch A, enable IP multicast routing.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Enable IGMP on VLAN-interface 100 (the interface that connects to the stub network).

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] quit
```

Enable PIM-DM on VLAN-interface 103.

```
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim dm
```

```
[SwitchA-Vlan-interface103] quit
```

Enable IP multicast routing, IGMP, and PIM-DM on Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

On Switch D, enable IP multicast routing, and enable PIM-DM on each interface.

```
<SwitchD> system-view
[SwitchD] multicast routing
[SwitchD-mrib] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] pim dm
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim dm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 101
[SwitchD-Vlan-interface101] pim dm
[SwitchD-Vlan-interface101] quit
[SwitchD] interface vlan-interface 102
[SwitchD-Vlan-interface102] pim dm
[SwitchD-Vlan-interface102] quit
```

Verifying the configuration

Display PIM information on Switch D.

```
[SwitchD] display pim interface
 Interface: Vlan300
   NbrCnt: 0
   HelloInt: 30
   DR priority: 1
   DR address: 10.110.5.1 (local)
 Interface: Vlan103
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: 192.168.1.2 (local)
 Interface: Vlan101
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: 192.168.2.2 (local)
 Interface: Vlan102
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: 192.168.3.2 (local)
```

Display PIM neighboring relationships on Switch D.

```
[SwitchD] display pim neighbor
 Total pim neighbors: 3
  Neighbor: 192.168.1.1
    Interface: Vlan103
    Uptime: 00:02:22
```

```
    Expiry time: 00:01:27
    DR priority: 1
    Mode: N/A
  Neighbor: 192.168.2.1
    Interface: Vlan101
    Uptime: 00:00:22
    Expiry time: 00:01:29
    DR priority: 3
    Mode: N/A
  Neighbor: 192.168.2.1
    Interface: Vlan102
    Uptime: 00:00:23
    Expiry time: 00:01:31
    DR priority: 5
    Mode: N/A
```

Send an IGMP report from Host A to join multicast group 225.1.1.1. (Details not shown.)

Send multicast data from multicast source 10.110.5.100/24 to multicast group 225.1.1.1. (Details not shown.)

Display PIM routing entries on Switch A.

```
[SwitchA] display pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
     Protocol: pim-dm, Flag: WC
     UpTime: 00:04:25
     Upstream interface: NULL
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstream interfaces: 1
         1: Vlan-interface100
             Protocol: igmp, UpTime: 00:04:25, Expires: -

 (10.110.5.100, 225.1.1.1)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:06:14
     Upstream interface: Vlan-interface103
         Upstream neighbor: 192.168.1.2
         RPF prime neighbor: 192.168.1.2
     Downstream interface(s) information:
     Total number of downstream interfaces: 1
         1: Vlan-interface100
             Protocol: pim-dm, UpTime: 00:04:25, Expires: -
```

Display PIM routing entries on Switch D.

```
[SwitchD] display pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 225.1.1.1)
```

```
Protocol: pim-dm, Flag: LOC ACT
UpTime: 00:03:27
Upstream interface: Vlan-interface300
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
Downstream interface(s) information:
Total number of downstream interfaces: 2
    1: Vlan-interface103
        Protocol: pim-dm, UpTime: 00:03:27, Expires: -
    2: Vlan-interface102
        Protocol: pim-dm, UpTime: 00:03:27, Expires: -
```

The output shows the following information:

- Switches on the SPT path (Switch A and Switch D) have the correct (S, G) entries.
- Switch A has the correct (*, G) entry.

Example: Configuring non-scoped PIM-SM

Network configuration

As shown in Figure 17:

- OSPF runs on the network.
- VOD streams are sent to receiver hosts in multicast. The receivers of different subnets form stub networks, and a minimum of one receiver host exist on each stub network.
- The entire PIM-SM domain contains only one BSR.
- Host A and Host C are multicast receivers on two stub networks N1 and N2.
- Specify VLAN-interface 102 on Switch E as a C-BSR and a C-RP. The C-RP is designated to multicast group range 225.1.1.0/24. Specify VLAN-interface 101 of Switch D as the static RP on all the switches to back up the dynamic RP.
- IGMPv2 runs between Switch A and N1, and between Switch B, Switch C, and N2.

Figure 17 Network diagram

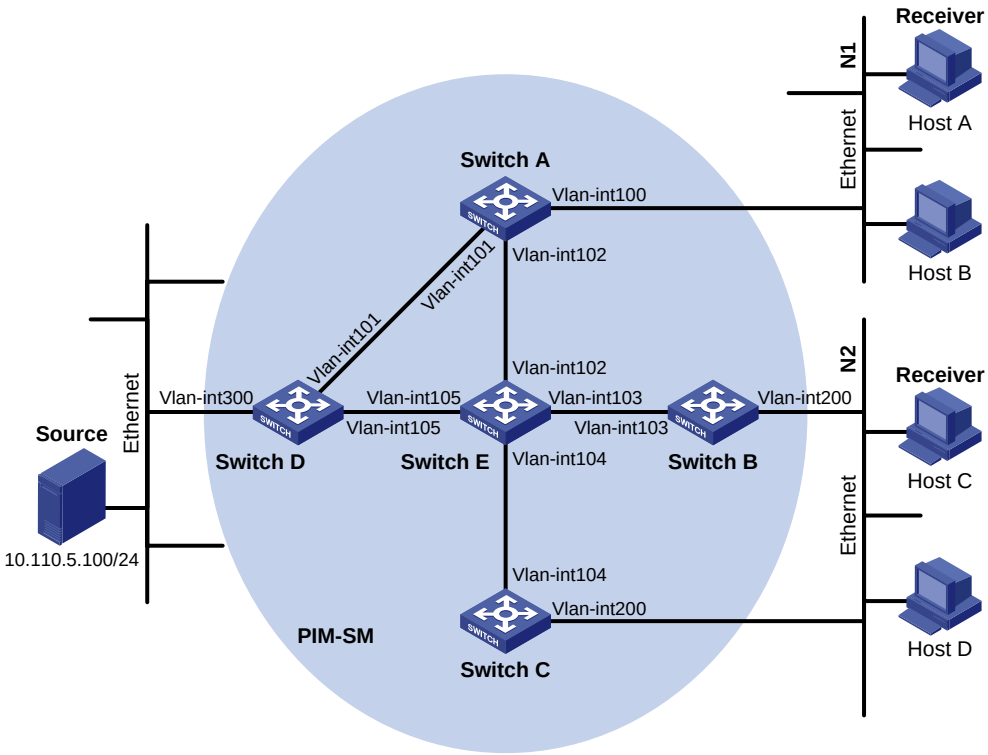


Table 2 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
Switch A	Vlan-int101	192.168.1.1/24	Switch D	Vlan-int101	192.168.1.2/24
Switch A	Vlan-int102	192.168.9.1/24	Switch D	Vlan-int105	192.168.4.2/24
Switch B	Vlan-int200	10.110.2.1/24	Switch E	Vlan-int104	192.168.3.2/24
Switch B	Vlan-int103	192.168.2.1/24	Switch E	Vlan-int103	192.168.2.2/24
Switch C	Vlan-int200	10.110.2.2/24	Switch E	Vlan-int102	192.168.9.2/24
Switch C	Vlan-int104	192.168.3.1/24	Switch E	Vlan-int105	192.168.4.1/24

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 17, and configure OSPF on all switches in the PIM-SM domain.

Procedure

1. Enable IP multicast routing, and enable IGMP and PIM-SM:

   # On Switch A, enable IP multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

   # Enable IGMP on VLAN-interface 100 (the interface that connects to the stub network).

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] igmp enable
   ```

```
[SwitchA-Vlan-interface100] quit
```

Enable PIM-SM on the other interfaces.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim sm
[SwitchA-Vlan-interface102] quit
```

Enable IP multicast routing, IGMP, and PIM-SM on Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

Enable IP multicast routing and PIM-SM on Switch D and Switch E in the same way Switch A is configured. (Details not shown.)

2. Configure C-BSRs, C-RPs, and the static RP:

On Switch E, configure the service scope of RP advertisements.

```
<SwitchE> system-view
[SwitchE] acl basic 2005
[SwitchE-acl-ipv4-basic-2005] rule permit source 225.1.1.0 0.0.0.255
[SwitchE-acl-ipv4-basic-2005] quit
```

Configure VLAN-interface 102 as a C-BSR and a C-RP.

```
[SwitchE] pim
[SwitchE-pim] c-bsr 192.168.9.2
[SwitchE-pim] c-rp 192.168.9.2 group-policy 2005
```

Configure VLAN-interface 101 of Switch D as the static RP.

```
[SwitchE-pim] static-rp 192.168.1.2
[SwitchE-pim] quit
```

On Switch A, configure VLAN-interface 101 of Switch D as the static RP.

```
<SwitchA> system-view
[SwitchA] pim
[SwitchA-pim] static-rp 192.168.1.2
[SwitchA-pim] quit
```

Configure the static RP on Switch B, Switch C, and Switch D in the same way Switch A is configured. (Details not shown.)

Verifying the configuration

Display PIM information on Switch A.

```
[SwitchA] display pim interface
 Interface: Vlan101
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: 192.168.1.2
 Interface: Vlan102
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: 192.168.9.2
```

Display BSR information on Switch A.

```
[SwitchA] display pim bsr-info
```

```
 Scope: non-scoped
     State: Accept Preferred
     Bootstrap timer: 00:01:44
     Elected BSR address: 192.168.9.2
       Priority: 64
       Hash mask length: 30
       Uptime: 00:11:18
```

Display BSR information on Switch E.

```
[SwitchE] display pim bsr-info
 Scope: non-scoped
     State: Elected
     Bootstrap timer: 00:01:44
     Elected BSR address: 192.168.9.2
       Priority: 64
       Hash mask length: 30
       Uptime: 00:11:18
     Candidate BSR address: 192.168.9.2
       Priority: 64
       Hash mask length: 30
```

Display RP information on Switch A.

```
[SwitchA] display pim rp-info
 BSR RP information:
   Scope: non-scoped
     Group/MaskLen: 225.1.1.0/24
       RP address               Priority  HoldTime  Uptime    Expires
       192.168.9.2              192       180       00:51:45  00:02:22

Static RP information:
       RP address               ACL   Mode    Preferred
       192.168.1.2              ----  pim-sm  No
```

Example: Configuring admin-scoped PIM-SM

Network configuration

As shown in Figure 18:

- VOD streams are sent to receiver hosts in multicast. The entire PIM-SM domain is divided into admin-scoped zone 1 and the global-scoped zone. Switch B and Switch C are ZBRs of admin-scoped zone 1.
- Source 1 sends multicast data to the multicast group 224.1.1.1, and Host A is a multicast receiver. Source 2 sends multicast data to multicast group 239.1.1.1, and Host B is a multicast receiver.
- VLAN-interface 20 of Switch B acts as a C-BSR and a C-RP for admin-scoped zone 1, and is designated to multicast group range 239.0.0.0/8. VLAN-interface 50 of Switch D acts as a C-BSR and a C-RP for the global-scoped zone, and is designated to all the multicast groups that are not in the range 239.0.0.0/8.
- IGMPv2 runs between Switch A, Switch C, and the receivers that directly connect to them, respectively.

Figure 18 Network diagram

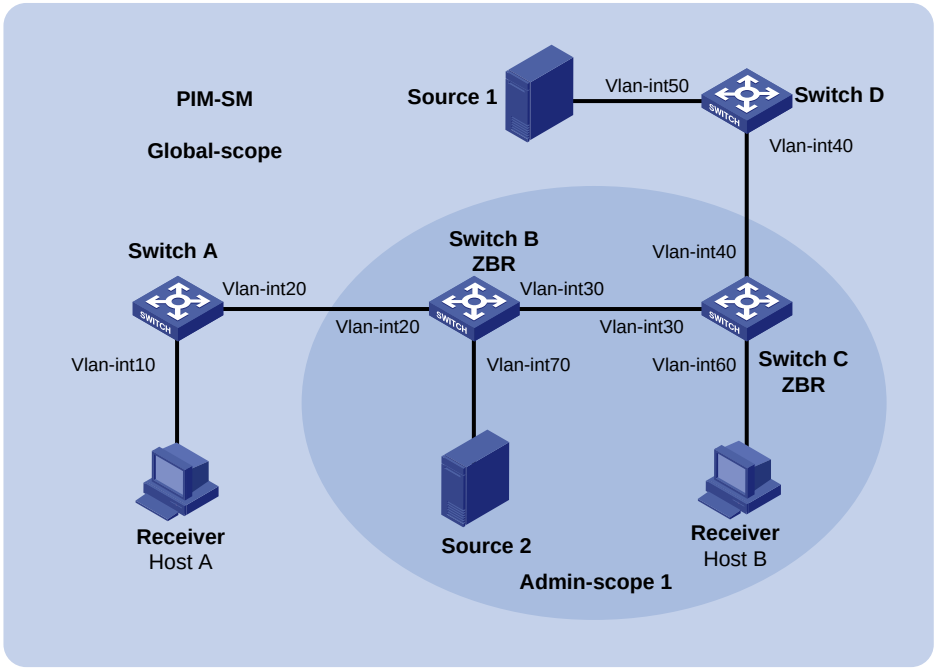


Table 3 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source 1	—	10.110.5.100/24	Source 2	—	10.110.6.100/24
Switch A	Vlan-int10	192.168.1.1/24	Switch D	Vlan-int50	10.110.5.2/24
Switch A	Vlan-int20	10.110.1.1/24	Switch D	Vlan-int40	10.110.7.1/24
Switch B	Vlan-int70	10.110.6.1/24	Switch C	Vlan-int30	10.110.2.2/24
Switch B	Vlan-int20	10.110.1.2/24	Switch C	Vlan-int60	10.110.4.1/24
Switch B	Vlan-int30	10.110.2.1/24	Switch C	Vlan-int40	10.110.7.2/24

Prerequisites

Assign an IP address and subnet mask to each interface, and make sure the switches in the PIM-DM domain can reach other.

Procedure

1. Enable IP multicast routing, and enable IGMP and PIM-SM:

   # On Switch A, enable IP multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

   # Enable IGMP on the receiver-side interface (VLAN-interface 20).

   ```
   [SwitchA] interface vlan-interface 20
   [SwitchA-Vlan-interface20] igmp enable
   [SwitchA-Vlan-interface20] quit
   ```

   # Enable PIM-SM on VLAN-interface 10.

   ```
   [SwitchA] interface vlan-interface 10
   ```

```
[SwitchA-Vlan-interface10] pim sm
[SwitchA-Vlan-interface10] quit
```

Enable IP multicast routing, IGMP, and PIM-SM on Switch C in the same way Switch A is configured. (Details not shown.)

On Switch B, enable IP multicast routing, and enable PIM-SM on each interface.

```
<SwitchB> system-view
[SwitchB] multicast routing
[SwitchB-mrib] quit
[SwitchB] interface vlan-interface 70
[SwitchB-Vlan-interface70] pim sm
[SwitchB-Vlan-interface70] quit
[SwitchB] interface vlan-interface 20
[SwitchB-Vlan-interface20] pim sm
[SwitchB-Vlan-interface20] quit
[SwitchB] interface vlan-interface 30
[SwitchB-Vlan-interface30] pim sm
[SwitchB-Vlan-interface30] quit
```

Enable IP multicast routing and PIM-SM on Switch D in the same way Switch B is configured. (Details not shown.)

2. Configure admin-scoped zone boundaries:

On Switch B, configure VLAN-interface 20 as the boundary of admin-scoped zone 1.

```
[SwitchB] interface vlan-interface 20
[SwitchB-Vlan-interface20] multicast boundary 239.0.0.0 8
[SwitchB-Vlan-interface20] quit
```

On Switch C, configure VLAN-interface 40 as the boundary of the global-scoped zone.

```
<SwitchC> system-view
[SwitchC] interface vlan-interface 40
[SwitchC-Vlan-interface40] multicast boundary 239.0.0.0 8
[SwitchC-Vlan-interface40] quit
```

3. Configure C-BSRs and C-RPs:

On Switch B, configure the service scope of RP advertisements, and configure VLAN-interface 20 as a C-BSR and a C-RP for admin-scoped zone 1.

```
[SwitchB] acl basic 2001
[SwitchB-acl-ipv4-basic-2001] rule permit source 239.0.0.0 0.255.255.255
[SwitchB-acl-ipv4-basic-2001] quit
[SwitchB] pim
[SwitchB-pim] c-bsr 10.110.1.2 scope 239.0.0.0 8
[SwitchB-pim] c-rp 10.110.1.2 group-policy 2001
[SwitchB-pim] quit
```

Configure VLAN-interface 50 as a C-BSR and a C-RP for admin-scoped zone 2.

```
[SwitchD] pim
[SwitchD-pim] c-bsr 10.110.5.2
[SwitchD-pim] c-rp 10.110.5.2
[SwitchD-pim] quit
```

Verifying the configuration

Display BSR information on Switch B.

```
[SwitchB] display pim bsr-info
```

```
 Scope: non-scoped
     State: Accept Preferred
     Bootstrap timer: 00:01:17
     Elected BSR address: 10.110.5.2
       Priority: 64
       Hash mask length: 30
       Uptime: 00:05:02
 Scope: 239.0.0.0/8
     State: Elected
     Bootstrap timer: 00:00:26
     Elected BSR address: 10.110.1.2
       Priority: 64
       Hash mask length: 30
       Uptime: 00:05:48
     Candidate BSR address: 10.110.1.2
       Priority: 64
       Hash mask length: 30
```

Display BSR information on Switch D.

```
[SwitchD] display pim bsr-info
 Scope: non-scoped
     State: Elected
     Bootstrap timer: 00:00:38
     Elected BSR address: 10.110.5.2
       Priority: 64
       Hash mask length: 30
       Uptime: 00:06:31
     Candidate BSR address: 10.110.5.2
       Priority: 64
       Hash mask length: 30
```

Display RP information on Switch B.

```
[SwitchB] display pim rp-info
 BSR RP information:
   Scope: non-scoped
     Group/MaskLen: 224.0.0.0/4
       RP address               Priority  HoldTime  Uptime    Expires
       10.110.5.2               192       180       00:07:05  00:02:54
   Scope: 239.0.0.0/8
     Group/MaskLen: 239.0.0.0/8
       RP address               Priority  HoldTime  Uptime    Expires
       10.110.1.2 (local)       192       180       00:07:47  00:02:13
```

Display RP information on Switch D.

```
[SwitchD] display pim rp-info
 BSR RP information:
   Scope: non-scoped
     Group/MaskLen: 224.0.0.0/4
       RP address               Priority  HoldTime  Uptime    Expires
       10.110.5.2 (local)       192       180       00:08:20  00:02:39
```

Example: Configuring BIDIR-PIM

Network configuration

As shown in Figure 19:

- OSPF runs on the network.
- VOD streams are sent to receiver hosts in multicast.
- Source 1 and Source 2 send multicast data to multicast group 225.1.1.1.
- Host A and Host B are receivers of this multicast group.
- VLAN-interface 102 of Switch C acts as the C-BSR. Loopback 0 of Switch C acts as the C-RP.
- IGMPv2 runs between Switch B and Host A, and between Switch D and Host B.

Figure 19 Network diagram

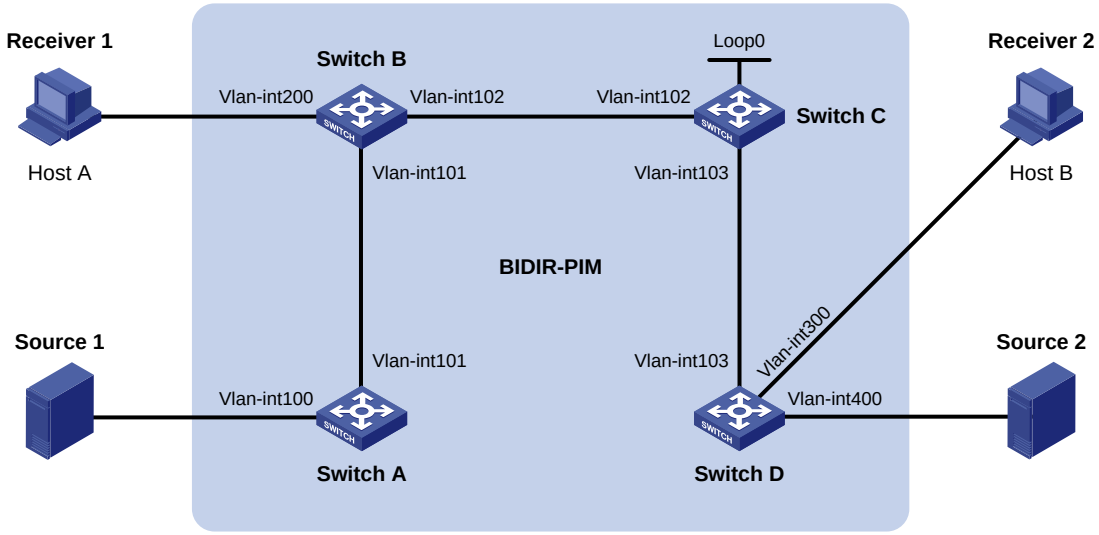


Table 4 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	192.168.1.1/24	Switch D	Vlan-int300	192.168.3.1/24
Switch A	Vlan-int101	10.110.1.1/24	Switch D	Vlan-int400	192.168.4.1/24
Switch B	Vlan-int200	192.168.2.1/24	Switch D	Vlan-int103	10.110.3.2/24
Switch B	Vlan-int101	10.110.1.2/24	Source 1	—	192.168.1.100/24
Switch B	Vlan-int102	10.110.2.1/24	Source 2	—	192.168.4.100/24
Switch C	Vlan-int102	10.110.2.2/24	Receiver 1	—	192.168.2.100/24
Switch C	Vlan-int103	10.110.3.1/24	Receiver 2	—	192.168.3.100/24
Switch C	Loop0	1.1.1.1/32			

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 19, and configure OSPF on the switches in the BIDIR-PIM domain. (Details not shown.)

Procedure

1. Enable IP multicast routing, PIM-SM, BIDIR-PIM, and IGMP:

On Switch A, enable IP multicast routing, enable PIM-SM on each interface, and enable BIDIR-PIM.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] pim
[SwitchA-pim] bidir-pim enable
[SwitchA-pim] quit
```

On Switch B, enable IP multicast routing.

```
<SwitchB> system-view
[SwitchB] multicast routing
[SwitchB-mrib] quit
```

Enable IGMP on the receiver-side interface (VLAN-interface 200).

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] igmp enable
[SwitchB-Vlan-interface200] quit
```

Enable PIM-SM on the other interfaces.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim sm
[SwitchB-Vlan-interface101] quit
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] pim sm
[SwitchB-Vlan-interface102] quit
```

Enable BIDIR-PIM.

```
[SwitchB] pim
[SwitchB-pim] bidir-pim enable
[SwitchB-pim] quit
```

On Switch C, enable IP multicast routing, enable PIM-SM on each interface, and enable BIDIR-PIM.

```
<SwitchC> system-view
[SwitchC] multicast routing
[SwitchC-mrib] quit
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] pim sm
[SwitchC-Vlan-interface102] quit
[SwitchC] interface vlan-interface 103
[SwitchC-Vlan-interface103] pim sm
[SwitchC-Vlan-interface103] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] pim sm
[SwitchC-LoopBack0] quit
[SwitchC] pim
```

```
[SwitchC-pim] bidir-pim enable
```

On Switch D, enable IP multicast routing.

```
<SwitchD> system-view
[SwitchD] multicast routing
[SwitchD-mrib] quit
```

Enable IGMP on the receiver-side interface (VLAN-interface 300).

```
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] igmp enable
[SwitchD-Vlan-interface300] quit
```

Enable PIM-SM on the other interfaces.

```
[SwitchD] interface vlan-interface 400
[SwitchD-Vlan-interface400] pim sm
[SwitchD-Vlan-interface400] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim sm
[SwitchD-Vlan-interface103] quit
```

Enable BIDIR-PIM.

```
[SwitchD] pim
[SwitchD-pim] bidir-pim enable
[SwitchD-pim] quit
```

2. On Switch C, configure VLAN interface 102 as a C-BSR, and Loopback 0 as a C-RP for the entire BIDIR-PIM domain.

```
<SwitchC> system-view
[SwitchC-pim] c-bsr 10.110.2.2
[SwitchC-pim] c-rp 1.1.1.1 bidir
[SwitchC-pim] quit
```

Verifying the configuration

1. Display BIDIR-PIM DF information:

Display BIDIR-PIM DF information on Switch A.

```
[SwitchA] display pim df-info
RP address: 1.1.1.1
  Interface: Vlan-interface100
    State     : Win        DF preference: 100
    DF metric : 2          DF uptime    : 01:08:50
    DF address: 192.168.1.1 (local)
  Interface: Vlan-interface101
    State     : Lose       DF preference: 100
    DF metric : 1          DF uptime    : 01:07:49
    DF address: 10.110.1.2
```

Display BIDIR-PIM DF information on Switch B.

```
[SwitchB] display pim df-info
RP address: 1.1.1.1
  Interface: Vlan-interface200
    State     : Win        DF preference: 100
    DF metric : 1          DF uptime    : 01:24:09
    DF address: 192.168.2.1 (local)
  Interface: Vlan-interface101
```

```
    State     : Win        DF preference: 100
    DF metric : 1          DF uptime    : 01:24:09
    DF address: 10.110.1.2 (local)
  Interface: Vlan-interface102
    State     : Lose       DF preference: 0
    DF metric : 0          DF uptime    : 01:23:12
    DF address: 10.110.2.2
```

Display BIDIR-PIM DF information on Switch C.

```
[SwitchC] display pim df-info
RP address: 1.1.1.1
  Interface: Loop0
    State     : -          DF preference: -
    DF metric : -          DF uptime    : -
    DF address: -
  Interface: Vlan-interface102
    State     : Win        DF preference: 0
    DF metric : 0          DF uptime    : 01:06:07
    DF address: 10.110.2.2 (local)
  Interface: Vlan-interface103
    State     : Win        DF preference: 0
    DF metric : 0          DF uptime    : 01:06:07
    DF address: 10.110.3.1 (local)
```

Display BIDIR-PIM DF information on Switch D.

```
[SwitchD] display pim df-info
RP address: 1.1.1.1
  Interface: Vlan-interface300
    State     : Win        DF preference: 100
    DF metric : 1          DF uptime    : 01:19:53
    DF address: 192.168.3.1 (local)
  Interface: Vlan-interface400
    State     : Win        DF preference: 100
    DF metric : 1          DF uptime    : 00:39:34
    DF address: 192.168.4.1 (local)
  Interface: Vlan-interface103
    State     : Lose       DF preference: 0
    DF metric : 0          DF uptime    : 01:21:40
    DF address: 10.110.3.1
```

2. Display information about DFs for multicast forwarding:

Display information about DFs for multicast forwarding on Switch A.

```
[SwitchA] display multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 1.1.1.1
     Flags: 0x0
     Uptime: 00:08:32
     RPF interface: Vlan-interface101
     List of 1 DF interfaces:
       1: Vlan-interface100
```

Display information about DFs for multicast forwarding on Switch B.

```
[SwitchB] display multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 1.1.1.1
     Flags: 0x0
     Uptime: 00:06:24
     RPF interface: Vlan-interface102
     List of 2 DF interfaces:
       1: Vlan-interface101
       2: Vlan-interface200
```

Display information about DFs for multicast forwarding on Switch C.

```
[SwitchC] display multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 1.1.1.1
     Flags: 0x0
     Uptime: 00:07:21
     RPF interface: LoopBack0
     List of 2 DF interfaces:
       1: Vlan-interface102
       2: Vlan-interface103
```

Display information about DFs for multicast forwarding on Switch D.

```
[SwitchD] display multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 1.1.1.1
     Flags: 0x0
     Uptime: 00:05:12
     RPF interface: Vlan-interface103
     List of 2 DF interfaces:
       1: Vlan-interface300
       2: Vlan-interface400
```

Example: Configuring PIM-SSM

Network configuration

As shown in Figure 20:

- OSPF runs on the network.
- The receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist on each stub network.
- The entire PIM domain operates in the SSM mode.
- Host A and Host C are multicast receivers on two stub networks.
- The SSM group range is 232.1.1.0/24.
- IGMPv3 runs between Switch A and N1, and between Switch B, Switch C, and N2.

Figure 20 Network diagram

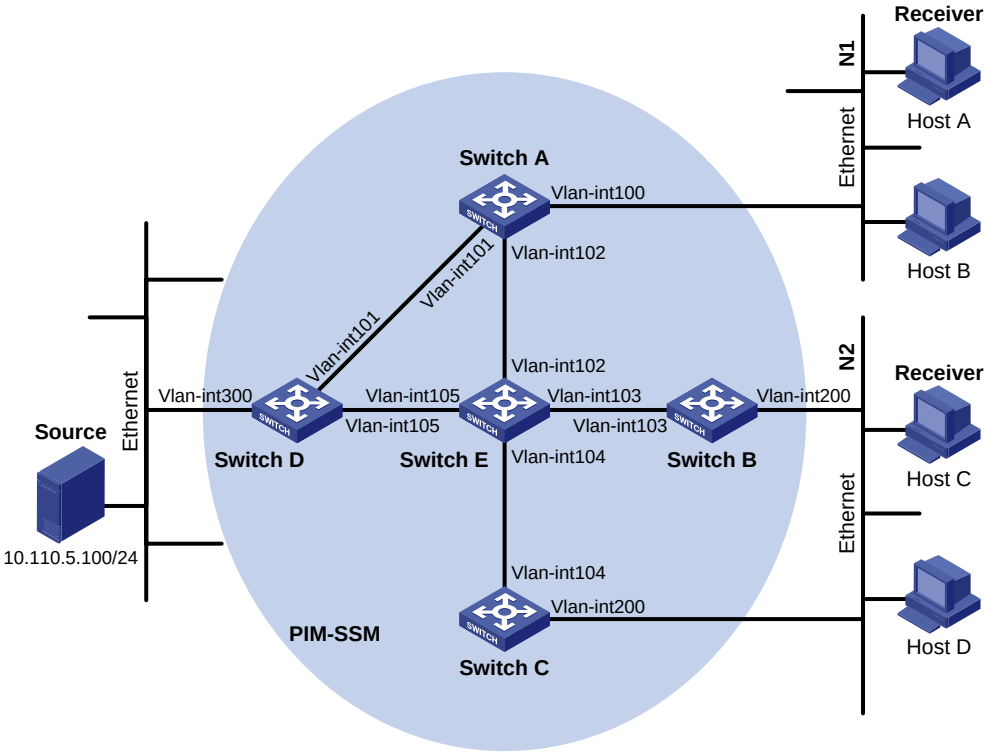


Table 5 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
Switch A	Vlan-int101	192.168.1.1/24	Switch D	Vlan-int101	192.168.1.2/24
Switch A	Vlan-int102	192.168.9.1/24	Switch D	Vlan-int105	192.168.4.2/24
Switch B	Vlan-int200	10.110.2.1/24	Switch E	Vlan-int104	192.168.3.2/24
Switch B	Vlan-int103	192.168.2.1/24	Switch E	Vlan-int103	192.168.2.2/24
Switch C	Vlan-int200	10.110.2.2/24	Switch E	Vlan-int102	192.168.9.2/24
Switch C	Vlan-int104	192.168.3.1/24	Switch E	Vlan-int105	192.168.4.1/24

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 20, and configure OSPF on the switches in the PIM-SSM domain.

Procedure

1. Enable IP multicast routing, IGMP, and PIM-SM:

   # On Switch A, enable IP multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

   # Enable IGMPv3 on VLAN-interface 100 (the interface that connects to the stub network).

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] igmp enable
   ```

```
[SwitchA-Vlan-interface100] igmp version 3
[SwitchA-Vlan-interface100] quit
```

Enable PIM-SM on the other interfaces.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim sm
[SwitchA-Vlan-interface102] quit
```

Enable IP multicast routing, IGMP, and PIM-SM on Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

Enable IP multicast routing and PIM-SM on Switch D and Switch E in the same way Switch A is configured. (Details not shown.)

2. Configure the SSM group range:

Configure the SSM group range 232.1.1.0/24 on Switch A.

```
[SwitchA] acl basic 2000
[SwitchA-acl-ipv4-basic-2000] rule permit source 232.1.1.0 0.0.0.255
[SwitchA-acl-ipv4-basic-2000] quit
[SwitchA] pim
[SwitchA-pim] ssm-policy 2000
[SwitchA-pim] quit
```

Configure the SSM group range on Switch B, Switch C, Switch D and Switch E in the same way Switch A is configured. (Details not shown.)

Verifying the configuration

Display PIM information on Switch A.

```
[SwitchA] display pim interface
 Interface: Vlan101
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: 192.168.1.2
 Interface: Vlan102
   NbrCnt: 0
   HelloInt: 30
   DR priority: 1
   DR address: 192.168.9.2
```

Send an IGMPv3 report from Host A to join the multicast source and group (10.110.5.100/24, 232.1.1.1). (Details not shown.)

Display the PIM routing table on Switch A.

```
[SwitchA] display pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 232.1.1.1)
     Protocol: pim-ssm, Flag: ACT
     UpTime: 00:13:25
     Upstream interface: Vlan-interface101
         Upstream neighbor: 192.168.1.2
```

```
        RPF prime neighbor: 192.168.1.2
    Downstream interface(s) information:
    Total number of downstream interfaces: 1
        1: Vlan-interface100
            Protocol: igmp, UpTime: 00:13:25, Expires: 00:03:25
```

Display the PIM routing table on Switch D.

```
[SwitchD] display pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 232.1.1.1)
    Protocol: pim-ssm, Flag: LOC
    UpTime: 00:12:05
    Upstream interface: Vlan-interface300
        Upstream neighbor: NULL
        RPF prime neighbor: NULL
    Downstream interface(s) information:
    Total number of downstream interfaces: 1
        1: Vlan-interface105
            Protocol: pim-ssm, UpTime: 00:12:05, Expires: 00:03:25
```

The output shows that switches on the SPT path (Switch A and Switch D) have generated the correct (S, G) entries.

Example: Configuring a multicast source policy

Network configuration

As shown in Figure 21:

- Source 1 and Source 3 send multicast data with different content to multicast group 225.0.0.1, and Source 2 sends multicast data to multicast group 225.0.0.2.
- Host A acts as a receiver and belongs to both multicast group 225.0.0.1 and multicast group 225.0.0.2.
- IGMPv2 runs between Device B and Host A.
- Device A discards multicast data from multicast source 10.110.7.2 and multicast data destined for multicast group 225.0.0.2.

Figure 21 Network diagram

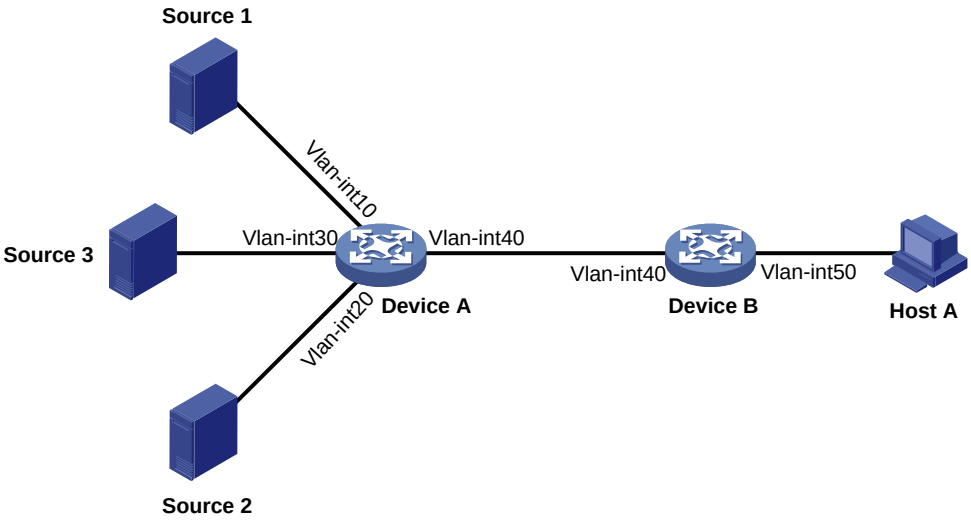


Table 6 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source 1		192.168.50.1/24	Device A	Vlan-int10	192.168.50.2/24
Source 2		192.168.56.1/24	Device A	Vlan-int20	192.168.56.2/24
Source 3		10.110.7.2/24	Device A	Vlan-int30	10.110.7.10/24
Device B	Vlan-int50	192.168.1.1/24	Device A	Vlan-int40	10.110.5.2/24
Device B	Vlan-int40	10.110.5.1/24			

Procedure

1. Configure Device A:

Enable IP multicast routing.

```
<DeviceA> system-view
[DeviceA] multicast routing
[DeviceA-mrib] quit
```

Configure OSPF.

```
[DeviceA] ospf 1
[DeviceA-ospf-1] area 0
[DeviceA-ospf-1-area-0.0.0.0] quit
[DeviceA-ospf-1] quit
```

Assign IP addresses to VLAN-interface 10, VLAN-interface 20, VLAN-interface 30, and VLAN-interface 40. Enable PIM-SM and OSPF on the interfaces.

```
[DeviceA] interface Vlan-interface10
[DeviceA-Vlan-interface10] ip address 192.168.50.2 255.255.255.0
[DeviceA-Vlan-interface10] pim sm
[DeviceA-Vlan-interface10] ospf 1 area 0
[DeviceA-Vlan-interface10] quit
[DeviceA] interface Vlan-interface20
[DeviceA-Vlan-interface20] ip address 192.168.56.2 255.255.255.0
[DeviceA-Vlan-interface20] pim sm
```

```
[DeviceA-Vlan-interface20] ospf 1 area 0
[DeviceA-Vlan-interface20] quit
[DeviceA] interface Vlan-interface30
[DeviceA-Vlan-interface30] ip address 10.110.7.10 255.255.255.0
[DeviceA-Vlan-interface30] pim sm
[DeviceA-Vlan-interface30] ospf 1 area 0
[DeviceA-Vlan-interface30] quit
[DeviceA] interface Vlan-interface40
[DeviceA-Vlan-interface40] ip address 10.110.5.2 255.255.255.0
[DeviceA-Vlan-interface40] pim sm
[DeviceA-Vlan-interface40] ospf 1 area 0
[DeviceA-Vlan-interface40] quit
```

Configure a multicast source policy to discard multicast data from multicast source 10.110.7.2 and multicast data destined for multicast group 225.0.0.2.

```
[DeviceA] acl advanced 3000
[DeviceA-acl-ipv4-adv-3000] rule 0 deny ip source 10.110.7.2 0
[DeviceA-acl-ipv4-adv-3000] rule 5 deny ip destination 225.0.0.2 0
[DeviceA-acl-ipv4-adv-3000] rule 10 permit ip
[DeviceA-acl-ipv4-adv-3000] quit
[DeviceA] pim
[DeviceA-pim] source-policy 3000
[DeviceA] quit
```

2. Configure Device B:

Enable IP multicast routing.

```
<DeviceB> system-view
[DeviceB] multicast routing
[DeviceB-mrib] quit
```

Configure OSPF.

```
[DeviceB] ospf 1
[DeviceB-ospf-1] area 0
[DeviceB-ospf-1-area-0.0.0.0] quit
[DeviceB-ospf-1] quit
```

Assign an IP address to VLAN-interface 50, and enable IGMP and OSPF on the interface.

```
[DeviceB] interface Vlan-interface50
[DeviceB-Vlan-interface50] ip address 192.168.1.1 255.255.255.0
[DeviceB-Vlan-interface50] igmp enable
[DeviceB-Vlan-interface50] ospf 1 area 0
[DeviceB-Vlan-interface50] quit
```

Assign an IP address to VLAN-interface 40, and enable IGMP and OSPF on the interface.

```
[DeviceB] interface Vlan-interface40
[DeviceB-Vlan-interface40] ip address 10.110.5.1 255.255.255.0
[DeviceB-Vlan-interface40] pim sm
[DeviceB-Vlan-interface40] ospf 1 area 0
[DeviceB-Vlan-interface40] quit
```

Assign an IP address to Loopback 0, and enable IGMP and OSPF on the interface.

```
[DeviceB] interface LoopBack0
[DeviceB-LoopBack0] ip address 1.1.1.1 255.255.255.255
[DeviceB-LoopBack0] pim sm
```

```
[DeviceB-LoopBack0] ospf 1 area 0
[DeviceB-LoopBack0] quit
```

Configure Loopback 0 as a C-BSR and a C-RP.

```
[DeviceB] pim
[DeviceB-pim] c-bsr 1.1.1.1
[DeviceB-pim] c-rp 1.1.1.1
[DeviceB-pim] quit
```

Verifying the configuration

Display the PIM routing table on Device B.

```
[DeviceB] display pim routing-table
 Total 2 (*, G) entries; 1 (S, G) entries

 (*, 225.0.0.1)
     RP: 1.1.1.1 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 16:27:44
     Upstream interface: Register-Tunnel0
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Vlan-interface50
             Protocol: igmp, UpTime: 16:27:44, Expires: -

 (192.168.50.1, 225.0.0.1)
     RP: 1.1.1.1 (local)
     Protocol: pim-sm, Flag: SPT 2MSDP ACT 2MVPN
     UpTime: 00:00:37
     Upstream interface: Vlan-interface40
         Upstream neighbor: 10.110.5.2
         RPF prime neighbor: 10.110.5.2
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Vlan-interface50
             Protocol: pim-sm, UpTime: 00:00:37, Expires: -

 (*, 225.0.0.2)
     RP: 1.1.1.1 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 16:27:52
     Upstream interface: Register-Tunnel0
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Vlan-interface50
             Protocol: igmp, UpTime: 16:27:52, Expires: -
```

The output shows that Host A can receive multicast data from only Source 1.

Troubleshooting PIM

A multicast distribution tree cannot be correctly built

Symptom

No multicast forwarding entries are established on the devices (including devices directly connected with multicast sources or receivers) in a PIM network. This means that a multicast distribution tree cannot be built correctly.

Solution

To resolve the problem:

1. Use the **display ip routing-table** command to verify that a unicast route to the multicast source or the RP is available.
2. Use the **display pim interface** command to verify PIM information on each interface, especially on the RPF interface. If PIM is not enabled on the interfaces, use the **pim dm** or **pim sm** command to enable PIM-DM or PIM-SM for the interfaces.
3. Use the **display pim neighbor** command to verify that the RPF neighbor is a PIM neighbor.
4. Verify that PIM and IGMP are enabled on the interfaces that directly connect to the multicast sources or the receivers.
5. Use the **display pim interface verbose** command to verify that the same PIM mode is enabled on the RPF interface on a device and the connected interface of the device's RPF neighbor.
6. Use the **display current-configuration** command to verify that the same PIM mode is enabled on all devices. For PIM-SM, verify that the BSR and C-RPs are correctly configured.
7. If the problem persists, contact H3C Support.

Multicast data is abnormally terminated on an intermediate device

Symptom

An intermediate device can receive multicast data successfully, but the data cannot reach the last-hop device. An interface on the intermediate device receives multicast data but does not create an (S, G) entry in the PIM routing table.

Solution

To resolve the problem:

1. Use the **display current-configuration** command to verify the multicast forwarding boundary settings. Use the **multicast boundary** command to change the multicast forwarding boundary settings to make the multicast packet able to cross the boundary.
2. Use the **display current-configuration** command to verify the multicast source policy. Change the ACL rule defined in the **source-policy** command so that the source/group address of the multicast data can pass ACL filtering.
3. If the problem persists, contact H3C Support.

An RP cannot join an SPT in PIM-SM

Symptom

An RPT cannot be correctly built, or an RP cannot join the SPT toward the multicast source.

Solution

To resolve the problem:

1. Use the **display ip routing-table** command to verify that a unicast route to the RP is available on each device.
2. Use the **display pim rp-info** command to verify that the dynamic RP information is consistent on all devices.
3. Use the **display pim rp-info** command to verify that the same static RPs are configured on all devices on the network.
4. If the problem persists, contact H3C Support.

An RPT cannot be built or multicast source registration fails in PIM-SM

Symptom

The C-RPs cannot unicast advertisement messages to the BSR. The BSR does not advertise BSMs containing C-RP information and has no unicast route to any C-RP. An RPT cannot be correctly established, or the source-side DR cannot register the multicast source with the RP.

Solution

To resolve the problem:

1. Use the **display ip routing-table** command on each device to view routing table information. Verify that unicast routes to the C-RPs and the BSR are available on each device and that a route is available between each C-RP and the BSR.
2. Use the **display pim bsr-info** command to verify that the BSR information exists on each device.
3. Use the **display pim rp-info** command to verify that the RP information is correct on each device.
4. Use the **display pim neighbor** command to verify that PIM neighboring relationship has been correctly established among the devices.
5. If the problem persists, contact H3C Support.

Contents

Configuring MSDP 1

About MSDP 1
How MSDP works 1
Anycast RP through MSDP 4
MSDP support for VPNs 5
Protocols and standards 5
MSDP tasks at a glance 5
Prerequisites for MSDP 6
Enabling MSDP 6
Creating an MSDP peer 7
Configuring a static RPF peer 7
Configuring an MSDP mesh group 7
Controlling MSDP peering connections 8
Tearing down MSDP peering connections 8
Setting the keepalive timer and peer hold timer for MSDP sessions 9
Setting the MSDP connection retry interval 9
Enabling MD5 authentication for TCP connection establishment 10
Configuring SA message-related parameters 10
Enabling multicast data encapsulation in SA messages 10
Configuring the originating RP of SA messages 11
Configuring SA request messages 11
Configuring SA message policies 12
Configuring the SA cache mechanism 12
Configuring MSDP-based Anycast RP 13
Enabling MSDP NSR 13
Setting the DSCP value for outgoing MSDP protocol packets 13
Display and maintenance commands for MSDP 14
MSDP configuration examples 15
Example: Configuring PIM-SM inter-domain multicast 15
Example: Configuring inter-AS multicast by leveraging static RPF peers 19
Example: Configuring Anycast RP 23
Example: Configuring SA message filtering 27
Troubleshooting MSDP 30
MSDP peers stay in disabled state 30
No SA entries exist in the device's SA message cache 30
No exchange of locally registered (S, G) entries between RPs 31

Configuring MSDP

About MSDP

Multicast Source Discovery Protocol (MSDP) is an inter-domain multicast solution that addresses the interconnection of PIM-SM domains. It discovers multicast source information in other PIM-SM domains.

MSDP is applicable only to the ASM model.

How MSDP works

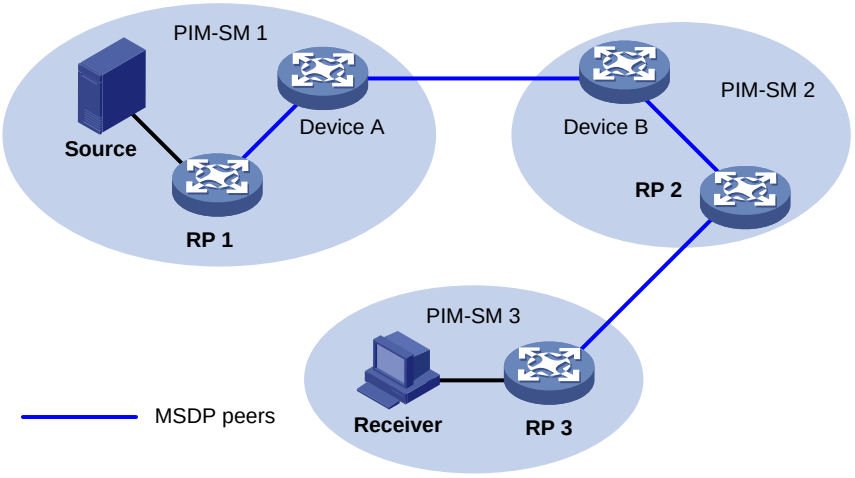
MSDP peer relationships are established between appropriate devices on the network, so that the RPs of different PIM-SM domains are interconnected with one another. These MSDP peers exchange source active (SA) messages, so that the multicast source information is shared among these domains.

For more information about DRs, BSRs, C-BSRs, RPs, C-RPs, SPTs, and RPTs, see "PIM overview."

MSDP peers

One or more pairs of MSDP peers on the network form an MSDP interconnection map.

Figure 1 MSDP peer locations on the network



As shown in Figure 1, an MSDP peer can be created on any PIM-SM device. MSDP peers created on PIM-SM devices that assume different roles function differently.

- MSDP peers created on RPs:
  - **Source-side MSDP peer**—MSDP peer closest to the multicast source, such as RP 1. The source-side RP creates and sends SA messages to its remote MSDP peer to notify the MSDP peer of the locally registered multicast source information.

    A source-side MSDP peer must be created on the source-side RP. Otherwise, it cannot advertise the multicast source information out of the PIM-SM domain.
  - **Receiver-side MSDP peer**—MSDP peer closest to the receivers, typically the receiver-side RP, such as RP 3. After receiving an SA message, the receiver-side MSDP peer resolves the multicast source information carried in the message. Then, it joins the SPT rooted at the multicast source across the PIM-SM domains. When multicast data from the multicast

source arrives, the receiver-side MSDP peer forwards the data to the receivers along the RPT.

- **Intermediate MSDP peer**—MSDP peer with multiple remote MSDP peers, such as RP 2. An intermediate MSDP peer forwards SA messages received from one remote MSDP peer to other remote MSDP peers. It acts as a relay for forwarding multicast source information.

- MSDP peers created on PIM-SM devices that are not RPs:

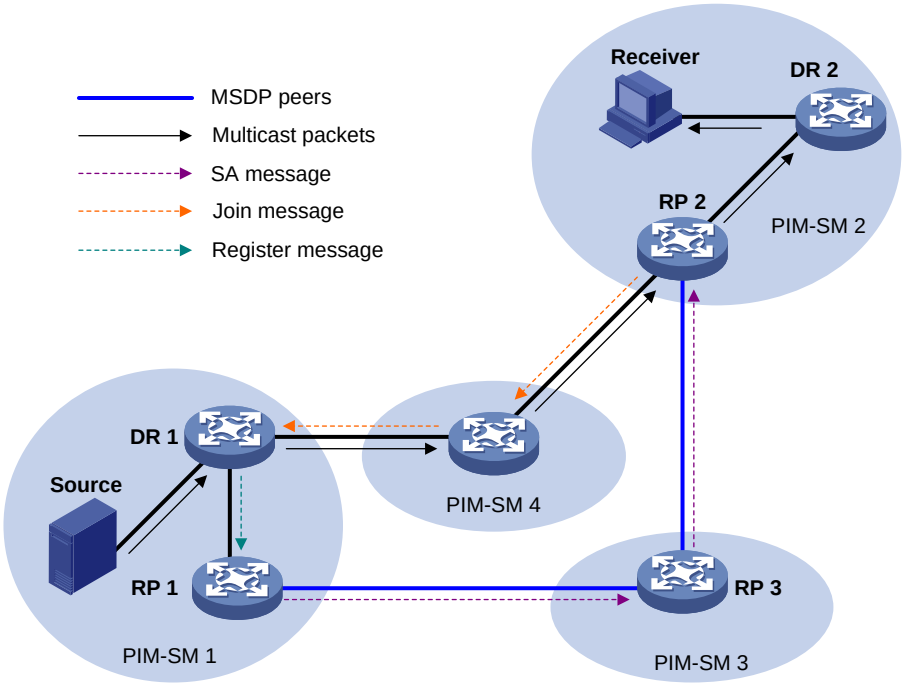
  Device A and Device B are MSDP peers on multicast devices that are not RPs. Such MSDP peers only forward SA messages.

In a PIM-SM network using the BSR mechanism, the RP is dynamically elected from C-RPs. A PIM-SM network typically has multiple C-RPs to ensure network robustness. Because the RP election result is unpredictable, MSDP peering relationships must be built between all C-RPs to always keep the winning C-RP on the MSDP interconnection map. Losing C-RPs assume the role of common PIM-SM devices on this map.

Inter-domain multicast delivery through MSDP

As shown in Figure 2, an active source (Source) exists in the domain PIM-SM 1, and RP 1 has learned the existence of Source through multicast source registration. RPs in PIM-SM 2 and PIM-SM 3 also seek the location of Source so that multicast traffic from Source can be sent to their receivers. MSDP peering relationships must be established between RP 1 and RP 3 and between RP 3 and RP 2.

Figure 2 Inter-domain multicast delivery through MSDP



The process of implementing PIM-SM inter-domain multicast delivery by leveraging MSDP peers is as follows:

1. When the multicast source in PIM-SM 1 sends the first multicast packet to multicast group G, DR 1 encapsulates the data within a register message. It sends the register message to RP 1, and RP 1 obtains information about the multicast source.
2. As the source-side RP, RP 1 creates SA messages and periodically sends them to its MSDP peer.

   An SA message contains the address of the multicast source (S), the multicast group address (G), and the address of the RP that has created this SA message (RP 1, in this example).

3. On MSDP peers, each SA message undergoes an RPF check and multicast policy-based filtering. Only SA messages that have arrived along the correct path and passed the filtering are received and forwarded. This avoids delivery loops of SA messages. In addition, you can configure MSDP peers into an MSDP mesh group to avoid SA message flooding between MSDP peers. An MSDP mesh group refers to a group of MSDP peers that establish MSDP peering relationships with each other and share the same group name.
4. SA messages are forwarded from one MSDP peer to another. Finally, information about the multicast source traverses all PIM-SM domains with MSDP peers (PIM-SM 2 and PIM-SM 3, in this example).
5. After receiving the SA message that RP 1 created, RP 2 in PIM-SM 2 examines whether any receivers for the multicast group exist in the domain.
    - If a receiver exists in the domain, the RPT for the multicast group G is maintained between RP 2 and the receivers. RP 2 creates an (S, G) entry and sends an (S, G) join message. The join message travels hop by hop toward the multicast source, and the SPT is established across the PIM-SM domains.

      The subsequent multicast data flows to RP 2 along the SPT, and from RP 2 to the receiver-side DR along the RPT. After receiving the multicast data, the receiver-side DR determines whether to initiate an RPT-to-SPT switchover process based on its configuration.
    - If no receivers exist in the domain, RP 2 neither creates an (S, G) entry nor sends a join message toward the multicast source.

In inter-domain multicasting using MSDP, once an RP gets information about a multicast source in another PIM-SM domain, it no longer relies on RPs in other PIM-SM domains. The receivers can override the RPs in other domains and directly join the multicast SPT rooted at the source.

MSDP peer-RPF forwarding

The MSDP peer-RPF check is used for forwarding SA messages on a network that runs MSDP. If the peer-RPF check succeeds, the SA message is accepted and forwarded. Otherwise, the SA message is discarded.

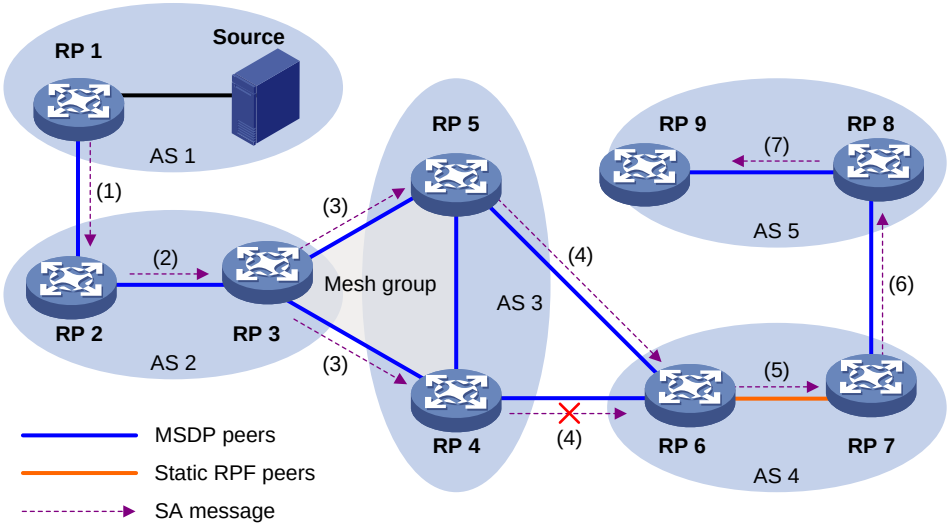
As shown in Figure 3:

- There are five ASs on the network. IGP runs within each AS, and BGP or MBGP runs between these ASs.
- Each AS contains a minimum of one PIM-SM domain, and each PIM-SM domain contains a minimum of one RP.
- MSDP peering relationship has been established among these RPs.

  RP 3, RP 4, and RP 5 are in the same MSDP mesh group.

  RP 6 is configured as the static RPF peer of RP 7.

Figure 3 MSDP peer-RPF forwarding



The process of peer-RPF forwarding is as follows:

1. RP 1 creates an SA message and forwards it to its peer RP 2.
2. RP 2 determines that RP 1 is the RP that creates the SA message because the RP address in the SA message is the same as that of RP 1. Then, RP 2 accepts and forwards the SA message.
3. RP 3 accepts and forwards the SA message, because RP 2 and RP 3 reside in the same AS and RP 2 is the next hop of RP 3 to RP 1.
4. RP 4 and RP 5 accept the SA message, because RP 3 is in the same mesh group with them. Then, RP 4 and RP 5 forward the SA message to their peer RP 6 rather than other members of the mesh group.
5. RP 4 and RP 5 reside in the closest AS in the route to RP 1. However, RP 6 accepts and forwards only the SA message from RP 5, because the IP address of RP 5 is higher than that of RP 4.
6. RP 7 accepts and forwards the SA message, because RP 6 is its static RPF peer.
7. RP 8 accepts and forwards the SA message, because RP 7 is the EBGP or MBGP next hop of the peer-RPF route to RP 1.
8. RP 9 accepts the SA message, because RP 8 is the only RP of RP 9.

Anycast RP through MSDP

PIM-SM requires only one active RP to serve each multicast group. If the active RP fails, the multicast traffic might be interrupted. The Anycast RP mechanism enables redundancy backup between two or more RPs by configuring multiple RPs with the same IP address for one multicast group. A multicast source registers with the closest RP or a receiver joins the closest RP to implement source information synchronization.

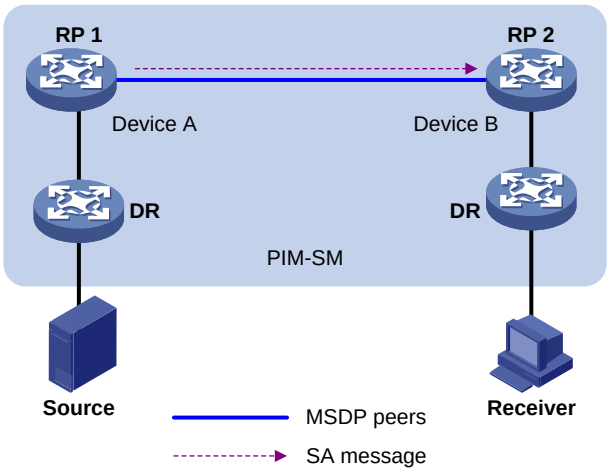
For Anycast RP implemented through MSDP, all member RPs in the Anycast RP set have the same IP address, and MSDP peering connections are established between the RPs.

As shown in Figure 4, within a PIM-SM domain, a multicast source sends multicast data to multicast group G, and the receiver joins the multicast group. To implement Anycast RP through MSDP, complete the following tasks:

1. Assign the same IP address (known as Anycast RP address, typically a private address) to an interface on Device A and Device B.
2. Configure the interfaces as C-RPs.

3. Establish an MSDP peering relationship between Device A and Device B.

Figure 4 Anycast RP through MSDP



The following describes how Anycast RP through MSDP is implemented:

4. After receiving the multicast data from Source, the source-side DR registers with the closest RP (RP 1 in this example).
5. After receiving the IGMP report message from the receiver, the receiver-side DR sends a join message toward the closest RP (RP 2 in this example). An RPT rooted at this RP is established.
6. The RPs share the registered multicast source information through SA messages. After obtaining the multicast source information, RP 2 sends an (S, G) source-specific join message toward the source to create an SPT.
7. When the multicast data reaches RP 2 along the SPT, the RP forwards the data along the RPT to the receiver. After receiving the multicast data, the receiver-side DR determines whether to initiate an RPT-to-SPT switchover process based on its configuration.

MSDP support for VPNs

Interfaces on the multicast devices in a VPN can set up MSDP peering relationships with each other. With the SA messages exchanged between MSDP peers, the multi-instance VPN implements the forwarding of multicast data across different PIM-SM domains.

To support MSDP for VPNs, a multicast device that runs MSDP maintains an independent set of MSDP mechanism for each VPN that it supports. These mechanisms include SA message cache, peering connection, timers, sending cache, and cache for exchanging PIM messages.

One VPN is isolated from another, and MSDP and PIM-SM messages can be exchanged only within the same VPN.

Protocols and standards

- RFC 3618, *Multicast Source Discovery Protocol (MSDP)*
- RFC 3446, *Anycast Rendezvous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)*

MSDP tasks at a glance

1. Enabling MSDP

2. Creating an MSDP peer
3. (Optional.) Configuring a static RPF peer
4. (Optional.) Configuring an MSDP mesh group
5. (Optional.) Controlling MSDP peering connections
   - Tearing down MSDP peering connections
   - Setting the keepalive timer and peer hold timer for MSDP sessions
   - Setting the MSDP connection retry interval
   - Enabling MD5 authentication for TCP connection establishment
6. (Optional.) Configuring SA message-related parameters
   - Enabling multicast data encapsulation in SA messages
   - Configuring the originating RP of SA messages
   - Configuring SA request messages
   - Configuring SA message policies
   - Configuring the SA cache mechanism
7. (Optional.) Configuring MSDP-based Anycast RP
8. (Optional.) Enabling MSDP NSR
9. (Optional.) Setting the DSCP value for outgoing MSDP protocol packets

Prerequisites for MSDP

Before you configure MSDP, complete the following tasks:

- Configure a unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Configure PIM-SM to enable intra-domain multicast.

Enabling MSDP

1. Enter system view.
   system-view
2. Enable IP multicast routing and enter MRIB view.
   multicast routing [**vpn-instance** *vpn-instance-name*]
   By default, IP multicast routing is disabled.
   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.
   quit
4. Enable MSDP and enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
   By default, MSDP is disabled.

Creating an MSDP peer

About this task

An MSDP peering relationship is identified by an address pair (the addresses of the local MSDP peer and the remote MSDP peer). To create an MSDP peering connection, you must perform the following operation on both devices that are a pair of MSDP peers.

You can configure a description for an MSDP peer to easily distinguish an MSDP peer from other MSDP peers.

Restrictions and guidelines

If an MSDP peer and a BGP or MBGP peer share the same interface, specify the MSDP peer and the BGP or MBGP peer by using the same IP address.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Create an MSDP peer.
   peer *peer-address* **connect-interface** *interface-type interface-number*
4. (Optional.) Configure a description for the MSDP peer.
5. **peer** *peer-address* **description** *text*
6. By default, an MSDP peer does not have a description.

Configuring a static RPF peer

About this task

You can configure a static RPF peer, so the static RPF peer does not perform RPF check on received SA messages. This simplifies the RPF check mechanism for SA messages.

Restrictions and guidelines

If only one MSDP peer is configured on a device, this MSDP peer is considered to be a static RPF peer.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Configure a static RPF peer.
   static-rpf-peer *peer-address* [**rp-policy** *ip-prefix-name*]

Configuring an MSDP mesh group

About this task

This feature avoids SA message flooding among MSDP peers within an AS. It also simplifies the RPF check mechanism because you do not need to run BGP or MBGP between these MSDP peers.

When receiving an SA message from outside the mesh group, a member MSDP peer performs the RPF check on the SA message. If the SA message passes the RPF check, the member MSDP peer floods the message to the other members in the mesh group. When receiving an SA message from another member, the MSDP peer neither performs an RPF check on the message nor forwards the message to the other members.

To organize multiple MSDP peers in a mesh group, assign the same mesh group name to these MSDP peers.

Restrictions and guidelines

Before you configure an MSDP mesh group, make sure the devices are interconnected with one another.

If you assign an MSDP peer to multiple MSDP mesh groups, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Configure an MSDP mesh group.
   peer *peer-address* **mesh-group** *name*
   By default, an MSDP peer does not belong to any mesh group.

Controlling MSDP peering connections

Tearing down MSDP peering connections

About this task

MSDP peers are interconnected over TCP (port number 639). You can tear down or re-establish MSDP peering connections to control SA message exchange between the MSDP peers. When the connection between two MSDP peers is torn down, SA messages are no longer delivered between them. No attempt is made to re-establish the connection. The configuration information for the peer remains unchanged.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Tear down an MSDP peering connection.
   shutdown *peer-address*
   By default, an MSDP peering connection is active.

Setting the keepalive timer and peer hold timer for MSDP sessions

About this task

MSDP peers periodically send keepalive messages to each other to keep a session alive. When a session is established, an MSDP peer sends a keepalive message to its peer and starts a keepalive timer and a peer hold timer. When the keepalive timer expires, the MSDP peer sends a new keepalive message. If the MSDP peer receives an MSDP message from its peer before the peer hold timer expires, it resets the peer hold timer. Otherwise, the MSDP peer tears down the session.

Because MSDP peers cannot negotiate the keepalive timer and peer hold timer, you must configure the same keepalive timer and peer hold timer for the MSDP peers of a session. The keepalive timer must be smaller than the peer hold timer.

Procedure

1. Enter system view.

   system-view
2. Enter MSDP view.

   msdp [**vpn-instance** *vpn-instance-name*]
3. Set the keepalive timer and peer hold timer for MSDP sessions.

   timer keepalive *keepalive holdtime*

   By default, the keepalive timer and peer hold timer are 60 seconds and 75 seconds, respectively.

   This command immediately takes effect on an established session.

Setting the MSDP connection retry interval

About this task

A TCP connection is required when one of the following conditions exists:

- A new MSDP peer is created.
- A previously deactivated MSDP peering connection is reactivated.
- A previously failed MSDP peer attempts to resume operation.

Perform this task to adjust the interval between MSDP peering connection attempts.

Procedure

1. Enter system view.

   system-view
2. Enter MSDP view.

   msdp [**vpn-instance** *vpn-instance-name*]
3. Set the MSDP connection retry interval.

   timer retry *interval*

   The default setting is 30 seconds.

Enabling MD5 authentication for TCP connection establishment

About this task

To enhance MSDP security, enable MD5 authentication for both MSDP peers to establish a TCP connection. If the MD5 authentication fails, the TCP connection cannot be established.

Restrictions and guidelines

Configure the same authentication method and key for the MSDP peers involved in MD5 authentication to avoid authentication and TCP connection failures.

Procedure

1. Enter system view.

   system-view

2. Enter MSDP view.

   msdp [**vpn-instance** *vpn-instance-name*]

3. Configure the device to perform MD5 authentication when establishing a TCP connection with an MSDP peer.

   peer *peer-address* **password** { **cipher** | **simple** } *password*

   By default, the device does not perform MD5 authentication when establishing a TCP connection with an MSDP peer.

Configuring SA message-related parameters

Enabling multicast data encapsulation in SA messages

About this task

Some multicast sources send multicast data at an interval longer than the aging time of (S, G) entries. In this case, the source-side DR must encapsulate multicast data packet-by-packet in register messages and send them to the source-side RP. The source-side RP transmits the (S, G) information to the remote RP through SA messages. Then, the remote RP sends join messages to the source-side DR and builds an SPT. Because the (S, G) entries have timed out, remote receivers can never receive the multicast data from the multicast source.

To avoid this problem, you can enable the source-side RP to encapsulate multicast data in SA messages. As a result, the source-side RP can forward the multicast data in SA messages to its remote MSDP peers. After receiving the SA messages, the remote RP decapsulates the SA messages and forwards the multicast data to the receivers in the local domain along the RPT.

Procedure

1. Enter system view.

   system-view

2. Enter MSDP view.

   msdp [**vpn-instance** *vpn-instance-name*]

3. Enable multicast data encapsulation in SA messages.

   encap-data-enable

   By default, an SA message contains only (S, G) entries, but not the multicast data.

Configuring the originating RP of SA messages

About this task

This feature enables an interface to originate SA messages and to use its IP address as the RP address in SA messages. It is typically used in the Anycast-RP application.

By default, the RP address in SA messages originated by a member Anycast-RP is the Anycast-RP address. The SA messages will fail the RPF check on the other members because the RP address in SA messages is the same as the local RP address. In this case, source information cannot be exchanged within the Anycast-RP. To solve the problem, you must specify an interface other than the interface where the member Anycast-RP resides as the originating RP of SA messages.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Configure an interface as the originating RP of SA messages.
   originating-rp *interface-type interface-number*
   By default, SA messages are originated by the actual RPs.

Configuring SA request messages

About this task

By default, after receiving a new join message, a device waits for an SA message to obtain the multicast source information and to join the SPT. You can enable the device to request source information by sending SA request messages to an MSDP peer. This reduces the join latency.

An SA request policy enables the device to filter SA request messages from an MSDP peer by using an ACL that specifies the multicast groups.

Prerequisites

Before you enable the device to send SA requests, make sure you have disabled the SA message cache mechanism by the **undo cache-sa-enable** command.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Enable the device to send SA request messages to an MSDP peer.
   peer *peer-address* **request-sa-enable**
   By default, after receiving a new join message, a device does not send an SA request message to any MSDP peer. Instead, it waits for the next SA message from its MSDP peer.
4. Configure an SA request policy for an MSDP peer.
   peer *peer-address* **sa-request-policy** [**acl** *ipv4-acl-number*]
   By default, no SA request policy exists, and all SA request from an MSDP peer are permitted.

Configuring SA message policies

About this task

To control the propagation of multicast source information, you can configure the following policies:

- **SA creation policy**—Limits the multicast source information advertised in SA messages. This policy enables the device to advertise (S, G) entries based on the used ACL that specifies the multicast sources and groups.
- **SA incoming or outgoing policy**—Limits the receipt or forwarding of SA messages. This policy enables the device to receive or forward SA messages based on the used ACL that specifies the multicast sources and groups.
- **The lower TTL threshold for multicast data packets encapsulated in SA messages**—By default, multicast data packets are encapsulated in SA messages and forwarded to MSDP peers only if the TTL values in the packets are larger than zero. You can set the lower TTL threshold for multicast data packets encapsulated in SA messages that are sent to an MSDP peer. Then, only multicast data packets whose TTL values are larger than or equal to the configured value are encapsulated in SA messages. Only SA messages whose TTL values are larger than or equal to the configured value are forwarded to the specified MSDP peer. This controls the multicast data packet encapsulation and limits the propagation range of the SA messages.

Procedure

1. Enter system view.

   system-view
2. Enter MSDP view.

   msdp [**vpn-instance** *vpn-instance-name*]
3. Configure an SA creation policy.

   import-source [**acl** *ipv4-acl-number*]
4. Configure an SA incoming or outgoing policy.

   peer *peer-address* **sa-policy** { **export** | **import** } [**acl** *ipv4-acl-number*]

   By default, no SA incoming or outgoing policy is configured. The device does not filter incoming or outgoing SA messages.
5. Set the lower TTL threshold for multicast data packets encapsulated in SA messages.

   peer *peer-address* **minimum-ttl** *ttl-value*

   The default setting is 0.

Configuring the SA cache mechanism

About this task

The SA cache mechanism enables the device to locally cache (S, G) entries contained in SA messages. It reduces the time for obtaining multicast source information, but increases memory occupation.

With the SA cache mechanism enabled, when the device receives a new (*, G) join message, it searches its SA message cache first.

- If no matching (S, G) entry is found, the device waits for the SA message that its MSDP peer sends in the next cycle.
- If a matching (S, G) entry is found in the cache, the device joins the SPT rooted at S.

To protect the device against DoS attacks, you can set a limit on the number of (S, G) entries in the SA cache from an MSDP peer.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Enable the SA cache mechanism.
   cache-sa-enable
   By default, the SA message cache mechanism is enabled. The device caches the (S, G) entries contained in the received SA messages.
4. (Optional.) Set the maximum number of (S, G) entries in the SA cache from an MSDP peer.
   peer *peer-address* **sa-cache-maximum** *sa-limit*
   The default setting is 4294967295.

Configuring MSDP-based Anycast RP

1. Configure the same RP on multiple devices.
   For more information about configuring RP, see "Configuring PIM."
2. Establish MSDP peer relationship between the devices. The RP address must be different from the MSDP peer address.
   For more information about establishing MSDP peer relationships, see "Creating an MSDP peer."
   The IP address for establishing MSDP peer relationship must be different from the Anycast RP address.

Enabling MSDP NSR

About this task

This feature enables MSDP to back up protocol state information and data, such as MSDP peer and remote multicast source information, from the active process to standby processes. A standby process seamlessly takes over when the active process fails. Use this feature to avoid forwarding interruption for MSDP when an active/standby switchover occurs.

Procedure

1. Enter system view.
   system-view
2. Enable MSDP NSR.
   msdp non-stop-routing
   By default, MSDP NSR is disabled.

Setting the DSCP value for outgoing MSDP protocol packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.
   system-view
2. Enter MSDP view.
   msdp [**vpn-instance** *vpn-instance-name*]
3. Set the DSCP value for outgoing MSDP protocol packets.
   dscp *dscp-value*
   By default, the DSCP value is 48 for outgoing MSDP protocol packets.

Display and maintenance commands for MSDP

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display brief information about MSDP peers.	**display msdp** [**vpn-instance** *vpn-instance-name*] **brief** [**state** { **connect** | **disabled** | **established** | **listen** | **shutdown** }]
Display MSDP NSR status information.	**display msdp non-stop-routing status**
Display detailed status of MSDP peers.	**display msdp** [**vpn-instance** *vpn-instance-name*] **peer-status** [*peer-address*]
Display (S, G) entries in the SA cache.	**display msdp** [**vpn-instance** *vpn-instance-name*] **sa-cache** [*group-address* | *source-address* | *as-number*] *
Display the number of (S, G) entries in the SA cache.	**display msdp** [**vpn-instance** *vpn-instance-name*] **sa-count** [*as-number*]
Reset the TCP connection with an MSDP peer and clear statistics for the MSDP peer.	**reset msdp** [**vpn-instance** *vpn-instance-name*] **peer** [*peer-address*]
Delete (S, G) entries in the SA cache.	**reset msdp** [**vpn-instance** *vpn-instance-name*] **sa-cache** [*group-address*]
Clear statistics for an MSDP peer without resetting the TCP connection with the MSDP peer.	**reset msdp** [**vpn-instance** *vpn-instance-name*] **statistics** [*peer-address*]

MSDP configuration examples

Example: Configuring PIM-SM inter-domain multicast

Network configuration

As shown in Figure 5:

- OSPF runs within AS 100 and AS 200 and BGP runs between them.
- Each PIM-SM domain has a minimum of one multicast source or receiver.

Set up MSDP peering relationships between the RPs in the PIM-SM domains to share multicast source information among the PIM-SM domains.

Figure 5 Network diagram

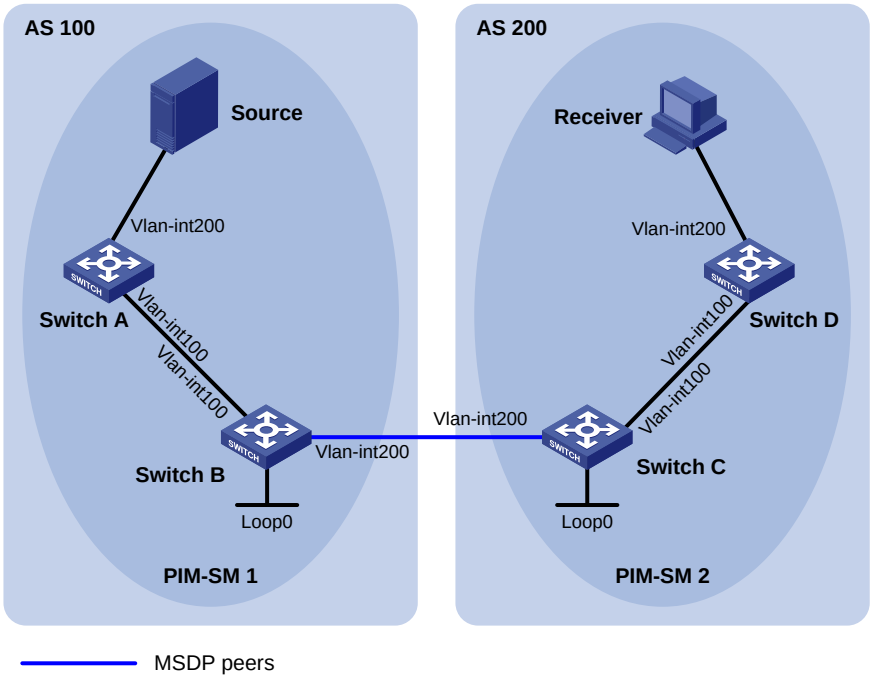


Table 1 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source	N/A	10.110.2.100/24	Switch C	Vlan-int100	10.110.4.1/24
Switch A	Vlan-int100	10.110.1.2/24	Switch C	Vlan-int200	192.168.1.2/24
Switch A	Vlan-int200	10.110.2.1/24	Switch C	Loop0	2.2.2.2/32
Switch B	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int100	10.110.4.2/24
Switch B	Vlan-int200	192.168.1.1/24	Switch D	Vlan-int200	10.110.5.1/24
Switch B	Loop0	1.1.1.1/32			

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 5, and configure OSPF on the switches.

Procedure

1. Enable IP multicast routing and PIM-SM, and configure PIM-SM domain borders:

   # On Switch A, enable IP multicast routing, and enable PIM-SM on interfaces.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] pim sm
[SwitchA-Vlan-interface200] quit
```

   # On Router D, enable IP multicast routing, enable IGMP on the receiver-side interface, and enable PIM-SM on the source-side interface.

```
<SwitchD> system-view
[SwitchD] multicast routing
[SwitchD-mrib] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] pim sm
[SwitchD-Vlan-interface100] quit
[SwitchD] interface vlan-interface 200
[SwitchD-Vlan-interface200] igmp enable
[SwitchD-Vlan-interface200] quit
```

   # On Router B, enable IP multicast routing, enable PIM-SM on interfaces, and configure the receiver-side interface as a PIM-SM domain border.

```
<SwitchB> system-view
[SwitchB] multicast routing
[SwitchB-mrib] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] pim sm
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] pim sm
[SwitchB-Vlan-interface200] pim bsr-boundary
[SwitchB-Vlan-interface200] quit
[SwitchB] interface LoopBack0
[SwitchB-LoopBack0] pim sm
[SwitchB-LoopBack0] quit
```

   # On Router C, enable IP multicast routing, enable PIM-SM on interfaces, and configure the receiver-side interface as a PIM-SM domain border.

```
<SwitchC> system-view
[SwitchC] multicast routing
[SwitchC-mrib] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] pim sm
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] pim sm
```

```
[SwitchC-Vlan-interface200] pim bsr-boundary
[SwitchC-Vlan-interface200] quit
[SwitchC] interface LoopBack0
[SwitchC-LoopBack0] pim sm
[SwitchC-LoopBack0] quit
```

2. Configure C-BSRs and C-RPs:

On Switch B, configure Loopback 0 as a C-BSR and a C-RP.

```
[SwitchB] pim
[SwitchB-pim] c-bsr 1.1.1.1
[SwitchB-pim] c-rp 1.1.1.1
[SwitchB-pim] quit
```

On Switch C, configure Loopback 0 as a C-BSR and a C-RP.

```
[SwitchC] pim
[SwitchC-pim] c-bsr 2.2.2.2
[SwitchC-pim] c-rp 2.2.2.2
[SwitchC-pim] quit
```

3. Configure BGP for mutual route redistribution between BGP and OSPF:

On Switch B, configure an EBGP peer, and redistribute OSPF routes.

```
[SwitchB] bgp 100
[SwitchB-bgp-default] router-id 1.1.1.1
[SwitchB-bgp-default] peer 192.168.1.2 as-number 200
[SwitchB-bgp-default] address-family ipv4
[SwitchB-bgp-default-ipv4] import-route ospf 1
[SwitchB-bgp-default-ipv4] peer 192.168.1.2 enable
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] address-family ipv4 multicast
[SwitchB-bgp-default-mul-ipv4] peer 192.168.1.2 enable
[SwitchB-bgp-default-mul-ipv4] quit
[SwitchB-bgp-default] quit
```

On Switch C, configure an EBGP peer, and redistribute OSPF routes.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 2.2.2.2
[SwitchC-bgp-default] peer 192.168.1.1 as-number 100
[SwitchC-bgp-default] address-family ipv4
[SwitchC-bgp-default-ipv4] import-route ospf 1
[SwitchC-bgp-default-ipv4] peer 192.168.1.1 enable
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] address-family ipv4 multicast
[SwitchC-bgp-default-mul-ipv4] peer 192.168.1.1 enable
[SwitchC-bgp-default-mul-ipv4] quit
[SwitchC-bgp-default] quit
```

Redistribute BGP routes into OSPF on Switch B.

```
[SwitchB] ospf 1
[SwitchB-ospf-1] import-route bgp
[SwitchB-ospf-1] quit
```

Redistribute BGP routes into OSPF on Switch C.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] import-route bgp
```

```
[SwitchC-ospf-1] quit
```

4. Configure MSDP peers:

Configure an MSDP peer on Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] peer 192.168.1.2 connect-interface vlan-interface 200
[SwitchB-msdp] quit
```

Configure an MSDP peer on Switch C.

```
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.1.1 connect-interface vlan-interface 200
[SwitchC-msdp] quit
```

Verifying the configuration

Display information about BGP peer groups on Switch B.

```
[SwitchB] display bgp peer ipv4

 BGP local router ID: 1.1.1.1
 Local AS number: 100
 Total number of peers: 1                 Peers in established state: 1

 * - Dynamically created peer
 Peer                    AS  MsgRcvd  MsgSent OutQ  PrefRcv Up/Down  State

 192.168.1.2             200 24       21      0    6        00:13:09 Established
```

Display information about BGP peer groups on Switch C.

```
[SwitchC] display bgp peer ipv4

 BGP local Switch ID: 2.2.2.2
 Local AS number: 200
 Total number of peers: 1                 Peers in established state: 1

 * - Dynamically created peer
 Peer                    AS  MsgRcvd  MsgSent OutQ  PrefRcv Up/Down  State

 192.168.1.1            100       18       16    0        1 00:12:04 Established
```

Display brief information about MSDP peers on Switch B.

```
[SwitchB] display msdp brief
Configured   Established  Listen       Connect      Shutdown     Disabled
1            1            0            0            0            0

Peer address    State       Up/Down time    AS         SA count   Reset count
192.168.1.2     Established 00:12:19        200        0          0
```

Display brief information about MSDP peers on Switch C.

```
[SwitchC] display msdp brief
Configured   Established  Listen       Connect      Shutdown     Disabled
1            1            0            0            0            0
Peer address    State       Up/Down time    AS         SA count   Reset count
192.168.1.1     Established 00:06:11        100        0          0
```

Display detailed MSDP peer information on Switch B.

```
[SwitchB] display msdp peer-status
MSDP Peer 192.168.1.2; AS 200
 Description:
 Information about connection status:
   State: Established
   Up/down time: 00:15:47
   Resets: 0
   Connection interface: Vlan-interface200 (192.168.1.1)
   Received/sent messages: 16/16
   Discarded input messages: 0
   Discarded output messages: 0
   Elapsed time since last connection or counters clear: 00:17:40
   Mesh group peer joined: momo
   Last disconnect reason: Hold timer expired with truncated message
   Truncated packet: 5 bytes in buffer, type: 1, length: 20, without packet time: 75s
 Information about (Source, Group)-based SA filtering policy:
   Import policy: None
   Export policy: None
 Information about SA-Requests:
   Policy to accept SA-Requests: None
   Sending SA-Requests status: Disable
 Minimum TTL to forward SA with encapsulated data: 0
 SAs learned from this peer: 0, SA cache maximum for the peer: 4294967295
 Input queue size: 0, Output queue size: 0
 Counters for MSDP messages:
   RPF check failure: 0
   Incoming/outgoing SA: 0/0
   Incoming/outgoing SA-Request: 0/0
   Incoming/outgoing SA-Response: 0/0
   Incoming/outgoing Keepalive: 867/867
   Incoming/outgoing Notification: 0/0
   Incoming/outgoing Traceroutes in progress: 0/0
   Incoming/outgoing Traceroute reply: 0/0
   Incoming/outgoing Unknown: 0/0
   Incoming/outgoing data packet: 0/0
```

Example: Configuring inter-AS multicast by leveraging static RPF peers

Network configuration

As shown in Figure 6:

- The network has two ASs: AS 100 and AS 200. OSPF runs within each AS. BGP runs between the two ASs.
- PIM-SM 1 belongs to AS 100, and PIM-SM 2 belongs to AS 200. Each PIM-SM domain has a minimum of one multicast source or receiver.

Configure inter-AS multicast as follows:

- Configure Loopback 0 as the C-BSR and C-RP of the related PIM-SM domain on Switch A and Switch D.
- According to the peer-RPF forwarding rule, the switches accept SA messages that pass the filtering policy from its static RPF peers. To share multicast source information among PIM-SM domains without changing the unicast topology structure, configure MSDP peering relationships for the RPs of the PIM-SM domains and configure the static RPF peering relationships.

Figure 6 Network diagram

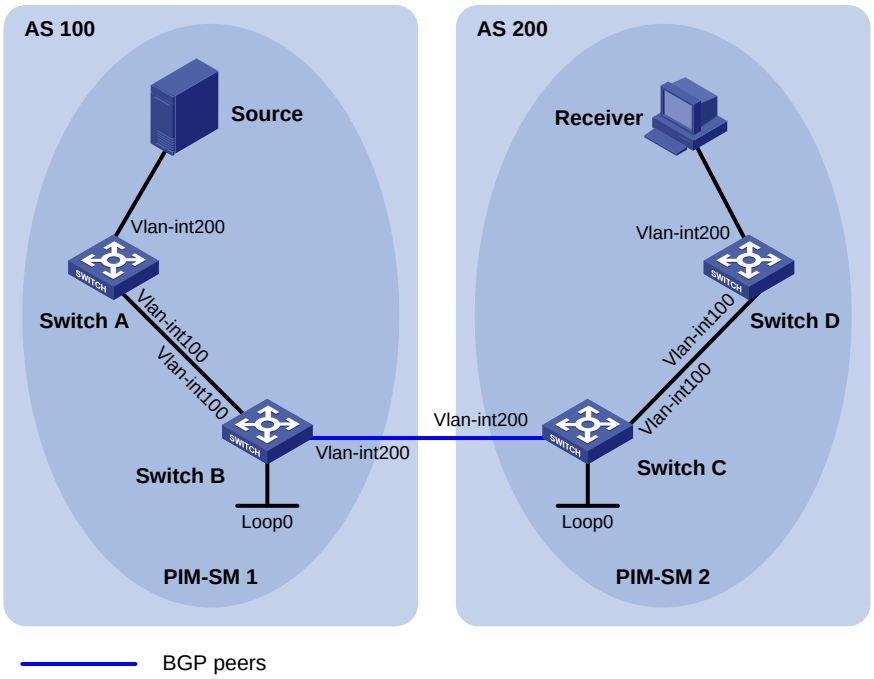


Table 2 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source	N/A	10.110.2.100/24	Switch C	Vlan-int100	10.110.4.1/24
Switch A	Vlan-int100	10.110.1.2/24	Switch C	Vlan-int200	192.168.1.2/24
Switch A	Vlan-int200	10.110.2.1/24	Switch C	Vlan-int100	10.110.4.2/24
Switch B	Loop0	1.1.1.1/32	Switch D	Vlan-int200	10.110.5.1/24
Switch B	Vlan-int100	10.110.1.1/24	Switch D	Loop0	2.2.2.2/32
Switch B	Vlan-int200	192.168.1.1/24			

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 6, and configure OSPF on the switches.

Procedure

1. Enable IP multicast routing, PIM-SM, and IGMP, and configure PIM-SM domain borders:

   # On Switch C, enable IP multicast routing, and enable PIM-SM on interfaces.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] pim sm
[SwitchA-Vlan-interface200] quit
[SwitchA] interface LoopBack0
[SwitchA-LoopBack0] pim sm
[SwitchA-LoopBack0] quit
```

On Switch C, enable IP multicast routing, enable PIM-SM on interfaces, and configure the receiver-side interface as a PIM-SM domain border.

```
<SwitchC> system-view
[SwitchC] multicast routing
[SwitchC-mrib] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] pim sm
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] pim sm
[SwitchC-Vlan-interface200] pim bsr-boundary
[SwitchC-Vlan-interface200] quit
```

On Switch D, enable IP multicast routing, enable IGMP on the receiver-side interface, and enable PIM-SM on other interfaces.

```
<SwitchD> system-view
[SwitchD] multicast routing
[SwitchD-mrib] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] pim sm
[SwitchD-Vlan-interface100] quit
[SwitchD] interface vlan-interface 200
[SwitchD-Vlan-interface200] igmp enable
[SwitchD-Vlan-interface200] quit
[SwitchD] interface LoopBack0
[SwitchD-LoopBack0] pim sm
[SwitchD-LoopBack0] quit
```

2. Configure C-BSRs and C-RPs:

On Switch A, configure Loopback 0 as a C-BSR and a C-RP.

```
[SwitchA] pim
[SwitchA-pim] c-bsr 1.1.1.1
[SwitchA-pim] c-rp 1.1.1.1
[SwitchA-pim] quit
```

On Switch D, configure Loopback 0 as a C-BSR and a C-RP.

```
[SwitchD] pim
[SwitchD-pim] c-bsr 2.2.2.2
[SwitchD-pim] c-rp 2.2.2.2
[SwitchD-pim] quit
```

3. Configure BGP, and redistribute BGP routing information into OSPF and OSPF routing information into BGP:

On Switch B, configure an EBGP peer, and redistribute OSPF routing information.

```
[SwitchB] bgp 100
[SwitchB-bgp-default] router-id 1.1.1.1
[SwitchB-bgp-default] peer 192.168.1.2 as-number 200
[SwitchB-bgp-default] address-family ipv4 unicast
[SwitchB-bgp-default-ipv4] peer 192.168.1.2 enable
[SwitchB-bgp-default-ipv4] import-route direct
[SwitchB-bgp-default-ipv4] import-route ospf 1
[SwitchB-bgp-default-ipv4] quit
[SwitchB-bgp-default] quit
```

On Switch C, configure an EBGP peer, and redistribute OSPF routing information.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 2.2.2.2
[SwitchC-bgp-default] peer 192.168.1.1 as-number 100
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] peer 192.168.1.1 enable
[SwitchC-bgp-default-ipv4] import-route direct
[SwitchC-bgp-default-ipv4] import-route ospf 1
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] quit
```

On Switch B, redistribute BGP routing information into OSPF.

```
[SwitchB] ospf 1
[SwitchB-ospf-1] import-route bgp
[SwitchB-ospf-1] quit
```

On Switch C, redistribute BGP routing information into OSPF.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] import-route bgp
[SwitchC-ospf-1] quit
```

4. Configure MSDP peers and static RPF peers:

On Switch A, configure Switch D and Switch G as the MSDP peers and static RPF peers.

```
[SwitchA] ip prefix-list list-dg permit 10.110.0.0 16 greater-equal 16 less-equal 32
[SwitchA] msdp
[SwitchA-msdp] peer 10.110.4.2 connect-interface vlan-interface 100
[SwitchA-msdp] static-rpf-peer 10.110.4.2 rp-policy list-dg
[SwitchA-msdp] quit
```

On Switch D, configure Switch A as the MSDP peer and static RPF peer.

```
[SwitchD] ip prefix-list list-a permit 10.110.0.0 16 greater-equal 16 less-equal 32
[SwitchD] msdp
[SwitchD-msdp] peer 10.110.1.2 connect-interface vlan-interface 100
[SwitchD-msdp] static-rpf-peer 10.110.1.2 rp-policy list-a
[SwitchD-msdp] quit
```

Verifying the configuration

Display the BGP peering relationships on Switch A.

```
[SwitchA] display bgp peer
```

No information is output, because no BGP peering relationship has been established between Switch A and Switch D. This means that the unicast topology is not changed.

Display brief information about MSDP peers on Switch A.

```
[SwitchA] display msdp brief
Configured    Established  Listen       Connect      Shutdown     Disabled
1             1            0            0            0            0

Peer address     State       Up/Down time    AS         SA count    Reset count
10.110.4.2       Established 01:07:08        ?          8           0
```

Display brief information about MSDP peers on Switch D.

```
[SwitchD] display msdp brief
Configured    Established  Listen       Connect      Shutdown     Disabled
1             1            0            0            0            0

Peer address     State       Up/Down time    AS         SA count    Reset count
10.110.1.2       Established 01:07:09        ?          8           0
```

Example: Configuring Anycast RP

Network configuration

As shown in Figure 7, OSPF runs within the domain to provide unicast routes.

Configure the Anycast RP application so that the receiver-side DRs and the source-side DRs can initiate a join process to their respective RPs that are topologically closest to them.

Configure the router IDs of Switch B and Switch D as 1.1.1.1 and 2.2.2.2, respectively. Set up an MSDP peering relationship between Switch B and Switch D.

Figure 7 Network diagram

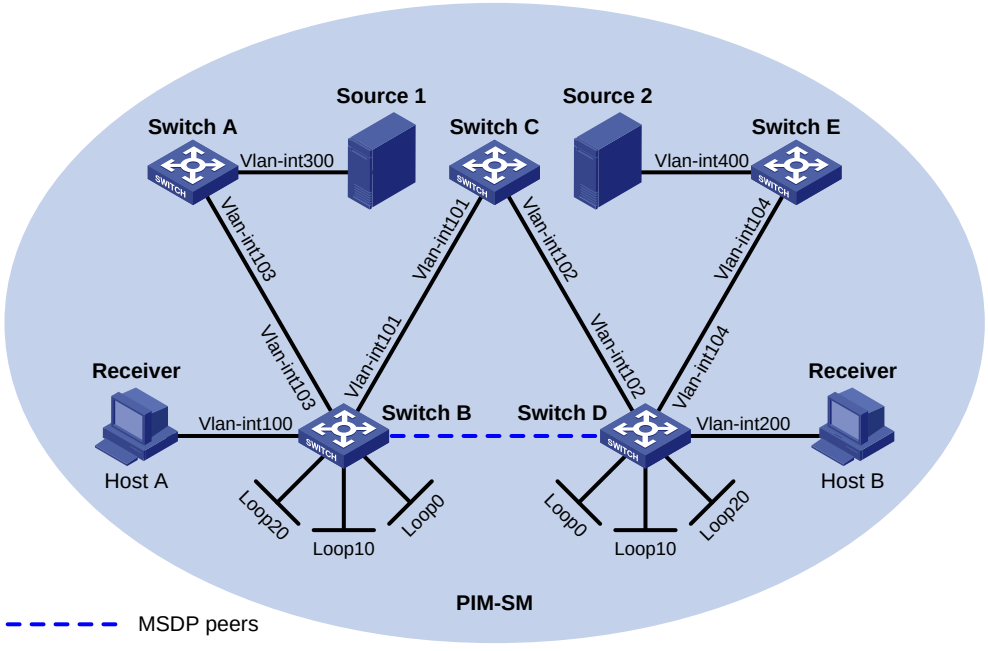


Table 3 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source 1	—	10.110.5.100/24	Switch C	Vlan-int101	192.168.1.2/24
Source 2	—	10.110.6.100/24	Switch C	Vlan-int102	192.168.2.2/24

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int300	10.110.5.1/24	Switch D	Vlan-int200	10.110.3.1/24
Switch A	Vlan-int103	10.110.2.2/24	Switch D	Vlan-int104	10.110.4.1/24
Switch B	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int102	192.168.2.1/24
Switch B	Vlan-int103	10.110.2.1/24	Switch D	Loop0	2.2.2.2/32
Switch B	Vlan-int101	192.168.1.1/24	Switch D	Loop10	4.4.4.4/32
Switch B	Loop0	1.1.1.1/32	Switch D	Loop20	10.1.1.1/32
Switch B	Loop10	3.3.3.3/32	Switch E	Vlan-int400	10.110.6.1/24
Switch B	Loop20	10.1.1.1/32	Switch E	Vlan-int104	10.110.4.2/24

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 7, and configure OSPF on the switches.

Procedure

1. Enable IP multicast routing, IGMP, and PIM-SM:

On Switch B, enable IP multicast routing.

```
<SwitchB> system-view
[SwitchB] multicast routing
[SwitchB-mrib] quit
```

Enable IGMP on the receiver-side interface (VLAN-interface 100).

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] igmp enable
[SwitchB-Vlan-interface100] quit
```

Enable PIM-SM on the other interfaces.

```
[SwitchB] interface vlan-interface 103
[SwitchB-Vlan-interface103] pim sm
[SwitchB-Vlan-interface103] quit
[SwitchB] interface Vlan-interface 101
[SwitchB-Vlan-interface101] pim sm
[SwitchB-Vlan-interface101] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] pim sm
[SwitchB-LoopBack0] quit
[SwitchB] interface loopback 10
[SwitchB-LoopBack10] pim sm
[SwitchB-LoopBack10] quit
[SwitchB] interface loopback 20
[SwitchB-LoopBack20] pim sm
[SwitchB-LoopBack20] quit
```

Enable IP multicast routing, IGMP, and PIM-SM on Switch A, Switch C, Switch D, and Switch E in the same way Switch B is configured. (Details not shown.)

2. Configure C-BSRs and C-RPs:

Configure Loopback 10 as a C-BSR and Loopback 20 as a C-RP on Switch B.

```
[SwitchB] pim
[SwitchB-pim] c-bsr 3.3.3.3
```

```
[SwitchB-pim] c-rp 10.1.1.1
[SwitchB-pim] quit
```

Configure a C-BSR and a C-RP on Switch D in the same way Switch B is configured. (Details not shown.)

3. Configure MSDP peers:

Configure an MSDP peer on Loopback 0 of Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] originating-rp loopback 0
[SwitchB-msdp] peer 2.2.2.2 connect-interface loopback 0
[SwitchB-msdp] quit
```

Configure an MSDP peer on Loopback 0 of Switch D.

```
[SwitchD] msdp
[SwitchD-msdp] originating-rp loopback 0
[SwitchD-msdp] peer 1.1.1.1 connect-interface loopback 0
[SwitchD-msdp] quit
```

Verifying the configuration

Display brief information about MSDP peers on Switch B.

```
[SwitchB] display msdp brief
Configured   Established  Listen       Connect      Shutdown     Disabled
1            1            0            0            0            0

Peer address    State       Up/Down time    AS         SA count   Reset count
2.2.2.2         Established 00:10:57        ?          0          0
```

Display brief information about MSDP peers on Switch D.

```
[SwitchD] display msdp brief
Configured   Established  Listen       Connect      Shutdown     Disabled
1            1            0            0            0            0

Peer address    State       Up/Down time    AS         SA count   Reset count
1.1.1.1         Established 00:10:57        ?          0          0
```

Send an IGMP report from Host A to join multicast group 225.1.1.1. (Details not shown.)

Send multicast data from Source 1 to multicast group 225.1.1.1. (Details not shown.)

Display the PIM routing table on Switch D.

```
[SwitchD] display pim routing-table
```

No information is output on Switch D.

Display the PIM routing table on Switch B.

```
[SwitchB] display pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:15:04
     Upstream interface: Register-Tunnel0
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
```

```
    Downstream interface(s) information:
    Total number of downstream interfaces: 1
        1: Vlan-interface100
            Protocol: igmp, UpTime: 00:15:04, Expires: -

 (10.110.5.100, 225.1.1.1)
    RP: 10.1.1.1 (local)
    Protocol: pim-sm, Flag: SPT 2MSDP ACT
    UpTime: 00:46:28
    Upstream interface: Vlan-interface103
        Upstream neighbor: 10.110.2.2
        RPF prime neighbor: 10.110.2.2
    Downstream interface(s) information:
    Total number of downstream interfaces: 1
        1: Vlan-interface100
            Protocol: pim-sm, UpTime:  - , Expires:  -
```

The output shows that Switch B now acts as the RP for Source 1 and Host A.

Send an IGMP leave message from Host A to leave multicast group 225.1.1.1. (Details not shown.),

Send an IGMP report from Host B to join multicast group 225.1.1.1. (Details not shown.)

Send multicast data from Source 2 to multicast group 225.1.1.1. (Details not shown.)

Display the PIM routing table on Switch B.

```
[SwitchB] display pim routing-table
```

No information is output on Switch B.

Display the PIM routing table on Switch D.

```
[SwitchD] display pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
    RP: 10.1.1.1 (local)
    Protocol: pim-sm, Flag: WC
    UpTime: 00:12:07
    Upstream interface: Register-Tunnel0
        Upstream neighbor: NULL
        RPF prime neighbor: NULL
    Downstream interface(s) information:
    Total number of downstream interfaces: 1
        1: Vlan-interface200
            Protocol: igmp, UpTime: 00:12:07, Expires: -

 (10.110.6.100, 225.1.1.1)
    RP: 10.1.1.1 (local)
    Protocol: pim-sm, Flag: SPT 2MSDP ACT
    UpTime: 00:40:22
    Upstream interface: Vlan-interface104
        Upstream neighbor: 10.110.4.2
        RPF prime neighbor: 10.110.4.2
```

```
Downstream interface(s) information:
Total number of downstream interfaces: 1
    1: Vlan-interface200
        Protocol: pim-sm, UpTime:  - , Expires:  -
```

The output shows that Switch D now acts as the RP for Source 2 and Host B.

Example: Configuring SA message filtering

Network configuration

As shown in Figure 8:

- OSPF runs within and among the PIM-SM domains to provide unicast routing.
- Set up an MSDP peering relationship between Switch A and Switch C and between Switch C and Switch D.
- Source 1 sends multicast data to multicast groups 225.1.1.0/30 and 226.1.1.0/30, and Source 2 sends multicast data to the multicast group 227.1.1.0/30.

Configure SA message policies so that:

- Host A and Host B receive the multicast data only addressed to multicast groups 225.1.1.0/30 and 226.1.1.0/30.
- Host C receives the multicast data only addressed to multicast groups 226.1.1.0/30 and 227.1.1.0/30.

Figure 8 Network diagram

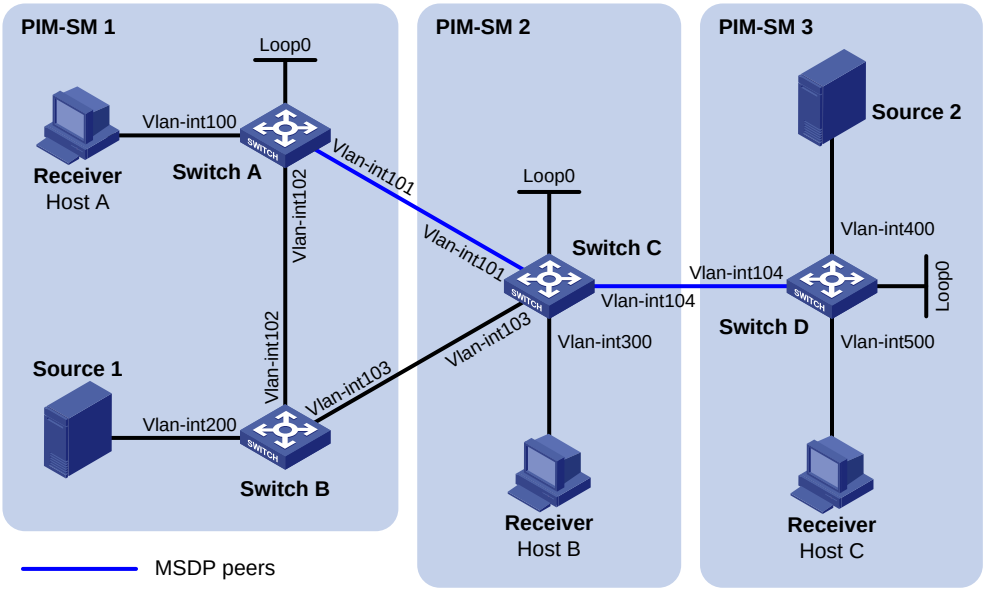


Table 4 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source 1	—	10.110.3.100/24	Switch C	Vlan-int300	10.110.4.1/24
Source 2	—	10.110.6.100/24	Switch C	Vlan-int104	10.110.5.1/24
Switch A	Vlan-int100	10.110.1.1/24	Switch C	Vlan-int101	192.168.1.2/24
Switch A	Vlan-int102	10.110.2.1/24	Switch C	Vlan-int103	192.168.2.2/24
Switch A	Vlan-int101	192.168.1.1/24	Switch C	Loop0	2.2.2.2/32

Device	Interface	IP address	Device	Interface	IP address
Switch A	Loop0	1.1.1.1/32	Switch D	Vlan-int400	10.110.6.1/24
Switch B	Vlan-int200	10.110.3.1/24	Switch D	Vlan-int500	10.110.7.1/24
Switch B	Vlan-int102	10.110.2.2/24	Switch D	Vlan-int104	10.110.5.2/24
Switch B	Vlan-int103	192.168.2.1/24	Switch D	Loop0	3.3.3.3/32

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 8, and configure OSPF on the switches in the PIM-SM domains.

Procedure

1. Enable IP multicast routing, IGMP, and PIM-SM, and configure PIM domain borders:

   # On Switch A, enable IP multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

   # Enable IGMP on the receiver-side interface (VLAN-interface 100).

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] igmp enable
   [SwitchA-Vlan-interface100] quit
   ```

   # Enable PIM-SM on the other interfaces.

   ```
   [SwitchA] interface vlan-interface 101
   [SwitchA-Vlan-interface101] pim sm
   [SwitchA-Vlan-interface101] quit
   [SwitchA] interface vlan-interface 102
   [SwitchA-Vlan-interface102] pim sm
   [SwitchA-Vlan-interface102] quit
   [SwitchA] interface loopback 0
   [SwitchA-LoopBack0] pim sm
   [SwitchA-LoopBack0] quit
   ```

   # Enable IP multicast routing, IGMP, and PIM-SM on Switch B, Switch C, and Switch D in the same way Switch A is configured. (Details not shown.)

   # Configure a PIM domain border on Switch C.

   ```
   [SwitchC] interface vlan-interface 101
   [SwitchC-Vlan-interface101] pim bsr-boundary
   [SwitchC-Vlan-interface101] quit
   [SwitchC] interface vlan-interface 103
   [SwitchC-Vlan-interface103] pim bsr-boundary
   [SwitchC-Vlan-interface103] quit
   [SwitchC] interface vlan-interface 104
   [SwitchC-Vlan-interface104] pim bsr-boundary
   [SwitchC-Vlan-interface104] quit
   ```

   # Configure PIM domain borders on Switch A, Switch B, and Switch D in the same way Switch C is configured. (Details not shown.)

2. Configure C-BSRs and C-RPs:

   # Configure Loopback 0 as a C-BSR and a C-RP on Switch A.

   ```
   [SwitchA] pim
   ```

```
[SwitchA-pim] c-bsr 1.1.1.1
[SwitchA-pim] c-rp 1.1.1.1
[SwitchA-pim] quit
```

Configure C-BSRs and C-RPs on Switch C and Switch D in the same way Switch A is configured. (Details not shown.)

3. Configure MSDP peers:

Configure an MSDP peer on Switch A.

```
[SwitchA] msdp
[SwitchA-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchA-msdp] quit
```

Configure MSDP peers on Switch C.

```
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchC-msdp] peer 10.110.5.2 connect-interface vlan-interface 104
[SwitchC-msdp] quit
```

Configure an MSDP peer on Switch D.

```
[SwitchD] msdp
[SwitchD-msdp] peer 10.110.5.1 connect-interface vlan-interface 104
[SwitchD-msdp] quit
```

4. Configure SA message policies:

Configure an SA accepting and forwarding policy on Switch C so that Switch C will not forward SA messages for (Source 1, 225.1.1.0/30) to Switch D.

```
[SwitchC] acl advanced 3001
[SwitchC-acl-ipv4-adv-3001] rule deny ip source 10.110.3.100 0 destination 225.1.1.0
0.0.0.3
[SwitchC-acl-ipv4-adv-3001] rule permit ip source any destination any
[SwitchC-acl-ipv4-adv-3001] quit
[SwitchC] msdp
[SwitchC-msdp] peer 10.110.5.2 sa-policy export acl 3001
[SwitchC-msdp] quit
```

Configure an SA creation policy on Switch D so that Switch D will not create SA messages for Source 2.

```
[SwitchD] acl basic 2001
[SwitchD-acl-ipv4-basic-2001] rule deny source 10.110.6.100 0
[SwitchD-acl-ipv4-basic-2001] quit
[SwitchD] msdp
[SwitchD-msdp] import-source acl 2001
[SwitchD-msdp] quit
```

Verifying the configuration

Display the (S, G) entries in the SA message cache on Switch C.

```
[SwitchC] display msdp sa-cache
 MSDP Total Source-Active Cache - 8 entries
 Matched 8 entries

Source        Group          Origin RP       Pro  AS     Uptime   Expires
10.110.3.100  225.1.1.0      1.1.1.1         ?    ?      02:03:30 00:05:31
10.110.3.100  225.1.1.1      1.1.1.1         ?    ?      02:03:30 00:05:31
10.110.3.100  225.1.1.2      1.1.1.1         ?    ?      02:03:30 00:05:31
```

```
10.110.3.100  225.1.1.3      1.1.1.1         ?    ?      02:03:30 00:05:31
10.110.3.100  226.1.1.0      1.1.1.1         ?    ?      02:03:30 00:05:31
10.110.3.100  226.1.1.1      1.1.1.1         ?    ?      02:03:30 00:05:31
10.110.3.100  226.1.1.2      1.1.1.1         ?    ?      02:03:30 00:05:31
10.110.3.100  226.1.1.3      1.1.1.1         ?    ?      02:03:30 00:05:31
```

Display the (S, G) entries in the SA message cache on Switch D.

```
[SwitchD] display msdp sa-cache
 MSDP Total Source-Active Cache - 4 entries
 Matched  4 entries

Source        Group          Origin RP       Pro  AS     Uptime   Expires
10.110.3.100  226.1.1.0      1.1.1.1         ?    ?      00:32:53 00:05:07
10.110.3.100  226.1.1.1      1.1.1.1         ?    ?      00:32:53 00:05:07
10.110.3.100  226.1.1.2      1.1.1.1         ?    ?      00:32:53 00:05:07
10.110.3.100  226.1.1.3      1.1.1.1         ?    ?      00:32:53 00:05:07
```

Troubleshooting MSDP

This section describes common MSDP problems and how to troubleshoot them.

MSDP peers stay in disabled state

Symptom

The configured MSDP peers stay in disabled state.

Solution

To resolve the problem:

1. Use the **display ip routing-table** command to verify that the unicast route between the devices is reachable.
2. Verify that a unicast route is available between the two devices that will become MSDP peers to each other.
3. Use the **display current-configuration** command to verify that the local interface address and the MSDP peer address of the remote device are the same.
4. If the problem persists, contact H3C Support.

No SA entries exist in the device's SA message cache

Symptom

MSDP fails to send (S, G) entries through SA messages.

Solution

To resolve the problem:

1. Use the **display ip routing-table** command to verify that the unicast route between the devices is reachable.
2. Verify that a unicast route is available between the two devices that will become MSDP peers to each other.
3. Verify the configuration of the **import-source** command and its *ipv4-acl-number* argument, and make sure the ACL rule filters appropriate (S, G) entries.

4. If the problem persists, contact H3C Support.

No exchange of locally registered (S, G) entries between RPs

Symptom

RPs fail to exchange their locally registered (S, G) entries with one another in the Anycast RP application.

Solution

To resolve the problem:

1. Use the **display ip routing-table** command to verify that the unicast route between the devices is reachable.
2. Verify that a unicast route is available between the two devices that will establish an MSDP peering relationship.
3. Verify the configuration of the **originating-rp** command. In the Anycast RP application environment, use the **originating-rp** command to configure the RP address in the SA messages, which must be the local interface address.
4. Verify that the C-BSR address is different from the Anycast RP address.
5. If the problem persists, contact H3C Support.

Contents

Configuring MLD snooping ····· 1

About MLD snooping ····· 1
Fundamentals of MLD snooping ····· 1
MLD snooping ports ····· 1
How MLD snooping works ····· 3
MLD snooping proxying ····· 4
Protocols and standards ····· 5
Restrictions and guidelines: MLD snooping configuration ····· 5
VLAN-based MLD snooping tasks at a glance ····· 6
VSI-based MLD snooping tasks at a glance ····· 7
Enabling the MLD snooping feature ····· 7
Enabling MLD snooping ····· 8
Enabling MLD snooping globally ····· 8
Enabling MLD snooping for VLANs ····· 8
Enabling MLD snooping for a VSI ····· 9
Configuring basic MLD snooping features ····· 9
Specifying an MLD snooping version ····· 9
Setting the maximum number of MLD snooping forwarding entries ····· 10
Configuring static IPv6 multicast MAC address entries ····· 11
Setting the MLD last listener query interval ····· 11
Configuring MLD snooping port features ····· 12
Setting aging timers for dynamic ports ····· 12
Configuring a static member port ····· 14
Configuring a static router port ····· 14
Configuring a port as a simulated member host ····· 14
Enabling fast-leave processing ····· 15
Disabling a port from becoming a dynamic router port ····· 16
Preventing the device from forwarding multicast data to router ports in a VLAN ····· 16
Specifying the Layer 2 IPv6 multicast forwarding mode ····· 17
About Layer 2 IPv6 multicast forwarding modes ····· 17
Restrictions and guidelines for specifying the Layer 2 IPv6 multicast forwarding mode ····· 17
Specifying the Layer 2 IPv6 multicast forwarding mode for a VLAN ····· 17
Configuring the MLD snooping querier ····· 18
Enabling the MLD snooping querier ····· 18
Enabling MLD snooping querier election ····· 18
Configuring parameters for MLD general queries and responses ····· 19
Enabling the device to send MLD general queries upon a path change ····· 20
Enabling the device to send MLD general queries upon a port state change ····· 21
Enabling MLD snooping proxying ····· 22
About MLD snooping proxying ····· 22
Restrictions and guidelines for enabling MLD snooping proxying ····· 22
Enabling MLD snooping proxying for a VLAN ····· 22
Enabling MLD snooping proxying for a VSI ····· 22
Configuring parameters for MLD messages ····· 23
Configuring source IPv6 addresses for MLD messages ····· 23
Setting the 802.1p priority for MLD messages ····· 24
Configuring MLD snooping policies ····· 25
Configuring an IPv6 multicast group policy ····· 25
Enabling dropping unknown IPv6 multicast data ····· 25
Enabling MLD report suppression ····· 26
Setting the maximum number of IPv6 multicast groups on a port ····· 27
Enabling IPv6 multicast group replacement ····· 27
Enabling host tracking ····· 28
Setting the DSCP value for outgoing MLD protocol packets ····· 29
Enabling SNMP notifications for MLD snooping ····· 29
Configuring MLD snooping logging ····· 29
Enabling MLD snooping logging ····· 29

Enabling MLD snooping logging suppression ····· 30
Display and maintenance commands for MLD snooping ····· 30
MLD snooping configuration examples ····· 32
Example: Configuring VLAN-based IPv6 group policy and simulated joining ····· 32
Example: Configuring VLAN-based static ports ····· 34
Example: Configuring the VLAN-based MLD snooping querier ····· 37
Example: Configuring VLAN-based MLD snooping proxying ····· 40
Troubleshooting MLD snooping ····· 42
Layer 2 multicast forwarding cannot function ····· 42
IPv6 multicast group policy does not work ····· 42

Configuring MLD snooping

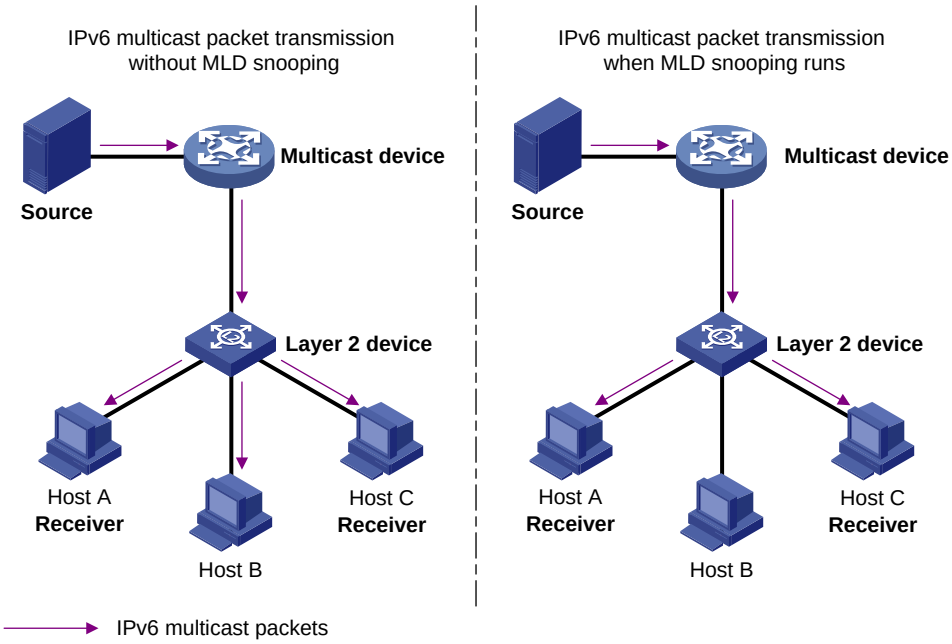
About MLD snooping

MLD snooping runs on a Layer 2 device as an IPv6 multicast constraining mechanism to improve multicast forwarding efficiency. It creates Layer 2 multicast forwarding entries from MLD messages that are exchanged between the hosts and the Layer 3 device.

Fundamentals of MLD snooping

As shown in Figure 1, when MLD snooping is not enabled, the Layer 2 switch floods IPv6 multicast packets all hosts in a VLAN or VSI. When MLD snooping is enabled, the Layer 2 switch forwards multicast packets of known IPv6 multicast groups to only the receivers.

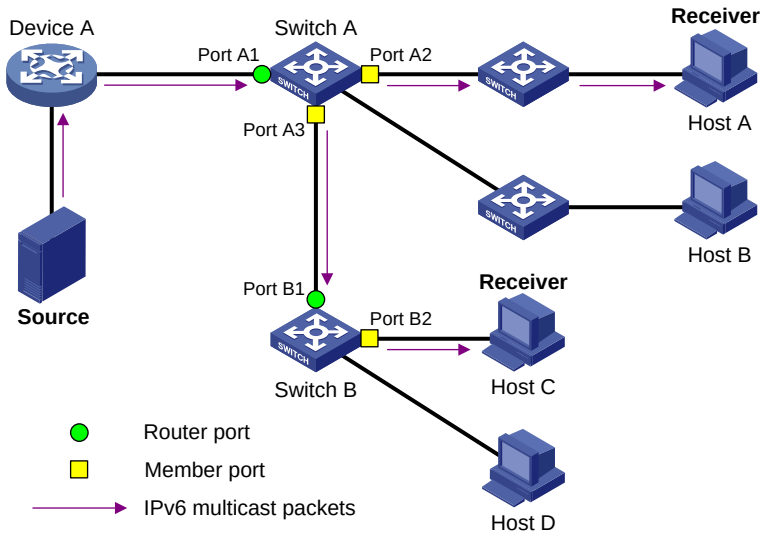
Figure 1 Multicast packet transmission processes without and with MLD snooping



MLD snooping ports

As shown in Figure 2, MLD snooping runs on Switch A and Switch B, and Host A and Host C are receiver hosts in an IPv6 multicast group. MLD snooping ports are divided into member ports and router ports.

Figure 2 MLD snooping ports



Router ports

On an MLD snooping Layer 2 device, the ports toward Layer 3 multicast devices are called router ports. In Figure 2, Port A1 of Switch A and Port B1 of Switch B are router ports.

Router ports contain the following types:

- **Dynamic router port**—When a port receives an MLD general query whose source address is not 0::0 or receives an IPv6 PIM hello message, the port is added into the dynamic router port list. At the same time, an aging timer is started for the port. If the port receives an MLD general query whose source address is not 0::0 or an IPv6 PIM hello message before the timer expires, the timer is reset. If the port does not receive either of the messages when the timer expires, the port is removed from the dynamic router port list.
- **Static router port**—When a port is statically configured as a router port, it is added into the static router port list. The static router port does not age out, and it can be deleted only manually.

Do not confuse the "router port" in MLD snooping with the "routed interface" commonly known as the "Layer 3 interface." The router port in MLD snooping is a Layer 2 interface.

Member ports

On an MLD snooping Layer 2 device, the ports toward receiver hosts are called member ports. In Figure 2, Port A2 and Port A3 of Switch A and Port B2 of Switch B are member ports.

Member ports contain the following types:

- **Dynamic member port**—When a port receives an MLD report, it is added to the associated dynamic MLD snooping forwarding entry as an outgoing interface. At the same time, an aging timer is started for the port. If the port receives an MLD report before the timer expires, the timer is reset. If the port does not receive an MLD report when the timer expires, the port is removed from the associated dynamic forwarding entry.
- **Static member port**—When a port is statically configured as a member port, it is added to the associated static MLD snooping forwarding entry as an outgoing interface. The static member port does not age out, and it can be deleted only manually.

Unless otherwise specified, router ports and member ports in this document include both static and dynamic router ports and member ports.

How MLD snooping works

The ports in this section are dynamic ports. For information about how to configure and remove static ports, see "Configuring a static member port" and "Configuring a static router port."

MLD messages include general query, MLD report, and done message. An MLD snooping-enabled Layer 2 device performs differently depending on the MLD message types.

General query

The MLD querier periodically sends MLD general queries to all hosts and devices on the local subnet to check for the existence of IPv6 multicast group members.

After receiving an MLD general query, the Layer 2 device forwards the query to all ports in the VLAN or VSI except the receiving port. The Layer 2 device also performs one of the following actions:

- If the receiving port is a dynamic router port in the dynamic router port list, the Layer 2 device restarts the aging timer for the router port.
- If the receiving port does not exist in the dynamic router port list, the Layer 2 device adds the port to the dynamic router port list. It also starts an aging timer for the port.

MLD report

A host sends an MLD report to the MLD querier for the following purposes:

- Responds to queries if the host is an IPv6 multicast group member.
- Applies for an IPv6 multicast group membership.

After receiving an MLD report from a host, the Layer 2 device forwards the report through all the router ports in the VLAN or VSI. It also resolves the IPv6 address of the reported IPv6 multicast group, and looks up the forwarding table for a matching entry as follows:

- If no match is found, the Layer 2 device creates a forwarding entry for the group with the receiving port an outgoing interface. It also marks the receiving port as a dynamic member port and starts an aging timer for the port.
- If a match is found but the matching forwarding entry does not contain the receiving port, the Layer 2 device adds the receiving port to the outgoing interface list. It also marks the port as a dynamic member port to the forwarding entry and starts an aging timer for the port.
- If a match is found and the matching forwarding entry contains the receiving port, the Layer 2 device restarts the aging timer for the port.

NOTE:

A Layer 2 device does not forward an MLD report through a non-router port because of the host MLD report suppression mechanism. If a non-router port has member host attached, the member hosts suppress their MLD reports upon receiving MLD reports forwarded by the non-router port. The Layer 2 device cannot know the existence of the member hosts attached to the non-router port. For more information about the MLD report suppression mechanism, see "Configuring MLD."

Done message

When a host leaves an IPv6 multicast group, the host sends an MLD done message to the Layer 3 devices. When the Layer 2 device receives the MLD done message on a dynamic member port, the Layer 2 device first examines whether a forwarding entry matches the IPv6 multicast group address in the message.

- If no match is found, the Layer 2 device discards the MLD done message.
- If a match is found but the receiving port is not an outgoing interface in the forwarding entry, the Layer 2 device discards the MLD done message.
- If a match is found and the receiving port is not the only outgoing interface in the forwarding entry, the Layer 2 device performs the following actions:

- Discards the MLD done message.
- Sends an MLD multicast-address-specific query to identify whether the group has active listeners attached to the receiving port.
- Sets the aging timer for the receiving port to twice the MLD last listener query interval.

- If a match is found and the receiving port is the only outgoing interface in the forwarding entry, the Layer 2 device performs the following actions:
  - Forwards the MLD done message to all router ports in the VLAN or VSI.
  - Sends an MLD multicast-address-specific query to identify whether the group has active listeners attached to the receiving port.
  - Sets the aging timer for the receiving port to twice the MLD last listener query interval.

After receiving the MLD done message on a port, the MLD querier resolves the IPv6 multicast group address in the message. Then, it sends an MLD multicast-address-specific query to the IPv6 multicast group through the receiving port.

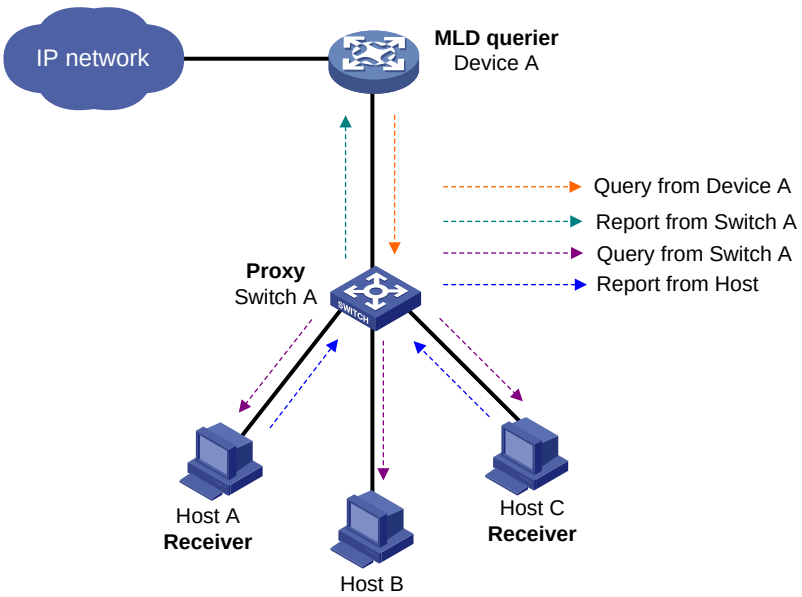
After receiving the MLD multicast-address-specific query, the Layer 2 device forwards the query through all router ports and member ports of the group in the VLAN or VSI. Then, it waits for the responding MLD report from the directly connected hosts. For the dynamic member port that received the done message, the Layer 2 device also performs one of the following actions:

- If the port receives an MLD report before the aging timer expires, the Layer 2 device resets the aging timer for the port.
- If the port does not receive any MLD report messages when the aging timer expires, the Layer 2 device removes the port from the forwarding entry for the IPv6 multicast group.

MLD snooping proxying

As shown in Figure 3, to reduce the number of MLD report and done messages received by the upstream device, you can enable MLD snooping proxying on the edge device. With MLD snooping proxying enabled, the edge device acts as a host for the upstream MLD snooping querier to send MLD report and done messages to Device A. The host MLD report suppression mechanism on the edge device does not take effect. For more information about the MLD report suppression mechanism, see "Configuring MLD."

Figure 3 MLD snooping proxying



The MLD snooping proxy device processes different MLD messages as follows:

- General query.

  After receiving an MLD general query, the device forwards the query to all ports in the VLAN or VSI except the receiving port. The device also generates an MLD report based on the local membership information and sends the report to all router ports.

- Multicast-address-specific query or multicast-address-and-source-specific query.

  After receiving an MLD multicast-address-specific query or multicast-address-and-source-specific query, the device forwards the query to all ports in the VLAN or VSI except the receiving port. If the forwarding entry has a member port, the device sends a response to all router ports in the VLAN or VSI.

- Report.

  After receiving an MLD report from a host, the device looks up the forwarding table for a matching entry as follows:

  - If a match is found and the matching forwarding entry contains the receiving port, the device resets the aging timer for the port.
  - If a match is found but the matching forwarding entry does not contain the receiving port, the device adds the receiving port to the outgoing interface list. It also marks the receiving port as a dynamic member port and starts an aging timer for the port.
  - If no match is found, the device creates a forwarding entry with the receiving port as an outgoing interface. It also marks the receiving port as a dynamic member port and starts an aging timer for the port. Then it sends the report to all router ports.

- Done message.

  After receiving the MLD done message on a port, the device sends an MLD multicast-address-specific query through the receiving port. The device sends the MLD done message to all router ports only when the last member port is removed from the forwarding entry.

Protocols and standards

RFC 4541, *Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches*

Restrictions and guidelines: MLD snooping configuration

For MLD reports received from secondary VLANs, the relevant MLD snooping forwarding entries are maintained by the primary VLAN. Therefore, you need to enable MLD snooping only for the primary VLAN. The configuration made in secondary VLANs will not take effect..

The MLD snooping configurations made on Layer 2 aggregate interfaces do not interfere with the configurations made on member ports. In addition, the configurations made on Layer 2 aggregate interfaces do not take part in aggregation calculations. The configuration made on a member port of the aggregate group takes effect after the port leaves the aggregate group.

Some features can be configured for a VLAN in VLAN view or for multiple VLANs in MLD-snooping view. The VLAN-specific configuration and the configuration made in MLD-snooping view have the same priority, and the most recent configuration takes effect.

Some features can be configured for a VLAN in VLAN view or globally for all VLANs in MLD-snooping view. The VLAN-specific configuration takes priority over the global configuration.

Some features can be configured for a VSI in VSI view or globally for all VSIs in MLD-snooping view. The VSI-specific configuration takes priority over the global configuration.

Some features can be configured for an interface in interface view or for all interfaces of the specified VLANs in MLD-snooping view. The interface-specific configuration takes priority over the configuration made in MLD-snooping view.

VLAN-based MLD snooping tasks at a glance

To configure MLD snooping for VLANs, perform the following tasks:

1. Enabling the MLD snooping feature
2. Enabling MLD snooping

   Choose the following tasks as needed:
   - Enabling MLD snooping globally
   - Enabling MLD snooping for VLANs
3. (Optional.) Configuring basic MLD snooping features
   - Specifying an MLD snooping version
   - Setting the maximum number of MLD snooping forwarding entries
   - Configuring static IPv6 multicast MAC address entries
   - Setting the MLD last listener query interval
4. (Optional.) Configuring MLD snooping port features
   - Setting aging timers for dynamic ports
   - Configuring a static member port
   - Configuring a static router port
   - Configuring a port as a simulated member host
   - Enabling fast-leave processing
   - Disabling a port from becoming a dynamic router port
   - Preventing the device from forwarding multicast data to router ports in a VLAN
5. (Optional.) Specifying the Layer 2 IPv6 multicast forwarding mode
6. (Optional.) Configuring the MLD snooping querier
   - Enabling the MLD snooping querier
   - Enabling MLD snooping querier election
   - Configuring parameters for MLD general queries and responses
   - Enabling the device to send MLD general queries upon a path change
   - Enabling the device to send MLD general queries upon a port state change
7. (Optional.) Enabling MLD snooping proxying
8. (Optional.) Configuring parameters for MLD messages
   - Configuring source IPv6 addresses for MLD messages
   - Setting the 802.1p priority for MLD messages
9. (Optional.) Configuring MLD snooping policies
   - Configuring an IPv6 multicast group policy
   - Enabling dropping unknown IPv6 multicast data
   - Enabling MLD report suppression
   - Setting the maximum number of IPv6 multicast groups on a port
   - Enabling IPv6 multicast group replacement
   - Enabling host tracking
10. (Optional.) Setting the DSCP value for outgoing MLD protocol packets

11. (Optional.) Enabling SNMP notifications for MLD snooping
12. (Optional.) Configuring MLD snooping logging
    - Enabling MLD snooping logging
    - Enabling MLD snooping logging suppression

VSI-based MLD snooping tasks at a glance

To configure MLD snooping for VSIs, perform the following tasks:

1. Enabling the MLD snooping feature
2. Enabling MLD snooping

   Choose the following tasks as needed:
   - Enabling MLD snooping globally
   - Enabling MLD snooping for a VSI
3. (Optional.) Configuring basic MLD snooping features
   - Specifying an MLD snooping version
   - Setting the maximum number of MLD snooping forwarding entries
   - Setting the MLD last listener query interval
4. (Optional.) Setting aging timers for dynamic ports
5. (Optional.) Configuring the MLD snooping querier
   - Enabling the MLD snooping querier
   - Enabling MLD snooping querier election
   - Configuring parameters for MLD general queries and responses
   - Enabling the device to send MLD general queries upon a path change
6. (Optional.) Enabling MLD snooping proxying
7. (Optional.) Configuring parameters for MLD messages
   - Configuring source IPv6 addresses for MLD messages
   - Setting the 802.1p priority for MLD messages
8. (Optional.) Configuring MLD snooping policies
   - Enabling dropping unknown IPv6 multicast data
   - Enabling MLD report suppression
9. (Optional.) Setting the DSCP value for outgoing MLD protocol packets
10. (Optional.) Enabling SNMP notifications for MLD snooping
11. (Optional.) Configuring MLD snooping logging
    - Enabling MLD snooping logging
    - Enabling MLD snooping logging suppression

Enabling the MLD snooping feature

About this task

You must enable the MLD snooping feature before you configure other MLD snooping features.

Procedure

1. Enter system view.

   `system-view`

2. Enable the MLD snooping feature and enter MLD-snooping view.

   mld-snooping

   By default, the MLD snooping feature is disabled.

Enabling MLD snooping

Enabling MLD snooping globally

About this task

After you enable MLD snooping globally, MLD snooping is enabled for all VLANs or VSIs. You can disable MLD snooping for a VLAN or VSI when MLD snooping is globally enabled.

Restrictions and guidelines

To configure other MLD snooping features for VLANs or VSIs, you must enable MLD snooping for the specific VLANs or VSIs even though MLD snooping is enabled globally.

The VLAN-specific or VSI-specific MLD snooping configuration takes priority over the global MLD snooping configuration. For example, if you enable MLD snooping globally and then use the **mld-snooping disable** command to disable MLD snooping for a VLAN, MLD snooping is disabled in the VLAN.

Procedure

1. Enter system view.

   system-view

2. Enter MLD-snooping view.

   mld-snooping

3. Enable MLD snooping globally.
4. **global-enable**

   By default, MLD snooping is disabled globally.

5. (Optional.) Disable MLD snooping for a VLAN or VSI.
   a. Return to system view.

      quit

   b. Enter VLAN or VSI view.

      vlan *vlan-id*

      vsi *vsi-name*

   c. Disable MLD snooping for the VLAN or VSI.

      mld-snooping disable

      By default, the MLD snooping status in a VLAN or VSI is consistent with the global MLD snooping status.

Enabling MLD snooping for VLANs

Restrictions and guidelines

You can enable MLD snooping for multiple VLANs by using the **enable vlan** command in MLD-snooping view or for a VLAN by using the **mld-snooping enable** command in VLAN view. The configuration in VLAN view has the same priority as the configuration in MLD-snooping view.

MLD snooping configuration in a VLAN takes effect only on the member ports in the VLAN.

Enabling MLD snooping for multiple VLANs

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Enable MLD snooping for multiple VLANs.
   enable vlan *vlan-list*
   By default, the MLD snooping status in a VLAN is consistent with the global MLD snooping status.

Enabling MLD snooping for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable MLD snooping for the VLAN.
   mld-snooping enable
   By default, the MLD snooping status in a VLAN is consistent with the global MLD snooping status.

Enabling MLD snooping for a VSI

Restrictions and guidelines

For VSIs, you can only enable MLD snooping for a VSI by using the **mld-snooping enable** command in VSI view.

MLD snooping configuration in a VSI takes effect only on the member ports in the VSI.

On an EVPN M-LAG network, you must enable both MLD snooping and MLD snooping proxying for VSIs. For more information about EVPN M-LAG, see EVPN VXLAN configuration in *EVPN Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable MLD snooping for the VSI.
   mld-snooping enable
   By default, the MLD snooping status in a VSI is consistent with the global MLD snooping status.

Configuring basic MLD snooping features

Specifying an MLD snooping version

About this task

Different MLD snooping versions can process different versions of MLD messages:

- MLDv1 snooping can process MLDv1 messages and MLDv2 queries, but it floods MLDv2 reports in the VLAN or VSI instead of processing them.
- MLDv2 snooping can process MLDv1 and MLDv2 messages.

Restrictions and guidelines

If you change the version of MLD snooping from 2 to 1, the system performs the following actions:

- Clears all MLD snooping forwarding entries that are dynamically created.
- Keeps static MLDv2 snooping forwarding entries (*, G).
- Clears static MLDv2 snooping forwarding entries (S, G), which will be restored when MLD snooping is switched back to MLDv2 snooping.

For more information about static MLD snooping forwarding entries, see "Configuring a static member port."

Specifying an MLD snooping version for multiple VLANs

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Specify an MLD snooping version for multiple VLANs.
   version *version-number* **vlan** *vlan-list*
   By default, the MLD snooping version for a VLAN is 1.

Specifying an MLD snooping version for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Specify an MLD snooping version for the VLAN.
   mld-snooping version *version-number*
   By default, the MLD snooping version for a VLAN is 1.

Specifying an MLD snooping version for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Specify an MLD snooping version for the VSI.
   mld-snooping version *version-number*
   By default, the MLD snooping version for a VSI is 1.

Setting the maximum number of MLD snooping forwarding entries

About this task

You can modify the maximum number of MLD snooping forwarding entries, including dynamic entries and static entries. When the number of forwarding entries on the device reaches the upper

limit, the device does not automatically remove any existing entries. To allow new entries to be created, remove some entries manually.

Procedure

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Set the maximum number of MLD snooping forwarding entries.
   entry-limit *limit*
   By default, the maximum number of MLD snooping forwarding entries is 4294967295.

Configuring static IPv6 multicast MAC address entries

About this task

In Layer 2 IPv6 multicast, IPv6 multicast MAC address entries can be dynamically created through Layer 2 multicast protocols (such as MLD snooping). You can also manually configure static IPv6 multicast MAC address entries by binding IPv6 multicast MAC addresses and ports to control the destination ports of the IPv6 multicast data.

Restrictions and guidelines

You must specify an unused multicast MAC address when configuring a static IPv6 multicast MAC address entry. A multicast MAC address is the MAC address in which the least significant bit of the most significant octet is 1.

Configuring a static IPv6 multicast MAC address entry in system view

1. Enter system view.
   system-view
2. Configure a static IPv6 multicast MAC address entry.
   mac-address multicast *mac-address* **interface** *interface-list* **vlan** *vlan-id*

Configuring a static IPv6 multicast MAC address entry in interface view

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Configure a static IPv6 multicast MAC address entry.
   mac-address multicast *mac-address* **vlan** *vlan-id*

Setting the MLD last listener query interval

About this task

A receiver host starts a report delay timer for an IPv6 multicast group when it receives an MLD multicast-address-specific query for the group. This timer is set to a random value in the range of 0 to

the maximum response time advertised in the query. When the timer value decreases to 0, the host sends an MLD report to the group.

The MLD last listener query interval defines the maximum response time advertised in MLD multicast-address-specific queries. Set an appropriate value for the MLD last listener query interval to speed up hosts' responses to MLD multicast-address-specific queries and avoid MLD report traffic bursts.

Setting the MLD last listener query interval globally

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Set the MLD last listener query interval globally.
   last-listener-query-interval *interval*
   By default, the MLD last listener query interval is 1 second.

Setting the MLD last listener query interval for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Set the MLD last listener query interval for the VLAN
   mld-snooping last-listener-query-interval *interval*
   By default, the MLD last listener query interval is 1 second for a VLAN.

Setting the MLD last listener query interval for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Set the MLD last listener query interval for the VSI.
   mld-snooping last-listener-query-interval *interval*
   By default, the MLD last listener query interval is 1 second for a VSI.

Configuring MLD snooping port features

Setting aging timers for dynamic ports

About this task

A dynamic router port is removed from the dynamic router port list if it does not receive an MLD general query or IPv6 PIM hello message when its aging timer expires.

A dynamic member port is removed from the dynamic member port if it does not receive an MLD report when its aging timer expires.

Restrictions and guidelines

Set an appropriate value for the aging timers of dynamic ports based on the actual network requirement. For example, if the memberships of IPv6 multicast groups frequently change, set a relatively small value for the aging timer of the dynamic member ports.

If a dynamic router port receives an IPv6 PIMv2 hello message, the aging timer for the port is specified by the hello message. In this case, the **mld-snooping router-aging-time** command does not take effect on the port.

MLD multicast-address-specific queries originated by the Layer 2 device trigger the adjustment of aging timers of dynamic member ports. If a dynamic member port receives such a query, its aging timer is set to twice the MLD last listener query interval. For more information about setting the MLD last listener query interval on the Layer 2 device, see "Setting the MLD last listener query interval."

Setting the aging timers for dynamic ports globally

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Set the aging timer for dynamic router ports globally.
   router-aging-time *seconds*
   By default, the aging timer for dynamic router ports is 260 seconds.
4. Set the aging timer for dynamic member ports globally.
   host-aging-time *seconds*
   By default, the aging timer for dynamic member ports is 260 seconds.

Setting the aging timers for dynamic ports in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Set the aging timer for dynamic router ports in the VLAN.
   mld-snooping router-aging-time *seconds*
   By default, a VLAN uses the global aging timer for dynamic router ports.
4. Set the aging timer for dynamic member ports in the VLAN.
   mld-snooping host-aging-time *seconds*
   By default, a VLAN uses the global aging timer for dynamic member ports.

Setting the aging timers for dynamic ports in a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Set the aging timer for dynamic router ports in the VSI.
   mld-snooping router-aging-time *seconds*
   By default, a VSI uses the global aging timer for dynamic router ports.
4. Set the aging timer for dynamic member ports in the VSI.
   mld-snooping host-aging-time *seconds*
   By default, a VSI uses the global aging timer for dynamic member ports.

Configuring a static member port

About this task

You can configure a port as a static member port for an IPv6 multicast group so that all hosts attached to the port can always receive IPv6 multicast data for the group. The static member port does not respond to MLD queries. When you complete or cancel this configuration, the port does not send an unsolicited report or done message.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Configure the port as a static member port.
   mld-snooping static-group *ipv6-group-address* [**source-ip** *ipv6-source-address*] { **vlan** *vlan-id* | **service-instance** *instance-id* }
   By default, a port is not a static member port.

Configuring a static router port

About this task

You can configure a port as a static router port for an IPv6 multicast group so that all IPv6 multicast data for the group received on the port will be forwarded.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Configure the port as a static router port.
   mld-snooping static-router-port vlan *vlan-id*
   By default, a port is not a static router port.

Configuring a port as a simulated member host

About this task

When a port is configured as a simulated member host, it is equivalent to an independent host in the following ways:

- It sends an unsolicited MLD report when you complete the configuration.
- It responds to MLD general queries with MLD reports.

- It sends an MLD done message when you remove the configuration.

The version of MLD running on the simulated member host is the same as the version of MLD snooping running on the port. The port ages out in the same ways as a dynamic member port.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Configure the port as a simulated member host.
   mld-snooping host-join *ipv6-group-address* [**source-ip** *ipv6-source-address*] **vlan** *vlan-id*
   By default, the port is not a simulated member host.

Enabling fast-leave processing

About this task

This feature enables the Layer 2 device to immediately remove a port from the forwarding entry for an IPv6 multicast group when the port receives a done message. The device no longer sends or forwards MLD multicast-address-specific queries for the group to the port.

Restrictions and guidelines

Do not enable fast-leave processing on a port that has multiple receiver hosts attached in a VLAN. If you do so, the remaining receivers cannot receive IPv6 multicast data for a group after a receiver leaves the group.

As a best practice to avoid multicast service interruption on other ports in the same VLAN, enable only MLD snooping fast-leave processing on a port on a network configured with both MLD snooping and MLD. Do not use the **mld fast-leave** command to enable MLD fast-leaving processing.

Enabling fast-leave processing globally

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Enable fast-leave processing globally.
   fast-leave [**vlan** *vlan-list*]
   By default, fast-leave processing is disabled globally.

Enabling fast-leave processing on a port

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.

interface bridge-aggregation *interface-number*

3. Enable fast-leave processing on the port.

   mld-snooping fast-leave [**vlan** *vlan-list*]

   By default, fast-leave processing is disabled on a port.

Disabling a port from becoming a dynamic router port

About this task

A receiver host might send MLD general queries or IPv6 PIM hello messages for testing purposes. On the Layer 2 device, the port that receives MLD general queries or IPv6 PIM hello messages becomes a dynamic router port. Before the aging timer for the port expires, the following problems might occur:

- All IPv6 multicast data for the VLAN to which the port belongs flows to the port. Then, the port forwards the data to attached receiver hosts. The receiver hosts will receive IPv6 multicast data that it does not expect.
- The port forwards the MLD general queries or IPv6 PIM hello messages to its upstream Layer 3 devices. These messages might affect the multicast routing protocol state (such as the MLD querier or DR election) on the Layer 3 devices. This might further cause network interruption.

To solve these problems, you can disable the port from becoming a dynamic router port when receiving MLD general queries or IPv6 PIM hello messages. This also improves network security and the control over receiver hosts.

Restrictions and guidelines

This configuration and the static router port configuration do not interfere with each other.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
3. Disable the port from becoming a dynamic router port.

   mld-snooping router-port-deny [**vlan** *vlan-list*]

   By default, a port is allowed to become a dynamic router port.

Preventing the device from forwarding multicast data to router ports in a VLAN

About this task

If all ports on an MLD snooping Layer 2 device are statically configured as router ports, the Layer 2 device will forwards multicast data from Layer 3 multicast devices to all its router ports. This wastes bandwidth resources. Perform this task to not forward multicast data to router ports in a VLAN if they are not interested in receiving the data.

Prerequisites

Before performing this task, you must enable MLD snooping in the VLAN.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Prevent the device from forwarding multicast data to static router ports in the VLAN.
   mld-snooping router-port-discard
   By default, the device forwards multicast data to static router ports in a VLAN.

Specifying the Layer 2 IPv6 multicast forwarding mode

About Layer 2 IPv6 multicast forwarding modes

The device supports the following Layer 2 IPv6 multicast forwarding modes:

- **MAC-based mode**—The device forwards Layer 2 IPv6 multicast packets based on the IPv6 multicast MAC addresses in the packets.
- **IP-based mode**—The device forwards Layer 2 IPv6 multicast packets based on the IPv6 addresses (IPv6 multicast source addresses and IPv6 multicast group addresses) in the packets.

To display Layer 2 IPv6 multicast IP forwarding entries, use the **display ipv6 l2-multicast ip forwarding** command.

To display Layer 2 IPv6 multicast MAC forwarding entries, use the **display ipv6 l2-multicast mac forwarding** command.

Restrictions and guidelines for specifying the Layer 2 IPv6 multicast forwarding mode

Multiple IPv6 multicast addresses might be mapped to the same IPv6 multicast MAC address. This causes that non-members of an IPv6 multicast group receive the multicast data of this IPv6 group if the device uses the MAC-based forwarding mode. As a best practice, specify the IP-based Layer 2 IPv6 multicast forwarding mode on Layer 3 multicast devices.

As a best practice, specify the IP-based Layer 2 IPv6 multicast forwarding mode on the device if a device running MLDv2 exists in the network.

Specifying the Layer 2 IPv6 multicast forwarding mode for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Specify the Layer 2 IPv6 multicast forwarding mode.
   mld-snooping forwarding-mode { **ip** | **mac** }

By default, the device forwards Layer 2 IPv6 multicast packets based on the MAC addresses in the packets.

Configuring the MLD snooping querier

Enabling the MLD snooping querier

About this task

This feature enables the Layer 2 device to periodically send MLD general queries to establish and maintain multicast forwarding entries at the data link Layer. You can configure an MLD snooping querier on a network without Layer 3 multicast devices.

Restrictions and guidelines

Do not enable the MLD snooping querier on an IPv6 multicast network that runs MLD. An MLD snooping querier does not participate in MLD querier elections. However, it might affect MLD querier elections if it sends MLD general queries with a low source IPv6 address. For more information about MLD queries, see "Configuring MLD."

On a VXLAN network, the MLD snooping querier in a VSI does not include VLAN tags in MLD general queries. As a best practice, do not enable the MLD snooping querier in a VSI if the VSI uses the Ethernet access mode. For more information about the Ethernet access mode, see *VXLAN Configuration Guide*.

Enabling the MLD snooping querier for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable the MLD snooping querier for the VLAN.
   mld-snooping querier
   By default, the MLD snooping querier is disabled for a VLAN.

Enabling the MLD snooping querier for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable the MLD snooping querier for the VSI.
   mld-snooping querier
   By default, the MLD snooping querier is disabled for a VSI.

Enabling MLD snooping querier election

About this task

To avoid traffic interruption caused by the failure of a single querier in a VLAN or VSI, configure multiple queriers in the VLAN or VSI and enable querier election. When the elected querier fails, the device starts a new querier election to ensure multicast forwarding. The mechanism for MLD snooping querier election is the same as that for MLD querier election. For more information about MLD querier election, see "Configuring MLD."

Prerequisites

Before you enable MLD snooping querier election, you must complete the following tasks:

- Enable the MLD snooping querier for a VLAN or VSI. For more information about enabling the MLD snooping querier, see "Enabling the MLD snooping querier."
- Configure the source IPv6 address for MLD general queries as an IPv6 address different from :: and the local querier IPv6 address. An MLD snooping querier performs querier election only if the source IPv6 address of a received MLD general query is not :: or its own IPv6 address.
- Make sure the candidate MLD snooping queriers run the same MLD snooping version. To specify the MLD snooping version, use the **mld-snooping version** command.

Enabling MLD snooping querier election for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable MLD snooping querier election for the VLAN.
   mld-snooping querier-election
   By default, MLD snooping querier election is disabled for a VLAN.

Enabling MLD snooping querier election for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable MLD snooping querier election for the VSI.
   mld-snooping querier-election
   By default, MLD snooping querier election is disabled for a VSI.

Configuring parameters for MLD general queries and responses

About this task

You can modify the MLD general query interval based on the actual network conditions.

A receiver host starts a report delay timer for each IPv6 multicast group that it has joined when it receives an MLD general query. This timer is set to a random value in the range of 0 to the maximum response time advertised in the query. When the timer value decreases to 0, the host sends an MLD report to the corresponding IPv6 multicast group.

Set an appropriate value for the maximum response time for MLD general queries to speed up hosts' responses to MLD general queries and avoid MLD report traffic bursts.

Restrictions and guidelines

To avoid mistakenly deleting IPv6 multicast group members, make sure the MLD general query interval is greater than the maximum response time for MLD general queries.

Configuring parameters for MLD general queries and responses globally

1. Enter system view.
   system-view
2. Enter MLD-snooping view.

mld-snooping

3. Set the maximum response time for MLD general queries.

   max-response-time *seconds*

   By default, the maximum response time for MLD general queries is 10 seconds.

Configuring parameters for MLD general queries and responses for a VLAN

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Set the MLD general query interval in the VLAN.

   mld-snooping query-interval *interval*

   By default, the MLD general query interval for a VLAN is 125 seconds.

4. Set the maximum response time for MLD general queries in the VLAN.

   mld-snooping max-response-time *seconds*

   By default, the maximum response time for MLD general queries for a VLAN is 10 seconds.

Configuring parameters for MLD general queries and responses for a VSI

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Set the MLD general query interval in the VSI.

   mld-snooping query-interval *interval*

   By default, the MLD general query interval for a VSI is 125 seconds.

4. Set the maximum response time for MLD general queries in the VSI.

   mld-snooping max-response-time *seconds*

   By default, the maximum response time for MLD general queries for a VSI is 10 seconds.

Enabling the device to send MLD general queries upon a path change

About this task

Perform this task to enable the device to immediately send an MLD general query to all ports in the affected VLAN or VSI when the path changes. This feature ensures that the device can update the member port information in IPv6 multicast forwarding entries and switch IPv6 multicast traffic to the new path promptly.

Restrictions and guidelines

Perform this task on the MLD snooping querier in a network of which the topology frequently changes.

Procedure

1. Enter system view.

   system-view

2. Enter MLD-snooping view.

   mld-snooping

3. Enable the device to send MLD general queries upon a path change.
   send-query enable
   By default, the device does not send MLD general queries upon a path change.

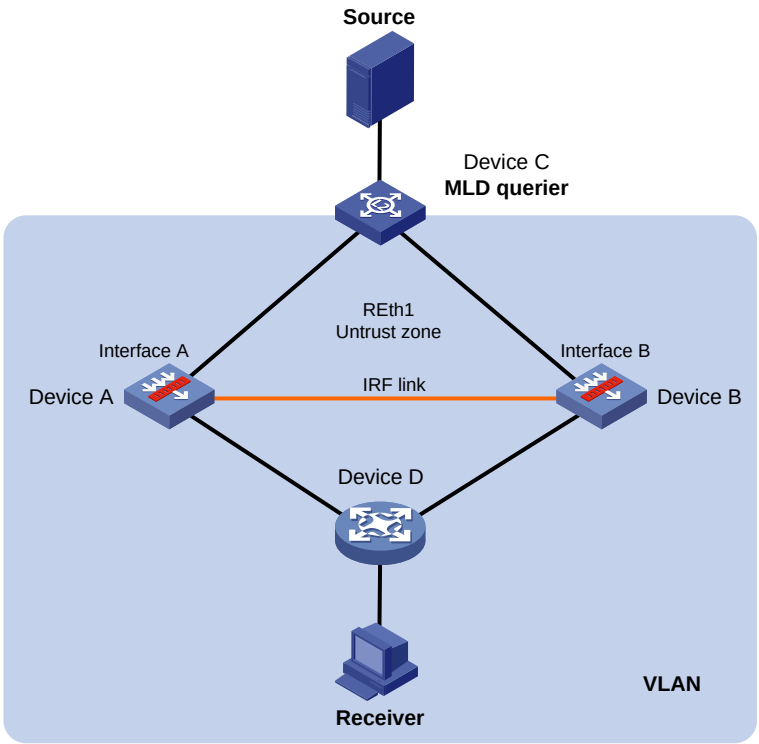
Enabling the device to send MLD general queries upon a port state change

About this task

As shown in Figure 4, the receiver side's devices connect to the Layer 2 multicast device through a Reth interface for high availability. Interface A and interface B are member interfaces of the Reth interface, and interface A is active. Multicast data is sent to the receivers through Device A. When interface A fails, Device C does not send an MLD general query immediately and cannot update interface B to multicast forwarding entries. For information about Reth interfaces, see *High Availability Configuration Guide*.

Perform this task to enable the device to immediately send an MLD general query to all ports in the affected VLAN when the state of a port changes. This feature ensures that the device can update the member port information in multicast forwarding entries and switch multicast traffic to the new link promptly.

Figure 4 Network diagram



Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable the device to send MLD general queries upon a port state change.
   mld-snooping triggered-query enable { **port-down** | **port-up** } *

By default, the device does not send MLD general queries upon a port state change.

Enabling MLD snooping proxying

About MLD snooping proxying

The device enabled with MLD snooping proxying is called an MLD snooping proxy. The MLD snooping proxy acts as a host to the upstream device. Enabled with MLD snooping querier, the MLD snooping proxy acts as the router to downstream devices and receives report and done messages on behalf of the upstream device. As a best practice, enable MLD snooping proxy on the edge device to alleviate the effect caused by excessive packets.

Restrictions and guidelines for enabling MLD snooping proxying

Before you enable MLD snooping proxying for a VLAN or VSI, you must first enable MLD snooping globally and enable MLD snooping for the VLAN or VSI. MLD snooping proxying does not take effect on sub VLANs of a multicast VLAN.

Use this feature with MLD snooping querier. For more information about enabling MLD snooping querier, see "Enabling the MLD snooping querier."

Enabling MLD snooping proxying for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable MLD snooping proxying for the VLAN.
   mld-snooping proxy enable
   By default, MLD snooping proxying is disabled for a VLAN.

Enabling MLD snooping proxying for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable MLD snooping proxying for the VSI.
   mld-snooping proxy enable
   By default, MLD snooping proxying is disabled for a VSI.

Configuring parameters for MLD messages

Configuring source IPv6 addresses for MLD messages

About this task

You can change the source IPv6 address of the MLD queries sent by an MLD snooping querier. This configuration might affect MLD querier election within the subnet.

You can also change the source IPv6 address of MLD reports or done messages sent by a simulated member host or an MLD snooping proxy.

Configuring the source IPv6 addresses for MLD messages for a VLAN

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Configure the source IPv6 address for MLD general queries.

   mld-snooping general-query source-ip *ipv6-address*

   By default, the source IPv6 address of MLD general queries is the IPv6 link-local address of the current VLAN interface. If the current VLAN interface does not have an IPv6 link-local address, the source IPv6 address is FE80::02FF:FFFF:FE00:0001.

4. Configure the source IPv6 address for MLD multicast-address-specific queries.

   mld-snooping special-query source-ip *ipv6-address*

   By default, the source IPv6 address of MLD multicast-address-specific queries is one of the following:

   - The source address of MLD general queries if the MLD snooping querier of the VLAN has received MLD general queries.
   - The IPv6 link-local address of the current VLAN interface if the MLD snooping querier does not receive an MLD general query.
   - FE80::02FF:FFFF:FE00:0001 if the MLD snooping querier does not receive an MLD general query and the current VLAN interface does not have an IPv6 link-local address.

5. Configure the source IPv6 address for MLD reports.

   mld-snooping report source-ip *ipv6-address*

   By default, the source IPv6 address of MLD reports is the IPv6 link-local address of the current VLAN interface. If the current VLAN interface does not have an IPv6 link-local address, the source IPv6 address is FE80::02FF:FFFF:FE00:0001.

6. Configure the source IPv6 address for MLD done messages.

   mld-snooping done source-ip *ipv6-address*

   By default, the source IPv6 address of MLD done messages is the IPv6 link-local address of the current VLAN interface. If the current VLAN interface does not have an IPv6 link-local address, the source IPv6 address is FE80::02FF:FFFF:FE00:0001.

Configuring the source IPv6 addresses for MLD messages for a VSI

1. Enter system view.

   system-view

2. Enter VLAN view.

   vsi *vsi-name*

3. Configure the source IPv6 address for MLD general queries.

mld-snooping general-query source-ip *ipv6-address*

By default, the source IPv6 address of MLD general queries is the IPv6 link-local address of the gateway interface for a VSI. If the gateway interface does not have an IPv6 link-local address, the source IPv6 address is FE80::02FF:FFFF:FE00:0001.

4. Configure the source IPv6 address for MLD multicast-address-specific queries.

   mld-snooping special-query source-ip *ipv6-address*

   By default, the source IPv6 address of MLD multicast-address-specific queries is one of the following:

   - The source address of MLD general queries if the MLD snooping querier of the VSI has received MLD general queries.
   - IPv6 link-local address of the gateway interface for the VSI if the MLD snooping querier does not receive an MLD general query.
   - FE80::02FF:FFFF:FE00:0001 if the gateway interface does not have an IPv6 link-local address.

5. Configure the source IPv6 address for MLD reports.

   mld-snooping report source-ip *ipv6-address*

   By default, the source IPv6 address of MLD reports is the IPv6 link-local address of the gateway interface for a VSI. If the gateway interface does not have an IPv6 link-local address, the source IPv6 address is FE80::02FF:FFFF:FE00:0001.

6. Configure the source IPv6 address for MLD done messages.

   mld-snooping done source-ip *ipv6-address*

   By default, the source IPv6 address of MLD done messages is the IPv6 link-local address of the gateway interface for a VSI. If the gateway interface does not have an IPv6 link-local address, the source IPv6 address is FE80::02FF:FFFF:FE00:0001.

Setting the 802.1p priority for MLD messages

About this task

When congestion occurs on outgoing ports of the Layer 2 device, it forwards MLD messages in their 802.1p priority order, from highest to lowest. You can assign a higher 802.1p priority to MLD messages that are created or forwarded by the device.

Restrictions and guidelines

The configuration in MLD-snooping view takes effect only for VLANs. For VSIs, you can set the 802.1p priority for MLD messages only in VSI view.

Setting the 802.1p priority for MLD messages globally

1. Enter system view.

   system-view

2. Enter MLD-snooping view.

   mld-snooping

3. Set the 802.1p priority for MLD messages globally.

   dot1p-priority *priority*

   By default, the global 802.1p priority is not set for MLD messages.

Setting the 802.1p priority for MLD messages for a VLAN

1. Enter system view.

   system-view

2. Enter VLAN view.

vlan *vlan-id*

3. Set the 802.1p priority for MLD messages in the VLAN.
   mld-snooping dot1p-priority *priority*
   By default, the 802.1p priority is not set for MLD messages.

Configuring MLD snooping policies

Configuring an IPv6 multicast group policy

About this task

This feature enables the Layer 2 device to filter MLD reports by using an ACL that specifies the IPv6 multicast groups and the optional sources. It is used to control the IPv6 multicast groups that receiver hosts can join. This configuration takes effect on the IPv6 multicast groups that ports join dynamically.

In an IPv6 multicast application, a host sends an unsolicited MLD report when a user requests an IPv6 multicast program. The Layer 2 device uses the IPv6 multicast group policy to filter the MLD report. The host can join the IPv6 multicast group only if the MLD report is permitted by the IPv6 multicast group policy.

Configuring an IPv6 multicast group policy globally

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Configure an IPv6 multicast group policy globally.
   group-policy *ipv6-acl-number* [**vlan** *vlan-list*]
   By default, no IPv6 multicast group policies exist. Hosts can join any IPv6 multicast groups.

Configuring an IPv6 multicast group policy on a port

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - *Enter* Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Configure an IPv6 multicast group policy on the port.
   mld-snooping group-policy *ipv6-acl-number* [**vlan** *vlan-list*]
   By default, no IPv6 multicast group policies exist on a port. Hosts attached to the port can join any IPv6 multicast groups.

Enabling dropping unknown IPv6 multicast data

About this task

Unknown IPv6 multicast data refers to IPv6 multicast data for which no forwarding entries exist in the MLD snooping forwarding table. This feature enables the device to forward unknown IPv6 multicast

data only to the router port. If the device does not have a router port, unknown IPv6 multicast data will be dropped.

If you do not enable this feature, the unknown IPv6 multicast data is flooded in the VLAN or VSI to which the data belongs.

Restrictions and guidelines

When dropping unknown IPv6 multicast data is enabled, the device also drops unknown IPv4 multicast data.

When this feature is enabled for a VLAN or VSI, the device still forwards unknown IPv6 multicast data out of router ports (except the receiving router port) in this VLAN or VSI.

If the outgoing ports of the IPv6 multicast forwarding entry for a Layer 3 IPv6 multicast packet are in the same VLAN, enable dropping unknown IPv6 multicast data for the VLAN as a best practice. This restriction prevents the device from flooding the Layer 3 IPv6 multicast packet in the VLAN.

Enabling dropping unknown IPv6 multicast data for a VLAN

1. Enter system view.

   system-view
2. Enter VLAN view.

   vlan *vlan-id*
3. Enable dropping unknown IPv6 multicast data for the VLAN.

   mld-snooping drop-unknown

   By default, dropping unknown IPv6 multicast data is disabled. Unknown IPv6 multicast data is flooded.

Enabling dropping unknown IPv6 multicast data for a VSI

1. Enter system view.

   system-view
2. Enter VSI view.

   vsi *vsi-name*
3. Enable dropping unknown IPv6 multicast data for the VSI.

   mld-snooping drop-unknown

   By default, dropping unknown IPv6 multicast data is disabled. Unknown IPv6 multicast data is flooded.

Enabling MLD report suppression

About this task

This feature enables the Layer 2 device to forward only the first MLD report for an IPv6 multicast group to its directly connected Layer 3 device. Other reports for the same group in the same query interval are discarded. Use this feature to reduce the multicast traffic.

Procedure

1. Enter system view.

   system-view
2. Enter MLD-snooping view.

   mld-snooping
3. Enable MLD report suppression.

   report-aggregation

   By default, MLD report suppression is enabled.

Setting the maximum number of IPv6 multicast groups on a port

About this task

You can set the maximum number of IPv6 multicast groups on a port to regulate the port traffic. This feature takes effect only on the IPv6 multicast groups that the port joins dynamically.

If the number of IPv6 multicast groups on a port exceeds the limit, the system removes all the forwarding entries related to that port. In this case, the receiver hosts attached to that port can join IPv6 multicast groups again before the number of IPv6 multicast groups on the port reaches the limit. When the number of IPv6 multicast groups on the port reaches the limit, the port automatically drops MLD reports for new IPv6 multicast groups.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Set the maximum number of IPv6 multicast groups on the port.
   mld-snooping group-limit *limit* [**vlan** *vlan-list*]
   By default, no limit is placed on the maximum number of IPv6 multicast groups on a port.

Enabling IPv6 multicast group replacement

About this task

When IPv6 multicast group replacement is enabled, the port does not drop MLD reports for new groups if the number of multicast groups on the port reaches the upper limit. Instead, the port leaves an IPv6 multicast group that has the lowest IPv6 address and joins the new group contained in the MLD report. The IPv6 multicast group replacement feature is typically used in the channel switching application.

Restrictions and guidelines

This feature takes effect only on the multicast groups that the port joins dynamically.

This feature does not take effect if the following conditions exist:

- The number of the MLD snooping forwarding entries on the device reaches the upper limit.
- The IPv6 multicast group that the port newly joins is not included in the multicast group list maintained by the device.

Enabling IPv6 multicast group replacement globally

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Enable IPv6 multicast group replacement globally.
   overflow-replace [**vlan** *vlan-list*]

By default, IPv6 multicast group replacement is disabled globally.

Enabling IPv6 multicast group replacement on a port

1. Enter system view.
   system-view
2. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Enable IPv6 multicast group replacement on the port.
   mld-snooping overflow-replace [**vlan** *vlan-list*]
   By default, IPv6 multicast group replacement is disabled on a port.

Enabling host tracking

About this task

This feature enables the Layer 2 device to record information about member hosts that are receiving IPv6 multicast data. The information includes IPv6 addresses of the hosts, length of time elapsed since the hosts joined IPv6 multicast groups, and remaining timeout time for the hosts. This feature facilitates monitoring and managing member hosts.

Restrictions and guidelines

The configuration made in MLD-snooping view does not take effect on VSIs.

Enabling host tracking globally

1. Enter system view.
   system-view
2. Enter MLD-snooping view.
   mld-snooping
3. Enable host tracking globally.
   host-tracking
   By default, host tracking is disabled globally.

Enabling host tracking in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable host tracking for the VLAN.
   mld-snooping host-tracking
   By default, host tracking is disabled in a VLAN.

Setting the DSCP value for outgoing MLD protocol packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.
   system-view
2. Enter MLD snooping view.
   mld-snooping
3. Set the DSCP value for outgoing MLD protocol packets.
   dscp *dscp-value*
   By default, the DSCP value is 48 for outgoing MLD protocol packets.

Enabling SNMP notifications for MLD snooping

About this task

To report critical MLD snooping events to an SNMP NMS, enable SNMP notifications for MLD snooping.

For MLD snooping SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about configuring SNMP, see Network Management and Monitoring Configuration Guide.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for MLD snooping.
   snmp-agent trap enable mld-snooping [**entry-refresh-failed**]
   By default, SNMP notifications are enabled for MLD snooping.

Configuring MLD snooping logging

Enabling MLD snooping logging

About this task

This feature enables the MLD snooping module to send log messages to the information center for troubleshooting. With the information center, you can set log message filtering and output rules, including output destinations. For more information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Excessive number of logs affects the device performance. You can disable MLD snooping logging or use the **mld-snooping log suppress** command to limit the rate of log messages sent.

Procedure

1. Enter system view.
   system-view
2. Enter MLD snooping view.
   mld-snooping
3. Enable MLD snooping logging.
   mld-snooping log enable [**group**]
   By default, MLD snooping logging is enabled.

Enabling MLD snooping logging suppression

About this task

Excessive number of logs affects the device performance. This feature limits the rate of log messages sent from the MLD snooping module to the information center.

Procedure

1. Enter system view.
   system-view
2. Enter MLD snooping view.
   mld-snooping
3. Enable MLD snooping logging suppression.
   mld-snooping log suppress [**group**]
   By default, MLD snooping logging suppression is enabled.

Display and maintenance commands for MLD snooping

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display Layer 2 IPv6 multicast fast forwarding entries.	**display ipv6 l2-multicast fast-forwarding cache** [**vlan** *vlan-id*] [*ipv6-source-address* | *ipv6-group-address*] * [**slot** *slot-number*]
Display information about Layer 2 IPv6 multicast groups.	**display ipv6 l2-multicast ip** [**group** *ipv6-group-address* | **source** *ipv6-source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display Layer 2 IPv6 multicast group entries.	**display ipv6 l2-multicast ip forwarding** [**group** *ipv6-group-address* | **source** *ipv6-source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display information about Layer 2 IPv6 MAC multicast groups.	**display ipv6 l2-multicast mac** [*mac-address*] [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display Layer 2 IPv6 MAC multicast	**display ipv6 l2-multicast mac forwarding**

Task	Command
group entries.	[*mac-address*] [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*]
Display static IPv6 multicast MAC address entries.	**display mac-address** [*mac-address* [**vlan** *vlan-id*] | [**multicast**] [**vlan** *vlan-id*] [**count**]]
Display MLD snooping status.	**display mld-snooping** [**global** | **vlan** *vlan-id* | **vsi** *vsi-name*]
Display dynamic MLD snooping group entries.	**display mld-snooping group** [*ipv6-group-address* | *ipv6-source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**interface** *interface-type interface-number* | [**verbose**] [**slot** *slot-number*]]
Display host tracking information.	**display mld-snooping host-tracking vlan** *vlan-id* **group** *ipv6-group-address* [**source** *ipv6-source-address*] [**slot** *slot-number*]
Display dynamic router port information.	**display mld-snooping router-port** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**verbose**] [**slot** *slot-number*]
Display static MLD snooping group entries.	**display mld-snooping static-group** [*ipv6-group-address* | *ipv6-source-address*] * [**vlan** *vlan-id* | **vsi** *vsi-name*] [**verbose**] [**slot** *slot-number*]
Display static router port information.	**display mld-snooping static-router-port** [**vlan** *vlan-id*] [**verbose**] [**slot** *slot-number*]
Display statistics for the MLD messages and IPv6 PIM hello messages learned through MLD snooping.	**display mld-snooping statistics**
Clear Layer 2 IPv6 multicast fast forwarding entries.	**reset ipv6 l2-multicast fast-forwarding cache** [**vlan** *vlan-id*] { { *ipv6-source-address* | *ipv6-group-address* } * | **all** } [**slot** *slot-number*]
Clear dynamic MLD snooping group entries.	**reset mld-snooping group** { *ipv6-group-address* [*ipv6-source-address*] | **all** } [**vlan** *vlan-id* | **vsi** *vsi-name*]
Clear dynamic router port information.	**reset mld-snooping router-port** { **all** | **vlan** *vlan-id* | **vsi** *vsi-name* }
Clear statistics for MLD messages and IPv6 PIM hello messages learned through MLD snooping.	**reset mld-snooping statistics**

MLD snooping configuration examples

Example: Configuring VLAN-based IPv6 group policy and simulated joining

Network configuration

As shown in Figure 5, Router A runs MLDv1 and acts as the MLD querier, and Switch A runs MLDv1 snooping.

Configure the group policy and simulate joining to meet the following requirements:

- Host A and Host B receive only the IPv6 multicast data addressed to the IPv6 multicast group FF1E::101. IPv6 multicast data can be forwarded through GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A uninterruptedly, even though Host A and Host B fail to receive the multicast data.
- Switch A will drop unknown IPv6 multicast data instead of flooding it in VLAN 100.

Figure 5 Network diagram

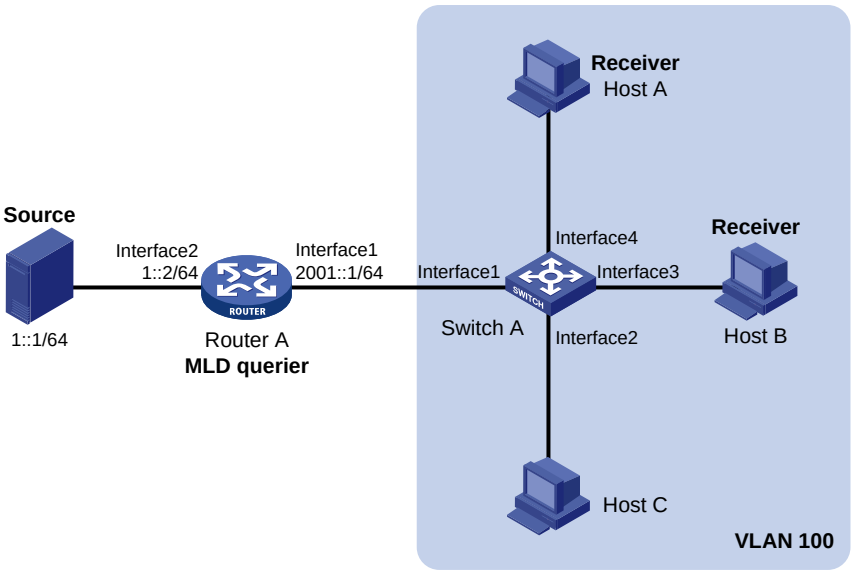


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Prerequisites

Assign an IPv6 address and prefix length to each interface, as shown in Figure 5.

Procedure

1. Configure Router A:

   # Enable IPv6 multicast routing.

```
<RouterA> system-view
[RouterA] ipv6 multicast routing
[RouterA-mrib6] quit
```

Enable MLD on GigabitEthernet 1/0/1.

```
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] mld enable
[RouterA-GigabitEthernet1/0/1] quit
```

Enable IPv6 PIM-DM on GigabitEthernet 1/0/2.

```
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] ipv6 pim dm
[RouterA-GigabitEthernet1/0/2] quit
```

2. Configure Switch A:

Enable the MLD snooping feature.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable MLD snooping, and enable dropping IPv6 unknown multicast data for VLAN 100.

```
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] mld-snooping drop-unknown
[SwitchA-vlan100] quit
```

Configure an IPv6 multicast group policy so that hosts in VLAN 100 can join only IPv6 multicast group FF1E::101.

```
[SwitchA] acl ipv6 basic 2001
[SwitchA-acl-ipv6-basic-2001] rule permit source ff1e::101 128
[SwitchA-acl-ipv6-basic-2001] quit
[SwitchA] mld-snooping
[SwitchA–mld-snooping] group-policy 2001 vlan 100
[SwitchA–mld-snooping] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 as simulated member hosts to join IPv6 multicast group FF1E::101.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] mld-snooping host-join ff1e::101 vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
[SwitchA] interface gigabitethernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] mld-snooping host-join ff1e::101 vlan 100
[SwitchA-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Send MLD reports from Host A and Host B to join IPv6 multicast groups FF1E::101 and FF1E::202. (Details not shown.)

Display dynamic MLD snooping group entries for VLAN 100 on Switch A.

```
[SwitchA] display mld-snooping group vlan 100
Total 1 entries.
```

```
VLAN 100: Total 1 entries.
  (::, FF1E::101)
    Host ports (2 in total):
      GE1/0/3                                  (00:03:23)
      GE1/0/4                                  (00:04:10)
```

The output shows the following information:

- Host A and Host B have joined IPv6 multicast group FF1E::101 through the member ports GigabitEthernet 1/0/4 and GigabitEthernet 1/0/3 on Switch A, respectively.
- Host A and Host B have failed to join the multicast group FF1E::202.

Example: Configuring VLAN-based static ports

Network configuration

As shown in Figure 6:

- Router A runs MLDv1 and acts as the MLD querier. Switch A, Switch B, and Switch C run MLDv1 snooping.
- Host A and Host C are permanent receivers of IPv6 multicast group FF1E::101.

Configure static ports to meet the following requirements:

- To enhance the reliability of IPv6 multicast traffic transmission, configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C as static member ports for IPv6 multicast group FF1E::101.
- Suppose the STP runs on the network. To avoid data loops, the forwarding path from Switch A to Switch C is blocked. IPv6 multicast data flows to the receivers attached to Switch C only along the path of Switch A—Switch B—Switch C. When this path is blocked, a minimum of one MLD query-response cycle must be completed before IPv6 multicast data flows to the receivers along the path of Switch A—Switch C. In this case, the multicast delivery is interrupted during the process. For more information about the STP, see *Layer 2—LAN Switching Configuration Guide*.

  Configure GigabitEthernet 1/0/3 on Switch A as a static router port. Then, IPv6 multicast data can flow to the receivers nearly uninterrupted along the path of Switch A—Switch C when the path of Switch A—Switch B—Switch C is blocked.

Figure 6 Network diagram

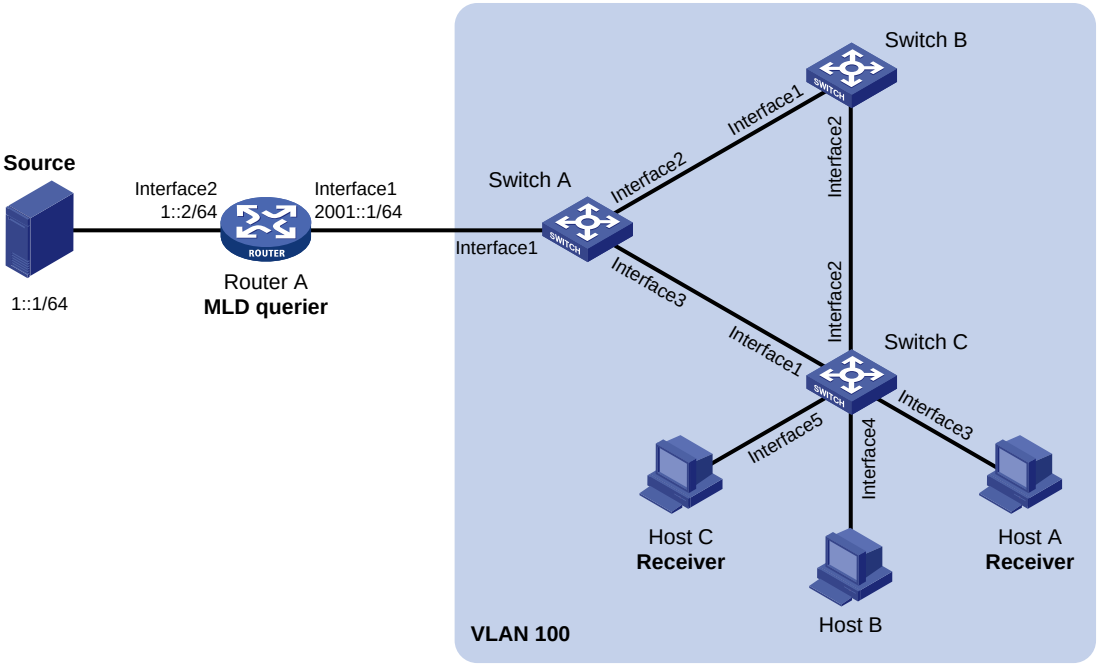


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Prerequisites

Assign an IPv6 address and prefix length to each interface, as shown in Figure 6.

Procedure

1. Configure Router A:

   # Enable IPv6 multicast routing.

   ```
   <RouterA> system-view
   [RouterA] ipv6 multicast routing
   [RouterA-mrib6] quit
   ```

   # Enable MLD on GigabitEthernet 1/0/1.

   ```
   [RouterA] interface gigabitethernet 1/0/1
   [RouterA-GigabitEthernet1/0/1] mld enable
   [RouterA-GigabitEthernet1/0/1] quit
   ```

   # Enable IPv6 PIM-DM on GigabitEthernet 1/0/2.

   ```
   [RouterA] interface gigabitethernet 1/0/2
   [RouterA-GigabitEthernet1/0/2] ipv6 pim dm
   [RouterA-GigabitEthernet1/0/2] quit
   ```

2. Configure Switch A:

Enable the MLD snooping feature.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
```

Enable MLD snooping for VLAN 100.

```
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] quit
```

Configure GigabitEthernet 1/0/3 as a static router port.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] mld-snooping static-router-port vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
```

3. Configure Switch B:

Enable the MLD snooping feature.

```
<SwitchB> system-view
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 gigabitethernet 1/0/2
```

Enable MLD snooping for VLAN 100.

```
[SwitchB-vlan100] mld-snooping enable
[SwitchB-vlan100] quit
```

4. Configure Switch C:

Enable the MLD snooping feature.

```
<SwitchC> system-view
[SwitchC] mld-snooping
[SwitchC-mld-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/5 to the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/5
```

Enable MLD snooping for VLAN 100.

```
[SwitchC-vlan100] mld-snooping enable
[SwitchC-vlan100] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 as static member ports for IPv6 multicast group FF1E::101.

```
[SwitchC] interface gigabitethernet 1/0/3
[SwitchC-GigabitEthernet1/0/3] mld-snooping static-group ff1e::101 vlan 100
[SwitchC-GigabitEthernet1/0/3] quit
[SwitchC] interface gigabitethernet 1/0/5
[SwitchC-GigabitEthernet1/0/5] mld-snooping static-group ff1e::101 vlan 100
[SwitchC-GigabitEthernet1/0/5] quit
```

Verifying the configuration

Display static router port information for VLAN 100 on Switch A.

```
[SwitchA] display mld-snooping static-router-port vlan 100
VLAN 100:
  Router ports (1 in total):
    GE1/0/3
```

The output shows that GigabitEthernet 1/0/3 on Switch A has become a static router port.

Display static MLD snooping group entries in VLAN 100 on Switch C.

```
[SwitchC] display mld-snooping static-group vlan 100
Total 1 entries).

VLAN 100: Total 1 entries).
  (::, FF1E::101)
    Host ports (2 in total):
      GE1/0/3
      GE1/0/5
```

The output shows that GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C have become static member ports of IPv6 multicast group FF1E::101.

Example: Configuring the VLAN-based MLD snooping querier

Network configuration

As shown in Figure 7:

- The network is a Layer 2-only network.
- Source 1 and Source 2 send multicast data to IPv6 multicast groups FF1E::101 and FF1E::102, respectively.
- Host A and Host C are receivers of IPv6 multicast group FF1E::101, and Host B and Host D are receivers of IPv6 multicast group FF1E::102.
- All host receivers run MLDv1 and all switches run MLDv1 snooping. Switch A (which is close to the multicast sources) acts as the MLD snooping querier.

To prevent the switches from flooding unknown IPv6 packets in the VLAN, enable all the switches to drop unknown IPv6 multicast packets.

Figure 7 Network diagram

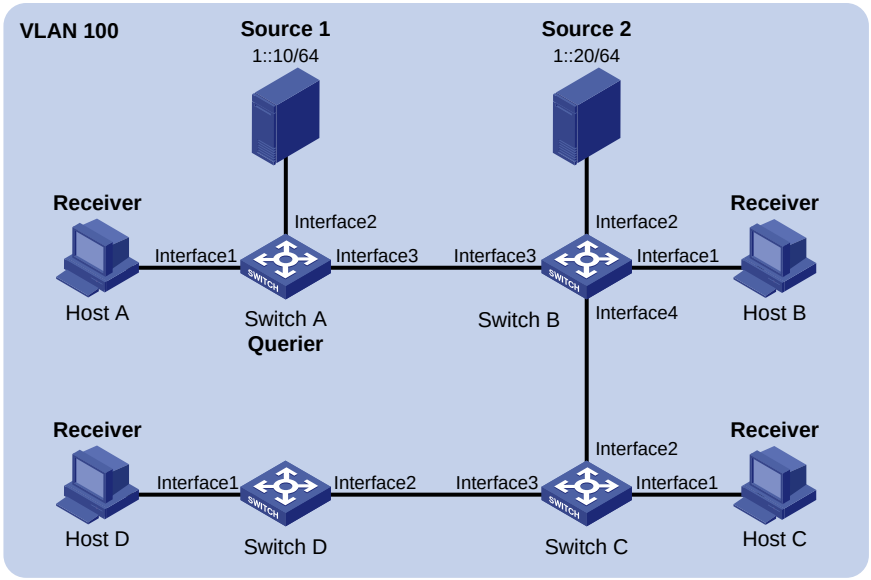


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Switch A:

   # Enable the MLD snooping feature.

   ```
   <SwitchA> system-view
   [SwitchA] mld-snooping
   [SwitchA-mld-snooping] quit
   ```

   # Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the VLAN.

   ```
   [SwitchA] vlan 100
   [SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
   ```

   # Enable MLD snooping, and enable dropping unknown IPv6 multicast data for VLAN 100.

   ```
   [SwitchA-vlan100] mld-snooping enable
   [SwitchA-vlan100] mld-snooping drop-unknown
   ```

   # Configure Switch A as the MLD snooping querier.

   ```
   [SwitchA-vlan100] MLD-snooping querier
   [SwitchA-vlan100] quit
   ```

2. Configure Switch B:

   # Enable the MLD snooping feature.

   ```
   <SwitchB> system-view
   [SwitchB] mld-snooping
   [SwitchB-mld-snooping] quit
   ```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable MLD snooping, and enable dropping unknown IPv6 multicast data for VLAN 100.

```
[SwitchB-vlan100] mld-snooping enable
[SwitchB-vlan100] mld-snooping drop-unknown
[SwitchB-vlan100] quit
```

3. Configure Switch C:

Enable the MLD snooping feature.

```
<SwitchC> system-view
[SwitchC] mld-snooping
[SwitchC-mld-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
```

Enable MLD snooping, and enable dropping unknown IPv6 multicast data for VLAN 100.

```
[SwitchC-vlan100] mld-snooping enable
[SwitchC-vlan100] mld-snooping drop-unknown
[SwitchC-vlan100] quit
```

4. Configure Switch D:

Enable the MLD snooping feature.

```
<SwitchD> system-view
[SwitchD] mld-snooping
[SwitchD-mld-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the VLAN.

```
[SwitchD] vlan 100
[SwitchD-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/2
```

Enable MLD snooping, and enable dropping unknown IPv6 multicast data for VLAN 100.

```
[SwitchD-vlan100] mld-snooping enable
[SwitchD-vlan100] mld-snooping drop-unknown
[SwitchD-vlan100] quit
```

Verifying the configuration

Display statistics for MLD messages and IPv6 PIM hello messages learned through MLD snooping on Switch B.

```
[SwitchB] display mld-snooping statistics
Received MLD messages:
  MLD general queries: 19
    M-LAG: 19
  MLDv1 specific queries: 0
    M-LAG: 0
  MLDv2 specific queries: 0
    M-LAG: 0
  MLDv2 specific sg queries: 0
    M-LAG: 0
  MLDv1 reports: 19
```

```
    M-LAG: 19
  MLDv2 reports: 0
    M-LAG: 0
  MLDv2 reports with right and wrong records: 0
    M-LAG: 0
  MLD dones: 5
     M-LAG: 5
  Error MLD messages: 1
    M-LAG: 1
Sent MLD messages:
  MLDv1 specific queries: 0
  MLDv2 specific queries: 0
  MLDv2 specific sg queries: 0
Received IPv6 PIM hello: 0
```

The output shows that all switches except Switch A can receive the MLD general queries after Switch A acts as the MLD snooping querier.

Example: Configuring VLAN-based MLD snooping proxying

Network configuration

As shown in Figure 8, Router A runs MLDv1 and acts as the MLD querier. Switch A runs MLDv1 snooping. Configure MLD snooping proxying so that Switch A can perform the following actions:

- Forward MLD report and done messages to Router A.
- Respond to MLD queries sent by Router A and forward the queries to downstream hosts.

Figure 8 Network diagram

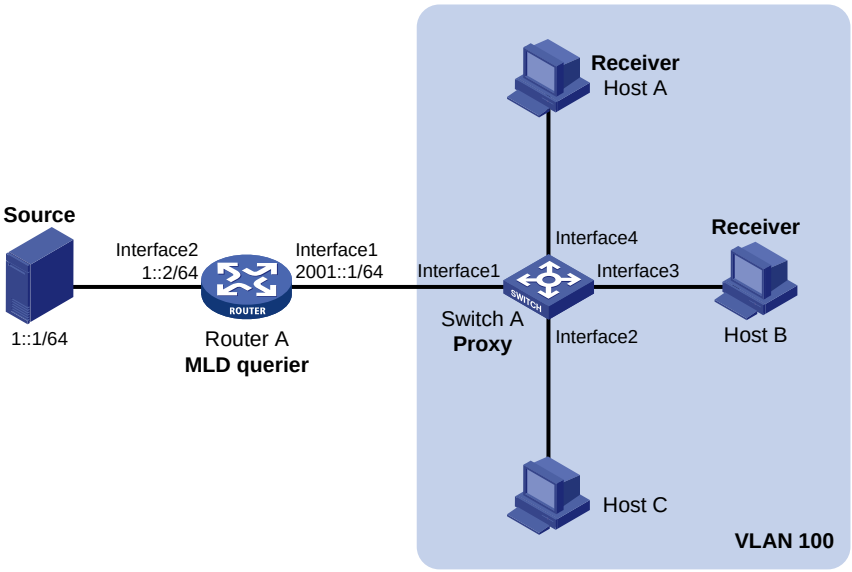


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Prerequisites

Assign an IPv6 address and subnet mask to each interface, as shown in Figure 8.

Procedure

1. Configure Router A:

   # Enable IPv6 multicast routing.

   ```
   <RouterA> system-view
   [RouterA] ipv6 multicast routing
   [RouterA-mrib6] quit
   ```

   # Enable MLD and IPv6 PIM-DM on GigabitEthernet 1/0/1.

   ```
   [RouterA] interface gigabitethernet 1/0/1
   [RouterA-GigabitEthernet1/0/1] mld enable
   [RouterA-GigabitEthernet1/0/1] ipv6 pim dm
   [RouterA-GigabitEthernet1/0/1] quit
   ```

   # Enable IPv6 PIM-DM on GigabitEthernet 1/0/2.

   ```
   [RouterA] interface gigabitethernet 1/0/2
   [RouterA-GigabitEthernet1/0/2] ipv6 pim dm
   [RouterA-GigabitEthernet1/0/2] quit
   ```

2. Configure Switch A:

   # Enable the MLD snooping feature.

   ```
   <SwitchA> system-view
   [SwitchA] mld-snooping
   [SwitchA-mld-snooping] quit
   ```

   # Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

   ```
   [SwitchA] vlan 100
   [SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
   ```

   # Enable MLD snooping and MLD snooping proxying for the VLAN.

   ```
   [SwitchA-vlan100] mld-snooping enable
   [SwitchA-vlan100] mld-snooping proxy enable
   [SwitchA-vlan100] quit
   ```

Verifying the configuration

Send MLD reports from Host A and Host B to join IPv6 multicast group FF1E::101. (Details not shown.)

Display MLD snooping group entries on Switch A.

```
[SwitchA] display mld-snooping group
Total 1 entries.

VLAN 100: Total 1 entries.
  (::, FF1E::101)
    Host ports (2 in total):
      GE1/0/3                                  (00:04:09)
      GE1/0/4                                  (00:03:06)
```

The output shows that GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 are member ports of IPv6 multicast group FF1E::101. Host A and Host B will receive IPv6 multicast data for the group.

Display MLD group membership information on Router A.

```
[RouterA] display mld group
MLD groups in total: 1
 GigabitEthernet1/0/1(2001::1):
  MLD groups reported in total: 1
   Group address: FF1E::101
    Last reporter: FE80::2FF:FFFF:FE00:1
    Uptime: 00:00:31
    Expires: 00:03:48
```

Send an MLD done message from Host A to leave IPv6 multicast group FF1E::101. (Details not shown.)

Display MLD snooping group entries on Switch A.

```
[SwitchA] display mld-snooping group
Total 1 entries.

VLAN 100: Total 1 entries.
  (::, FF1E::101)
    Host ports (1 in total):
      GE1/0/3                                  ( 00:01:23 )
```

The output shows that GigabitEthernet 1/0/3 is the only member port of IPv6 multicast group FF1E::101. Only Host B will receive IPv6 multicast data for the group.

Troubleshooting MLD snooping

Layer 2 multicast forwarding cannot function

Symptom

Layer 2 multicast forwarding cannot function through MLD snooping.

Solution

To resolve the problem:

1. Use the **display mld-snooping** command to display MLD snooping status.
2. If MLD snooping is not enabled, use the **mld-snooping** command in system view to enable the MLD snooping feature. Then, use the **mld-snooping enable** command in VLAN view or VSI view to enable MLD snooping for the VLAN or VSI.
3. If MLD snooping is enabled globally but not enabled for the VLAN or VSI, use the **mld-snooping enable** command in VLAN view or VSI view to enable MLD snooping for the VLAN or VSI.
4. If the problem persists, contact H3C Support.

IPv6 multicast group policy does not work

Symptom

Hosts can receive IPv6 multicast data for IPv6 multicast groups that are not permitted by the IPv6 multicast group policy.

Solution

To resolve the problem:

1. Use the **display acl ipv6** command to verify that the configured IPv6 ACL meets the IPv6 multicast group policy requirements.
2. Use the **display this** command in MLD-snooping view or in interface view to verify that the correct IPv6 multicast group policy has been applied. If the applied policy is not correct, use the **group-policy** or **mld-snooping group-policy** command to apply the correct IPv6 multicast group policy.
3. Use the **display mld-snooping** command to verify that dropping unknown IPv6 multicast data is enabled. If it is not, use the **drop-unknown** or **mld-snooping drop-unknown** command to enable dropping unknown IPv6 multicast data.
4. If the problem persists, contact H3C Support.

Contents

Configuring IPv6 PIM snooping 1

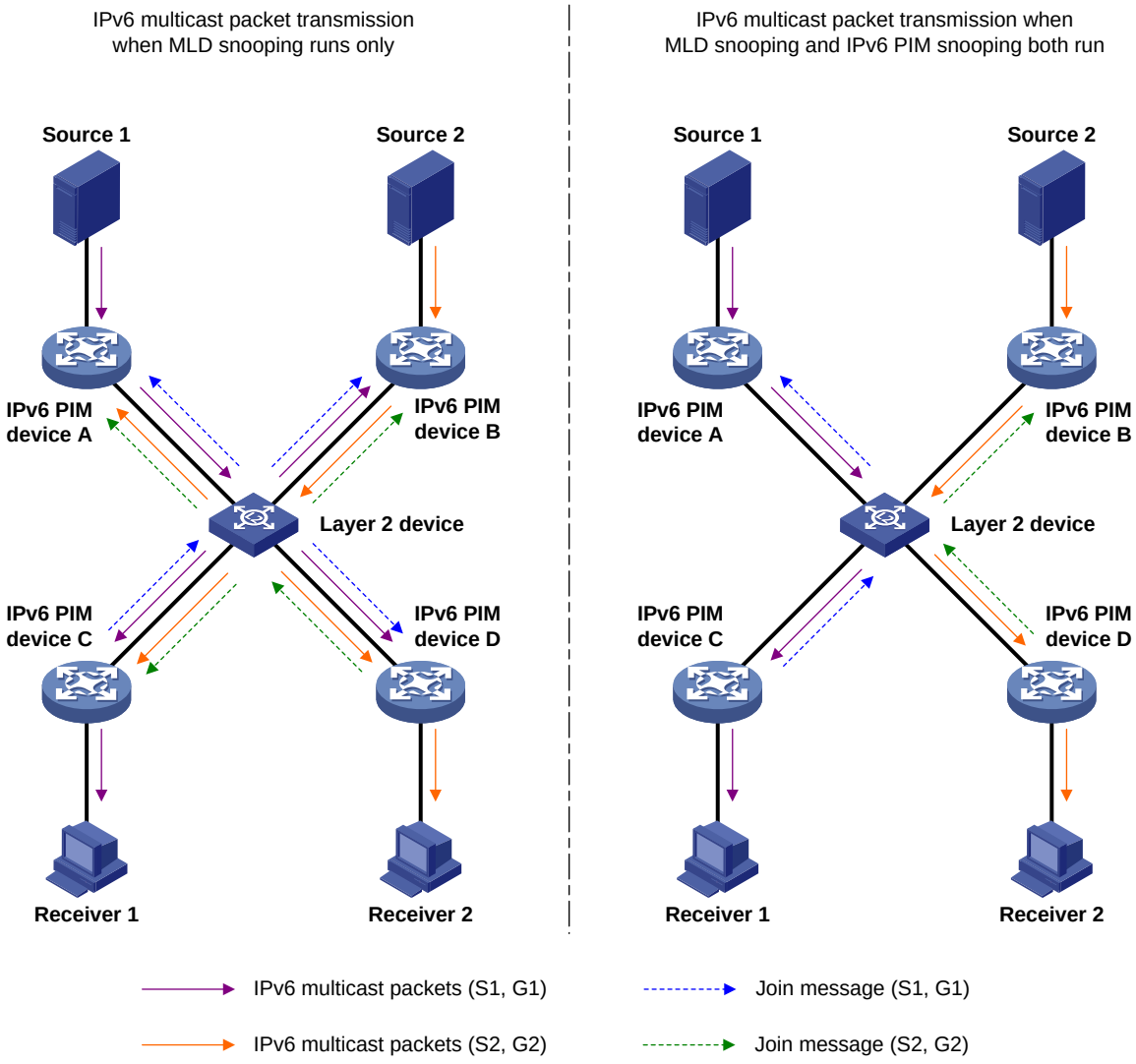
About IPv6 PIM snooping 1
Restrictions and guidelines: IPv6 PIM snooping configuration 2
IPv6 PIM snooping tasks at a glance 2
Enabling IPv6 PIM snooping 2
Setting the aging time for global ports after a master/subordinate switchover 3
About global ports 3
Restrictions and guidelines 3
Setting the aging time for global neighbor ports 3
Setting the aging time for global downstream ports and global router ports 4
Display and maintenance commands for IPv6 PIM snooping 4
IPv6 PIM snooping configuration examples 5
Example: Configuring IPv6 PIM snooping 5
Troubleshooting IPv6 PIM snooping 8
IPv6 PIM snooping does not work on a Layer 2 device 9

Configuring IPv6 PIM snooping

About IPv6 PIM snooping

IPv6 PIM snooping runs on Layer 2 devices. It works with MLD snooping to analyze received IPv6 PIM messages, and adds the ports that are interested in specific multicast data to an IPv6 PIM snooping routing entry. In this way, the multicast data can be forwarded to only the ports that are interested in the data.

Figure 1 Multicast packet transmission without or with IPv6 PIM snooping



As shown in Figure 1, Source 1 sends multicast data to multicast group G1, and Source 2 sends multicast data to multicast group G2. Receiver 1 belongs to G1, and Receiver 2 belongs to G2. The Layer 2 switch's interfaces that connect to the IPv6 PIM-capable routers are in the same VLAN or VSI.

- When the Layer 2 switch only runs MLD snooping, it performs the following actions:
  a. Maintains the router ports according to the received IPv6 PIM hello messages that IPv6 PIM-capable routers send.

b. Floods all other types of received IPv6 PIM messages except PIM hello messages in the VLAN or VSI.

c. Forwards all multicast data to all router ports in the VLAN or VSI.

Each IPv6 PIM-capable router in the VLAN, whether interested in the multicast data or not, can receive all multicast data and all IPv6 PIM messages except IPv6 PIM hello messages.

- When the Layer 2 switch runs both MLD snooping and IPv6 PIM snooping, it performs the following actions:

d. Examines whether an IPv6 PIM router is interested in the multicast data destined for a multicast group according to the received IPv6 PIM messages that the router sends.

e. Adds only the ports that connect to the router and are interested in the data to an IPv6 PIM snooping routing entry.

f. Forwards IPv6 PIM messages and the multicast data only to the router according to the multicast forwarding entry, which saves network bandwidth.

For more information about MLD snooping and the router port, see "Configuring MLD snooping."

Restrictions and guidelines: IPv6 PIM snooping configuration

As a best practice, do not configure IPv6 PIM snooping for secondary VLANs because IPv6 PIM snooping does not take effect on secondary VLANs. For more information about secondary VLANs, see *Layer 2—LAN Switching Configuration Guide*.

Make sure the maximum size of an IPv6 PIM join or prune message is smaller than the path MTU of devices connected to IPv6 PIM snooping devices. Otherwise, fragmented IPv6 PIM join or prune messages prevent IPv6 PIM snooping from correctly forwarding IPv6 multicast data. For more information about setting the maximum size of a join or prune message, see "Configuring IPv6 PIM."

IPv6 PIM snooping is supported only for IPv6 PIM-SM and IPv6 PIM-SSM. As a best practice, do not configure IPv6 PIM snooping in the IPv6 PIM-DM or IPv6 BIDIR-PIM network. For more information about IPv6 PIM, see "IPv6 PIM overview."

After you enable IPv6 PIM snooping for a VLAN or VSI, IPv6 PIM snooping takes effect only on ports that belong to the VLAN or VSI.

IPv6 PIM snooping tasks at a glance

To configure IPv6 PIM snooping, perform the following tasks:

1. Enabling IPv6 PIM snooping
2. (Optional.) Setting the aging time for global ports after a master/subordinate switchover
   - Setting the aging time for global neighbor ports
   - Setting the aging time for global downstream ports and global router ports

Enabling IPv6 PIM snooping

1. Enter system view.

   `system-view`
2. Enable the MLD snooping feature and enter MLD -snooping view.

   `mld-snooping`

   By default, MLD snooping is disabled.

For more information about this command, see *IP Multicast Command Reference*.

3. Return to system view.
   quit
4. Enter VLAN view or VSI view.
   - Enter VLAN view.
     vlan *vlan-id*
   - Enter VSI view.
     vsi *vsi-name*
5. Enable MLD snooping for the VLAN or VSI.
   mld-snooping enable

   By default, MLD snooping is disabled in a VLAN or VSI.

   For more information about this command, see *IP Multicast Command Reference*.
6. Enable IPv6 PIM snooping for the VLAN or VSI.
   ipv6 pim-snooping enable

   By default, IPv6 PIM snooping is disabled in a VLAN or VSI.

Setting the aging time for global ports after a master/subordinate switchover

About global ports

A global port is a virtual port on the master device, such as a Layer 2 aggregate interface, AC, N-PW, and U-PW. A global port that acts as a neighbor port, downstream port, or router port is called a global neighbor port, global downstream port, and global router port, respectively.

Perform this task to decrease Layer 2 IPv6 multicast data interruption caused by the aging of IPv6 PIM snooping entries after a master/subordinate switchover.

Restrictions and guidelines

For a global neighbor port, the set aging time does not take effect when the port receives an IPv6 PIM hello message after a master/subordinate switchover. The aging time for the port is determined by the aging time in the IPv6 PIM hello message.

For a global router port or global downstream port, the set aging time does not take effect when the port receives an IPv6 PIM join message after a master/subordinate switchover. The aging time for the port is determined by the aging time in the IPv6 PIM join message.

Setting the aging time for global neighbor ports

1. Enter system view.
   system-view
2. Enter VLAN view or VSI view.
   - Enter VLAN view.
     vlan *vlan-id*
   - Enter VSI view.
     vsi *vsi-name*

3. Set the aging time for global neighbor ports after a master/subordinate switchover.

   pim-snooping graceful-restart neighbor-aging-time *seconds*

   By default, the aging time for global neighbor ports after a master/subordinate switchover is 105 seconds.

Setting the aging time for global downstream ports and global router ports

1. Enter system view.

   system-view

2. Enter VLAN view or VSI view.
   - Enter VLAN view.

     vlan *vlan-id*
   - Enter VSI view.

     vsi *vsi-name*

3. Set the aging time for global downstream ports and global router ports after a master/subordinate switchover.

   pim-snooping graceful-restart join-aging-time *seconds*

   By default, the aging time for downstream ports and global router ports after a master/subordinate switchover is 210 seconds.

Display and maintenance commands for IPv6 PIM snooping

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display IPv6 PIM snooping neighbor information.	**display ipv6 pim-snooping neighbor** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*] [**verbose**]
Display IPv6 PIM snooping router port information.	**display ipv6 pim-snooping router-port** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*] [**verbose**]
Display IPv6 PIM snooping routing entries.	**display ipv6 pim-snooping routing-table** [**vlan** *vlan-id* | **vsi** *vsi-name*] [**slot** *slot-number*] [**verbose**]
Display statistics for the IPv6 PIM messages learned through IPv6 PIM snooping.	**display ipv6 pim-snooping statistics**
Clear statistics for the IPv6 PIM messages learned through IPv6 PIM snooping.	**reset ipv6 pim-snooping statistics**

IPv6 PIM snooping configuration examples

Example: Configuring IPv6 PIM snooping

Network configuration

As shown in Figure 2:

- OSPFv3 runs on the network.
- Source 1 and Source 2 send IPv6 multicast data to IPv6 multicast groups FF1E::101 and FF2E::101, respectively.
- Receiver 1 and Receiver 2 belong to IPv6 multicast groups FF1E::101 and FF2E::101, respectively.
- Router C and Router D run MLD on GigabitEthernet 1/0/1.
- Router A, Router B, Router C, and Router D run IPv6 PIM-SM. GigabitEthernet 1/0/2 on Router A acts as a C-BSR and a C-RP.

Configure MLD snooping and IPv6 PIM snooping on Switch A. Then, Switch A forwards IPv6 PIM protocol packets and IPv6 multicast data packets only to routers that are connected to receivers.

Figure 2 Network diagram

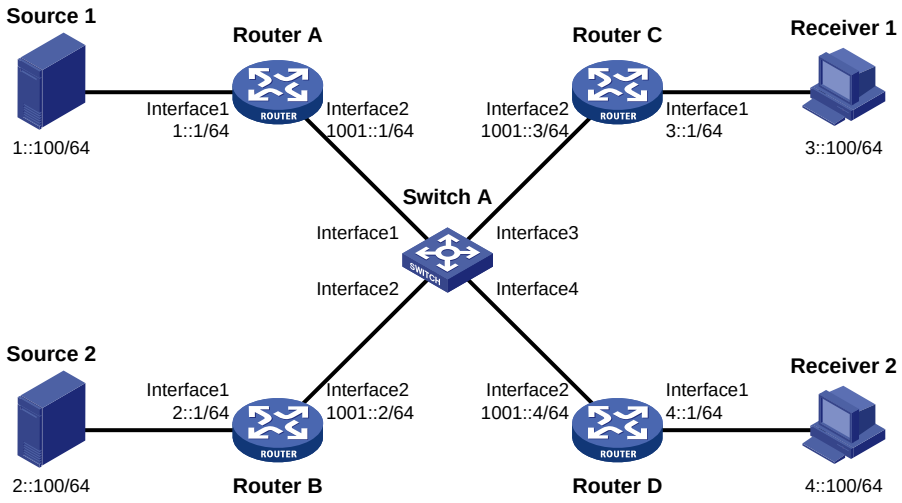


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Prerequisites

Assign an IPv6 address and prefix length to each interface as shown in Figure 2, and then configure OSPFv3 on the routers.

Procedure

1. Configure Router A:

Enable IPv6 multicast routing.

```
<RouterA> system-view
[RouterA] ipv6 multicast routing
[RouterA-mrib6] quit
```

Enable IPv6 PIM-SM on each interface.

```
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] ipv6 pim sm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] ipv6 pim sm
[RouterA-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/2 as a C-BSR and a C-RP.

```
[RouterA] ipv6 pim
[RouterA-pim6] c-bsr 1001::1
[RouterA-pim6] c-rp 1001::1
[RouterA-pim6] quit
```

2. Configure Router B:

Enable IPv6 multicast routing.

```
<RouterB> system-view
[RouterB] ipv6 multicast routing
[RouterB-mrib6] quit
```

Enable IPv6 PIM-SM on each interface.

```
[RouterB] interface gigabitethernet 1/0/1
[RouterB-GigabitEthernet1/0/1] ipv6 pim sm
[RouterB-GigabitEthernet1/0/1] quit
[RouterB] interface gigabitethernet 1/0/2
[RouterB-GigabitEthernet1/0/2] ipv6 pim sm
[RouterB-GigabitEthernet1/0/2] quit
```

3. Configure Router C:

Enable IPv6 multicast routing.

```
<RouterC> system-view
[RouterC] ipv6 multicast routing
[RouterC-mrib6] quit
```

Enable MLD on GigabitEthernet 1/0/1.

```
[RouterC] interface gigabitethernet 1/0/1
[RouterC-GigabitEthernet1/0/1] mld enable
[RouterC-GigabitEthernet1/0/1] quit
```

Enable IPv6 PIM-SM on GigabitEthernet 1/0/2.

```
[RouterC] interface gigabitethernet 1/0/2
[RouterC-GigabitEthernet1/0/2] ipv6 pim sm
[RouterC-GigabitEthernet1/0/2] quit
```

4. Configure Router D:

Enable IPv6 multicast routing.

```
<RouterD> system-view
[RouterD] ipv6 multicast routing
[RouterD-mrib6] quit
```

Enable MLD on GigabitEthernet 1/0/1.

```
[RouterD] interface gigabitethernet 1/0/1
[RouterD-GigabitEthernet1/0/1] mld enable
[RouterD-GigabitEthernet1/0/1] quit
```

Enable IPv6 PIM-SM on GigabitEthernet 1/0/2.

```
[RouterD] interface gigabitethernet 1/0/2
[RouterD-GigabitEthernet1/0/2] ipv6 pim sm
[RouterD-GigabitEthernet1/0/2] quit
```

5. Configure Switch A:

Enable the MLD snooping feature.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable MLD snooping and IPv6 PIM snooping for VLAN 100.

```
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] ipv6 pim-snooping enable
[SwitchA-vlan100] quit
```

Verifying the configuration

On Switch A, display IPv6 PIM snooping neighbor information for VLAN 100.

```
[SwitchA] display ipv6 pim-snooping neighbor vlan 100
Total 4 neighbors.

VLAN 100: Total 4 neighbors.
  FE80::1
    Ports (1 in total):
      GE1/0/1                 (00:32:43)
  FE80::2
   Ports (1 in total):
      GE1/0/2                 (00:32:43)
  FE80::3
    Ports (1 in total):
      GE1/0/3                 (00:32:43)
  FE80::4
    Ports (1 in total):
      GE1/0/4                 (00:32:43)
```

The output shows that Router A, Router B, Router C, and Router D are IPv6 PIM snooping neighbors.

On Switch A, display IPv6 PIM snooping routing entries for VLAN 100.

```
[SwitchA] display ipv6 pim-snooping routing-table vlan 100
Total 4 entries.
FSM Flag: NI-no info, J-join, PP-prune pending

VLAN 100: Total 4 entries.
  (*, FF1E::101)
```

```
  Upstream neighbor: FE80::8CE5:7AFF:FE3F:107
    Upstream Slots (0 in total):
    Upstream Ports (1 in total):
      GE1/0/1
    Downstream Slots (0 in total):
    Downstream Ports (1 in total):
      GE1/0/3
        Expires: 00:02:32, FSM: J
(1::100, FF1E::101)
  Upstream neighbor: FE80::8CE5:7AFF:FE3F:107
    Upstream Slots (0 in total):
    Upstream Ports (1 in total):
      GE1/0/1
    Downstream Slots (0 in total):
    Downstream Ports (1 in total):
      GE1/0/3
        Expires: 00:02:57, FSM: J
(*, FF2E::101)
  Upstream neighbor: FE80::8CE5:7AFF:FE3F:107
    Upstream Slots (0 in total):
    Upstream Ports (1 in total):
      GE1/0/1
    Downstream Slots (0 in total):
    Downstream Ports (1 in total):
      GE1/0/4
        Expires: 00:02:32, FSM: J
(2::100, FF2E::101)
  Upstream neighbor: FE80::8CE5:7EFF:FE3B:207
    Upstream Slots (0 in total):
    Upstream Ports (1 in total):
      GE1/0/2
    Downstream Slots (0 in total):
    Downstream Ports (1 in total):
      GE1/0/4
        Expires: 00:03:05, FSM: J
```

The output shows the following information:

- Switch A will forward the multicast data intended for IPv6 multicast group FF1E::101 to only Router C.
- Switch A will forward the multicast data intended for IPv6 multicast group FF2E::101 to only Router D.

Troubleshooting IPv6 PIM snooping

This section describes common IPv6 PIM snooping problems and how to troubleshoot them.

IPv6 PIM snooping does not work on a Layer 2 device

Symptom

IPv6 PIM snooping does not work on a Layer 2 device.

Solution

To resolve the problem:

1. Use the **`display current-configuration`** command to display information about MLD snooping and IPv6 PIM snooping.
2. If MLD snooping is not enabled, enable the MLD snooping feature, and then enable MLD snooping and IPv6 PIM snooping for the VLAN or VSI.
3. If IPv6 PIM snooping is not enabled, enable IPv6 PIM snooping for the VLAN or VSI.
4. If the problem persists, contact H3C Support.

Contents

Configuring IPv6 multicast VLANs 1

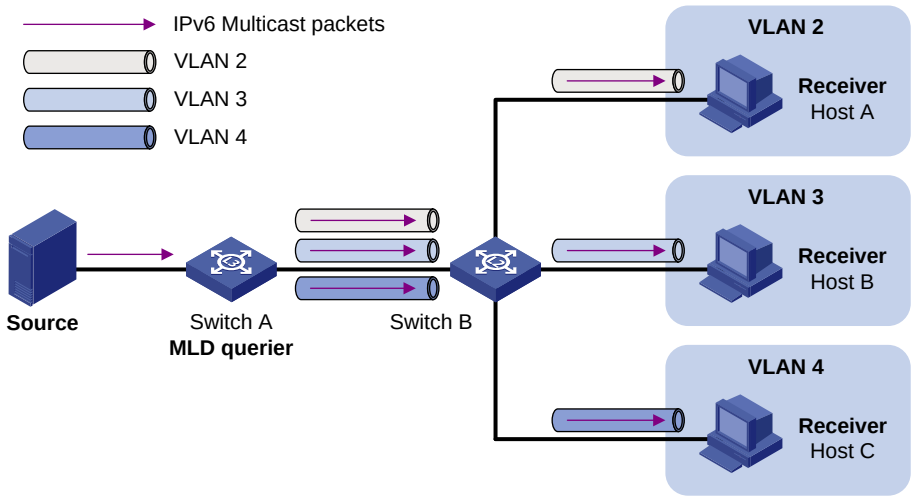
IPv6 multicast VLAN feature 1
IPv6 multicast VLAN methods 1
Sub-VLAN-based IPv6 multicast VLAN 1
Port-based IPv6 multicast VLAN 2
Restrictions and guidelines: IPv6 multicast VLAN configuration 3
Configuring a sub-VLAN-based IPv6 multicast VLAN 3
Configuring a port-based IPv6 multicast VLAN 3
Setting the maximum number of IPv6 multicast VLAN forwarding entries 4
Display and maintenance commands for IPv6 multicast VLANs 5
IPv6 multicast VLAN configuration examples 5
Example: Configuring sub-VLAN-based IPv6 multicast VLAN 5
Example: Configuring port-based IPv6 multicast VLAN 8

Configuring IPv6 multicast VLANs

IPv6 multicast VLAN feature

As shown in Figure 1, Host A, Host B, and Host C are in different VLANs and the same IPv6 multicast group. When Switch A (Layer 3 device) receives IPv6 multicast data for that group, it forwards three copies of the data to Switch B (Layer 2 device). This occupies a large amount of bandwidth and increases the burden on the Layer 3 device.

Figure 1 Multicast transmission without the IPv6 multicast VLAN feature



After an IPv6 multicast VLAN is configured on Switch B, Switch A sends one copy of the IPv6 multicast data to the IPv6 multicast VLAN on Switch B. This saves network bandwidth and lessens the burden on the Layer 3 device.

IPv6 multicast VLAN methods

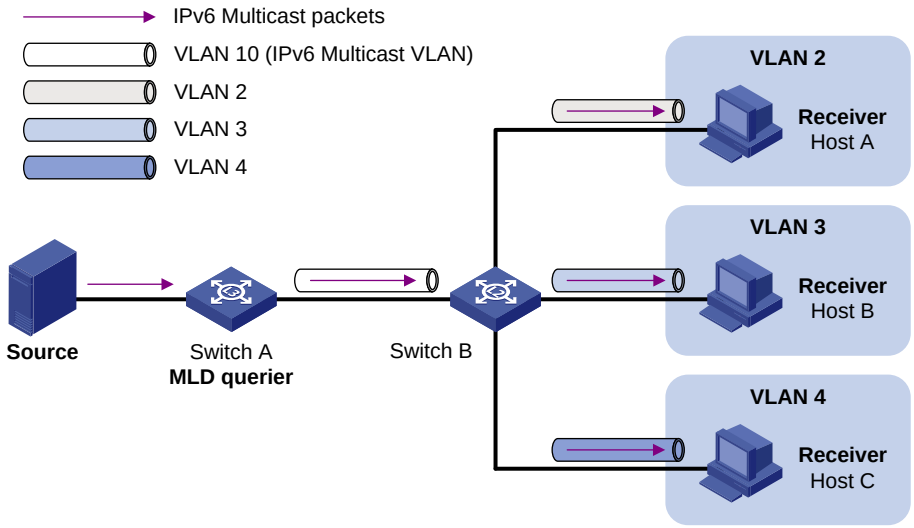
IPv6 multicast VLANs support sub-VLAN-based IPv6 multicast VLANs and port-based IPv6 multicast VLANs.

Sub-VLAN-based IPv6 multicast VLAN

As shown in Figure 2:

- Host A, Host B, and Host C are in VLAN 2 through VLAN 4, respectively.
- On Switch B, VLAN 10 is an IPv6 multicast VLAN. VLAN 2 through VLAN 4 are sub-VLANs of VLAN 10.
- MLD snooping is enabled for the multicast VLAN and its sub-VLANs.

Figure 2 Sub-VLAN-based multicast VLAN



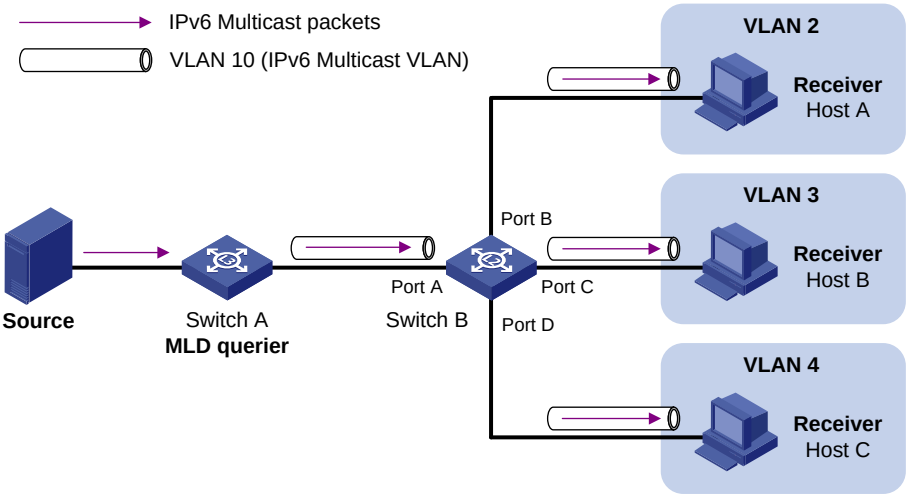
MLD snooping manages router ports in the IPv6 multicast VLAN and member ports in each sub-VLAN. When Switch A receives IPv6 multicast data from the IPv6 multicast source, it sends only one copy of the IPv6 multicast data to the IPv6 multicast VLAN on Switch B. Then, Switch B sends a separate copy to each sub-VLAN of the IPv6 multicast VLAN.

Port-based IPv6 multicast VLAN

As shown in Figure 3:

- Host A, Host B, and Host C are in VLAN 2 through VLAN 4, respectively. All the user ports (ports with attached hosts) on Switch B are hybrid ports.
- On Switch B, VLAN 10 is an IPv6 multicast VLAN. All the user ports are assigned to VLAN 10.
- MLD snooping is enabled for the IPv6 multicast VLAN and its sub-VLANs.

Figure 3 Port-based IPv6 multicast VLAN



MLD snooping manages the router ports and member ports in the IPv6 multicast VLAN. When Switch A receives IPv6 multicast data from the IPv6 multicast source, it sends only one copy of the IPv6 multicast data to the IPv6 multicast VLAN on Switch B. Then, Switch B sends a separate copy to each user port in the IPv6 multicast VLAN.

Restrictions and guidelines: IPv6 multicast VLAN configuration

The VLAN to be configured as an IPv6 multicast VLAN must exist.

If you have configured both a sub-VLAN-based IPv6 multicast VLAN and a port-based IPv6 multicast VLAN on a device, the port-based IPv6 multicast VLAN configuration takes effect.

Do not configure an IPv6 multicast VLAN on a device that is enabled with IPv6 multicast routing. Otherwise, the system displays errors. Do not enable IPv6 multicast routing on a device that is configured with IPv6 multicast VLANs. Otherwise, IPv6 multicast forwarding exceptions occur.

Do not configure a VLAN as an IPv6 multicast VLAN or an IPv6 sub-VLAN if the VLAN interface is enabled with IPv6 PIM or MLD. Otherwise, the system displays errors. Do not enable IPv6 PIM or MLD on a VLAN interface if the VLAN interface belongs to an IPv6 multicast VLAN or an IPv6 multicast sub-VLAN. Otherwise, IPv6 multicast forwarding exceptions occur.

Configuring a sub-VLAN-based IPv6 multicast VLAN

Restrictions and guidelines

The VLANs to be configured as sub-VLANs of an IPv6 multicast VLAN must exist and cannot be IPv6 multicast VLANs or sub-VLANs of any other IPv6 multicast VLANs.

Prerequisites

Before you configure a sub-VLAN-based IPv6 multicast VLAN, you must complete the following tasks:

- Create VLANs as required.
- Enable MLD snooping for the VLAN to be configured as the IPv6 multicast VLAN, and for the VLANs to be configured as sub-VLANs.

Procedure

1. Enter system view.

   system-view
2. Configure a VLAN as an IPv6 multicast VLAN and enter IPv6 multicast VLAN view.

   ipv6 multicast-vlan *vlan-id*

   By default, a VLAN is not an IPv6 multicast VLAN.
3. Assign VLANs to the IPv6 multicast VLAN as sub-VLANs.

   subvlan *vlan-list*

   By default, an IPv6 multicast VLAN does not have any sub-VLANs.

Configuring a port-based IPv6 multicast VLAN

Restrictions and guidelines

You can assign user ports to an IPv6 multicast VLAN in IPv6 multicast VLAN view or assign a user port to an IPv6 multicast VLAN in interface view.

A user port can belong to only one IPv6 multicast VLAN.

Prerequisites

Before you configure a port-based IPv6 multicast VLAN, you must complete the following tasks:

- Create VLANs as required.
- Enable MLD snooping for the VLAN to be configured as the IPv6 multicast VLAN.
- Enable MLD snooping for all the VLANs that contain the multicast receivers.
- Configure the attributes of user ports. Make sure the ports can forward packets from the VLAN to be configured as the IPv6 multicast VLAN and send the packets with the VLAN tag removed. For more information about configuring port attributes, see VLAN configuration in *Layer 2—LAN Switching Configuration Guide*.

Assigning user ports to an IPv6 multicast VLAN in IPv6 multicast VLAN view

1. Enter system view.
   system-view
2. Configure an IPv6 VLAN as an IPv6 multicast VLAN and enter IPv6 multicast VLAN view.
   ipv6 multicast-vlan *vlan-id*
   By default, a VLAN is not an IPv6 multicast VLAN.
3. Assign ports to the IPv6 multicast VLAN as user ports.
   port *interface-list*

Assigning user ports to an IPv6 multicast VLAN in interface view

1. Enter system view.
   system-view
2. Configure an IPv6 VLAN as an IPv6 multicast VLAN and enter IPv6 multicast VLAN view.
   ipv6 multicast-vlan *vlan-id*
   By default, a VLAN is not an IPv6 multicast VLAN.
3. Return to system view.
   quit
4. Enter Layer 2 interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
5. Assign the port to the IPv6 multicast VLAN as a user port.
   ipv6 port multicast-vlan *vlan-id*
   By default, a port does not belong to any IPv6 multicast VLAN.

Setting the maximum number of IPv6 multicast VLAN forwarding entries

About this task

You can set the maximum number of IPv6 multicast VLAN forwarding entries on the device. When the upper limit is reached, the device does not create IPv6 multicast VLAN forwarding entries until some entries age out or are manually removed.

Procedure

1. Enter system view.

system-view

2. Set the maximum number of IPv6 multicast VLAN forwarding entries.
   ipv6 multicast-vlan entry-limit *limit*
   By default, the maximum number of IPv6 multicast VLAN forwarding entries is 4096.

Display and maintenance commands for IPv6 multicast VLANs

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about IPv6 multicast VLANs.	**display ipv6 multicast-vlan** [*vlan-id*]
Display IPv6 multicast VLAN forwarding entries.	**display ipv6 multicast-vlan forwarding-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **slot** *slot-number* | **subvlan** *vlan-id* | **vlan** *vlan-id*] *
Display information about information about IPv6 multicast group entries in IPv6 multicast VLANs.	**display ipv6 multicast-vlan group** [*ipv6-source-address* | *ipv6-group-address* | **slot** *slot-number* | **verbose** | **vlan** *vlan-id*] *
Clear IPv6 multicast group entries in IPv6 multicast VLANs.	**reset ipv6 multicast-vlan group** [*ipv6-group-address* [*prefix-length*] | *ipv6-source-address* [*prefix-length*] | **vlan** *vlan-id*] *

IPv6 multicast VLAN configuration examples

Example: Configuring sub-VLAN-based IPv6 multicast VLAN

Network configuration

As shown in Figure 4:

- Layer 3 device Switch A runs MLD and acts as the MLD querier. Layer 2 device Switch B runs MLDv1 snooping.
- The IPv6 multicast source sends IPv6 multicast data to IPv6 multicast group FF1E::101. Receivers Host A, Host B, and Host C belong to VLAN 2, VLAN 3, and VLAN 4, respectively.

Configure a sub-VLAN-based IPv6 multicast VLAN on Switch B to meet the following requirements:

- Switch A sends the IPv6 multicast data to Switch B through the IPv6 multicast VLAN.
- Switch B forwards the IPv6 multicast data to the receivers in different user VLANs.

Figure 4 Network diagram

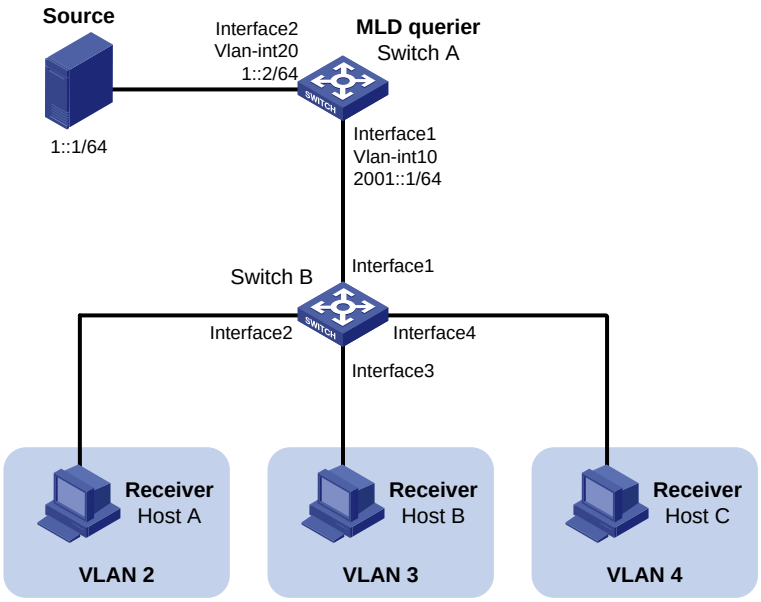


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Switch A:

Enable IPv6 multicast routing.

```
<SwitchA> system-view
[SwitchA] ipv6 multicast routing
[SwitchA-mrib6] quit
```

Create VLAN 20, and assign GigabitEthernet 1/0/2 to the VLAN.

```
[SwitchA] vlan 20
[SwitchA-vlan20] port gigabitethernet 1/0/2
[SwitchA-vlan20] quit
```

Assign an IPv6 address to VLAN-interface 20, and enable IPv6 PIM-DM on the interface.

```
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] ipv6 address 1::2 64
[SwitchA-Vlan-interface20] ipv6 pim dm
[SwitchA-Vlan-interface20] quit
```

Create VLAN 10.

```
[SwitchA] vlan 10
[SwitchA-vlan10] quit
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign the port to VLAN 10 as a tagged VLAN member.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-type hybrid
[SwitchA-GigabitEthernet1/0/1] port hybrid vlan 10 tagged
[SwitchA-GigabitEthernet1/0/1] quit
```

Assign an IPv6 address to VLAN-interface 10, and enable MLD on the interface.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ipv6 address 2001::1 64
[SwitchA-Vlan-interface10] mld enable
[SwitchA-Vlan-interface10] quit
```

2. Configure Switch B:

Enable the MLD snooping feature.

```
<SwitchB> system-view
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

Create VLAN 2, assign GigabitEthernet 1/0/2 to the VLAN, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/2
[SwitchB-vlan2] mld-snooping enable
[SwitchB-vlan2] quit
```

Create VLAN 3, assign GigabitEthernet 1/0/3 to the VLAN, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/3
[SwitchB-vlan3] mld-snooping enable
[SwitchB-vlan3] quit
```

Create VLAN 4, assign GigabitEthernet 1/0/4 to the VLAN, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 4
[SwitchB-vlan4] port gigabitethernet 1/0/4
[SwitchB-vlan4] mld-snooping enable
[SwitchB-vlan4] quit
```

Create VLAN 10, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 10
[SwitchB-vlan10] mld-snooping enable
[SwitchB-vlan10] quit
```

Configure GigabitEthernet 1/0/1 as a hybrid port, and assign the port to VLAN 10 as a tagged VLAN member.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-type hybrid
[SwitchB-GigabitEthernet1/0/1] port hybrid vlan 10 tagged
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure VLAN 10 as an IPv6 multicast VLAN, and assign VLAN 2 through VLAN 4 as sub-VLANs to multicast VLAN 10.

```
[SwitchB] ipv6 multicast-vlan 10
[SwitchB-ipv6-mvlan-10] subvlan 2 to 4
[SwitchB-ipv6-mvlan-10] quit
```

Verifying the configuration

Display information about all IPv6 multicast VLANs on Switch B.

```
[SwitchB] display ipv6 multicast-vlan
Total 1 IPv6 multicast VLANs.

IPv6 multicast VLAN 10:
  Sub-VLAN list(3 in total):
     2-4
  Port list(0 in total):
```

Display information about IPv6 multicast groups in IPv6 multicast VLANs on Switch B.

```
[SwitchB] display ipv6 multicast-vlan group
Total 1 entries.

IPv6 multicast VLAN 10: Total 1 entries.
  (::, FF1E::101)
    Sub-VLANs (3 in total):
      VLAN 2
      VLAN 3
      VLAN 4
```

The output shows that IPv6 multicast group FF1E::101 belongs to IPv6 multicast VLAN 10. IPv6 multicast VLAN 10 contains sub-VLANs VLAN 2 through VLAN 4. Switch B will replicate the IPv6 multicast data of VLAN 10 to VLAN 2 through VLAN 4.

Example: Configuring port-based IPv6 multicast VLAN

Network configuration

As shown in Figure 5:

- Layer 3 device Switch A runs MLDv1 and acts as the MLD querier. Layer 2 device Switch B runs MLDv1 snooping.
- The IPv6 multicast source sends IPv6 multicast data to IPv6 multicast group FF1E::101. Receivers Host A, Host B, and Host C belong to VLAN 2, VLAN 3, and VLAN 4, respectively.

Configure a port-based IPv6 multicast VLAN on Switch B to meet the following requirements:

- Switch A sends IPv6 multicast data to Switch B through the IPv6 multicast VLAN.
- Switch B forwards the IPv6 multicast data to the receivers in different user VLANs.

Figure 5 Network diagram

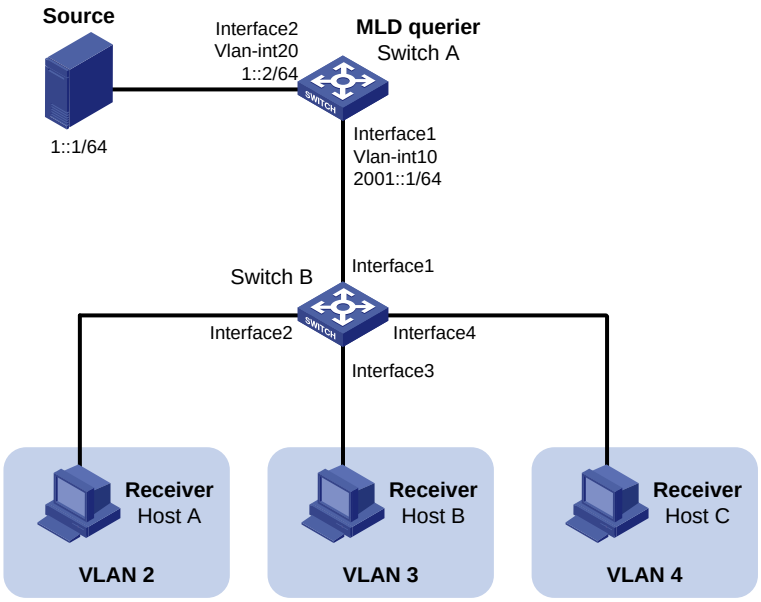


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Switch A:

   # Enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Create VLAN 20, and assign GigabitEthernet 1/0/2 to the VLAN.

   ```
   [SwitchA] vlan 20
   [SwitchA-vlan20] port gigabitethernet 1/0/2
   [SwitchA-vlan20] quit
   ```

   # Assign an IPv6 address to VLAN-interface 20, and enable IPv6 PIM-DM on the interface.

   ```
   [SwitchA] interface vlan-interface 20
   [SwitchA-Vlan-interface20] ipv6 address 1::2 64
   [SwitchA-Vlan-interface20] ipv6 pim dm
   [SwitchA-Vlan-interface20] quit
   ```

   # Create VLAN 10, and assign GigabitEthernet 1/0/1 to the VLAN.

   ```
   [SwitchA] vlan 10
   [SwitchA-vlan10] port gigabitethernet 1/0/1
   [SwitchA-vlan10] quit
   ```

   # Assign an IPv6 address to VLAN-interface 10, and enable MLD on the interface.

```
[SwitchA] interface vlan-interface 10
[SwitchA-Vlan-interface10] ipv6 address 2001::1 64
[SwitchA-Vlan-interface10] mld enable
[SwitchA-Vlan-interface10] quit
```

2. Configure Switch B:

Enable the MLD snooping feature.

```
<SwitchB> system-view
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

Create VLAN 10, assign GigabitEthernet 1/0/1 to the VLAN, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 10
[SwitchB-vlan10] port gigabitethernet 1/0/1
[SwitchB-vlan10] mld-snooping enable
[SwitchB-vlan10] quit
```

Create VLAN 2, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 2
[SwitchB-vlan2] mld-snooping enable
[SwitchB-vlan2] quit
```

Create VLAN 3, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 3
[SwitchB-vlan3] mld-snooping enable
[SwitchB-vlan3] quit
```

Create VLAN 4, and enable MLD snooping for the VLAN.

```
[SwitchB] vlan 4
[SwitchB-vlan4] mld-snooping enable
[SwitchB-vlan4] quit
```

Configure GigabitEthernet 1/0/2 as a hybrid port, and configure VLAN 2 as the PVID of the hybrid port.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-type hybrid
[SwitchB-GigabitEthernet1/0/2] port hybrid pvid vlan 2
```

Assign GigabitEthernet 1/0/2 to VLAN 2 and VLAN 10 as an untagged VLAN member.

```
[SwitchB-GigabitEthernet1/0/2] port hybrid vlan 2 untagged
[SwitchB-GigabitEthernet1/0/2] port hybrid vlan 10 untagged
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a hybrid port, and configure VLAN 3 as the PVID of the hybrid port.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port link-type hybrid
[SwitchB-GigabitEthernet1/0/3] port hybrid pvid vlan 3
```

Assign GigabitEthernet 1/0/3 to VLAN 3 and VLAN 10 as an untagged VLAN member.

```
[SwitchB-GigabitEthernet1/0/3] port hybrid vlan 3 untagged
[SwitchB-GigabitEthernet1/0/3] port hybrid vlan 10 untagged
[SwitchB-GigabitEthernet1/0/3] quit
```

Configure GigabitEthernet 1/0/4 as a hybrid port, and configure VLAN 4 as the PVID of the hybrid port.

```
[SwitchB] interface gigabitethernet 1/0/4
```

```
[SwitchB-GigabitEthernet1/0/4] port link-type hybrid
[SwitchB-GigabitEthernet1/0/4] port hybrid pvid vlan 4
```

Assign GigabitEthernet 1/0/4 to VLAN 4 and VLAN 10 as an untagged VLAN member.

```
[SwitchB-GigabitEthernet1/0/4] port hybrid vlan 4 untagged
[SwitchB-GigabitEthernet1/0/4] port hybrid vlan 10 untagged
[SwitchB-GigabitEthernet1/0/4] quit
```

Configure VLAN 10 as an IPv6 multicast VLAN.

```
[SwitchB] ipv6 multicast-vlan 10
```

Assign GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 to VLAN 10.

```
[SwitchB-ipv6-mvlan-10] port gigabitethernet 1/0/2 to gigabitethernet 1/0/3
[SwitchB-ipv6-mvlan-10] quit
```

Assign GigabitEthernet 1/0/4 to VLAN 10.

```
[SwitchB] interface gigabitethernet 1/0/4
[SwitchB-GigabitEthernet1/0/4] ipv6 port multicast-vlan 10
[SwitchB-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Display information about IPv6 multicast VLANs on Switch B.

```
[SwitchB] display ipv6 multicast-vlan
Total 1 IPv6 multicast VLANs.

IPv6 multicast VLAN 10:
  Sub-VLAN list(0 in total):
  Port list(3 in total):
    GE1/0/2
    GE1/0/3
    GE1/0/4
```

Display dynamic MLD snooping forwarding entries on Switch B.

```
[SwitchB] display mld-snooping group
Total 1 entries.

VLAN 10: Total 1 entries.
  (::, FF1E::101)
    Host ports (3 in total):
      GE1/0/2         (00:03:23)
      GE1/0/3         (00:04:07)
      GE1/0/4         (00:04:16)
```

The output shows that MLD snooping maintains the user ports in the multicast VLAN (VLAN 10). Switch B will forward the IPv6 multicast data of VLAN 10 through these user ports.

Contents

IPv6 PIM overview 1
IPv6 PIM modes 1
IPv6 PIM-DM 1
Neighbor discovery 1
SPT building 1
Graft 2
Assert 3
IPv6 PIM-SM 3
Neighbor discovery 4
DR election 4
RP discovery 4
Embedded RP 6
Anycast RP 7
RPT building 8
IPv6 multicast source registration 8
Switchover to SPT 9
Assert 10
IPv6 BIDIR-PIM 10
Neighbor discovery 10
RP discovery 10
DF election 11
Bidirectional RPT building 11
IPv6 administrative scoping 13
IPv6 administrative scoping mechanism 13
Relationship between IPv6 admin-scoped zones and the IPv6 global-scoped zone 14
IPv6 PIM-SSM 15
Neighbor discovery 15
DR election 16
SPT building 16
Relationship among IPv6 PIM protocols 16
IPv6 PIM support for VPNs 17
Protocols and standards 17
Configuring IPv6 PIM 18
Restrictions and guidelines: IPv6 PIM configuration 18
Configuring IPv6 PIM-DM 18
IPv6 PIM-DM tasks at a glance 18
Prerequisites for IPv6 PIM-DM 18
Enabling IPv6 PIM-DM 18
Configuring the state refresh feature 19
Setting the IPv6 PIM-DM graft retry timer 20
Configuring IPv6 PIM-SM 20
IPv6 PIM-SM tasks at a glance 20
Prerequisites for IPv6 PIM-SM 21
Enabling IPv6 PIM-SM 21
Configuring static RPs 21
Configuring C-RPs 22
Configuring C-BSRs 23
Configuring the interval for C-BSRs to send BSMs 23
Configuring the timer for C-BSRs to wait BSMs from the BSR 24
Configuring an IPv6 PIM domain border 24
Disabling BSM semantic fragmentation 25
Disabling the device from forwarding BSMs out of their incoming interfaces 25
Enabling embedded RP 26
Configuring Anycast RP 26
Configuring IPv6 multicast source registration 27
Configuring the switchover to SPT 28

Configuring the delay time for DR election 28
Configuring IPv6 BIDIR-PIM 28
IPv6 BIDIR-PIM tasks at a glance 28
Prerequisites for IPv6 BIDIR-PIM 29
Enabling IPv6 BIDIR-PIM 29
Configuring static RPs 30
Configuring C-RPs 30
Configuring C-BSRs 31
Configuring the interval for C-BSRs to send BSMs 31
Configuring the timer for C-BSRs to wait BSMs from the BSR 32
Configuring an IPv6 PIM domain border 33
Disabling BSM semantic fragmentation 33
Disabling the device from forwarding BSMs out of their incoming interfaces 34
Setting the maximum number of IPv6 BIDIR-PIM RPs 34
Configuring IPv6 PIM-SSM 34
IPv6 PIM-SSM tasks at a glance 34
Prerequisites for IPv6 PIM-SSM 35
Enabling IPv6 PIM-SM 35
Specifying the IPv6 SSM group range 35
Configuring the delay time for DR election 36
Configuring common IPv6 PIM features 36
Common IPv6 PIM feature tasks at a glance 36
Configuring an IPv6 multicast source policy 37
Configuring an IPv6 PIM hello policy 37
Configuring IPv6 PIM hello message options 38
Dropping hello messages without the Generation ID option 39
Enabling the device to send hello messages with a different Generation ID upon a port state change 40
Configuring common IPv6 PIM timers 41
Setting the maximum size of a join or prune message 42
Configuring an IPv6 PIM join policy 42
Setting the DSCP value for outgoing IPv6 PIM protocol packets 43
Enabling BFD for IPv6 PIM 43
Enabling IPv6 PIM passive mode 44
Enabling IPv6 PIM NSR 44
Enabling SNMP notifications for IPv6 PIM 44
Enabling IPv6 PIM logging 45
Enabling log message suppression for IPv6 PIM 45
Display and maintenance commands for IPv6 PIM 45
IPv6 PIM configuration examples 46
Example: Configuring IPv6 PIM-DM 46
Example: Configuring non-scoped IPv6 PIM-SM 50
Example: Configuring admin-scoped IPv6 PIM-SM 53
Example: Configuring IPv6 BIDIR-PIM 58
Example: Configuring IPv6 PIM-SSM 62
Troubleshooting IPv6 PIM 65
A multicast distribution tree cannot be correctly built 65
IPv6 multicast data is abnormally terminated on an intermediate device 66
An RP cannot join an SPT in IPv6 PIM-SM 66
An RPT cannot be built or IPv6 multicast source registration fails in IPv6 PIM-SM 66

IPv6 PIM overview

IPv6 Protocol Independent Multicast (IPv6 PIM) provides IPv6 multicast forwarding by leveraging IPv6 unicast static routes or IPv6 unicast routing tables generated by any IPv6 unicast routing protocol. IPv6 PIM uses the underlying IPv6 unicast routing to generate an IPv6 multicast routing table without relying on any particular IPv6 unicast routing protocol.

IPv6 PIM uses the RPF mechanism to implement IPv6 multicast forwarding. For more information about RPF, see "Configuring IPv6 multicast routing and forwarding."

IPv6 PIM modes

Based on the implementation mechanism, IPv6 PIM includes the following modes:

- IPv6 Protocol Independent Multicast–Dense Mode (IPv6 PIM-DM).
- IPv6 Protocol Independent Multicast–Sparse Mode (IPv6 PIM-SM).
- IPv6 Bidirectional Protocol Independent Multicast (IPv6 BIDIR-PIM).
- IPv6 Protocol Independent Multicast Source-Specific Multicast (IPv6 PIM-SSM).

In this document, an IPv6 PIM domain refers to a network composed of IPv6 PIM devices.

IPv6 PIM-DM

IPv6 PIM-DM uses the push mode for multicast forwarding and is suitable for small networks with densely distributed IPv6 multicast members.

IPv6 PIM-DM assumes that all downstream nodes want to receive IPv6 multicast data from a source, so IPv6 multicast data is flooded to all downstream nodes on the network. Branches without downstream receivers are pruned from the forwarding trees, leaving only those branches that contain receivers. When the pruned branch has new receivers, the graft mechanism turns the pruned branch into a forwarding branch.

In IPv6 PIM-DM, the multicast forwarding paths for an IPv6 multicast group constitute a forwarding tree. The forwarding tree is rooted at the IPv6 multicast source and has multicast group members as its "leaves." Because the forwarding tree consists of the shortest paths from the IPv6 multicast source to the receivers, it is also called a "shortest path tree (SPT)."

IPv6 PIM-DM mechanisms include neighbor discovery, SPT building, graft, and assert.

Neighbor discovery

In an IPv6 PIM domain, each IPv6 PIM interface periodically multicasts IPv6 PIM hello messages to all other IPv6 PIM devices on the local subnet. Through the exchanging of hello messages, all IPv6 PIM devices determine their IPv6 PIM neighbors, maintain IPv6 PIM neighboring relationship with other devices, and build and maintain SPTs.

SPT building

The process of building an SPT is the flood-and-prune process:

1. In an IPv6 PIM-DM domain, the IPv6 multicast data from the IPv6 multicast source S to the IPv6 multicast group G is flooded throughout the domain. A device performs an RPF check on the IPv6 multicast data. If the check succeeds, the device creates an (S, G) entry and forwards the data to all downstream nodes in the network. In the flooding process, all the devices in the IPv6 PIM-DM domain create the (S, G) entry.

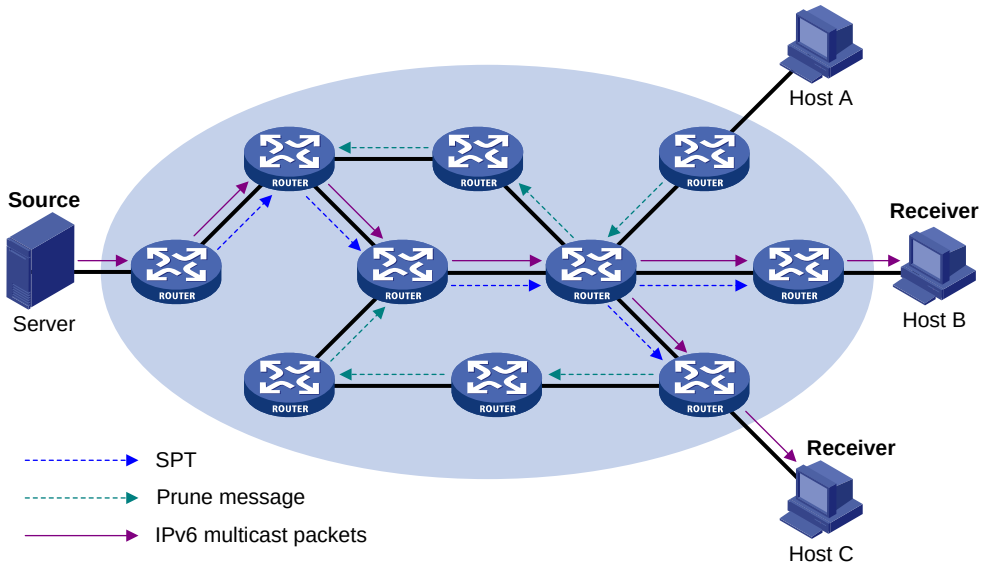
2. The nodes without downstream receivers are pruned. A device that has no downstream receivers multicasts a prune message to all IPv6 PIM devices on the subnet. When the upstream node receives the prune message, it removes the receiving interface from the (S, G) entry. In this way, the upstream stream node stops forwarding subsequent packets addressed to that IPv6 multicast group down to this node.

NOTE:

An (S, G) entry contains an IPv6 multicast source address S, an IPv6 multicast group address G, an outgoing interface list, and an incoming interface.

A prune process is initiated by a leaf device. As shown in Figure 1, the device interface that does not have any downstream receivers initiates a prune process by sending a prune message toward the IPv6 multicast source. This prune process goes on until only necessary branches are left in the IPv6 PIM-DM domain, and these necessary branches constitute an SPT.

Figure 1 SPT building



The pruned state of a branch has a finite holdtime timer. When the timer expires, IPv6 multicast data is again forwarded to the pruned branch. The flood-and-prune cycle takes place periodically to maintain the forwarding branches.

Graft

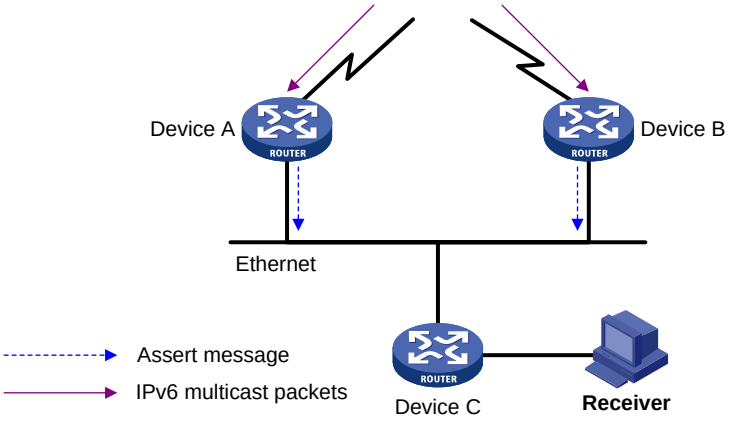
A previously pruned branch might have new downstream receivers. To reduce the latency for resuming the forwarding capability of this branch, a graft mechanism is used as follows:

1. The node that needs to receive the IPv6 multicast data sends a graft message to its upstream node, telling it to rejoin the SPT.
2. After receiving this graft message on an interface, the upstream node adds the receiving interface to the outgoing interface list of the (S, G) entry. It also sends a graft-ack message to the graft sender.
3. If the graft sender receives a graft-ack message, the graft process finishes. Otherwise, the graft sender continues to send graft messages at a graft retry interval until it receives an acknowledgment from its upstream node.

Assert

On a subnet with more than one multicast device, the assert mechanism shuts off duplicate multicast flows to the network. It does this by electing a unique multicast forwarder for the subnet.

Figure 2 Assert mechanism



As shown in Figure 2, after Device A and Device B receive an (S, G) packet from the upstream node, they both forward the packet to the local subnet. As a result, the downstream node Device C receives two identical multicast packets. In addition, both Device A and Device B, on their downstream interfaces, receive a duplicate packet forwarded by the other. After detecting this condition, both devices send an assert message to all IPv6 PIM devices on the local subnet through the interface that received the packet. The assert message contains the IPv6 multicast source address (S), the IPv6 multicast group address (G), and the metric preference and metric of the IPv6 unicast route/MBGP route/static multicast route to the IPv6 multicast source. By comparing these parameters, either Device A or Device B becomes the unique forwarder of the subsequent (S, G) packets on the subnet. The comparison process is as follows:

2. The device with a higher metric preference to the IPv6 multicast source wins.
3. If both devices have the same metric preference to the IPv6 multicast source, the device with a smaller metric to the IPv6 multicast source wins.
4. If both devices have the same metric, the device with a higher IPv6 link-local address on the downstream interface wins.

IPv6 PIM-SM

IPv6 PIM-DM uses the flood-and-prune cycles to build SPTs for IPv6 multicast data forwarding. Although an SPT has the shortest paths from the IPv6 multicast source to the receivers, it is built with a low efficiency. Therefore, IPv6 PIM-DM is not suitable for large- and medium-sized networks.

IPv6 PIM-SM uses the pull mode for IPv6 multicast forwarding, and it is suitable for large-sized and medium-sized networks with sparsely and widely distributed IPv6 multicast group members.

IPv6 PIM-SM assumes that no hosts need IPv6 multicast data. A multicast receiver must express its interest in the IPv6 multicast data for an IPv6 multicast group before the data is forwarded to it. A rendezvous point (RP) is the core of an IPv6 PIM-SM domain. Relying on the RP, SPTs and rendezvous point trees (RPTs) are established and maintained to implement IPv6 multicast data forwarding. An SPT is rooted at the IPv6 multicast source and has the RPs as its leaves. An RPT is rooted at the RP and has the receiver hosts as its leaves.

IPv6 PIM-SM mechanisms include neighbor discovery, DR election, RP discovery, Anycast RP, RPT building, multicast source registration, switchover to SPT, and assert.

Neighbor discovery

IPv6 PIM-SM uses the same neighbor discovery mechanism as IPv6 PIM-DM does. For more information, see "Neighbor discovery."

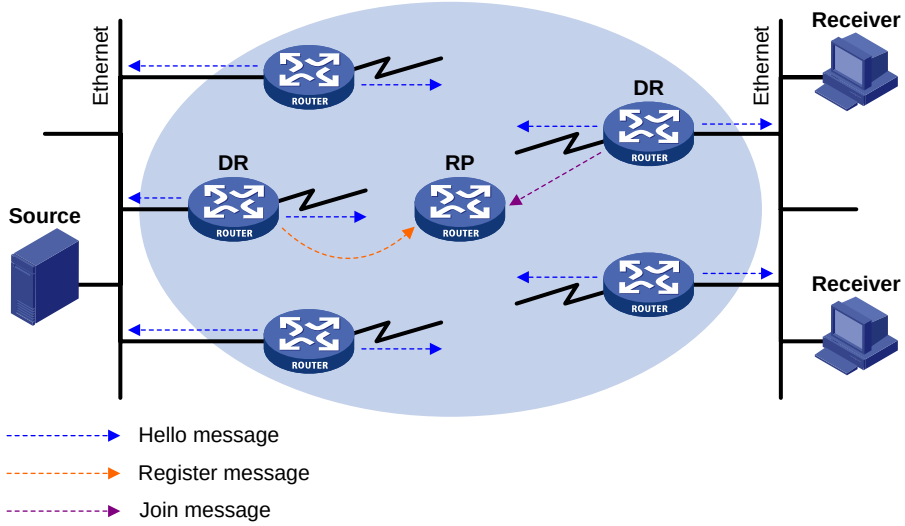
DR election

A designated router (DR) is required on both the source-side network and receiver-side network. A source-side DR acts on behalf of the IPv6 multicast source to send register messages to the RP. The receiver-side DR acts on behalf of the receiver hosts to send join messages to the RP.

IMPORTANT:

MLD must be enabled on the device that acts as the receiver-side DR. Otherwise, the receiver hosts attached to the DR cannot join any IPv6 multicast groups. For more information about MLD, see "Configuring MLD."

Figure 3 DR election



As shown in Figure 3, the DR election process is as follows:

2. The devices on the shared-media LAN send hello messages to one another. The hello messages contain the DR priority for DR election. The device with the highest DR priority is elected as the DR.
3. The device with the highest IPv6 link-local address wins the DR election under one of the following conditions:
   - All the devices have the same DR election priority.
   - A device does not support carrying the DR priority in hello messages.

If the DR fails, its IPv6 PIM neighbor lifetime expires, and the other devices initiate a new DR election.

RP discovery

An RP is the core of an IPv6 PIM-SM domain. An IPv6 multicast group can have only one RP for IPv6 multicast forwarding, and an RP can be designated to multiple IPv6 multicast groups.

RP selection mechanism

An RP can be statically configured, dynamically elected, or extracted from an IPv6 multicast address with embedded RP. The embedded RP takes precedence over the static RP and the dynamic RP. The priorities of the static RP and the dynamic RP depend on whether the static RP is preferred when the static RP is configured.

- If the static RP is preferred, the static RP takes precedence over the dynamic RP. The dynamic RP takes services over only when the static RP fails.
- If the static RP is not preferred, the dynamic RP takes precedence over the static RP. The static RP takes services over only when the dynamic RP fails.

Static RPs

A static RP can avoid a single point of failure and save network bandwidth that is consumed by frequent communications between C-RPs and the BSR.

Dynamic RP election

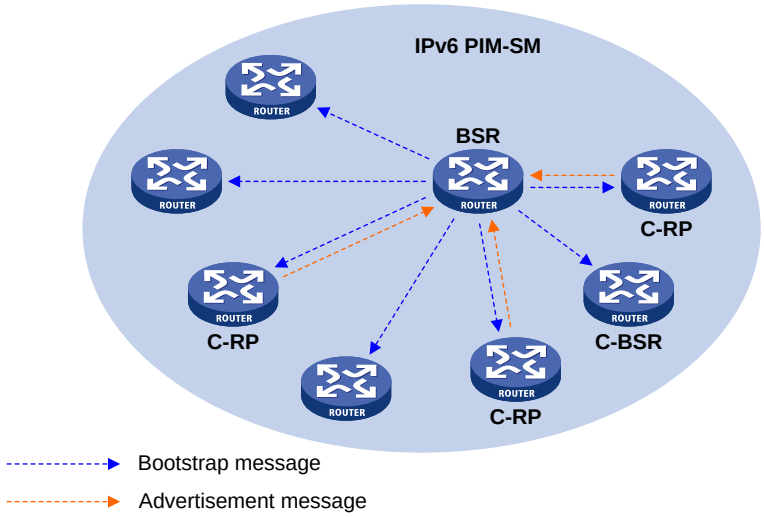
Dynamic RP election is implemented through the BSR mechanism. BSR mechanism includes the following roles:

- **Candidate-RPs (C-RPs)**—An RP is dynamically elected from C-RPs to provide services to an IPv6 multicast group.
- **BSR**—A BSR is the core of the administrative core of the IPv6 PIM-SM domain. It is responsible for collecting and advertising RP information in the whole domain. An IPv6 PIM-SM domain has only one BSR, and the BSR is elected from C-BSRs.
- **Candidate-BSRs (C-BSRs)**—Any devices in the IPv6 PIM-SM domain can act as C-BSRs and the BSR is elected from the C-BSRs. Once the BSR fails, a new BSR is elected from the C-BSRs to avoid multicast traffic interruption.

NOTE:

A device can act as a C-RP and a C-BSR at the same time.

Figure 4 Information exchange between C-RPs and BSR



As shown in Figure 4, an RP is elected as follows:

2. Each C-BSR sends a BSM to other devices in the IPv6 PIM-SM domain.
3. When a C-BSR receives a BSM from another C-BSR, it compares its own priority with the priority carried in the message. The C-BSR with a higher priority wins the BSR election. If a tie exists in the priority, the C-BSR with a higher IPv6 address wins. The loser uses the winner's

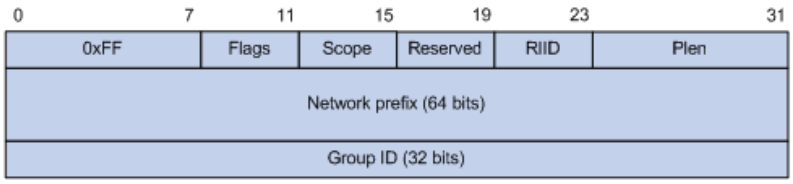
BSR address to replace its own BSR address and no longer regards itself as the BSR. The winner retains its own BSR address and continues to regard itself as the BSR.

4. Each C-RP periodically unicasts an advertisement message to the BSR. An advertisement message contains the address of the advertising C-RP and the IPv6 multicast group range to which it is designated.
5. The BSR collects these advertisement messages and organizes the C-RP information into an RP-set, which is a database of mappings between IPv6 multicast groups and RPs. The BSR encapsulates the RP-set information in the BSMs and advertises the BSMs to the entire IPv6 PIM-SM domain.
6. All devices in the IPv6 PIM-SM domain select an RP for an IPv6 multicast group based on the following rules:
   a. The C-RP that is designated to the smallest IPv6 multicast group range wins.
   b. If the C-RPs are designated to the same group ranges, the C-RP with the highest priority wins.
   c. If the C-RPs have the same priority, the C-RP with the largest hash value wins. The hash value is calculated through the hash algorithm.
   d. If the C-RPs have the same hash value, the C-RP with the highest IP address wins.

Embedded RP

The embedded RP mechanism enables a device to resolve the RP address from an IPv6 multicast group address to map the IPv6 multicast group to an RP. This RP can take the place of the configured static RP or the RP dynamically elected by the bootstrap mechanism. A DR does not need to learn the RP address beforehand. Figure 5 shows the format of an IPv6 multicast group address with embedded RP.

Figure 5 Format of an IPv6 multicast address with embedded RP



As shown in Figure 5, the requirements for each field are as follows:

- **First 8 bits**—Fixed to 0xFF.
- **Flags**—The R, P, and T bits must be set to 1.
- **Reserved**—Contains 4 bits and must be set to 0.
- **RIID**—Contains 4 bits and indicates the RP interface ID.
- **Plen**—Contains 8 bits and indicates the valid length (in bits) of the RP address prefix. This field value cannot be greater than 64 and cannot be 0.
- **Network prefix**—Contains 64 bits and indicates the prefix of the RP address. The valid length of the prefix is determined by the Plen field.
- **Group ID**—Reduced to 32 bits and indicates the ID of the IPv6 multicast group.

The process of discovering an embedded RP is as follows:

- At the receiver side, a receiver host initiates an MLD report to join an IPv6 multicast group. Upon receiving the MLD report, the receiver-side DR resolves the RP address embedded in the IPv6 multicast group address and uses the RP for the IPv6 multicast group.
- At the IPv6 multicast source side, the IPv6 multicast source sends an IPv6 multicast packet to an IPv6 multicast group. Upon receiving the IPv6 multicast packet, the source-side DR resolves

the RP address embedded in the IPv6 multicast address and uses the RP for the IPv6 multicast group.

Anycast RP

IPv6 PIM-SM requires only one active RP to serve each IPv6 multicast group. If the active RP fails, the multicast traffic might be interrupted. The Anycast RP mechanism enables redundancy backup among RPs by configuring multiple RPs with the same IPv6 address for one multicast group. An IPv6 multicast source registers with the closest RP or a receiver-side DR joins the closest RP to implement source information synchronization.

Anycast RP has the following benefits:

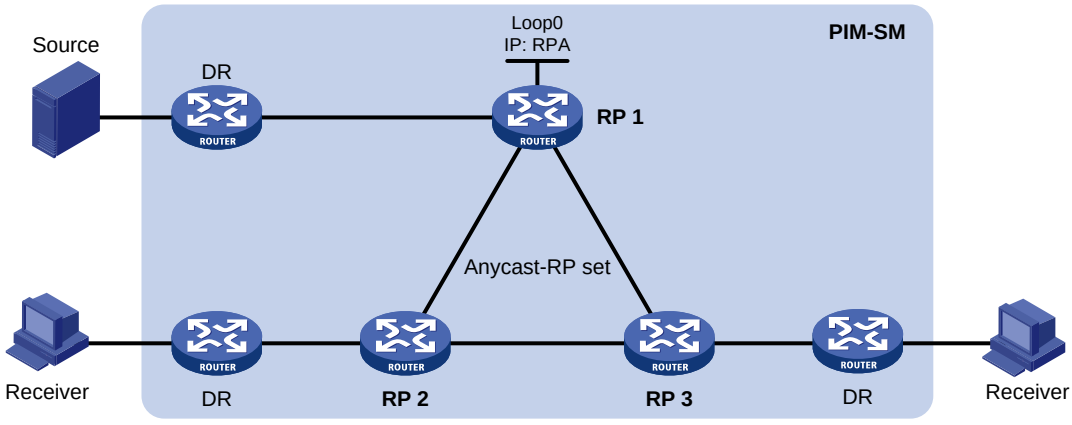
- **Optimal RP path**—An IPv6 multicast source registers with the closest RP to build an optimal SPT. A receiver joins the closest RP to build an optimal RPT.
- **Redundancy backup among RPs**—When an RP fails, the RP-related sources will register with the closest available RPs and the receiver-side DRs will join the closest available RPs. This provides redundancy backup among RPs.

Anycast RP is implemented by using either of the following methods:

- **Anycast RP through MSDP**—In this method, you can configure multiple RPs with the same IP address for one multicast group and configure MSDP peering relationships between them. For more information about Anycast RP through MSDP, see "Configuring MSDP."
- **Anycast RP through IPv6 PIM-SM**—In this method, you can configure multiple RPs for one IPv6 multicast group and add them to an Anycast RP set. This method introduces the following concepts:
  - **Anycast RP set**—A set of RPs that are designated to the same IPv6 multicast group.
  - **Anycast RP member**—Each RP in the Anycast RP set.
  - **Anycast RP member address**—IPv6 address of each Anycast RP member for communication among the RP members.
  - **Anycast RP address**—IPv6 address of the Anycast RP set for communication within the IPv6 PIM-SM domain. It is also known as RPA.

  As shown in Figure 6, RP 1, RP 2, and RP 3 are members of an Anycast RP set.

Figure 6 Anycast RP through IPv6 PIM-SM



The following describes how Anycast RP through IPv6 PIM-SM is implemented:

a. RP 1 receives a register message destined to RPA. Because the message is not from other Anycast RP members (RP 2 or RP 3), RP 1 considers that the register message is from the DR. RP 1 changes the source IPv6 address of the register message to its own IPv6 address and sends the message to the other members (RP 2 and RP 3).

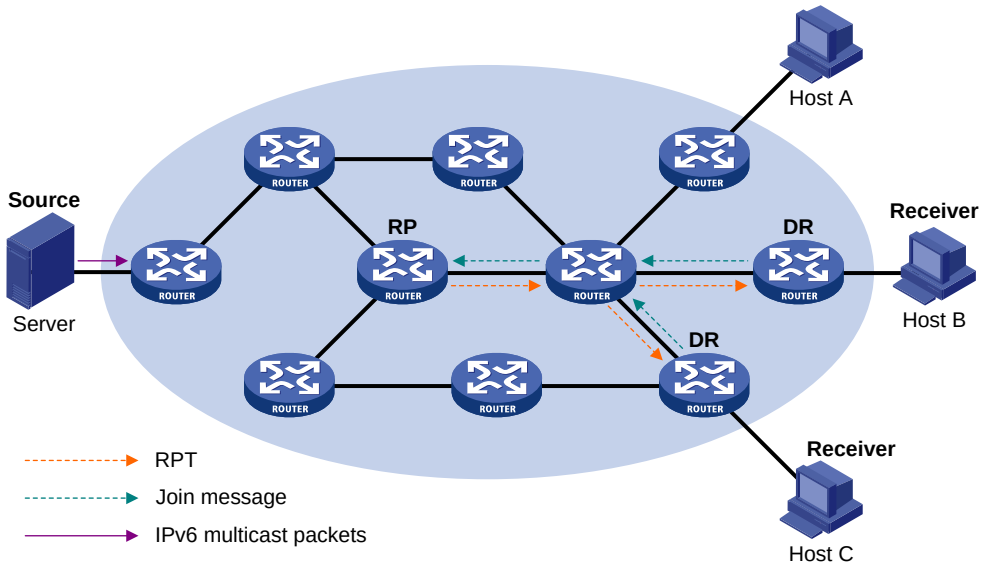
If a device acts as both a DR and an RP, it creates a register message, and then forwards the message to the other RP members.

b. After receiving the register message, RP 2 and RP 3 find out that the source address of the register message is an Anycast RP member address. They stop forwarding the message to other devices.

In Anycast RP implementation, an RP must forward the register message from the DR to other Anycast RP members to synchronize IPv6 multicast source information.

RPT building

Figure 7 RPT building in an IPv6 PIM-SM domain



As shown in Figure 7, the process of building an RPT is as follows:

2. When a receiver wants to join the IPv6 multicast group G, it uses an MLD message to inform the receiver-side DR.

3. After getting the receiver information, the DR sends a join message, which travels hop by hop to the RP for the IPv6 multicast group.

4. The devices along the path from the DR to the RP form an RPT branch. Each device on this branch adds to its forwarding table a (*, G) entry, where the asterisk (*) represents any IPv6 multicast source. The RPT is rooted at the RP and has the DR as its leaf.

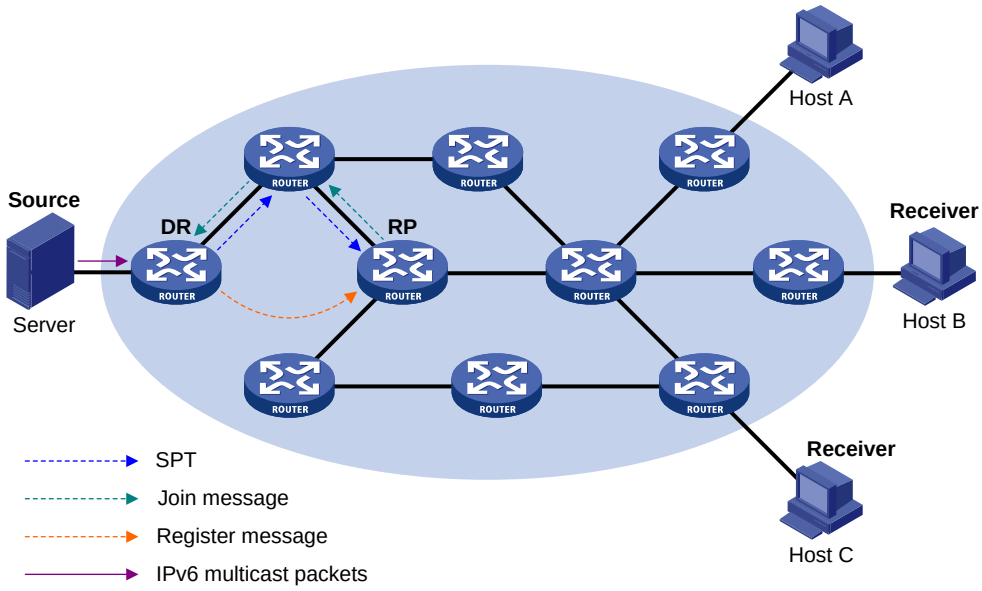
When the IPv6 multicast data addressed to the IPv6 multicast group G reaches the RP, the RP forwards the data to the DR along the established RPT. Finally, the DR forwards the IPv6 multicast data to the receiver hosts.

When a receiver is no longer interested in the IPv6 multicast data addressed to the IPv6 multicast group G, the receiver-side DR sends a prune message. The prune message goes hop by hop along the RPT to the RP. After receiving the prune message, the upstream node deletes the interface that connects to this downstream node from the outgoing interface list. At the same time, the upstream device checks for the existence of receivers for that IPv6 multicast group. If no receivers for the IPv6 multicast group exist, the device continues to forward the prune message to its upstream device.

IPv6 multicast source registration

The IPv6 multicast source uses the registration process to inform an RP of its presence.

Figure 8 IPv6 multicast source registration



As shown in Figure 8, the IPv6 multicast source registers with the RP as follows:

2. The IPv6 multicast source S sends the first multicast packet to the IPv6 multicast group G. When receiving the multicast packet, the source-side DR that directly connects to the IPv6 multicast source encapsulates the packet into a register message and unicasts the message to the RP.
3. After the RP receives the register message, it decapsulates it and forwards it down to the RPT. Meanwhile, it sends an (S, G) source-specific join message toward the IPv6 multicast source. The devices along the path from the RP to the IPv6 multicast source constitute an SPT branch. Each device on this branch creates an (S, G) entry in its forwarding table.
4. The subsequent IPv6 multicast data from the IPv6 multicast source are forwarded to the RP along the SPT. When the IPv6 multicast data reaches the RP along the SPT, the RP forwards the data to the receivers along the RPT. Meanwhile, it unicasts a register-stop message to the source-side DR to prevent the DR from unnecessarily encapsulating the data.

Switchover to SPT

In an IPv6 PIM-SM domain, only one RP and one RPT provide services for a specific IPv6 multicast group. Before the switchover to SPT occurs, the source-side DR encapsulates all IPv6 multicast data in register messages and sends them to the RP. After receiving these register messages, the RP decapsulates them and forwards them to the receiver-side DR along the RPT.

IPv6 multicast forwarding along the RPT has the following weaknesses:

- Encapsulation and decapsulation are complex on the source-side DR and the RP.
- The path for an IPv6 multicast packet might not be the shortest one.
- The RP might be overloaded by IPv6 multicast traffic bursts.

To eliminate these weaknesses, IPv6 PIM-SM allows an RP or the receiver-side DR to initiate the switchover to SPT when the RP or DR receives IPv6 multicast traffic.

- The RP initiates the switchover to SPT:

  If the RP receives IPv6 multicast traffic, it sends an (S, G) source-specific join message toward the IPv6 multicast source. The devices along the path from the RP to the IPv6 multicast source constitute an SPT branch. The subsequent IPv6 multicast data is forwarded to the RP along the SPT without being encapsulated into register messages.

For more information about the switchover to SPT initiated by the RP, see "IPv6 multicast source registration."

- The receiver-side DR initiates the switchover to SPT:

  If the receiver-side DR receives IPv6 multicast traffic, it initiates the switchover to SPT as follows:

  a. The receiver-side DR sends an (S, G) source-specific join message toward the IPv6 multicast source. The devices along the path create an (S, G) entry in their forwarding table to constitute an SPT branch.

  b. When the multicast packets reach the device where the RPT and the SPT branches, the device drops the multicast packets that travel along the RPT. It then sends a prune message with the RP bit toward the RP.

  c. After receiving the prune message, the RP forwards it toward the IPv6 multicast source (supposed only one receiver exists). Thus, the switchover to SPT is completed. The subsequent IPv6 multicast packets for the IPv6 multicast group travel along the SPT from the IPv6 multicast source to the receiver hosts.

With the switchover to SPT, IPv6 PIM-SM builds SPTs more economically than IPv6 PIM-DM does.

Assert

IPv6 PIM-SM uses a similar assert mechanism as IPv6 PIM-DM does. For more information, see "Assert."

IPv6 BIDIR-PIM

In some many-to-many applications, such as a multi-side video conference, multiple receivers of an IPv6 multicast group might be interested in the IPv6 multicast data from multiple IPv6 multicast sources. With IPv6 PIM-DM or IPv6 PIM-SM, each device along the SPT must create an (S, G) entry for each IPv6 multicast source, consuming a lot of system resources.

IPv6 BIDIR-PIM addresses the problem. Derived from IPv6 PIM-SM, IPv6 BIDIR-PIM builds and maintains a bidirectional RPT, which is rooted at the RP and connects the IPv6 multicast sources and the receivers. Along the bidirectional RPT, the IPv6 multicast sources send IPv6 multicast data to the RP, and the RP forwards the data to the receivers. Each device along the bidirectional RPT needs to maintain only one (*, G) entry, saving system resources.

IPv6 BIDIR-PIM is suitable for a network with dense IPv6 multicast sources and receivers.

IPv6 BIDIR-PIM mechanisms include neighbor discovery, RP discovery, DF election, and bidirectional RPT building.

Neighbor discovery

IPv6 BIDIR-PIM uses the same neighbor discovery mechanism as IPv6 PIM-SM does. For more information, see "Neighbor discovery."

RP discovery

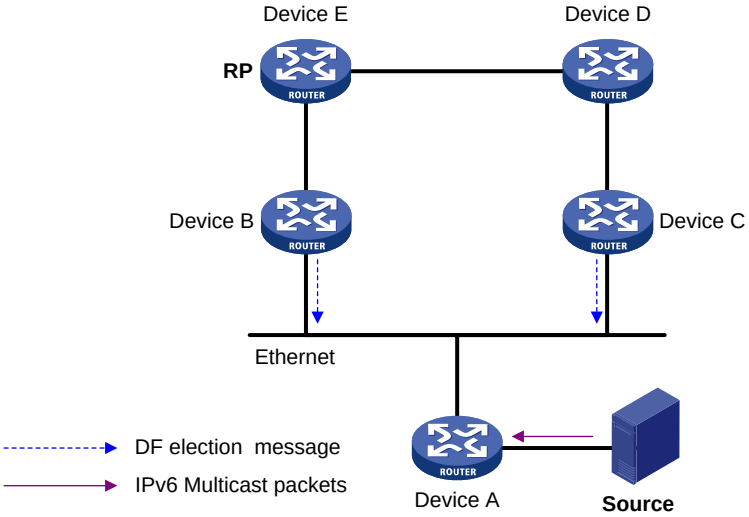
IPv6 BIDIR-PIM supports static RPs and dynamic RP election, and uses the same RP discovery mechanism as IPv6 PIM-SM, except that IPv6 BIDIR-PIM does not support embedded RP. For more information, see "RP discovery." In IPv6 BIDIR-PIM, an RPF interface is the interface toward an RP, and an RPF neighbor is the address of the next hop to the RP.

DF election

On a subnet with multiple multicast devices, duplicate multicast packets might be forwarded to the RP. To address this issue, IPv6 BIDIR-PIM uses a designated forwarder (DF) election mechanism to elect a unique DF on each subnet. Only the DFs can forward IPv6 multicast data to the RP.

DF election is not necessary for an RPL.

Figure 9 DF election



As shown in Figure 9, without the DF election mechanism, both Device B and Device C can receive IPv6 multicast packets from Device A. They also can forward the packets to downstream devices on the local subnet. As a result, the RP (Device E) receives duplicate IPv6 multicast packets.

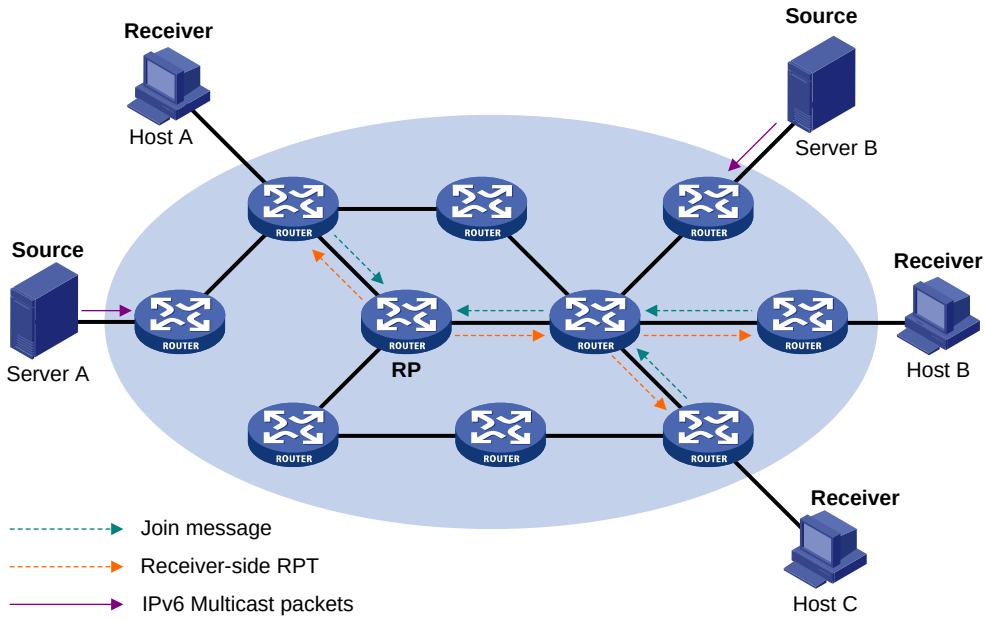
With the DF election mechanism, once receiving the RP information, Device B and Device C multicast a DF election message to all IPv6 PIM devices on the subnet to initiate a DF election process. The election message carries the RP's address, and the route preference and metric of the IPv6 unicast route, IPv6 MBGP route, or IPv6 multicast static route to the RP. A DF is elected as follows:

2. The device with higher route preference becomes the DF.
3. If the devices have the same route preference, the device with lower metric becomes the DF.
4. If the devices have the same metric, the device with the higher IP address becomes the DF.

Bidirectional RPT building

A bidirectional RPT comprises a receiver-side RPT and a source-side RPT. The receiver-side RPT is rooted at the RP and takes the devices that directly connect to the receivers as leaves. The source-side RPT is also rooted at the RP but takes the devices that directly connect to the IPv6 multicast sources as leaves. The processes for building these two RPTs are different.

Figure 10 RPT building at the receiver side

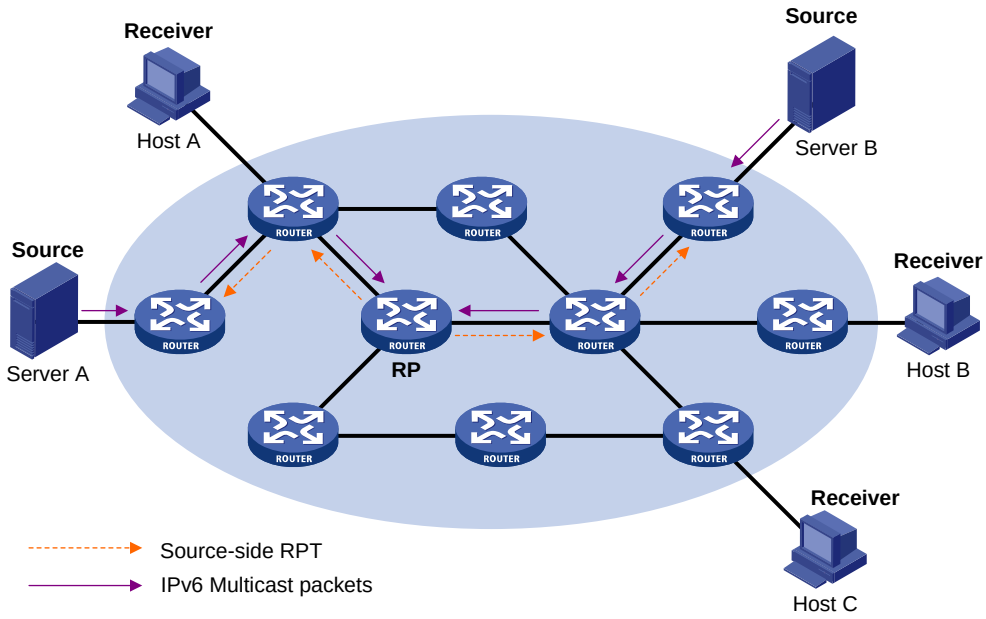


As shown in Figure 10, the process for building a receiver-side RPT is the same as the process for building an RPT in IPv6 PIM-SM:

2. When a receiver wants to join the IPv6 multicast group G, it uses an MLD message to inform the directly connected device.
3. After receiving the message, the device sends a join message, which is forwarded hop by hop to the RP for the IPv6 multicast group.
4. The devices along the path from the receiver's directly connected device to the RP form an RPT branch. Each device on this branch adds a (*, G) entry to its forwarding table.

After a receiver host leaves the IPv6 multicast group G, the directly connected device multicasts a prune message to all IPv6 PIM devices on the subnet. The prune message goes hop by hop along the reverse direction of the RPT to the RP. After receiving the prune message, an upstream node removes the interface that connects to the downstream node from the outgoing interface list. At the same time, the upstream device checks the existence of receivers for that IPv6 multicast group. If no receivers for the IPv6 multicast group exist, the device continues to forward the prune message to its upstream device.

Figure 11 RPT building at the IPv6 multicast source side



As shown in Figure 11, the process for building a source-side RPT is relatively simple:

5. When an IPv6 multicast source sends multicast packets to the IPv6 multicast group G, the DF in each subnet unconditionally forwards the packets to the RP.
6. The devices along the path from the source's directly connected device to the RP constitute an RPT branch. Each device on this branch adds a (*, G) entry to its forwarding table.

After a bidirectional RPT is built, the IPv6 multicast sources send multicast traffic to the RP along the source-side RPT. Then, the RP forwards the traffic to the receivers along the receiver-side RPT.

IMPORTANT:

If a receiver and a source are at the same side of the RP, the source-side RPT and the receiver-side RPT might meet at a node before reaching the RP. In this case, the multicast packets from the IPv6 multicast source to the receiver are directly forwarded by the node, instead of by the RP.

IPv6 administrative scoping

IPv6 administrative scoping mechanism

IPv6 administrative scoping mechanisms for IPv6 PIM-SM/IPv6 BIDIR-PIM include IPv6 non-scoped BSR mechanism and IPv6 administrative scoping mechanism.

- **IPv6 non-scoped BSR mechanism**—A domain contains only one BSR, which is responsible for advertising RP-set information within the entire domain. The information about all multicast groups is forwarded within the network that the BSR administers.
- **IPv6 administrative scoping mechanism**—To implement refined management, you can divide a domain into a global-scoped zone and multiple administratively-scoped zones (IPv6 admin-scoped zones). The administrative scoping mechanism effectively releases stress on the management in a single-BSR domain and enables provision of zone-specific services through private group addresses.

An IPv6 admin-scoped zone is designated to particular IPv6 multicast groups with the same scope field value in their group addresses. Zone border routers (ZBRs) form the boundary of an IPv6 admin-scoped zone. Each IPv6 admin-scoped zone maintains one BSR for IPv6 multicast groups

with the same scope field value. IPv6 multicast protocol packets, such as assert messages and BSMs, of these IPv6 multicast groups cannot cross the boundary of the IPv6 admin-scoped zone for the group range. The IPv6 multicast group ranges to which different IPv6 admin-scoped zones are designated can have intersections. However, the IPv6 multicast groups in an IPv6 admin-scoped zone are valid only within its local zone, and theses IPv6 multicast groups are regarded as private group addresses.

The IPv6 global-scoped zone can be regarded as a special IPv6 admin-scoped zone, and it maintains a BSR for the IPv6 multicast groups with the scope field value as 14.

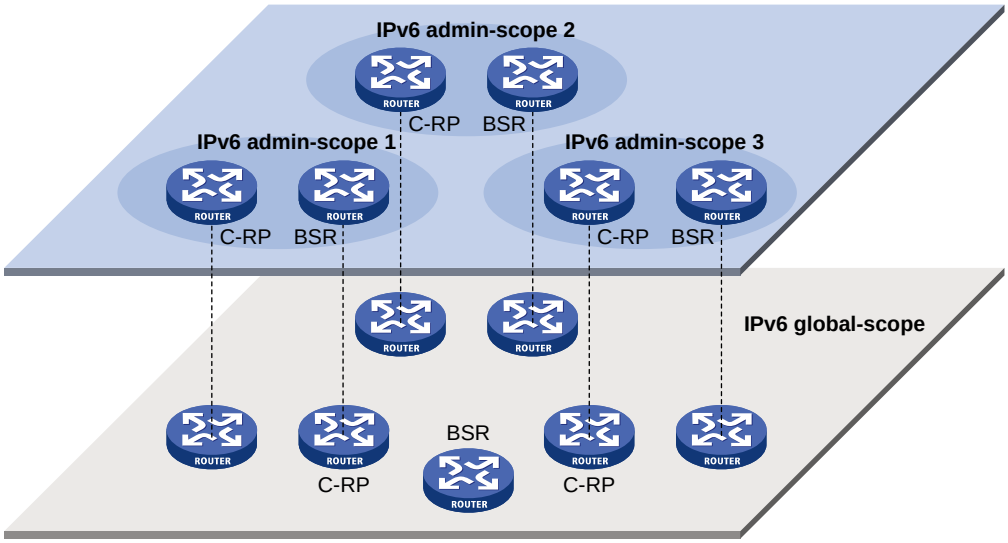
Relationship between IPv6 admin-scoped zones and the IPv6 global-scoped zone

The IPv6 global-scoped zone and each IPv6 admin-scoped zone have their own C-RPs and BSRs. These devices are effective only on their respective zones, and the BSR election and the RP election are implemented independently. Each IPv6 admin-scoped zone has its own boundary. The IPv6 multicast information within a zone cannot cross this boundary in either direction. You can have a better understanding of the IPv6 global-scoped zone and IPv6 admin-scoped zones based on geographical locations and the scope field values.

- In view of geographical locations:

  An IPv6 admin-scoped zone is a logical zone for particular IPv6 multicast groups with the same scope field value. The IPv6 multicast packets for such IPv6 multicast groups are confined within the local IPv6 admin-scoped zone and cannot cross the boundary of the zone.

  Figure 12 Relationship in view of geographical locations

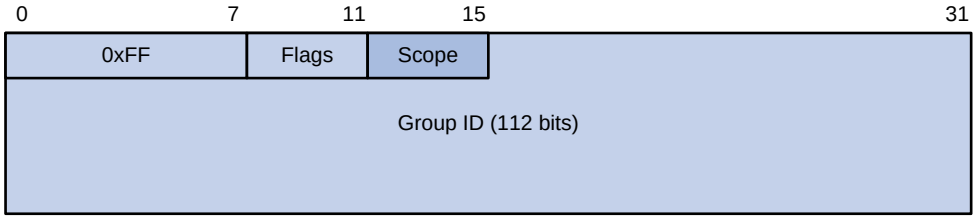


  As shown in Figure 12, for the IPv6 multicast groups with the same scope field value, the IPv6 admin-scoped zones must be geographically separated and isolated. The IPv6 global-scoped zone includes all devices in the IPv6 PIM-SM domain or IPv6 BIDIR-PIM domain. IPv6 multicast packets that do not belong to any IPv6 admin-scoped zones are forwarded in the entire IPv6 PIM-SM domain or IPv6 BIDIR-PIM domain.

- In view of the scope field values:

  In terms of the scope field values, the scope field in an IPv6 multicast group address shows the zone to which the IPv6 multicast group belongs.

Figure 13 IPv6 multicast address format



An IPv6 admin-scoped zone with a larger scope field value contains an IPv6 admin-scoped zone with a smaller scope field value. The zone with the scope field value of E is the IPv6 global-scoped zone. Table 1 lists the possible values of the scope field.

Table 1 Values of the Scope field

Value	Meaning	Remarks
0, F	Reserved	N/A
1	Interface-local scope	N/A
2	Link-local scope	N/A
3	Subnet-local scope	IPv6 admin-scoped zone.
4	Admin-local scope	IPv6 admin-scoped zone.
5	Site-local scope	IPv6 admin-scoped zone.
6, 7, 9 through D	Unassigned	IPv6 admin-scoped zone.
8	Organization-local scope	IPv6 admin-scoped zone.
E	Global scope	IPv6 global-scoped zone.

IPv6 PIM-SSM

The ASM model includes IPv6 PIM-DM and IPv6 PIM-SM. The SSM model can be implemented by leveraging part of the IPv6 PIM-SM technique. It is also called "IPv6 PIM-SSM."

The SSM model provides a solution for source-specific multicast. It maintains the relationship between hosts and devices through MLDv2.

In actual applications, part of MLDv2 and IPv6 PIM-SM techniques are adopted to implement the SSM model. In the SSM model, because receivers have located an IPv6 multicast source, no RP or RPT is required. No source registration process is required for discovering IPv6 multicast sources in other IPv6 PIM domains.

IPv6 PIM-SSM mechanisms include neighbor discovery, DR election, and SPT building.

Neighbor discovery

IPv6 PIM-SSM uses the same neighbor discovery mechanism as IPv6 PIM-SM. For more information, see "Neighbor discovery."

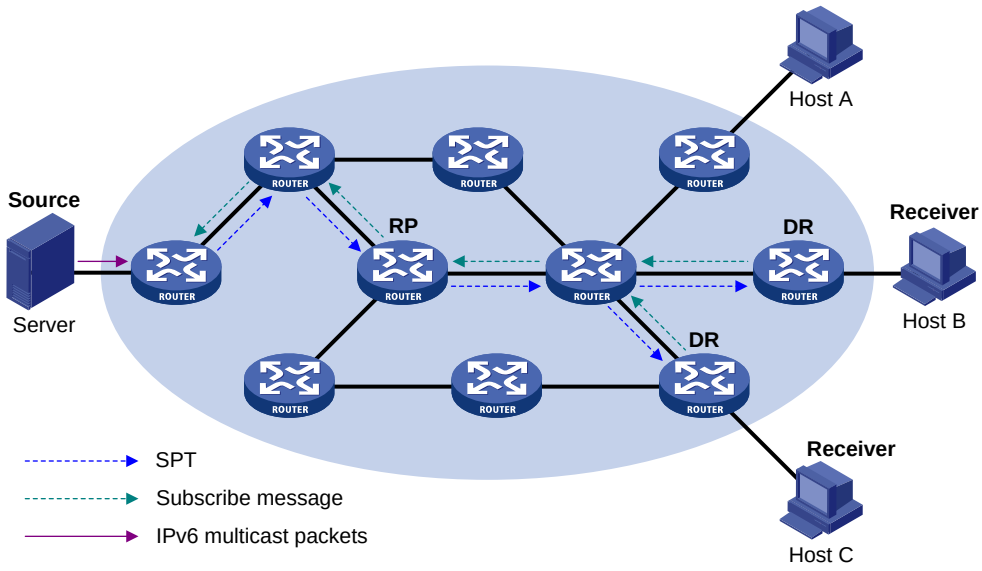
DR election

IPv6 PIM-SSM uses the same DR election mechanism as IPv6 PIM-SM. For more information, see "DR election."

SPT building

The decision to build an RPT for IPv6 PIM-SM or an SPT for IPv6 PIM-SSM depends on whether the IPv6 multicast group that the receiver host joins is in the IPv6 SSM group range. The IPv6 SSM group range reserved by IANA is FF3x::/32, where "x" represents any legal address scope.

Figure 14 SPT building in IPv6 PIM-SSM



As shown in Figure 14, Host B and Host C are receivers. They send MLDv2 report messages to their DRs to express their interest in the multicast information that the IPv6 multicast source S sends to the IPv6 multicast group G.

After receiving a report message, the DR first checks whether the group address in the message is in the IPv6 SSM group range and does the following:

- If the group address is in the IPv6 SSM group range, the DR sends a subscribe message hop by hop toward the IPv6 multicast source S. All devices along the path from the DR to the IPv6 multicast source create an (S, G) entry to build an SPT. The SPT is rooted the IPv6 multicast source S and has the receivers as its leaves. This SPT is the transmission channel in IPv6 PIM-SSM.
- If the group address is not in the IPv6 SSM group range, the receiver-side DR sends a (*, G) join message to the RP. The IPv6 multicast source registers with the source-side DR.

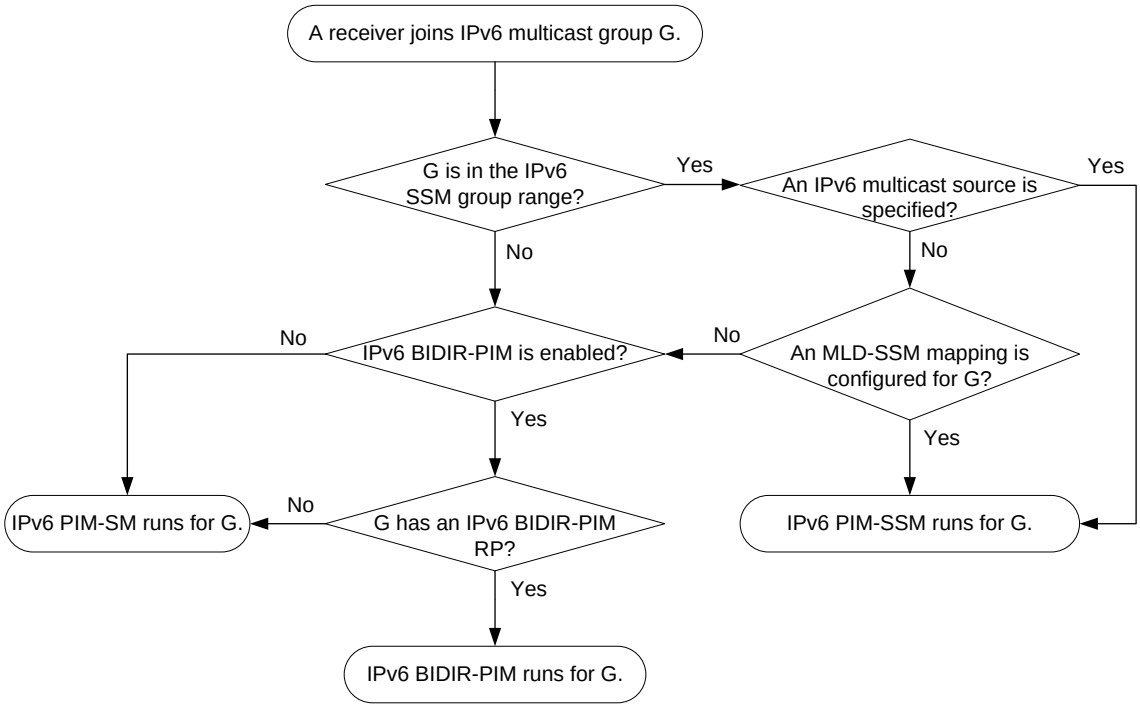
In IPv6 PIM-SSM, the term "subscribe message" refers to a join message.

Relationship among IPv6 PIM protocols

In an IPv6 PIM network, IPv6 PIM-DM cannot run together with IPv6 PIM-SM, IPv6 BIDIR-PIM, or IPv6 PIM-SSM. However, IPv6 PIM-SM, IPv6 BIDIR-PIM, and IPv6 PIM-SSM can run together. Figure 15 shows how the device selects one protocol from among them for a receiver trying to join a group.

For more information about MLD SSM mapping, see "Configuring MLD."

Figure 15 Relationship among IPv6 PIM protocols



IPv6 PIM support for VPNs

To support IPv6 PIM for VPNs, a multicast device that runs IPv6 PIM maintains an independent set of IPv6 PIM neighbor table, IPv6 multicast routing table, BSR information, and RP-set information for each VPN.

After receiving an IPv6 multicast data packet, the multicast device checks which VPN the IPv6 data packet belongs to. Then, the device forwards the IPv6 packet according to the IPv6 multicast routing table for that VPN or creates an IPv6 multicast routing entry for that VPN.

Protocols and standards

- RFC 3973, *Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification(Revised)*
- RFC 4601, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised)*
- RFC 4610, *Anycast-RP Using Protocol Independent Multicast (PIM)*
- RFC 3956, *Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address*
- RFC 5015, *Bidirectional Protocol Independent Multicast (BIDIR-PIM)*
- RFC 5059, *Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)*
- RFC 4607, *Source-Specific Multicast for IP*
- Draft-ietf-ssm-overview-05, *An Overview of Source-Specific Multicast (SSM)*

Configuring IPv6 PIM

Restrictions and guidelines: IPv6 PIM configuration

All the interfaces on a device must operate in the same IPv6 PIM mode on the public network or the same VPN instance.

To avoid IPv6 PIM routing table exceptions, do not use the same RP to provide services for IPv6 PIM-SM and IPv6 BIDIR-PIM when both of them run on the IPv6 PIM network.

Configuring IPv6 PIM-DM

IPv6 PIM-DM tasks at a glance

To configure IPv6 PIM-DM, perform the following tasks:

1. Enabling IPv6 PIM-DM
2. (Optional.) Configuring the state refresh feature
3. (Optional.) Setting the IPv6 PIM-DM graft retry timer
4. (Optional.) Configuring common IPv6 PIM features

Prerequisites for IPv6 PIM-DM

Before you configure IPv6 PIM-DM, you must complete the following tasks:

- Configure an IPv6 unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable IPv6 multicast routing.

Enabling IPv6 PIM-DM

About this task

With IPv6 PIM-DM enabled on interfaces, devices can establish IPv6 PIM neighbor relationship and process IPv6 PIM messages from their IPv6 PIM neighbors.

Restrictions and guidelines

As a best practice, enable IPv6 PIM-DM on all non-border interfaces of devices when you deploy an IPv6 PIM-DM domain.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 multicast routing and enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
   By default, IPv6 multicast routing is disabled.
   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.

quit

4. Enter interface view.
   interface *interface-type interface-number*
5. Enable IPv6 PIM-DM.
   ipv6 pim dm
   By default, IPv6 PIM-DM is disabled.

Configuring the state refresh feature

About this task

- **State refresh capability**—Enables the IPv6 PIM device that is directly connected to the source to periodically send state refresh messages. It also enables other PIM devices to refresh pruned state timers after receiving the state refresh messages. Use this feature to prevent the pruned interfaces from resuming multicast forwarding.
- **State refresh interval**—Determines the interval at which a device sends state refresh messages.
- **Wait time before accepting a new state refresh message**—A device might receive duplicate state refresh messages within a short time. To prevent this situation, you can adjust the amount of time that the device must wait to accept a new state refresh message. If the device receives a new state refresh message before the timer expires, it discards the message. If the device receives a new state refresh message after the timer expires, it accepts the message, refreshes its own IPv6 PIM-DM state, and resets the waiting timer.
- **Hop limit value of state refresh messages**—The hop limit value of a state refresh message decrements by 1 whenever it passes a device before it is forwarded to the downstream node. The state refresh message is not forwarded when the hop limit comes down to 0. A state refresh message with a large hop limit value might cycle on a small network. To control the propagation scope of state refresh messages, configure an appropriate hop limit value based on the network size on the device directly connected with the IPv6 multicast source.

Restrictions and guidelines

Perform this task on all devices in the IPv6 PIM-DM domain.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the state refresh feature.
   ipv6 pim state-refresh-capable
   By default, the state refresh feature is enabled.
4. Return to system view.
   quit
5. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
6. Set the state refresh interval.
   state-refresh-interval *interval*
   The default setting is 60 seconds.
7. Set the amount of time that the device must wait to accept a new state refresh message.
   state-refresh-rate-limit *time*

The default setting is 30 seconds.

8. Set the hop limit value of state refresh messages.
   state-refresh-hoplimit *hoplimit-value*
   The default setting is 255.

Setting the IPv6 PIM-DM graft retry timer

About this task

Perform this task to adjust the interval at which the device retransmits a graft message if it does not receive a graft-ack message from the upstream device.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the IPv6 PIM-DM graft retry timer.
   ipv6 pim timer graft-retry *interval*
   By default, the IPv6 PIM-DM graft retry timer is 3 seconds.

Configuring IPv6 PIM-SM

IPv6 PIM-SM tasks at a glance

To configure IPv6 PIM-SM, perform the following tasks:

1. Enabling IPv6 PIM-SM
2. Configuring static RPs
   As a best practice, configure a static RP when only one dynamic RP exists in the network.
3. Configure dynamic RP election:
   - Configuring C-RPs
   - Configuring C-BSRs
   - (Optional.) Configuring the interval for C-BSRs to send BSMs
   - (Optional.) Configuring the timer for C-BSRs to wait BSMs from the BSR
   - (Optional.) Configuring an IPv6 PIM domain border
   - (Optional.) Disabling BSM semantic fragmentation
   - (Optional.) Disabling the device from forwarding BSMs out of their incoming interfaces

   As a best practice, configure dynamic RPs when multiple IPv6 PIM devices exist in the network.
4. Enabling embedded RP
   Enable embedded RP on each device for inter-domain communication when the devices are located in different IPv6 PIM-SM domains.
5. (Optional.) Configuring Anycast RP
6. (Optional.) Configuring IPv6 multicast source registration
7. (Optional.) Configuring the switchover to SPT
8. (Optional.) Configuring the delay time for DR election
9. (Optional.) Configuring common IPv6 PIM features

Prerequisites for IPv6 PIM-SM

Before you configure IPv6 PIM-SM, you must complete the following tasks:

- Configure an IPv6 unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable IPv6 multicast routing.

Enabling IPv6 PIM-SM

About this task

With IPv6 PIM-SM enabled on interfaces, devices can establish IPv6 PIM neighbor relationship and process IPv6 PIM messages from their IPv6 PIM neighbors.

Restrictions and guidelines

As a best practice, enable IPv6 PIM-SM on all non-border interfaces of devices when you deploy an IPv6 PIM-SM domain.

Procedure

1. Enter system view.

   system-view
2. Enable IPv6 multicast routing and enter IPv6 MRIB view.

   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]

   By default, IPv6 multicast routing is disabled.

   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.

   quit
4. Enter interface view.

   interface *interface-type interface-number*
5. Enable IPv6 PIM-SM.

   ipv6 pim sm

   By default, IPv6 PIM-SM is disabled.

Configuring static RPs

About this task

If only one dynamic RP exists on a network, you can configure a static RP to avoid communication interruption caused by a single point of failure. The static RP can also avoid waste of bandwidth caused by frequent message exchanges between C-RPs and the BSR.

Restrictions and guidelines

In the IPv6 PIM-SM domain, you can configure the same static RP for multiple IPv6 multicast groups by using the same RP address but different ACLs.

You do not need to enable IPv6 PIM for an interface to be configured as a static RP.

If you configure multiple static RPs for an IPv6 multicast group, only the static RP with the highest IPv6 address takes effect.

The static RP configuration must be the same on all devices in the IPv6 PIM-SM domain.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Configure a static RP for IPv6 PIM-SM.

   static-rp *ipv6-rp-address* [{ *ipv6-acl-number* | **name** *ipv6-acl-name* }] | **preferred**] *

   When both a static RP and a dynamically elected RP exist on the network, you can specify the **preferred** keyword to give priority to the static RP. The dynamic RP is used only when the static RP fails. If you do not specify the **preferred** keyword, the dynamic RP takes priority and the static RP is used only when the dynamic RP fails.

Configuring C-RPs

About this task

To avoid C-RP spoofing, configure a C-RP policy to filter C-RP advertisement messages by using an ACL that specifies the packet source address range and multicast group addresses.

Restrictions and guidelines

Configure C-RPs on devices that reside in the backbone network.

Because the RP and other devices exchange a large amount of information in the IPv6 PIM-SM domain, reserve a large bandwidth between C-RPs and other devices.

You must configure the same C-RP policy on all C-BSRs in the IPv6 PIM-SM domain because every C-BSR might become the BSR.

The device might use the BSR RP hash algorithm described in RFC 4601 or in RFC 2362 to calculate the RP for a multicast group. To ensure consistent group-to-RP mappings on all PIM devices in the IPv6 PIM-SM domain, specify the same BSR RP hash algorithm on the devices.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-RP.

   c-rp *ipv6-address* [**advertisement-interval** *adv-interval* | { **group-policy** { *ipv6-acl-number* | **name** *ipv6-acl-name* } | **scope** *scope-id* } | **holdtime** *hold-time* | **priority** *priority*] *

4. (Optional.) Configure a C-RP policy.

   crp-policy *ipv6-acl-number*

   By default, no C-RP policies are configured. All C-RP advertisement messages are regarded as legal.

5. (Optional.) Configure the device to use the BSR RP hash algorithm described in RFC 2362.

   bsr-rp-mapping rfc2362

   By default, the device uses the BSR RP hash algorithm described in RFC 4601.

Configuring C-BSRs

About this task

You must configure C-BSRs when you configure dynamic RP election.

To prevent a legal BSR from being replaced by a malicious host, configure a BSR policy to filter BSR messages by using an ACL that specifies the legal BSR addresses.

Restrictions and guidelines

Configure C-BSRs on devices that reside in the backbone network.

Because the BSR and other devices exchange a large amount of information in the IPv6 PIM-SM domain, reserve a large bandwidth between the C-BSR and other devices.

The C-BSR configuration on the devices in the IPv6 PIM-SM domain must be the same.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-BSR.

   c-bsr *ipv6-address* [**scope** *scope-id*] [**hash-length** *hash-length* | **priority** *priority*] *

4. (Optional.) Configure a BSR policy.

   bsr-policy *ipv6-acl-number*

   By default, no BSR policies are configured. All bootstrap messages are regarded as legal.

Configuring the interval for C-BSRs to send BSMs

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires (configured by using the **c-bsr holdtime** command). During this period of time, the BSR periodically sends BSMs to all other devices in the PIM domain. The BSMs carry the IP address of the BSR, RP-set information, and RP aging time. Other devices in the PIM domain select an RP for a multicast group based on the RP-set information. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure the interval for C-BSRs to send BSMs.
   c-bsr interval *interval*
   By default, a C-BSR sends BSMs at 60-second intervals.

Configuring the timer for C-BSRs to wait BSMs from the BSR

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure the timer for C-BSRs to wait BSMs from the BSR.
   c-bsr holdtime *hold-time*
   By default, a C-BSR waits for BSMs from the BSR for 130 seconds before sending BSMs.

Configuring an IPv6 PIM domain border

About this task

An IPv6 PIM domain border determines the transmission boundary of bootstrap messages. Bootstrap messages cannot cross the domain border in either direction. A number of IPv6 PIM domain border interfaces partition a network into different IPv6 PIM-SM domains.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.

interface *interface-type interface-number*

3. Configure an IPv6 PIM domain border.

   ipv6 pim bsr-boundary

   By default, an interface is not an IPv6 PIM domain border.

Disabling BSM semantic fragmentation

About this task

BSM semantic fragmentation enables a BSR to split a BSM into multiple BSM fragments (BSMFs) if the BSM exceeds the MTU. In this way, a non-BSR device can update the RP-set information for a group range after receiving all BSMFs for the group range. The loss of one BSMF only affects the RP-set information of the group ranges that the fragment contains.

Restrictions and guidelines

If a device does not support BSM semantic fragmentation, it regards a BSMF as a BSM and updates the RP-set information each time it receives a BSMF. It learns only part of the RP-set information, which further affects the RP election. Therefore, if such a device exists in the IPv6 PIM-SM domain, you must disable BSM semantic fragmentation on all C-BSRs.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Disable BSM semantic fragmentation.

   undo bsm-fragment enable

   By default, BSM semantic fragmentation is enabled.

Disabling the device from forwarding BSMs out of their incoming interfaces

About this task

By default, the device forwards BSMs out of their incoming interfaces. This default setting ensures that all devices on the subnet can receive BSMs even when they have inconsistent routing information. However, this default setting results in duplicated IPv6 multicast traffic. If you are sure that all the devices on the subnet have consistent routing information, you can disable the device from forwarding BSMs out of their incoming interfaces.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Disable the device from forwarding BSMs out of their incoming interfaces.

   undo bsm-reflection enable

   By default, the device forwards BSMs out of their incoming interfaces.

Enabling embedded RP

Restrictions and guidelines

This feature is supported only for IPv6 multicast group addresses in the range of FF7x::/12 or FFFx::/12 and in compliance with RFC 3956.

Enable this feature on all devices in the IPv6 PIM-SM domain.

IPv6 BIDIR-PIM does not support embedded RP.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim[**vpn-instance** *vpn-instance-name*]
3. Enable embedded RP.
   embedded-rp [*ipv6-acl-number*]
   By default, embedded RP is disabled.

Configuring Anycast RP

Restrictions and guidelines

In an Anycast RP set, RP members must have an interface (for example, loopback interface) with the same IPv6 address. The IPv6 address must be configured as both a C-RP and the Anycast RP address.

To prevent the other Anycast RP member devices from discarding the BSM sent by the BSR, make sure the Anycast RP address is different from the BSR address.

As a best practice to ensure network performance, configure a maximum of 16 Anycast RP members for an Anycast RP set.

As a best practice, configure the IPv6 address of a loopback interface as the RP member address.

If you configure IPv6 addresses of multiple interfaces on the same device as RP member addresses, the lowest IPv6 address takes effect. The rest of the interface addresses become backup RP member addresses.

You must add the device where the Anycast RP resides as an RP member to the Anycast RP set.

Prerequisites

You must configure a static RPs or C-RPs in the IPv6 PIM-SM domain before you configure the Anycast RP. Use the address of the static RP or the dynamically elected RP as the Anycast RP address.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure Anycast RP.
   anycast-rp *ipv6-anycast-rp-address ipv6-member-address*
   You can repeat this command to add multiple RP member addresses to an Anycast RP set.

Configuring IPv6 multicast source registration

About this task

- **IPv6 PIM register policy**—An IPv6 PIM register policy enables an RP to filter register messages by using an ACL that specifies the IPv6 multicast sources and groups. The policy limits the multicast groups to which the RP is designated. If a register message is denied by the ACL or does not match the ACL, the RP discards the register message and sends a register-stop message to the source-side DR. The registration process stops.
- **Checksum computing method for register messages**—For information integrity of a register message, you can configure the device to calculate the checksum based on the entire register message. If a device cannot calculate the checksum based on the entire register message, you can configure the device to calculate the checksum based on the register message header.
- **Register suppression time**—The source-side DR stops sending register messages encapsulated with IPv6 multicast data and starts a register-stop timer upon receiving a register-stop message from the RP. Before the register-stop timer expires, the DR sends a null register message (a register message without encapsulated IPv6 multicast data) to the RP and starts a register probe timer. If the DR receives a register-stop message before the register probe timer expires, it resets its register-stop timer. If the DR does not receive a register-stop message when the register probe timer expires, the DR starts sending register messages with encapsulated data again.

  The register-stop timer is set to a random value chosen uniformly from (0.5 × register_suppression_time minus register_probe_time) to (1.5 × register_suppression_time minus register_probe_time). The register_probe_time is (5 seconds). You can adjust the register suppression time.

Restrictions and guidelines

On all C-RP devices, configure an IPv6 PIM register policy and the checksum computing method for register messages.

On all devices that might become the source-side DR, set the register suppression time.

As a best practice, set the suppression time to be longer than or equal to 11 seconds.

Procedure

1. Enter system view.

   system-view
2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure an IPv6 PIM register policy.

   register-policy { *ipv6-acl-number* | **name** *ipv6-acl-name* }

   By default, no IPv6 register policy is configured. All IPv6 register messages are regarded as legal.
4. Configure the device to calculate the checksum based on the entire register message.

   register-whole-checksum

   By default, the device calculates the checksum based on the header of a register message.
5. Set the register suppression time.

   register-suppression-timeout *interval*

   The default setting is 60 seconds.

Configuring the switchover to SPT

Restrictions and guidelines

Some devices cannot encapsulate IPv6 multicast data in register messages destined to the RP. As a best practice to avoid IPv6 multicast data forwarding failures, do not disable the switchover to SPT on C-RPs that might become the RP.

Procedure

1. Enter system view.

   system-view
2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure the switchover to SPT.

   spt-switch-threshold { **immediacy** | **infinity** } [**group-policy** *ipv6-acl-number*]

   By default, the first IPv6 multicast data packet triggers the RPT to SPT switchover.

Configuring the delay time for DR election

About this task

After a failed DR recovers, it will send IPv6 PIM hello messages and will be elected as the new DR immediately. In this case, transient multicast traffic loss will occur because the new DR has not generated a complete multicast forwarding table.

Perform this task to delay a recovered DR from being elected as the new DR. Before the delay timer expires, the current DR continues to forward multicast traffic, minimizing multicast traffic loss.

Restrictions and guidelines

As a best practice, perform this task on the user-side interface (interface enabled with MLD) only on the DR. If you perform this task on both the DR and a non-DR device and both of them fail, traffic loss will occur before the delay timer expires.

As a best practice to avoid sending repeated packets after the failed DR recovers, set the delay time to be longer than the interval for sending IPv6 PIM hello messages.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the delay time for DR election.

   ipv6 pim timer dr-elect-delay *delay-time*

   By default, a failed DR is elected as the new DR immediately after it recovers.

Configuring IPv6 BIDIR-PIM

IPv6 BIDIR-PIM tasks at a glance

To configure IPv6 BIDIR-PIM, perform the following tasks:

1. Enabling IPv6 BIDIR-PIM

2. Configuring static RPs

   As a best practice, configure a static R when only one dynamic RP exists in the network.
3. Configuring dynamic RPs
   - Configuring C-RPs
   - Configuring C-BSRs
   - (Optional.) Configuring the interval for C-BSRs to send BSMs
   - (Optional.) Configuring the timer for C-BSRs to wait BSMs from the BSR
   - (Optional.) Configuring an IPv6 PIM domain border
   - (Optional.) Disabling BSM semantic fragmentation
   - (Optional.) Disabling the device from forwarding BSMs out of their incoming interfaces
   - (Optional.) Setting the maximum number of IPv6 BIDIR-PIM RPs

   As a best practice, configure dynamic RPs when multiple IPv6 PIM devices exist in the network.

   You can configure static RPs, dynamic RPs, or both.
4. (Optional.) Configuring common IPv6 PIM features

Prerequisites for IPv6 BIDIR-PIM

Before you configure IPv6 BIDIR-PIM, you must complete the following tasks:

- Configure an IPv6 unicast routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable IPv6 PIM-SM.

Enabling IPv6 BIDIR-PIM

Restrictions and guidelines

As a best practice, enable IPv6 BIDIR-PIM on all non-border interfaces of devices when you deploy an IPv6 BIDIR-PIM.

Procedure

1. Enter system view.

   `system-view`
2. Enable IPv6 multicast routing and enter MRIB view.

   `ipv6 multicast routing` [**`vpn-instance`** *`vpn-instance-name`*]

   By default, IPv6 multicast routing is disabled.

   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.

   `quit`
4. Enter interface view.

   `interface` *`interface-type interface-number`*
5. Enable IPv6 PIM-SM.

   `ipv6 pim sm`

   By default, IPv6 PIM-SM is disabled.
6. Return to system view.

   `quit`
7. Enter IPv6 PIM view

ipv6 pim [**vpn-instance** *vpn-instance-name*]

8. Enable IPv6 BIDIR-PIM.

   bidir-pim enable

   By default, IPv6 BIDIR-SM is disabled.

Configuring static RPs

About this task

If only one dynamic RP exists on a network, you can configure a static RP to avoid communication interruption caused by single-point failures. The static RP can also avoid bandwidth waste caused by frequent message exchanges between C-RPs and the BSR.

Restrictions and guidelines

The static RP configuration must be the same on all devices in the IPv6 BIDIR-PIM domain.

In an IPv6 BIDIR-PIM domain, you can configure the same static RP for different IPv6 multicast groups by using the same RP address but different ACLs.

You do not need to enable IPv6 PIM for an interface to be configured as a static RP.

If you configure multiple static RPs for an IPv6 multicast group, only the static RP with the highest IPv6 address takes effect.

You can specify an unused IP address for a static RP. This address must be on the same subnet as the link on which the static RP is configured. For example, if the IPv6 addresses of the interfaces at the two ends of a link are 1001::1/64 and 1001::2/64, you can specify the interface with IPv6 address 1001::100/64 as a static RP. As a result, the link becomes an RPL.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Configure a static RP for IPv6 BIDIR-PIM.

   static-rp *ipv6-rp-address* **bidir** [{ *ipv6-acl-number* | **name** *ipv6-acl-name* } | **preferred**] *

   When both a static RP and a dynamically elected RP exist on the network, you can specify the **preferred** keyword to give priority to the static RP. The dynamic RP is used only when the static RP fails. If you do not specify the **preferred** keyword, the dynamic RP takes priority and the static RP is used only when the dynamic RP fails.

Configuring C-RPs

About this task

To guard against C-RP spoofing, configure a C-RP policy to filter C-RP advertisement messages by using an ACL that specifies the packet source address range and IPv6 multicast groups.

Restrictions and guidelines

Configure C-RPs on devices that reside in the backbone network.

Because the RP and other devices exchange a large amount of information in the IPv6 BIDIR-PIM domain, reserve a large bandwidth between C-RPs and other devices.

You must configure the same C-RP policy on all C-BSRs in the IPv6 BIDIR-PIM domain.

The device might use the BSR RP hash algorithm described in RFC 4601 or in RFC 2362 to calculate the RP for a multicast group. To ensure consistent group-to-RP mappings on all PIM devices in the IPv6 BIDIR-PIM domain, specify the same BSR RP hash algorithm on the devices.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-RP to provide services for IPv6 BIDIR-PIM.

   c-rp *ipv6-address* [**advertisement-interval** *adv-interval* | { **group-policy** { *ipv6-acl-number* | **name** *ipv6-acl-name* } | **scope** *scope-id* } | **holdtime** *hold-time* | **priority** *priority*] * **bidir**

4. (Optional.) Configure the device to use the RP hash algorithm described in RFC 2362.

   bsr-rp-mapping rfc2362

   By default, the device uses the BSR RP hash algorithm described in RFC 4601.

Configuring C-BSRs

About this task

You must configure C-BSRs when you configure dynamic RP election.

To prevent a legal BSR from being replaced by a malicious host, configure a BSR policy to filter BSR messages by using an ACL that specifies the legal BSR addresses.

Restrictions and guidelines

Configure C-BSRs on devices that reside in the backbone network.

Because the BSRs and other devices exchange a large amount of information in the IPv6 BIDIR-PIM domain, reserve a large bandwidth between the C-BSRs and other devices.

The C-BSR configuration on all devices in the same IPv6 BIDIR-PIM domain must be the same.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Configure a C-BSR.

   c-bsr *ipv6-address* [**scope** *scope-id*] [**hash-length** *hash-length* | **priority** *priority*] *

   By default, no C-BSRs exist.

4. (Optional.) Configure a BSR policy.

   bsr-policy *ipv6-acl-number*

   By default, no BSR policy is configured. All bootstrap messages are regarded as legal.

Configuring the interval for C-BSRs to send BSMs

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires (configured by using the **c-bsr holdtime** command). During this

period of time, the BSR periodically sends BSMs to all other devices in the PIM domain. The BSMs carry the IP address of the BSR, RP-set information, and RP aging time. Other devices in the PIM domain select an RP for a multicast group based on the RP-set information. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

In an IPv6 BIDIR-PIM network with active and backup RPs, when the link for the active RP fails, the backup RP must wait the active RP to age out before taking over. The aging time of the active RP is long because it must be longer than the sending interval of bootstrap messages, and the sending interval of BSMs is not configurable. You can use this command to set a smaller sending interval.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Configure the interval for C-BSRs to send BSMs.

   c-bsr interval *interval*

   By default, a C-BSR sends BSMs at 60-second intervals.

Configuring the timer for C-BSRs to wait BSMs from the BSR

About this task

After a C-BSR is elected as the BSR, all other C-BSRs stop sending BSMs to other devices until the timer for waiting BSMs expires. After the timer for waiting BSMs expires, all C-BSRs send BSMs to one another, triggering a new BSR election process.

Restrictions and guidelines

If you configure both the sending interval and the hold time, make sure the former is smaller than the latter. Otherwise, BSR election processes will be triggered frequently.

If you configure only one of them, the value of the other parameter can be calculated by using the formula: value of the unconfigured parameter = (value of the configured parameter–10)/2

If you configure only the hold time and the sending interval calculated by using the formula is smaller than 1 second, the actual sending interval is 1 second.

If you configure only the sending interval and the hold time calculated by using the formula is greater than the maximum value (2147483647 seconds), the actual hold time is 2147483647 seconds.

You must configure the same sending interval and hold time on all C-BSRs in one PIM domain. Otherwise, BSR election will be unstable, which causes multicast traffic forwarding problems.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure the timer for C-BSRs to wait BSMs from the BSR.
   c-bsr holdtime *hold-time*
   By default, a C-BSR waits BSMs from the BSR for 130 seconds before sending BSMs.

Configuring an IPv6 PIM domain border

About this task

An IPv6 PIM domain border determines the transmission boundary of bootstrap messages. Bootstrap messages cannot cross the domain border in either direction. A number of PIM domain border interfaces partition a network into different IPv6 BIDIR-PIM domains.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure an IPv6 PIM domain border.
   ipv6 pim bsr-boundary
   By default, an interface is not an IPv6 PIM domain border.

Disabling BSM semantic fragmentation

About this task

BSM semantic fragmentation enables a BSR to split a BSM into multiple BSM fragments (BSMFs) if the BSM exceeds the MTU. In this way, a non-BSR device can update the RP-set information for a group range after receiving all BSMFs for the group range. The loss of one BSMF only affects the RP-set information of the group ranges that the fragment contains.

Restrictions and guidelines

If a device does not support BSM semantic fragmentation, it regards a BSMF as a BSM and updates the RP-set information each time it receives a BSMF. It learns only part of the RP-set information, which further affects the RP election. Therefore, if such a device exists in the IPv6 BIDIR-PIM domain, you must disable BSM semantic fragmentation on all C-BSRs.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Disable BSM semantic fragmentation.
   undo bsm-fragment enable
   By default, BSM semantic fragmentation is enabled.

Disabling the device from forwarding BSMs out of their incoming interfaces

About this task

By default, the device forwards BSMs out of their incoming interfaces. This default setting ensures that all devices on the subnet can receive BSMs even when they have inconsistent routing information. However, this default setting results in duplicated IPv6 multicast traffic. If you are sure that all the devices on the subnet have consistent routing information, you can disable the device from forwarding BSMs out of their incoming interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Disable the device from forwarding BSMs out of their incoming interfaces.
   undo bsm-reflection enable
   By default, the device forwards BSMs out of their incoming interfaces.

Setting the maximum number of IPv6 BIDIR-PIM RPs

About this task

In an IPv6 BIDIR-PIM domain, one DF election per RP is implemented on all IPv6 PIM-enabled interfaces. To avoid unnecessary DF elections, do not configure multiple RPs for IPv6 BIDIR-PIM.

This configuration sets a limit on the number of IPv6 BIDIR-PIM RPs. If the number of RPs exceeds the limit, excess RPs do not take effect and can be used only for DF election rather than IPv6 multicast data forwarding. The system does not delete these excess RPs. They must be deleted manually.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Set the maximum number of IPv6 BIDIR-PIM RPs.
   bidir-rp-limit *limit*
   By default, the maximum number of IPv6 BIDIR-PIM RPs is 6.

Configuring IPv6 PIM-SSM

IPv6 PIM-SSM tasks at a glance

To configure IPv6 PIM-SSM, perform the following tasks:

1. Enabling IPv6 PIM-SM
2. (Optional.) Specifying the IPv6 SSM group range
3. (Optional.) Configuring the delay time for DR election

4. (Optional.) Configuring common IPv6 PIM features

Prerequisites for IPv6 PIM-SSM

Before you configure IPv6 PIM-SSM, you must complete the following tasks:

- Configure an IPv6 unicast IPv6 routing protocol so that all devices in the domain can interoperate at the network layer.
- Enable IPv6 PIM-SM.
- Enable MLDv2 on the IPv6 PIM devices that connect to IPv6 multicast receivers.

Enabling IPv6 PIM-SM

Restrictions and guidelines

As a best practice, enable IPv6 PIM-SM on all non-border interfaces of devices when you deploy an IPv6 PIM-SSM domain.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 multicast routing and enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
   By default, IPv6 multicast routing is disabled.
   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.
   quit
4. Enter interface view.
   interface *interface-type interface-number*
5. Enable IPv6 PIM-SM.
   ipv6 pim sm
   By default, IPv6 PIM-SM is disabled.

Specifying the IPv6 SSM group range

About this task

When an IPv6 PIM-SM enabled interface receives an IPv6 multicast packet, it checks whether the IPv6 multicast group address of the packet is in the IPv6 SSM group range. If the IPv6 multicast group address is in this range, the IPv6 PIM mode for this packet is IPv6 PIM-SSM. If the IPv6 multicast group address is not in this range, the IPv6 PIM mode is IPv6 PIM-SM.

Restrictions and guidelines

Configure the same IPv6 SSM group range on all devices in the entire IPv6 PIM-SM domain. Otherwise, IPv6 multicast information cannot be delivered through the IPv6 SSM model.

When a member of an IPv6 multicast group in the IPv6 SSM group range sends an MLDv1 report message, the device does not trigger a (*, G) join.

Procedure

1. Enter system view.
   system-view

2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure the IPv6 SSM group range.
   ssm-policy { *ipv6-acl-number* | **name** *ipv6-acl-name* }
   The default range is FF3x::/32, where x can be any valid scope.

Configuring the delay time for DR election

About this task

After a failed DR recovers, it will send IPv6 PIM hello messages and will be elected as the new DR immediately. In this case, transient multicast traffic loss will occur because the new DR has not generated a complete multicast forwarding table.

Perform this task to delay a recovered DR from being elected as the new DR. Before the delay timer expires, the current DR continues to forward multicast traffic, minimizing multicast traffic loss.

Restrictions and guidelines

As a best practice, perform this task on the user-side interface (interface enabled with MLD) only on the DR. If you perform this task on both the DR and a non-DR device and both of them fail, traffic loss will occur before the delay timer expires.

As a best practice to avoid sending repeated packets after the failed DR recovers, set the delay time to be longer than the interval for sending IPv6 PIM hello messages.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the delay time for DR election.
   ipv6 pim timer dr-elect-delay *delay-time*
   By default, a failed DR is elected as the new DR immediately after it recovers.

Configuring common IPv6 PIM features

Common IPv6 PIM feature tasks at a glance

All the following tasks are optional.

- Configuring an IPv6 multicast source policy
- Configuring an IPv6 PIM hello policy
- Configuring IPv6 PIM hello message options
- Dropping hello messages without the Generation ID option
- Enabling the device to send hello messages with a different Generation ID upon a port state change
- Configuring common IPv6 PIM timers
- Setting the maximum size of a join or prune message
- Configuring an IPv6 PIM join policy
- Setting the DSCP value for outgoing IPv6 PIM protocol packets

- Enabling BFD for IPv6 PIM
- Enabling IPv6 PIM passive mode
- Enabling IPv6 PIM NSR
- Enabling SNMP notifications for IPv6 PIM
- Enabling IPv6 PIM logging
- Enabling log message suppression for IPv6 PIM

Configuring an IPv6 multicast source policy

About this task

This feature enables the device to filter IPv6 multicast data by using an ACL that specifies the IPv6 multicast sources and the optional groups. It filters not only IPv6 multicast data packets but also IPv6 PIM register messages with IPv6 multicast data encapsulated. This also reduces the IPv6 multicast traffic on the network.

Procedure

1. Enter system view.

   system-view
2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Configure an IPv6 multicast source policy.

   source-policy { *ipv6-acl-number* | **name** *ipv6-acl-name* }

   By default, no IPv6 multicast source policy is configured. The device does not filter IPv6 multicast data packets.

Configuring an IPv6 PIM hello policy

About this task

This feature enables the device to filter IPv6 PIM hello messages by using an ACL that specifies the packet source addresses. It is used to guard against IPv6 PIM message attacks and to establish correct IPv6 PIM neighboring relationships.

If hello messages of an existing IPv6 PIM neighbor are filtered out by the policy, the neighbor is automatically removed when its aging timer expires.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure an IPv6 PIM hello policy.

   ipv6 pim neighbor-policy { *ipv6-acl-number* | **name** *ipv6-acl-name* }

   By default, no IPv6 PIM hello policy is configured on an interface. All IPv6 PIM hello messages are regarded as legal.

Configuring IPv6 PIM hello message options

About this task

In either an IPv6 PIM-DM domain or an IPv6 PIM-SM domain, hello messages exchanged among devices contain the following configurable options:

- **DR_Priority** (for IPv6 PIM-SM only)—Priority for DR election. The device with the highest priority wins the DR election. You can configure this option for all the devices in a shared-media LAN that directly connects to the IPv6 multicast source or the receivers.
- **Holdtime**—IPv6 PIM neighbor lifetime. If a device receives no hello message from a neighbor when the neighbor lifetime expires, it regards the neighbor failed or unreachable.
- **LAN_Prune_Delay**—Delay of pruning a downstream interface on a shared-media LAN. This option has the LAN delay field, the override interval field, and the T bit.

  The LAN delay defines the IPv6 PIM message propagation delay. The override interval defines a time period for a downstream device to override a prune message. The T bit specifies the capability of tracking downstream device status on upstream devices.

  On the shared-media LAN, the propagation delay and override interval are used as follows:

  - If a device receives a prune message on its upstream interface, it means that there are downstream devices on the shared-media LAN. If this device still needs to receive multicast data, it must send a join message to override the prune message within the override interval.
  - When a device receives a prune message from its downstream interface, it does not immediately prune this interface. Instead, it starts a timer (the propagation delay plus the override interval). If interface receives a join message before the timer expires, the device does not prune the interface. Otherwise, the device prunes the interface.
- **Neighbor tracking**—If you enable neighbor tracking on an upstream device, this device can track the states of the downstream nodes for which the joined state holdtime timer has not expired. All join messages from downstream devices are accepted.

Restrictions and guidelines

If the propagation delay or override interval on different IPv6 PIM devices on a shared-media LAN is different, the largest ones apply.

If you enable neighbor tracking, you must enable it on all IPv6 PIM devices on a shared-media LAN or the upstream device to track join messages from every downstream device.

You can configure hello message options for all interfaces in IPv6 PIM view or for the current interface in interface view. The configuration made in interface view takes priority over the configuration made in IPv6 PIM view.

Configuring hello message options globally

1. Enter system view.

   system-view
2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Set the DR priority.

   hello-option dr-priority *priority*

   The default setting is 1.
4. Set the neighbor lifetime.

   hello-option holdtime *time*

   The default setting is 105 seconds.
5. Set the IPv6 PIM message propagation delay for a shared-media LAN.

hello-option lan-delay *delay*

The default setting is 500 milliseconds.

6. Set the override interval.

   hello-option override-interval *interval*

   The default setting is 2500 milliseconds.

7. Enable neighbor tracking.

   hello-option neighbor-tracking

   By default, neighbor tracking is disabled.

Configuring hello message options on an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the DR priority.

   ipv6 pim hello-option dr-priority *priority*

   The default setting is 1.

4. Set the neighbor lifetime.

   ipv6 pim hello-option holdtime *time*

   The default setting is 105 seconds.

5. Set the IPv6 PIM message propagation delay.

   ipv6 pim hello-option lan-delay *delay*

   The default setting is 500 milliseconds.

6. Set the override interval.

   ipv6 pim hello-option override-interval *interval*

   The default setting is 2500 milliseconds.

7. Enable neighbor tracking.

   ipv6 pim hello-option neighbor-tracking

   By default, neighbor tracking is disabled.

Dropping hello messages without the Generation ID option

About this task

A device generates a generation ID for hello messages when an interface is enabled with IPv6 PIM. The generation ID is a random value, but it changes only when the status of the device changes. If an IPv6 PIM device finds that the generation ID in a hello message from the upstream device has changed, it assumes that the status of the upstream device has changed. In this case, it sends a join message to the upstream device for status update. You can configure an interface to drop hello messages without the generation ID options to promptly know the status of an upstream device.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the interface to drop hello messages without the Generation ID option.

ipv6 pim require-genid

By default, an interface accepts hello messages without the Generation ID option.

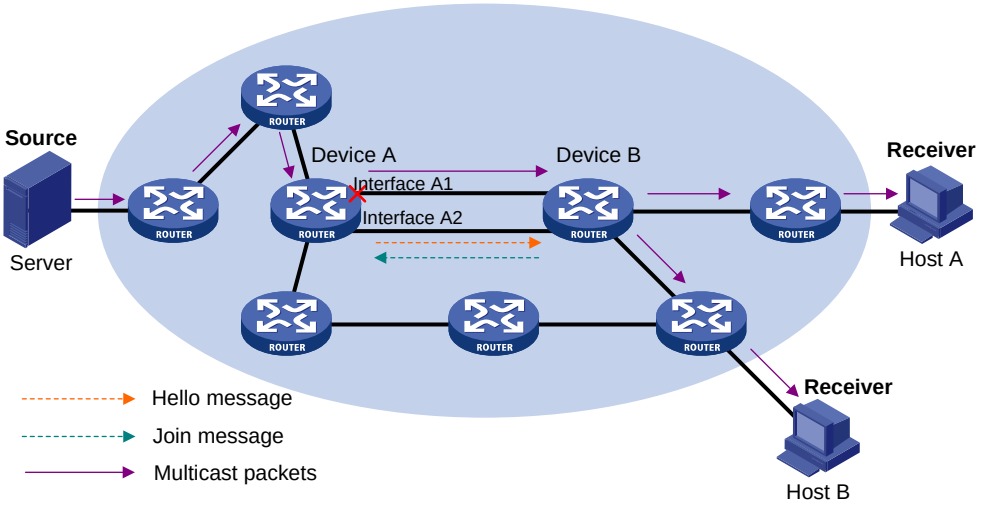
Enabling the device to send hello messages with a different Generation ID upon a port state change

About this task

As shown in Figure 16, Device B connects to Device A through two interfaces in the same VLAN (interface A1 and interface A2) for high availability. Device A sends multicast data to Device B through interface A. When interface A1 fails or recovers from failure, Device A sends multicast data to Device B through interface A2. When interface A1 recovers from failure, Device A sends multicast data to Device B through interface A1 again. In both cases, Device A does not immediately send hello messages with a different Generation ID and multicast data cannot be switched to the new path immediately.

After you perform this task, Device A will immediately send hello messages with a different Generation ID when the port state in the VLAN changes. Then, Device B will send an IPv6 PIM join message and update the multicast forwarding entry. Multicast data can be switched to the new path immediately.

Figure 16 Network diagram



Procedure

1. Enter system view.

   system-view

2. Enter VLAN interface view.

   interface vlan-interface *interface-number*

3. Enable the interface to send hello messages with a different Generation ID upon a port state change.

   ipv6 pim triggered-hello enable { **port-down** | **port-up** } *

   By default, the device does not send hello messages with a different Generation ID upon a port state change.

Configuring common IPv6 PIM timers

About this task

The following are common IPv6 PIM timers:

- **Hello interval**—Interval at which an IPv6 PIM device sends hello messages to discover IPv6 PIM neighbors and maintain IPv6 PIM neighbor relationship.
- **Triggered hello delay**—Maximum delay for sending a hello message to avoid collisions caused by simultaneous hello messages. After receiving a hello message, an IPv6 PIM device waits for a random time before sending a hello message. This random time is in the range of 0 to the triggered hello delay.
- **Join or prune interval**—Interval at which an IPv6 PIM device sends join or prune messages to its upstream devices for state update.
- **Joined/Pruned state holdtime**—Time for which an IPv6 PIM device keeps the joined/pruned state for the downstream interfaces. This joined/pruned state holdtime is contained in a join or prune message.
- **IPv6 multicast source lifetime**—Lifetime that an IPv6 PIM device maintains for an IPv6 multicast source. If a device does not receive subsequent IPv6 multicast data from the IPv6 multicast source S when the timer expires, it deletes the (S, G) entry for the IPv6 multicast source.

Restrictions and guidelines

To prevent upstream neighbors from aging out, set the join or prune interval to be less than the join or pruned state holdtime.

You can configure common IPv6 PIM timers for all interfaces in IPv6 PIM view or for the current interface in interface view. The configuration made in interface view takes priority over the configuration made in IPv6 PIM view.

As a best practice, use the defaults for a network without special requirements.

Configuring common IPv6 PIM timers globally

1. Enter system view.

   system-view
2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Set the hello interval.

   timer hello *interval*

   By default, the interval to send hello messages is 30 seconds.
4. Set the join or prune interval.

   timer join-prune *interval*

   By default, the interval to send join or prune messages is 60 seconds.

   This configuration takes effect after the current interval ends.
5. Set the joined/pruned state holdtime.

   holdtime join-prune *time*

   By default, the joined/pruned state holdtime timer is 210 seconds.
6. Set the IPv6 multicast source lifetime.

   source-lifetime *time*

   By default, the IPv6 multicast source lifetime is 210 seconds.

Configuring common IPv6 PIM timers on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the hello interval.
   ipv6 pim timer hello *interval*
   The default setting is 30 seconds.
4. Set the triggered hello delay.
   ipv6 pim triggered-hello-delay *delay*
   The default setting is 5 seconds.
5. Set the join or prune interval.
   ipv6 pim timer join-prune *interval*
   The default setting is 60 seconds.
   This configuration takes effect after the current interval ends.
6. Set the joined/pruned state holdtime.
   ipv6 pim holdtime join-prune *time*
   The default setting is 210 seconds.

Setting the maximum size of a join or prune message

Restrictions and guidelines

The loss of an oversized join or prune message might result in loss of massive information. You can set a small value for the size of a join or prune message to reduce the impact.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 PIM view.
   ipv6 pim [**vpn-instance** *vpn-instance-name*]
3. Set the maximum size of a join or prune message.
   jp-pkt-size *size*
   By default, the maximum size of a join or prune message is 1200 bytes.

Configuring an IPv6 PIM join policy

About this task

The IPv6 PIM join policy filters joined IPv6 multicast sources and groups in IPv6 PIM join or prune messages to protect the device against invalid IPv6 PIM join or prune message attacks. The policy identifies valid joined IPv6 multicast source and groups in IPv6 PIM join or prune messages and filters out invalid joined IPv6 multicast source and groups. The device creates (*, G) or (S, G) entries only for valid joined IPv6 multicast sources and groups.

Procedure

1. Enter system view.
   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure an IPv6 PIM join policy to filter joined IPv6 multicast sources and groups in IPv6 PIM join or prune messages.

   ipv6 pim join-policy { { *ipv6-acl-number* | **name** *ipv6-acl-name* } | [**asm** { *ipv6-acl-number* | **name** *ipv6-acl-name* } | **ssm** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] * }

   By default, no IPv6 PIM join policy is configured and joined IPv6 multicast sources and groups in IPv6 PIM join or prune messages are not filtered.

Setting the DSCP value for outgoing IPv6 PIM protocol packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.

   system-view

2. Enter IPv6 PIM view.

   ipv6 pim [**vpn-instance** *vpn-instance-name*]

3. Set the DSCP value for outgoing IPv6 PIM protocol packets.

   dscp *dscp-value*

   By default, the DSCP value is 48 for outgoing IPv6 PIM protocol packets.

Enabling BFD for IPv6 PIM

About this task

If a DR on a shared-media network fails, a new DR election process will start after the DR ages out. However, it might take a long period of time before other devices detect the link failures and trigger a new DR election. To start a new DR election process immediately after the original DR fails, you can enable BFD for IPv6 PIM to detect link failures among IPv6 PIM neighbors.

You must enable BFD for IPv6 PIM on all IPv6 PIM devices on a shared-media network. For more information about BFD, see *High Availability Configuration Guide*.

You must enable IPv6 PIM-DM or IPv6 PIM-SM on an interface before you configure this feature on the interface.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable BFD for IPv6 PIM.

   ipv6 pim bfd enable

   By default, BFD is disabled for IPv6 PIM.

Enabling IPv6 PIM passive mode

About this task

To guard against IPv6 PIM hello spoofing, you can enable IPv6 PIM passive mode on a receiver-side interface. The interface cannot receive or forward IPv6 PIM protocol messages (excluding register, register-stop and C-RP-Adv messages), and acts as the DR on the subnet.

Restrictions and guidelines

This feature takes effect only when IPv6 PIM-DM or IPv6 PIM-SM is enabled on the interface.

To avoid duplicate multicast data transmission and flow loop, do not enable this feature on a shared-media LAN with multiple IPv6 PIM devices.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IPv6 PIM passive mode on the interface.
   ipv6 pim passive
   By default, IPv6 PIM passive mode is disabled on an interface.

Enabling IPv6 PIM NSR

About this task

This feature enables IPv6 PIM to back up protocol state information, including IPv6 PIM neighbor information and routes, from the active process to the standby process. The standby process immediately takes over when the active process fails. Use this feature to avoid route flapping and forwarding interruption for IPv6 PIM when an active/standby switchover occurs.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 PIM NSR.
   ipv6 pim non-stop-routing
   By default, IPv6 PIM NSR is disabled.

Enabling SNMP notifications for IPv6 PIM

About this task

To report critical IPv6 PIM events to an NMS, enable SNMP notifications for IPv6 PIM. For PIM event notifications to be sent correctly, you must also configure SNMP as described in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for IPv6 PIM.

snmp-agent trap enable pim6 [**candidate-bsr-win-election** | **elected-bsr-lost-election** | **invalid-join-prune** | **invalid-register** | **interface-election** | **neighbor-loss** | **private-interface-election** | **private-neighbor-loss** | **private-new-neighbor** | **private-rp-mapping-change** | **rp-mapping-change**] *

By default, SNMP notifications for IPv6 PIM are enabled.

Enabling IPv6 PIM logging

About this task

To report critical IPv6 PIM events to the information center, enable IPv6 PIM logging. For IPv6 PIM log messages to be sent correctly, you must also configure the information center on the device. For more information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Outputting a large amount of log messages degrades the device performance. To resolve this issue, disable IPv6 PIM logging or use the **ipv6 pim log suppress** command to enable log message suppression for IPv6 PIM.

Procedure

1. Enter system view.

   system-view
2. Enable IPv6 PIM logging.

   ipv6 pim log enable [**dr** | **error-packet** | **neighbor** | **rp**] *

   IPv6 PIM logging is enabled by default.

Enabling log message suppression for IPv6 PIM

About this task

Outputting a large amount of log messages degrades the device performance. To resolve this issue, perform this task to enable log message suppression for IPv6 PIM. With this feature configured, the speed of sending IPv6 PIM log messages to the information center will be decreased to reduce the load on the device.

Restrictions and guidelines

Before you configure this feature, make sure IPv6 PIM logging has been enabled using the **ipv6 pim log enable** command.

Procedure

1. Enter system view.

   system-view
2. Enable log message suppression for IPv6 PIM.

   ipv6 pim log suppress [**dr** | **error-packet** | **neighbor** | **rp**] *

   By default, log message suppression is enabled for IPv6 PIM.

Display and maintenance commands for IPv6 PIM

Execute **display** commands in any view.

Task	Command
Display register-tunnel interface information.	**display interface** [**register-tunnel** [*interface-number*]] [**brief** [**description**| **down**]]
Display BSR information in the IPv6 PIM-SM domain.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **bsr-info**
Display information about the routes used by IPv6 PIM.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **claimed-route** [*ipv6-source-address*]
Display C-RP information in the IPv6 PIM-SM domain.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **c-rp** [**local**]
Display DF information in the IPv6 BIDIR-PIM domain.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **df-info** [*ipv6-rp-address*]
Display IPv6 PIM information on an interface.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **interface** [*interface-type interface-number*] [**verbose**]
Display IPv6 PIM neighbor information.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **neighbor** [*ipv6-neighbor-address* | **interface** *interface-type interface-number* | **verbose**] *
Display IPv6 PIM routing entries.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **routing-table** [*ipv6-group-address* [*prefix-length*] | *ipv6-source-address* [*prefix-length*] | **flags** *flag-value* | **fsm** | **incoming-interface** *interface-type interface-number* | **mode** *mode-type* | **outgoing-interface** { **exclude** | **include** | **match** } *interface-type interface-number* | **extranet** { **source-vpn-instance** *source-vpn-instance-name* | **source-public-instance** | **receive-vpn-instance** *receive-vpn-instance-name* | **receive-public-instance** }] *
Display RP information in the IPv6 PIM-SM domain.	**display ipv6 pim** [**vpn-instance** *vpn-instance-name*] **rp-info** [*ipv6-group-address*]
Display statistics for IPv6 PIM packets.	**display ipv6 pim statistics**

IPv6 PIM configuration examples

Example: Configuring IPv6 PIM-DM

Network configuration

As shown in Figure 17:

- OSPFv3 runs on the network.

- VOD streams are sent to receiver hosts in multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire IPv6 PIM domain is operating in the dense mode.
- Host A and Host C are multicast receivers in two stub networks N1 and N2.
- MLDv1 runs between Switch A and N1, and between Switch B, Switch C, and N2.

Figure 17 Network diagram

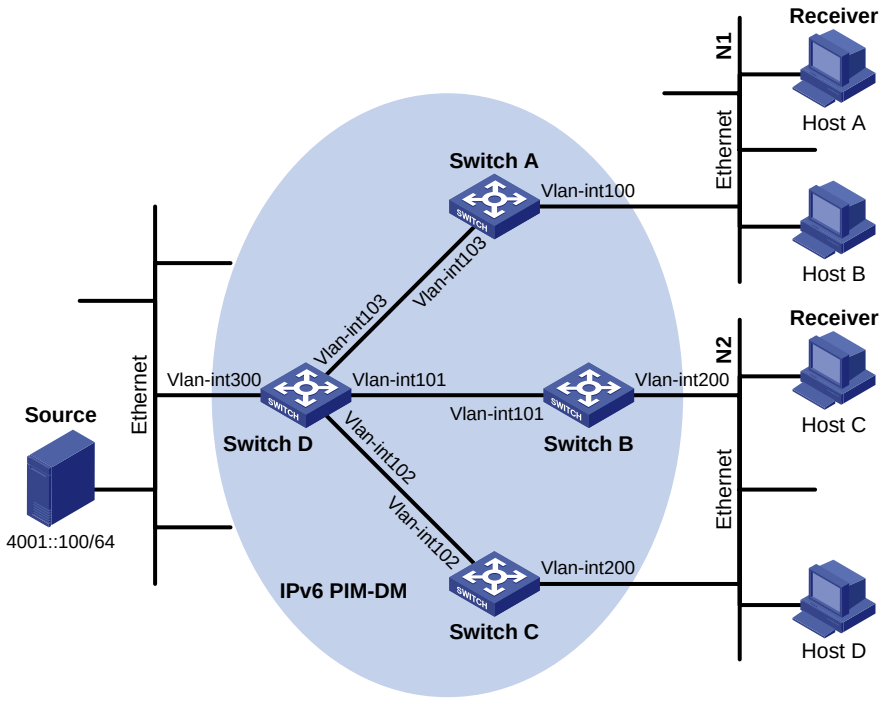


Table 2 Interface and IPv6 address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Switch A	Vlan-int100	1001::1/64	Switch D	Vlan-int300	4001::1/64
Switch A	Vlan-int103	1002::1/64	Switch D	Vlan-int103	1002::2/64
Switch B	Vlan-int200	2001::1/64	Switch D	Vlan-int101	2002::2/64
Switch B	Vlan-int101	2002::1/64	Switch D	Vlan-int102	3001::2/64
Switch C	Vlan-int200	2001::2/64	Switch D	Vlan-int300	4001::1/64
Switch C	Vlan-int102	3001::1/64	Switch D	Vlan-int103	1002::2/64

Prerequisites

Assign an IPv6 address and prefix length to each interface as shown in Figure 17, and configure OSPFv3 on the switches in the IPv6 PIM-DM domain.

Procedure

1. Enable IPv6 multicast routing, MLD, and IPv6 PIM-DM:

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

Enable MLD on VLAN-interface 100 (the interface that is connected to the stub network).

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld enable
[SwitchA-Vlan-interface100] quit
```

Enable IPv6 PIM-DM on VLAN-interface 103.

```
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] ipv6 pim dm
[SwitchA-Vlan-interface103] quit
```

Enable IPv6 multicast routing, MLD, and IPv6 PIM-DM on Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

On Switch D, enable IPv6 multicast routing, and enable IPv6 PIM-DM on each interface.

```
<SwitchD> system-view
[SwitchD] ipv6 multicast routing
[SwitchD-mrib6] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] ipv6 pim dm
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] ipv6 pim dm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 101
[SwitchD-Vlan-interface101] ipv6 pim dm
[SwitchD-Vlan-interface101] quit
[SwitchD] interface vlan-interface 102
[SwitchD-Vlan-interface102] ipv6 pim dm
[SwitchD-Vlan-interface102] quit
```

Verifying the configuration

Display IPv6 PIM information on Switch D.

```
[SwitchD] display ipv6 pim interface
 Interface: Vlan300
   NbrCnt: 0
   HelloInt: 30
   DR priority: 1
   DR address: FE80::A01:201:1    (local)
 Interface: Vlan103
   NbrCnt: 0
   HelloInt: 30
   DR priority: 1
   DR address: FE80::A01:201:2    (local)
 Interface: Vlan101
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: FE80::A01:201:3    (local)
 Interface: Vlan102
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
```

```
   DR address: FE80::A01:201:4    (local)
```

Display IPv6 PIM neighboring relationship on Switch D.

```
[SwitchD] display ipv6 pim neighbor
 Total IPv6 pim neighbors: 3
  Neighbor: FE80::A01:101:1
    Interface: Vlan103
    Uptime: 00:04:00
    Expiry time: 00:01:29
    DR priority: 1
    Mode: N/A
  Neighbor: FE80::B01:102:2
    Interface: Vlan101
    Uptime: 00:04:16
    Expiry time: 00:01:29
    DR priority: 3
    Mode: N/A
  Neighbor: FE80::C01:103:3
    Interface: Vlan102
    Uptime: 00:03:54
    Expiry time: 00:01:17
    DR priority: 5
    Mode: N/A
```

Send an MLD report from Host A to join IPv6 multicast group FF0E::101. (Details not shown.)

Send IPv6 multicast data from IPv6 multicast source 4001::100/64 to IPv6 multicast group FF0E::101. (Details not shown.)

Display IPv6 PIM multicast routing table information on Switch A.

```
[SwitchA] display ipv6 pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::101)
     Protocol: pim-dm, Flag: WC
     UpTime: 00:01:24
     Upstream interface: NULL
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstream interfaces: 1
         1: Vlan-interface100
             Protocol: mld, UpTime: 00:01:20, Expires: -

 (4001::100, FF0E::101)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:01:20
     Upstream interface: Vlan-interface103
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
     Downstream interface(s) information:
     Total number of downstream interfaces: 1
```

```
        1: Vlan-interface100
            Protocol: pim-dm, UpTime: 00:01:20, Expires: -
```

Display IPv6 PIM multicast routing table information on Switch D.

```
[SwitchD] display ipv6 pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF0E::101)
     Protocol: pim-dm, Flag: LOC ACT
     UpTime: 00:02:19
     Upstream interface: Vlan-interface300
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstream interfaces: 2
         1: Vlan-interface103
             Protocol: pim-dm, UpTime: 00:02:19, Expires: -
         2: Vlan-interface102
             Protocol: pim-dm, UpTime: 00:02:19, Expires: -
```

The output shows the following information:

- Switches on the SPT path (Switch A and Switch D) have the correct (S, G) entries.
- Switch A has a correct (*, G) entry.

Example: Configuring non-scoped IPv6 PIM-SM

Network configuration

As shown in Figure 18:

- OSPFv3 runs on the network.
- VOD streams are sent to receiver hosts in multicast. The receivers of different subnets form stub networks, and a minimum of one receiver host exist in each stub network. The entire IPv6 PIM-SM domain contains only one BSR.
- Host A and Host C are multicast receivers in the stub networks N1 and N2.
- Specify VLAN-interface 102 on Switch E as a C-BSR and a C-RP. The C-RP is designated to the IPv6 multicast group range FF0E::101/64. Specify VLAN-interface101 of Switch D as the static RP on all the switches to back up the dynamic RP.
- MLDv1 runs between Switch A and N1, and between Switch B, Switch C, and N2.

Figure 18 Network diagram

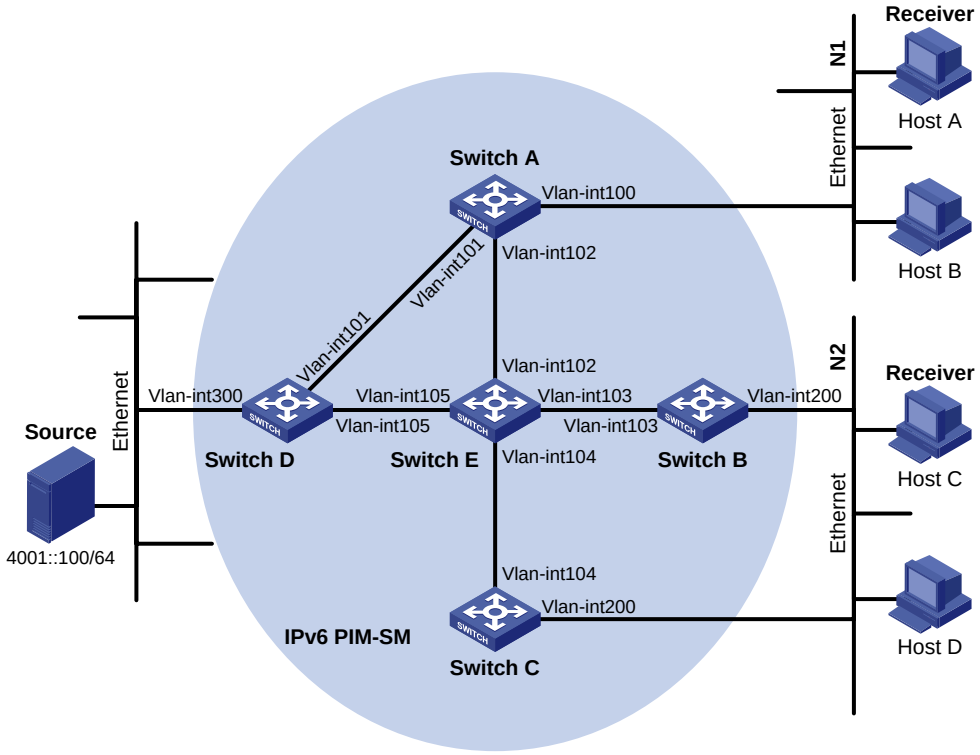


Table 3 Interface and IPv6 address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Switch A	Vlan-int100	1001::1/64	Switch D	Vlan-int300	4001::1/64
Switch A	Vlan-int101	1002::1/64	Switch D	Vlan-int101	1002::2/64
Switch A	Vlan-int102	1003::1/64	Switch D	Vlan-int105	4002::1/64
Switch B	Vlan-int200	2001::1/64	Switch E	Vlan-int104	3001::2/64
Switch B	Vlan-int103	2002::1/64	Switch E	Vlan-int103	2002::2/64
Switch C	Vlan-int200	2001::2/64	Switch E	Vlan-int102	1003::2/64
Switch C	Vlan-int104	3001::1/64	Switch E	Vlan-int105	4002::2/64

Prerequisites

Assign an IPv6 address and prefix length to each interface as shown in Figure 18, and configure OSPFv3 on all switches in the IPv6 PIM-SM domain.

Procedure

1. Enable IPv6 multicast routing, MLD, and IPv6 PIM-SM:

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Enable MLD on VLAN-interface 100 (the interface that is connected to the stub network).

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] mld enable
   ```

```
[SwitchA-Vlan-interface100] quit
```

Enable IPv6 PIM-SM on other interfaces.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] ipv6 pim sm
[SwitchA-Vlan-interface102] quit
```

Enable IPv6 multicast routing, MLD, and IPv6 PIM-SM on Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

Enable IPv6 multicast routing and IPv6 PIM-SM on Switch D and Switch E in the same way Switch A is configured. (Details not shown.)

2. Configure C-BSRs, C-RPs, and static RPs:

On Switch E, configure the service scope of RP advertisements.

```
<SwitchE> system-view
[SwitchE] acl ipv6 basic 2005
[SwitchE-acl-ipv6-basic-2005] rule permit source ff0e::101 64
[SwitchE-acl-ipv6-basic-2005] quit
```

Configure VLAN-interface 102 as a C-BSR and a C-RP, and configure VLAN-interface 101 of Switch D as the static RP.

```
[SwitchE] ipv6 pim
[SwitchE-pim6] c-bsr 1003::2
[SwitchE-pim6] c-rp 1003::2 group-policy 2005
[SwitchE-pim6] static-rp 1002::2
[SwitchE-pim6] quit
```

On Switch A, configure VLAN-interface 101 of Switch D as a static RP.

```
[SwitchA] ipv6 pim
[SwitchA-pim6] static-rp 1002::2
[SwitchA-pim6] quit
```

Configure a static RP on Switch B, Switch C, and Switch D in the same way Switch A is configured. (Details not shown.)

Verifying the configuration

Display IPv6 PIM information on Switch A.

```
[SwitchA] display ipv6 pim interface
 Interface: Vlan101
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: FE80::A01:201:2
 Interface: Vlan102
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: FE80::A01:201:3
```

Display BSR information on Switch A.

```
[SwitchA] display ipv6 pim bsr-info
 Scope: non-scoped
```

```
    State: Accept Preferred
    Bootstrap timer: 00:01:44
    Elected BSR address: 1003::2
      Priority: 64
      Hash mask length: 126
      Uptime: 00:11:18
```

Display BSR information on Switch E.

```
[SwitchE] display ipv6 pim bsr-info
 Scope: non-scoped
    State: Elected
    Bootstrap timer: 00:01:44
    Elected BSR address: 1003::2
      Priority: 64
      Hash mask length: 126
      Uptime: 00:11:18
    Candidate BSR address: 1003::2
      Priority: 64
      Hash mask length: 126
```

Display RP information on Switch A.

```
[SwitchA] display ipv6 pim rp-info
 BSR RP information:
   Scope: non-scoped
     Group/MaskLen: FF0E::101/64
       RP address               Priority  HoldTime  Uptime    Expires
       1003::2                  192       180       00:05:19  00:02:11

Static RP information:
       RP address               ACL   Mode    Preferred
       1002::2                  ----  pim-sm  No
```

Example: Configuring admin-scoped IPv6 PIM-SM

Network configuration

As shown in Figure 19:

- VOD streams are sent to receiver hosts in multicast. The entire IPv6 PIM-SM domain is divided into IPv6 admin-scoped zone 1 and the IPv6 global-scoped zone. Router B and Router C are ZBRs of IPv6 admin-scoped zone 1.
- Source 2 sends IPv6 multicast data to the IPv6 multicast group FF14::101, and Host A is an IPv6 multicast receiver. Source 1 sends IPv6 multicast data to the IPv6 multicast group FF1E::202, and Host A is an IPv6 multicast receiver.
- VLAN-interface 20 of Switch B acts as a C-BSR and a C-RP for IPv6 admin-scoped zone 1, and is designated to the IPv6 multicast groups with the scope field value of 4. VLAN-interface 50 of Switch D acts as a C-BSR and a C-RP for the IPv6 global-scoped zone, and it is designated to the IPv6 multicast groups with the scope field value of 14.
- MLDv1 runs between Switch A, Switch C, and the receivers that directly connect to them, respectively.

Figure 19 Network diagram

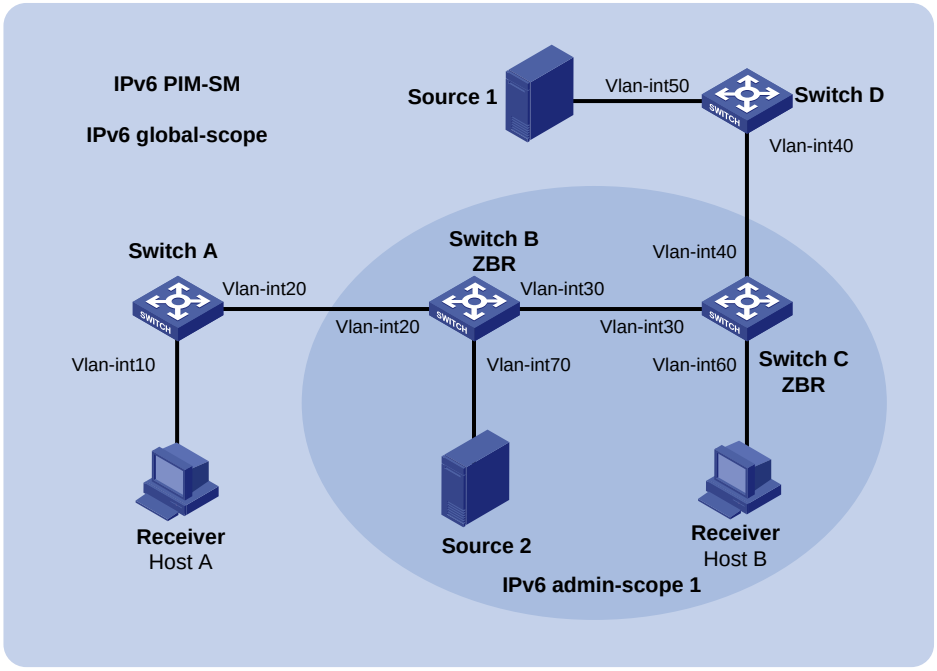


Table 4 Interface and IPv6 address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Source 1	—	500::1/64	Source 2	—	600::1/64
Switch A	Vlan-int10	1001::1/64	Switch D	Vlan-int50	500::2/64
Switch A	Vlan-int20	1002::1/64	Switch D	Vlan-int40	6001::1/64
Switch B	Vlan-int70	600::2/64	Switch C	Vlan-int30	2002::2/64
Switch B	Vlan-int20	1002::2/64	Switch C	Vlan-int60	3002::2/64
Switch B	Vlan-int30	2002::1/64	Switch C	Vlan-int40	6001::2/64

Prerequisites

Assign an IPv6 address and prefix length to each interface, and make sure the switches in the IPv6 PIM-DM domain can reach each other.

Procedure

1. Enable IPv6 multicast routing, MLD, and IPv6 PIM-SM:

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Enable MLD on the receiver-side interface (VLAN-interface 20).

   ```
   [SwitchA] interface vlan-interface 20
   [SwitchA-Vlan-interface20] mld enable
   [SwitchA-Vlan-interface20] quit
   ```

   # Enable IPv6 PIM-SM on VLAN-interface 10.

   ```
   [SwitchA] interface vlan-interface 10
   ```

```
[SwitchA-Vlan-interface10] ipv6 pim sm
[SwitchA-Vlan-interface10] quit
```

Enable IPv6 multicast routing, MLD, and IPv6 PIM-SM on Switch C in the same way Switch A is configured. (Details not shown.)

On Switch B, enable IPv6 multicast routing, and enable IPv6 PIM-SM on each interface.

```
<SwitchB> system-view
[SwitchB] ipv6 multicast routing
[SwitchB-mrib6] quit
[SwitchB] interface vlan-interface 70
[SwitchB-Vlan-interface70] ipv6 pim sm
[SwitchB-Vlan-interface70] quit
[SwitchB] interface vlan-interface 20
[SwitchB-Vlan-interface20] ipv6 pim sm
[SwitchB-Vlan-interface20] quit
[SwitchB] interface vlan-interface 30
[SwitchB-Vlan-interface30] ipv6 pim sm
[SwitchB-Vlan-interface30] quit
```

Enable IPv6 multicast routing and IPv6 PIM-SM on Switch D in the same way Switch B is configured. (Details not shown.)

2. Configure IPv6 admin-scoped zone boundaries:

On Switch B, configure VLAN-interface 20 as the boundary of IPv6 admin-scoped zone 1.

```
[SwitchB] interface vlan-interface 20
[SwitchB-Vlan-interface20] ipv6 multicast boundary scope 4
[SwitchB-Vlan-interface20] quit
```

On Switch C, configure VLAN-interface 40 as the boundary of IPv6 admin-scoped zone 2.

```
<SwitchC> system-view
[SwitchC] interface vlan-interface 40
[SwitchC-Vlan-interface40] ipv6 multicast boundary scope 4
[SwitchC-Vlan-interface40] quit
```

3. Configure C-BSRs and C-RPs:

On Switch D, configure VLAN-interface 50 as a C-BSR and a C-RP for the IPv6 global-scoped zone.

```
[SwitchD] ipv6 pim
[SwitchD-pim6] c-bsr 6001::1
[SwitchD-pim6] c-rp 6001::1
[SwitchD-pim6] quit
```

Verifying the configuration

Display BSR information on Switch B.

```
[SwitchB] display ipv6 pim bsr-info
 Scope: non-scoped
     State: Accept Preferred
     Bootstrap timer: 00:01:25
     Elected BSR address: 6001::1
       Priority: 64
       Hash mask length: 126
       Uptime: 00:01:45
 Scope: 4
     State: Elected
```

```
    Bootstrap timer: 00:00:06
    Elected BSR address: 1002::2
      Priority: 64
      Hash mask length: 126
      Uptime: 00:04:54
    Candidate BSR address: 1002::2
      Priority: 64
      Hash mask length: 126
```

Display BSR information on Switch D.

```
[SwitchD] display ipv6 pim bsr-info
 Scope: non-scoped
    State: Elected
    Bootstrap timer: 00:00:49
    Elected BSR address: 6001::1
      Priority: 64
      Hash mask length: 126
      Uptime: 00:01:11
    Candidate BSR address: 6001::1
      Priority: 64
      Hash mask length: 126
```

Display RP information on Switch B.

```
[SwitchB] display ipv6 pim rp-info
 BSR RP information:
   Scope: non-scoped
     Group/MaskLen: FF00::/8
       RP address               Priority  HoldTime  Uptime    Expires
       6001::1                  192       180       00:38:59  00:02:07
   Scope: 4
     Group/MaskLen: FF04::/16
       RP address               Priority  HoldTime  Uptime    Expires
       1002::2 (local)          10        180       00:39:38  00:02:31
       6001::1                  192       180       00:38:28  00:02:31
     Group/MaskLen: FF14::/16
       RP address               Priority  HoldTime  Uptime    Expires
       1002::2 (local)          10        180       00:39:38  00:02:31
       6001::1                  192       180       00:38:28  00:02:31
     Group/MaskLen: FF24::/16
       RP address               Priority  HoldTime  Uptime    Expires
       1002::2 (local)          10        180       00:39:38  00:02:31
       6001::1                  192       180       00:38:28  00:02:31
     Group/MaskLen: FF34::/16
       RP address               Priority  HoldTime  Uptime    Expires
       1002::2 (local)          10        180       00:39:38  00:02:31
       6001::1                  192       180       00:38:28  00:02:31
     Group/MaskLen: FF44::/16
       RP address               Priority  HoldTime  Uptime    Expires
       1002::2 (local)          10        180       00:39:38  00:02:31
       6001::1                  192       180       00:38:28  00:02:31
```

```
Group/MaskLen: FF54::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FF64::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FF74::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FF84::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FF94::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FFA4::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FFB4::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FFC4::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FFD4::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FFE4::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
Group/MaskLen: FFF4::/16
  RP address               Priority  HoldTime  Uptime    Expires
  1002::2 (local)          10        180       00:39:38  00:02:31
  6001::1                  192       180       00:38:28  00:02:31
```

Example: Configuring IPv6 BIDIR-PIM

Network configuration

As shown in Figure 20:

- OSPFv3 runs on the network.
- VOD streams are sent to receiver hosts in IPv6 multicast.
- Source 1 and Source 2 send IPv6 multicast data to IPv6 multicast group FF14::101.
- Host A and Host B are receivers of this IPv6 multicast group.
- VLAN-interface 102 of Switch C acts as a C-BSR. Loopback 0 of Router C acts as the C-RP.
- MLDv1 runs between Switch B and Host A, and between Switch D and Host B.

Figure 20 Network diagram

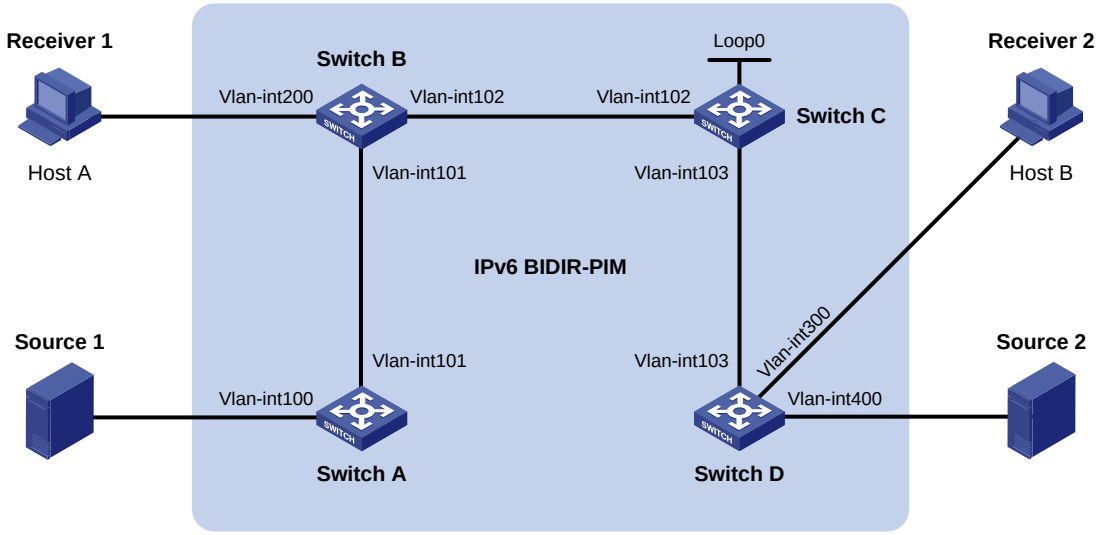


Table 5 Interface and IPv6 address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Switch A	Vlan-int100	1001::1/64	Switch D	Vlan-int300	4001::1/64
Switch A	Vlan-int101	1002::1/64	Switch D	Vlan-int400	5001::1/64
Switch B	Vlan-int200	2001::1/64	Switch D	Vlan-int103	3001::2/64
Switch B	Vlan-int101	1002::2/64	Source 1	—	1001::2/64
Switch B	Vlan-int102	2002::1/64	Source 2	—	5001::2/64
Switch C	Vlan-int102	2002::2/64	Receiver 1	—	2001::2/64
Switch C	Vlan-int103	3001::1/64	Receiver 2	—	4001::2/64
Switch C	Loop0	6001::1/128			

Prerequisites

Assign an IPv6 address and prefix length to each interface as shown in Figure 20, and configure OSPFv3 on the switches in the IPv6 BIDIR-PIM domain.

Procedure

1. Enable IPv6 multicast routing, IPv6 PIM-SM, IPv6 BIDIR-PIM, and MLD:

On Switch A, enable IPv6 multicast routing, enable IPv6 PIM-SM on each interface, and enable IPv6 BIDIR-PIM.

```
<SwitchA> system-view
[SwitchA] ipv6 multicast routing
[SwitchA-mrib6] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipv6 pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] ipv6 pim
[SwitchA-pim6] bidir-pim enable
[SwitchA-pim6] quit
```

On Switch B, enable IPv6 multicast routing.

```
<SwitchB> system-view
[SwitchB] ipv6 multicast routing
[SwitchB-mrib6] quit
```

Enable MLD on the receiver-side interface (VLAN-interface 200).

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] mld enable
[SwitchB-Vlan-interface200] quit
```

Enable IPv6 PIM-SM on the other interfaces.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 pim sm
[SwitchB-Vlan-interface101] quit
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] ipv6 pim sm
[SwitchB-Vlan-interface102] quit
```

Enable IPv6 BIDIR-PIM.

```
[SwitchB] ipv6 pim
[SwitchB-pim6] bidir-pim enable
[SwitchB-pim6] quit
```

On Switch C, enable IPv6 multicast routing, enable IPv6 PIM-SM on each interface, and enable IPv6 BIDIR-PIM.

```
<SwitchC> system-view
[SwitchC] ipv6 multicast routing
[SwitchC-mrib6] quit
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] ipv6 pim sm
[SwitchC-Vlan-interface102] quit
[SwitchC] interface vlan-interface 103
[SwitchC-Vlan-interface103] ipv6 pim sm
[SwitchC-Vlan-interface103] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ipv6 pim sm
[SwitchC-LoopBack0] quit
[SwitchC] ipv6 pim
```

```
[SwitchC-pim6] bidir-pim enable
```

On Switch D, enable IPv6 multicast routing.

```
<SwitchD> system-view
[SwitchD] ipv6 multicast routing
[SwitchD-mrib6] quit
```

Enable MLD on the receiver-side interface (VLAN-interface 300).

```
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] mld enable
[SwitchD-Vlan-interface300] quit
```

Enable IPv6 PIM-SM on the other interfaces.

```
[SwitchD] interface vlan-interface 400
[SwitchD-Vlan-interface400] ipv6 pim sm
[SwitchD-Vlan-interface400] quit
```

Enable IPv6 BIDIR-PIM.

```
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] ipv6 pim sm
[SwitchD-Vlan-interface103] quit
[SwitchD] ipv6 pim
[SwitchD-pim6] bidir-pim enable
[SwitchD-pim6] quit
```

2. On Switch C, configure VLAN interface 102 as a C-BSR, and loopback interface 0 as a C-RP for the entire IPv6 BIDIR-PIM domain.

```
[SwitchC-pim6] c-bsr 2002::2
[SwitchC-pim6] c-rp 6001::1 bidir
[SwitchC-pim6] quit
```

Verifying the configuration

1. Display IPv6 BIDIR-PIM DF information:

Display IPv6 BIDIR-PIM DF information on Switch A.

```
[SwitchA] display ipv6 pim df-info
RP address: 6001::1
  Interface: Vlan-interface100
    State     : Win        DF preference: 100
    DF metric : 2          DF uptime    : 01:08:50
    DF address: FE80::200:5EFF:FE71:2800 (local)
  Interface: Vlan-interface101
    State     : Lose       DF preference: 100
    DF metric : 1          DF uptime    : 01:07:49
    DF address: FE80::20F:E2FF:FE38:4E01
```

Display IPv6 BIDIR-PIM DF information on Switch B.

```
[SwitchB] display ipv6 pim df-info
RP address: 6001::1
  Interface:Vlan-interface200
    State     : Win        DF preference: 100
    DF metric : 1          DF uptime    : 01:24:09
    DF address: FE80::200:5EFF:FE71:2801 (local)
  Interface: Vlan-interface101
    State     : Win        DF preference: 100
```

```
    DF metric : 1          DF uptime    : 01:24:09
    DF address: FE80::20F:E2FF:FE38:4E01 (local)
  Interface: Vlan-interface102
    State     : Lose       DF preference: 0
    DF metric : 0          DF uptime    : 01:23:12
    DF address: FE80::20F:E2FF:FE15:5601
```

Display IPv6 BIDIR-PIM DF information on Switch C.

```
[SwitchC] display ipv6 pim df-info
RP address: 6001::1
  Interface: Loop0
    State     : -          DF preference: -
    DF metric : -          DF uptime    : -
    DF address: -
  Interface: Vlan-interface102
    State     : Win        DF preference: 0
    DF metric : 0          DF uptime    : 01:06:07
    DF address: FE80::20F:E2FF:FE15:5601 (local)
  Interface: Vlan-interface103
    State     : Win        DF preference: 0
    DF metric : 0          DF uptime    : 01:06:07
    DF address: FE80::20F:E2FF:FE15:5602 (local)
```

Display IPv6 BIDIR-PIM DF information on Switch D.

```
[SwitchD] display ipv6 pim df-info
RP address: 6001::1
  Interface: Vlan-interface300
    State     : Win        DF preference: 100
    DF metric : 1          DF uptime    : 01:19:53
    DF address: FE80::200:5EFF:FE71:2803 (local)
  Interface: Vlan-interface400
    State     : Win        DF preference: 100
    DF metric : 1          DF uptime    : 00:39:34
    DF address: FE80::200:5EFF:FE71:2802 (local)
  Interface: Vlan-interface103
    State     : Lose       DF preference: 0
    DF metric : 0          DF uptime    : 01:21:40
    DF address: FE80::20F:E2FF:FE15:5602
```

2. Display information about DFs for IPv6 multicast forwarding:

Display information about DFs for IPv6 multicast forwarding on Switch A.

```
[SwitchA] display ipv6 multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 6001::1
     Flags: 0x0
     Uptime: 00:08:32
     RPF interface: Vlan-interface101
     List of 1 DF interfaces:
       1: Vlan-interface100
```

Display information about DFs for IPv6 multicast forwarding on Switch B.

```
[SwitchB] display ipv6 multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 6001::1
     Flags: 0x0
     Uptime: 00:06:24
     RPF interface: Vlan-interface102
     List of 2 DF interfaces:
       1: Vlan-interface101
       2: Vlan-interface200
```

Display information about DFs for IPv6 multicast forwarding on Switch C.

```
[SwitchC] display ipv6 multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 6001::1
     Flags: 0x0
     Uptime: 00:07:21
     RPF interface: LoopBack0
     List of 2 DF interfaces:
       1: Vlan-interface102
       2: Vlan-interface103
```

Display information about DFs for IPv6 multicast forwarding on Switch D.

```
[SwitchD] display ipv6 multicast forwarding df-info
Total 1 RP, 1 matched

00001. RP address: 6001::1
     Flags: 0x0
     Uptime: 00:05:12
     RPF interface: Vlan-interface103
     List of 2 DF interfaces:
       1: Vlan-interface300
       2: Vlan-interface400
```

Example: Configuring IPv6 PIM-SSM

Network configuration

As shown in Figure 21:

- OSPFv3 runs on the network.
- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire IPv6 PIM domain operates in the SSM mode.
- Host A and Host C are IPv6 multicast receivers on two stub networks N1 and N2.
- The SSM group range is FF3E::/64.
- MLDv2 runs between Switch A and N1, and between Switch B, Switch C, and N2.

Figure 21 Network diagram

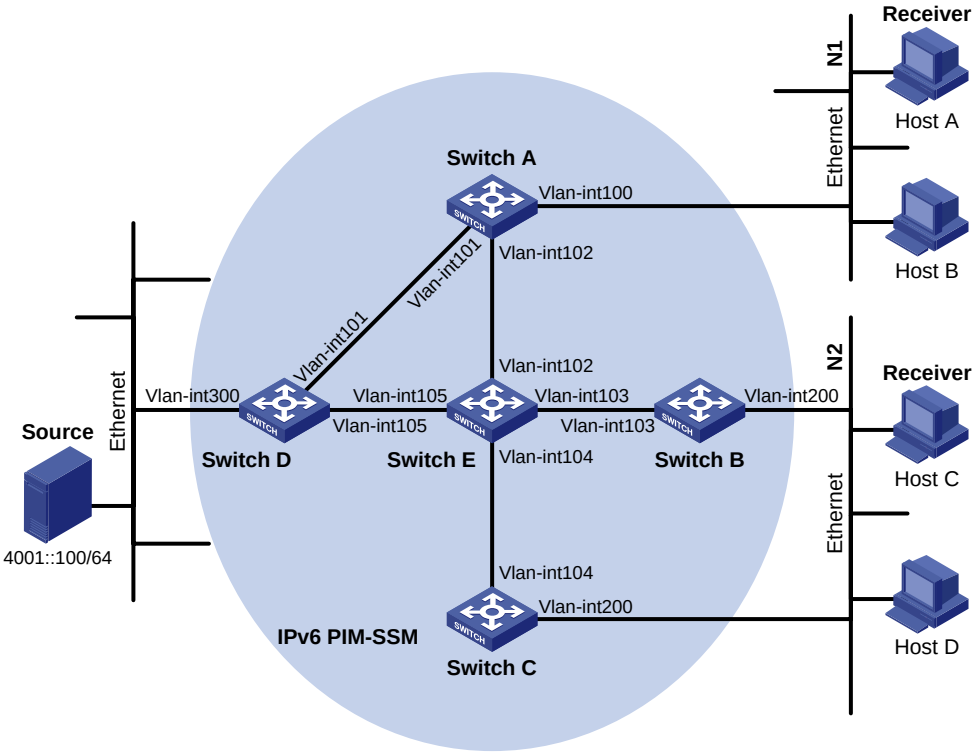


Table 6 Interface and IPv6 address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Switch A	Vlan-int100	1001::1/64	Switch D	Vlan-int300	4001::1/64
Switch A	Vlan-int101	1002::1/64	Switch D	Vlan-int101	1002::2/64
Switch A	Vlan-int102	1003::1/64	Switch D	Vlan-int105	4002::1/64
Switch B	Vlan-int200	2001::1/64	Switch E	Vlan-int104	3001::2/64
Switch B	Vlan-int103	2002::1/64	Switch E	Vlan-int103	2002::2/64
Switch C	Vlan-int200	2001::2/64	Switch E	Vlan-int102	1003::2/64
Switch C	Vlan-int104	3001::1/64	Switch E	Vlan-int105	4002::2/64

Prerequisites

Assign an IPv6 address and prefix length to each interface as shown in Figure 21, and configure OSPFv3 on the switches in the IPv6 PIM-SSM domain.

Procedure

1. Enable IPv6 multicast routing, MLD, and IPv6 PIM-SM:

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Enable MLDv2 on VLAN-interface 100 (the interface that connects to the stub network).

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] mld enable
   ```

```
[SwitchA-Vlan-interface100] mld version 2
[SwitchA-Vlan-interface100] quit
```

Enable IPv6 PIM-SM on the other interfaces.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] ipv6 pim sm
[SwitchA-Vlan-interface102] quit
```

Enable IPv6 multicast routing, MLD, and IPv6 PIM-SM on Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

Enable IPv6 multicast routing and IPv6 PIM-SM on Switch D and Switch E in the same way Switch A is configured. (Details not shown.)

2. Configure the IPv6 SSM group range FF3E::/64 on Switch A.

```
[SwitchA] acl ipv6 basic 2000
[SwitchA-acl-ipv6-basic-2000] rule permit source ff3e:: 64
[SwitchA-acl-ipv6-basic-2000] quit
[SwitchA] ipv6 pim
[SwitchA-pim6] ssm-policy 2000
[SwitchA-pim6] quit
```

3. Configure the IPv6 SSM group range on Switch B, Switch C, Switch D, and Switch E in the same way Switch B is configured. (Details not shown.)

Verifying the configuration

Display IPv6 PIM information on Switch A.

```
[SwitchA] display ipv6 pim interface
 Interface: Vlan101
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: FE80::A01:201:2
 Interface: Vlan102
   NbrCnt: 1
   HelloInt: 30
   DR priority: 1
   DR address: FE80::A01:201:3
```

Send an MLDv2 report from Host A to join the IPv6 multicast source and group (4001::100/64, FF3E::101). (Details not shown.)

Display IPv6 PIM multicast routing table information on Switch A.

```
[SwitchA] display ipv6 pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry


 (4001::100, FF3E::101)
     Protocol: pim-ssm, Flag: ACT
     UpTime: 00:00:11
     Upstream interface: Vlan-interface101
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
```

```
    Downstream interface(s) information:
    Total number of downstream interfaces: 1
        1: Vlan-interface100
            Protocol: mld, UpTime: 00:00:11, Expires: 00:03:25
```

Display IPv6 PIM multicast routing table information on Switch B.

```
[SwitchD] display ipv6 pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF3E::101)
    Protocol: pim-ssm, Flag: LOC
    UpTime: 00:08:02
    Upstream interface: Vlan-interface300
        Upstream neighbor: NULL
        RPF prime neighbor: NULL
    Downstream interface(s) information:
    Total number of downstream interfaces: 1
        1: Vlan-interface105
            Protocol: pim-ssm, UpTime: 00:08:02, Expires: 00:03:25
```

The output shows that switches on the SPT path (Switch A and Switch D) have generated the correct (S, G) entries.

Troubleshooting IPv6 PIM

A multicast distribution tree cannot be correctly built

Symptom

No IPv6 multicast forwarding entries are established on the devices (including devices directly connected with multicast sources or receivers) in an IPv6 PIM network. This means that an IPv6 multicast distribution tree cannot be correctly built.

Solution

To resolve the problem:

1. Use the **`display ipv6 routing-table`** command to verify that an IPv6 unicast route to the IPv6 multicast source or the RP is available.
2. Use the **`display ipv6 pim interface`** command to verify IPv6 PIM information on each interface, especially on the RPF interface. If IPv6 PIM is not enabled on the interfaces, use **`ipv6 pim dm`** or **`ipv6 pim sm`** to command enable IPv6 PIM-DM or IPv6 PIM-SM for the interfaces.
3. Use the **`display ipv6 pim neighbor`** command to verify that the RPF neighbor is an IPv6 PIM neighbor.
4. Verify that IPv6 PIM and MLD are enabled on the interfaces that directly connect to the IPv6 multicast sources or the receivers.
5. Use the **`display ipv6 pim interface verbose`** command to verify that the same IPv6 PIM mode is enabled on the RPF interface on a device and the connected interface of the device's RPF neighbor.
6. Use the **`display current-configuration`** command to verify that the same IPv6 PIM mode is enabled on all devices on the network. For IPv6 PIM-SM, verify that the BSR and C-RPs are correctly configured.

7. If the problem persists, contact H3C Support.

IPv6 multicast data is abnormally terminated on an intermediate device

Symptom

An intermediate device can receive IPv6 multicast data successfully, but the data cannot reach the last-hop device. An interface on the intermediate device receives IPv6 multicast data but does not create an (S, G) entry in the IPv6 PIM routing table.

Solution

To resolve the problem:

1. Use the **display current-configuration** command to verify the IPv6 multicast forwarding boundary settings. Use the **ipv6 multicast boundary** command to change the multicast forwarding boundary settings to make the IPv6 multicast packet able to cross the boundary.
2. Use the **display current-configuration** command to verify the IPv6 multicast source policy. Change the ACL rule defined in the **source-policy** command so that the source/group address of the IPv6 multicast data can pass ACL filtering.
3. If the problem persists, contact H3C Support.

An RP cannot join an SPT in IPv6 PIM-SM

Symptom

An RPT cannot be correctly built, or an RP cannot join the SPT toward the IPv6 multicast source.

Solution

To resolve the problem:

1. Use the **display ipv6 routing-table** command to verify that an IPv6 unicast route to the RP is available on each device.
2. Use the **display ipv6 pim rp-info** command to verify that the dynamic RP information is consistent on all devices.
3. Use the **display ipv6 pim rp-info** command to verify that the same static RPs are configured on all devices on the network.
4. If the problem persists, contact H3C Support.

An RPT cannot be built or IPv6 multicast source registration fails in IPv6 PIM-SM

Symptom

The C-RPs cannot unicast advertisement messages to the BSR. The BSR does not advertise BSMs containing C-RP information and has no IPv6 unicast route to any C-RP. An RPT cannot be correctly established, or the source-side DR cannot register the IPv6 multicast source with the RP.

Solution

To resolve the problem:

1. Use the **display ipv6 routing-table** command on each device to view routing table information. Verify that IPv6 unicast routes to the C-RPs and the BSR are available on each device and that a route is available between each C-RP and the BSR.

2. Use the **display ipv6 pim bsr-info** command to verify that the BSR information exists on each device.
3. Use the **display ipv6 pim rp-info** command to verify that the RP information is correct on each device.
4. Use the **display ipv6 pim neighbor** command to verify that IPv6 PIM neighboring relationship has been correctly established among the devices.
5. If the problem persists, contact H3C Support.

Contents

Multicast VPN overview 1

Typical network diagram 1
MVPN scheme 1
Basic concepts in MDT-based MVPN 1
How MDT-based MVPN works 2
Default MDT establishment 2
Default MDT-based delivery 6
MDT switchover 8
Inter-AS MDT-based MVPN 9
MDT-based MVPN inter-AS option A 9
MDT-based MVPN inter-AS option B 10
MDT-based MVPN inter-AS option C 11
M6VPE 12
MVPN extranet 12
Protocols and standards 15

Configuring multicast VPN 16

Restrictions and guidelines: Multicast VPN configuration 16
MDT-based MVPN tasks at a glance 16
Configuring MDT-based MVPN 16
Prerequisites for configuring MDT-based MVPN 16
Enabling IP multicast routing for a VPN instance 16
Creating an MDT-based MVPN instance 17
Creating an MVPN address family 18
Specifying the default group 18
Specifying the MVPN source interface 18
Configuring MDT switchover parameters 19
Configuring the RPF vector feature 20
Enabling data group reuse logging 20
Setting the DSCP value for outgoing data group switchover packets 21
Configuring BGP MDT 21
Configuring BGP MDT peers or peer groups 21
Configuring a BGP MDT route reflector 22
Configuring BGP MDT optimal route selection delay 23
Configuring an MVPN extranet RPF selection policy 23
Display and maintenance commands for multicast VPN 24
Multicast VPN configuration examples 26
Example: Configuring intra-AS MDT-based MVPN 26
Example: Configuring intra-AS M6VPE 39
Example: Configuring MDT-based MVPN inter-AS option B 54
Example: Configuring MDT-based MVPN inter-AS option C 80
Example: Configuring receiver-PE-based MVPN extranet 106
Troubleshooting MDT-based MVPN 121
A default MDT cannot be established 121
An MVRF cannot be created 122

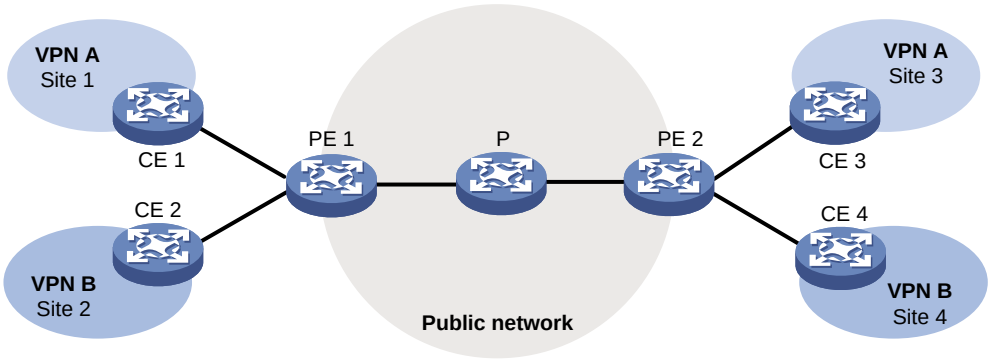
Multicast VPN overview

Multicast VPN implements multicast delivery in VPNs.

Typical network diagram

As shown in Figure 1, VPN A contains Site 1 and Site 3, and VPN B contains Site 2 and Site 4.

Figure 1 Typical network diagram for multicast VPN



VPN multicast traffic between the PEs and the CEs is transmitted on a per-VPN-instance basis. The public network multicast traffic between the PEs and the P device is transmitted through the public network. Multicast VPN provides independent multicast services for the public network, VPN A, and VPN B.

For more information about CEs, PEs and Ps, see *MPLS Configuration Guide*.

MVPN scheme

MVPN is used to implement multicast VPN. MVPN only requires the PEs to support multiple VPN instances and the public network provided by the service provider to support multicast. There is no need to upgrade CEs and Ps or change their PIM configurations. The MVPN solution is transparent to CEs and Ps.

The device supports MDT-based MVPN.

Basic concepts in MDT-based MVPN

This section introduces the following basic concepts in MDT-based MVPN:

- **MVPN**—An MVPN logically defines the transmission boundary of the multicast traffic of a VPN over the public network. It also physically identifies all the PEs that support that VPN instance on the public network. Different VPN instances correspond to different MVPNs.
- **Multicast distribution tree (MDT)**—An MDT is a multicast distribution tree constructed by all PEs in the same VPN. MDT includes default MDT and data MDT.
- **Multicast tunnel (MT)**—An MT is a tunnel that interconnects all PEs in an MVPN. The local PE encapsulates a VPN multicast packet into a public network multicast packet and forwards it through the MT over the public network. The remote PE decapsulates the public network multicast packet to get the original VPN multicast packet.
- **Multicast tunnel interface (MTI)**—An MTI is the entrance or exit of an MT, equivalent to an entrance or exit of an MVPN. MTIs are automatically created when the MVPN for the VPN

instance is created. PEs use the MTI to access the MT. The local PE sends VPN data out of the MTI. The remote PEs receive the private data from their MTIs. An MTI runs the same PIM mode as the VPN instance to which the MTI belongs. PIM is enabled on MTIs when a minimum of one interface in the VPN instance is enabled with PIM. When PIM is disabled on all interfaces in the VPN instance, PIM is also disabled on MTIs.

- **Default-group**—A default group is a unique multicast address assigned to each MVPN on the public network. It is the unique identifier of an MVPN on the public network and helps build the default MDT for an MVPN on the public network. A PE encapsulates a VPN multicast packet (a multicast protocol packet or a multicast data packet) into a public network multicast packet. The default group address is used as the public network multicast group.
- **Default MDT**—A default MDT uses a default group address as its group address. In a VPN, the default MDT is uniquely identified by the default group. A default MDT is automatically created after the default group is specified and will always exist on the public network, regardless of the presence of any multicast services on the public network or the VPN.
- **Data group**—An MVPN is assigned a unique data group for MDT switchover. The ingress PE selects a least used address from the data group range to encapsulate the VPN multicast packets when the multicast traffic of the VPN arrives at the PE. Other PEs are notified to use the address to forward the multicast traffic for that VPN. This initiates the switchover to the data MDT.
- **Data MDT**—A data MDT is an MDT that uses a data group as it group address. At MDT switchover, PEs with downstream receivers join a data group to build a data MDT. The ingress PE forwards the encapsulated VPN multicast traffic along the data MDT over the public network.

How MDT-based MVPN works

For a VPN instance, multicast data transmission on the public network is transparent. The VPN data is exchanged between the MTIs of the local PE and the remote PE. This implements the seamless transmission of the VPN data over the public network. However, the multicast data transmission process (the MDT transmission process) over the public network is very complicated.

NOTE:

The following types of PIM neighboring relationships exist in MVPN:

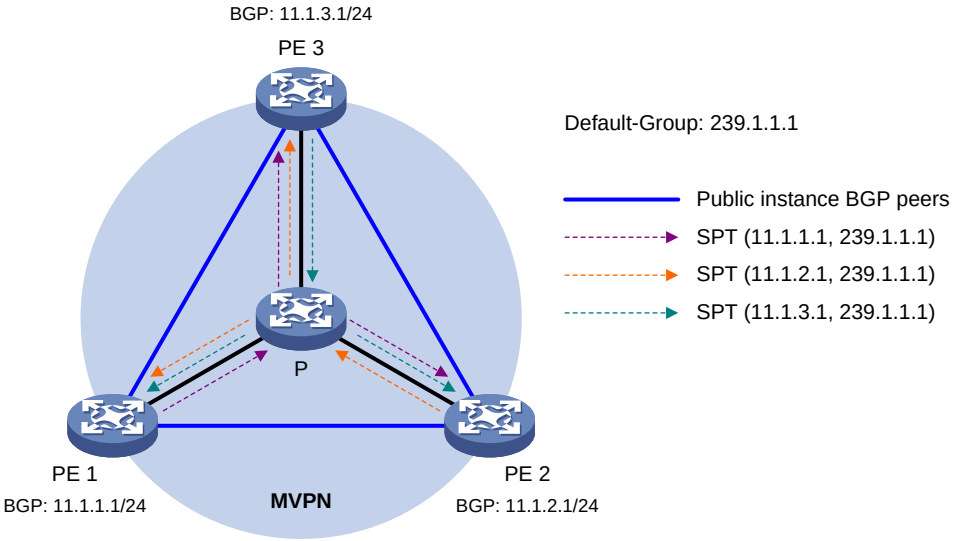
- **PE-P PIM neighboring relationship**—Established between the public network interface on a PE and the peer interface on the P device over the link.
- **PE-PE PIM neighboring relationship**—Established between PEs that are in the same VPN instance after they receive the PIM hello packets.
- **PE-CE PIM neighboring relationship**—Established between a PE interface that is bound with the VPN instance and the peer interface on the CE over the link.

Default MDT establishment

The multicast routing protocol running on the public network can be PIM-DM, PIM-SM, BIDIR-PIM, or PIM-SSM. The process of creating a Default MDT is different in these PIM modes.

Default MDT establishment in a PIM-DM network

Figure 2 Default MDT establishment in a PIM-DM network



As shown in Figure 2, PIM-DM is enabled on the network, and all the PEs support VPN instance A. The process of establishing a default MDT is as follows:

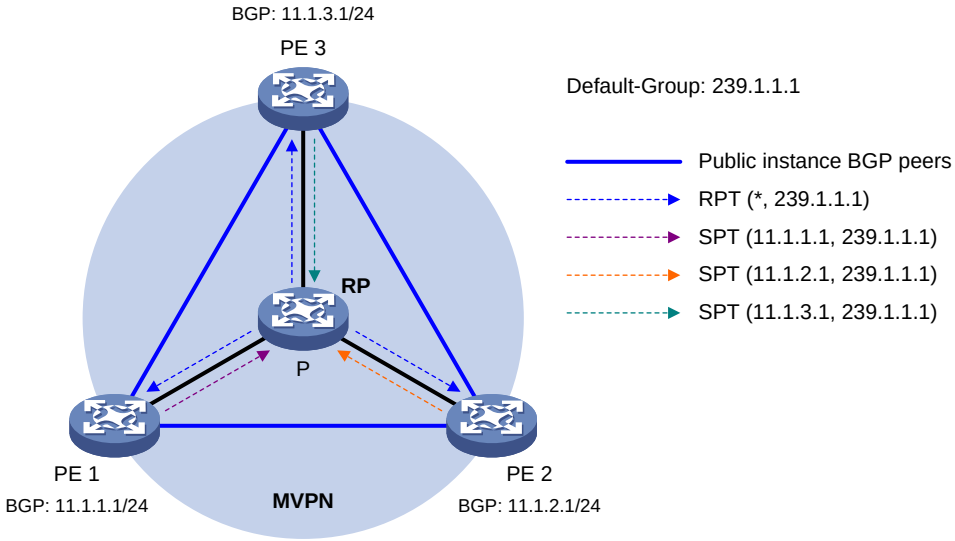
1. To establish PIM neighboring relationships with PE 2 and PE 3 through the MTI for VPN instance A, PE 1 does the following:
   a. Encapsulates the PIM protocol packet of the private network into a public network multicast data packet. PE 1 does this by specifying the source address as the IP address of the MVPN source interface and the multicast group address as the default group address.
   b. Sends the multicast data packet to the public network.

   For other PEs that support VPN instance A as default group members, PE 1 of VPN instance A initiates a flood-prune process in the entire public network. A (11.1.1.1, 239.1.1.1) state entry is created on each device along the path on the public network. This forms an SPT with PE 1 as the root, and PE 2 and PE 3 as leaves.
2. At the same time, PE 2 and PE 3 separately initiate a similar flood-prune process.

   Finally, three independent SPTs are established in the MVPN, constituting the default MDT in the PIM-DM network.

Default MDT establishment in a PIM-SM network

Figure 3 Default MDT establishment in a PIM-SM network



As shown in Figure 3, PIM-SM is enabled on the network, and all the PEs support VPN instance A. The process of establishing a default MDT is as follows:

1. PE 1 initiates a join to the public network RP by specifying the multicast group address as the default group address in the join message. A (*, 239.1.1.1) state entry is created on each device along the path on the public network.
2. At the same time, PE 2 and PE 3 separately initiate a similar join process.

   Finally, an RPT is established in the MVPN, with the public network RP as the root and PE 1, PE 2, and PE 3 as leaves.
3. To establish PIM neighboring relationships with PE 2 and PE 3 through the MTI for VPN instance A, PE 1 does the following:

   a. Encapsulates the PIM protocol packet of the private network into a public network multicast data packet. PE 1 does this by specifying the source address as the IP address of the MVPN source interface and the multicast group address as the default group address.

   b. Sends the multicast data packet to the public network.

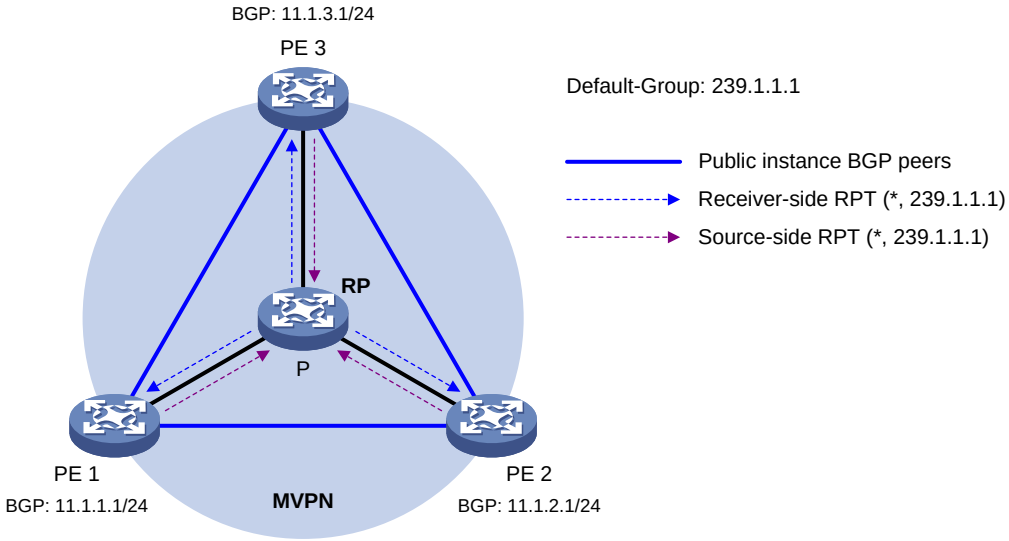
   The public network interface of PE 1 registers the multicast source with the public network RP, and the public network RP initiates a join to PE 1. A (11.1.1.1, 239.1.1.1) state entry is created on each device along the path on the public network.
4. At the same time, PE 2 and PE 3 separately initiate a similar register process.

   Finally, three SPTs between the PEs and the RP are established in the MVPN.

In the PIM-SM network, the RPT, or the (*, 239.1.1.1) tree, and the three independent SPTs constitute the default MDT.

Default MDT establishment in a BIDIR-PIM network

Figure 4 Default MDT establishment in a BIDIR-PIM network



As shown in Figure 4, BIDIR-PIM runs on the network, and all the PEs support VPN instance A. The process of establishing a default MDT is as follows:

1. PE 1 initiates a join to the public network RP by specifying the multicast group address as the default group address in the join message. A (*, 239.1.1.1) state entry is created on each device along the path on the public network.

   At the same time, PE 2 and PE 3 separately initiate a similar join process. Finally, a receiver-side RPT is established in the MVPN, with the public network RP as the root and PE 1, PE 2, and PE 3 as leaves.

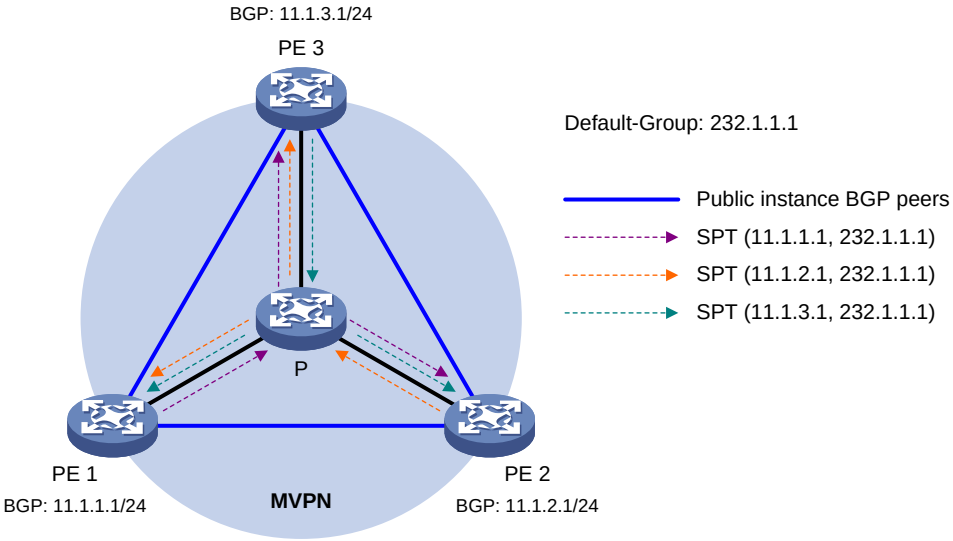
2. PE 1 sends a multicast packet with the default group address as the multicast group address. The DF of each network segment on the public network forwards the multicast packet to the RP. Each device on the path creates a (*, 239.1.1.1) state entry.

   At the same time, PE 2 and PE 3 separately initiate a similar process. Finally, three source-side RPTs are established in the MVPN, with PE 1, PE 2, and PE 3 as the roots and as the public network RP as the leave.

3. The receiver-side RPT and the three source-side RPTs constitute the default MDT in the BIDIR-PIM network.

Default MDT establishment in a PIM-SSM network

Figure 5 Default MDT establishment in a PIM-SSM network



As shown in Figure 5, PIM-SSM runs on the network, and all the PEs support VPN instance A. The process of establishing a default MDT is as follows:

1. PE 1, PE 2, and PE 3 exchange MDT route information (including BGP interface address and the default group address) through BGP.
2. PE 1 sends a subscribe message to PE 2 and PE 3. Each device on the public network creates an (S, G) entry. An SPT is established in the MVPN with PE 1 as the root and PE 2 and PE 3 as the leaves.

   At the same time, PE 2 and PE 3 separately initiate a similar process, and establish an SPT with itself as the root and the other PEs as the leaves.
3. The three independent SPTs constitute the default MDT in the PIM-SSM network.

In PIM-SSM, the term "subscribe message" refers to a join message.

Default MDT characteristics

No matter which PIM mode is running on the public network, the default MDT has the following characteristics:

- All PEs that support the same VPN instance join the default MDT.
- All multicast packets that belong to this VPN are forwarded along the default MDT to every PE on the public network, even if no active downstream receivers exist.

Default MDT-based delivery

After the default MDT is established, the multicast source forwards the VPN multicast data to the receivers in each site along the default MDT. The VPN multicast packets are encapsulated into public network multicast packets on the local PE, and transmitted along the default MDT. Then, they are decapsulated on the remote PE and transmitted in that VPN site.

VPN multicast data packets are forwarded across the public network differently in the following circumstances:

- If PIM-DM or PIM-SSM is running in the VPN, the multicast source forwards multicast data packets to the receivers along the VPN SPT across the public network.
- When PIM-SM is running in the VPN:

- Before the RPT-to-SPT switchover, if the multicast source and the VPN RP are in different sites, the VPN multicast data packets travel to the VPN RP along the VPN SPT across the public network. If the VPN RP and the receivers are in different sites, the VPN multicast data packets travel to the receivers along the VPN RPT over the public network.
- After the RPT-to-SPT switchover, if the multicast source and the receivers are in different sites, the VPN multicast data packets travel to the receivers along the VPN SPT across the public network.

- When BIDIR-PIM is running in the VPN, if the multicast source and the VPN RP are in different sites, the multicast source sends multicast data to the VPN RP across the public network along the source-side RPT. If the VPN RP and the receivers are in different sites, the multicast data packets travel to the receivers across the public network along the receiver-side RPT.

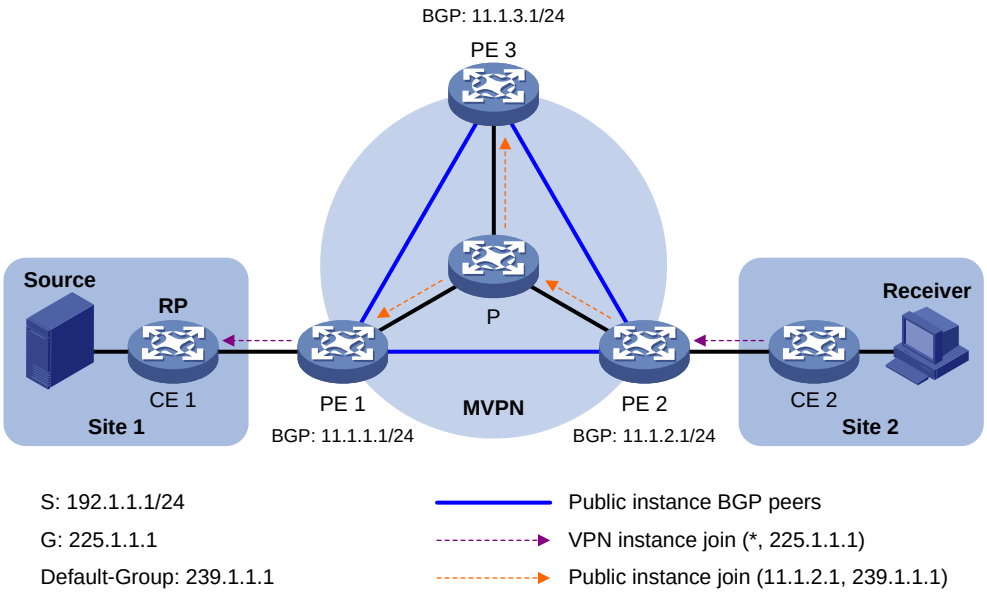
  For more information about RPT-to-SPT switchover, see "PIM overview."

The following example explains how multicast data packets are delivered based on the default MDT when PIM-DM is running in both the public network and the VPN network.

As shown in Figure 6:

- PIM-DM is running in both the public network and the VPN sites.
- Receiver of the VPN multicast group G (225.1.1.1) in Site 2 is attached to CE 2.
- Source in Site 1 sends multicast data to multicast group (G).
- The default group address used to forward public network multicast data is 239.1.1.1.

Figure 6 Multicast data packet delivery



A VPN multicast data packet is delivered across the public network as follows:

1. Source sends a VPN multicast data packet (192.1.1.1, 225.1.1.1) to CE 1.
2. CE 1 forwards the VPN multicast data packet along an SPT to PE 1, and the VPN instance on PE 1 examines the MVRF.

   If the outgoing interface list of the forwarding entry contains an MTI, PE 1 processes the VPN multicast data packet as described in step 3. The VPN instance on PE 1 considers the VPN multicast data packet to have been sent out of the MTI, because step 3 is transparent to it.
3. PE 1 encapsulates the VPN multicast data packet into a public network multicast packet (11.1.1.1, 239.1.1.1) by using the GRE method. The source IP address of the packet is the MVPN source interface 11.1.1.1, and the destination address is the default group address 239.1.1.1. PE 1 then forwards it to the public network.

4. The default MDT forwards the multicast data packet (11.1.1.1, 239.1.1.1) to the public network instance on all the PEs. After receiving this packet, every PE decapsulates it to get the original VPN multicast data packet, and passes it to the corresponding VPN instance. If a PE has a downstream interface for an SPT, it forwards the VPN multicast packet down the SPT. Otherwise, it discards the packet.
5. The VPN instance on PE 2 looks up the MVRF and finally delivers the VPN multicast data to Receiver.

   By now, the process of transmitting a VPN multicast data packet across the public network is completed.

MDT switchover

Switching from default MDT to data MDT

When a multicast packet of a VPN is transmitted through the default MDT on the public network, the packet is forwarded to all PEs that support that VPN instance. This occurs whether or not any active receivers exist in the attached sites. When the rate of the multicast traffic of that VPN is high, multicast data might get flooded on the public network. This increases the bandwidth use and brings extra burden on the PEs.

To optimize multicast transmission of large VPN multicast traffic that enters the public network, the MDT-based MVPN solution introduces a dedicated data MDT. The data MDT is built between the PEs that connect VPN multicast receivers and multicast sources. When specific network criteria are met, a switchover from the default MDT to the data MDT occurs to forward VPN multicast traffic to receivers.

The device initiates a switchover of the default MDT to the data MDT as follows:

1. The source-side PE (PE 1, for example) starts a data-delay timer when receiving the first multicast VPN packet that matches the switchover criterion.
2. When the data-delay timer expires, PE 1 sends an MDT switchover message to all the other PE devices down the default-MDT. This message contains the multicast source address, the multicast group address, and the least-used data-group address in the data-group range.
3. Each PE that receives this message examines whether it interfaces with a VPN that has receivers of that VPN multicast stream.

   If so, it joins the data MDT rooted at PE 1. Otherwise, it caches the message and will join the data MDT when it has attached receivers.
4. After sending the MDT switchover message, PE 1 starts the data-delay timer. When the timer expires, PE 1 uses the default group address to encapsulate the VPN multicast data. The multicast data is then forwarded down the data MDT.
5. After the multicast traffic is switched from the default MDT to the data MDT, PE 1 continues sending MDT switchover messages periodically. Subsequent PEs with attached receivers can then join the data MDT. When a downstream PE no longer has active receivers attached to it, it leaves the data MDT.

For a given VPN instance, the default MDT and the data MDT are both forwarding tunnels in the same MVPN. A default MDT is uniquely identified by a default group address, and a data MDT is uniquely identified by a data group address. Each default group is uniquely associated with a data group range.

Backward switching from data MDT to default MDT

After the VPN multicast traffic is switched to the data MDT, the multicast traffic conditions might change and no longer meet the switchover criterion. In this case, PE 1, as in the preceding example, initiates a backward MDT switchover process when any of the following criteria are met:

- The traffic rate of the VPN multicast data has dropped below the switchover threshold. In addition, the traffic rate has stayed lower than the threshold for a certain length of time (known as the data holddown period).

- The associated data group range is changed, and the data group address for encapsulating the VPN multicast data is out of the new address range.
- The ACL rule for controlling the switchover from the default MDT to the data MDT has changed, and the VPN multicast data fails to pass the new ACL rule.

Inter-AS MDT-based MVPN

In an inter-AS VPN networking scenario, VPN sites are located in multiple ASs. These sites must be interconnected. Inter-AS VPN provides the following solutions:

- **VRF-to-VRF connections between ASBRs**—This solution is also called inter-AS option A.
- **EBGP redistribution of labeled VPN-IPv4 routes between ASBRs**—ASBRs advertise VPN-IPv4 routes to each other through MP-EBGP. This solution is also called inter-AS option B.
- **Multihop EBGP redistribution of labeled VPN-IPv4 routes between PE devices**—PEs advertise VPN-IPv4 routes to each other through MP-EBGP. This solution is also called inter-AS option C.

For more information about the three inter-AS VPN solutions, see MPLS L3VPN configuration in *MPLS Configuration Guide*.

Based on these solutions, there are three ways to implement inter-AS MVPN:

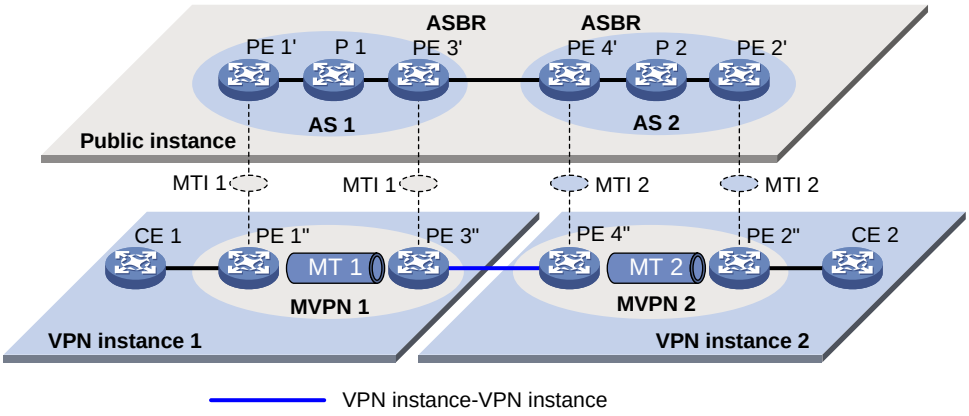
- MDT-based MVPN inter-AS option A.
- MDT-based MVPN inter-AS option B.
- MDT-based MVPN inter-AS option C.

MDT-based MVPN inter-AS option A

As shown in Figure 7:

- Two VPN instances are in AS 1 and AS 2.
- PE 3 and PE 4 are ASBRs for AS 1 and AS 2, respectively.
- PE 3 and PE 4 are interconnected through their respective VPN instance and treat each other as a CE.

Figure 7 MVPN inter-AS option A



To implement MVPN inter-AS option A, a separate MVPN must be created in each AS. Multicast data is transmitted between the VPNs in different ASs through the MDs.

Multicast packets of VPN instance 1 are delivered as follows:

1. CE 1 forwards the multicast packet of VPN instance 1 to PE 1.

2. PE 1 encapsulates the multicast packet into a public network packet and forwards it to PE 3 through the MTI interface in MVPN 1.
3. PE 3 considers PE 4 as a CE of VPN instance 1, so PE 3 forwards the multicast packet to PE 4.
4. PE 4 considers PE 3 as a CE of VPN instance 2, so it forwards the multicast packet to PE 2 through the MTI interface in MVPN 2 on the public network.
5. PE 2 forwards the multicast packet to CE 2.

Because only VPN multicast data is forwarded between ASBRs, different PIM modes can run within different ASs. However, the same PIM mode must run on all interfaces that belong to the same VPN (including interfaces with VPN bindings on ASBRs).

MDT-based MVPN inter-AS option B

In MVPN inter-AS option B, RPF vector and BGP connector are introduced:

- **RPF vector**—Attribute encapsulated in a PIM join message. It is the next hop of BGP MDT route from the local PE to the remote PE. Typically, it is the ASBR in the local AS.

  When a device receives the join message with the RPF vector, it first checks whether the RPF vector is its own IP address. If so, the device removes the RPF vector, and sends the message to its upstream neighbor according to the route to the remote PE. Otherwise, it keeps the RPF vector, looks up the route to the RPF vector, and sends the message to the next hop of the route. In this way, the PIM message can be forwarded across the ASs and an MDT is established.
- **BGP connector**—Attribute shared by BGP peers when they exchange IPv4 VPN routes. It is the IP address of the remote PE.

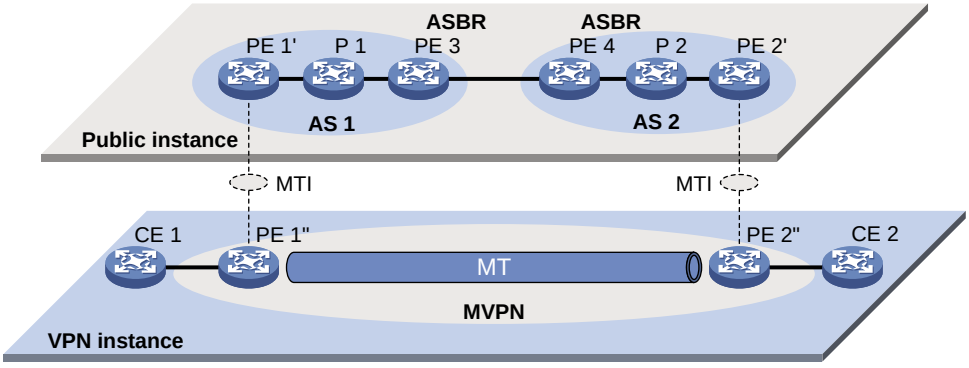
  The local PE fills the upstream neighbor address field with the BGP connector in a join message. This ensures that the message can pass the RPF check on the remote PE after it travels along the MT.

To implement MVPN inter-AS option B, only one MVPN needs to be established for the two ASs. VPN multicast data is transmitted between different ASs on the public network within this MVPN.

As shown in Figure 8:

- A VPN network involves AS 1 and AS 2.
- PE 3 and PE 4 are the ASBRs for AS 1 and AS 2, respectively.
- PE 3 and PE 4 are interconnected through MP-EBGP and treat each other as a P device.
- PE 3 and PE 4 advertise VPN-IPv4 routes to each other through MP-EBGP.
- An MT is established between PE 1 and PE 2 for delivering VPN multicast traffic across the ASs.

Figure 8 MVPN inter-AS option B



The establishment of the MDT on the public network is as follows:

1. PE 1 originates a PIM join message to join the SPT rooted at PE 2. In the join message, the upstream neighbor address is the IP address of PE 2 (the BGP connector). The RPF vector attribute is the IP address of PE 3. PE 1 encapsulates the join message as a public network packet and forwards it through the MTI.
2. P 1 determines that the RPF vector is not an IP address of its own. It looks up the routing table for a route to PE 3, and forwards the packet to PE 3.
3. PE 3 removes the RPF vector because the RPF vector is its own IP address. It fails to find a BGP MDT route to PE 2, so it encapsulates a new RPF vector (IP address of PE 4) in the packet and forwards it to PE 4.
4. PE 4 removes the RPF vector because the RPF vector is its own IP address. It has a local route to PE 2, so it forwards the packet to P 2, which is the next hop of the route to PE 2.
5. P 2 sends the packet to PE 2.
6. PE 2 receives the packet on the MTI and decapsulates the packet. The receiving interface is the RPF interface of the RPF route back to PE 1 for the join message, and the join message passes the RPF check. The SPT from PE 1 to PE 2 is established.

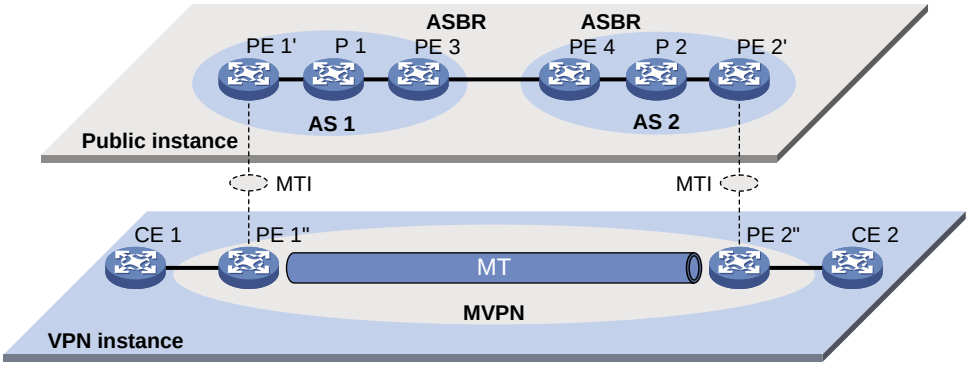
When PE 1 joins the SPT rooted at PE 1, PE 2 also initiates a join process to the SPT rooted at PE 1. A MDT is established when the two SPTs are finished.

MDT-based MVPN inter-AS option C

As shown in Figure 9:

- A VPN network involves AS 1 and AS 2.
- PE 3 and PE 4 are the ASBRs for AS 1 and AS 2, respectively.
- PE 3 and PE 4 are interconnected through MP-EBGP and treat each other as a P device.
- PEs in different ASs establish a multihop MP-EBGP session to advertise VPN-IPv4 routes to each other.

Figure 9 MVPN inter-AS option C



To implement MVPN inter-AS option C, only one MVPN needs to be created for the two ASs. Multicast data is transmitted between the two ASs through the MVPN.

Multicast packets are delivered as follows:

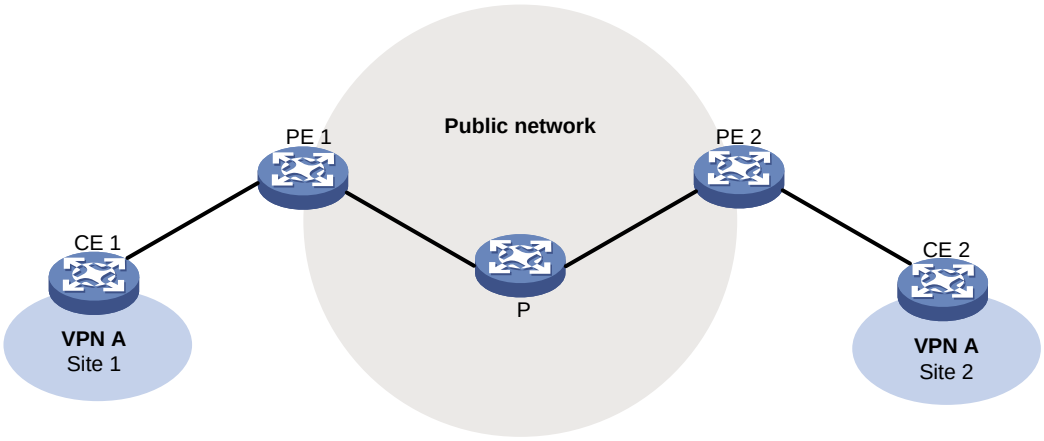
1. CE 1 forwards the VPN instance multicast packet to PE 1.
2. PE 1 encapsulates the multicast packet into a public network multicast packet and forwards it to PE 3 through the MTI interface on the public network.
3. PE 3 and PE 4 are interconnected through MP-EBGP, so PE 3 forwards the public network multicast packet to PE 4 along the VPN IPv4 route.
4. The public network multicast packet arrives at the MTI interface of PE 2 in AS 2. PE 2 decapsulates the public network multicast packet and forwards the VPN multicast packet to CE 2.

M6VPE

The multicast IPv6 VPN provider edge (M6VPE) feature enables PEs to transmit IPv6 multicast traffic of a VPN instance over the public network. Only the IPv4 network is available for the backbone network.

As shown in Figure 10, the public network runs IPv4 protocols, and sites of VPN instance **VPN A** run IPv6 multicast protocols. To transmit IPv6 multicast traffic between CE 1 and CE 2, configure M6VPE on the PEs.

Figure 10 M6VPE network



IPv6 multicast traffic forwarding over the IPv4 public network is as follows:

1. CE 1 forwards an IPv6 multicast packet for VPN instance **VPN A** to PE 1.
2. PE 1 encapsulates the IPv6 multicast packet with an IPv4 packet header and transmits the IPv4 packet in the IPv4 backbone network.
3. PE 2 decapsulates the IPv4 packet and forwards the IPv6 multicast packet to CE 2.

MVPN extranet

MVPN extranet implements inter-VPN multicast traffic transmission. In MVPN extranet, the multicast source and receivers are in different VPNs. In actual application, the MVPN extranet solution ensures that the service provider in a VPN provides multicast services for users in other VPNs.

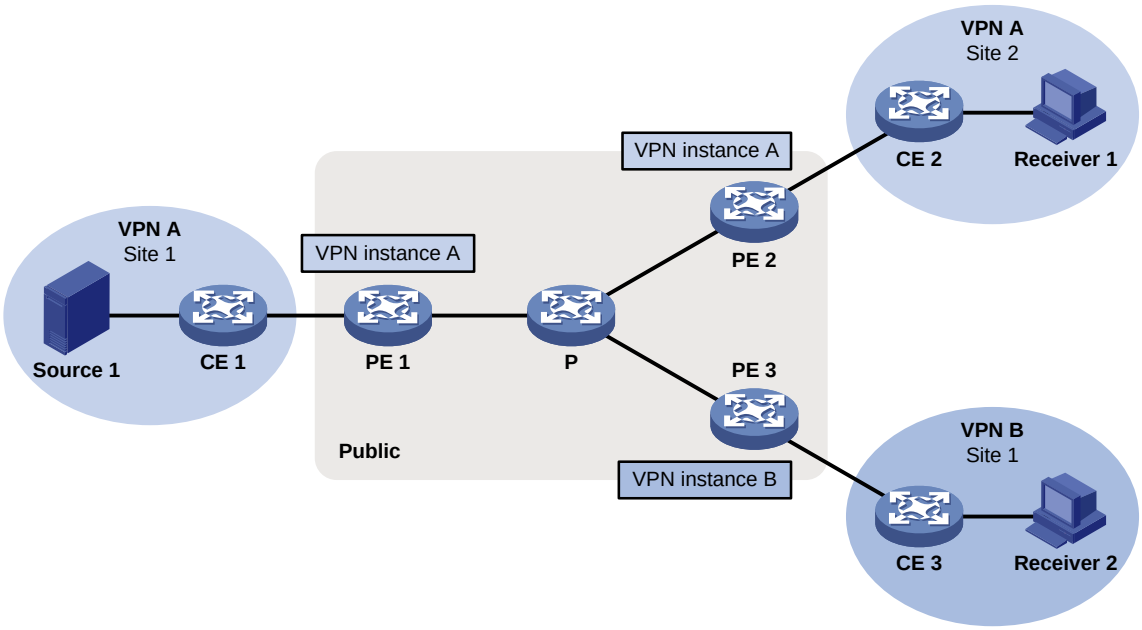
The following terms are used in the MVPN extranet solution:

- **Source VPN instance**—VPN instance to which the multicast source belongs.
- **Receiver VPN instance**—VPN instance to which the multicast receiver belongs.
- **Source PE**—PE directly connected to the multicast source.
- **Receiver PE**—PE directly connected to multicast receivers.

As shown in Figure 11, multicast source Source 1 is in Site 1 of VPN A. Receiver 1 and Receiver 2 are located in Site 2 of VPN A and Site 1 of VPN B, respectively. MVPN enables Receiver 1 to receive multicast data from Source 1, and MVPN extranet enables Receiver 2 to receive multicast data from Source 1.

The MVPN extranet solution can be implemented through the source-PE-based MVPN extranet option or the receiver-PE-based MVPN extranet option.

Figure 11 MVPN extranet



Source-PE-based MVPN extranet option

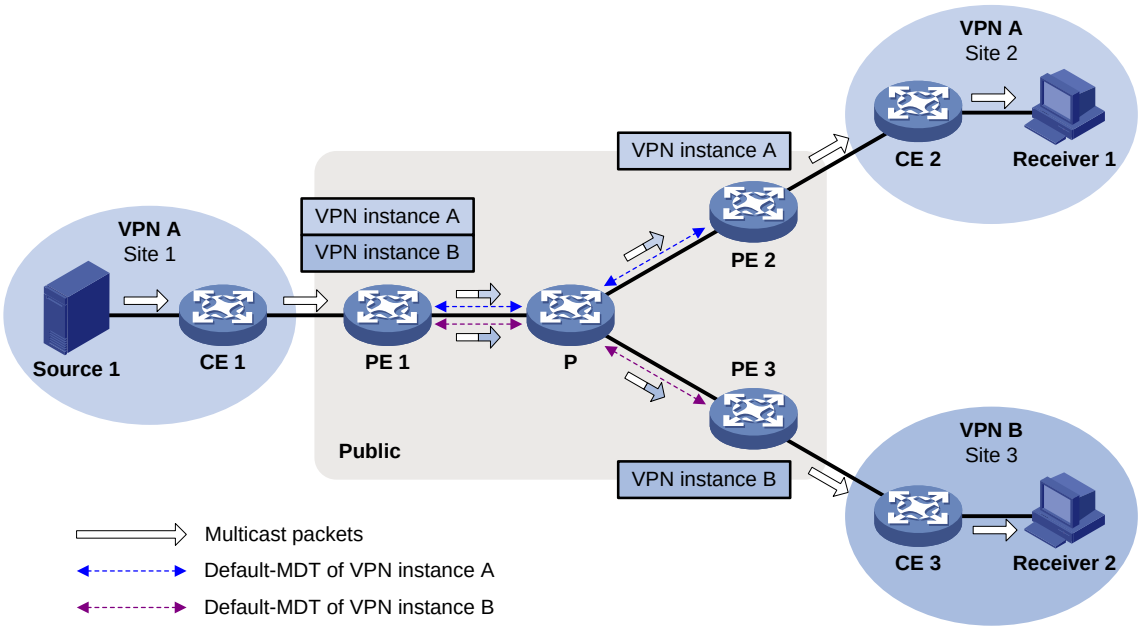
To use this option, create a receiver VPN instance on the source PE and configure MVPN for this VPN instance.

This option is available only in MDT-based MVPN.

As shown in Figure 12, perform the following tasks:

1. For multicast traffic transmission from Site 1 to Site 2, configure MVPN for VPN A on PE 1 and PE 2.
2. For multicast traffic transmission from Site 1 to Site 3, create VPN B and configure MVPN for VPN B on PE 1.
3. Configure an MVPN extranet RPF selection policy for VPN B on PE 1.

Figure 12 Source-PE-based MVPN extranet option



When Receiver 2 in VPN B joins the multicast group that matches the policy, multicast packets from Source 1 are transmitted as follows:

4. PE 1 replicates and encapsulates the packets for VPN A and VPN B and forwards them.
5. The packets travel along default MDTs of VPN A and VPN B and arrive at PE 2 and PE 3.
6. PE 2 and PE 3 decapsulate and forward the packets to receivers.

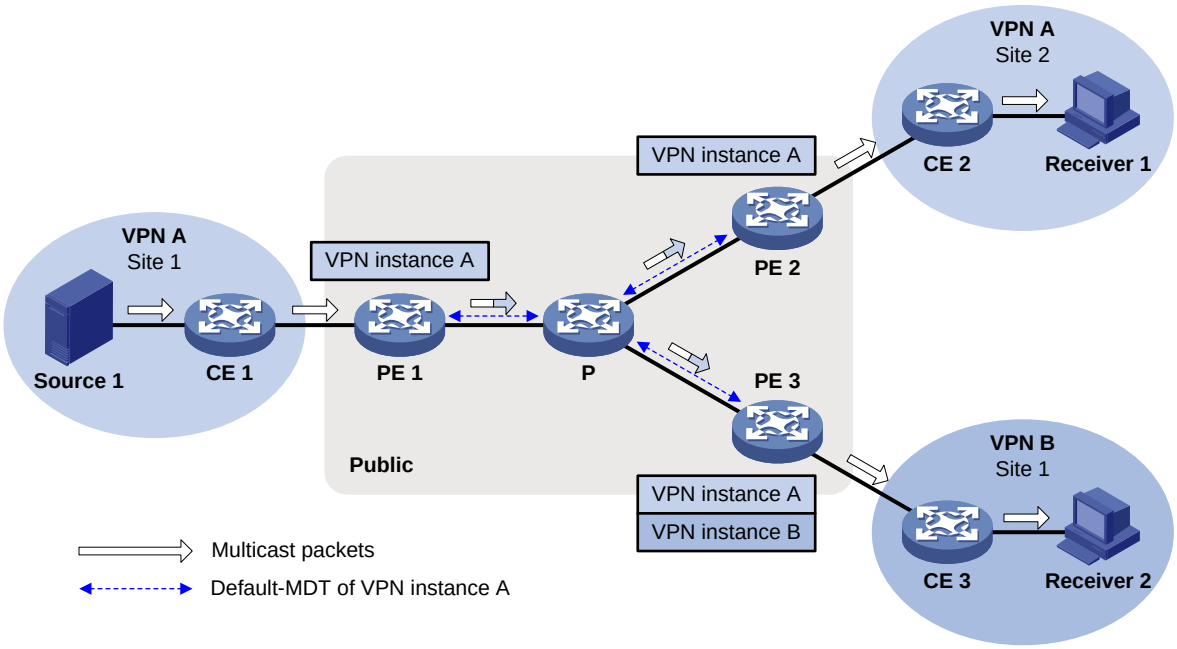
Receiver-PE-based MVPN extranet option

To use this option, create a source VPN instance on the receiver PE and configure MVPN. This option is available in MDT-based MVPN and RSVP-TE-based MVPN. The following steps describe how to configure this option for MDT-based MVPN.

As shown in Figure 13, configure PEs as follows:

1. For multicast traffic transmission from Site 1 to Site 2, configure MVPN for VPN A on PE 1 and PE 2.
2. For multicast traffic transmission from Site 1 to Site 3, create VPN A and configure MVPN for VPN A on PE 3.
3. Configure an MVPN extranet RPF selection policy for VPN B on PE 3.

Figure 13 Receiver-PE-based MVPN extranet option



When Receiver 2 in VPN B joins the multicast group that matches the RPF selection policy, multicast packets from Source 1 are encapsulated on PE 1 and transmitted along the default MDT.

- When the packets arrive at PE 2, PE 2 decapsulates the packets and forwards them to Receiver 1.
- When the packets arrive at PE 3, PE 3 decapsulates the packets and forwards them to Receiver 2 based on the MVPN extranet RPF selection policy.

Protocols and standards

- RFC 6037, *Cisco Systems' Solution for Multicast in BGP/MPLS IP VPNs*
- RFC 6513, *Multicast in MPLS/BGP IP VPNs*
- RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs*

Configuring multicast VPN

Restrictions and guidelines: Multicast VPN configuration

PIM FRR and IPv6 PIM FRR periodically monitor multicast traffic on the primary and backup incoming interfaces of each multicast forwarding entry to detect link failures. Use this feature only in a scenario with continued high-speed forwarding of multicast traffic. To avoid multicast traffic forwarding failures when multicast VPN is used, do not enable PIM FRR and IPv6 PIM FRR. For more information about PIM FRR and IPv6 PIM FRR, see "Configuring PIM" and "Configuring IPv6 PIM", respectively.

MDT-based MVPN tasks at a glance

To configure MDT-based MVPN, perform the following tasks on PEs:

1. Configuring MDT-based MVPN
   a. Enabling IP multicast routing for a VPN instance
   b. Creating an MDT-based MVPN instance
   c. Creating an MVPN address family
   d. Specifying the default group
   e. Specifying the MVPN source interface
   f. (Optional.) Configuring MDT switchover parameters
   g. (Optional.) Configuring the RPF vector feature
   h. (Optional.) Enabling data group reuse logging
   i. (Optional.) Setting the DSCP value for outgoing data group switchover packets
2. Configuring BGP MDT

   If PIM-SSM is running on the public network, you must configure BGP MDT.
   a. Configuring BGP MDT peers or peer groups
   b. (Optional.) Configuring a BGP MDT route reflector
   c. (Optional.) Configuring BGP MDT optimal route selection delay

Configuring MDT-based MVPN

Prerequisites for configuring MDT-based MVPN

Before you configure MDT-based MVPN, complete the following tasks:

- Configure a unicast routing protocol on the public network.
- Configure MPLS L3VPN on the public network.
- Configure PIM-DM, PIM-SM, BIDIR-PIM, or PIM-SSM on the public network.

Enabling IP multicast routing for a VPN instance

1. Enter system view.

system-view

2. Create a VPN instance and enter its view.
   ip vpn-instance *vpn-instance-name*
   For more information about this command, see *MPLS Command Reference*.
3. Configure an RD for the VPN instance.
   route-distinguisher *route-distinguisher*
   For more information about this command, see *MPLS Command Reference*.
4. Return to system view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Associate the interface with the VPN instance.
   ip binding vpn-instance *vpn-instance-name*
   By default, an interface is associated with no VPN instance and belongs to the public network.
   For more information about this command, see *MPLS Command Reference*.
7. Return to system view.
   quit
8. Enable IP multicast routing for the VPN instance and enter MRIB view of the VPN instance.
   IPv4:
   multicast routing vpn-instance *vpn-instance-name*
   By default, IPv4 multicast routing is disabled for a VPN instance.
   For more information about this command, see *IP Multicast Command Reference*.
   IPv6:
   ipv6 multicast routing vpn-instance *vpn-instance-name*
   By default, IPv6 multicast routing is disabled for a VPN instance.
   For more information about this command, see *IP Multicast Command Reference*.

Creating an MDT-based MVPN instance

About this task

To provide multicast services for a VPN instance, you must create an MDT-based MVPN instance on PEs that belong to the VPN instance. After the MVPN instance is created, the system automatically creates MTIs and binds them with the VPN instance.

You can create one or more MDT-based MVPN instances on a PE.

Procedure

1. Enter system view.
   system-view
2. Create an MDT-based MVPN instance and enter MVPN view.
   multicast-vpn vpn-instance *vpn-instance-name* **mode mdt**

Creating an MVPN address family

About this task

You must create an MVPN IPv4 or IPv6 address family for a VPN instance before you can perform other MVPN VPN configuration tasks for the VPN instance. For a VPN instance, configurations in MVPN IPv4 address family view apply to its IPv4 multicast packets, and configurations in IPv6 address family view apply to its IPv6 multicast packets.

Procedure

1. Enter system view.
   system-view
2. Enter MVPN view of a VPN instance.
   multicast-vpn vpn-instance *vpn-instance-name* **mode mdt**
3. Create an MVPN address family and enter MVPN address family view.
   IPv4:
   address-family ipv4
   IPv6:
   address-family ipv6

Specifying the default group

Restrictions and guidelines

You must specify the same default group on all PEs that belong to the same MVPN.

The default group for an MVPN must be different from the default group and the data group used by any other MVPN.

For an MVPN that transmits both IPv4 and IPv6 multicast packets, you must specify the same default group in MVPN IPv4 address family view and IPv6 address family view.

Procedure

1. Enter system view.
   system-view
2. Enter MVPN view.
   multicast-vpn vpn-instance *vpn-instance-name* **mode mdt**
3. Enter MVPN address family view.
   IPv4:
   address-family ipv4
   IPv6:
   address-family ipv6
4. Specify the default group.
   default-group *group-address*

Specifying the MVPN source interface

About this task

An MTI of a VPN instance uses the IP address of the MVPN source interface as the source address to encapsulate multicast packets for the VPN instance.

Restrictions and guidelines

For the PE to obtain correct routing information, you must specify the interface used for establishing BGP peer relationship as the MVPN source interface.

For an MVPN that transmits both IPv4 and IPv6 multicast packets, you must specify the same MVPN source interface in MVPN IPv4 and IPv6 address family views.

The MTI takes effect only after the default group and MVPN source interface are specified and the MTI obtains the public IP address of the MVPN source interface.

Procedure

1. Enter system view.
   system-view
2. Enter MVPN view.
   multicast-vpn vpn-instance *vpn-instance-name* **mode mdt**
3. Enter MVPN address family view.
   IPv4:
   address-family ipv4
   IPv6:
   address-family ipv6
4. Specify the MVPN source interface.
   source *interface-type interface-number*
   By default, no MVPN source interface is specified.

Configuring MDT switchover parameters

About this task

The traffic rate of the private network multicast data might fluctuate around the MDT switchover threshold. To avoid frequent switching of multicast traffic between the default MDT and the data MDT, set the data-delay period and the data holddown period.

- MDT switchover does not take place immediately after the multicast traffic rate exceeds the switchover threshold. It takes place after a data-delay period, during which the traffic rate must stay higher than the switchover threshold.
- Likewise, a backward switchover does not take place immediately after the multicast traffic rate drops below the MDT switchover threshold. It takes place after a data holddown period, during which the traffic rate must stay lower than the switchover threshold.

Restrictions and guidelines

On a PE, the data group range for an MVPN cannot include the default group or data groups of any other MVPN.

All VPN instances share the data group resources. As a best practice to avoid data group resource exhaustion, specify a reasonable data group range for a VPN instance.

For an MVPN that transmits both IPv4 and IPv6 multicast packets, the data group range in MVPN IPv4 and IPv6 address family views cannot overlap.

If the public network runs PIM-SSM, the data group range for an MVPN on a PE can overlap with data group ranges for other MDs on other PEs.

If BIDIR-PIM runs in a VPN instance, the switchover from the default MDT to a data MDT is not supported.

Procedure

1. Enter system view.
   system-view
2. Enter MVPN view.
   multicast-vpn vpn-instance *vpn-instance-name* **mode mdt**
3. Enter MVPN address family view.
   IPv4:
   address-family ipv4
   IPv6:
   address-family ipv6
4. Configure the data group range and the switchover criteria.
   data-group *group-address* { *mask-length* | *mask* } [**acl** *acl-number*]
   By default, no data group range exists, and the default MDT to data MDT switchover never occurs.
5. Set the data-delay period.
   data-delay *delay*
   By default, the data-delay period is 3 seconds.

Configuring the RPF vector feature

About this task

This feature enables the device to insert the RPF vector (IP address of the ASBR in the local AS) in PIM join messages for other devices to perform RPF check.

Restrictions and guidelines

Perform this task on PEs that have attached receivers.

For the device to work with other manufacturers' products on the RPF vector, you must enable RPF vector compatibility for all H3C P devices and H3C PE devices on the public network.

Procedure

1. Enter system view.
   system-view
2. Enter MRIB view of a VPN instance.
   multicast routing vpn-instance *vpn-instance-name*
3. Enable the RPF vector feature.
   rpf proxy vector
   By default, the RPF vector feature is disabled.
4. Enable RPF vector compatibility.
   multicast rpf-proxy-vector compatible
   By default, RPF vector compatibility is disabled.

Enabling data group reuse logging

About this task

For a given VPN, the number of VPN multicast streams to be switched to data MDTs might exceed the number of addresses in the data group range. In this case, the VPN instance on the source-side

PE can reuse the addresses in the address range. With data group reuse logging enabled, the address reuse information will be logged. Perform this task on PEs.

The group address reuse logging information has a severity level **informational**. For more information about the logging information, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter MVPN view.
   multicast-vpn vpn-instance *vpn-instance-name* **mode mdt**
3. Enter MVPN address family view.
   IPv4:
   address-family ipv4
   IPv6:
   address-family ipv6
4. Enable data group reuse logging.
   log data-group-reuse
   By default, data group reuse logging is disabled.

Setting the DSCP value for outgoing data group switchover packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.
   system-view
2. Enter MVPN view.
   multicast-vpn vpn-instance *vpn-instance-name* **mode mdt**
3. Setting the DSCP value for outgoing data group switchover packets.
   dscp *dscp-value*
   By default, the DSCP value is 48 for outgoing data group switchover packets.

Configuring BGP MDT

Configuring BGP MDT peers or peer groups

About this task

Perform this task so that the PE can exchange MDT information with the BGP peer or peer group. MDT information includes the IP address of the PE and default group to which the PE belongs. On a public network running PIM-SSM, the multicast VPN establishes a default MDT rooted at the PE (multicast source) based on the MDT information.

Prerequisites

Before you configure a BGP MDT peer or peer group, you must create a BGP peer or peer group in BGP instance view. For more information about creating a BGP peer or peer group, see BGP configuration in *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Create a BGP IPv4 MDT address family and enter its view.

   address-family ipv4 mdt

   By default, no BGP IPv4 address family exists.
4. Enable the device to exchange MDT routing information with the BGP peer or the peer group.

   peer { *group-name* | *ip-address* [*mask-length*] } **enable**

   By default, the device cannot exchange BGP MDT routing information with a BGP peer or peer group.

   For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Configuring a BGP MDT route reflector

About this task

- **Configuring a BGP MDT route reflector**—BGP MDT peers in the same AS must be fully meshed to maintain connectivity. However, when multiple BGP MDT peers exist in an AS, connection establishment among them might result in increased costs. To reduce connections between BGP MDT peers, you can configure one of them as a route reflector and specify other devices as clients.
- **Disabling routing reflection between clients**—When clients establish BGP MDT connections with the route reflector, the route reflector forwards (or reflects) BGP MDT routing information between clients. The clients are not required to be fully meshed. To save bandwidth if the clients have been fully meshed, you can disable the routing reflection between clients by using the **undo reflect between-clients** command.
- **Configuring the cluster ID of the route reflector**—The route reflector and its clients form a cluster. Typically, a cluster has only one route reflector whose router ID identifies the cluster. However, you can configure several route reflectors in a cluster to improve network reliability. To avoid routing loops, make sure the route reflectors in a cluster have the same cluster ID.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP IPv4 MDT address family view.

   address-family ipv4 mdt
4. Configure the device as a route reflector and specify its peers or peer groups as clients.

   peer { *group-name* | *ip-address* [*mask-length*] } **reflect-client**

   By default, neither route reflectors nor clients exist.
5. (Optional.) Disable route reflection between clients.

undo reflect between-clients

By default, route reflection between clients is disabled.

For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.

6. (Optional.) Configure the cluster ID of the route reflector.

   reflector cluster-id { *cluster-id* | *ip-address* }

   By default, a route reflector uses its router ID as the cluster ID.

   For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Configuring BGP MDT optimal route selection delay

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP IPv4 MDT address family view.

   address-family ipv4 mdt

4. Set the optimal route selection delay timer. Choose the options to configure as needed:
   - Set the optimal route selection delay timer for all BGP routes.

     route-select delay *delay-value*
   - Set the optimal route selection delay timer for BGP routes from only neighbors that change from down state to up state.

     route-select suppress on-peer-up *milliseconds*

   By default, optimal route selection is not delayed.

   For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Configuring an MVPN extranet RPF selection policy

About this task

MVPN extranet RPF routing policies are used for multicast transmission when multicast sources and receivers are located in different VPNs.

Restrictions and guidelines for configuring MVPN extranet RPF selection policies

The PIM modes in the source VPN instance and the receiver VPN instance must be the same. Only PIM-SM and PIM-SSM are supported.

Multicast packets can only be forwarded between two VPNs. The receiver VPN instance cannot also be the source VPN instance.

In PIM-SM mode, you can configure only one RPF selection policy for a multicast group in a VPN instance.

If an IPv4 MVPN extranet RPF selection policy with only the multicast group address specified is configured in the receiver VPN instance, the multicast traffic for the intra-VPN transmission will be interrupted.

To implement source-specific RPF selection in MVPN extranet, you must configure two MVPN extranet RPF routing policies as follows:

- In one policy, specify the address of the RP designated to the multicast group that requires inter-VPN multicast communication as the source address.
- In the other policy, specify the multicast source in the source VPN instance as the source address.

To implement source-and-group-specific RPF selection in MVPN extranet, you must configure two MVPN extranet RPF routing policies as follows:

- In one policy, specify the address of the RP designated to the multicast group as the source address, and specify the multicast group.
- In the other policy, specify the multicast source in the source VPN instance as the source address, and specify the multicast group.
- Make sure the multicast groups in the two policies are the same to avoid inter-VPN multicast transmission failure.

Prerequisites for configuring MVPN extranet RPF selection policies

Before you configure MVPN extranet RPF selection policies, you must configure MVPN.

For the source-PE-based MVPN extranet option, see "Configuring MDT-based MVPN."

For the receiver-PE-based MVPN extranet option, see "Configuring MDT-based MVPN."

Configuring an IPv4 MVPN extranet RPF selection policy

1. Enter system view.

   system-view

2. Enter MRIB view.

   multicast routing [**vpn-instance** *vpn-instance-name*]

3. Configure an IPv4 MVPN extranet RPF selection policy.

   multicast extranet select-rpf [**vpn-instance** *vpn-instance-name*] { **source** *source-address* { *mask* | *mask-length* } | **group** *group-address* { *mask* | *mask-length* } } *

   By default, no IPv4 MVPN extranet RPF selection policies are configured.

Configuring an IPv6 MVPN extranet RPF selection policy

1. Enter system view.

   system-view

2. Enter IPv6 MRIB view.

   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]

3. Configure an IPv6 MVPN extranet RPF selection policy.

   ipv6 multicast extranet select-rpf { **vpn-instance** *vpn-instance-name* } { **source** *ipv6-source-address prefix-length* | **group** *ipv6-group-address prefix-length* }*

   By default, no IPv6 MVPN extranet RPF selection policies are configured.

Display and maintenance commands for multicast VPN

Execute **display** commands in any view and **reset** commands in user view.

Display and maintenance commands for MDT-based MVPN:

Task	Command
Display BGP MDT peer group information.	**display bgp** [**instance** *instance-name*] **group ipv4 mdt** [**group-name** *group-name*]
Display information about BGP MDT peers or peer groups.	**display bgp** [**instance** *instance-name*] **peer ipv4 mdt** [*ip-address mask-length* | { *ip-address* | **group-name** *group-name* } **log-info** | [*ip-address*] **verbose**]
Display information about BGP MVPN peers or peer groups.	**display bgp** [**instance** *instance-name*] **peer ipv4 mvpn** [*ip-address mask-length* | { *ip-address* | **group-name** *group-name* } **log-info** | [*ip-address*] **verbose**]
Display BGP MDT routing information.	**display bgp** [**instance** *instance-name*] **routing-table ipv4 mdt** [**route-distinguisher** *route-distinguisher*] [*ip-address* [**advertise-info** | **as-path** | **cluster-list**]]
Display information about BGP update groups for the BGP IPv4 MDT address family.	**display bgp** [**instance** *instance-name*] **update-group ipv4 mdt** [*ip-address*]
Display information about data groups for IPv4 multicast transmission that are received in a VPN instance.	**display multicast-vpn vpn-instance** *vpn-instance-name* **data-group receive** [**brief** | [**active** | **group** *group-address* | **sender** *source-address* | *vpn-source-address* [**mask** { *mask-length* | *mask* }] | *vpn-group-address* [**mask** { *mask-length* | *mask* }]] *]
Display information about data groups for IPv6 multicast transmission that are received in a VPN instance.	**display multicast-vpn vpn-instance** *vpn-instance-name* **ipv6 data-group receive** [**brief** | [**active** | **group** *group-address* | **sender** *source-address* | *vpn-source-address* [*mask-length*] | *vpn-group-address* [*mask-length*]] *]
Display information about data groups for IPv4 multicast transmission that are sent in a VPN instance.	**display multicast-vpn vpn-instance** *vpn-instance-name* **data-group send** [**group** *group-address* | **reuse** *interval* | *vpn-source-address* [**mask** { *mask-length* | *mask* }] | *vpn-group-address* [**mask** { *mask-length* | *mask* }]] *
Display information about data groups for IPv6 multicast transmission that are sent in a VPN instance.	**display multicast-vpn vpn-instance** *vpn-instance-name* **ipv6 data-group send** [**group** *group-address* | **reuse** *interval* | *vpn-source-address* [*mask-length*] | *vpn-group-address* [*mask-length*]] *
Display information about default groups for IPv4 multicast transmission.	**display multicast-vpn** [**vpn-instance** *vpn-instance-name*] **default-group** { **local** | **remote** }
Display information about default groups for IPv6 multicast transmission.	**display multicast-vpn** [**vpn-instance** *vpn-instance-name*] **ipv6 default-group** { **local** | **remote** }
Reset BGP sessions for BGP	**reset bgp** [**instance** *instance-name*] { *as-number*

Task	Command
IPv4 MDT address family.	| *ip-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **ipv4 mdt**

For more information about the **display bgp group**, **display bgp peer**, **display bgp update-group**, and **reset bgp** commands, see BGP commands in *Layer 3—IP Routing Command Reference*.

Multicast VPN configuration examples

Example: Configuring intra-AS MDT-based MVPN

Network configuration

As shown in Figure 14, configure intra-AS MDT-based MVPN to meet the following requirements:

Item	Network configuration
Multicast sources and receivers	• In VPN instance **a**, S 1 is a multicast source, and R 1 is a receiver. • For VPN instance **a**, the default group is 239.1.1.1, and the data group range is 225.2.2.0 to 225.2.2.15.
VPN instances to which PE interfaces belong	• PE 1: VLAN-interface 20 belongs to VPN instance **a**. VLAN-interface 10 and Loopback 1 belong to the public network instance. • PE 2: VLAN-interface 60 belongs to VPN instance **a**. VLAN-interface 40 and Loopback 1 belong to the public network instance.
Unicast routing protocols and MPLS	• Configure OSPF on the public network, and configure RIP between the PEs and CEs. • Establish BGP peer connections between PE 1 and PE 2 on their respective Loopback 1. • Configure MPLS on the public network.
IP multicast routing	• Enable IP multicast routing on the P device. • Enable IP multicast routing on the public network on PE 1 and PE 2. • Enable IP multicast routing for VPN instance **a** on PE 1 and PE 2. • Enable IP multicast routing on CE 1 and CE 2.
IGMP	Enable IGMPv2 on VLAN-interface 60 of CE 2.
PIM	Enable PIM-SM on the public network and for VPN instances **a** and **b**: • Enable PIM-SM on all interfaces of the P device. • Enable PIM-SM on all public and private network interfaces of PE 1 and PE 2. • Enable PIM-SM on all interfaces that have no receiver hosts connected of CE 1 and CE 2. • Configure Loopback 1 of P as a C-BSR and a C-RP for the public network to provide services for all multicast groups. • Configure Loopback 2 of PE 1 as a C-BSR and a C-RP for VPN instance **a** to provide services for all multicast groups.

Figure 14 Network diagram

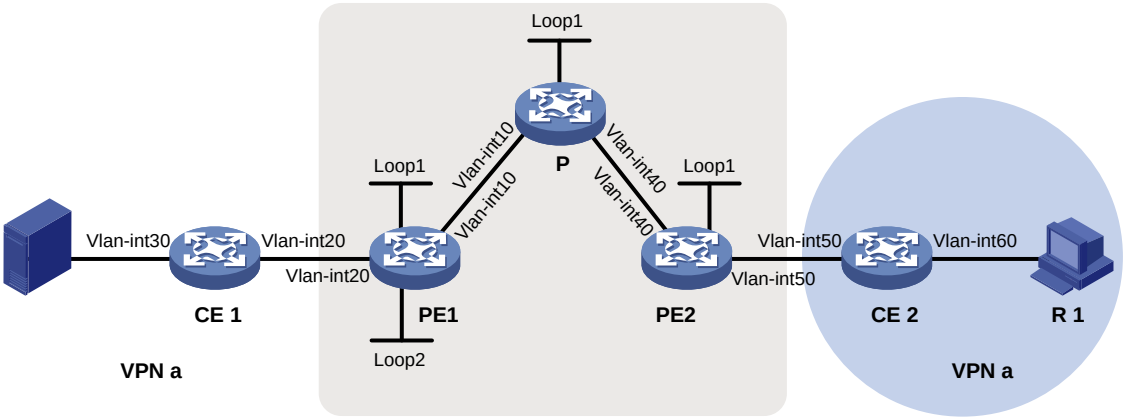


Table 1 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
S 1	—	10.110.7.2/24	PE 3	—	10.110.9.2/24
P	Vlan-int10	192.168.6.2/24	CE a1	Vlan-int30	10.110.7.1/24
P	Vlan-int40	192.168.7.2/24	CE a2	Vlan-int20	10.110.2.2/24
P	Loop1	2.2.2.2/32	CE a2	Vlan-int60	10.110.9.1/24
PE 1	Vlan-int10	192.168.6.1/24	CE a2	Vlan-int50	10.110.4.2/24
PE 1	Vlan-int20	10.110.2.1/24	CE a3	Vlan-int40	192.168.7.1/24
PE 1	Loop1	1.1.1.1/32	CE a3	Vlan-int50	10.110.4.1/24
PE 1	Loop2	33.33.33.33/32	CE a3	Loop1	1.1.1.2/32

Procedure

1. Configure PE 1:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE1
[PE1] router id 1.1.1.1
[PE1] multicast routing
[PE1-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE1] mpls lsr-id 1.1.1.1
[PE1] mpls ldp
[PE1-ldp] quit
```

Create a VPN instance named **a** and configure an RD and route targets for the VPN instance.

```
[PE1] ip vpn-instance a
[PE1-vpn-instance-a] route-distinguisher 100:1
[PE1-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE1-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE1-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE1] multicast routing vpn-instance a
[PE1-mrib-a] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.1 32
[PE1-LoopBack1] pim sm
[PE1-LoopBack1] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE1] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 family address for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a-ipv4] default-group 239.1.1.1
[PE1-mvpn-vpn-instance-a-ipv4] source loopback 1
[PE1-mvpn-vpn-instance-a-ipv4] data-group 225.2.2.0 28
[PE1-mvpn-vpn-instance-a-ipv4] quit
[PE1-mvpn-vpn-instance-a] quit
```

Assign an IP address to VLAN-interface 10.

```
[PE1] vlan 10
[PE1-vlan10] port gigabitethernet 1/0/1
[PE1-vlan10] quit
[PE1] interface vlan-interface 10
[PE1-Vlan-interface10] ip address 192.168.6.1 24
```

#Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 10.

```
[PE1-Vlan-interface10] pim sm
[PE1-Vlan-interface10] mpls enable
[PE1-Vlan-interface10] mpls ldp enable
[PE1-Vlan-interface10] quit
```

Associate VLAN-interface 20 with VPN instance **a**.

```
[PE1] vlan 20
[PE1-vlan20] port gigabitethernet 1/0/2
[PE1-vlan20] quit
[PE1] interface vlan-interface 20
[PE1-Vlan-interface20] ip binding vpn-instance a
```

Assign an IP address to VLAN-interface 20, and enable PIM-SM on the interface.

```
[PE1-Vlan-interface20] ip address 10.110.2.1 24
[PE1-Vlan-interface20] pim sm
[PE1-Vlan-interface20] quit
```

Associate Loopback 2 with VPN instance **a**.

```
[PE1] interface loopback 2
[PE1-LoopBack2] ip binding vpn-instance a
```

Assign an IP address to Loopback 2, and enable PIM-SM on the interface.

```
[PE1-LoopBack2] ip address 33.33.33.33 32
[PE1-LoopBack2] pim sm
[PE1-LoopBack2] quit
```

Configure Loopback 2 as a C-BSR and a C-RP.

```
[PE1] pim vpn-instance a
[PE1-pim-a] c-bsr 33.33.33.33
[PE1-pim-a] c-rp 33.33.33.33
```

```
[PE1-pim-a] quit
```

Configure BGP.

```
[PE1] bgp 100
[PE1-bgp-default] group vpn-g internal
[PE1-bgp-default] peer vpn-g connect-interface loopback 1
[PE1-bgp-default] peer 1.1.1.2 group vpn-g
[PE1–bgp-default] ip vpn-instance a
[PE1-bgp-default-a] address-family ipv4
[PE1-bgp-default-ipv4-a] import-route rip 2
[PE1-bgp-default-ipv4-a] import-route direct
[PE1-bgp-default-ipv4-a] quit
[PE1-bgp-default-a] quit
[PE1–bgp-default] address-family vpnv4
[PE1–bgp-default-vpnv4] peer vpn-g enable
[PE1–bgp-default-vpnv4] quit
[PE1-bgp-default] address-family ipv4 mdt
[PE1-bgp-default-mdt] peer vpn-g enable
[PE1-bgp-default-mdt] quit
[PE1–bgp-default] quit
```

Configure OSPF.

```
[PE1] ospf 1
[PE1-ospf-1] area 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 192.168.6.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

Configure RIP.

```
[PE1] rip 2 vpn-instance a
[PE1-rip-2] network 33.33.33.33 0.0.0.0
[PE1-rip-2] network 10.110.2.0 0.0.0.255
[PE1-rip-2] import-route bgp
[PE1-rip-2] quit
```

2. Configure PE 2:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] router id 1.1.1.2
[PE2] multicast routing
[PE2-mrib] quit
```

Configure an MPLS LSR ID, and enable LDP globally.

```
[PE2] mpls lsr-id 1.1.1.2
[PE2] mpls ldp
[PE2-ldp] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 1.1.1.2 32
[PE2-LoopBack1] pim sm
[PE2-LoopBack1] quit
```

Create a VPN instance named **a**, and configure an RD and route targets for VPN instance.

```
[PE2] ip vpn-instance a
[PE2-vpn-instance-a] route-distinguisher 100:1
[PE2-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE2-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE2-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE2] multicast routing vpn-instance a
[PE2-mrib-a] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE2] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 address family for VPN instance **a**.

```
[PE2-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE2-mvpn-vpn-instance-a-ipv4] default-group 239.1.1.1
[PE2-mvpn-vpn-instance-a-ipv4] source loopback 1
[PE2-mvpn-vpn-instance-a-ipv4] data-group 225.2.2.0 28
[PE2-mvpn-vpn-instance-a-ipv4] quit
[PE2-mvpn-vpn-instance-a] quit
```

Assign an IP address to VLAN-interface 40.

```
[PE2] vlan 40
[PE2-vlan40] port gigabitethernet 1/0/1
[PE2-vlan40] quit
[PE2] interface vlan-interface 40
[PE2-Vlan-interface40] ip address 192.168.7.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 40.

```
[PE2-Vlan-interface40] pim sm
[PE2-Vlan-interface40] mpls enable
[PE2-Vlan-interface40] mpls ldp enable
[PE2-Vlan-interface40] quit
```

Associate VLAN-interface 50 with VPN instance **a**, assign an IP address to VLAN-interface 50, and enable PIM-SM on the interface.

```
[PE2] vlan 50
[PE2-vlan50] port gigabitethernet 1/0/2
[PE2-vlan50] quit
[PE2] interface vlan-interface 50
[PE2-Vlan-interface50] ip binding vpn-instance a
[PE2-Vlan-interface50] ip address 10.110.4.1 24
[PE2-Vlan-interface50] pim sm
[PE2-Vlan-interface50] quit
```

Configure BGP.

```
[PE2] bgp 100
[PE2-bgp-default] group vpn-g internal
[PE2-bgp-default] peer vpn-g connect-interface loopback 1
[PE2-bgp-default] peer 1.1.1.1 group vpn-g
[PE2–bgp-default] ip vpn-instance a
[PE2-bgp-default-a] address-family ipv4
```

```
[PE2-bgp-default-ipv4-a] import-route rip 2
[PE2-bgp-default-ipv4-a] import-route direct
[PE2-bgp-default-ipv4-a] quit
[PE2-bgp-default-a] quit
[PE2–bgp-default] address-family vpnv4
[PE2–bgp-default-vpnv4] peer vpn-g enable
[PE2–bgp-default-vpnv4] quit
[PE2-bgp-default] address-family ipv4 mdt
[PE2-bgp-default-mdt] peer vpn-g enable
[PE2-bgp-default-mdt] quit
[PE2–bgp-default] quit
```

Configure OSPF.

```
[PE2] ospf 1
[PE2-ospf-1] area 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 1.1.1.2 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 192.168.7.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

Configure RIP.

```
[PE2] rip 2 vpn-instance a
[PE2-rip-2] network 10.110.4.0 0.0.0.255
[PE2-rip-2] import-route bgp
[PE2-rip-2] quit
```

3. Configure P:

Enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname P
[P] multicast routing
[P-mrib] quit
```

Configure an MPLS LSR ID, and enable LDP globally.

```
[P] mpls lsr-id 2.2.2.2
[P] mpls ldp
[P-ldp] quit
```

Assign an IP address to VLAN-interface 10.

```
[P] vlan 10
[P-vlan10] port gigabitethernet 1/0/1
[P-vlan10] quit
[P] interface vlan-interface 10
[P-Vlan-interface10] ip address 192.168.6.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 10.

```
[P-Vlan-interface10] pim sm
[P-Vlan-interface10] mpls enable
[P-Vlan-interface10] mpls ldp enable
[P-Vlan-interface10] quit
```

Assign an IP address to VLAN-interface 40.

```
[P] vlan 40
[P-vlan40] port gigabitethernet 1/0/2
[P-vlan40] quit
```

```
[P] interface vlan-interface 40
[P-Vlan-interface40] ip address 192.168.7.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 40.

```
[P-Vlan-interface40] pim sm
[P-Vlan-interface40] mpls enable
[P-Vlan-interface40] mpls ldp enable
[P-Vlan-interface40] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[P] interface loopback 1
[P-LoopBack1] ip address 2.2.2.2 32
[P-LoopBack1] pim sm
[P-LoopBack1] quit
```

Configure Loopback 1 as a C-BSR and a C-RP.

```
[P] pim
[P-pim] c-bsr 2.2.2.2
[P-pim] c-rp 2.2.2.2
[P-pim] quit
```

Configure OSPF.

```
[P] ospf 1
[P-ospf-1] area 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 192.168.6.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 192.168.7.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

4. Configure CE 1:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CE1
[CE1] multicast routing
[CE1-mrib] quit
```

Assign an IP address to VLAN-interface 30, and enable PIM-SM on the interface.

```
[CE1] vlan 30
[CE1-vlan30] port gigabitethernet 1/0/1
[CE1-vlan30] quit
[CE1] interface vlan-interface 30
[CE1-Vlan-interface30] ip address 10.110.7.1 24
[CE1-Vlan-interface30] pim sm
[CE1-Vlan-interface30] quit
```

Assign an IP address to VLAN-interface 20, and enable PIM-SM on the interface.

```
[CE1] vlan 20
[CE1-vlan20] port gigabitethernet 1/0/2
[CE1-vlan20] quit
[CE1] interface vlan-interface 20
[CE1-Vlan-interface20] ip address 10.110.2.2 24
[CE1-Vlan-interface20] pim sm
[CE1-Vlan-interface20] quit
```

Configure RIP.

```
[CE1] rip 2
[CE1-rip-2] network 10.110.2.0 0.0.0.255
[CE1-rip-2] network 10.110.7.0 0.0.0.255
[CE1-rip-2] quit
```

5. Configure CE 2:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CE2
[CE2] multicast routing
[CE2-mrib] quit
```

Assign an IP address to VLAN-interface 60, and enable IGMP on the interface.

```
[CE2] vlan 60
[CE2-vlan60] port gigabitethernet 1/0/1
[CE2-vlan60] quit
[CE2] interface vlan-interface 60
[CE2-Vlan-interface60] ip address 10.110.9.1 24
[CE2-Vlan-interface60] igmp enable
[CE2-Vlan-interface60] quit
```

Assign an IP address to VLAN-interface 50, and enable PIM-SM on the interface.

```
[CE2] vlan 50
[CE2-vlan50] port gigabitethernet 1/0/2
[CE2-vlan50] quit
[CE2] interface vlan-interface 50
[CE2-Vlan-interface50] ip address 10.110.4.2 24
[CE2-Vlan-interface50] pim sm
[CE2-Vlan-interface50] quit
```

Configure RIP.

```
[CE2] rip 2
[CE2-rip-2] network 10.110.4.0 0.0.0.255
[CE2-rip-2] quit
```

Verifying the configuration

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 1.

```
[PE1] display multicast-vpn default-group local
MVPN local default-group information:
 Group address    Source address   Interface     VPN instance
 239.1.1.1        1.1.1.1          MTunnel0      a
```

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 2.

```
[PE2] display multicast-vpn default-group local
MVPN local default-group information:
 Group address    Source address   Interface     VPN instance
 239.1.1.1        1.1.1.2          MTunnel0      a
```

Configuration files

- CE 1:

```
#
```

```
 sysname CE1
#
rip 2
 network 10.110.2.0 0.0.0.255
 network 10.110.7.0 0.0.0.255
#
vlan 20
#
vlan 30
#
interface Vlan-interface20
 ip address 10.110.2.2 255.255.255.0
 pim sm
#
interface Vlan-interface30
 ip address 10.110.7.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 30
#
interface GigabitEthernet1/0/2
 port access vlan 20
#
multicast routing
#
return
```

- CE 2:

```
#
 sysname CE2
#
rip 2
 network 10.110.4.0 0.0.0.255
#
vlan 50
#
vlan 60
#
interface Vlan-interface50
 ip address 10.110.4.2 255.255.255.0
 pim sm
#
interface Vlan-interface60
 ip address 10.110.9.1 255.255.255.0
 igmp enable
#
interface GigabitEthernet1/0/1
 port access vlan 60
```

```
#
interface GigabitEthernet1/0/2
 port access vlan 50
#
multicast routing
#
return
```

- P:

```
#
 sysname P
#
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 192.168.6.0 0.0.0.255
  network 192.168.7.0 0.0.0.255
#
 mpls lsr-id 2.2.2.2
#
vlan 10
#
vlan 40
#
mpls ldp
#
interface LoopBack1
 ip address 2.2.2.2 255.255.255.255
 pim sm
#
interface Vlan-interface10
 ip address 192.168.6.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface40
 ip address 192.168.7.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port access vlan 10
#
interface GigabitEthernet1/0/2
 port access vlan 40
#
multicast routing
```

```
#
pim
 c-bsr 2.2.2.2
 c-rp 2.2.2.2
#
return
```

- PE 1:

```
#
 sysname PE1
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
 router id 1.1.1.1
#
ospf 1
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 192.168.6.0 0.0.0.255
#
rip 2 vpn-instance a
 network 10.110.2.0 0.0.0.255
 network 33.33.33.33 0.0.0.0
 import-route bgp
#
 mpls lsr-id 1.1.1.1
#
vlan 10
#
vlan 20
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.1 255.255.255.255
 pim sm
#
interface LoopBack2
 ip binding vpn-instance a
 ip address 33.33.33.33 255.255.255.255
 pim sm
#
interface Vlan-interface10
 ip address 192.168.6.1 255.255.255.0
 pim sm
 mpls enable
```

```
 mpls ldp enable
#
interface Vlan-interface20
 ip binding vpn-instance a
 ip address 10.110.2.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 10
#
interface GigabitEthernet1/0/2
 port access vlan 20
#
bgp 100
 group vpn-g internal
 peer vpn-g connect-interface LoopBack1
 peer 1.1.1.2 group vpn-g
 #
 address-family ipv4 unicast
  import-route direct
  import-route rip 2
 #
 address-family vpnv4
  peer vpn-g enable
 #
 address-family ipv4 mdt
  peer vpn-g enable
 #
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
   import-route rip 2
#
multicast routing
#
multicast routing vpn-instance a
#
pim vpn-instance a
 c-bsr 33.33.33.33
 c-rp 33.33.33.33
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack1
  default-group 239.1.1.1
  data-group 225.2.2.0 255.255.255.240
#
```

```
return
```

- PE 2:

```
#
 sysname PE2
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
 router id 1.1.1.2
#
ospf 1
 area 0.0.0.0
  network 1.1.1.2 0.0.0.0
  network 192.168.7.0 0.0.0.255
#
rip 2 vpn-instance a
 network 10.110.4.0 0.0.0.255
 import-route bgp
#
 mpls lsr-id 1.1.1.2
#
vlan 40
#
vlan 50
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.2 255.255.255.255
 pim sm
#
interface Vlan-interface40
 ip address 192.168.7.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface50
 ip binding vpn-instance a
 ip address 10.110.4.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 40
#
interface GigabitEthernet1/0/2
```

```
 port access vlan 50
#
bgp 100
 group vpn-g internal
 peer vpn-g connect-interface LoopBack1
 peer 1.1.1.1 group vpn-g
 #
 address-family vpnv4
  peer vpn-g enable
 #
 address-family ipv4 mdt
  peer vpn-g enable
 #
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
   import-route rip 2
#
multicast routing
#
multicast routing vpn-instance a
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack1
  default-group 239.1.1.1
  data-group 225.2.2.0 255.255.255.240
#
return
```

Example: Configuring intra-AS M6VPE

Network configuration

As shown in Figure 15, configure intra-AS M6VPE to meet the following requirements:

Item	Network configuration
Multicast sources and receivers	• In VPN instance **a**, S 1 is an IPv6 multicast source, and R 1 is a receiver. • For VPN instance **a**, the default group is 239.1.1.1, and the data group range is 225.2.2.0 to 225.2.2.15.
VPN instances to which PE interfaces belong	• PE 1: VLAN-interface 20 belongs to VPN instance **a**. VLAN-interface 10 and Loopback 1 belong to the public network. • PE 2: VLAN-interface 50 belongs to VPN instance **a**. VLAN-interface 40 and Loopback 1 belong to the public network.
Unicast routing protocols and MPLS	• Configure OSPF on the public network, and configure OSPFv3 between the PEs and the CEs. • Establish BGP peer connections between PE 1 and PE 2 on their respective Loopback 1.

Item	Network configuration
	• Configure MPLS on the public network.
IP multicast routing and IPv6 multicast routing	• Enable IP multicast routing on P. • Enable IP multicast routing for the public network on PE 1 and PE 2. • Enable IPv6 multicast routing for VPN instance **a** on PE 1 and PE 2. • Enable IPv6 multicast routing on CE 1 and CE 2.
MLDv1	Enable MLDv1 on VLAN-interface 60 of CE 2.
PIM and IPv6 PIM	Enable PIM-SM on the public network and IPv6 PIM-SM for VPN instances **a** and **b**: • Enable PIM-SM on all interfaces of P. • On PE 1 and PE 2, enable PIM-SM on all public network interfaces and IPv6 PIM-SM on private network interfaces. • Enable IPv6 PIM-SM on all interfaces that do not have attached receivers on CE 1 and CE 2. • Configure Loopback 1 of P as a C-BSR and a C-RP for the public network to provide services for all IPv4 multicast groups. • Configure Loopback 2 of PE 1 as a C-BSR and a C-RP for VPN instance **a** to provide services for all IPv6 multicast groups.

Figure 15 Network diagram

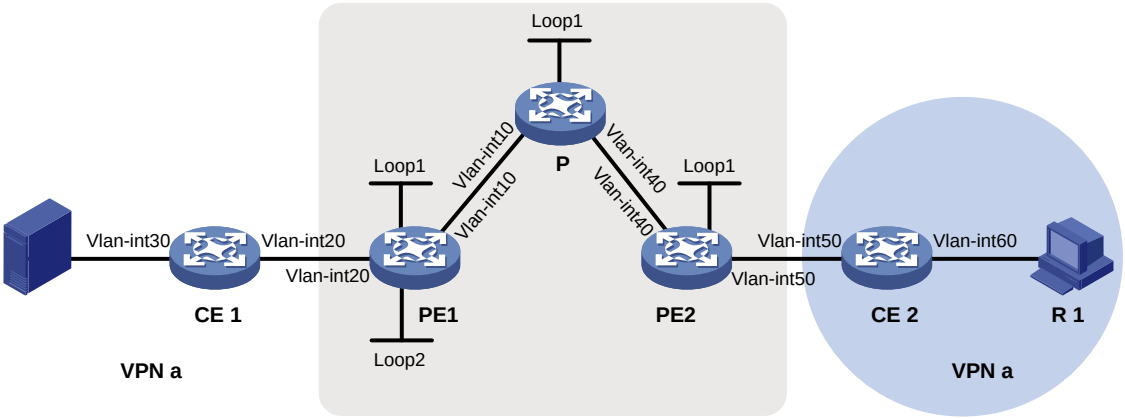


Table 2 Interface and IP address assignment

Device	Interface	IPv4/IPv6 address	Device	Interface	IPv4/IPv6 address
S 1	—	10:110:7::2/64	R 1	—	10:110:9::2/64
P	Vlan-int10	192.168.6.2/24	CE 1	Vlan-int30	10:110:7::1/64
P	Vlan-int40	192.168.7.2/24	CE 1	Vlan-int20	10:110:2::2/64
P	Loop1	2.2.2.2/32	CE 2	Vlan-int60	10:110:9::1/64
PE 1	Vlan-int10	192.168.6.1/24	CE 2	Vlan-int50	10:110:4::2/64
PE 1	Vlan-int20	10:110:2::1/64	PE 2	Vlan-int40	192.168.7.1/24
PE 1	Loop1	1.1.1.1/32	PE 2	Vlan-int50	10:110:4::1/64
PE 1	Loop2	33:33:33::33/128	PE 2	Loop1	1.1.1.2/32

Procedure

1. Configure PE 1:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE1
[PE1] router id 1.1.1.1
[PE1] multicast routing
[PE1-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE1] mpls lsr-id 1.1.1.1
[PE1] mpls ldp
[PE1-ldp] quit
```

Create a VPN instance named **a**, and configure the RD and route targets for the VPN instance.

```
[PE1] ip vpn-instance a
[PE1-vpn-instance-a] route-distinguisher 100:1
[PE1-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE1-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE1-vpn-instance-a] quit
```

Enable IPv6 multicast routing for VPN instance **a**.

```
[PE1] ipv6 multicast routing vpn-instance a
[PE1-mrib6-a] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.1 32
[PE1-LoopBack1] pim sm
[PE1-LoopBack1] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE1] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv6 address family for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a] address-family ipv6
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a-ipv6] default-group 239.1.1.1
[PE1-mvpn-vpn-instance-a-ipv6] source loopback 1
[PE1-mvpn-vpn-instance-a-ipv6] data-group 225.2.2.0 28
[PE1-mvpn-vpn-instance-a-ipv6] quit
[PE1-mvpn-vpn-instance-a] quit
```

Assign an IP address to VLAN-interface 10.

```
[PE1] vlan 10
[PE1-vlan10] port gigabitethernet 1/0/1
[PE1-vlan10] quit
[PE1] interface vlan-interface 10
[PE1-Vlan-interface10] ip address 192.168.6.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 10.

```
[PE1-Vlan-interface10] pim sm
[PE1-Vlan-interface10] mpls enable
[PE1-Vlan-interface10] mpls ldp enable
[PE1-Vlan-interface10] quit
```

Associate VLAN-interface 20 with VPN instance **a**, and assign an IPv6 address to the interface.

```
[PE1] vlan 20
[PE1-vlan20] port gigabitethernet 1/0/2
[PE1-vlan20] quit
[PE1] interface vlan-interface 20
[PE1-Vlan-interface20] ip binding vpn-instance a
[PE1-Vlan-interface20] ipv6 address 10:110:2::1 64
```

Configure VLAN-interface 20 to run OSPFv3 process 2 in Area 0, and enable IPv6 PIM-SM on the interface.

```
[PE1-Vlan-interface20] ospfv3 2 area 0.0.0.0
[PE1-Vlan-interface20] ipv6 pim sm
[PE1-Vlan-interface20] quit
```

Associate Loopback 2 with VPN instance **a**, and assign an IPv6 address to the interface.

```
[PE1] interface loopback 2
[PE1-LoopBack2] ip binding vpn-instance a
[PE1-LoopBack2] ipv6 address 33:33:33::33 128
```

Configure Loopback 2 to run OSPFv3 process 3 in Area 0, and enable IPv6 PIM-SM on the interface.

```
[PE1-LoopBack2] ospfv3 3 area 0.0.0.0
[PE1-LoopBack2] ipv6 pim sm
[PE1-LoopBack2] quit
```

Configure Loopback 2 as a C-BSR and a C-RP.

```
[PE1] ipv6 pim vpn-instance a
[PE1-pim6-a] c-bsr 33:33:33::33
[PE1-pim6-a] c-rp 33:33:33::33
[PE1-pim6-b] quit
```

Configure BGP.

```
[PE1] bgp 100
[PE1-bgp-default] group vpn-g internal
[PE1-bgp-default] peer vpn-g connect-interface loopback 1
[PE1-bgp-default] peer 1.1.1.2 group vpn-g
[PE1–bgp-default] ip vpn-instance a
[PE1-bgp-default-a] address-family ipv6
[PE1-bgp-default-ipv6-a] import-route ospfv3 2
[PE1-bgp-default-ipv6-a] import-route direct
[PE1-bgp-default-ipv6-a] quit
[PE1-bgp-default-a] quit
[PE1–bgp-default] address-family vpnv6
[PE1–bgp-default-vpnv6] peer vpn-g enable
[PE1–bgp-default-vpnv6] quit
[PE1-bgp-default] address-family ipv4 mdt
[PE1-bgp-default-mdt] peer vpn-g enable
[PE1-bgp-default-mdt] quit
[PE1–bgp-default] quit
```

Configure OSPF.

```
[PE1] ospf 1
[PE1-ospf-1] area 0.0.0.0
```

```
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 192.168.6.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

Configure OSPFv3.

```
[PE1] ospfv3 2 vpn-instance a
[PE1-ospfv3-2] router-id 1.1.1.1
[PE1-ospfv3-2] import-route bgp4+
[PE1-ospfv3-2] import-route direct
[PE1-ospfv3-2] area 0
[PE1-ospfv3-2-area-0.0.0.0] quit
[PE1-ospfv3-2] quit
```

2. Configure PE 2:

Configure a global RD, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] router id 1.1.1.2
[PE2] multicast routing
[PE2-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE2] mpls lsr-id 1.1.1.2
[PE2] mpls ldp
[PE2-ldp] quit
```

Create a VPN instance named **a**, and configure the RD and route targets for the VPN instance.

```
[PE2] ip vpn-instance a
[PE2-vpn-instance-a] route-distinguisher 100:1
[PE2-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE2-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE2-vpn-instance-a] quit
```

Enable IPv6 multicast routing for VPN instance **a**.

```
[PE2] ipv6 multicast routing vpn-instance a
[PE2-mrib6-a] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 1.1.1.2 32
[PE2-LoopBack1] pim sm
[PE2-LoopBack1] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE2] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv6 address family for VPN instance **a**.

```
[PE2-mvpn-vpn-instance-a] address-family ipv6
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE2-mvpn-vpn-instance-a-ipv6] default-group 239.1.1.1
[PE2-mvpn-vpn-instance-a-ipv6] source loopback 1
[PE2-mvpn-vpn-instance-a-ipv6] data-group 225.2.2.0 28
[PE2-mvpn-vpn-instance-a-ipv6] quit
```

```
[PE2-mvpn-vpn-instance-a] quit
```

Assign an IP address to VLAN-interface 40.

```
[PE2] vlan 40
[PE2-vlan40] port gigabitethernet 1/0/1
[PE2-vlan40] quit
[PE2] interface vlan-interface 40
[PE2-Vlan-interface40] ip address 192.168.7.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 40.

```
[PE2-Vlan-interface40] pim sm
[PE2-Vlan-interface40] mpls enable
[PE2-Vlan-interface40] mpls ldp enable
[PE2-Vlan-interface40] quit
```

Associate VLAN-interface 50 with VPN instance **a**, and assign an IPv6 address to the interface.

```
[PE2] vlan 50
[PE2-vlan50] port gigabitethernet 1/0/2
[PE2-vlan50] quit
[PE2] interface vlan-interface 50
[PE2-Vlan-interface50] ip binding vpn-instance a
[PE2-Vlan-interface50] ipv6 address 10:110:4::1 64
```

Configure VLAN-interface 50 to run OSPFv3 process 3 in Area 0, and enable IPv6 PIM-SM on the interface.

```
[PE2-Vlan-interface50] ospfv3 2 area 0.0.0.0
[PE2-Vlan-interface50] ipv6 pim sm
[PE2-Vlan-interface50] quit
```

Configure BGP.

```
[PE2] bgp 100
[PE2-bgp-default] group vpn-g internal
[PE2-bgp-default] peer vpn-g connect-interface loopback 1
[PE2-bgp-default] peer 1.1.1.1 group vpn-g
[PE2–bgp-default] ip vpn-instance a
[PE2-bgp-default-a] address-family ipv6
[PE2-bgp-default-ipv6-a] import-route ospfv3 2
[PE2-bgp-default-ipv6-a] import-route direct
[PE2-bgp-default-ipv6-a] quit
[PE2-bgp-default-a] quit
[PE2–bgp-default] address-family vpnv6
[PE2–bgp-default-vpnv6] peer vpn-g enable
[PE2–bgp-default-vpnv6] quit
[PE2-bgp-default] address-family ipv4 mdt
[PE2-bgp-default-mdt] peer vpn-g enable
[PE2-bgp-default-mdt] quit
[PE2–bgp-default] quit
```

Configure OSPF.

```
[PE2] ospf 1
[PE2-ospf-1] area 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 1.1.1.2 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 192.168.7.0 0.0.0.255
```

```
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

Configure OSPFv3.

```
[PE2] ospfv3 2 vpn-instance a
[PE2-ospfv3-2] router-id 2.2.2.2
[PE2-ospfv3-2] import-route bgp4+
[PE2-ospfv3-2] import-route direct
[PE2-ospfv3-2] area 0
[PE2-ospfv3-2-area-0.0.0.0] quit
[PE2-ospfv3-2] quit
```

3. Configure P:

Enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname P
[P] multicast routing
[P-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[P] mpls lsr-id 2.2.2.2
[P] mpls ldp
[P-ldp] quit
```

Assign an IP address to VLAN-interface 10.

```
[P] vlan 10
[P-vlan10] port gigabitethernet 1/0/1
[P-vlan10] quit
[P] interface vlan-interface 10
[P-Vlan-interface10] ip address 192.168.6.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 10.

```
[P-Vlan-interface10] pim sm
[P-Vlan-interface10] mpls enable
[P-Vlan-interface10] mpls ldp enable
[P-Vlan-interface10] quit
```

Assign an IP address to VLAN-interface 40.

```
[P] vlan 40
[P-vlan40] port gigabitethernet 1/0/2
[P-vlan40] quit
[P] interface vlan-interface 40
[P-Vlan-interface40] ip address 192.168.7.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 40.

```
[P-Vlan-interface40] pim sm
[P-Vlan-interface40] mpls enable
[P-Vlan-interface40] mpls ldp enable
[P-Vlan-interface40] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[P] interface loopback 1
[P-LoopBack1] ip address 2.2.2.2 32
[P-LoopBack1] pim sm
[P-LoopBack1] quit
```

Configure Loopback 1 as a C-BSR and a C-RP.

```
[P] pim
[P-pim] c-bsr 2.2.2.2
[P-pim] c-rp 2.2.2.2
[P-pim] quit
```

Configure OSPF.

```
[P] ospf 1
[P-ospf-1] area 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 192.168.6.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 192.168.7.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

4. Configure CE 1:

Enable IPv6 multicast routing.

```
<Sysname> system-view
[Sysname] sysname CE1
[CE1] ipv6 multicast routing
[CE1-mrib6] quit
```

Assign an IPv6 address to VLAN-interface 30.

```
[CE1-vlan30] port gigabitethernet 1/0/1
[CE1-vlan30] quit
[CE1] interface vlan-interface 30
[CE1-Vlan-interface30] ipv6 address 10:110:7::1 64
```

Configure VLAN-interface 30 to run OSPFv3 process 2 in Area 0, and enable IPv6 PIM-SM on the interface.

```
[CE1-Vlan-interface30] ospfv3 2 area 0.0.0.0
[CE1-Vlan-interface30] ipv6 pim sm
[CE1-Vlan-interface30] quit
```

Assign an IPv6 address to VLAN-interface 11.

```
[CE1] interface vlan-interface 11
[CE1-Vlan-interface11] ipv6 address 10:110:2::2 64
```

Configure VLAN-interface 20 to run OSPFv3 process 2 in Area 0, and enable IPv6 PIM-SM on the interface.

```
[CE1] vlan 20
[CE1-vlan20] port gigabitethernet 1/0/2
[CE1-vlan20] quit
[CE1-Vlan-interface20] ospfv3 2 area 0.0.0.0
[CE1-Vlan-interface20] ipv6 pim sm
[CE1-Vlan-interface20] quit
```

Configure OSPFv3.

```
[CE1] ospfv3 2
[CE1-ospfv3-2] router-id 6.6.6.6
[CE1-ospfv3-2] area 0.0.0.0
[CE1-ospfv3-2-area-0.0.0.0] quit
[CE1-ospfv3-2] quit
```

5. Configure CE 2:

Enable IPv6 multicast routing.

```
<Sysname> system-view
[Sysname] sysname CE2
[CE2] ipv6 multicast routing
[CE2-mrib6] quit
```

Assign an IPv6 address to VLAN-interface 60.

```
[CE2] vlan 60
[CE2-vlan60] port gigabitethernet 1/0/1
[CE2-vlan60] quit
[CE2] interface vlan-interface 60
[CE2-Vlan-interface60] ipv6 address 10:110:9::1 64
```

Configure VLAN-interface 60 to run OSPFv3 process 2 in Area 0, and enable MLD on the interface.

```
[CE2-Vlan-interface60] ospfv3 2 area 0.0.0.0
[CE2-Vlan-interface60] mld enable
[CE2-Vlan-interface60] quit
```

Assign an IPv6 address to VLAN-interface 50.

```
[CE2] vlan 50
[CE2-vlan50] port gigabitethernet 1/0/2
[CE2-vlan50] quit
[CE2] interface vlan-interface 50
[CE2-Vlan-interface50] ipv6 address 10:110:4::2 64
```

Configure VLAN-interface 50 to run OSPFv3 process 2 in Area 0, and enable IPv6 PIM-SM on the interface.

```
[CE2-Vlan-interface50] ospfv3 2 area 0.0.0.0
[CE2-Vlan-interface50] ipv6 pim sm
[CE2-Vlan-interface50] quit
```

Configure OSPFv3.

```
[CE2] ospfv3 2
[CE2-ospfv3-2] router-id 8.8.8.8
[CE2-ospfv3-2] area 0.0.0.0
[CE2-ospfv3-2-area-0.0.0.0] quit
[CE2-ospfv3-2] quit
```

Verifying the configuration

Display information about the local default group for IPv6 multicast transmission in each VPN instance on PE 1.

```
[PE1] display multicast-vpn default-group ipv6 local
MVPN local default-group information:
Group address    Source address   Interface     VPN instance
239.1.1.1        1.1.1.1          MTunnel0      a
```

Display information about the local default group for IPv6 multicast transmission in each VPN instance on PE 2.

```
[PE2] display multicast-vpn default-group ipv6 local
MVPN local default-group information:
Group address    Source address   Interface     VPN instance
239.1.1.1        1.1.1.2          MTunnel0      a
```

Configuration files

- CE 1:

```
#
 sysname CE1
#
ospfv3 2
 router-id 6.6.6.6
 area 0.0.0.0
#
vlan 20
#
vlan 30
#
interface Vlan-interface20
 ospfv3 2 area 0.0.0.0
 ipv6 pim sm
 ipv6 address 10:110:2::2/64
#
interface Vlan-interface30
 ospfv3 2 area 0.0.0.0
 ipv6 pim sm
 ipv6 address 10:110:7::1/64
#
interface GigabitEthernet1/0/1
 port access vlan 30
#
interface GigabitEthernet1/0/2
 port access vlan 20
#
ipv6 multicast routing
#
return
```

- CE 2:

```
#
 sysname CE2
#
ospfv3 2
 router-id 8.8.8.8
 area 0.0.0.0
#
vlan 50
#
vlan 60
#
interface Vlan-interface50
 ospfv3 2 area 0.0.0.0
 ipv6 pim sm
 ipv6 address 10:110:4::2/64
#
interface Vlan-interface60
```

```
 ospfv3 2 area 0.0.0.0
 ipv6 address 10:110:9::1/64
 mld enable
#
interface GigabitEthernet1/0/1
 port access vlan 60
#
interface GigabitEthernet1/0/2
 port access vlan 50
#
ipv6 multicast routing
#
return
```

- P:

```
#
 sysname P
#
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 192.168.6.0 0.0.0.255
  network 192.168.7.0 0.0.0.255
#
 mpls lsr-id 2.2.2.2
#
vlan 10
#
vlan 40
#
mpls ldp
#
interface LoopBack1
 ip address 2.2.2.2 255.255.255.255
 pim sm
#
interface Vlan-interface10
 ip address 192.168.6.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface40
 ip address 192.168.7.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
```

```
 port access vlan 10
#
interface GigabitEthernet1/0/2
 port access vlan 40
#
multicast routing
#
pim
 c-bsr 2.2.2.2
 c-rp 2.2.2.2
#
return
```

- PE 1:

```
#
 sysname PE1
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
 router id 1.1.1.1
#
ospf 1
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 192.168.6.0 0.0.0.255
#
ospfv3 2 vpn-instance a
 router-id 1.1.1.1
 import-route direct
 import-route bgp4+
 area 0.0.0.0
#
ospfv3 3 vpn-instance a
 area 0.0.0.0
#
 mpls lsr-id 1.1.1.1
#
vlan 10
#
vlan 20
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.1 255.255.255.255
 pim sm
```

```
#
interface LoopBack2
 ip binding vpn-instance a
 ospfv3 3 area 0.0.0.0
 ipv6 pim sm
 ipv6 address 33:33:33::33/128
#
interface Vlan-interface10
 ip address 192.168.6.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface20
 ip binding vpn-instance a
 ospfv3 2 area 0.0.0.0
 ipv6 pim sm
 ipv6 address 10:110:2::1/64
#
interface GigabitEthernet1/0/1
 port access vlan 10
#
interface GigabitEthernet1/0/2
 port access vlan 20
#
bgp 100
 group vpn-g internal
 peer vpn-g connect-interface LoopBack1
 peer 1.1.1.2 group vpn-g
 #
 address-family ipv4 mdt
  peer vpn-g enable
 #
 address-family vpnv6
  peer vpn-g enable
 #
 ip vpn-instance a
  #
  address-family ipv6 unicast
   import-route direct
   import-route ospfv3 2
#
multicast routing
#
ipv6 multicast routing vpn-instance a
#
ipv6 pim vpn-instance a
 c-bsr 33:33:33::33
```

```
 c-rp 33:33:33::33
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv6
  source LoopBack1
  default-group 239.1.1.1
  data-group 225.2.2.0 255.255.255.240
#
return
```

- PE 2:

```
#
 sysname PE2
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
 router id 1.1.1.2
#
ospf 1
 area 0.0.0.0
  network 1.1.1.2 0.0.0.0
  network 192.168.7.0 0.0.0.255
#
ospfv3 2 vpn-instance a
 router-id 2.2.2.2
 import-route direct
 import-route bgp4+
 area 0.0.0.0
#
 mpls lsr-id 1.1.1.2
#
vlan 40
#
vlan 50
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.2 255.255.255.255
 pim sm
#
interface Vlan-interface40
 ip address 192.168.7.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
```

```
#
interface Vlan-interface50
 ip binding vpn-instance a
 ospfv3 2 area 0.0.0.0
 ipv6 pim sm
 ipv6 address 10:110:4::1/64
#
interface GigabitEthernet1/0/1
 port access vlan 40
#
interface GigabitEthernet1/0/2
 port access vlan 50
#
interface GigabitEthernet1/0/3
#
bgp 100
 group vpn-g internal
 peer vpn-g connect-interface LoopBack1
 peer 1.1.1.1 group vpn-g
 #
 address-family ipv4 mdt
  peer vpn-g enable
 #
 address-family vpnv6
  peer vpn-g enable
 #
 ip vpn-instance a
  #
  address-family ipv6 unicast
   import-route direct
   import-route ospfv3 2
#
multicast routing
#
ipv6 multicast routing vpn-instance a
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv6
  source LoopBack1
  default-group 239.1.1.1
  data-group 225.2.2.0 255.255.255.240
#
return
```

Example: Configuring MDT-based MVPN inter-AS option B

Network configuration

As shown in Figure 16, configure inter-AS option B MDT-based MVPN to meet the following requirements:

Item	Network configuration
Multicast sources and receivers	• In VPN instance **a**, S 1 is a multicast source, and R 2 is a receiver. • In VPN instance **b**, S 2 is a multicast source, and R 1 is a receiver. • For VPN instance **a**, the default group is 232.1.1.1, and the data group range is 232.2.2.0 to 232.2.2.15. They are in the SSM group range. • For VPN instance **b**, the default group is 232.3.3.3, and the data group range is 232.4.4.0 to 232.4.4.15. They are in the SSM group range.
VPN instances to which PE interfaces belong	• PE 1: VLAN-interface 11 belongs to VPN instance **a**. VLAN-interface 12 belongs to VPN instance **b**. VLAN-interface 2 and Loopback 1 belong to the public network instance. • PE 2: VLAN-interface 3, VLAN-interface 4, and Loopback 1 belong to the public network instance. • PE 3: VLAN-interface 4, VLAN-interface 5, and Loopback 1 belong to the public network instance. • PE 4: VLAN-interface 13 belongs to VPN instance **a**. VLAN-interface 14 belongs to VPN instance **b**. VLAN-interface 6 and Loopback 1 belong to the public network instance.
Unicast routing protocols and MPLS	• Configure OSPF separately in AS 100 and AS 200, and configure OSPF between the PEs and CEs. • Establish IBGP peer connections between PE 1, PE 2, PE 3, and PE 4 on their respective Loopback 1. • Configure BGP MDT peer connections between PE 1, PE 2, PE 3, and PE 4 on their respective Loopback 1 and between PE 2 and PE 3 on their respective VLAN-interface 4. • Establish EBGP peer connections between VLAN-interface 4 on PE 2 and PE 3. • Configure MPLS separately in AS 100 and AS 200.
IP multicast routing	• Enable IP multicast routing on P 1 and P 2. • Enable IP multicast routing on the public network on PE 1, PE 2, PE 3, and PE 4. • Enable IP multicast routing for VPN instance **a** on PE 1 and PE 4. • Enable IP multicast routing for VPN instance **b** on PE 1 and PE 4. • Enable IP multicast routing on CE a1, CE a2, CE b1, and CE b2.
IGMP	• Enable IGMPv2 on VLAN-interface 23 of CE a2. • Enable IGMPv2 on VLAN-interface 24 of CE b2.
PIM	Enable PIM-SSM on the public network and PIM-SM on VPN instances **a** and **b**: • Enable PIM-SM on all interfaces of P 1 and P 2. • Enable PIM-SM on all public network interfaces of PE 2 and PE 3. • Enable PIM-SM on all public and private network interfaces of PE 1 and PE 4. • Enable PIM-SM on all interfaces that do not have attached receiver hosts on CE a1, CE a2, CE b1, and CE b2. • Configure VLAN-interface 11 of CE a1 as a C-BSR and a C-RP for VPN instance **a** to provide services for all multicast groups. • Configure VLAN-interface 12 of CE b1 as a C-BSR and a C-RP for VPN instance **b** to provide services for all multicast groups.
RPF vector	Enable the RPF vector feature on PE 1 and PE 4.

Figure 16 Network diagram

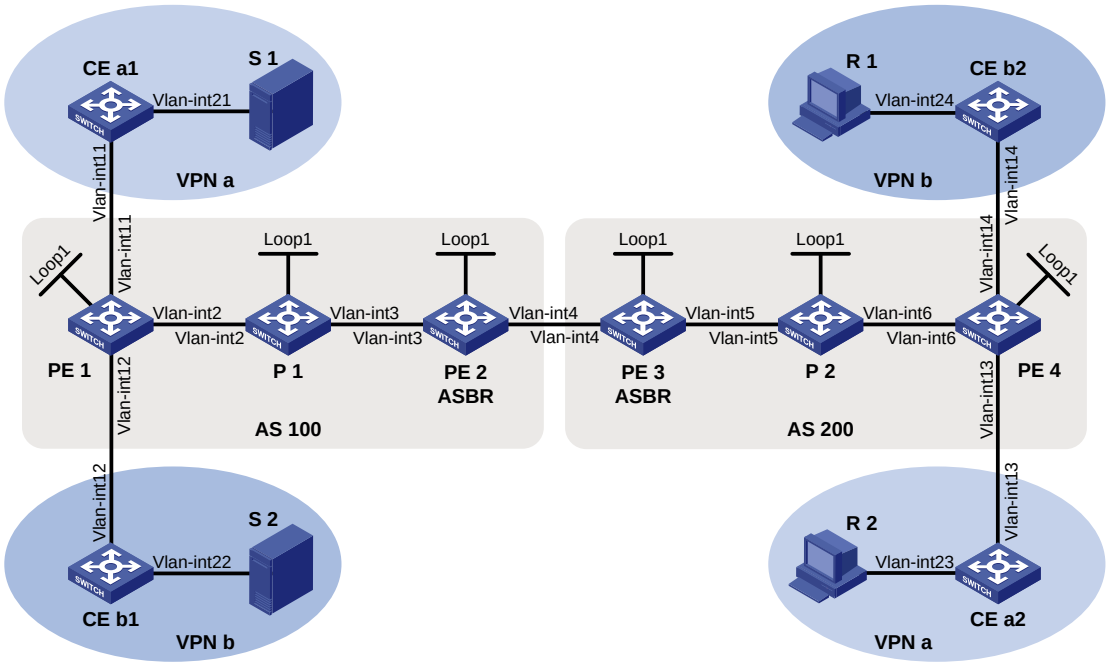


Table 3 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
S 1	—	12.1.1.100/24	R 1	—	12.4.1.100/24
S 2	—	12.2.1.100/24	R 2	—	12.3.1.100/24
PE 1	Vlan-int2	10.1.1.1/24	PE 3	Vlan-int5	10.4.1.1/24
PE 1	Vlan-int11	11.1.1.1/24	PE 3	Vlan-int4	10.3.1.2/24
PE 1	Vlan-int12	11.2.1.1/24	PE 3	Loop1	3.3.3.3/32
PE 1	Loop1	1.1.1.1/32	PE 4	Vlan-int6	10.5.1.2/24
PE 2	Vlan-int3	10.2.1.2/24	PE 4	Vlan-int13	11.3.1.1/24
PE 2	Vlan-int4	10.3.1.1/24	PE 4	Vlan-int14	11.4.1.1/24
PE 2	Loop1	2.2.2.2/32	PE 4	Loop1	4.4.4.4/24
P 1	Vlan-int2	10.1.1.2/24	P 2	Vlan-int6	10.5.1.1/24
P 1	Vlan-int3	10.2.1.1/24	P 2	Vlan-int5	10.4.1.2/24
P 1	Loop1	5.5.5.5/32	P 2	Loop1	6.6.6.6/32
CE a1	Vlan-int21	12.1.1.1/24	CE b1	Vlan-int22	12.2.1.1/24
CE a1	Vlan-int11	11.1.1.2/24	CE b1	Vlan-int12	11.2.1.2/24
CE a2	Vlan-int23	12.3.1.1/24	CE b2	Vlan-int24	12.4.1.1/24
CE a2	Vlan-int13	11.3.1.2/24	CE b2	Vlan-int14	11.4.1.2/24

Procedure

1. Configure PE 1:

   # Configure a global router ID, and enable IP multicast routing on the public network.

   ```
   <Sysname> system-view
   ```

```
[Sysname] sysname PE1
[PE1] router id 1.1.1.1
[PE1] multicast routing
[PE1-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE1] mpls lsr-id 1.1.1.1
[PE1] mpls ldp
[PE1-ldp] quit
```

Create a VPN instance named **a**, and configure the RD and route targets for the VPN instance.

```
[PE1] ip vpn-instance a
[PE1-vpn-instance-a] route-distinguisher 100:1
[PE1-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE1-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE1-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE1] multicast routing vpn-instance a
```

Enable the RPF vector feature.

```
[PE1-mrib-a] rpf proxy vector
[PE1-mrib-a] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.1 32
[PE1-LoopBack1] pim sm
[PE1-LoopBack1] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE1] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 address family for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a-ipv4] default-group 232.1.1.1
[PE1-mvpn-vpn-instance-a-ipv4] source loopback 1
[PE1-mvpn-vpn-instance-a-ipv4] data-group 232.2.2.0 28
[PE1-mvpn-vpn-instance-a-ipv4] quit
[PE1-mvpn-vpn-instance-a] quit
```

Create a VPN instance named **b**, and configure the RD and route targets for VPN instance.

```
[PE1] ip vpn-instance b
[PE1-vpn-instance-b] route-distinguisher 200:1
[PE1-vpn-instance-b] vpn-target 200:1 export-extcommunity
[PE1-vpn-instance-b] vpn-target 200:1 import-extcommunity
[PE1-vpn-instance-b] quit
```

Enable IP multicast routing for VPN instance **b**.

```
[PE1] multicast routing vpn-instance b
```

Enable the RPF vector feature.

```
[PE1-mrib-b] rpf proxy vector
[PE1-mrib-b] quit
```

Create an MDT-based MVPN for VPN instance **b**.

```
[PE1] multicast-vpn vpn-instance b mode mdt
```

Create an MVPN IPv4 address family for VPN instance **b**.

```
[PE1-mvpn-vpn-instance-b] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **b**.

```
[PE1-mvpn-vpn-instance-b-ipv4] default-group 232.3.3.3
[PE1-mvpn-vpn-instance-b-ipv4] source loopback 1
[PE1-mvpn-vpn-instance-b-ipv4] data-group 232.4.4.0 28
[PE1-mvpn-vpn-instance-b-ipv4] quit
[PE1-mvpn-vpn-instance-b] quit
```

Assign an IP address to VLAN-interface 2.

```
[PE1] vlan 2
[PE1-vlan2] port gigabitethernet 1/0/1
[PE1-vlan2] quit
[PE1] interface vlan-interface 2
[PE1-Vlan-interface2] ip address 10.1.1.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 2.

```
[PE1-Vlan-interface2] pim sm
[PE1-Vlan-interface2] mpls enable
[PE1-Vlan-interface2] mpls ldp enable
[PE1-Vlan-interface2] quit
```

Associate VLAN-interface 11 with VPN instance **a**.

```
[PE1] vlan 11
[PE1-vlan11] port gigabitethernet 1/0/2
[PE1-vlan11] quit
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip binding vpn-instance a
```

Assign an IP address to VLAN-interface 11, and enable PIM-SM on the interface.

```
[PE1-Vlan-interface11] ip address 11.1.1.1 24
[PE1-Vlan-interface11] pim sm
[PE1-Vlan-interface11] quit
```

Associate VLAN-interface 12 with VPN instance **b**.

```
[PE1] vlan 12
[PE1-vlan12] port gigabitethernet 1/0/3
[PE1-vlan12] quit
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance b
```

Assign an IP address to VLAN-interface 12, and enable PIM-SM on the interface.

```
[PE1-Vlan-interface12] ip address 11.2.1.1 24
[PE1-Vlan-interface12] pim sm
[PE1-Vlan-interface12] quit
```

Configure BGP.

```
[PE1] bgp 100
[PE1-bgp-default] peer 2.2.2.2 as-number 100
[PE1-bgp-default] peer 2.2.2.2 connect-interface loopback 1
[PE1–bgp-default] ip vpn-instance a
[PE1-bgp-default-a] address-family ipv4
[PE1-bgp-default-ipv4-a] import-route ospf 2
```

```
[PE1-bgp-default-ipv4-a] import-route direct
[PE1-bgp-default-ipv4-a] quit
[PE1-bgp-default-a] quit
[PE1–bgp-default] ip vpn-instance b
[PE1-bgp-default-b] address-family ipv4
[PE1-bgp-default-ipv4-b] import-route ospf 3
[PE1-bgp-default-ipv4-b] import-route direct
[PE1-bgp-default-ipv4-b] quit
[PE1-bgp-default-b] quit
[PE1–bgp-default] address-family vpnv4
[PE1–bgp-default-vpnv4] peer 2.2.2.2 enable
[PE1–bgp-default-vpnv4] quit
[PE1-bgp-default] address-family ipv4 mdt
[PE1-bgp-default-mdt] peer 2.2.2.2 enable
[PE1-bgp-default-mdt] quit
[PE1–bgp-default] quit
```

Configure OSPF.

```
[PE1] ospf 1
[PE1-ospf-1] area 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
[PE1] ospf 2 vpn-instance a
[PE1-ospf-2] import-route bgp
[PE1-ospf-2] area 0.0.0.0
[PE1-ospf-2-area-0.0.0.0] network 11.1.1.0 0.0.0.255
[PE1-ospf-2-area-0.0.0.0] quit
[PE1-ospf-2] quit
[PE1] ospf 3 vpn-instance b
[PE1-ospf-3] import-route bgp
[PE1-ospf-3] area 0.0.0.0
[PE1-ospf-3-area-0.0.0.0] network 11.2.1.0 0.0.0.255
[PE1-ospf-3-area-0.0.0.0] quit
[PE1-ospf-3] quit
```

2. Configure PE 2:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] router id 2.2.2.2
[PE2] multicast routing
[PE2-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE2] mpls lsr-id 2.2.2.2
[PE2] mpls ldp
[PE2-ldp] quit
```

Assign an IP address to VLAN-interface 3.

```
[PE2] vlan 3
```

```
[PE2-vlan3] port gigabitethernet 1/0/1
[PE2-vlan3] quit
[PE2] interface vlan-interface 3
[PE2-Vlan-interface3] ip address 10.2.1.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 3.

```
[PE2-Vlan-interface3] pim sm
[PE2-Vlan-interface3] mpls enable
[PE2-Vlan-interface3] mpls ldp enable
[PE2-Vlan-interface3] quit
```

Assign an IP address to VLAN-interface 4.

```
[PE2] vlan 4
[PE2-vlan4] port gigabitethernet 1/0/2
[PE2-vlan4] quit
[PE2] interface vlan-interface 4
[PE2-Vlan-interface4] ip address 10.3.1.1 24
```

Enable PIM-SM and MPLS on VLAN-interface 4.

```
[PE2-Vlan-interface4] pim sm
[PE2-Vlan-interface4] mpls enable
[PE2-Vlan-interface4] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 2.2.2.2 32
[PE2-LoopBack1] pim sm
[PE2-LoopBack1] quit
```

Configure BGP.

```
[PE2] bgp 100
[PE2-bgp-default] peer 1.1.1.1 as-number 100
[PE2-bgp-default] peer 1.1.1.1 connect-interface loopback 1
[PE2-bgp-default] peer 10.3.1.2 as-number 200
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] undo policy vpn-target
[PE2-bgp-default-vpnv4] peer 1.1.1.1 enable
[PE2-bgp-default-vpnv4] peer 10.3.1.2 enable
[PE2-bgp-default-vpnv4] quit
[PE2-bgp-default] address-family ipv4 mdt
[PE2-bgp-default-mdt] peer 1.1.1.1 enable
[PE2-bgp-default-mdt] peer 10.3.1.2 enable
[PE2-bgp-default-mdt] quit
[PE2–bgp-default] quit
```

Configure OSPF.

```
[PE2] ospf 1
[PE2-ospf-1] area 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

3. Configure PE 3:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE3
[PE3] router id 3.3.3.3
[PE3] multicast routing
[PE3-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE3] mpls lsr-id 3.3.3.3
[PE3] mpls ldp
[PE3-ldp] quit
```

Assign an IP address to VLAN-interface 5.

```
[PE3] vlan 5
[PE3-vlan5] port gigabitethernet 1/0/1
[PE3-vlan5] quit
[PE3] interface vlan-interface 5
[PE3-Vlan-interface5] ip address 10.4.1.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 5.

```
[PE3-Vlan-interface5] pim sm
[PE3-Vlan-interface5] mpls enable
[PE3-Vlan-interface5] mpls ldp enable
[PE3-Vlan-interface5] quit
```

Assign an IP address to VLAN-interface 4.

```
[PE3] vlan 4
[PE3-vlan4] port gigabitethernet 1/0/2
[PE3-vlan4] quit
[PE3] interface vlan-interface 4
[PE3-Vlan-interface4] ip address 10.3.1.2 24
```

Enable PIM-SM and MPLS on VLAN-interface 4.

```
[PE3-Vlan-interface4] pim sm
[PE3-Vlan-interface4] mpls enable
[PE3-Vlan-interface4] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE3] interface loopback 1
[PE3-LoopBack1] ip address 3.3.3.3 32
[PE3-LoopBack1] pim sm
[PE3-LoopBack1] quit
```

Configure BGP.

```
[PE3] bgp 200
[PE3-bgp-default] peer 4.4.4.4 as-number 200
[PE3-bgp-default] peer 4.4.4.4 connect-interface loopback 1
[PE3-bgp-default] peer 10.3.1.1 as-number 100
[PE3-bgp-default] address-family vpnv4
[PE3-bgp-default-vpnv4] undo policy vpn-target
[PE3-bgp-default-vpnv4] peer 4.4.4.4 enable
[PE3-bgp-default-vpnv4] peer 10.3.1.1 enable
[PE3-bgp-default-vpnv4] quit
[PE3-bgp-default] address-family ipv4 mdt
[PE3-bgp-default-mdt] peer 4.4.4.4 enable
[PE3-bgp-default-mdt] peer 10.3.1.1 enable
```

```
[PE3-bgp-default-mdt] quit
[PE3–bgp-default] quit
```

Configure OSPF.

```
[PE3] ospf 1
[PE3-ospf-1] area 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[PE3-ospf-1-area-0.0.0.0] quit
[PE3-ospf-1] quit
```

4. Configure PE 4:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE4
[PE4] router id 4.4.4.4
[PE4] multicast routing
[PE4-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE4] mpls lsr-id 4.4.4.4
[PE4] mpls ldp
[PE4-ldp] quit
```

Create a VPN instance named **a**, and configure the RD and route targets for the VPN instance.

```
[PE4] ip vpn-instance a
[PE4-vpn-instance-a] route-distinguisher 100:1
[PE4-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE4-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE4-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE4] multicast routing vpn-instance a
```

Enable the RPF vector feature.

```
[PE4-mrib-a] rpf proxy vector
[PE4-mrib-a] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE4] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 address family for VPN instance **a**.

```
[PE4-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE4-mvpn-vpn-instance-a-ipv4] default-group 232.1.1.1
[PE4-mvpn-vpn-instance-a-ipv4] source loopback 1
[PE4-mvpn-vpn-instance-a-ipv4] data-group 232.2.2.0 28
[PE4-mvpn-vpn-instance-a-ipv4] quit
[PE4-mvpn-vpn-instance-a] quit
```

Create a VPN instance named **b**, and configure the RD and route targets for the VPN instance.

```
[PE4] ip vpn-instance b
[PE4-vpn-instance-b] route-distinguisher 200:1
[PE4-vpn-instance-b] vpn-target 200:1 export-extcommunity
```

```
[PE4-vpn-instance-b] vpn-target 200:1 import-extcommunity
[PE4-vpn-instance-b] quit
```

Enable IP multicast routing for VPN instance **b**.

```
[PE4] multicast routing vpn-instance b
```

Enable the RPF vector feature.

```
[PE4-mrib-b] rpf proxy vector
[PE4-mrib-b] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE4] interface loopback 1
[PE4-LoopBack1] ip address 4.4.4.4 32
[PE4-LoopBack1] pim sm
[PE4-LoopBack1] quit
```

Create an MDT-based MVPN for VPN instance **b**.

```
[PE4] multicast-vpn vpn-instance b mode mdt
```

Create an MVPN IPv4 address family for VPN instance **b**.

```
[PE4-mvpn-vpn-instance-b] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **b**.

```
[PE4-mvpn-vpn-instance-b-ipv4] default-group 232.3.3.3
[PE4-mvpn-vpn-instance-b-ipv4] source loopback 1
[PE4-mvpn-vpn-instance-b-ipv4] data-group 232.4.4.0 28
[PE4-mvpn-vpn-instance-b-ipv4] quit
[PE4-mvpn-vpn-instance-b] quit
```

Assign an IP address to VLAN-interface 6.

```
[PE4] vlan 6
[PE4-vlan6] port gigabitethernet 1/0/1
[PE4-vlan6] quit
[PE4] interface vlan-interface 6
[PE4-Vlan-interface6] ip address 10.5.1.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 6.

```
[PE4-Vlan-interface6] pim sm
[PE4-Vlan-interface6] mpls enable
[PE4-Vlan-interface6] mpls ldp enable
[PE4-Vlan-interface6] quit
```

Associate VLAN-interface 13 with VPN instance **a**.

```
[PE4] vlan 13
[PE4-vlan13] port gigabitethernet 1/0/2
[PE4-vlan13] quit
[PE4] interface vlan-interface 13
[PE4-Vlan-interface13] ip binding vpn-instance a
```

Assign an IP address to VLAN-interface 13, and enable PIM-SM on the interface.

```
[PE4-Vlan-interface13] ip address 11.3.1.1 24
[PE4-Vlan-interface13] pim sm
[PE4-Vlan-interface13] quit
```

Associate VLAN-interface 14 with VPN instance **b.**

```
[PE4] vlan 14
[PE4-vlan14] port gigabitethernet 1/0/3
```

```
[PE4-vlan14] quit
[PE4] interface vlan-interface 14
[PE4-Vlan-interface14] ip binding vpn-instance b
```

Assign an IP address to VLAN-interface 14, and enable PIM-SM on the interface.

```
[PE4-Vlan-interface14] ip address 11.4.1.1 24
[PE4-Vlan-interface14] pim sm
[PE4-Vlan-interface14] quit
```

Configure BGP.

```
[PE4] bgp 200
[PE4-bgp-default] peer 3.3.3.3 as-number 200
[PE4-bgp-default] peer 3.3.3.3 connect-interface loopback 1
[PE4–bgp-default] ip vpn-instance a
[PE4-bgp-default-a] address-family ipv4
[PE4-bgp-default-ipv4-a] import-route ospf 2
[PE4-bgp-default-ipv4-a] import-route direct
[PE4-bgp-default-ipv4-a] quit
[PE4-bgp-default-a] quit
[PE4–bgp-default] ip vpn-instance b
[PE4-bgp-default-b] address-family ipv4
[PE4-bgp-default-ipv4-b] import-route ospf 3
[PE4-bgp-default-ipv4-b] import-route direct
[PE4-bgp-default-ipv4-b] quit
[PE4-bgp-default-b] quit
[PE4–bgp-default] address-family vpnv4
[PE4–bgp-default-vpnv4] peer 3.3.3.3 enable
[PE4–bgp-default-vpnv4] quit
[PE4-bgp-default] address-family ipv4 mdt
[PE4-bgp-default-mdt] peer 3.3.3.3 enable
[PE4-bgp-default-mdt] quit
[PE4–bgp-default] quit
```

Configure OSPF.

```
[PE4] ospf 1
[PE4-ospf-1] area 0.0.0.0
[PE4-ospf-1-area-0.0.0.0] network 4.4.4.4 0.0.0.0
[PE4-ospf-1-area-0.0.0.0] network 10.5.1.0 0.0.0.255
[PE4-ospf-1-area-0.0.0.0] quit
[PE4-ospf-1] quit
[PE4] ospf 2 vpn-instance a
[PE4-ospf-2] import-route bgp
[PE4-ospf-2] area 0.0.0.0
[PE4-ospf-2-area-0.0.0.0] network 11.3.1.0 0.0.0.255
[PE4-ospf-2-area-0.0.0.0] quit
[PE4-ospf-2] quit
[PE4] ospf 3 vpn-instance b
[PE4-ospf-3] import-route bgp
[PE4-ospf-3] area 0.0.0.0
[PE4-ospf-3-area-0.0.0.0] network 11.4.1.0 0.0.0.255
[PE4-ospf-3-area-0.0.0.0] quit
```

```
[PE4-ospf-3] quit
```

5. Configure P 1:

Enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname P1
[P1] multicast routing
[P1-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[P1] mpls lsr-id 5.5.5.5
[P1] mpls ldp
[P1-ldp] quit
```

Assign an IP address to VLAN-interface 2.

```
[P1] vlan 2
[P1-vlan2] port gigabitethernet 1/0/1
[P1-vlan2] quit
[P1] interface vlan-interface 2
[P1-Vlan-interface2] ip address 10.1.1.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 2.

```
[P1-Vlan-interface2] pim sm
[P1-Vlan-interface2] mpls enable
[P1-Vlan-interface2] mpls ldp enable
[P1-Vlan-interface2] quit
```

Assign an IP address to VLAN-interface 3.

```
[P1] vlan 3
[P1-vlan3] port gigabitethernet 1/0/2
[P1-vlan3] quit
[P1] interface vlan-interface 3
[P1-Vlan-interface3] ip address 10.2.1.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 3.

```
[P1-Vlan-interface3] pim sm
[P1-Vlan-interface3] mpls enable
[P1-Vlan-interface3] mpls ldp enable
[P1-Vlan-interface3] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[P1] interface loopback 1
[P1-LoopBack1] ip address 5.5.5.5 32
[P1-LoopBack1] pim sm
[P1-LoopBack1] quit
```

Configure OSPF.

```
[P1] ospf 1
[P1-ospf-1] area 0.0.0.0
[P1-ospf-1-area-0.0.0.0] network 5.5.5.5 0.0.0.0
[P1-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[P1-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[P1-ospf-1-area-0.0.0.0] quit
[P1-ospf-1] quit
```

6. Configure P 2:

Enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname P2
[P2] multicast routing
[P2-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[P2] mpls lsr-id 6.6.6.6
[P2] mpls ldp
[P2-ldp] quit
```

Assign an IP address to VLAN-interface 6.

```
[P2] vlan 6
[P2-vlan6] port gigabitethernet 1/0/1
[P2-vlan6] quit
[P2] interface vlan-interface 6
[P2-Vlan-interface6] ip address 10.5.1.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 6.

```
[P2-Vlan-interface6] pim sm
[P2-Vlan-interface6] mpls enable
[P2-Vlan-interface6] mpls ldp enable
[P2-Vlan-interface6] quit
```

Assign an IP address to VLAN-interface 5.

```
[P2] vlan 5
[P2-vlan5] port gigabitethernet 1/0/2
[P2-vlan5] quit
[P2] interface vlan-interface 5
[P2-Vlan-interface5] ip address 10.4.1.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 5.

```
[P2-Vlan-interface5] pim sm
[P2-Vlan-interface5] mpls enable
[P2-Vlan-interface5] mpls ldp enable
[P2-Vlan-interface5] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[P2] interface loopback 1
[P2-LoopBack1] ip address 6.6.6.6 32
[P2-LoopBack1] pim sm
[P2-LoopBack1] quit
```

Configure OSPF.

```
[P2] ospf 1
[P2-ospf-1] area 0.0.0.0
[P2-ospf-1-area-0.0.0.0] network 6.6.6.6 0.0.0.0
[P2-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[P2-ospf-1-area-0.0.0.0] network 10.5.1.0 0.0.0.255
[P2-ospf-1-area-0.0.0.0] quit
[P2-ospf-1] quit
```

7. Configure CE a1:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEa1
```

```
[CEa1] multicast routing
[CEa1-mrib] quit
```

Assign an IP address to VLAN-interface 21, and enable PIM-SM on the interface.

```
[CEa1] vlan 21
[CEa1-vlan21] port gigabitethernet 1/0/1
[CEa1-vlan21] quit
[CEa1] interface vlan-interface 21
[CEa1-Vlan-interface21] ip address 12.1.1.1 24
[CEa1-Vlan-interface21] pim sm
[CEa1-Vlan-interface21] quit
```

Assign an IP address to VLAN-interface 11, and enable PIM-SM on the interface.

```
[CEa1] vlan 11
[CEa1-vlan11] port gigabitethernet 1/0/2
[CEa1-vlan11] quit
[CEa1] interface vlan-interface 11
[CEa1-Vlan-interface11] ip address 11.1.1.2 24
[CEa1-Vlan-interface11] pim sm
[CEa1-Vlan-interface11] quit
```

Configure VLAN-interface 11 as a C-BSR and a C-RP.

```
[CEa1] pim
[CEa1-pim] c-bsr 11.1.1.2
[CEa1-pim] c-rp 11.1.1.2
[CEa1-pim] quit
```

Configure OSPF.

```
[CEa1] ospf 1
[CEa1-ospf-1] area 0.0.0.0
[CEa1-ospf-1-area-0.0.0.0] network 12.1.1.0 0.0.0.255
[CEa1-ospf-1-area-0.0.0.0] network 11.1.1.0 0.0.0.255
[CEa1-ospf-1-area-0.0.0.0] quit
[CEa1-ospf-1] quit
```

8. Configure CE b1:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEb1
[CEb1] multicast routing
[CEb1-mrib] quit
```

Assign an IP address to VLAN-interface 22, and enable PIM-SM on the interface.

```
[CEb1] vlan 22
[CEb1-vlan22] port gigabitethernet 1/0/1
[CEb1-vlan22] quit
[CEb1] interface vlan-interface 22
[CEb1-Vlan-interface22] ip address 12.2.1.1 24
[CEb1-Vlan-interface22] pim sm
[CEb1-Vlan-interface22] quit
```

Assign an IP address to VLAN-interface 12, and enable PIM-SM on the interface.

```
[CEb1] vlan 12
[CEb1-vlan12] port gigabitethernet 1/0/2
[CEb1-vlan12] quit
```

```
[CEb1] interface vlan-interface 12
[CEb1-Vlan-interface12] ip address 11.2.1.2 24
[CEb1-Vlan-interface12] pim sm
[CEb1-Vlan-interface12] quit
```

Configure VLAN-interface 12 as a C-BSR and a C-RP.

```
[CEb1] pim
[CEb1-pim] c-bsr 11.2.1.2
[CEb1-pim] c-rp 11.2.1.2
[CEb1-pim] quit
```

Configure OSPF.

```
[CEb1] ospf 1
[CEb1-ospf-1] area 0.0.0.0
[CEb1-ospf-1-area-0.0.0.0] network 12.2.1.0 0.0.0.255
[CEb1-ospf-1-area-0.0.0.0] network 11.2.1.0 0.0.0.255
[CEb1-ospf-1-area-0.0.0.0] quit
[CEb1-ospf-1] quit
```

9. Configure CE a2:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEa2
[CEa2] multicast routing
[CEa2-mrib] quit
```

Assign an IP address to VLAN-interface 23, and enable IGMP on the interface.

```
[CEa2] vlan 23
[CEa2-vlan23] port gigabitethernet 1/0/1
[CEa2-vlan23] quit
[CEa2] interface vlan-interface 23
[CEa2-Vlan-interface23] ip address 12.3.1.1 24
[CEa2-Vlan-interface23] igmp enable
[CEa2-Vlan-interface23] quit
```

Assign an IP address to VLAN-interface 13, and enable PIM-SM on the interface.

```
[CEa2] vlan 13
[CEa2-vlan13] port gigabitethernet 1/0/2
[CEa2-vlan13] quit
[CEa2] interface vlan-interface 13
[CEa2-Vlan-interface13] ip address 11.3.1.2 24
[CEa2-Vlan-interface13] pim sm
[CEa2-Vlan-interface13] quit
```

Configure OSPF.

```
[CEa2] ospf 1
[CEa2-ospf-1] area 0.0.0.0
[CEa2-ospf-1-area-0.0.0.0] network 12.3.1.0 0.0.0.255
[CEa2-ospf-1-area-0.0.0.0] network 11.3.1.0 0.0.0.255
[CEa2-ospf-1-area-0.0.0.0] quit
[CEa2-ospf-1] quit
```

10. Configure CE b2:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEb2
[CEb2] multicast routing
[CEb2-mrib] quit
```

Assign an IP address to VLAN-interface 24, and enable IGMP on the interface.

```
[CEb2] vlan 24
[CEb2-vlan24] port gigabitethernet 1/0/1
[CEb2-vlan24] quit
[CEb2] interface vlan-interface 24
[CEb2-Vlan-interface24] ip address 12.4.1.1 24
[CEb2-Vlan-interface24] igmp enable
[CEb2-Vlan-interface24] quit
```

Assign an IP address to VLAN-interface 14, and enable PIM-SM on the interface.

```
[CEb2] vlan 14
[CEb2-vlan14] port gigabitethernet 1/0/2
[CEb2-vlan14] quit
[CEb2] interface vlan-interface 14
[CEb2-Vlan-interface14] ip address 11.4.1.2 24
[CEb2-Vlan-interface14] pim sm
[CEb2-Vlan-interface14] quit
```

Configure OSPF.

```
[CEb2] ospf 1
[CEb2-ospf-1] area 0.0.0.0
[CEb2-ospf-1-area-0.0.0.0] network 12.4.1.0 0.0.0.255
[CEb2-ospf-1-area-0.0.0.0] network 11.4.1.0 0.0.0.255
[CEb2-ospf-1-area-0.0.0.0] quit
[CEb2-ospf-1] quit
```

Verifying the configuration

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 1.

```
[PE1] display multicast-vpn default-group local
MVPN local default-group information:
 Group address    Source address   Interface     VPN instance
 232.1.1.1        1.1.1.1          MTunnel0      a
 232.3.3.3        1.1.1.1          MTunnel1      b
```

Display information about the remote default group for IPv4 multicast transmission in each VPN instance on PE 1.

```
[PE1] display multicast-vpn default-group remote
MD remote default-group information:
 Group address   Source address  Next hop         VPN instance
 232.1.1.1       4.4.4.4         2.2.2.2          a
 232.3.3.3       4.4.4.4         2.2.2.2          b
```

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 4.

```
[PE4] display multicast-vpn default-group local
MVPN local default-group information:
 Group address    Source address   Interface     VPN instance
 232.1.1.1        4.4.4.4          MTunnel0      a
```

```
 232.3.3.3        4.4.4.4          MTunnel1      b
```

Display information about the remote default group for IPv4 multicast transmission in each VPN instance on PE 4.

```
[PE4] display multicast-vpn default-group remote
MD remote default-group information:
 Group address   Source address  Next hop         VPN instance
 232.1.1.1       1.1.1.1         3.3.3.3          a
 232.3.3.3       1.1.1.1         3.3.3.3          b
```

Configuration files

- CE a1:

```
#
 sysname CEa1
#
ospf 1
 area 0.0.0.0
  network 11.1.1.0 0.0.0.255
  network 12.1.1.0 0.0.0.255
#
vlan 11
#
vlan 21
#
interface Vlan-interface11
 ip address 11.1.1.2 255.255.255.0
 pim sm
#
interface Vlan-interface21
 ip address 12.1.1.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 21
#
interface GigabitEthernet1/0/2
 port access vlan 11
#
multicast routing
#
pim
 c-bsr 11.1.1.2
 c-rp 11.1.1.2
#
return
```

- CE a2:

```
#
 sysname CEa2
#
ospf 1
```

```
 area 0.0.0.0
  network 11.3.1.0 0.0.0.255
  network 12.3.1.0 0.0.0.255
#
vlan 13
#
vlan 23
#
interface Vlan-interface13
 ip address 11.3.1.2 255.255.255.0
 pim sm
#
interface Vlan-interface23
 ip address 12.3.1.1 255.255.255.0
 igmp enable
#
interface GigabitEthernet1/0/1
 port access vlan 23
#
interface GigabitEthernet1/0/2
 port access vlan 13
#
multicast routing
#
return
```

- CE b1:

```
#
 sysname CEb1
#
ospf 1
 area 0.0.0.0
  network 11.2.1.0 0.0.0.255
  network 12.2.1.0 0.0.0.255
#
vlan 12
#
vlan 22
#
interface Vlan-interface12
 ip address 11.2.1.2 255.255.255.0
 pim sm
#
interface Vlan-interface22
 ip address 12.2.1.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 22
```

```
#
interface GigabitEthernet1/0/2
 port access vlan 12
#
multicast routing
#
pim
 c-bsr 11.2.1.2
 c-rp 11.2.1.2
#
return
```

- CE b2:

```
#
 sysname CEb2
#
ospf 1
 area 0.0.0.0
  network 11.4.1.0 0.0.0.255
  network 12.4.1.0 0.0.0.255
#
vlan 14
#
vlan 24
#
interface Vlan-interface14
 ip address 11.4.1.2 255.255.255.0
 pim sm
#
interface Vlan-interface24
 ip address 12.4.1.1 255.255.255.0
 igmp enable
#
interface GigabitEthernet1/0/1
 port access vlan 24
#
interface GigabitEthernet1/0/2
 port access vlan 14
#
multicast routing
#
return
```

- P 1:

```
#
 sysname P1
#
ospf 1
 area 0.0.0.0
  network 5.5.5.5 0.0.0.0
```

```
  network 10.1.1.0 0.0.0.255
  network 10.2.1.0 0.0.0.255
#
 mpls lsr-id 5.5.5.5
#
vlan 2 to 3
#
mpls ldp
#
interface LoopBack1
 ip address 5.5.5.5 255.255.255.255
 pim sm
#
interface Vlan-interface2
 ip address 10.1.1.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface3
 ip address 10.2.1.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port access vlan 3
#
multicast routing
#
return
```

- P 2:

```
#
 sysname P2
#
ospf 1
 area 0.0.0.0
  network 6.6.6.6 0.0.0.0
  network 10.4.1.0 0.0.0.255
  network 10.5.1.0 0.0.0.255
#
 mpls lsr-id 6.6.6.6
#
vlan 5 to 6
#
```

```
mpls ldp
#
interface LoopBack1
 ip address 6.6.6.6 255.255.255.255
 pim sm
#
interface Vlan-interface5
 ip address 10.4.1.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface6
 ip address 10.5.1.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port access vlan 6
#
interface GigabitEthernet1/0/2
 port access vlan 5
#
multicast routing
#
return
```

- PE 1:

```
#
 sysname PE1
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
ip vpn-instance b
 route-distinguisher 200:1
 vpn-target 200:1 import-extcommunity
 vpn-target 200:1 export-extcommunity
#
 router id 1.1.1.1
#
ospf 1
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 10.1.1.0 0.0.0.255
#
```

```
ospf 2 vpn-instance a
 import-route bgp
 area 0.0.0.0
  network 11.1.1.0 0.0.0.255
#
ospf 3 vpn-instance b
 import-route bgp
 area 0.0.0.0
  network 11.2.1.0 0.0.0.255
#
 mpls lsr-id 1.1.1.1
#
vlan 2
#
vlan 11 to 12
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.1 255.255.255.255
 pim sm
#
interface Vlan-interface2
 ip address 10.1.1.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface11
 ip binding vpn-instance a
 ip address 11.1.1.1 255.255.255.0
 pim sm
#
interface Vlan-interface12
 ip binding vpn-instance b
 ip address 11.2.1.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port access vlan 11
#
interface GigabitEthernet1/0/3
 port access vlan 12
#
bgp 100
```

```
 peer 2.2.2.2 as-number 100
 peer 2.2.2.2 connect-interface LoopBack1
 #
 address-family vpnv4
  peer 2.2.2.2 enable
 #
 address-family ipv4 mdt
  peer 2.2.2.2 enable
 #
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 2
 #
 ip vpn-instance b
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 3
#
multicast routing
#
multicast routing vpn-instance a
 rpf proxy vector
#
multicast routing vpn-instance b
 rpf proxy vector
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack1
  default-group 232.1.1.1
  data-group 232.2.2.0 255.255.255.240
#
multicast-vpn vpn-instance b mode mdt
 address-family ipv4
  source LoopBack1
  default-group 232.3.3.3
  data-group 232.4.4.0 255.255.255.240
#
return
```

- PE 2:

```
#
 sysname PE2
#
 router id 2.2.2.2
#
```

```
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 10.2.1.0 0.0.0.255
#
 mpls lsr-id 2.2.2.2
#
vlan 3 to 4
#
mpls ldp
#
interface LoopBack1
 ip address 2.2.2.2 255.255.255.255
 pim sm
#
interface Vlan-interface3
 ip address 10.2.1.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface4
 ip address 10.3.1.1 255.255.255.0
 pim sm
 mpls enable
#
interface GigabitEthernet1/0/1
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port access vlan 4
#
bgp 100
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 connect-interface LoopBack1
 peer 10.3.1.2 as-number 200
 #
 address-family vpnv4
  undo policy vpn-target
  peer 1.1.1.1 enable
  peer 10.3.1.2 enable
 #
 address-family ipv4 mdt
  peer 1.1.1.1 enable
  peer 10.3.1.2 enable
#
multicast routing
#
```

```
return
```

- PE 3:

```
#
 sysname PE3
#
 router id 3.3.3.3
#
ospf 1
 area 0.0.0.0
  network 3.3.3.3 0.0.0.0
  network 10.4.1.0 0.0.0.255
#
 mpls lsr-id 3.3.3.3
#
vlan 4 to 5
#
mpls ldp
#
interface LoopBack1
 ip address 3.3.3.3 255.255.255.255
 pim sm
#
interface Vlan-interface4
 ip address 10.3.1.2 255.255.255.0
 pim sm
 mpls enable
#
interface Vlan-interface5
 ip address 10.4.1.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port access vlan 5
#
interface GigabitEthernet1/0/2
 port access vlan 4
#
bgp 200
 peer 4.4.4.4 as-number 200
 peer 4.4.4.4 connect-interface LoopBack1
 peer 10.3.1.1 as-number 100
 #
 address-family vpnv4
  undo policy vpn-target
  peer 4.4.4.4 enable
  peer 10.3.1.1 enable
```

```
 #
 address-family ipv4 mdt
  peer 4.4.4.4 enable
  peer 10.3.1.1 enable
#
multicast routing
#
return
```

- PE 4:

```
#
 sysname PE4
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
ip vpn-instance b
 route-distinguisher 200:1
 vpn-target 200:1 import-extcommunity
 vpn-target 200:1 export-extcommunity
#
 router id 4.4.4.4
#
ospf 1
 area 0.0.0.0
  network 4.4.4.4 0.0.0.0
  network 10.5.1.0 0.0.0.255
#
ospf 2 vpn-instance a
 import-route bgp
 area 0.0.0.0
  network 11.3.1.0 0.0.0.255
#
ospf 3 vpn-instance b
 import-route bgp
 area 0.0.0.0
  network 11.4.1.0 0.0.0.255
#
 mpls lsr-id 4.4.4.4
#
vlan 6
#
vlan 13 to 14
#
mpls ldp
#
interface LoopBack1
```

```
 ip address 4.4.4.4 255.255.255.255
 pim sm
#
interface Vlan-interface6
 ip address 10.5.1.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface13
 ip binding vpn-instance a
 ip address 11.3.1.1 255.255.255.0
 pim sm
#
interface Vlan-interface14
 ip binding vpn-instance b
 ip address 11.4.1.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 6
#
interface GigabitEthernet1/0/2
 port access vlan 13
#
bgp 200
 peer 3.3.3.3 as-number 200
 peer 3.3.3.3 connect-interface LoopBack1
 #
 address-family vpnv4
  peer 3.3.3.3 enable
 #
 address-family ipv4 mdt
  peer 3.3.3.3 enable
 #
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 2
 #
 ip vpn-instance b
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 3
#
multicast routing
```

```
#
multicast routing vpn-instance a
 rpf proxy vector
#
multicast routing vpn-instance b
 rpf proxy vector
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack1
  default-group 232.1.1.1
  data-group 232.2.2.0 255.255.255.240
#
multicast-vpn vpn-instance b mode mdt
 address-family ipv4
  source LoopBack1
  default-group 232.3.3.3
  data-group 232.4.4.0 255.255.255.240
#
return
```

Example: Configuring MDT-based MVPN inter-AS option C

Network configuration

As shown in Figure 17, configure inter-AS option C MDT-based MVPN to meet the following requirements:

Item	Network configuration
Multicast sources and receivers	• In VPN instance **a**, S 1 is a multicast source, and R 2 is a receiver. • In VPN instance **b**, S 2 is a multicast source, and R 1 is a receiver. • For VPN instance **a**, the default group is 239.1.1.1, and the data group range is 225.1.1.0 to 225.1.1.15. • For VPN instance **b**, the default group is 239.4.4.4, and the data group range is 225.4.4.0 to 225.4.4.15.
VPN instances to which PEs belong	• PE 1: VLAN-interface 11 belongs to VPN instance **b**. VLAN-interface 12 belongs to VPN instance **a**. VLAN-interface 2 and Loopback 1 belong to the public network instance. • PE 2: VLAN-interface 2, VLAN-interface 3, Loopback 1, and Loopback 2 belong to the public network instance. • PE 3: VLAN-interface 3, VLAN-interface 4, Loopback 1, and Loopback 2 belong to the public network instance. • PE 4: VLAN-interface 13 belongs to VPN instance **a**. VLAN-interface 14 belongs to VPN instance **b**. VLAN-interface 4 and Loopback 1 belong to the public network instance.
Unicast routing protocols and MPLS	• Configure OSPF separately in AS 100 and AS 200, and configure OSPF between the PEs and CEs. • Establish BGP peer connections between PE 1, PE 2, PE 3, and PE 4 on their respective Loopback 1. • Configure MPLS separately in AS 100 and AS 200.
IP multicast routing	• Enable IP multicast routing on the public network on PE 1, PE 2, PE 3, and PE 4. • Enable IP multicast routing for VPN instance **a** on PE 1 and PE 4.

Item	Network configuration
	• Enable IP multicast routing for VPN instance **b** on PE 1 and PE 4. • Enable IP multicast routing on CE a1, CE a2, CE b1, and CE b2.
IGMPv2	• Enable IGMPv2 on VLAN-interface 30 of CE a2. • Enable IGMPv2 on VLAN-interface 40 of CE b2.
PIM-SM	Enable PIM-SM on the public network and for VPN instances **a** and **b**: • Enable PIM-SM on all public network interfaces of PE 2 and PE 3. • Enable PIM-SM on all public and private network interfaces of PE 1 and PE 4. • Enable PIM-SM on all interfaces that do not have attached receiver hosts of CE a1, CE a2, CE b1, and CE b2. • Configure Loopback 2 of PE 2 and PE 3 as a C-BSR and a C-RP for their own AS to provide services for all multicast groups. • Configure Loopback 0 of CE a1 as a C-BSR and a C-RP for VPN instance **a** to provide services for all multicast groups. • Configure Loopback 0 of CE b1 as a C-BSR and a C-RP for VPN instance **b** to provide services for all multicast groups.
MSDP	Establish an MSDP peering relationship between PE 2 and PE 3 on their Loopback 1.

Figure 17 Network diagram

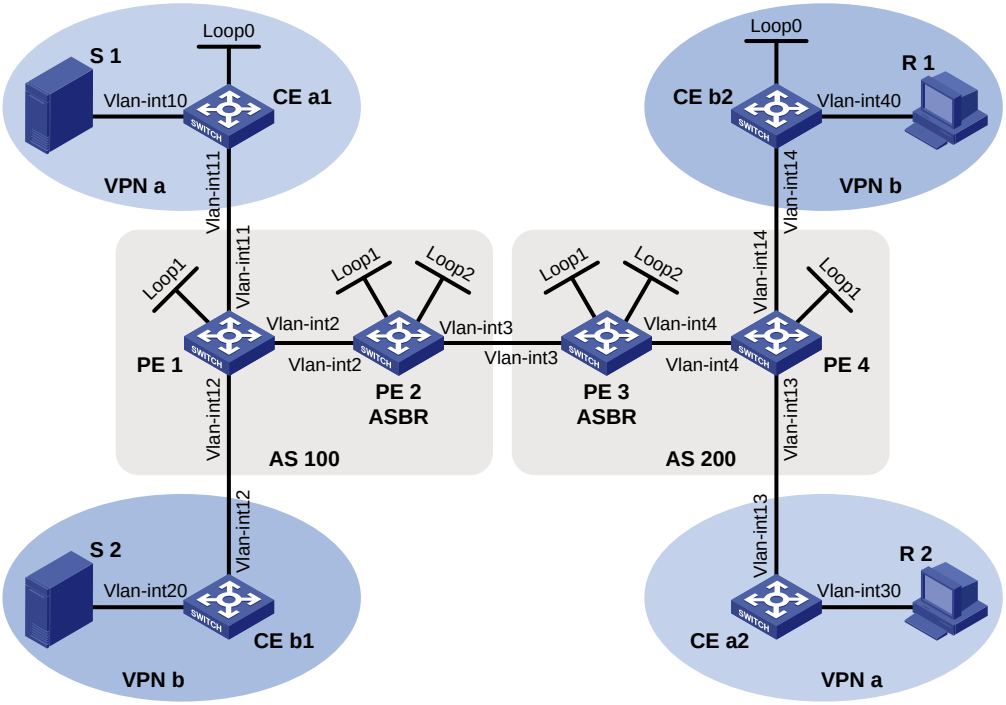


Table 4 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
S 1	—	10.11.5.2/24	R 1	—	10.11.8.2/24
S 2	—	10.11.6.2/24	R 2	—	10.11.7.2/24
PE 1	Vlan-int2	10.10.1.1/24	PE 3	Vlan-int4	10.10.2.1/24
PE 1	Vlan-int11	10.11.1.1/24	PE 3	Vlan-int3	192.168.1.2/24
PE 1	Vlan-int12	10.11.2.1/24	PE 3	Loop1	1.1.1.3/32

Device	Interface	IP address	Device	Interface	IP address
PE 1	Loop1	1.1.1.1/32	PE 3	Loop2	22.22.22.22/32
PE 2	Vlan-int2	10.10.1.2/24	PE 4	Vlan-int4	10.10.2.2/24
PE 2	Vlan-int3	192.168.1.1/24	PE 4	Vlan-int13	10.11.3.1/24
PE 2	Loop1	1.1.1.2/32	PE 4	Vlan-int14	10.11.4.1/32
PE 2	Loop2	11.11.11.11/32	PE 4	Loop1	1.1.1.4/32
CE a1	Vlan-int10	10.11.5.1/24	CE b1	Vlan-int20	10.11.6.1/24
CE a1	Vlan-int11	10.11.1.2/24	CE b1	Vlan-int12	10.11.2.2/24
CE a1	Loop0	2.2.2.2/32	CE b2	Vlan-int40	10.11.8.1/24
CE a2	Vlan-int30	10.11.7.1/24	CE b2	Vlan-int14	10.11.4.2/24
CE a2	Vlan-int13	10.11.3.2/24	CE b2	Loop0	3.3.3.3/32

Procedure

1. Configure PE 1:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE1
[PE1] router id 1.1.1.1
[PE1] multicast routing
[PE1-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE1] mpls lsr-id 1.1.1.1
[PE1] mpls ldp
[PE1-ldp] quit
```

Create a VPN instance named **a**, and configure an RD and route targets for the VPN instance.

```
[PE1] ip vpn-instance a
[PE1-vpn-instance-a] route-distinguisher 100:1
[PE1-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE1-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE1-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE1] multicast routing vpn-instance a
[PE1-mrib-a] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.1 32
[PE1-LoopBack1] pim sm
[PE1-LoopBack1] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE1] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 address family for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a-ipv4] default-group 239.1.1.1
[PE1-mvpn-vpn-instance-a-ipv4] source loopback 1
[PE1-mvpn-vpn-instance-a-ipv4] data-group 225.1.1.0 28
[PE1-mvpn-vpn-instance-a-ipv4] quit
[PE1-mvpn-vpn-instance-a] quit
```

Create a VPN instance named **b**, and configure an RD and route targets for the VPN instance.

```
[PE1] ip vpn-instance b
[PE1-vpn-instance-b] route-distinguisher 200:1
[PE1-vpn-instance-b] vpn-target 200:1 export-extcommunity
[PE1-vpn-instance-b] vpn-target 200:1 import-extcommunity
[PE1-vpn-instance-b] quit
```

Enable IP multicast routing for VPN instance **b**.

```
[PE1] multicast routing vpn-instance b
[PE1-mrib-b] quit
```

Create an MDT-based MVPN for VPN instance **b**.

```
[PE1] multicast-vpn vpn-instance b mode mdt
```

Create an MVPN IPv4 address family for VPN instance **b**.

```
[PE1-mvpn-vpn-instance-b] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **b**.

```
[PE1-mvpn-vpn-instance-b-ipv4] default-group 239.4.4.4
[PE1-mvpn-vpn-instance-b-ipv4] source loopback 1
[PE1-mvpn-vpn-instance-b-ipv4] data-group 225.4.4.0 28
[PE1-mvpn-vpn-instance-b-ipv4] quit
[PE1-mvpn-vpn-instance-b] quit
```

Assign an IP address to VLAN-interface 2.

```
[PE1] vlan 2
[PE1-vlan2] port gigabitethernet 1/0/1
[PE1-vlan2] quit
[PE1] interface vlan-interface 2
[PE1-Vlan-interface2] ip address 10.10.1.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 2.

```
[PE1-Vlan-interface2] pim sm
[PE1-Vlan-interface2] mpls enable
[PE1-Vlan-interface2] mpls ldp enable
[PE1-Vlan-interface2] quit
```

Associate VLAN-interface 11 with VPN instance **a**.

```
[PE1] vlan 11
[PE1-vlan11] port gigabitethernet 1/0/2
[PE1-vlan11] quit
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip binding vpn-instance a
```

Assign an IP address to VLAN-interface 11, and enable PIM-SM on the interface.

```
[PE1-Vlan-interface11] ip address 10.11.1.1 24
[PE1-Vlan-interface11] pim sm
[PE1-Vlan-interface11] quit
```

Associate VLAN-interface 12 with VPN instance **b**.

```
[PE1] vlan 12
```

```
[PE1-vlan12] port gigabitethernet 1/0/3
[PE1-vlan12] quit
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance b
```

Assign an IP address to VLAN-interface 12, and enable PIM-SM on the interface.

```
[PE1-Vlan-interface12] ip address 10.11.2.1 24
[PE1-Vlan-interface12] pim sm
[PE1-Vlan-interface12] quit
```

Configure BGP.

```
[PE1] bgp 100
[PE1-bgp-default] group pe1-pe2 internal
[PE1-bgp-default] peer pe1-pe2 connect-interface loopback 1
[PE1-bgp-default] peer 1.1.1.2 group pe1-pe2
[PE1-bgp-default] group pe1-pe4 external
[PE1-bgp-default] peer pe1-pe4 as-number 200
[PE1-bgp-default] peer pe1-pe4 ebgp-max-hop 255
[PE1-bgp-default] peer pe1-pe4 connect-interface loopback 1
[PE1-bgp-default] peer 1.1.1.4 group pe1-pe4
[PE1–bgp-default] ip vpn-instance a
[PE1-bgp-default-a] address-family ipv4
[PE1-bgp-default-ipv4-a] import-route ospf 2
[PE1-bgp-default-ipv4-a] import-route direct
[PE1-bgp-default-ipv4-a] quit
[PE1-bgp-default-a] quit
[PE1–bgp-default] ip vpn-instance b
[PE1-bgp-default-b] address-family ipv4
[PE1-bgp-default-ipv4-b] import-route ospf 3
[PE1-bgp-default-ipv4-b] import-route direct
[PE1-bgp-default-ipv4-b] quit
[PE1-bgp-default-b] quit
[PE1–bgp-default] address-family ipv4
[PE1-bgp-default-ipv4] peer pe1-pe2 enable
[PE1-bgp-default-ipv4] peer pe1-pe2 label-route-capability
[PE1-bgp-default-ipv4] quit
[PE1–bgp-default] address-family vpnv4
[PE1–bgp-default-vpnv4] peer pe1-pe4 enable
[PE1–bgp-default-vpnv4] quit
[PE1–bgp-default] quit
```

Configure OSPF.

```
[PE1] ospf 1
[PE1-ospf-1] area 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 10.10.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
[PE1] ospf 2 vpn-instance a
[PE1-ospf-2] import-route bgp
[PE1-ospf-2] area 0.0.0.0
```

```
[PE1-ospf-2-area-0.0.0.0] network 10.11.1.0 0.0.0.255
[PE1-ospf-2-area-0.0.0.0] quit
[PE1-ospf-2] quit
[PE1] ospf 3 vpn-instance b
[PE1-ospf-3] import-route bgp
[PE1-ospf-3] area 0.0.0.0
[PE1-ospf-3-area-0.0.0.0] network 10.11.2.0 0.0.0.255
[PE1-ospf-3-area-0.0.0.0] quit
[PE1-ospf-3] quit
```

2. Configure PE 2:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] router id 1.1.1.2
[PE2] multicast routing
[PE2-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE2] mpls lsr-id 1.1.1.2
[PE2] mpls ldp
[PE2-ldp] quit
```

Assign an IP address to VLAN-interface 2.

```
[PE2] vlan 2
[PE2-vlan2] port gigabitethernet 1/0/1
[PE2-vlan2] quit
[PE2] interface vlan-interface 2
[PE2-Vlan-interface2] ip address 10.10.1.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 2.

```
[PE2-Vlan-interface2] pim sm
[PE2-Vlan-interface2] mpls enable
[PE2-Vlan-interface2] mpls ldp enable
[PE2-Vlan-interface2] quit
```

Assign an IP address to VLAN-interface 3.

```
[PE2] vlan 3
[PE2-vlan3] port gigabitethernet 1/0/2
[PE2-vlan3] quit
[PE2] interface vlan-interface 3
[PE2-Vlan-interface3] ip address 192.168.1.1 24
```

Enable PIM-SM and MPLS on VLAN-interface 3.

```
[PE2-Vlan-interface3] pim sm
[PE2-Vlan-interface3] mpls enable
[PE2-Vlan-interface3] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 1.1.1.2 32
[PE2-LoopBack1] pim sm
[PE2-LoopBack1] quit
```

Assign an IP address to Loopback 2, and enable PIM-SM on the interface.

```
[PE2] interface loopback 2
[PE2-LoopBack2] ip address 11.11.11.11 32
[PE2-LoopBack2] pim sm
[PE2-LoopBack2] quit
```

Configure Loopback 2 as a C-BSR and a C-RP.

```
[PE2] pim
[PE2-pim] c-bsr 11.11.11.11
[PE2-pim] c-rp 11.11.11.11
[PE2-pim] quit
```

Configure VLAN-interface 3 as a PIM-SM domain border.

```
[PE2] interface vlan-interface 3
[PE2-Vlan-interface3] pim bsr-boundary
[PE2-Vlan-interface3] quit
```

Establish an MSDP peering relationship.

```
[PE2] msdp
[PE2-msdp] encap-data-enable
[PE2-msdp] peer 1.1.1.3 connect-interface loopback 1
```

Configure a routing policy.

```
[PE2] route-policy map1 permit node 1
[PE2-route-policy-map1-1] apply mpls-label
[PE2-route-policy-map1-1] quit
[PE2] route-policy map2 permit node 1
[PE2-route-policy-map2-1] apply mpls-label
[PE2-route-policy-map2-1] if-match mpls-label
[PE2-route-policy-map2-1] quit
```

Configure a static route.

```
[PE2] ip route-static 1.1.1.3 32 vlan-interface 3 192.168.1.2
```

Configure BGP.

```
[PE2] bgp 100
[PE2-bgp-default] group pe2-pe1 internal
[PE2-bgp-default] peer pe2-pe1 connect-interface loopback 1
[PE2-bgp-default] peer 1.1.1.1 group pe2-pe1
[PE2-bgp-default] group pe2-pe3 external
[PE2-bgp-default] peer pe2-pe3 as-number 200
[PE2-bgp-default] peer 192.168.1.2 group pe2-pe3
[PE2-bgp-default] address-family ipv4
[PE2-bgp-default-ipv4] peer pe2-pe1 enable
[PE2-bgp-default-ipv4] peer pe2-pe1 route-policy map2 export
[PE2-bgp-default-ipv4] peer pe2-pe1 label-route-capability
[PE2-bgp-default-ipv4] peer pe2-pe3 enable
[PE2-bgp-default-ipv4] peer pe2-pe3 route-policy map1 export
[PE2-bgp-default-ipv4] peer pe2-pe3 label-route-capability
[PE2-bgp-default-ipv4] import-route ospf 1
[PE2-bgp-default-ipv4] quit
[PE2–bgp-default] quit
```

Configure OSPF.

```
[PE2] ospf 1
[PE2-ospf-1] area 0.0.0.0
```

```
[PE2-ospf-1-area-0.0.0.0] network 1.1.1.2 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 11.11.11.11 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 10.10.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

3. Configure PE 3:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE3
[PE3] router id 1.1.1.3
[PE3] multicast routing
[PE3-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE3] mpls lsr-id 1.1.1.3
[PE3] mpls ldp
[PE3-ldp] quit
```

Assign an IP address to VLAN-interface 4.

```
[PE3] vlan 4
[PE3-vlan4] port gigabitethernet 1/0/1
[PE3-vlan4] quit
[PE3] interface vlan-interface 4
[PE3-Vlan-interface4] ip address 10.10.2.1 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 4.

```
[PE3-Vlan-interface4] pim sm
[PE3-Vlan-interface4] mpls enable
[PE3-Vlan-interface4] mpls ldp enable
[PE3-Vlan-interface4] quit
```

Assign an IP address to VLAN-interface 3.

```
[PE3] vlan 3
[PE3-vlan3] port gigabitethernet 1/0/2
[PE3-vlan3] quit
[PE3] interface vlan-interface 3
[PE3-Vlan-interface3] ip address 192.168.1.2 24
```

Enable PIM-SM and MPLS on VLAN-interface 3.

```
[PE3-Vlan-interface3] pim sm
[PE3-Vlan-interface3] mpls enable
[PE3-Vlan-interface3] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE3] interface loopback 1
[PE3-LoopBack1] ip address 1.1.1.3 32
[PE3-LoopBack1] pim sm
[PE3-LoopBack1] quit
```

Assign an IP address to Loopback 2, and enable PIM-SM on the interface.

```
[PE3] interface loopback 2
[PE3-LoopBack2] ip address 22.22.22.22 32
[PE3-LoopBack2] pim sm
[PE3-LoopBack2] quit
```

Configure Loopback 2 as a C-BSR and a C-RP.

```
[PE3] pim
[PE3-pim] c-bsr 22.22.22.22
[PE3-pim] c-rp 22.22.22.22
[PE3-pim] quit
```

Configure VLAN-interface 3 as a PIM-SM domain border.

```
[PE3] interface vlan-interface 3
[PE3-Vlan-interface3] pim bsr-boundary
[PE3-Vlan-interface3] quit
```

Establish an MSDP peering relationship.

```
[PE3] msdp
[PE3-msdp] encap-data-enable
[PE3-msdp] peer 1.1.1.2 connect-interface loopback 1
```

Configure a static route.

```
[PE3] ip route-static 1.1.1.2 32 vlan-interface 3 192.168.1.1
```

Configure a routing policy.

```
[PE3] route-policy map1 permit node 1
[PE3-route-policy-map1-1] apply mpls-label
[PE3-route-policy-map1-1] quit
[PE3] route-policy map2 permit node 1
[PE3-route-policy-map2-1] apply mpls-label
[PE3-route-policy-map2-1] if-match mpls-label
[PE3-route-policy-map2-1] quit
```

Configure BGP.

```
[PE3] bgp 200
[PE3-bgp-default] group pe3-pe4 internal
[PE3-bgp-default] peer pe3-pe4 connect-interface loopback 1
[PE3-bgp-default] peer 1.1.1.4 group pe3-pe4
[PE3-bgp-default] group pe3-pe2 external
[PE3-bgp-default] peer pe3-pe2 as-number 100
[PE3-bgp-default] peer 192.168.1.1 group pe3-pe2
[PE3-bgp-default] address-family ipv4
[PE3-bgp-default-ipv4] peer pe3-pe4 enable
[PE3-bgp-default-ipv4] peer pe3-pe4 route-policy map2 export
[PE3-bgp-default-ipv4] peer pe3-pe4 label-route-capability
[PE3-bgp-default-ipv4] peer pe3-pe2 enable
[PE3-bgp-default-ipv4] peer pe3-pe2 route-policy map1 export
[PE3-bgp-default-ipv4] peer pe3-pe2 label-route-capability
[PE3-bgp-default-ipv4] import-route ospf 1
[PE3-bgp-default-ipv4] quit
[PE3–bgp-default] quit
```

Configure OSPF.

```
[PE3] ospf 1
[PE3-ospf-1] area 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 1.1.1.3 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 22.22.22.22 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 10.10.2.0 0.0.0.255
[PE3-ospf-1-area-0.0.0.0] quit
```

```
[PE3-ospf-1] quit
```

4. Configure PE 4:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE4
[PE4] router id 1.1.1.4
[PE4] multicast routing
[PE4-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE4] mpls lsr-id 1.1.1.4
[PE4] mpls ldp
[PE4-ldp] quit
```

Create a VPN instance named **a**, and configure an RD and route targets for the VPN instance.

```
[PE4] ip vpn-instance a
[PE4-vpn-instance-a] route-distinguisher 100:1
[PE4-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE4-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE4-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE4] multicast routing vpn-instance a
[PE4-mrib-a] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE4] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 address family for VPN instance **a**.

```
[PE4-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE4-mvpn-vpn-instance-a-ipv4] default-group 239.1.1.1
[PE4-mvpn-vpn-instance-a-ipv4] source loopback 1
[PE4-mvpn-vpn-instance-a-ipv4] data-group 225.1.1.0 28
[PE4-mvpn-vpn-instance-a-ipv4] quit
[PE4-mvpn-vpn-instance-a] quit
```

Create a VPN instance named **b**, and configure an RD and route targets for the VPN instance.

```
[PE4] ip vpn-instance b
[PE4-vpn-instance-b] route-distinguisher 200:1
[PE4-vpn-instance-b] vpn-target 200:1 export-extcommunity
[PE4-vpn-instance-b] vpn-target 200:1 import-extcommunity
[PE4-vpn-instance-b] quit
```

Enable IP multicast routing for VPN instance **b**.

```
[PE4] multicast routing vpn-instance b
[PE4-mrib-b] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[PE4] interface loopback 1
[PE4-LoopBack1] ip address 1.1.1.4 32
[PE4-LoopBack1] pim sm
[PE4-LoopBack1] quit
```

Create an MDT-based MVPN for VPN instance **b**.

```
[PE4] multicast-vpn vpn-instance b mode mdt
```

Create an MVPN IPv4 address family for VPN instance **b**.

```
[PE4-mvpn-vpn-instance-b] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **b**.

```
[PE4-mvpn-vpn-instance-b-ipv4] default-group 239.4.4.4
[PE4-mvpn-vpn-instance-b-ipv4] source loopback 1
[PE4-mvpn-vpn-instance-b-ipv4] data-group 225.4.4.0 28
[PE4-mvpn-vpn-instance-b-ipv4] quit
[PE4-mvpn-vpn-instance-b] quit
```

Assign an IP address to VLAN-interface 4.

```
[PE4] vlan 4
[PE4-vlan4] port gigabitethernet 1/0/1
[PE4-vlan4] quit
[PE4] interface vlan-interface 4
[PE4-Vlan-interface4] ip address 10.10.2.2 24
```

Enable PIM-SM, MPLS, and IPv4 LDP on VLAN-interface 4.

```
[PE4-Vlan-interface4] pim sm
[PE4-Vlan-interface4] mpls enable
[PE4-Vlan-interface4] mpls ldp enable
[PE4-Vlan-interface4] quit
```

Associate VLAN-interface 13 with VPN instance **a**.

```
[PE4] vlan 13
[PE4-vlan13] port gigabitethernet 1/0/2
[PE4-vlan13] quit
[PE4] interface vlan-interface 13
[PE4-Vlan-interface13] ip binding vpn-instance a
```

Assign an IP address to VLAN-interface 13, and enable PIM-SM on the interface.

```
[PE4-Vlan-interface13] ip address 10.11.3.1 24
[PE4-Vlan-interface13] pim sm
[PE4-Vlan-interface13] quit
```

Associate VLAN-interface 14 with VPN instance **b**.

```
[PE4] vlan 14
[PE4-vlan14] port gigabitethernet 1/0/3
[PE4-vlan14] quit
[PE4] interface vlan-interface 14
[PE4-Vlan-interface14] ip binding vpn-instance b
```

Assign an IP address to VLAN-interface 14, and enable PIM-SM on the interface.

```
[PE4-Vlan-interface14] ip address 10.11.4.1 24
[PE4-Vlan-interface14] pim sm
[PE4-Vlan-interface14] quit
```

Configure BGP.

```
[PE4] bgp 200
[PE4-bgp-default] group pe4-pe3 internal
[PE4-bgp-default] peer pe4-pe3 connect-interface loopback 1
[PE4-bgp-default] peer 1.1.1.3 group pe4-pe3
[PE4-bgp-default] group pe4-pe1 external
[PE4-bgp-default] peer pe4-pe1 as-number 100
```

```
[PE4-bgp-default] peer pe4-pe1 ebgp-max-hop 255
[PE4-bgp-default] peer pe4-pe1 connect-interface loopback 1
[PE4-bgp-default] peer 1.1.1.1 group pe4-pe1
[PE4–bgp-default] ip vpn-instance a
[PE4-bgp-default-a] address-family ipv4
[PE4-bgp-default-ipv4-a] import-route ospf 2
[PE4-bgp-default-ipv4-a] import-route direct
[PE4-bgp-default-ipv4-a] quit
[PE4-bgp-default-a] quit
[PE4–bgp-default] ip vpn-instance b
[PE4-bgp-default-b] address-family ipv4
[PE4-bgp-default-ipv4-b] import-route ospf 3
[PE4-bgp-default-ipv4-b] import-route direct
[PE4-bgp-default-ipv4-b] quit
[PE4-bgp-default-b] quit
[PE4–bgp-default] address-family ipv4
[PE4-bgp-default-ipv4] peer pe4-pe3 enable
[PE4-bgp-default-ipv4] peer pe4-pe3 label-route-capability
[PE4-bgp-default-ipv4] quit
[PE4–bgp-default] address-family vpnv4
[PE4–bgp-default-vpnv4] peer pe4-pe1 enable
[PE4–bgp-default-vpnv4] quit
[PE4–bgp-default] quit
```

Configure OSPF.

```
[PE4] ospf 1
[PE4-ospf-1] area 0.0.0.0
[PE4-ospf-1-area-0.0.0.0] network 1.1.1.4 0.0.0.0
[PE4-ospf-1-area-0.0.0.0] network 10.10.2.0 0.0.0.255
[PE4-ospf-1-area-0.0.0.0] quit
[PE4-ospf-1] quit
[PE4] ospf 2 vpn-instance a
[PE4-ospf-2] import-route bgp
[PE4-ospf-2] area 0.0.0.0
[PE4-ospf-2-area-0.0.0.0] network 10.11.3.0 0.0.0.255
[PE4-ospf-2-area-0.0.0.0] quit
[PE4-ospf-2] quit
[PE4] ospf 3 vpn-instance b
[PE4-ospf-3] import-route bgp
[PE4-ospf-3] area 0.0.0.0
[PE4-ospf-3-area-0.0.0.0] network 10.11.4.0 0.0.0.255
[PE4-ospf-3-area-0.0.0.0] quit
[PE4-ospf-3] quit
```

5. Configure CE a1:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEa1
[CEa1] multicast routing
[CEa1-mrib] quit
```

Assign an IP address to VLAN-interface 10, and enable PIM-SM on the interface.

```
[CEa1] vlan 10
[CEa1-vlan10] port gigabitethernet 1/0/1
[CEa1-vlan10] quit
[CEa1] interface vlan-interface 10
[CEa1-Vlan-interface10] ip address 10.11.5.1 24
[CEa1-Vlan-interface10] pim sm
[CEa1-Vlan-interface10] quit
```

Assign an IP address to VLAN-interface 11, and enable PIM-SM on the interface.

```
[CEa1] vlan 11
[CEa1-vlan11] port gigabitethernet 1/0/2
[CEa1-vlan11] quit
[CEa1] interface vlan-interface 11
[CEa1-Vlan-interface11] ip address 10.11.1.2 24
[CEa1-Vlan-interface11] pim sm
[CEa1-Vlan-interface11] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[CEa1] interface loopback 1
[CEa1-LoopBack1] ip address 2.2.2.2 32
[CEa1-LoopBack1] pim sm
[CEa1-LoopBack1] quit
```

Configure Loopback 1 as a C-BSR and a C-RP.

```
[CEa1] pim
[CEa1-pim] c-bsr 2.2.2.2
[CEa1-pim] c-rp 2.2.2.2
[CEa1-pim] quit
```

Configure OSPF.

```
[CEa1] ospf 1
[CEa1-ospf-1] area 0.0.0.0
[CEa1-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[CEa1-ospf-1-area-0.0.0.0] network 10.11.1.0 0.0.0.255
[CEa1-ospf-1-area-0.0.0.0] network 10.11.5.0 0.0.0.255
[CEa1-ospf-1-area-0.0.0.0] quit
[CEa1-ospf-1] quit
```

6. Configure CE b1:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEb1
[CEb1] multicast routing
[CEb1-mrib] quit
```

Assign an IP address to VLAN-interface 20, and enable PIM-SM on the interface.

```
[CEb1] vlan 20
[CEb1-vlan20] port gigabitethernet 1/0/1
[CEb1-vlan20] quit
[CEb1] interface vlan-interface 20
[CEb1-Vlan-interface20] ip address 10.11.6.1 24
[CEb1-Vlan-interface20] pim sm
[CEb1-Vlan-interface20] quit
```

Assign an IP address to VLAN-interface 12, and enable PIM-SM on the interface.

```
[CEb1] vlan 12
[CEb1-vlan12] port gigabitethernet 1/0/2
[CEb1-vlan12] quit
[CEb1] interface vlan-interface 12
[CEb1-Vlan-interface12] ip address 10.11.2.2 24
[CEb1-Vlan-interface12] pim sm
[CEb1-Vlan-interface12] quit
```

Configure OSPF.

```
[CEb1] ospf 1
[CEb1-ospf-1] area 0.0.0.0
[CEb1-ospf-1-area-0.0.0.0] network 10.11.2.0 0.0.0.255
[CEb1-ospf-1-area-0.0.0.0] network 10.11.6.0 0.0.0.255
[CEb1-ospf-1-area-0.0.0.0] quit
[CEb1-ospf-1] quit
```

7. Configure CE a2:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEa2
[CEa2] multicast routing
[CEa2-mrib] quit
```

Assign an IP address to VLAN-interface 30, and enable IGMP on the interface.

```
[CEa2] vlan 30
[CEa2-vlan30] port gigabitethernet 1/0/1
[CEa2-vlan30] quit
[CEa2] interface vlan-interface 30
[CEa2-Vlan-interface30] ip address 10.11.7.1 24
[CEa2-Vlan-interface30] igmp enable
[CEa2-Vlan-interface30] quit
```

Assign an IP address to VLAN-interface 13, and enable PIM-SM on the interface.

```
[CEa2] vlan 13
[CEa2-vlan13] port gigabitethernet 1/0/2
[CEa2-vlan13] quit
[CEa2] interface vlan-interface 13
[CEa2-Vlan-interface13] ip address 10.11.3.2 24
[CEa2-Vlan-interface13] pim sm
[CEa2-Vlan-interface13] quit
```

Configure OSPF.

```
[CEa2] ospf 1
[CEa2-ospf-1] area 0.0.0.0
[CEa2-ospf-1-area-0.0.0.0] network 10.11.3.0 0.0.0.255
[CEa2-ospf-1-area-0.0.0.0] network 10.11.7.0 0.0.0.255
[CEa2-ospf-1-area-0.0.0.0] quit
[CEa2-ospf-1] quit
```

8. Configure CE b2:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CEb2
```

```
[CEb2] multicast routing
[CEb2-mrib] quit
```

Assign an IP address to VLAN-interface 40, and enable IGMP on the interface.

```
[CEb2] vlan 40
[CEb2-vlan40] port gigabitethernet 1/0/1
[CEb2-vlan40] quit
[CEb2] interface vlan-interface 40
[CEb2-Vlan-interface40] ip address 10.11.8.1 24
[CEb2-Vlan-interface40] igmp enable
[CEb2-Vlan-interface40] quit
```

Assign an IP address to VLAN-interface 14, and enable PIM-SM on the interface.

```
[CEb2] vlan 14
[CEb2-vlan14] port gigabitethernet 1/0/2
[CEb2-vlan14] quit
[CEb2] interface vlan-interface 14
[CEb2-Vlan-interface14] ip address 10.11.4.2 24
[CEb2-Vlan-interface14] pim sm
[CEb2-Vlan-interface14] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[CEb2] interface loopback 1
[CEb2-LoopBack1] ip address 3.3.3.3 32
[CEb2-LoopBack1] pim sm
[CEb2-LoopBack1] quit
```

Configure Loopback 1 as a C-BSR and a C-RP.

```
[CEb2] pim
[CEb2-pim] c-bsr 3.3.3.3
[CEb2-pim] c-rp 3.3.3.3
[CEb2-pim] quit
```

Configure OSPF.

```
[CEb2] ospf 1
[CEb2-ospf-1] area 0.0.0.0
[CEb2-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[CEb2-ospf-1-area-0.0.0.0] network 10.11.4.0 0.0.0.255
[CEb2-ospf-1-area-0.0.0.0] network 10.11.8.0 0.0.0.255
[CEb2-ospf-1-area-0.0.0.0] quit
[CEb2-ospf-1] quit
```

Verifying the configuration

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 1.

```
[PE1] display multicast-vpn default-group local
MVPN local default-group information:
 Group address    Source address   Interface     VPN instance
 239.1.1.1        1.1.1.1          MTunnel0      a
 239.4.4.4        1.1.1.1          MTunnel1      b
```

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 4.

```
[PE4] display multicast-vpn default-group local
```

```
MVPN local default-group information:
 Group address    Source address   Interface     VPN instance
 239.1.1.1        1.1.1.4          MTunnel0      a
 239.4.4.4        1.1.1.4          MTunnel1      b
```

Configuration files

- CE a1:

```
#
 sysname CEa1
#
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 10.11.1.0 0.0.0.255
  network 10.11.5.0 0.0.0.255
#
vlan 10 to 11
#
interface LoopBack1
 ip address 2.2.2.2 255.255.255.255
 pim sm
#
interface Vlan-interface10
 ip address 10.11.5.1 255.255.255.0
 pim sm
#
interface Vlan-interface11
 ip address 10.11.1.2 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 10
#
interface GigabitEthernet1/0/2
 port access vlan 11
#
multicast routing
#
pim
 c-bsr 2.2.2.2
 c-rp 2.2.2.2
#
return
```

- CE a2:

```
#
 sysname CEa2
#
ospf 1
 area 0.0.0.0
```

```
  network 10.11.3.0 0.0.0.255
  network 10.11.7.0 0.0.0.255
#
vlan 13
#
vlan 30
#
interface Vlan-interface13
 ip address 10.11.3.2 255.255.255.0
 pim sm
#
interface Vlan-interface30
 ip address 10.11.7.1 255.255.255.0
 igmp enable
#
interface GigabitEthernet1/0/1
 port access vlan 30
#
interface GigabitEthernet1/0/2
 port access vlan 13
#
multicast routing
#
return
```

- CE b1:

```
#
 sysname CEb1
#
ospf 1
 area 0.0.0.0
  network 10.11.2.0 0.0.0.255
  network 10.11.6.0 0.0.0.255
#
vlan 12
#
vlan 20
#
interface Vlan-interface12
 ip address 10.11.2.2 255.255.255.0
 pim sm
#
interface Vlan-interface20
 ip address 10.11.6.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 20
#
```

```
interface GigabitEthernet1/0/2
 port access vlan 12
#
multicast routing
#
return
```

- CE b2:

```
#
 sysname CEb2
#
ospf 1
 area 0.0.0.0
  network 3.3.3.3 0.0.0.0
  network 10.11.4.0 0.0.0.255
  network 10.11.8.0 0.0.0.255
#
vlan 14
#
vlan 40
#
interface LoopBack1
 ip address 3.3.3.3 255.255.255.255
 pim sm
#
interface Vlan-interface14
 ip address 10.11.4.2 255.255.255.0
 pim sm
#
interface Vlan-interface40
 ip address 10.11.8.1 255.255.255.0
 igmp enable
#
interface GigabitEthernet1/0/1
 port access vlan 40
#
interface GigabitEthernet1/0/2
 port access vlan 14
#
multicast routing
#
pim
 c-bsr 3.3.3.3
 c-rp 3.3.3.3
#
return
```

- PE 1:

```
#
 sysname PE1
```

```
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
ip vpn-instance b
 route-distinguisher 200:1
 vpn-target 200:1 import-extcommunity
 vpn-target 200:1 export-extcommunity
 #
 address-family ipv4
#
 router id 1.1.1.1
#
ospf 1
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 10.10.1.0 0.0.0.255
#
ospf 2 vpn-instance a
 import-route bgp
 area 0.0.0.0
  network 10.11.1.0 0.0.0.255
#
ospf 3 vpn-instance b
 import-route bgp
 area 0.0.0.0
  network 10.11.2.0 0.0.0.255
#
 mpls lsr-id 1.1.1.1
#
vlan 2
#
vlan 11 to 12
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.1 255.255.255.255
 pim sm
#
interface Vlan-interface2
 ip address 10.10.1.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
```

```
interface Vlan-interface11
 ip binding vpn-instance a
 ip address 10.11.1.1 255.255.255.0
 pim sm
#
interface Vlan-interface12
 ip binding vpn-instance b
 ip address 10.11.2.1 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port access vlan 11
#
interface GigabitEthernet1/0/3
 port access vlan 12
#
bgp 100
 group pe1-pe2 internal
 peer pe1-pe2 connect-interface LoopBack1
 group pe1-pe4 external
 peer pe1-pe4 as-number 200
 peer pe1-pe4 connect-interface LoopBack1
 peer pe1-pe4 ebgp-max-hop 255
 peer 1.1.1.2 group pe1-pe2
 peer 1.1.1.4 group pe1-pe4
 #
 address-family ipv4 unicast
  peer pe1-pe2 enable
  peer pe1-pe2 label-route-capability
 #
 address-family vpnv4
  peer pe1-pe4 enable
 #
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 2
 #
 ip vpn-instance b
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 3
#
```

```
multicast routing
#
multicast routing vpn-instance a
#
multicast routing vpn-instance b
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack1
  default-group 239.1.1.1
  data-group 225.1.1.0 255.255.255.240
#
multicast-vpn vpn-instance b mode mdt
 address-family ipv4
  source LoopBack1
  default-group 239.4.4.4
  data-group 225.4.4.0 255.255.255.240
#
return
```

- PE 2:

```
#
 sysname PE2
#
 router id 1.1.1.2
#
ospf 1
 area 0.0.0.0
  network 1.1.1.2 0.0.0.0
  network 10.10.1.0 0.0.0.255
  network 11.11.11.11 0.0.0.0
#
 mpls lsr-id 1.1.1.2
#
vlan 2 to 3
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.2 255.255.255.255
 pim sm
#
interface LoopBack2
 ip address 11.11.11.11 255.255.255.255
 pim sm
#
interface Vlan-interface2
 ip address 10.10.1.2 255.255.255.0
 pim sm
```

```
 mpls enable
 mpls ldp enable
#
interface Vlan-interface3
 ip address 192.168.1.1 255.255.255.0
 pim sm
 pim bsr-boundary
 mpls enable
#
interface GigabitEthernet1/0/1
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port access vlan 3
#
bgp 100
 group pe2-pe1 internal
 peer pe2-pe1 connect-interface LoopBack1
 group pe2-pe3 external
 peer pe2-pe3 as-number 200
 peer 1.1.1.1 group pe2-pe1
 peer 192.168.1.2 group pe2-pe3
 #
 address-family ipv4 unicast
  import-route ospf 1
  peer pe2-pe1 enable
  peer pe2-pe1 route-policy map2 export
  peer pe2-pe1 label-route-capability
  peer pe2-pe3 enable
  peer pe2-pe3 route-policy map1 export
  peer pe2-pe3 label-route-capability
#
multicast routing
#
pim
 c-bsr 11.11.11.11
 c-rp 11.11.11.11
#
route-policy map1 permit node 1
 apply mpls-label
#
route-policy map2 permit node 1
 if-match mpls-label
 apply mpls-label
#
msdp
 encap-data-enable
 peer 1.1.1.3 connect-interface LoopBack1
```

```
#
 ip route-static 1.1.1.3 32 Vlan-interface3 192.168.1.2
#
return
```

- PE 3:

```
#
 sysname PE3
#
 router id 1.1.1.3
#
ospf 1
 area 0.0.0.0
  network 1.1.1.3 0.0.0.0
  network 10.10.2.0 0.0.0.255
  network 22.22.22.22 0.0.0.0
#
 mpls lsr-id 1.1.1.3
#
vlan 3 to 4
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.3 255.255.255.255
 pim sm
#
interface LoopBack2
 ip address 22.22.22.22 255.255.255.255
 pim sm
#
interface Vlan-interface3
 ip address 192.168.1.2 255.255.255.0
 pim sm
 pim bsr-boundary
 mpls enable
#
interface Vlan-interface4
 ip address 10.10.2.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port access vlan 4
#
interface GigabitEthernet1/0/2
 port access vlan 3
#
```

```
bgp 200
 group pe3-pe2 external
 peer pe3-pe2 as-number 100
 group pe3-pe4 internal
 peer pe3-pe4 connect-interface LoopBack1
 peer 1.1.1.4 group pe3-pe4
 peer 192.168.1.1 group pe3-pe2
 #
 address-family ipv4 unicast
  import-route ospf 1
  peer pe3-pe2 enable
  peer pe3-pe2 route-policy map1 export
  peer pe3-pe2 label-route-capability
  peer pe3-pe4 enable
  peer pe3-pe4 route-policy map2 export
  peer pe3-pe4 label-route-capability
#
multicast routing
#
pim
 c-bsr 22.22.22.22
 c-rp 22.22.22.22
#
route-policy map1 permit node 1
 apply mpls-label
#
route-policy map2 permit node 1
 if-match mpls-label
 apply mpls-label
#
msdp
 encap-data-enable
 peer 1.1.1.2 connect-interface LoopBack1
#
 ip route-static 1.1.1.2 32 Vlan-interface3 192.168.1.1
#
return
```

- PE 4:

```
#
 sysname PE4
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
ip vpn-instance b
 route-distinguisher 200:1
```

```
 vpn-target 200:1 import-extcommunity
 vpn-target 200:1 export-extcommunity
#
 router id 1.1.1.4
#
ospf 1
 area 0.0.0.0
  network 1.1.1.4 0.0.0.0
  network 10.10.2.0 0.0.0.255
#
ospf 2 vpn-instance a
 import-route bgp
 area 0.0.0.0
  network 10.11.3.0 0.0.0.255
#
ospf 3 vpn-instance b
 import-route bgp
 area 0.0.0.0
  network 10.11.4.0 0.0.0.255
#
 mpls lsr-id 1.1.1.4
#
vlan 4
#
vlan 13 to 14
#
mpls ldp
#
interface LoopBack1
 ip address 1.1.1.4 255.255.255.255
 pim sm
#
interface Vlan-interface4
 ip address 10.10.2.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface13
 ip binding vpn-instance a
 ip address 10.11.3.1 255.255.255.0
 pim sm
#
interface Vlan-interface14
 ip binding vpn-instance b
 ip address 10.11.4.1 255.255.255.0
 pim sm
#
```

```
interface GigabitEthernet1/0/1
 port access vlan 4
#
interface GigabitEthernet1/0/2
 port access vlan 13
#
interface GigabitEthernet1/0/3
 port access vlan 14
#
bgp 200
 group pe4-pe1 external
 peer pe4-pe1 as-number 100
 peer pe4-pe1 connect-interface LoopBack1
 peer pe4-pe1 ebgp-max-hop 255
 group pe4-pe3 internal
 peer pe4-pe3 connect-interface LoopBack1
 peer 1.1.1.1 group pe4-pe1
 peer 1.1.1.3 group pe4-pe3
 #
 address-family ipv4 unicast
  peer pe4-pe3 enable
  peer pe4-pe3 label-route-capability
 #
 address-family vpnv4
  peer pe4-pe1 enable
 #
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 2
 #
 ip vpn-instance b
  #
  address-family ipv4 unicast
   import-route direct
   import-route ospf 3
#
multicast routing
#
multicast routing vpn-instance a
#
multicast routing vpn-instance b
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack1
  default-group 239.1.1.1
```

```
  data-group 225.1.1.0 255.255.255.240
#
multicast-vpn vpn-instance b mode mdt
 address-family ipv4
  source LoopBack1
  default-group 239.4.4.4
  data-group 225.4.4.0 255.255.255.240
#
return1
```

Example: Configuring receiver-PE-based MVPN extranet

Network configuration

As shown in Figure 18, configure MVPN extranet to meet the following requirements:

Item	Network configuration
Multicast sources and receivers	• In VPN instance **a**, S 1 is a multicast source. • In VPN instance **b**, R 1 is a receiver. • For VPN instance **a**, the default group is 239.1.1.1, and the data group range is 225.2.2.0 to 225.2.2.15.
VPN instances to which PE interfaces belong	• PE 1: VLAN-interface 11 belongs to VPN instance **a**. VLAN-interface 12 and Loopback 0 belong to the public network instance. • PE 2: VLAN-interface 12 belongs to VPN instance **b**. VLAN-interface 11 and Loopback 0 belong to the public network instance.
Unicast routing protocols and MPLS	• Configure OSPF on the public network and configure RIP between PEs and CEs. • Establish BGP peer connections between PE 1 and PE 2 on their respective Loopback 0. • Configure MPLS on the public network.
IP multicast routing	• Enable IP multicast routing on the P device. • Enable IP multicast routing for the public network on PE 1 and PE 2. • Enable IP multicast routing for VPN instance **a** on PE 1. • Enable IP multicast routing for VPN instance **b** on PE 2. • Enable IP multicast routing on CE 1 and CE 2.
IGMPv2	Enable IGMPv2 on VLAN-interface 12 of CE 2.
PIM-SM	Enable PIM-SM on the public network and for VPN instances **a** and **b**: • Enable PIM-SM on all interfaces of the P device. • Enable PIM-SM on all public and private network interfaces of PE 1 and PE 2. • Enable PIM-SM on all interfaces that do not have attached receiver hosts on CE 1 and CE 2. • Configure Loopback 0 of the P device as a C-BSR and a C-RP for the public network to provide services for all multicast groups. • Configure Loopback 1 of CE 1 as a C-BSR and a C-RP for VPN instance **a** to provide services for all multicast groups. • Configure Loopback 1 of PE 2 as a C-BSR and a C-RP for VPN instance **b** to provide services for all multicast groups.

Figure 18 Network diagram

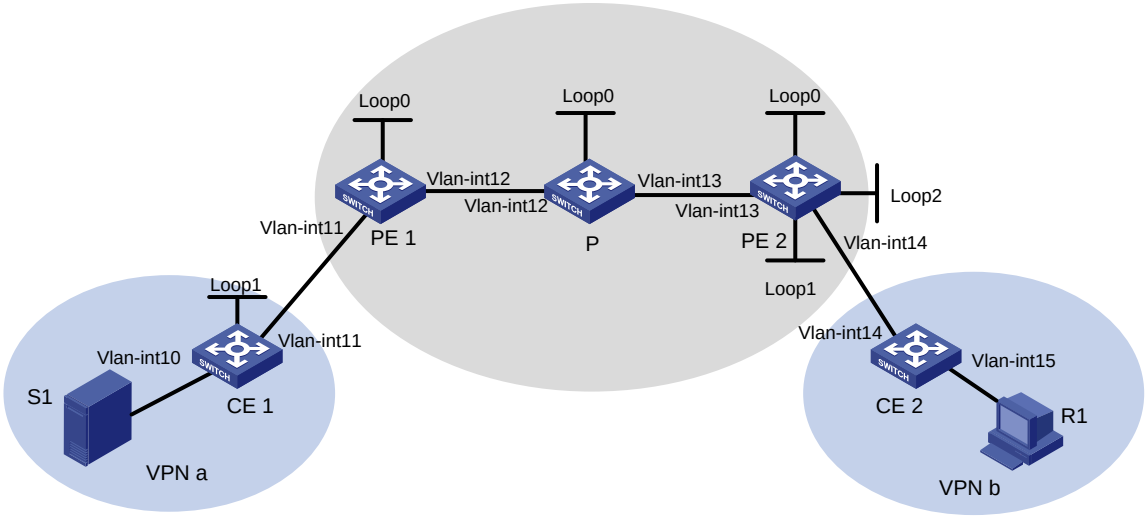


Table 5 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
S 1	—	10.110.7.2/24	PE 2	Vlan-int13	192.168.7.1/24
R 1	—	10.110.11.2/24	PE 2	Vlan-int14	10.110.3.2/24
P	Vlan-int12	192.168.6.1/24	PE 2	Loop0	1.1.1.2/32
P	Vlan-int13	192.168.7.2/24	PE 2	Loop1	20.20.20.20/32
P	Loop0	2.2.2.2/32	PE 2	Loop2	100.100.100.100/32
PE 1	Vlan-int11	10.110.2.1/24	CE 1	Vlan-int10	10.110.7.1/24
PE 1	Vlan-int12	192.168.6.2/24	CE 1	Vlan-int11	10.110.2.2/24
PE 1	Loop0	1.1.1.1/32	CE 1	Loop1	11.11.11.11/32
			CE 2	Vlan-int14	10.110.3.1/24
			CE 2	Vlan-int15	10.110.11.1/24

Procedure

1. Configure PE 1:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE2
[PE1] router id 1.1.1.1
[PE1] multicast routing
[PE1-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE1] mpls lsr-id 1.1.1.1
[PE1] mpls ldp
[PE1-ldp] quit
```

Create a VPN instance named **a**, and configure an RD and route targets for the VPN instance.

```
[PE1] ip vpn-instance a
[PE1-vpn-instance-a] route-distinguisher 100:1
[PE1-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE1-vpn-instance-a] vpn-target 100:1 import-extcommunity
```

```
[PE1-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE1] multicast routing vpn-instance a
[PE1-mrib-a] quit
```

Assign an IP address to Loopback 0, and enable PIM-SM on the interface.

```
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 1.1.1.1 32
[PE1-LoopBack0] pim sm
[PE1-LoopBack0] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE1] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 address family for VPN instance **a**.

```
[PE1-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE1-mvpn-a-ipv4] default-group 239.1.1.1
[PE1-mvpn-a-ipv4] source loopback 0
[PE1-mvpn-a-ipv4] data-group 225.2.2.0 28
[PE1-mvpn-a-ipv4] quit
[PE1-mvpn-a] quit
```

Assign an IP address to VLAN-interface 12, and enable PIM-SM, MPLS, and IPv4 LDP on the interface.

```
[PE1] vlan 12
[PE1-vlan12] port gigabitethernet 1/0/2
[PE1-vlan12] quit
[PE1] interface vlan-interface12
[PE1-Vlan-interface12] ip address 192.168.6.2 24
[PE1-Vlan-interface12] pim sm
[PE1-Vlan-interface12] mpls enable
[PE1-Vlan-interface12] mpls ldp enable
[PE1-Vlan-interface12] quit
```

Associate VLAN-interface 11 with VPN instance **a**, assign an IP address to the interface and enable PIM-SM on the interface.

```
[PE1] vlan 11
[PE1-vlan11] port gigabitethernet 1/0/1
[PE1-vlan11] quit
[PE1] interface vlan-interface11
[PE1-Vlan-interface11] ip binding vpn-instance a
[PE1-Vlan-interface11] ip address 10.110.2.1 24
[PE1-Vlan-interface11] pim sm
[PE1-Vlan-interface11] quit
```

Configure BGP.

```
[PE1] bgp 100
[PE1-bgp-default] group vpn-g internal
[PE1-bgp-default] peer vpn-g connect-interface loopback 0
[PE1-bgp-default] peer 1.1.1.2 group vpn-g
[PE1–bgp-default] ip vpn-instance a
[PE1-bgp-default-a] address-family ipv4
```

```
[PE1-bgp-default-ipv4-a] import-route rip 2
[PE1-bgp-default-ipv4-a] import-route direct
[PE1-bgp-default-ipv4-a] quit
[PE1-bgp-default-a] quit
[PE1–bgp-default] address-family vpnv4
[PE1–bgp-default-vpnv4] peer vpn-g enable
[PE2–bgp-default-vpnv4] quit
[PE1–bgp-default] address-family ipv4 mdt
[PE1–bgp-default-mdt] peer vpn-g enable
[PE1–bgp-default-mdt] quit
[PE1–bgp-default] quit
```

Configure OSPF.

```
[PE1] ospf 1
[PE1-ospf-1] area 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 192.168.6.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

Configure RIP.

```
[PE1] rip 2 vpn-instance a
[PE1-rip-2] network 10.110.2.0 0.0.0.255
[PE1-rip-2] import-route bgp
[PE1-rip-2] quit
```

2. Configure PE 2:

Configure a global router ID, and enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] router id 1.1.1.2
[PE2] multicast routing
[PE2-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[PE2] mpls lsr-id 1.1.1.2
[PE2] mpls ldp
[PE2-ldp] quit
```

Create a VPN instance named **a**, and configure an RD and route targets for the VPN instance.

```
[PE2] ip vpn-instance a
[PE2-vpn-instance-a] route-distinguisher 100:1
[PE2-vpn-instance-a] vpn-target 100:1 export-extcommunity
[PE2-vpn-instance-a] vpn-target 100:1 import-extcommunity
[PE2-vpn-instance-a] quit
```

Enable IP multicast routing for VPN instance **a**.

```
[PE2] multicast routing vpn-instance a
[PE2-mrib-a] quit
```

Assign an IP address to Loopback 0, and enable PIM-SM on the interface.

```
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 1.1.1.2 32
[PE2-LoopBack0] pim sm
[PE2-LoopBack0] quit
```

Create an MDT-based MVPN for VPN instance **a**.

```
[PE2] multicast-vpn vpn-instance a mode mdt
```

Create an MVPN IPv4 address family for VPN instance **a**.

```
[PE2-mvpn-vpn-instance-a] address-family ipv4
```

Specify the default group, the MVPN source interface, and the data group range for VPN instance **a**.

```
[PE2-mvpn-a-ipv4] default-group 239.1.1.1
[PE2-mvpn-a-ipv4] source loopback 0
[PE2-mvpn-a-ipv4] data-group 225.2.2.0 28
[PE2-mvpn-a-ipv4] quit
[PE2-mvpn-a] quit
```

Create a VPN instance named **b**.

```
[PE2] ip vpn-instance b
[PE2-vpn-instance-b] quit
```

Enable IP multicast routing for VPN instance **b**.

```
[PE2] multicast routing vpn-instance b
```

Configure IPv4 MVPN extranet RPF selection policies.

```
[PE2-mrib-b] multicast extranet select-rpf vpn-instance a source 10.110.7.2 32 group 226.1.1.0 24
[PE2-mrib-b] multicast extranet select-rpf vpn-instance a source 20.20.20.20 32 group 226.1.1.0 24
[PE2-mrib-b] quit
```

Assign an IP address to VLAN-interface 13, and enable PIM-SM, MPLS, and IPv4 LDP on the interface.

```
[PE2] vlan 13
[PE2-vlan13] port gigabitethernet 1/0/1
[PE2-vlan13] quit
[PE2] interface vlan-interface13
[PE2-Vlan-interface13] ip address 192.168.7.1 24
[PE2-Vlan-interface13] pim sm
[PE2-Vlan-interface13] mpls enable
[PE2-Vlan-interface13] mpls ldp enable
[PE2-Vlan-interface13] quit
```

Associate VLAN-interface 14 with VPN instance **b**, assign an IP address to the interface, and enable PIM-SM on the interface.

```
[PE2] vlan 14
[PE2-vlan14] port gigabitethernet 1/0/2
[PE2-vlan14] quit
[PE2] interface vlan-interface14
[PE2-Vlan-interface14] ip binding vpn-instance b
[PE2-Vlan-interface14] ip address 10.110.3.2 24
[PE2-Vlan-interface14] pim sm
[PE2-Vlan-interface14] quit
```

Associate Loopback 1 with VPN instance **b**, assign an IP address to the interface, and enable PIM-SM on the interface.

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip binding vpn-instance b
[PE2-LoopBack1] ip address 20.20.20.20 32
[PE2-LoopBack1] pim sm
```

```
[PE2-LoopBack1] quit
```

Configure Loopback 1 as a C-BSR and a C-RP for VPN instance **b**.

```
[PE2] pim vpn-instance b
[PE2-pim-b] c-bsr 20.20.20.20
[PE2-pim-b] c-rp 20.20.20.20
[PE2-pim-b] quit
```

Associate Loopback 2 with VPN instance **a**, assign an IP address to the interface, and enable PIM-SM on the interface.

```
[PE2] interface loopback 2
[PE2-LoopBack2] ip binding vpn-instance a
[PE2-LoopBack2] ip address 100.100.100.100 32
[PE2-LoopBack2] pim sm
[PE2-LoopBack2] quit
```

Configure BGP.

```
[PE2] bgp 100
[PE2-bgp-default] group vpn-g internal
[PE2-bgp-default] peer vpn-g connect-interface loopback 0
[PE2-bgp-default] peer 1.1.1.1 group vpn-g
[PE2–bgp-default] ip vpn-instance a
[PE2-bgp-default-a] address-family ipv4
[PE2-bgp-default-ipv4-a] import-route direct
[PE2-bgp-default-ipv4-a] quit
[PE2-bgp-default-a] quit
[PE2–bgp-default] address-family vpnv4
[PE2–bgp-default-vpnv4] peer vpn-g enable
[PE2–bgp-default-vpnv4] quit
[PE2–bgp-default] address-family ipv4 mdt
[PE2–bgp-default-mdt] peer vpn-g enable
[PE2–bgp-default-mdt] quit
[PE2–bgp-default] quit
```

Configure OSPF.

```
[PE2] ospf 1
[PE2-ospf-1] area 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 1.1.1.2 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 192.168.7.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

Configure RIP.

```
[PE2] rip 3 vpn-instance b
[PE2-rip-3] network 10.110.3.0 0.0.0.255
[PE2-rip-3] network 20.20.20.20 0.0.0.0
[PE2-rip-3] import-route bgp
[PE2-rip-3] quit
```

3. Configure the P device:

Enable IP multicast routing on the public network.

```
<Sysname> system-view
[Sysname] sysname P
[P] multicast routing
```

```
[P-mrib] quit
```

Configure an LSR ID, and enable LDP globally.

```
[P] mpls lsr-id 2.2.2.2
[P] mpls ldp
[P-ldp] quit
```

Assign an IP address to VLAN-interface 12, and enable PIM-SM, MPLS, and IPv4 LDP on the interface.

```
[P] vlan 12
[P-vlan12] port gigabitethernet 1/0/1
[P-vlan12] quit
[P] interface vlan-interface12
[P-Vlan-interface12] ip address 192.168.6.1 24
[P-Vlan-interface12] pim sm
[P-Vlan-interface12] mpls enable
[P-Vlan-interface12] mpls ldp enable
[P-Vlan-interface12] quit
```

Assign an IP address to VLAN-interface 13, and enable PIM-SM, MPLS, and IPv4 LDP on the interface.

```
[P] vlan 13
[P-vlan13] port gigabitethernet 1/0/2
[P-vlan13] quit
[P] interface vlan-interface13
[P-Vlan-interface13] ip address 192.168.7.2 24
[P-Vlan-interface13] pim sm
[P-Vlan-interface13] mpls enable
[P-Vlan-interface13] mpls ldp enable
[P-Vlan-interface13] quit
```

Assign an IP address to Loopback 0, and enable PIM-SM on the interface.

```
[P] interface loopback 0
[P-LoopBack0] ip address 2.2.2.2 32
[P-LoopBack0] pim sm
[P-LoopBack0] quit
```

Configure Loopback 0 as a C-BSR and a C-RP for the public network.

```
[P] pim
[P-pim] c-bsr 2.2.2.2
[P-pim] c-rp 2.2.2.2
[P-pim] quit
```

Configure OSPF.

```
[P] ospf 1
[P-ospf-1] area 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 192.168.6.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 192.168.7.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] quit
```

4. Configure CE 1:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CE1
```

```
[CE1] multicast routing
[CE1-mrib] quit
```

Assign an IP address to VLAN-interface 10, and enable PIM-SM on the interface.

```
[CE1] vlan 10
[CE1-vlan10] port gigabitethernet 1/0/1
[CE1-vlan10] quit
[CE1] interface vlan-interface10
[CE1-Vlan-interface10] ip address 10.100.7.1 24
[CE1-Vlan-interface10] pim sm
[CE1-Vlan-interface10] quit
```

Assign an IP address to VLAN-interface 11, and enable PIM-SM on the interface.

```
[CE1] vlan 11
[CE1-vlan11] port gigabitethernet 1/0/2
[CE1-vlan11] quit
[CE1] interface vlan-interface11
[CE1-Vlan-interface11] ip address 10.110.2.2 24
[CE1-Vlan-interface11] pim sm
[CE1-Vlan-interface11] quit
```

Assign an IP address to Loopback 1, and enable PIM-SM on the interface.

```
[CE1] interface loopback 1
[CE1-LoopBack1] ip address 11.11.11.11 32
[CE1-LoopBack1] pim sm
[CE1-LoopBack1] quit
```

Configure Loopback 1 as a C-BSR and a C-RP for VPN instance **a**.

```
[CE1] pim
[CE1-pim] c-bsr 11.11.11.11
[CE1-pim] c-rp 11.11.11.11
[CE1-pim] quit
```

Configure RIP.

```
[CE1] rip 2
[CE1-rip-2] network 10.110.2.0 0.0.0.255
[CE1-rip-2] network 10.110.7.0 0.0.0.255
[CE1-rip-2] network 11.11.11.11 0.0.0.0
[CE1-rip-2] quit
```

5. Configure CE 2:

Enable IP multicast routing.

```
<Sysname> system-view
[Sysname] sysname CE2
[CE2] multicast routing
[CE2-mrib] quit
```

Assign an IP address to VLAN-interface 14, and enable IGMP on the interface.

```
[CE2] vlan 14
[CE2-vlan14] port gigabitethernet 1/0/1
[CE2-vlan14] quit
[CE2] interface vlan-interface14
[CE2-Vlan-interface14] ip address 10.110.3.1 24
[CE2-Vlan-interface14] pim sm
[CE2-Vlan-interface14] quit
```

Assign an IP address to VLAN-interface 15, and enable PIM-SM on the interface.

```
[CE2] vlan 15
[CE2-vlan15] port gigabitethernet 1/0/2
[CE2-vlan15] quit
[CE2] interface vlan-interface15
[CE2-Vlan-interface15] ip address 10.110.11.1 24
[CE2-Vlan-interface15] igmp enable
[CE2-Vlan-interface15] quit
```

Configure RIP.

```
[CE2] rip 3
[CE2-rip-3] network 10.110.3.0 0.0.0.255
[CE2-rip-3] network 10.110.11.0 0.0.0.255
[CE2-rip-3] quit
```

Verifying the configuration

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 1.

```
[PE1] display multicast-vpn default-group local
MVPN local default-group information:
 Group address     Source address    Interface     VPN instance
 239.1.1.1         1.1.1.1           MTunnel0      a
```

Display information about the local default group for IPv4 multicast transmission in each VPN instance on PE 2.

```
[PE2] display multicast-vpn default-group local
MVPN local default-group information:
 Group address     Source address    Interface     VPN instance
 239.1.1.1         1.1.1.2           MTunnel0      a
```

Display PIM routing entries for MVPN extranet of VPN instance **a** on PE 2.

```
[PE2] display pim vpn-instance a routing-table
 Total 1 (*, G) entries; 1 (S, G) entries
 (*, 226.1.1.1)
     RP: 11.11.11.11
     Protocol: pim-sm, Flag: WC
     UpTime: 07:06:11
     Upstream interface: MTunnel0
         Upstream neighbor: 1.1.1.1
         RPF prime neighbor: 1.1.1.1
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Extranet (VPN: b)
             Protocol: MD, UpTime: 01:12:52, Expires: -

(10.110.7.2, 226.1.1.1)
     RP: 11.11.11.11
     Protocol: pim-sm, Flag: SPT ACT RQ
     UpTime: 07:06:10
     Upstream interface: MTunnel0
         Upstream neighbor: 1.1.1.1
         RPF prime neighbor: 1.1.1.1
```

```
    Downstream interface information:
    Total number of downstream interfaces: 1
        1: Extranet (VPN: b)
            Protocol: MD, UpTime: 01:12:52, Expires: -
```

Display PIM routing entries for MVPN extranet of VPN instance **b** on PE 2.

```
[PE2] display pim vpn-instance b routing-table
 Total 1 (*, G) entries; 1 (S, G) entries
 (*, 226.1.1.1)
     RP: 20.20.20.20 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 07:06:11
     Upstream interface: Extranet (VPN: a)
         Upstream neighbor: 127.0.0.1
         RPF prime neighbor: 127.0.0.1
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Vlan-interface14
             Protocol: pim-sm, UpTime: 01:12:52, Expires: -
(10.110.7.2, 226.1.1.1)
     RP: 20.20.20.20 (local)
     Protocol: pim-sm, Flag: SPT ACT 2MVPN
     UpTime: 07:06:10
     Upstream interface: Extranet (VPN: a)
         Upstream neighbor: 127.0.0.1
         RPF prime neighbor: 127.0.0.1
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Vlan-interface14
             Protocol: pim-sm, UpTime: 01:12:52, Expires: -
```

Configuration files

- CE 1:

```
#
 sysname CE1
#
rip 2
 network 10.110.2.0 0.0.0.255
 network 10.110.7.0 0.0.0.255
 network 11.11.11.11 0.0.0.0
#
vlan 10 to 11
#
interface LoopBack1
 ip address 11.11.11.11 255.255.255.255
 pim sm
#
interface Vlan-interface10
 ip address 10.110.7.1 255.255.255.0
 pim sm
```

```
#
interface Vlan-interface11
 ip address 10.110.2.2 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 10
#
interface GigabitEthernet1/0/2
 port access vlan 11
#
multicast routing
#
pim
 c-bsr 11.11.11.11
 c-rp 11.11.11.11
#
return
```

- CE 2:

```
#
 sysname CE2
#
rip 3
 network 10.110.3.0 0.0.0.255
 network 10.110.11.0 0.0.0.255
#
vlan 14 to 15
#
interface Vlan-interface14
 ip address 10.110.3.1 255.255.255.0
 pim sm
#
interface Vlan-interface15
 ip address 10.110.11.1 255.255.255.0
 igmp enable
#
interface GigabitEthernet1/0/1
 port access vlan 14
#
interface GigabitEthernet1/0/2
 port access vlan 15
#
multicast routing
#
return
```

- P:

```
#
 sysname P
```

```
#
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 192.168.6.0 0.0.0.255
  network 192.168.7.0 0.0.0.255
#
 mpls lsr-id 2.2.2.2
#
vlan 12 to 13
#
mpls ldp
#
interface LoopBack0
 ip address 2.2.2.2 255.255.255.255
 pim sm
#
interface Vlan-interface12
 ip address 192.168.6.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface13
 ip address 192.168.7.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port access vlan 12
#
interface GigabitEthernet1/0/2
 port access vlan 13
#
multicast routing
#
pim
 c-bsr 2.2.2.2
 c-rp 2.2.2.2
#
return
```

- PE 1:

```
#
 sysname PE1
#
ip vpn-instance a
 route-distinguisher 100:1
```

```
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
 router id 1.1.1.1
#
ospf 1
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 192.168.6.0 0.0.0.255
#
rip 2 vpn-instance a
 network 10.110.2.0 0.0.0.255
 import-route bgp
#
 mpls lsr-id 1.1.1.1
#
vlan 11 to 12
#
mpls ldp
#
interface LoopBack0
 ip address 1.1.1.1 255.255.255.255
 pim sm
#
interface Vlan-interface11
 ip binding vpn-instance a
 ip address 10.110.2.1 255.255.255.0
 pim sm
#
interface Vlan-interface12
 ip address 192.168.6.2 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port access vlan 11
#
interface GigabitEthernet1/0/2
 port access vlan 12
#
bgp 100
 group vpn-g internal
 peer vpn-g connect-interface LoopBack0
 peer 1.1.1.2 group vpn-g
 #
 address-family ipv4 unicast
  import-route direct
```

```
  import-route rip 2
 #
 address-family vpnv4
  peer vpn-g enable
 #
 address-family ipv4 mdt
  peer vpn-g enable
 #
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
   import-route rip 2
#
multicast routing
#
multicast routing vpn-instance a
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack0
  default-group 239.1.1.1
  data-group 225.2.2.0 255.255.255.240
#
return
```

- PE 2:

```
#
 sysname PE2
#
ip vpn-instance a
 route-distinguisher 100:1
 vpn-target 100:1 import-extcommunity
 vpn-target 100:1 export-extcommunity
#
ip vpn-instance b
#
 router id 1.1.1.2
#
ospf 1
 area 0.0.0.0
  network 1.1.1.2 0.0.0.0
  network 192.168.7.0 0.0.0.255
#
rip 3 vpn-instance b
 network 10.110.3.0 0.0.0.255
 network 20.20.20.20 0.0.0.0
 import-route bgp
#
```

```
 mpls lsr-id 1.1.1.2
#
vlan 13 to 14
#
mpls ldp
#
interface LoopBack0
 ip address 1.1.1.2 255.255.255.255
 pim sm
#
interface LoopBack1
 ip binding vpn-instance b
 ip address 20.20.20.20 255.255.255.255
 pim sm
#
interface LoopBack2
 ip binding vpn-instance a
 ip address 100.100.100.100 255.255.255.255
 pim sm
#
interface Vlan-interface13
 ip address 192.168.7.1 255.255.255.0
 pim sm
 mpls enable
 mpls ldp enable
#
interface Vlan-interface14
 ip binding vpn-instance b
 ip address 10.110.3.2 255.255.255.0
 pim sm
#
interface GigabitEthernet1/0/1
 port access vlan 13
#
interface GigabitEthernet1/0/2
 port access vlan 14
#
bgp 100
 group vpn-g internal
 peer vpn-g connect-interface LoopBack0
 peer 1.1.1.1 group vpn-g
 #
 address-family vpnv4
  peer vpn-g enable
 #
 address-family ipv4 mdt
  peer vpn-g enable
 #
```

```
 ip vpn-instance a
  #
  address-family ipv4 unicast
   import-route direct
#
multicast routing
#
multicast routing vpn-instance a
#
multicast routing vpn-instance b
 multicast extranet select-rpf vpn-instance a source 10.110.7.2 32 group 226.1.1.0
24
 multicast extranet select-rpf vpn-instance a source 20.20.20.20 32 group 226.1.1.0
24
#
pim vpn-instance b
 c-bsr 20.20.20.20
 c-rp 20.20.20.20
#
multicast-vpn vpn-instance a mode mdt
 address-family ipv4
  source LoopBack0
  default-group 239.1.1.1
  data-group 225.2.2.0 255.255.255.240
#
return
```

Troubleshooting MDT-based MVPN

This section describes common MDT-based MVPN problems and how to troubleshoot them.

A default MDT cannot be established

Symptom

The default MDT cannot be established. PIM neighboring relationship cannot be established between PEs' interfaces that are in the same VPN instance.

Solution

To resolve the problem:

1. Use the **`display interface`** command to examine the MTI interface state and address encapsulation on the MTI.
2. Use the **`display multicast-vpn default-group`** command to verify that the same default group address has been configured for the same VPN instance on different PEs.
3. Use the **`display pim interface`** command to verify the following:
   - PIM is enabled on a minimum of one interface of the same VPN on different PEs.
   - The same PIM mode is running on all the interfaces of the same VPN instance on different PEs and on all the interfaces of the P device.

4. Use the **display ip routing-table** command to verify that a unicast route exists from the VPN instance on the local PE to the same VPN instance on each remote PE.
5. Use the **display bgp peer** command to verify that the BGP peer connections have been correctly configured.
6. If the problem persists, contact H3C Support.

An MVRF cannot be created

Symptom

A VPN instance cannot create an MVRF correctly.

Solution

To resolve the problem:

1. Use the **display pim bsr-info** command to verify that the BSR information exists on the public network and VPN instance. If it does not, verify that a unicast route exists to the BSR.
2. Use the **display pim rp-info** command to examine the RP information. If no RP information is available, verify that a unicast route exists to the RP. Use the **display pim neighbor** command to verify that the PIM adjacencies have been correctly established on the public network and the VPN.
3. Use the **ping** command to examine the connectivity between the VPN DR and the VPN RP.
4. If the problem persists, contact H3C Support.

Contents

Configuring IGMP 1

About IGMP 1
IGMP versions 1
IGMPv1 overview 1
IGMPv2 enhancements 2
IGMPv3 enhancements 3
IGMP SSM mapping 4
IGMP proxying 5
IGMP support for VPNs 6
Protocols and standards 6
IGMP tasks at a glance 6
Prerequisites for IGMP 7
Enabling IGMP 7
Configuring basic IGMP features 8
Specifying an IGMP version 8
Configuring a static group member 8
Configuring a multicast group policy 8
Adjusting IGMP performance 9
Configuring IGMP query and response parameters 9
Enabling fast-leave processing 11
Configuring IGMP SSM mappings 12
Configuring IGMP proxying 12
Prerequisites for configuring IGMP proxying 12
Enabling IGMP proxying 12
Enabling multicast forwarding on a non-querier interface 13
Configuring multicast load splitting for IGMP proxy interfaces 13
Enabling IGMP NSR 14
Setting the DSCP value for outgoing IGMP protocol packets 14
Enabling SNMP notifications for IGMP 14
Configuring IGMP logging 15
Enabling IGMP logging 15
Enabling log message suppression for IGMP 15
Display and maintenance commands for IGMP 15
IGMP configuration examples 16
Example: Configuring basic IGMP features 16
Example: Configuring IGMP SSM mapping 18
Example: Configuring IGMP proxying 21
Troubleshooting IGMP 23
No membership information on the receiver-side device 23
Inconsistent membership information on the devices on the same subnet 23

Configuring IGMP

About IGMP

Internet Group Management Protocol (IGMP) establishes and maintains the multicast group memberships between a Layer 3 multicast device and the hosts on the directly connected subnet.

IGMP versions

IGMP has the following versions:

- IGMPv1 (defined by RFC 1112).
- IGMPv2 (defined by RFC 2236).
- IGMPv3 (defined by RFC 3376).

All IGMP versions support the ASM model. IGMPv3 can directly implement the SSM model. IGMPv1 and IGMPv2 must work with the IGMP SSM mapping feature to implement the SSM model. For more information about the ASM and SSM models, see "Multicast overview."

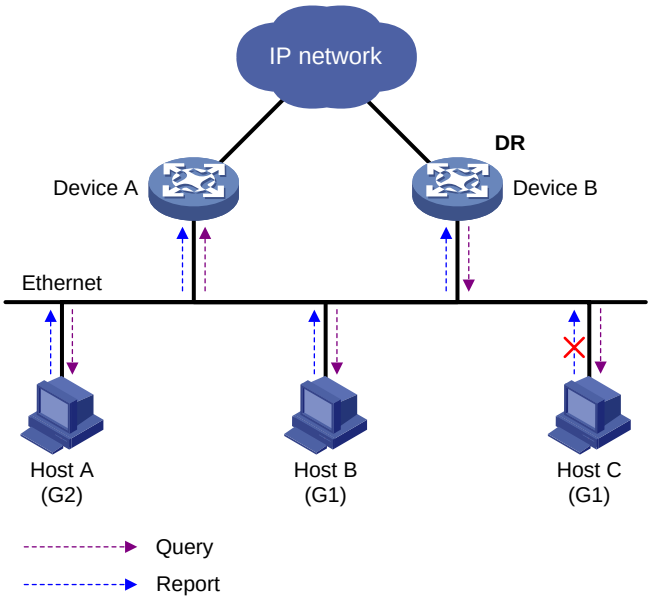
IGMPv1 overview

IGMPv1 manages multicast group memberships based on the query and response mechanism.

All devices that run IGMP on the same subnet can get IGMP membership report messages (called reports) from hosts. However, only one device can act as the IGMP querier to send IGMP query messages (called queries). The querier election mechanism determines which device acts as the IGMP querier on the subnet.

In IGMPv1, the DR elected by the multicast routing protocol (such as PIM) acts as the IGMP querier. For more information about DR, see "PIM overview."

Figure 1 IGMP queries and reports



As shown in Figure 1, Host B and Host C are interested in the multicast data addressed to the multicast group G1. Host A is interested in the multicast data addressed to G2. The following process

describes how the hosts join the multicast groups and how the IGMP querier (Device B in Figure 1) maintains the multicast group memberships:

1. The hosts send unsolicited IGMP reports to the multicast groups they want to join without having to wait for the IGMP queries.
2. The IGMP querier periodically multicasts IGMP queries (with the destination address of 224.0.0.1) to all hosts and devices on the local subnet.
3. After receiving a query message, the host whose report delay timer expires first sends an IGMP report to multicast group G1 to announce its membership for G1. In this example, Host B sends the report message. After receiving the report from Host B, Host C suppresses its own report for G1.

   Because IGMP devices already know that G1 has a minimum of one member, other members do not need to report their memberships. This mechanism, known as the host IGMP report suppression, helps reduce traffic on the local subnet.
4. At the same time, Host A sends a report to the multicast group G2 after receiving a query.
5. Through the query and response process, the IGMP devices (Device A and Device B) determine that the local subnet has members of G1 and G2. The multicast routing protocol (PIM, for example) on the devices generates (*, G1) and (*, G2) multicast forwarding entries, where asterisk (*) represents any multicast source. These entries are the basis for subsequent multicast forwarding.
6. When the multicast data addressed to G1 or G2 reaches an IGMP device, the device looks up the multicast forwarding table. Based on the (*, G1) or (*, G2) entries, the device forwards the multicast data to the local subnet. Then, the receivers on the subnet can receive the data.

IGMPv1 does not define a leave group message (often called a leave message). When an IGMPv1 host is leaving a multicast group, it stops sending reports to that multicast group. If the subnet has no members for a multicast group, the IGMP devices will not receive any report addressed to that multicast group. In this case, the devices clear the information for that multicast group after a period of time.

IGMPv2 enhancements

Backwards-compatible with IGMPv1, IGMPv2 has introduced a querier election mechanism and a leave-group mechanism.

Querier election mechanism

In IGMPv1, the DR elected by the Layer 3 multicast routing protocol (such as PIM) acts as the querier.

IGMPv2 introduced an independent querier election mechanism. The querier election process is as follows:

1. Initially, every IGMPv2 device assumes itself to be the querier. Each device sends IGMP general query messages (called general queries) to all hosts and devices on the local subnet. The destination address is 224.0.0.1.
2. After receiving a general query, every IGMPv2 device compares the source IP address of the query with its own interface address. The device with the lowest IP address becomes the querier. All the other IGMPv2 devices become non-queriers.
3. All the non-queriers start the other querier present timer. If a device receives an IGMP query from the querier before the timer expires, it resets this timer. Otherwise, the device considers that the querier has timed out. In this case, the device initiates a new querier election process.

"Leave group" mechanism

In IGMPv1, when a host leaves a multicast group, it does not send any notification to the multicast devices. The multicast devices determine whether a group has members by using the maximum response time. This adds to the leave latency.

In IGMPv2, when a host is leaving a multicast group, the following process occurs:

1. The host sends a leave message to all devices on the local subnet. The destination address of leave messages is 224.0.0.2.
2. After receiving the leave message, the querier sends a configurable number of IGMP group-specific queries to the group that the host is leaving. Both the destination address field and the group address field of the message are the address of the multicast group that is being queried.
3. One of the remaining members (if any on the subnet) in the group should send a report within the maximum response time advertised in the group-specific queries.
4. If the querier receives a report for the group before the maximum response timer expires, it maintains the memberships for the group. Otherwise, the querier assumes that the local subnet has no member hosts for the group and stops maintaining the memberships for the group.

IGMPv3 enhancements

IGMPv3 is based on and is compatible with IGMPv1 and IGMPv2. It enhances the control capabilities of hosts and the query and report capabilities of IGMP devices.

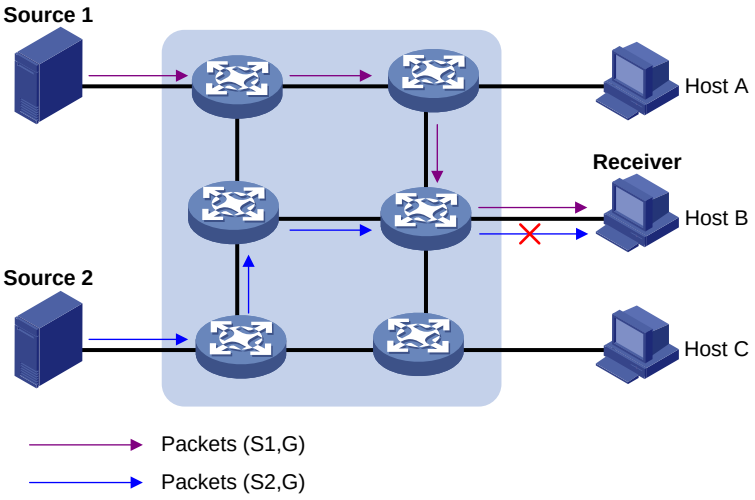
Enhancements in control capability of hosts

IGMPv3 introduced two source filtering modes (Include and Exclude). These modes allow a host to receive or reject multicast data from the specified multicast sources. When a host joins a multicast group, one of the following occurs:

- If the host expects to receive multicast data from specific sources like S1, S2, …, it sends a report with the Filter-Mode denoted as "Include Sources (S1, S2, …)."
- If the host expects to reject multicast data from specific sources like S1, S2, …, it sends a report with the Filter-Mode denoted as "Exclude Sources (S1, S2, …)."

As shown in Figure 2, the network has two multicast sources: Source 1 (S1) and Source 2 (S2). Both of these sources can send multicast data to the multicast group G. Host B wants to receive the multicast data addressed to G from Source 1 but not from Source 2.

Figure 2 Flow paths of source-and-group-specific multicast traffic



In IGMPv1 or IGMPv2, Host B cannot select multicast sources when it joins the multicast group G. The multicast streams from both Source 1 and Source 2 flow to Host B whether or not it needs them.

In IGMPv3, Host B can explicitly express that it needs to receive multicast data destined to the multicast group G from Source 1 but not from Source 2.

Enhancements in query and report capabilities

IGMPv3 introduces IGMP group-and-source queries and IGMP reports carrying group records.

- Query message carrying the source addresses

  IGMPv3 is compatible with IGMPv1 and IGMPv2 and supports IGMP general queries and IGMP group-specific queries. It also introduces IGMP group-and-source-specific queries.

  - A general query does not carry a group address or a source address.
  - A group-specific query carries a group address, but no source address.
  - A group-and-source-specific query carries a group address and one or more source addresses.

- Reports containing multiple group records

  Unlike an IGMPv1 or IGMPv2 report, an IGMPv3 report is destined to 224.0.0.22 and contains one or more group records. Each group record contains a multicast group address and a multicast source address list.

  Group records include the following categories:

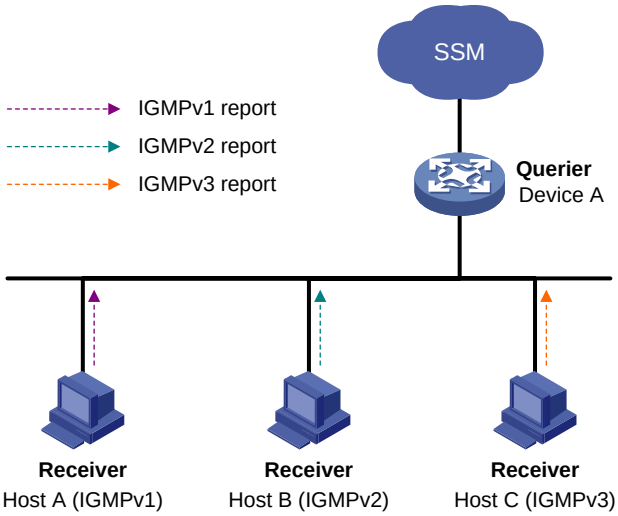
  - **IS_IN**—The current filtering mode is Include. The report sender requests the multicast data only from the sources specified in the Source Address field.
  - **IS_EX**—The current filtering mode is Exclude. The report sender requests the multicast data from any sources except those specified in the Source Address field.
  - **TO_IN**—The filtering mode has changed from Exclude to Include.
  - **TO_EX**—The filtering mode has changed from Include to Exclude.
  - **ALLOW**—The Source Address field contains a list of additional sources from which the receiver wants to obtain data. If the current filtering mode is Include, these sources are added to the multicast source list. If the current filtering mode is Exclude, these sources are deleted from the multicast source list.
  - **BLOCK**—The Source Address field contains a list of the sources from which the receiver no longer wants to obtain data. If the current filtering mode is Include, these sources are deleted from the multicast source list. If the current filtering mode is Exclude, these sources are added to the multicast source list.

IGMP SSM mapping

An IGMPv3 host can explicitly specify multicast sources in its IGMPv3 reports. From the reports, the IGMP device can obtain the multicast source addresses and directly provide the SSM service. However, an IGMPv1 or IGMPv2 host cannot specify multicast sources in its IGMPv1 or IGMPv2 reports.

The IGMP SSM mapping feature enables the IGMP device to provide SSM support for IGMPv1 or IGMPv2 hosts. The device translates (*, G) in IGMPv1 or IGMPv2 reports into (G, INCLUDE, (S1, S2...)) based on the configured IGMP SSM mappings.

Figure 3 IGMP SSM mapping



As shown in Figure 3, on an SSM network, Host A, Host B, and Host C run IGMPv1, IGMPv2, and IGMPv3, respectively. To provide the SSM service for Host A and Host B, you must configure the IGMP SSM mapping feature on Device A.

After IGMP SSM mappings are configured, Device A checks the multicast group address G in the received IGMPv1 or IGMPv2 report, and performs the following operations:

- If G is not in the SSM group range, Device A provides the ASM service.
- If G is in the SSM group range but does not match any IGMP SSM mapping, Device A drops the report.
- If G is in the SSM group range and matches IGMP SSM mappings, Device A translates (*, G) in the report into (G, INCLUDE, (S1, S2...)) to provide SSM services.

NOTE:

The IGMP SSM mapping feature does not process IGMPv3 reports.

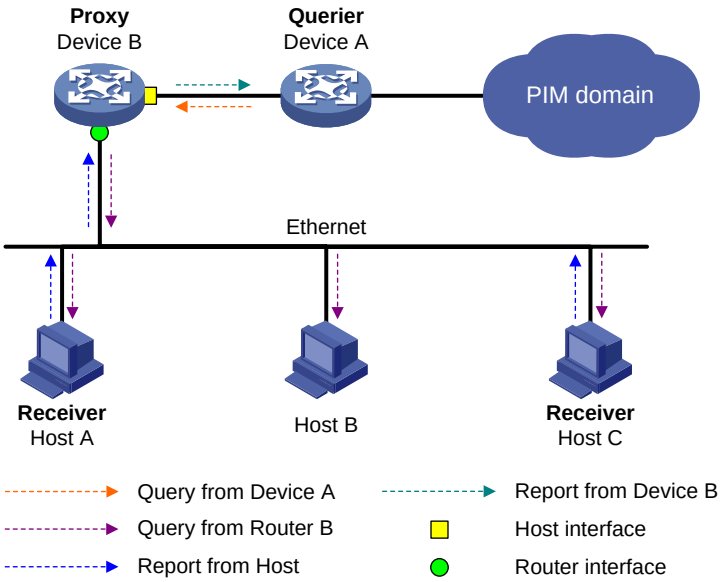
For more information about SSM group ranges, see "PIM overview" and "Configuring PIM."

IGMP proxying

As shown in Figure 4, in a simple tree-shaped topology, it is not necessary to run multicast routing protocols, such as PIM, on edge devices. Instead, you can configure IGMP proxying on these devices. With IGMP proxying configured, the edge device acts as an IGMP proxy:

- For the upstream IGMP querier, the IGMP proxy device acts as a host.
- For the downstream receiver hosts, the IGMP proxy device acts as an IGMP querier.

Figure 4 IGMP proxying



The following types of interfaces are defined in IGMP proxying:

- **Host interface**—An interface that is in the direction toward the root of the multicast forwarding tree. A host interface acts as a receiver host that is running IGMP. IGMP proxying must be enabled on this interface. This interface is also called the "proxy interface."
- **Router interface**—An interface that is in the direction toward the leaf of the multicast forwarding tree. A router interface acts as a router that is running IGMP. IGMP must be configured on this interface.

An IGMP proxy device maintains a group membership database, which stores the group memberships on all the router interfaces. The host interfaces and router interfaces perform actions based on this membership database.

- The host interfaces respond to queries according to the membership database or send join/leave messages when the database changes.
- The router interfaces participate in the querier election, send queries, and maintain memberships based on received IGMP reports.

IGMP support for VPNs

IGMP maintains group memberships on a per-interface basis. After receiving an IGMP message on an interface, IGMP processes the packet within the VPN to which the interface belongs. IGMP only communicates with other multicast protocols within the same VPN instance.

Protocols and standards

- RFC 1112, *Host Extensions for IP Multicasting*
- RFC 2236, *Internet Group Management Protocol, Version 2*
- RFC 3376, *Internet Group Management Protocol, Version 3*

IGMP tasks at a glance

To configure IGMP, perform the following tasks:

1. Enabling IGMP
2. (Optional.) Configuring basic IGMP features
   - Specifying an IGMP version
   - Configuring a static group member
   - Configuring a multicast group policy
3. (Optional.) Adjusting IGMP performance
   - Configuring IGMP query and response parameters
   - Enabling fast-leave processing
4. (Optional.) Configuring IGMP SSM mappings
5. (Optional.) Configuring IGMP proxying
   - Enabling IGMP proxying
   - Enabling multicast forwarding on a non-querier interface
   - Configuring multicast load splitting for IGMP proxy interfaces
6. (Optional.) Enabling IGMP NSR
7. (Optional.) Setting the DSCP value for outgoing IGMP protocol packets
8. (Optional.) Enabling SNMP notifications for IGMP
9. (Optional.) Configuring IGMP logging
   - Enabling IGMP logging
   - Enabling log message suppression for IGMP

Prerequisites for IGMP

Before you configure IGMP, you must configure any unicast routing protocol so that all devices can interoperate at the network layer.

Enabling IGMP

Restrictions and guidelines

Enable IGMP on interfaces where the multicast group memberships are established and maintained.

For IGMP to take effect on an interface, you must configure the primary IP address for the interface.

Procedure

1. Enter system view.
   `system-view`
2. Enable IP multicast routing and enter MRIB view.
   `multicast routing` [**`vpn-instance`** *`vpn-instance-name`*]
   By default, IP multicast routing is disabled.
   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.
   `quit`
4. Enter interface view.
   `interface` *`interface-type interface-number`*
5. Enable IGMP.
   `igmp enable`

By default, IGMP is disabled.

Configuring basic IGMP features

Specifying an IGMP version

Restrictions and guidelines

For IGMP to operate correctly, specify the same IGMP version for all devices on the same subnet.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify an IGMP version on the interface.
   igmp version *version-number*
   By default, the IGMP version on an interface is IGMPv2.

Configuring a static group member

About this task

You can configure an interface as a static group member of a multicast group. Then, the interface can always receive multicast data addressed to the specified multicast group.

A static group member does not respond to IGMP queries. When you complete or cancel this configuration on an interface, the interface does not send an unsolicited IGMP report or leave message.

Restrictions and guidelines

The interface to be configured as a static group member has the following restrictions:

- If the interface is IGMP and PIM-SM enabled, it must be a PIM-SM DR.
- If the interface is IGMP enabled but not PIM-SM enabled, it must be an IGMP querier.

For more information about PIM-SM and DR, see "PIM overview."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the interface as a static group member.
   igmp static-group *group-address* [**source** *source-address*]

Configuring a multicast group policy

About this task

This feature enables an interface to filter IGMP reports by using an ACL that specifies multicast groups and the optional sources. It is used to control the multicast groups that the hosts attached to an interface can join.

Restrictions and guidelines

This configuration does not take effect on static group members, because static group members do not send IGMP reports.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure a multicast group policy.
   igmp group-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* } [*version-number*]

Adjusting IGMP performance

Configuring IGMP query and response parameters

About this task

The following are IGMP query and response parameters:

- **IGMP querier's robustness variable**—Number of times for retransmitting IGMP queries in case of packet loss. A higher robustness variable makes the IGMP querier more robust, but increases the timeout time for multicast groups.
- **IGMP startup query interval**—Interval at which an IGMP querier sends IGMP general queries at startup.
- **IGMP startup query count**—Number of IGMP general queries that an IGMP querier sends at startup.
- **IGMP general query interval**—Interval at which an IGMP querier sends IGMP general queries to check for multicast group members on the network.
- **IGMP last member query interval**—In IGMPv2, it sets the interval at which a querier sends group-specific queries after receiving a leave message. In IGMPv3, it sets the interval at which a querier sends group-and-source-specific queries after receiving a report that changes multicast source and group mappings.
- **IGMP last member query count**—In IGMPv2, it sets the number of group-specific queries that a querier sends after receiving a leave message. In IGMPv3, it sets the number of group-and-source-specific queries that a querier sends after receiving a report that changes multicast source and group mappings.
- **IGMP maximum response time**—Maximum time before a receiver responds with a report to an IGMP general query. This per-group timer is initialized to a random value in the range of 0 to the maximum response time specified in the IGMP query. When the timer value for a group decreases to 0, the receiver sends an IGMP report to the group.
- **IGMP other querier present timer**—Lifetime for an IGMP querier after a non-querier receives an IGMP general query. If the non-querier does not receive a new query when this timer expires, the non-querier considers that the querier has failed and starts a new querier election.

Restrictions and guidelines

To avoid frequent IGMP querier changes, set the IGMP other querier present timer greater than the IGMP general query interval. In addition, configure the same IGMP other querier present timer for all IGMP devices on the same subnet.

To avoid mistakenly deleting multicast receivers, set the IGMP general query interval greater than the maximum response time for IGMP general queries.

To speed up the response to IGMP queries and avoid simultaneous timer expirations that cause IGMP report traffic bursts, set an appropriate maximum response time.

- For IGMP general queries, the maximum response time is set by the **max-response-time** command.
- For IGMP group-specific queries and IGMP group-and-source-specific queries, the maximum response time equals the IGMP last member query interval.

The following configurations take effect only on the devices that run IGMPv2 and IGMPv3:

- Maximum response time for IGMP general queries.
- IGMP last member query interval.
- IGMP last member query count.
- IGMP other querier present interval.

You can configure the IGMP query and response parameters globally for all interfaces in IGMP view or for an interface in interface view. The interface-specific configuration takes priority over the global configuration.

Configuring the IGMP query and response parameters globally

1. Enter system view.
   system-view
2. Enter IGMP view.
   igmp [**vpn-instance** *vpn-instance-name*]
3. Set the IGMP querier's robustness variable.
   robust-count *count*
   By default, the IGMP querier's robustness variable is 2.
4. Set the IGMP startup query interval.
   startup-query-interval *interval*
   By default, the IGMP startup query interval is equal to one quarter of the IGMP general query interval.
5. Set the IGMP startup query count.
   startup-query-count *count*
   By default, the IGMP startup query count is equal to the IGMP querier's robustness variable.
6. Set the IGMP general query interval.
   query-interval *interval*
   By default, the IGMP general query interval is 125 seconds.
7. Set the IGMP last member query interval.
   last-member-query-interval *interval*
   By default, the IGMP last member query interval is 1 second.
8. Set the IGMP last member query count.
   last-member-query-count *count*
   By default, the IGMP last member query count is equal to the IGMP querier's robustness variable.
9. Set the maximum response time for IGMP general queries.
   max-response-time *time*
   By default, the maximum response time for IGMP general queries is 10 seconds.
10. Set the IGMP other querier present timer.
    other-querier-present-interval *interval*

By default, the IGMP other querier present timer is calculated by using the following formula: [IGMP general query interval] × [IGMP robustness variable] + [maximum response time for IGMP general queries] / 2.

Configuring the IGMP query and response parameters on an interface

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set the IGMP querier's robustness variable.

   igmp robust-count *count*

   By default, the IGMP querier's robustness variable is 2.
4. Set the IGMP startup query interval.

   igmp startup-query-interval *interval*

   By default, the IGMP startup query interval is equal to one quarter of the IGMP general query interval.
5. Set the IGMP startup query count.

   igmp startup-query-count *count*

   By default, the IGMP startup query count is equal to the IGMP querier's robustness variable.
6. Set the IGMP general query interval.

   igmp query-interval *interval*

   By default, the IGMP general query interval is 125 seconds.
7. Set the IGMP last member query interval.

   igmp last-member-query-interval *interval*

   By default, the IGMP last member query interval is 1 second.
8. Set the IGMP last member query count.

   igmp last-member-query-count *count*

   By default, the IGMP last member query count is equal to the IGMP querier's robustness variable.
9. Set the maximum response time for IGMP general queries.

   igmp max-response-time *time*

   By default, the maximum response time for IGMP general queries is 10 seconds.
10. Set the IGMP other querier present timer.

    igmp other-querier-present-interval *interval*

    By default, the IGMP other querier present timer is calculated by using the following formula:

    [IGMP general query interval] × [IGMP robustness variable] + [maximum response time for IGMP general queries] / 2.

Enabling fast-leave processing

About this task

This feature enables an IGMP querier to send leave notifications to the upstream without sending group-specific or group-and-source-specific queries after receiving a leave message. Use this feature to reduce leave latency and to preserve the network bandwidth.

Restrictions and guidelines

The fast-leave processing configuration takes effect only when the device runs IGMPv2 or IGMPv3.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable fast-leave processing.
   igmp fast-leave [**group-policy** *ipv4-acl-number*]
   By default, fast-leave processing is disabled.

Configuring IGMP SSM mappings

About this task

This feature enables the device to provide SSM services for IGMPv1 or IGMPv2 hosts.

Restrictions and guidelines

This feature does not process IGMPv3 messages. Enable IGMPv3 on the receiver-side interface to ensure that IGMPv3 reports can be processed..

Procedure

1. Enter system view.
   system-view
2. Enter IGMP view.
   igmp [**vpn-instance** *vpn-instance-name*]
3. Configure an IGMP SSM mapping.
   ssm-mapping *source-address* { *ipv4-acl-number* | **name** *ipv4-acl-name* }

Configuring IGMP proxying

Prerequisites for configuring IGMP proxying

Before you configure IGMP proxying, determine the router interfaces and host interfaces based on the network topology, and enable IGMP on the router interfaces.

Enabling IGMP proxying

Restrictions and guidelines

If IGMP proxying is enabled on multiple interfaces, only the interface with the highest IP address will forward multicast traffic.

Enable IGMP proxying on the receiver-side interfaces.

On an interface enabled with IGMP proxying, only the **igmp version** command takes effect and other IGMP commands do not take effect.

If you enable both IGMP proxying and a multicast routing protocol (such as PIM or MSDP) on the same device, the multicast routing protocol does not take effect.

In IGMPv1, the DR is elected by PIM and acts as the IGMP querier. Because PIM does not take effect on a proxy device, a router interface running IGMPv1 cannot be elected as the DR. To ensure that the downstream receiver hosts on the router interface can receive multicast data, you must

enable multicast forwarding on the interface. For more information, see "Enabling multicast forwarding on a non-querier interface."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IGMP proxying.
   igmp proxy enable
   By default, IGMP proxying is disabled.

Enabling multicast forwarding on a non-querier interface

About this task

Typically, only IGMP queriers can forward multicast traffic and non-queriers cannot. This prevents multicast data from being repeatedly forwarded. If a router interface on an IGMP proxy device failed in the querier election, enable multicast forwarding on the interface to forward multicast data to attached receiver hosts.

Restrictions and guidelines

A shared-media network might have multiple IGMP proxies, including one proxy acting as a querier. To avoid duplicate multicast traffic, do not enable multicast forwarding on any of the non-querier IGMP proxies for the network.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable multicast forwarding on the interface.
   igmp proxy forwarding
   By default, multicast forwarding is disabled on a non-querier interface.

Configuring multicast load splitting for IGMP proxy interfaces

About this task

If multiple IGMP proxy interfaces exist on the device, only the proxy interface with the highest IP address forwards multicast traffic. You can enable multicast load splitting for IGMP proxy interfaces so that all the proxy interfaces can share multicast traffic.

Procedure

1. Enter system view.
   system-view
2. Enter IGMP view.
   igmp [**vpn-instance** *vpn-instance-name*]
3. Enable multicast load splitting for IGMP proxy interfaces.
   proxy multipath
   By default, multicast load splitting is disabled for IGMP proxy interfaces.

Enabling IGMP NSR

About this task

This feature backs up information about IGMP interfaces and IGMP multicast groups to the standby process. The device recovers the information without cooperation of other devices when an active/standby switchover occurs. Use this feature to prevent an active/standby switchover from affecting the multicast service.

Procedure

1. Enter system view.
   system-view
2. Enable IGMP NSR.
   igmp non-stop-routing
   By default, IGMP NSR is disabled.

Setting the DSCP value for outgoing IGMP protocol packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.
   system-view
2. Enter IGMP view.
   igmp [**vpn-instance** *vpn-instance-name*]
3. Set the DSCP value for outgoing IGMP protocol packets.
   dscp *dscp-value*
   By default, the DSCP value is 48 for outgoing IGMP protocol packets.

Enabling SNMP notifications for IGMP

About this task

To report critical IGMP events to an SNMP NMS, enable SNMP notifications for IGMP.

For IGMP SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about configuring SNMP, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for IGMP.
   snmp-agent trap enable igmp [**join** | **leave**] *
   By default, SNMP notifications are enabled for IGMP.

Configuring IGMP logging

Enabling IGMP logging

About this task

To report critical IGMP events to the information center, enable IGMP logging. For IGMP log messages to be sent correctly, you must also configure the information center on the device. For more information about configuring the information center, see Network Management and Monitoring Configuration Guide.

Restrictions and guidelines

Outputting a large amount of log messages degrades the device performance. To resolve this issue, disable IGMP logging or use the **igmp log suppress** command to enable IGMP log message suppression.

Procedure

1. Enter system view.
   system-view
2. Enable IGMP logging.
   igmp log enable [**group**]
   IGMP logging is enabled by default.

Enabling log message suppression for IGMP

About this task

Outputting a large amount of log messages degrades the device performance. To resolve this issue, perform this task to enable log message suppression for IGMP. With this feature configured, the speed of sending IGMP log messages to the information center will be decreased to reduce the load on the device.

Restrictions and guidelines

Before you configure this feature, make sure IGMP logging has been enabled using the **igmp log enable** command.

Procedure

1. Enter system view.
   system-view
2. Enable log message suppression for IGMP.
   igmp log suppress [**group**]
   By default, log message suppression is enabled for IGMP.

Display and maintenance commands for IGMP

CAUTION:
The **reset igmp group** command might cause multicast data transmission failures.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about IGMP multicast groups.	**display igmp** [**vpn-instance** *vpn-instance-name*] **group** [*group-address* | **interface** *interface-type interface-number*] [**evpn**] [**static** | **verbose**]
Display IGMP information for interfaces.	**display igmp** [**vpn-instance** *vpn-instance-name*] **interface** [*interface-type interface-number*] [**host** | **proxy**] [**evpn**] [**verbose**]
Display multicast group membership information maintained by the IGMP proxy.	**display igmp** [**vpn-instance** *vpn-instance-name*] **proxy group** [*group-address* | **interface** *interface-type interface-number*] [**verbose**]
Display multicast routing entries maintained by the IGMP proxy.	**display igmp** [**vpn-instance** *vpn-instance-name*] **proxy routing-table** [*source-address* [**mask** { *mask-length* | *mask* }] | *group-address* [**mask** { *mask-length* | *mask* }]] * [**verbose**]
Display IGMP SSM mappings.	**display igmp** [**vpn-instance** *vpn-instance-name*] **ssm-mapping** *group-address*
Display authorization information for IGMP users.	**display igmp user-authorization** [**interface** *interface-type interface-number*]
Clear dynamic IGMP multicast group entries.	**reset igmp** [**vpn-instance** *vpn-instance-name*] **group** { **all** | **interface** *interface-type interface-number* { **all** | *group-address* [**mask** { *mask* | *mask-length* }] [*source-address* [**mask** { *mask* | *mask-length* }]] } }

IGMP configuration examples

Example: Configuring basic IGMP features

Network configuration

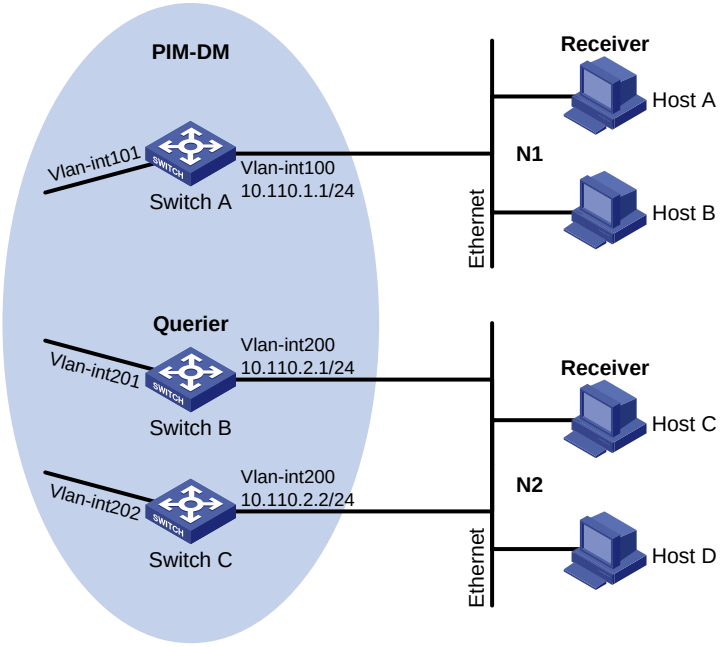
As shown in Figure 5:

- OSPF and PIM-DM run on the network.
- VOD streams are sent to receiver hosts in multicast. Receiver hosts of different organizations form stub networks N1 and N2. Host A and Host C are receiver hosts in N1 and N2, respectively.
- IGMPv2 runs between Switch A and N1, and between the other two switches and N2. Switch A acts as the IGMP querier in N1. Switch B acts as the IGMP querier in N2 because it has a lower IP address.

Configure the routers to meet the following requirements:

- The hosts in N1 can join only multicast group 224.1.1.1.
- The hosts in N2 can join any multicast groups.

Figure 5 Network diagram



Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 5, and configure OSPF on the switches in the PIM-DM domain. (Details not shown.)

Procedure

1. Enable IP multicast routing, IGMP, and PIM-DM:

   # On Switch A, enable IP multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

   # Enable IGMP on VLAN-interface 100.

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] igmp enable
   [SwitchA-Vlan-interface100] quit
   ```

   # Enable PIM-DM on VLAN-interface 101.

   ```
   [SwitchA] interface vlan-interface 101
   [SwitchA-Vlan-interface101] pim dm
   [SwitchA-Vlan-interface101] quit
   ```

   # On Switch B, enable IP multicast routing.

   ```
   <SwitchB> system-view
   [SwitchB] multicast routing
   [SwitchB-mrib] quit
   ```

   # Enable IGMP on VLAN-interface 200.

   ```
   [SwitchB] interface vlan-interface 200
   [SwitchB-Vlan-interface200] igmp enable
   ```

```
[SwitchB-Vlan-interface200] quit
```

Enable PIM-DM on VLAN-interface 201.

```
[SwitchB] interface vlan-interface 201
[SwitchB-Vlan-interface201] pim dm
[SwitchB-Vlan-interface201] quit
```

On Switch C, enable IP multicast routing.

```
<SwitchC> system-view
[SwitchC] multicast routing
[SwitchC-mrib] quit
```

Enable IGMP on VLAN-interface 200.

```
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] igmp enable
[SwitchC-Vlan-interface200] quit
```

Enable PIM-DM on VLAN-interface 202.

```
[SwitchC] interface vlan-interface 202
[SwitchC-Vlan-interface202] pim dm
[SwitchC-Vlan-interface202] quit
```

2. Configure a multicast group policy on Switch A so that hosts connected to VLAN-interface 100 can join only multicast group 224.1.1.1.

```
[SwitchA] acl basic 2001
[SwitchA-acl-ipv4-basic-2001] rule permit source 224.1.1.1 0
[SwitchA-acl-ipv4-basic-2001] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp group-policy 2001
[SwitchA-Vlan-interface100] quit
```

Verifying the configuration

Display IGMP information for VLAN-interface 200 on Switch B.

```
[SwitchB] display igmp interface vlan-interface 200
 Vlan-interface200(10.110.2.1):
   IGMP is enabled.
   IGMP version: 2
   Query interval for IGMP: 125s
   Other querier present time for IGMP: 255s
   Maximum query response time for IGMP: 10s
   Querier for IGMP: 10.110.2.1 (This router)
  IGMP groups reported in total: 1
```

Example: Configuring IGMP SSM mapping

Network configuration

As shown in Figure 6:

- OSPF runs on the network.
- The PIM-SM domain uses the SSM model for multicast delivery. The SSM group range is 232.1.1.0/24.
- IGMPv3 runs on VLAN-interface 400 of Switch D. The receiver host runs IGMPv2, and does not support IGMPv3. The receiver host cannot specify multicast sources in its membership reports.

- Source 1, Source 2, and Source 3 send multicast packets to multicast groups in the SSM group range 232.1.1.0/24.

Configure the IGMP SSM mapping feature on Switch D so that the receiver host can receive multicast data only from Source 1 and Source 3.

Figure 6 Network diagram

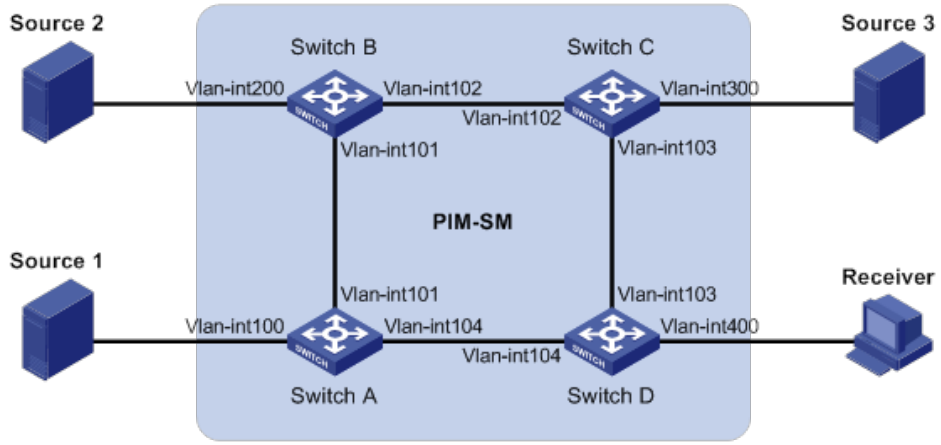


Table 1 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Source 1	—	133.133.1.1/24	Source 3	—	133.133.3.1/24
Source 2	—	133.133.2.1/24	Receiver	—	133.133.4.1/24
Switch A	Vlan-int100	133.133.1.2/24	Switch C	Vlan-int300	133.133.3.2/24
Switch A	Vlan-int101	192.168.1.1/24	Switch C	Vlan-int103	192.168.3.1/24
Switch A	Vlan-int104	192.168.4.2/24	Switch C	Vlan-int102	192.168.2.2/24
Switch B	Vlan-int200	133.133.2.2/24	Switch D	Vlan-int400	133.133.4.2/24
Switch B	Vlan-int101	192.168.1.2/24	Switch D	Vlan-int103	192.168.3.2/24
Switch B	Vlan-int102	192.168.2.1/24	Switch D	Vlan-int104	192.168.4.1/24

Prerequisites

Assign an IP address and subnet mask to each interface as shown in Figure 6, and configure OSPF on the switches in the PIM-SM domain. (Details not shown.)

Procedure

1. Enable IP multicast routing, PIM-SM, and IGMP:

On Switch D, enable IP multicast routing.

```
<SwitchD> system-view
[SwitchD] multicast routing
[SwitchD-mrib] quit
```

Enable IGMPv3 on the receiver-side interface (VLAN-interface 400).

```
[SwitchD] interface vlan-interface 400
[SwitchD-Vlan-interface400] igmp enable
[SwitchD-Vlan-interface400] igmp version 3
[SwitchD-Vlan-interface400] quit
```

Enable PIM-SM on the other interfaces.

```
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim sm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 104
[SwitchD-Vlan-interface104] pim sm
[SwitchD-Vlan-interface104] quit
```

On Switch A, enable IP multicast routing.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Enable PIM-SM on each interface.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 104
[SwitchA-Vlan-interface104] pim sm
[SwitchA-Vlan-interface104] quit
```

Configure Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

2. Configure the SSM group range:

On Switch D, specify 232.1.1.0/24 as the SSM group range.

```
[SwitchD] acl basic 2000
[SwitchD-acl-ipv4-basic-2000] rule permit source 232.1.1.0 0.0.0.255
[SwitchD-acl-ipv4-basic-2000] quit
[SwitchD] pim
[SwitchD-pim] ssm-policy 2000
[SwitchD-pim] quit
```

Configure the SSM group range on Switch A, Switch B, and Switch C in the same way Switch D is configured. (Details not shown.)

3. Configure IGMP SSM mappings on Switch D.

```
[SwitchD] igmp
[SwitchD-igmp] ssm-mapping 133.133.1.1 2000
[SwitchD-igmp] ssm-mapping 133.133.3.1 2000
[SwitchD-igmp] quit
```

Verifying the configuration

Display IGMP SSM mappings for multicast group 232.1.1.1 on Switch D.

```
[SwitchD] display igmp ssm-mapping 232.1.1.1
 Group: 232.1.1.1
 Source list:
        133.133.1.1
        133.133.3.1
```

On Switch D, display information about IGMP multicast groups that hosts have dynamically joined.

```
[SwitchD] display igmp group
IGMP groups in total: 1
 Vlan-interface400(133.133.4.2):
```

```
  IGMP groups reported in total: 1
   Group address   Last reporter   Uptime      Expires
   232.1.1.1       133.133.4.1     00:02:04    off
```

Display PIM routing entries on Switch D.

```
[SwitchD] display pim routing-table
 Total 0 (*, G) entry; 2 (S, G) entry

 (133.133.1.1, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface104
         Upstream neighbor: 192.168.4.2
         RPF prime neighbor: 192.168.4.2
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Vlan-interface400
             Protocol: igmp, UpTime: 00:13:25, Expires: -

 (133.133.3.1, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface103
         Upstream neighbor: 192.168.3.1
         RPF prime neighbor: 192.168.3.1
     Downstream interface information:
     Total number of downstream interfaces: 1
         1: Vlan-interface400
             Protocol: igmp, UpTime: 00:13:25, Expires: -
```

Example: Configuring IGMP proxying

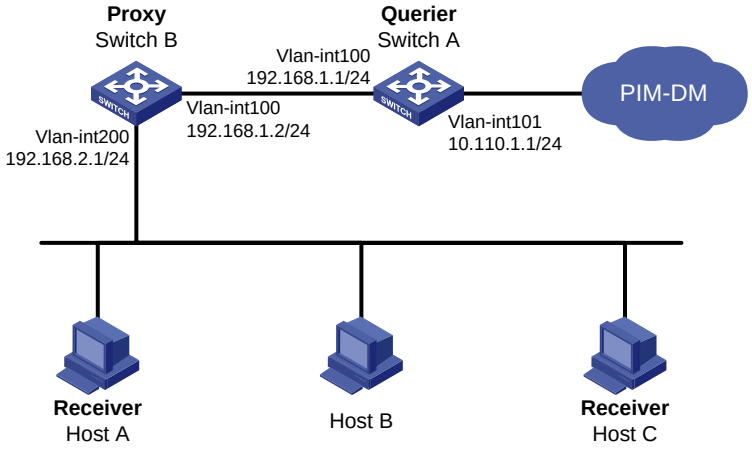
Network configuration

As shown in Figure 7:

- PIM-DM runs on the core network.
- Host A and Host C on the stub network receive VOD information sent to multicast group 224.1.1.1.

Configure the IGMP proxying feature on Switch B so that Switch B can maintain group memberships and forward multicast traffic without running PIM-DM.

Figure 7 Network diagram



Prerequisites

Assign an IP address and subnet mask to each interface, as shown in Figure 7.

Procedure

1. Enable IP multicast routing, PIM-DM, IGMP, and IGMP proxying:

   # On Switch A, enable IP multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] multicast routing
   [SwitchA-mrib] quit
   ```

   # Enable PIM-DM on VLAN-interface 101.

   ```
   [SwitchA] interface vlan-interface 101
   [SwitchA-Vlan-interface101] pim dm
   [SwitchA-Vlan-interface101] quit
   ```

   # Enable IGMP on VLAN-interface 100.

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] igmp enable
   [SwitchA-Vlan-interface100] quit
   ```

   # On Switch B, enable IP multicast routing.

   ```
   <SwitchB> system-view
   [SwitchB] multicast routing
   [SwitchB-mrib] quit
   ```

   # Enable IGMP proxying on VLAN-interface 100.

   ```
   [SwitchB] interface vlan-interface 100
   [SwitchB-Vlan-interface100] igmp proxy enable
   [SwitchB-Vlan-interface100] quit
   ```

   # Enable IGMP on VLAN-interface 200.

   ```
   [SwitchB] interface vlan-interface 200
   [SwitchB-Vlan-interface200] igmp enable
   [SwitchB-Vlan-interface200] quit
   ```

Verifying the configuration

Display multicast group membership information maintained by the IGMP proxy on Switch B.

```
[SwitchB] display igmp proxy group
IGMP proxy group records in total: 1
```

```
Vlan-interface100(192.168.1.2):
 IGMP proxy group records in total: 1
  Group address      Member state      Expires
  224.1.1.1          Delay             00:00:02
```

Troubleshooting IGMP

No membership information on the receiver-side device

Symptom

When a host sends a report for joining multicast group G, no membership information of multicast group G exists on the device closest to that host.

Solution

To resolve the problem:

1. Use the **display igmp interface** command to verify that the networking, interface connection, and IP address configuration are correct.
2. Use the **display current-configuration** command to verify that multicast routing is enabled. If it is not enabled, use the **multicast routing** command in system view to enable IP multicast routing. In addition, verify that IGMP is enabled on the associated interfaces.
3. Use the **display igmp interface** command to verify that the IGMP version on the interface is lower than that on the host.
4. Use the **display current-configuration interface** command to verify that no multicast group policies have been configured to filter IGMP reports for multicast group G.
5. If the problem persists, contact H3C Support.

Inconsistent membership information on the devices on the same subnet

Symptom

Different memberships are maintained on different IGMP devices on the same subnet.

Solution

To resolve the problem:

1. Use the **display current-configuration** command to verify the IGMP information on the interfaces. Make sure the devices on the subnet have the same IGMP settings on their interfaces.
2. Use the **display igmp interface** command on all devices on the same subnet to verify the IGMP-related timer settings. Make sure the settings are consistent on all the devices.
3. Use the **display igmp interface** command to verify that all devices on the same subnet are running the same IGMP version.
4. If the problem persists, contact H3C Support.

Contents

Configuring MLD 1

About MLD 1
MLD versions 1
How MLDv1 works 1
MLDv2 enhancements 3
MLD SSM mapping 4
MLD proxying 5
MLD support for VPNs 5
Protocols and standards 5
MLD tasks at a glance 6
Prerequisites for MLD 6
Enabling MLD 6
Configuring basic MLD features 7
Specifying an MLD version 7
Configuring a static group member 7
Configuring an IPv6 multicast group policy 8
Adjusting MLD performance 8
Configuring MLD query and response parameters 8
Enabling fast-leave processing 10
Configuring MLD SSM mappings 11
Configuring MLD proxying 11
Prerequisites for MLD proxying 11
Enabling MLD proxying 11
Enabling IPv6 multicast forwarding on a non-querier interface 12
Enabling IPv6 multicast load splitting for MLD proxy interfaces 12
Enabling MLD NSR 13
Setting the DSCP value for outgoing MLD protocol packets 13
Enabling SNMP notifications for MLD 13
Configuring MLD logging 14
Enabling MLD logging 14
Enabling log message suppression for MLD 14
Display and maintenance commands for MLD 14
MLD configuration examples 15
Example: Configuring basic MLD features 15
Example: Configuring MLD SSM mapping 17
Example: Configuring MLD proxying 20
Troubleshooting MLD 22
No member information exists on the receiver-side device 22
Inconsistent membership information on the devices on the same subnet 22

Configuring MLD

About MLD

Multicast Listener Discovery (MLD) establishes and maintains IPv6 multicast group memberships between a Layer 3 multicast device and the hosts on the directly connected subnet.

MLD versions

MLD has the following versions:

- MLDv1 (defined by RFC 2710), which is derived from IGMPv2.
- MLDv2 (defined by RFC 3810), which is derived from IGMPv3.

MLDv1 and MLDv2 support the ASM model. MLDv2 can directly implement the SSM model, but MLDv1 must work with the MLD SSM mapping feature to implement the SSM model. For more information about the ASM and SSM models, see "Multicast overview."

How MLDv1 works

MLDv1 implements IPv6 multicast listener management based on the query and response mechanism.

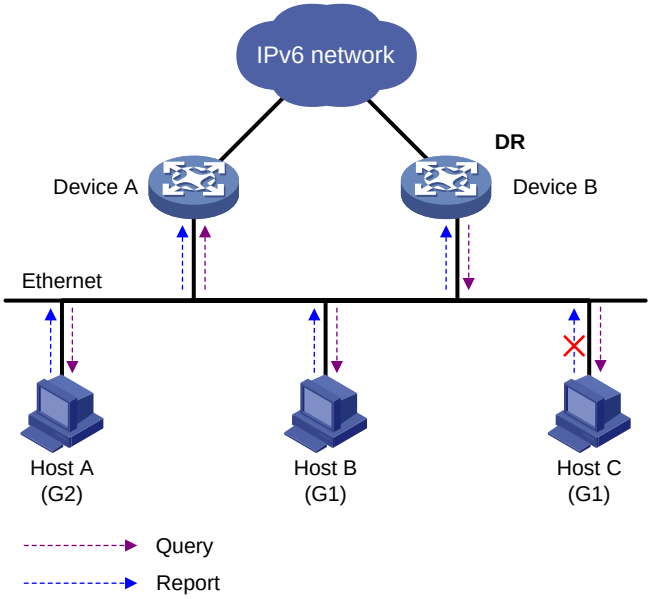
Electing the MLD querier

All IPv6 multicast devices that run MLD on the same subnet can monitor MLD listener report messages (often called reports) from hosts. However, only one device can act as the MLD querier to send MLD query messages (often called queries). A querier election mechanism determines which device acts as the MLD querier on the subnet.

1. Initially, every MLD device assumes itself as the querier. Each device sends MLD general query messages (often called general queries) to all hosts and devices on the local subnet. The destination address of the general queries is FF02::1.
2. After receiving a general query, every MLD device compares the source IPv6 address of the query with its own link-local interface address. The device with the lowest IPv6 address wins the querier election and becomes the querier. All the other devices become non-queriers.
3. All the non-queriers start a timer called the "other querier present timer." If a device receives an MLD query from the querier before the timer expires, it resets this timer. Otherwise, it considers that the querier has timed out. In this case, the device initiates a new querier election process.

Joining an IPv6 multicast group

Figure 1 MLD queries and reports



As shown in Figure 1, Host B and Host C want to receive the IPv6 multicast data addressed to IPv6 multicast group G1. Host A wants to receive the IPv6 multicast data addressed to G2. The following process describes how the hosts join the IPv6 multicast groups and how the MLD querier (Device B in Figure 1) maintains the IPv6 multicast group memberships:

1. The hosts send unsolicited MLD reports to the IPv6 multicast groups they want to join without having to wait for the MLD queries.
2. The MLD querier periodically multicasts MLD queries (with the destination address FF02::1) to all hosts and devices on the local subnet.
3. After receiving a query, the host whose report delay timer expires first sends an MLD report to the IPv6 multicast group G1 to announce its membership for G1. In this example, Host B sends the report. After hearing the report from Host B, Host C, which is on the same subnet as Host B, suppresses its own report for G1.

   Because the MLD devices already know that G1 has a minimum of one member, other members do not need to report their memberships. This mechanism, known as the host MLD report suppression, helps reduce traffic on the local subnet.
4. At the same time, because Host A is interested in G2, it sends a report to the IPv6 multicast group G2.
5. Through the query/report process, the MLD devices determine that G1 and G2 have members on the local subnet. The IPv6 multicast routing protocol (for example, IPv6 PIM) that is running on the devices generates (*, G1) and (*, G2) multicast forwarding entries. These entries are the basis for subsequent IPv6 multicast forwarding. The asterisk (*) represents any IPv6 multicast source.
6. When the IPv6 multicast data addressed to G1 or G2 reaches an MLD device, the device looks up the IPv6 multicast forwarding table. Based on the (*, G1) and (*, G2) entries, the device forwards the IPv6 multicast data to the local subnet. Then, the receivers on the subnet receive the data.

Leaving an IPv6 multicast group

When a host is leaving an IPv6 multicast group, the following process occurs:

1. The host sends an MLD done message to all IPv6 multicast devices on the local subnet. The destination address of done messages is FF02::2.

2. After receiving the MLD done message, the querier sends a configurable number of multicast-address-specific queries to the group that the host is leaving. The IPv6 multicast addresses queried include both the destination address field and the group address field of the message.
3. One of the remaining members (if any on the subnet) in the group sends a report within the time of the maximum response time advertised in the multicast-address-specific queries.
4. If the querier receives a report for the group within the maximum response time, it maintains the memberships of the IPv6 multicast group. Otherwise, the querier assumes that no hosts on the subnet are interested in IPv6 multicast traffic addressed to that group and stops maintaining the memberships of the group.

MLDv2 enhancements

MLDv2 is based on and backwards-compatible with MLDv1. MLDv2 provides hosts with enhanced control capabilities and enhances the MLD state.

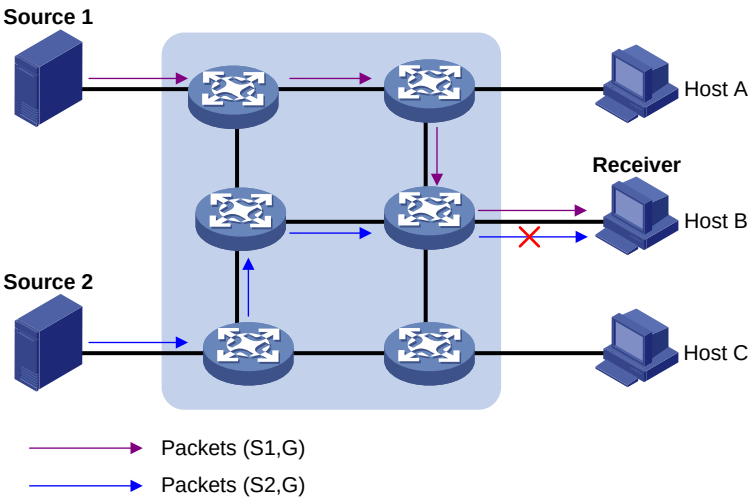
Enhancements in control capability of hosts

MLDv2 has introduced IPv6 multicast source filtering modes (Include and Exclude). These modes allow a host to receive or reject multicast data from the specified IPv6 multicast sources. When a host joins an IPv6 multicast group, one of the following occurs:

- If the host expects IPv6 multicast data from specific IPv6 multicast sources like S1, S2, …, it sends a report with Filter-Mode denoted as "Include Sources (S1, S2, …)."
- If the host does not expect IPv6 multicast data from specific IPv6 multicast sources like S1, S2, …, it sends a report with Filter-Mode denoted as "Exclude Sources (S1, S2, …)."

As shown in Figure 2, the network has two IPv6 multicast sources, Source 1 (S1) and Source 2 (S2). Both of the sources can send IPv6 multicast data to IPv6 multicast group G. Host B wants to receive IPv6 multicast data addressed to G from Source 1 but not from Source 2.

Figure 2 Flow paths of multicast-address-and-source-specific multicast traffic



In MLDv1, Host B cannot select IPv6 multicast sources when it joins IPv6 multicast group G. The IPv6 multicast streams from both Source 1 and Source 2 flow to Host B whether it needs them or not.

In MLDv2, Host B can explicitly express its interest in IPv6 multicast data destined to G from Source 1 but not from Source 2. Then, Host B receives only IPv6 multicast data from Source 1.

Enhancement in MLD state

A multicast device that is running MLDv2 maintains the multicast address state for each multicast address on each attached subnet. The multicast address state consists of the following information:

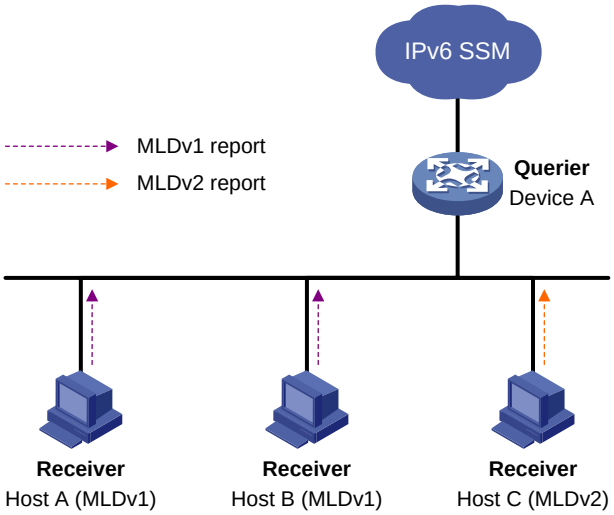
- **Filter mode**—Device keeps tracing the Include or Exclude state.
- **List of sources**—Device keeps tracing the newly added or deleted IPv6 multicast source.
- **Timers**—Filter timers, which include the time that the device waits before switching to the Include mode after an IPv6 multicast address times out, and source timers for source recording.

MLD SSM mapping

An MLDv2 host can explicitly specify multicast sources in its MLDv2 reports. From the reports, the MLD device can obtain the multicast source addresses and directly provide the SSM service. However, an MLDv1 host cannot specify multicast sources in its MLDv1 reports.

The MLD SSM mapping feature enables the MLD device to provide SSM support for MLDv1 receiver host. The device translates (*, G) in MLDv1 reports into (G, INCLUDE, (S1, S2...)) based on the configured MLD SSM mappings.

Figure 3 Network diagram



As shown in Figure 3, Host A and Host B on the IPv6 SSM network run MLDv1, and Host C runs MLDv2. To provide the SSM service for Host A and Host B, you must configure the MLD SSM mapping feature on Device A.

After MLD SSM mappings are configured, Device A checks the IPv6 multicast group address G carried in the message, and performs the following operations:

- If G is not in the IPv6 SSM group range, Device A provides the ASM service.
- If G is in the IPv6 SSM group range but does not match any MLD SSM mapping, Device A drops the report.
- If G is in the IPv6 SSM group range and matches MLD SSM mappings, Device A translates (*, G) in the report to (G, INCLUDE, (S1, S2...)) to provide SSM services.

NOTE:

The MLD SSM mapping feature does not process MLDv2 reports.

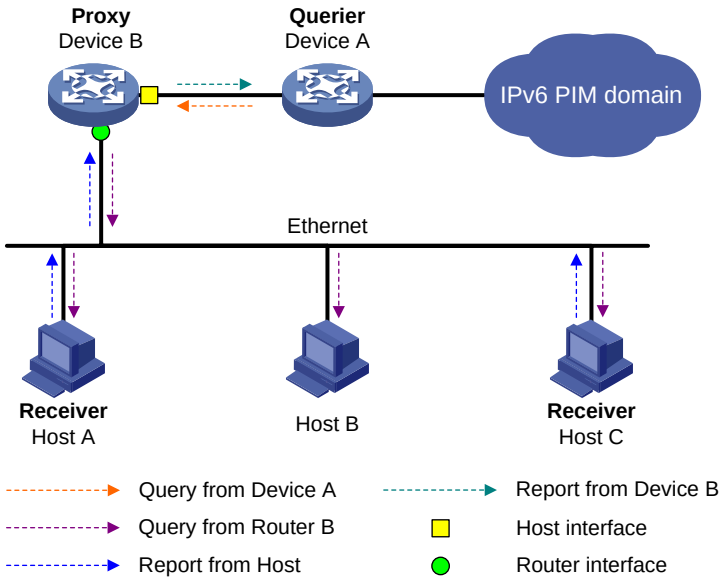
For more information about the IPv6 SSM group ranges, see "IPv6 PIM overview" and "Configuring IPv6 PIM."

MLD proxying

As shown in Figure 4, in a simple tree-shaped topology, it is not necessary to configure IPv6 multicast routing protocols, such as IPv6 PIM, on edge devices. Instead, you can configure MLD proxying on these devices. With MLD proxying configured, the edge device acts as an MLD proxy:

- For the upstream MLD querier, the MLD proxy device acts as a host.
- For the downstream receiver hosts, the MLD proxy device acts as an MLD querier.

Figure 4 Network diagram



The following interfaces are defined in MLD proxying:

- **Host interface**—An interface that is in the direction toward the root of the multicast forwarding tree. A host interface acts as a receiver host that is running MLD. MLD proxying must be enabled on this interface. This interface is also called the "proxy interface."
- **Router interface**—An interface that is in the direction toward the leaf of the multicast forwarding tree. A router interface acts as a router that is running MLD. MLD must be configured on this interface.

An MLD proxy device maintains a group membership database, which stores the group memberships on all the router interfaces. The host interfaces and router interfaces perform actions based on this membership database.

- The host interfaces respond to queries according to the membership database or sends join/done messages when the database changes.
- The router interfaces participate in the querier election, send queries, and maintain memberships based on received MLD reports.

MLD support for VPNs

MLD maintains group memberships on a per-interface basis. After receiving an MLD message on an interface, MLD processes the packet within the VPN to which the interface belongs. MLD only communicates with other multicast protocols within the same VPN instance.

Protocols and standards

- RFC 2710, *Multicast Listener Discovery (MLD) for IPv6*

- RFC 3810, *Multicast Listener Discovery Version 2 (MLDv2) for IPv6*

MLD tasks at a glance

To configure MLD, perform the following tasks:

1. Enabling MLD
2. (Optional.) Configuring basic MLD features
   - Specifying an MLD version
   - Configuring a static group member
   - Configuring an IPv6 multicast group policy
3. (Optional.) Adjusting MLD performance
   - Configuring MLD query and response parameters
   - Enabling fast-leave processing
4. (Optional.) Configuring MLD SSM mappings
5. (Optional.) Configuring MLD proxying
   - Enabling MLD proxying
   - Enabling IPv6 multicast forwarding on a non-querier interface
   - Enabling IPv6 multicast load splitting for MLD proxy interfaces
6. (Optional.) Enabling MLD NSR
7. (Optional.) Setting the DSCP value for outgoing MLD protocol packets
8. (Optional.) Enabling SNMP notifications for MLD
9. (Optional.) Configuring MLD logging
   - Enabling MLD logging
   - Enabling log message suppression for MLD

Prerequisites for MLD

Before you configure MLD, you must configure an IPv6 unicast routing protocol so that all devices can interoperate at the network layer.

Enabling MLD

Restrictions and guidelines

Perform this task on interfaces where IPv6 multicast group memberships are created and maintained.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 multicast routing and enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
   By default, IPv6 multicast routing is disabled.
   For more information about this command, see *IP Multicast Command Reference*.
3. Return to system view.
   quit

4. Enter interface view.
   interface *interface-type interface-number*
5. Enable MLD.
   mld enable
   By default, MLD is disabled.

Configuring basic MLD features

Specifying an MLD version

Restrictions and guidelines

For MLD to operate correctly, specify the same MLD version for all devices on the same subnet.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify an MLD version on the interface.
   mld version *version-number*
   By default, the MLD version on an interface is MLDv1.

Configuring a static group member

About this task

You can configure an interface as a static member of an IPv6 multicast group. Then, the interface can always receive IPv6 multicast data for the group.

A static group member does not respond to MLD queries. When you complete or cancel this configuration on an interface, the interface does not send an unsolicited MLD report or done message.

Restrictions and guidelines

The interface to be configured as a static member of an IPv6 multicast group has the following restrictions:

- If the interface is MLD and IPv6 PIM-SM enabled, it must be an IPv6 PIM-SM DR.
- If the interface is MLD enabled but not IPv6 PIM-SM enabled, it must be an MLD querier.

For more information about IPv6 PIM-SM and DR, see "IPv6 PIM overview."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure a static group member.
   mld static-group *ipv6-group-address* [**source** *ipv6-source-address*]
   By default, an interface is not a static group member.

Configuring an IPv6 multicast group policy

About this task

This feature enables an interface to filter MLD reports by using an ACL that specifies IPv6 multicast groups and the optional sources. It is used to control the IPv6 multicast groups that the hosts attached to an interface can join.

Restrictions and guidelines

This configuration does not take effect on static group members, because static group members do not send MLD reports.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure an IPv6 multicast group policy on the interface.

   mld group-policy { *ipv6-acl-number* | **name** *ipv6-acl-name* } [*version-number*]

   By default, the hosts attached to an interface can join any IPv6 multicast group policy.

Adjusting MLD performance

Configuring MLD query and response parameters

About this task

The following are MLD query and response parameters:

- **MLD querier's robustness variable**—Number of times for retransmitting MLD queries in case of packet loss. A higher robustness variable makes the MLD querier more robust, but increases the timeout time for IPv6 multicast groups.
- **MLD startup query interval**—Interval at which an MLD querier sends MLD general queries at startup.
- **MLD startup query count**—Number of MLD general queries that an MLD querier sends at startup.
- **MLD general query interval**—Interval at which an MLD querier sends MLD general queries to check for IPv6 multicast group members on the network.
- **MLD last listener query interval**—In MLDv1, it sets the interval at which a querier sends multicast-address-specific queries after receiving a done message. In MLDv2, it sets the interval at which a querier sends multicast-address-and-source-specific queries after receiving a report that changes IPv6 multicast source and group mappings.
- **MLD last listener query count**—In MLDv1, it sets the number of multicast-address-specific queries that the querier sends after receiving a done message. In MLDv2, it sets the number of multicast-address-and-source-specific queries that the querier sends after receiving a report that changes IPv6 multicast group and source mappings.
- **MLD maximum response time**—Maximum time before a receiver responds with a report to an MLD general query. This per-group timer is initialized to a random value in the range of 0 to the maximum response time specified in the MLD query. When the timer value decreases to 0, the receiver sends an MLD report to the group.

- **MLD other querier present timer**—Lifetime for an MLD querier after a non-querier receives an MLD general query. If the non-querier does not receive a new query when this timer expires, the non-querier considers that the querier has failed and starts a new querier election.

Restrictions and guidelines

- To avoid frequent MLD querier changes, set the MLD other querier present timer greater than the MLD general query interval. In addition, configure the same MLD other querier present timer for all MLD devices on the same subnet.
- To speed up the response to MLD queries and avoid simultaneous timer expirations that cause MLD report traffic bursts, you must set an appropriate maximum response time.
  - For MLD general queries, the maximum response time is set by the **max-response-time** command.
  - For MLD multicast-address-specific queries or MLD multicast-address-and-source-specific queries, the maximum response time equals the MLD last listener query interval.
- You can configure MLD query and response parameters globally for all interfaces in MLD view or for an interface in interface view. The interface-specific configuration takes priority over the global configuration.

Configuring the MLD query and response parameters globally

1. Enter system view.
   system-view
2. Enter MLD view.
   mld [**vpn-instance** *vpn-instance-name*]
3. Set the MLD querier's robustness variable.
   robust-count *count*
   By default, the MLD querier's robustness variable is 2.
4. Set the MLD startup query interval.
   startup-query-interval *interval*
   By default, the MLD startup query interval is equal to one quarter of the MLD general query interval.
5. Set the MLD startup query count.
   startup-query-count *count*
   By default, the MLD startup query count is equal to the MLD querier's robustness variable.
6. Set the MLD general query interval.
   query-interval *interval*
   By default, the MLD general query interval is 125 seconds.
7. Set the MLD last listener query interval.
   last-listener-query-interval *interval*
   By default, the MLD last listener query interval is 1 second.
8. Set the MLD last listener query count.
   last-listener-query-count *count*
   By default, the MLD last listener query count is equal to the MLD querier's robustness variable.
9. Set the maximum response time for MLD general queries.
   max-response-time *time*
   By default, the maximum response time for MLD general queries is 10 seconds.
10. Set the MLD other querier present timer.
    other-querier-present-timeout *time*

By default, the MLD other querier present timer is calculated by using the following formula: [MLD general query interval] × [MLD robustness variable] + [maximum response time for MLD general queries] / 2.

Configuring the MLD query and response parameters on an interface

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set the MLD querier's robustness variable.

   mld robust-count *count*

   By default, the MLD querier's robustness variable is 2.
4. Set the MLD startup query interval.

   mld startup-query-interval *interval*

   By default, the MLD startup query interval is equal to one quarter of the MLD general query interval.
5. Set the MLD startup query count.

   mld startup-query-count *count*

   By default, the MLD startup query count is equal to the MLD querier's robustness variable.
6. Set the MLD general query interval.

   mld query-interval *interval*

   By default, the MLD general query interval is 125 seconds.
7. Set the MLD last listener query interval.

   mld last-listener-query-interval *interval*

   By default, the MLD last listener query interval is 1 second.
8. Set the MLD last listener query count.

   mld last-listener-query-count *count*

   By default, the MLD last listener query count is equal to the MLD querier's robustness variable.
9. Set the maximum response time for MLD general queries.

   mld max-response-time *time*

   By default, the maximum response time for MLD general queries is 10 seconds.
10. Set the MLD other querier present timer.

    mld other-querier-present-timeout *time*

    By default, the MLD other querier present timer is calculated by using the following formula: [MLD general query interval] × [MLD robustness variable] + [maximum response time for MLD general queries] / 2.

Enabling fast-leave processing

About this task

This feature enables an MLD querier to send leave notifications to the upstream without sending multicast-address-specific or multicast-address-and-source-specific queries after receiving a done message. Use this feature to reduce leave latency and to preserve the network bandwidth.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Enable fast-leave processing.
   mld fast-leave [**group-policy** *ipv6-acl-number*]
   By default, fast-leave processing is disabled.

Configuring MLD SSM mappings

About this task

This feature enables the device to provide SSM services for MLDv1 hosts.

Restrictions and guidelines

This feature does not process MLDv2 messages. Enable MLDv2 on the receiver-side interface to ensure that MLDv2 reports from MLDv2 receiver hosts can be processed.

Procedure

1. Enter system view.
   system-view
2. Enter MLD view.
   mld [**vpn-instance** *vpn-instance-name*]
3. Configure an MLD SSM mapping.
   ssm-mapping *ipv6-source-address* { *ipv6-acl-number* | **name** *ipv6-acl-name* }
   By default, no MLD SSM mapping is configured.

Configuring MLD proxying

Prerequisites for MLD proxying

Before you configure MLD proxying, determine the router interfaces and host interface based on the network topology. Then, enable MLD on the router interfaces.

Enabling MLD proxying

Restrictions and guidelines

- If MLD proxying is enabled on multiple interfaces, only the interface with the highest IP address will forward IPv6 multicast traffic.
- Enable MLD proxying on the receiver-side interfaces.
- On an interface enabled with MLD proxying, only the **mld version** command takes effect and other MLD commands do not take effect.
- If you enable both MLD proxying and an IPv6 multicast routing protocol on the same device, the IPv6 multicast routing protocol does not take effect.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.

interface *interface-type interface-number*

3. Enable MLD proxying.

   mld proxy enable

   By default, MLD proxying is disabled.

Enabling IPv6 multicast forwarding on a non-querier interface

About this task

Typically, only MLD queriers can forward IPv6 multicast traffic and non-queriers cannot. This prevents IPv6 multicast data from being repeatedly forwarded. If a router interface on the MLD proxy failed the querier election, enable IPv6 multicast forwarding on the interface to forward IPv6 multicast data to downstream receivers.

Restrictions and guidelines

A shared-media network might have multiple MLD proxies, including one proxy acting as a querier. To avoid duplicate IPv6 multicast traffic, do not enable IPv6 multicast forwarding on any of the non-querier MLD proxies for the network.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable IPv6 multicast forwarding on a non-querier interface.

   mld proxy forwarding

   By default, IPv6 multicast forwarding is disabled on a non-querier interface.

Enabling IPv6 multicast load splitting for MLD proxy interfaces

About this task

If multiple MLD proxy interfaces exist on the device, only the proxy interface with the highest IP address forwards IPv6 multicast traffic. You can enable IPv6 multicast load splitting on the device so that all the proxy interfaces can share IPv6 multicast traffic.

Procedure

1. Enter system view.

   system-view

2. Enter MLD view.

   mld [**vpn-instance** *vpn-instance-name*]

3. Enable IPv6 multicast load splitting for MLD proxy interfaces.

   proxy multipath

   By default, IPv6 multicast load splitting is disabled for MLD proxy interfaces.

Enabling MLD NSR

About this task

This feature backs up information about MLD interfaces and MLD multicast groups to the standby process. The device recovers the information without cooperation of other devices when an active/standby switchover occurs. Use this feature to prevent an active/standby switchover from affecting the IPv6 multicast service.

Procedure

1. Enter system view.
   system-view
2. Enable MLD NSR.
   mld non-stop-routing
   By default, MLD NSR is disabled.

Setting the DSCP value for outgoing MLD protocol packets

About this task

The DSCP value determines the packet transmission priority. A greater DSCP value represents a higher priority.

Procedure

1. Enter system view.
   system-view
2. Enter MLD view.
   mld [**vpn-instance** *vpn-instance-name*]
3. Set the DSCP value for outgoing MLD protocol packets.
   dscp *dscp-value*
   By default, the DSCP value is 48 for outgoing MLD protocol packets.

Enabling SNMP notifications for MLD

About this task

To report critical MLD events to an SNMP NMS, enable SNMP notifications for MLD.

For MLD SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about configuring SNMP, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for MLD.
   snmp-agent trap enable mld [**join** | **leave**] *
   By default, SNMP notifications are enabled for MLD.

Configuring MLD logging

Enabling MLD logging

About this task

To report critical MLD events to the information center, enable MLD logging. For MLD log messages to be sent correctly, you must also configure the information center on the device. For more information about configuring the information center, see Network Management and Monitoring Configuration Guide.

Restrictions and guidelines

Outputting a large amount of log messages degrades the device performance. To resolve this issue, disable MLD logging or use the **mld log suppress** command to enable log message suppression for MLD.

Procedure

1. Enter system view.
   system-view
2. Enable MLD logging.
   mld log enable [**group**]
   MLD logging is enabled by default.

Enabling log message suppression for MLD

About this task

Outputting a large amount of log messages degrades the device performance. To resolve this issue, perform this task to enable log message suppression for MLD. With this feature configured, the speed of sending MLD log messages to the information center will be decreased to reduce the load on the device.

Restrictions and guidelines

Before you configure this feature, make sure MLD logging has been enabled using the **mld log enable** command.

Procedure

1. Enter system view.
   system-view
2. Enable log message suppression for MLD.
   mld log suppress [**group**]
   By default, log message suppression is enabled for MLD.

Display and maintenance commands for MLD

CAUTION:
The **reset mld group** command might cause IPv6 multicast data transmission failures.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about MLD multicast groups.	**display mld** [**vpn-instance** *vpn-instance-name*] **group** [*ipv6-group-address* | **interface** *interface-type interface-number*] [**static** | **verbose**]
Display MLD information for interfaces.	**display mld** [**vpn-instance** *vpn-instance-name*] **interface** [*interface-type interface-number*] [**proxy**] [**verbose**]
Display IPv6 multicast routing entries maintained by the MLD proxy.	**display mld** [**vpn-instance** *vpn-instance-name*] **proxy group** [*ipv6-group-address* | **interface** *interface-type interface-number*] [**verbose**]
Display information about the MLD proxy routing table.	**display mld** [**vpn-instance** *vpn-instance-name*] **proxy routing-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*]] * [**verbose**]
Display MLD SSM mappings.	**display mld** [**vpn-instance** *vpn-instance-name*] **ssm-mapping** *ipv6-group-address*
Clear dynamic MLD multicast group entries.	**reset mld** [**vpn-instance** *vpn-instance-name*] **group** { **all** | **interface** *interface-type interface-number* { **all** | *ipv6-group-address* [*prefix-length*] [*ipv6-source-address* [*prefix-length*]] } }

MLD configuration examples

Example: Configuring basic MLD features

Network configuration

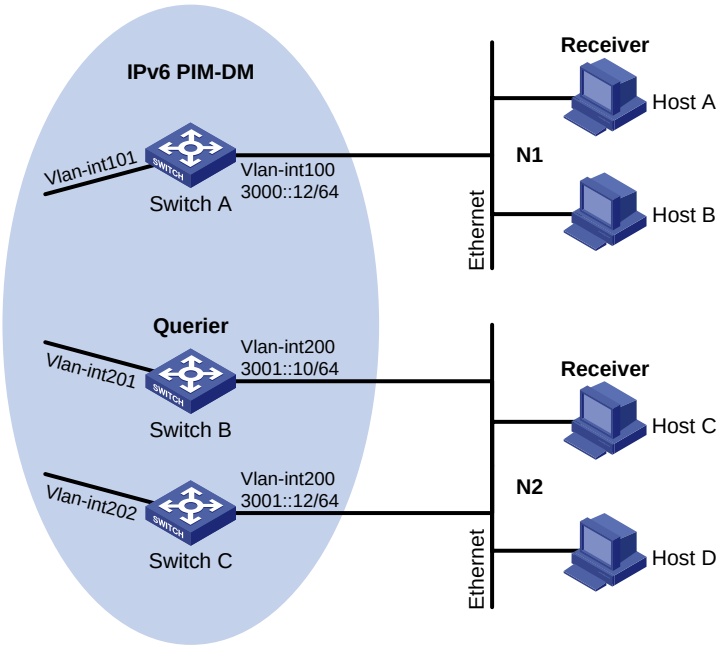
As shown in Figure 5:

- OSPFv3 and IPv6 PIM-DM run on the network.
- VOD streams are sent to receiver hosts in multicast. Receiver hosts of different organizations form stub networks N1 and N2. Host A and Host C are multicast receiver hosts in N1 and N2, respectively.
- MLDv1 runs between Switch A and N1, and between the other two switches (Switch B and Switch C) and N2.
- Switch A acts as the MLD querier in N1. Switch B acts as the MLD querier in N2 because it has a lower IPv6 address.

Configure the switches to meet the following requirements:

- The hosts in N1 can join only IPv6 multicast group FF1E::101.
- The hosts in N2 can join any IPv6 multicast groups.

Figure 5 Network diagram



Prerequisites

Assign an IPv6 address and prefix length to each interface as shown in Figure 5, and configure OSPFv3 on the switches in the IPv6 PIM-DM domain.

Procedure

1. Enable the IPv6 multicast routing, MLD, and IPv6 PIM-DM:

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Enable MLD on VLAN-interface 100.

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] mld enable
   [SwitchA-Vlan-interface100] quit
   ```

   # Enable IPv6 PIM-DM on VLAN-interface 101.

   ```
   [SwitchA] interface vlan-interface 101
   [SwitchA-Vlan-interface101] ipv6 pim dm
   [SwitchA-Vlan-interface101] quit
   ```

   # On Switch B, enable IPv6 multicast routing.

   ```
   <SwitchB> system-view
   [SwitchB] ipv6 multicast routing
   [SwitchB-mrib6] quit
   ```

   # Enable MLD on VLAN-interface 200.

   ```
   [SwitchB] interface vlan-interface 200
   [SwitchB-Vlan-interface200] mld enable
   [SwitchB-Vlan-interface200] quit
   ```

Enable IPv6 PIM-DM on VLAN-interface 201.

```
[SwitchB] interface vlan-interface 201
[SwitchB-Vlan-interface201] ipv6 pim dm
[SwitchB-Vlan-interface201] quit
```

On Switch C, enable IPv6 multicast routing.

```
<SwitchC> system-view
[SwitchC] ipv6 multicast routing
[SwitchC-mrib6] quit
```

Enable MLD on VLAN-interface 200.

```
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] mld enable
[SwitchC-Vlan-interface200] quit
```

Enable IPv6 PIM-DM on VLAN-interface 202.

```
[SwitchC] interface vlan-interface 202
[SwitchC-Vlan-interface202] ipv6 pim dm
[SwitchC-Vlan-interface202] quit
```

2. Configure an IPv6 multicast group policy on Switch A so that the hosts connected to VLAN-interface 100 can join only IPv6 multicast group FF1E::101.

```
[SwitchA] acl ipv6 basic 2001
[SwitchA-acl-ipv6-basic-2001] rule permit source ff1e::101 128
[SwitchA-acl-ipv6-basic-2001] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld group-policy 2001
[SwitchA-Vlan-interface100] quit
```

Verifying the configuration

Display MLD information for VLAN-interface 200 on Switch B.

```
[SwitchB] display mld interface vlan-interface 200
 Vlan-interface200(FE80::200:5EFF:FE66:5100):
   MLD is enabled.
   MLD version: 1
   Query interval for MLD: 125s
   Other querier present time for MLD: 255s
   Maximum query response time for MLD: 10s
   Querier for MLD: FE80::200:5EFF:FE66:5100 (This router)
  MLD groups reported in total: 1
```

Example: Configuring MLD SSM mapping

Network configuration

As shown in Figure 6:

- OSPFv3 runs on the network.
- The IPv6 PIM-SM domain uses the SSM model for IPv6 multicast delivery. The SSM group range is FF3E::/64.
- MLDv2 runs on VLAN-interface 400 of Switch D. The receiver host runs MLDv1, and does not support MLDv2. The receiver host cannot specify IPv6 multicast sources in its membership reports.

- Source 1, Source 2, and Source 3 send IPv6 multicast packets to multicast groups in IPv6 SSM group range FF3E::/64.

Configure the MLD SSM mapping feature on Switch D so that the receiver host will receive IPv6 multicast data only from Source 1 and Source 3.

Figure 6 Network diagram

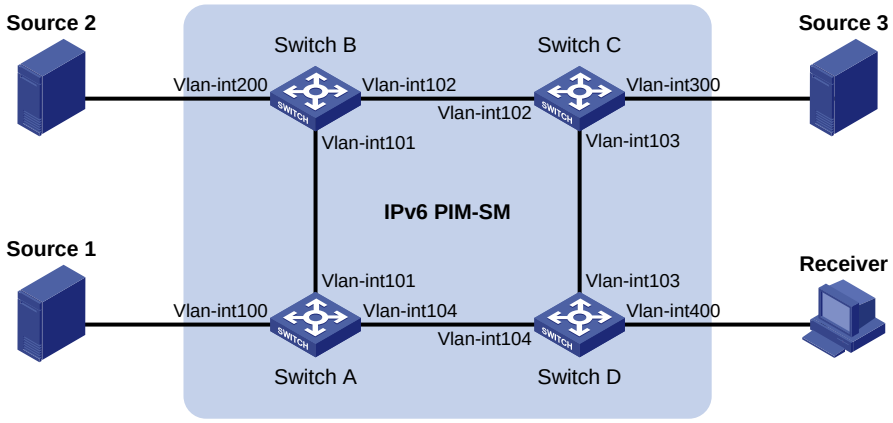


Table 1 Interface and IPv6 address assignment

Device	Interface	IPv6 address	Device	Interface	IPv6 address
Source 1	—	1001::1/64	Source 3	—	3001::1/64
Source 2	—	2001::1/64	Receiver	—	4001::1/64
Switch A	Vlan-int100	1001::2/64	Switch C	Vlan-int300	3001::2/64
Switch A	Vlan-int101	1002::1/64	Switch C	Vlan-int103	3002::1/64
Switch A	Vlan-int104	1003::1/64	Switch C	Vlan-int102	2002::2/64
Switch B	Vlan-int200	2001::2/64	Switch D	Vlan-int400	4001::2/64
Switch B	Vlan-int101	1002::2/64	Switch D	Vlan-int103	3002::2/64
Switch B	Vlan-int102	2002::1/64	Switch D	Vlan-int104	1003::2/64

Prerequisites

Assign an IPv6 address and prefix length to each interface as shown in Table 1, and configure OSPFv3 on the switches in the IPv6 PIM-SM domain.

Procedure

1. Enable IPv6 multicast routing, and enable IPv6 PIM-SM and MLD:

   # On Switch D, enable IPv6 multicast routing.

   ```
   <SwitchD> system-view
   [SwitchD] ipv6 multicast routing
   [SwitchD-mrib6] quit
   ```

   # Enable MLDv2 on the receiver-side interface (VLAN-interface 400).

   ```
   [SwitchD] interface vlan-interface 400
   [SwitchD-Vlan-interface400] mld enable
   [SwitchD-Vlan-interface400] mld version 2
   [SwitchD-Vlan-interface400] quit
   ```

   # Enable IPv6 PIM-SM on the other interfaces.

```
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] ipv6 pim sm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 104
[SwitchD-Vlan-interface104] ipv6 pim sm
[SwitchD-Vlan-interface104] quit
```

On Switch A, enable IPv6 multicast routing.

```
<SwitchA> system-view
[SwitchA] ipv6 multicast routing
[SwitchA-mrib6] quit
```

Enable IPv6 PIM-SM on each interface.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipv6 pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 104
[SwitchA-Vlan-interface104] ipv6 pim sm
[SwitchA-Vlan-interface104] quit
```

Configure Switch B and Switch C in the same way Switch A is configured. (Details not shown.)

2. Configure the IPv6 SSM group range:

On Switch D, specify FF3E::/64 as the IPv6 SSM group range.

```
[SwitchD] acl ipv6 basic 2000
[SwitchD-acl-ipv6-basic-2000] rule permit source ff3e:: 64
[SwitchD-acl-ipv6-basic-2000] quit
[SwitchD] pim ipv6
[SwitchD-pim6] ssm-policy 2000
[SwitchD-pim6] quit
```

Configure Switch A, Switch B and Switch C in the same way Switch D is configured. (Details not shown.)

3. Configure MLD SSM mappings on Switch D.

```
[SwitchD] mld
[SwitchD-mld] ssm-mapping 1001::1 2000
[SwitchD-mld] ssm-mapping 3001::1 2000
[SwitchD-mld] quit
```

Verifying the configuration

Display MLD SSM mappings for IPv6 multicast group FF3E::101 on Switch D.

```
[SwitchD] display mld ssm-mapping ff3e::101
 Group: FF3E::101
 Source list:
        1001::1
        3001::1
```

On Switch D, display information about MLD multicast groups that hosts have dynamically joined.

```
[SwitchD] display mld group
MLD groups in total: 1
 Vlan-interface400(FE80::101):
```

```
  MLD groups reported in total: 1
   Group address: FF3E::101
    Last reporter: FE80::1
    Uptime: 00:02:04
    Expires: Off
```

Display IPv6 PIM routing entries on Switch D.

```
[SwitchD] display ipv6 pim routing-table
 Total 0 (*, G) entry; 2 (S, G) entry

 (1001::1, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface104
         Upstream neighbor: FE80::1
         RPF prime neighbor: FE80::1
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface400
             Protocol: mld, UpTime: 00:13:25, Expires: -

 (3001::1, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface103
         Upstream neighbor: FE80::1
         RPF prime neighbor: FE80::1
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface400
             Protocol: mld, UpTime: 00:13:25, Expires: -
```

Example: Configuring MLD proxying

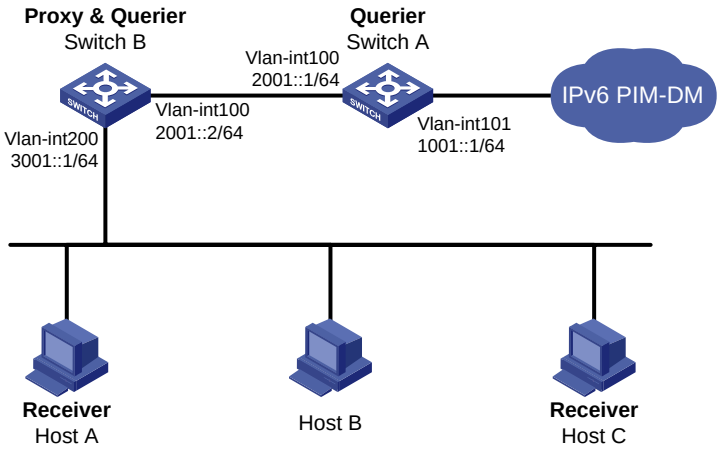
Network configuration

As shown in Figure 7:

- IPv6 PIM-DM runs on the core network.
- Host A and Host C on the stub network receive VOD information destined to IPv6 multicast group FF3E::101.

Configure the MLD proxying feature on Switch B so that Switch B can maintain group memberships and forward IPv6 multicast traffic without running IPv6 PIM-DM.

Figure 7 Network diagram



Prerequisites

Assign an IPv6 address and prefix length to each interface, as shown in Figure 7.

Procedure

1. Enable IPv6 multicast routing, IPv6 PIM-DM, MLD, and MLD proxying:

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Enable IPv6 PIM-DM on VLAN-interface 101.

   ```
   [SwitchA] interface vlan-interface 101
   [SwitchA-Vlan-interface101] ipv6 pim dm
   [SwitchA-Vlan-interface101] quit
   ```

   # Enable MLD on VLAN-interface 100.

   ```
   [SwitchA] interface vlan-interface 100
   [SwitchA-Vlan-interface100] mld enable
   [SwitchA-Vlan-interface100] quit
   ```

   # On Switch B, enable IPv6 multicast routing.

   ```
   <SwitchB> system-view
   [SwitchB] multicast ipv6 routing-enable
   [SwitchB-mrib6] quit
   ```

   # Enable MLD proxying on VLAN-interface 100.

   ```
   [SwitchB] interface vlan-interface 100
   [SwitchB-Vlan-interface100] mld proxying enable
   [SwitchB-Vlan-interface100] quit
   ```

   # Enable MLD on VLAN-interface 200.

   ```
   [SwitchB] interface vlan-interface 200
   [SwitchB-Vlan-interface200] mld enable
   [SwitchB-Vlan-interface200] quit
   ```

Verifying the configuration

Display IPv6 multicast group membership information maintained by the MLD proxy on Switch B.

```
[SwitchB] display mld proxy group
MLD proxy group records in total: 1
```

```
Vlan-interface100(FE80::16:1):
 MLD proxy group records in total: 1
  Group address: FF1E::1
   Member state: Delay
   Expires: 00:00:02
```

Troubleshooting MLD

No member information exists on the receiver-side device

Symptom

When a host sends a message to announce that it is joining IPv6 multicast group G, no member information of multicast group G exists on the immediate device.

Solution

To resolve the problem:

1. Use the **`display mld interface`** command to verify that the networking, interface connections, and IP address configuration are correct.
2. Use the **`display current-configuration`** command to verify that the IPv6 multicast routing is enabled. If it is not enabled, use the **`ipv6 multicast routing`** command in system view to enable IPv6 multicast routing. In addition, verify that MLD is enabled on the associated interfaces.
3. Use the **`display mld interface`** command to verify that the MLD version on the interface is lower than that on the host.
4. Use the **`display current-configuration interface`** command to verify that no IPv6 multicast group policies have been configured to filter MLD reports for IPv6 multicast group G.
5. If the problem persists, contact H3C Support.

Inconsistent membership information on the devices on the same subnet

Symptom

Different memberships are maintained on different MLD devices on the same subnet.

Solution

To resolve the problem:

1. Use the **`display current-configuration`** command to verify the MLD information on the interface. Make sure the devices on the subnet have the same MLD settings on their interfaces.
2. Use the **`display mld interface`** command on all devices on the same subnet to check the MLD timers for inconsistent configurations.
3. Use the **`display mld interface`** command to verify that all devices are running the same MLD version.
4. If the problem persists, contact H3C Support.

Contents

Configuring multicast routing and forwarding 1

About multicast routing and forwarding 1
RPF check mechanism 1
Usages of static multicast routes 3
Mtrace 4
Restrictions and guidelines: Multicast routing and forwarding configuration 5
Multicast routing and forwarding tasks at a glance 5
Prerequisites for multicast routing and forwarding 5
Enabling IP multicast routing 5
Configuring static multicast routes 6
Specifying the longest prefix match principle 6
Configuring multicast load splitting 6
Configuring an IPv4 multicast flow policy 7
About this task 7
Restrictions and guidelines for configuring an IPv4 multicast flow policy 7
Creating an IPv4 multicast flow policy 7
Configuring the multicast group range 8
Configuring the estimated bandwidth 8
Configuring the IPv4 unicast reserved bandwidth percentage 9
Configuring a multicast forwarding boundary 9
Using mtrace to trace a multicast path 10
Using mtrace1 to trace a multicast path 10
Using mtrace2 to trace a multicast path 10
Setting the maximum number of cached unknown multicast packets 10
Setting the maximum number of copied multicast packets during software forwarding 11
Enabling multicast replication resource reuse 11
Display and maintenance commands for multicast routing and forwarding 12
Multicast routing and forwarding configuration examples 14
Example: Changing an RPF route 14
Example: Creating an RPF route 16
Troubleshooting multicast routing and forwarding 18
Static multicast route failure 18

Configuring multicast routing and forwarding

About multicast routing and forwarding

Each multicast routing protocol has its own routing table. Multicast routing information in routing entries generated by the multicast routing protocols and statically configured multicast routing entries are summarized in a set of (S, G) and (*, G) entries. All the (S, G) and (*, G) entries form a general multicast routing table. The optimal multicast routing entries in the general multicast routing table are added to the multicast forwarding table to guide multicast data forwarding.

RPF check mechanism

A multicast routing protocol uses reverse path forwarding (RPF) check to ensure the multicast data delivery along the correct path and to avoid data loops.

RPF check process

A multicast device performs the RPF check on a multicast packet as follows:

1. Chooses an optimal route back to the packet source separately from the unicast, MBGP, and static multicast routing tables.

   The term "packet source" means different things in different situations:

   - For a packet that travels along the SPT, the packet source is the multicast source.
   - For a packet that travels along the RPT, the packet source is the RP.
   - For a bootstrap message originated from the BSR, the packet source is the BSR.

   For more information about the concepts of SPT, RPT, source-side RPT, RP, and BSR, see "PIM overview."
2. Selects one of the three optimal routes as the RPF route as follows:
   - If the device uses the longest prefix match principle, the route with the highest subnet mask becomes the RPF route. If the routes have the same mask, the route with the highest route preference becomes the RPF route. If the routes have the same route preference, the unicast route becomes the RPF route. If equal cost routes exist, the equal cost route with the highest next hop IP address becomes the RPF route.

     For more information about the route preference, see *Layer 3—IP Routing Configuration Guide*.
   - If the device does not use the longest prefix match principle, the route with the highest route preference becomes the RPF route. If the routes have the same preference, the unicast route becomes the RPF route. If equal cost routes exist, the equal cost route with the highest next hop IP address becomes the RPF route.

   The RPF route contains the RPF interface and RPF neighbor information.
   - If the RPF route is a unicast route or MBGP route, the outgoing interface is the RPF interface and the next hop is the RPF neighbor.
   - If the RPF route is a static multicast route, the RPF interface and RPF neighbor are specified in the route.
3. Determines whether the packet arrived at the RPF interface.
   - If the packet arrived at the RPF interface, the RPF check succeeds and the packet is forwarded.

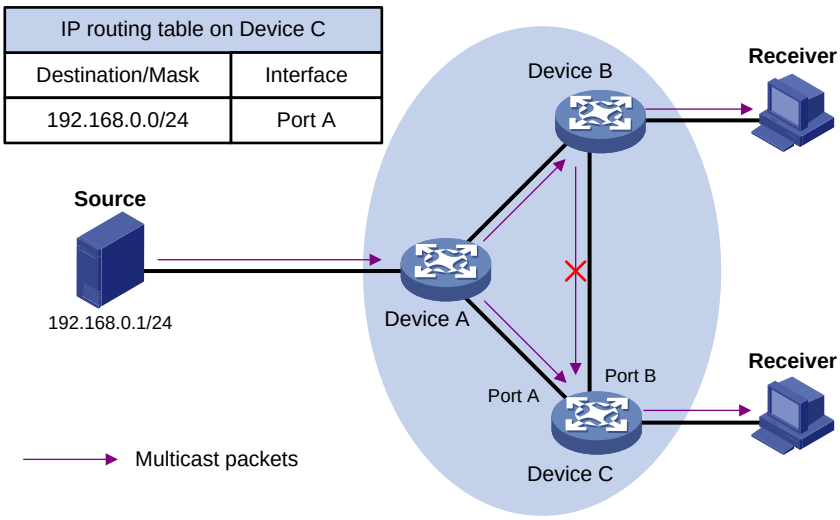
- If the packet arrived at the non-RPF interface, the RPF check fails and the packet is discarded.

RPF check implementation in multicast

Implementing an RPF check on each received multicast packet brings a big burden to the device. The use of a multicast forwarding table is the solution to this issue. When the device creates a multicast forwarding entry for an (S, G) packet, it sets the RPF interface of the packet as the incoming interface of the (S, G) entry. After the device receives another (S, G) packet, it looks up the multicast forwarding table for a matching (S, G) entry.

- If no match is found, the device first determines the RPF route back to the packet source and the RPF interface. Then, it creates a forwarding entry with the RPF interface as the incoming interface and makes the following judgments:
  - If the receiving interface is the RPF interface, the RPF check succeeds and the device forwards the packet out of all the outgoing interfaces.
  - If the receiving interface is not the RPF interface, the RPF check fails and the device discards the packet.
- If a match is found and the matching forwarding entry contains the receiving interface, the device forwards the packet out of all the outgoing interfaces.
- If a match is found but the matching forwarding entry does not contain the receiving interface, the device determines the RPF route back to the packet source. Then, the device performs one of the following actions:
  - If the RPF interface is the incoming interface, it means that the forwarding entry is correct but the packet traveled along a wrong path. The packet fails the RPF check, and the device discards the packet.
  - If the RPF interface is not the incoming interface, it means that the forwarding entry has expired. The device replaces the incoming interface with the RPF interface and matches the receiving interface against the RPF interface. If the receiving interface is the RPF interface, the device forwards the packet out of all outgoing interfaces. Otherwise, it discards the packet.

Figure 1 RPF check process



As shown in Figure 1, assume that unicast routes are available on the network, MBGP is not configured, and no static multicast routes have been configured on Device C. Multicast packets travel along the SPT from the multicast source to the receivers. The multicast forwarding table on Device C contains the (S, G) entry, with Port A as the incoming interface.

- If a multicast packet arrives at Device C on Port A, the receiving interface is the incoming interface of the (S, G) entry. Device C forwards the packet out of all outgoing interfaces.

- If a multicast packet arrives at Device C on Port B, the receiving interface is not the incoming interface of the (S, G) entry. Device C searches its unicast routing table and finds that the outgoing interface to the source (the RPF interface) is Port A. In this case, the (S, G) entry is correct, but the packet traveled along a wrong path. The packet fails the RPF check and Device C discards the packet.

Usages of static multicast routes

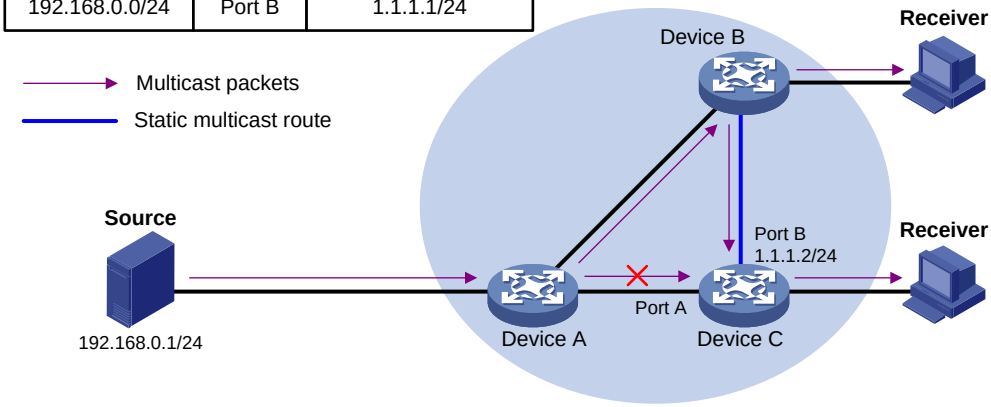
A static multicast route can change an RPF route or create an RPF route.

Changing an RPF route

Typically, the topology structure of a multicast network is the same as that of a unicast network, and multicast traffic follows the same transmission path as unicast traffic does. You can configure a static multicast route for a multicast source to change the RPF route. As a result, the device creates a transmission path for multicast traffic that is different from the transmission path for unicast traffic.

Figure 2 Changing an RPF route

Static multicast routing table on Device C		
Source/Mask	Interface	RPF neighbor/Mask
192.168.0.0/24	Port B	1.1.1.1/24



As shown in Figure 2, when no static multicast route is configured, Device C's RPF neighbor on the path back to the source is Device A. The multicast data from the source travels through Device A to Device C. You can configure a static multicast route on Device C and specify Device B as Device C's RPF neighbor on the path back to the source. The multicast data from the source travels along the path: Device A to Device B and then to Device C.

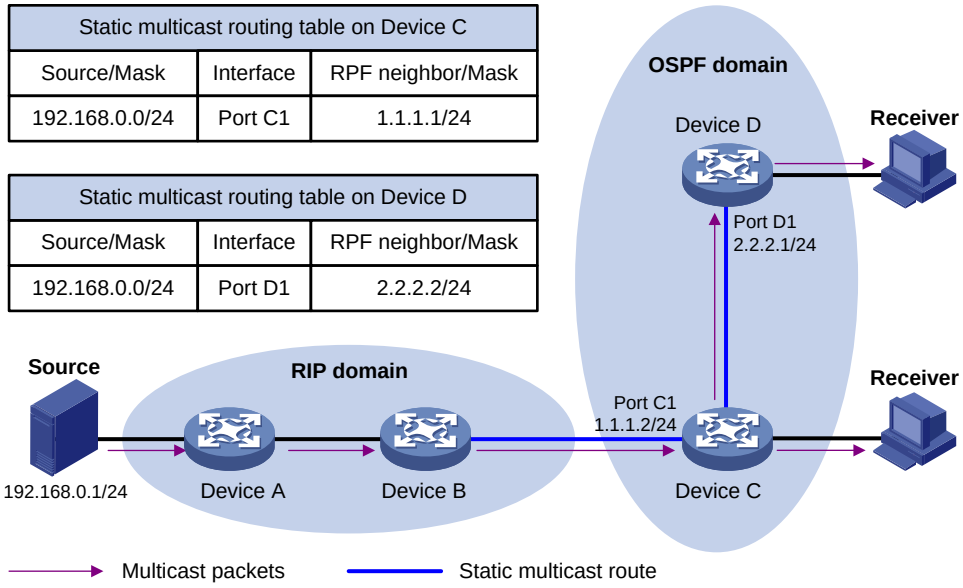
Creating an RPF route

When a unicast route is blocked, multicast forwarding might be stopped due to lack of an RPF route. You can configure a static multicast route to create an RPF route. In this way, a multicast routing entry is created to guide multicast forwarding.

Figure 3 Creating an RPF route

Static multicast routing table on Device C		
Source/Mask	Interface	RPF neighbor/Mask
192.168.0.0/24	Port C1	1.1.1.1/24

Static multicast routing table on Device D		
Source/Mask	Interface	RPF neighbor/Mask
192.168.0.0/24	Port D1	2.2.2.2/24



As shown in Figure 3, the RIP domain and the OSPF domain are unicast isolated from each other. For the receiver hosts in the OSPF domain to receive multicast packets from the multicast source in the RIP domain, you must configure Device C and Device D as follows:

- On Device C, configure a static multicast route for the multicast source and specify Device B as the RPF neighbor.
- On Device D, configure a static multicast route for the multicast source and specify Device C as the RPF neighbor.

Mtrace

Mtrace is a traceroute facility used to trace the path along which multicast group data travels from a source to a destination.

Device roles

Mtrace includes the following roles:

- **Last-hop router (LHR)**—An LHR is a router that has a multicast-enabled interface on the same subnet as the destination and can forward specific multicast data to the subnet.
- **First-hop router (FHR)**—An FHR is a router that is directly connected to the multicast source.
- **Client**—A client is a router that initiates an mtrace.

Process

The mtrace process is as follows:

1. The client sends an mtrace Query message (with a **hops** field indicating the maximum number of hops to trace) to the destination.
2. The LHR turns the received Query message to an mtrace Request message by adding local forwarding information and sends the Request message to the upstream neighbor.
3. Each router along the traced path adds its local forwarding information to the received Request message and sends the Request message to its upstream neighbor.
4. The FHR adds its local forwarding information to the received Request message. Then, it turns the Request message to an mtrace Reply message and sends the Reply message to the client.
5. The client interprets forwarding information in the Reply message and displays the information.

If the client does not receive a Reply message within the waiting time, the client initiates a hop-by-hop trace to determine which device on the path encountered an error. It sends a Query message with the **hops** field set to 1 and waits for a Reply message. If it does not receive a Reply message within the waiting time, the client determines that this hop encountered an error. If the client receives a Reply message within the waiting time, it sends a Query message with the **hops** field value increased by 1 and waits for a Reply message. This process continues until the client does not receive a Reply message within the waiting time any more.

Restrictions and guidelines: Multicast routing and forwarding configuration

During ISSU on a multichassis IRF, if the incoming interface of Layer 3 multicast traffic contains non-local member ports, Layer 3 multicast traffic interruption might occur when the device is rebooting.

Multicast routing and forwarding tasks at a glance

To configure multicast routing and forwarding, perform the following tasks:

1. Enabling IP multicast routing
2. (Optional.) Configuring static multicast routes
3. (Optional.) Specifying the longest prefix match principle
4. (Optional.) Configuring multicast load splitting
5. (Optional.) Configuring a multicast forwarding boundary
6. (Optional.) Using mtrace to trace a multicast path
7. (Optional.) Setting the maximum number of cached unknown multicast packets
8. (Optional.) Setting the maximum number of copied multicast packets during software forwarding
9. (Optional.) Enabling multicast replication resource reuse

Prerequisites for multicast routing and forwarding

Before you configure multicast routing and forwarding, configure a unicast routing protocol so that all devices in the domain can interoperate at the network layer.

Enabling IP multicast routing

About this task

Enable IP multicast routing before you configure any Layer 3 multicast functionality on the public network or VPN instance.

Procedure

1. Enter system view.
   system-view
2. Enable IP multicast routing and enter MRIB view.
   multicast routing [**vpn-instance** *vpn-instance-name*]
   By default, IP multicast routing is disabled.

Configuring static multicast routes

About this task

To configure a static multicast route for a multicast source, you can specify an RPF interface or an RPF neighbor for the multicast traffic from that source.

Restrictions and guidelines

Static multicast routes take effect only on the multicast devices on which they are configured, and will not be advertised throughout the network or redistributed to other devices.

Procedure

1. Enter system view.

   system-view

2. Configure a static multicast route.

   ip rpf-route-static [**vpn-instance** *vpn-instance-name*] *source-address* { *mask-length* | *mask* } { *gateway-address* | *interface-type interface-number* } [**preference** *preference*]

3. (Optional.) Delete all static multicast routes.

   delete ip rpf-route-static [**vpn-instance** *vpn-instance-name*]

   You can use the **undo ip rpf-route-static** command to delete a specific static multicast route or use the **delete ip rpf-route-static** command to delete all static multicast routes.

Specifying the longest prefix match principle

About this task

You can enable the device to use the longest prefix match principle for RPF route selection. For more information about RPF route selection, see "RPF check mechanism."

Procedure

1. Enter system view.

   system-view

2. Enter MRIB view.

   multicast routing [**vpn-instance** *vpn-instance-name*]

3. Specify the longest prefix match principle.

   longest-match

   By default, the route preference principle is used.

Configuring multicast load splitting

About this task

You can enable the device to split multiple data flows on a per-source basis or on a per-source-and-group basis. This optimizes the traffic delivery.

Procedure

1. Enter system view.

   system-view

2. Enter MRIB view.

   multicast routing [**vpn-instance** *vpn-instance-name*]

3. Configure multicast load splitting.

   load-splitting { **balance-ecmp** | **balance-ucmp** | **ecmp** | **flow-ucmp** | **source** | **source-group** | **ucmp** }

   By default, multicast load splitting is disabled.

Configuring an IPv4 multicast flow policy

About this task

You can configure the estimated bandwidth for matching multicast flows through an IPv4 multicast flow policy on a downstream device if the following conditions exist:

- ECMP routes between the downstream device and an upstream device.
- The multicast load-splitting mode is **flow-ucmp**.

When the downstream device selects an incoming interface on the upstream device, it selects the link with the smallest bandwidth usage calculated based on the estimated bandwidth. If two links have the same bandwidth usage, the link with the higher next-hop IP address is selected.

The bandwidth usage is calculated according to the following formula: Bandwidth usage = (Used bandwidth+Estimated bandwidth)/(Total interface bandwidth x (1–Unicast reserved bandwidth)).

- The estimated bandwidth is configured by the **bandwidth** command in IPv4 multicast flow policy view.
- The total interface bandwidth is configured by the **bandwidth** command in interface view.
- The unicast reserved bandwidth is configured by the **flow-ucmp unicast reserve-bandwidth** or **multicast flow-ucmp unicast reserve-bandwidth** command.

Restrictions and guidelines for configuring an IPv4 multicast flow policy

For this feature to take effect, you must configure the multicast load-splitting mode as **flow-ucmp**.

Adding, deleting, or modifying an IPv4 multicast flow policy does not affect the link selection results of existing multicast flows and affects link selection for only new multicast flows. Before configuring the multicast load-splitting mode as **flow-ucmp**, you must plan the configuration of an IPv4 multicast flow policy and the unicast reserved bandwidth configuration.

The change of link bandwidth and the change in the number of ECMP routes do not affect the route selection of existing multicast flows.

This feature does not affect multicast source-side link selection, link selection in PIM-DM, RPT link selection in PIM-SM, or local RP link selection.

If both IPv4 and IPv6 multicast flows exist in the network, increase the unicast reserved bandwidth as needed to avoid congestion.

Creating an IPv4 multicast flow policy

1. Enter system view.

   system-view

2. Enter MRIB view.
   multicast routing [**vpn-instance** *vpn-instance-name*]
3. Create an IPv4 multicast flow policy.
   - Create a custom IPv4 multicast flow policy.
     flow-policy name *policy-name*
   - Create the default IPv4 multicast flow policy.
     flow-policy default

Configuring the multicast group range

About this task

Perform this task to configure a multicast group range in which an IPv4 multicast flow policy takes effect. If a multicast group matches the ACL specified in an IPv4 multicast flow policy, the estimated bandwidth configured in the policy is used for the multicast flows of the multicast group.

Restrictions and guidelines

This feature is not supported by the default IPv4 multicast flow policy.

Procedure

1. Enter system view.
   system-view
2. Enter MRIB view.
   multicast routing [**vpn-instance** *vpn-instance-name*]
3. Enter IPv4 multicast flow policy view.
   flow-policy { **default** | **name** *policy-name* }
4. Configure the multicast group range.
   acl { *ipv4-acl-number* | **name** *ipv4-acl-name* }
   By default, an IPv4 multicast flow policy does not take effect on any multicast groups.

Configuring the estimated bandwidth

About this task

The estimated bandwidth is selected for IPv4 multicast flows as follows:

1. If the multicast group of an IPv4 multicast flow matches the ACL specified in a custom IPv4 multicast flow policy, the estimated bandwidth configured in the policy is selected for the multicast flow.
2. If the multicast group of an IPv4 multicast flow does not match an ACL in any custom IPv4 multicast flow policy, the estimated bandwidth configured in the default policy is selected for the multicast flow.
3. If the default policy is not created or the estimated bandwidth is not configured in the default policy, the estimated bandwidth for the multicast flow is 0.

Procedure

1. Enter system view.
   system-view
2. Enter MRIB view.
   multicast routing [**vpn-instance** *vpn-instance-name*]
3. Enter IPv4 multicast flow policy view.

flow-policy { **default** | **name** *policy-name* }

4. Configure the estimated bandwidth.
   bandwidth *bandwidth* { **gbps** | **kbps** | **mbps** }
   By default, the estimated bandwidth is 0.

Configuring the IPv4 unicast reserved bandwidth percentage

About this task

Perform this task when IPv4 unicast traffic and IPv4 multicast traffic coexist in the network. The available bandwidth for multicast traffic is calculated based on the total interface bandwidth and the configured IPv4 unicast reserved bandwidth percentage. For example, if the total interface bandwidth is 100 kbps and the unicast reserved bandwidth percentage is 20% (20 kbps), the available bandwidth for multicast traffic is 80 kbps.

Restrictions and guidelines

For this feature to take effect, you must configure the multicast load-splitting mode as **flow-ucmp**.

The **flow-ucmp unicast reserve-bandwidth** command takes effect on all interfaces, and the **multicast flow-ucmp unicast reserve-bandwidth** command takes effect only on the current interface. If you execute both commands, the latter has higher priority.

Setting the global IPv4 unicast reserved bandwidth percentage

1. Enter system view.
   system-view
2. Enter MRIB view.
   multicast routing [**vpn-instance** *vpn-instance-name*]
3. Set the global IPv4 unicast reserved bandwidth percentage.
   flow-ucmp unicast reserve-bandwidth *percentage*
   By default, the global IPv4 unicast reserved bandwidth percentage is not set.

Setting the IPv4 unicast reserved bandwidth percentage on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the IPv4 unicast reserved bandwidth percentage on the interface.
   multicast flow-ucmp unicast reserve-bandwidth *percentage*
   By default, the IPv4 unicast reserved bandwidth percentage is not set on an interface.

Configuring a multicast forwarding boundary

About this task

You can configure an interface as a multicast forwarding boundary for a multicast group range. The interface cannot receive or forward multicast packets for the group range.

Restrictions and guidelines

You do not need to enable IP multicast before this configuration.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the interface as a multicast forwarding boundary for a multicast group range.
   multicast boundary *group-address* { *mask-length* | *mask* }
   By default, an interface is not a multicast forwarding boundary.

Using mtrace to trace a multicast path

Using mtrace1 to trace a multicast path

To use mtrace version 1 (mtrace1) to trace a multicast path, execute the following command in any view:

mtrace [**v1**] [**vpn-instance** *vpn-instance-name*] { *source-address* | *group-address* } * [**destination** *address*] [**verbose**]

Using mtrace2 to trace a multicast path

Restrictions and guidelines

For successful mtrace, do not specify a UDP port number used by other modules.

You must specify the same UDP port number on all devices on the traced path.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Specify the UDP port number used by mtrace.
   mtrace-service port *number*
   By default, mtrace uses UDP port number 10240.
3. Use mtrace version 2 (mtrace2) to trace a multicast path.
   mtrace v2 [**vpn-instance** *vpn-instance-name*] { *source-address* | *group-address* } * [**destination** *address* | **port** *number* | **wait-time** *time* | **max-hop** *count*] * [**verbose**]
   Execute this command in any view.
   The UDP port number specified in this command must be the same as that specified in the **mtrace-service port** command.

Setting the maximum number of cached unknown multicast packets

About this task

The device caches a multicast packet for a period of time if no matching multicast forwarding entry is found for the packet. If a multicast forwarding entry is established for the packet within the time period, the device forwards the packet. This mechanism prevents the device from mistakenly

dropping multicast packets when the multicast forwarding entries for these packets are to be established.

You can set the maximum number of unknown multicast packets that can be cached for an (S, G) entry, in total, or both.

Restrictions and guidelines

As a best practice, set the value in the **multicast forwarding-table cache-unknown total** command to be far greater than the value in the **multicast forwarding-table cache-unknown per-entry** command.

Procedure

1. Enter system view.
   system-view
2. Set the maximum number of unknown multicast packets that can be cached for an (S, G) entry.
   multicast forwarding-table cache-unknown per-entry *per-entry-limit*
   By default, the device can cache only one unknown multicast packet for an (S, G) entry.
3. Set the maximum number of unknown multicast packets that can be cached in total.
   multicast forwarding-table cache-unknown total *total-limit*
   By default, the device can cache 1024 unknown multicast packets in total.

Setting the maximum number of copied multicast packets during software forwarding

About this task

Setting a greater limit will cause high CPU usage and degrade forwarding performance. Setting a small limit will cause multicast packets to be dropped.

CAUTION:
Perform this task under the guidance of H3C Support.

Procedure

1. Enter system view.
   system-view
2. Set the maximum number of copied multicast packets during software forwarding.
   multicast cpu-forwarding max-copy-count *count*
   The default setting is 5.

Enabling multicast replication resource reuse

About this task

Multicast replication resources are hardware resources used to replicate multicast traffic. By default, multiple (S, G) entries with the same outgoing interface use multiple multicast replication resources.

This feature allows (S, G) entries with the same outgoing interface to use the same multicast replication resource.

Restrictions and guidelines

For this feature to take effect ,you must use the **reboot** command to reboot the device.

Procedure

1. Enter system view.
   system-view
2. Enable multicast replication resource reuse.
   multicast replication-share enable
   By default, multicast replication resource reuse is disabled.

Display and maintenance commands for multicast routing and forwarding

CAUTION:
The **reset** commands might cause multicast data transmission failures.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about the interfaces maintained by the MRIB.	**display mrib** [**vpn-instance** *vpn-instance-name*] **interface** [*interface-type interface-number*]
Display link selection information for IPv4 multicast flow policies.	**display multicast** [**vpn-instance** *vpn-instance-name*] **flow-policy info** [**interface** *interface-type interface-number* | **source** *source-address* | **group** *group-address* | **policy** { **default** | **name** *policy-name* }] *
Display multicast boundary information.	**display multicast** [**vpn-instance** *vpn-instance-name*] **boundary** [*group-address* [*mask-length* | *mask*]] [**interface** *interface-type interface-number*]
Display multicast fast forwarding entries.	**display multicast** [**vpn-instance** *vpn-instance-name*] **fast-forwarding cache** [*source-address* | *group-address*] * [**slot** *slot-number*]
Display statistics for multicast forwarding events.	**display multicast** [**vpn-instance** *vpn-instance-name*] **forwarding event** [**slot** *slot-number*]
Display multicast forwarding entries.	**display multicast** [**vpn-instance** *vpn-instance-name*] **forwarding-table** [*source-address* [**mask** { *mask-length* | *mask* }] | *group-address* [**mask** { *mask-length* | *mask* }] | **incoming-interface** *interface-type interface-number*| **outgoing-interface** { **exclude** | **include** |

Task	Command
	match } *interface-type interface-number* | **replicate** *replicate-index* | **slot** *slot-number* | **statistics**] *
Display multicast forwarding capability information.	**display multicast forwarding-table capability [slot slot-number]**
Display information about the DF list in the multicast forwarding table.	**display multicast** [**vpn-instance** *vpn-instance-name*] **forwarding-table df-list** [*group-address*] [**verbose**] [**slot** *slot-number*]
Display multicast routing entries.	**display multicast** [**vpn-instance** *vpn-instance-name*] **routing-table** [*source-address* [**mask** { *mask-length* | *mask* }] | *group-address* [**mask** { *mask-length* | *mask* }] | **incoming-interface** *interface-type interface-number* | **outgoing-interface** { **exclude** | **include** | **match** } *interface-type interface-number*] *
Display static multicast routing entries.	**display multicast** [**vpn-instance** *vpn-instance-name*] **routing-table static** [*source-address* { *mask-length* | *mask* }]
Display RPF information for a multicast source.	**display multicast** [**vpn-instance** *vpn-instance-name*] **rpf-info** *source-address* [*group-address*]
Display multicast replication resource information.	**display multicast replication-share info** [**outgoing-interface** { **exclude** | **include** | **match** } *interface-type interface-number* | **replicate** *replicate-index*] * [**slot** *slot-number*]
Clear multicast fast forwarding entries.	**reset multicast** [**vpn-instance** *vpn-instance-name*] **fast-forwarding cache** { { *source-address* | *group-address* } * | **all** } [**slot** *slot-number*]
Clear statistics for multicast forwarding events.	**reset multicast** [**vpn-instance** *vpn-instance-name*] **forwarding event**
Clear multicast forwarding entries.	**reset multicast** [**vpn-instance** *vpn-instance-name*] **forwarding-table** { { *source-address* [**mask** { *mask-length* | *mask* }] | *group-address* [**mask** { *mask-length* | *mask* }] | **incoming-interface** { *interface-type interface-number* } } * | **all** }
Clear multicast routing entries.	**reset multicast** [**vpn-instance** *vpn-instance-name*] **routing-table** { { *source-address* [**mask** { *mask-length* | *mask* }] | *group-address* [**mask** { *mask-length* | *mask* }] | **incoming-interface** *interface-type interface-number* } * | **all** }

NOTE:

- When you clear a multicast routing entry, the associated multicast forwarding entry is also cleared.
- When you clear a multicast forwarding entry, the associated multicast routing entry is also cleared.

Multicast routing and forwarding configuration examples

Example: Changing an RPF route

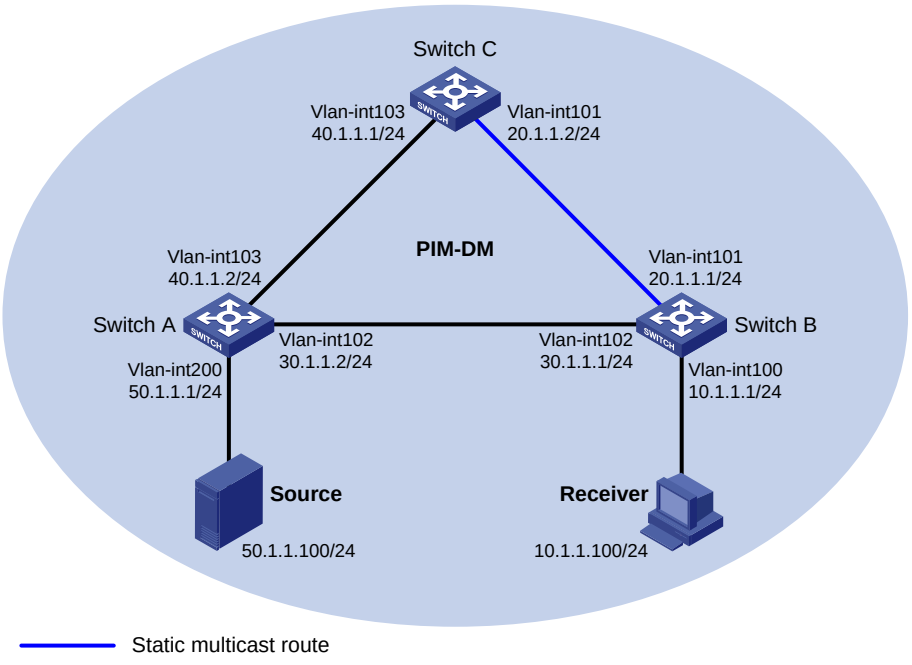
Network configuration

As shown in Figure 4:

- PIM-DM runs on the network.
- All switches on the network support multicast.
- Switch A, Switch B and Switch C run OSPF.
- Typically, the receiver host can receive the multicast data from the source through the path: Switch A to Switch B, which is the same as the unicast route.

Configure the switches so that the multicast data from the source travels to the receiver through the path: Switch A to Switch C to Switch B. This is different from the unicast route.

Figure 4 Network diagram



Prerequisites

Assign an IP address and subnet mask for each interface as shown in Figure 4, and configure OSPF on the switches in the PIM-DM domain.

Procedure

1. Enable IP multicast routing, and enable IGMP and PIM-DM:

   # On Switch B, enable IP multicast routing.

```
<SwitchB> system-view
[SwitchB] multicast routing
[SwitchB-mrib] quit
```

Enable IGMP on the receiver-side interface VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] igmp enable
[SwitchB-Vlan-interface100] quit
```

Enable PIM-DM on other interfaces.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim dm
[SwitchB-Vlan-interface101] quit
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] pim dm
[SwitchB-Vlan-interface102] quit
```

On Switch A, enable IP multicast routing.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Enable PIM-DM on each interface.

```
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] pim dm
[SwitchA-Vlan-interface200] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim dm
[SwitchA-Vlan-interface102] quit
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim dm
[SwitchA-Vlan-interface103] quit
```

Enable IP multicast routing and PIM-DM on Switch C in the same way Switch A is configured. (Details not shown.)

2. Display RPF information for the source on Switch B.

```
[SwitchB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface102, RPF neighbor: 30.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: igp
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

The output shows that the current RPF route on Switch B is contributed by a unicast routing protocol and the RPF neighbor is Switch A.

3. Configure a static multicast route on Switch B and specify Switch C as its RPF neighbor on the route to the source.

```
[SwitchB] ip rpf-route-static 50.1.1.0 24 20.1.1.2
```

Verifying the configuration

Display RPF information for the source on Switch B.

```
[SwitchB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface101, RPF neighbor: 20.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

The output shows the following information:

- The RPF route on Switch B is the configured static multicast route.
- The RPF neighbor of Switch B is Switch C.

Example: Creating an RPF route

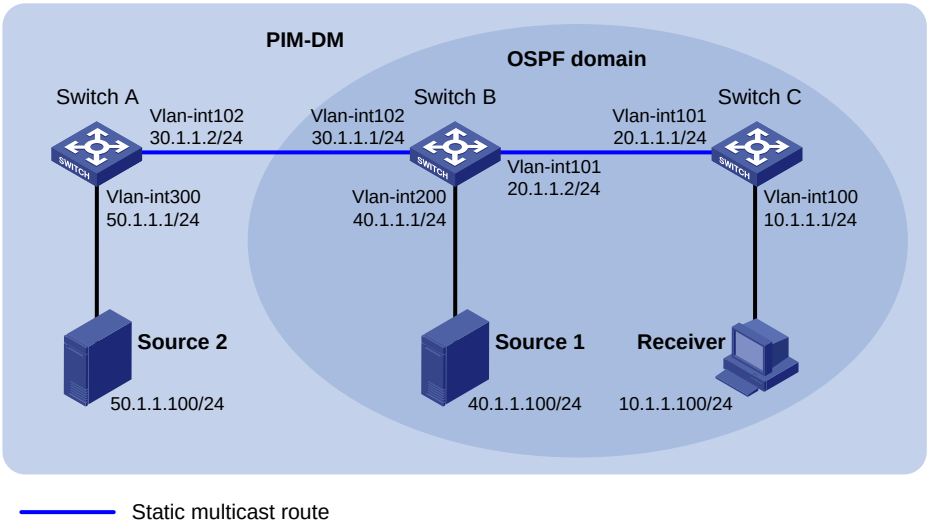
Network configuration

As shown in Figure 5:

- PIM-DM runs on the network.
- All switches on the network support IP multicast.
- Switch B and Switch C run OSPF, and have no unicast routes to Switch A.
- Typically, the receiver host receives the multicast data from Source 1 in the OSPF domain.

Configure the switches so that the receiver host receives multicast data from Source 2, which is outside the OSPF domain.

Figure 5 Network diagram



Prerequisites

Assign an IP address and subnet mask for each interface as shown in Figure 5, and configure OSPF on Switch B and Switch C.

Procedure

1. Enable IP multicast routing, and enable IGMP and PIM-DM:

   # On Switch C, enable IP multicast routing.

```
<SwitchC> system-view
[SwitchC] multicast routing
[SwitchC-mrib] quit
```

Enable IGMP on the receiver-side interface VLAN-interface 100.

```
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] igmp enable
[SwitchC-Vlan-interface100] quit
```

Enable PIM-DM on VLAN-interface 101.

```
[SwitchC] interface vlan-interface 101
[SwitchC-Vlan-interface101] pim dm
[SwitchC-Vlan-interface101] quit
```

On Switch A, enable IP multicast routing.

```
<SwitchA> system-view
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Enable PIM-DM on each interface.

```
[SwitchA] interface vlan-interface 300
[SwitchA-Vlan-interface300] pim dm
[SwitchA-Vlan-interface300] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim dm
[SwitchA-Vlan-interface102] quit
```

Enable IP multicast routing and PIM-DM on Switch B in the same way Switch A is configured. (Details not shown.)

2. Display RPF information for Source 2 on Switch B and Switch C.

```
[SwitchB] display multicast rpf-info 50.1.1.100
[SwitchC] display multicast rpf-info 50.1.1.100
```

No output is displayed because no RPF routes to Source 2 exist on Switch B or Switch C.

3. Configure a static multicast route:

Configure a static multicast route on Switch B and specify Switch A as its RPF neighbor on the route to Source 2.

```
[SwitchB] ip rpf-route-static 50.1.1.0 24 30.1.1.2
```

Configure a static multicast route on Switch C and specify Switch B as its RPF neighbor on the route to Source 2.

```
[SwitchC] ip rpf-route-static 10.1.1.0 24 20.1.1.2
```

Verifying the configuration

Display RPF information for Source 2 on Switch B.

```
[SwitchB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface102, RPF neighbor: 30.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

Display RPF information for Source 2 on Switch C.

```
[SwitchC] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface101, RPF neighbor: 20.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

The output shows that the RPF routes to Source 2 exist on Switch B and Switch C. The routes are the configured static routes.

Troubleshooting multicast routing and forwarding

Static multicast route failure

Symptom

No dynamic routing protocol is enabled on the device, and the physical status and link layer status of interfaces are both up, but the static multicast route fails.

Solution

To resolve the problem:

1. Use the **`display multicast routing-table static`** command to display information about static multicast routes. Verify that the static multicast route has been correctly configured and that the route entry exists in the static multicast routing table.
2. Check the type of interface that connects the static multicast route to the RPF neighbor. If the interface is not a point-to-point interface, be sure to specify the address for the RPF neighbor.
3. If the problem persists, contact H3C Support.

Contents

Configuring IPv6 multicast routing and forwarding 1

About IPv6 multicast routing and forwarding 1
RPF check mechanism 1
IPv6 mtrace 3
Restrictions and guidelines: IPv6 multicast routing and forwarding configuration 3
IPv6 multicast routing and forwarding tasks at a glance 3
Prerequisites for IPv6 multicast routing and forwarding 4
Enabling IPv6 multicast routing 4
Configuring static IPv6 multicast routes 4
Specifying the longest prefix match principle 5
Configuring IPv6 multicast load splitting 5
Configuring an IPv6 multicast flow policy 5
About this task 5
Restrictions and guidelines for configuring an IPv6 multicast flow policy 6
Creating an IPv6 multicast flow policy 6
Configuring the multicast group range 6
Configuring the estimated bandwidth 7
Configuring the IPv6 unicast reserved bandwidth percentage 7
Configuring an IPv6 multicast forwarding boundary 8
Using mtrace2 to trace an IPv6 multicast path 8
Setting the maximum number of cached unknown IPv6 multicast packets 9
Setting the maximum number of copied IPv6 multicast packets during software forwarding 9
Enabling IPv6 multicast replication resource reuse 10
Display and maintenance commands for IPv6 multicast routing and forwarding 10
IPv6 multicast routing and forwarding configuration examples 12
Example: Changing an RPF route 12
Example: Creating an RPF route 14

Configuring IPv6 multicast routing and forwarding

About IPv6 multicast routing and forwarding

Each IPv6 multicast routing protocol has its own routing table. Multicast routing information in routing entries generated by the IPv6 multicast routing protocols are summarized in a set of (S, G) and (*, G) entries. All the (S, G) and (*, G) entries form a general IPv6 multicast routing table. The optimal IPv6 multicast routing entries in the general IPv6 multicast routing table are added to the IPv6 multicast forwarding table to guide IPv6 multicast data forwarding.

RPF check mechanism

An IPv6 multicast routing protocol uses the reverse path forwarding (RPF) check mechanism to ensure IPv6 multicast data delivery along the correct path and to avoid data loops.

RPF check process

An IPv6 multicast device performs the RPF check on an IPv6 multicast packet as follows:

1. Chooses an optimal route back to the packet source separately from the IPv6 unicast and IPv6 MBGP routing tables.

   In RPF check, the "packet source" means difference things in difference situations:

   - For a packet that travels along the SPT, the packet source is the IPv6 multicast source.
   - For a packet that travels along the RPT, the packet source is the RP.
   - For a bootstrap message originated from the BSR, the packet source is the BSR.

   For more information about the concepts of SPT, RPT, source-side RPT, RP, and BSR, see "IPv6 PIM overview."
2. Selects one of the optimal routes as the RPF route as follows:
   - If the device uses the longest prefix match principle, the route with a higher prefix length becomes the RPF route. If the routes have the same prefix length, the route with a higher route preference becomes the RPF route. If the routes have the same route preference, the IPv6 MBGP route becomes the RPF route. If equal cost routes exist, the equal cost route with the highest next hop IPv6 address becomes the RPF route.

     For more information about the route preference, see *Layer 3—IP Routing Configuration Guide*.
   - If the device does not use the longest prefix match principle, the route with a higher route preference becomes the RPF route. If the routes have the same route preference, the IPv6 MBGP route becomes the RPF route. If equal cost routes exist, the equal cost route with the highest next hop IPv6 address becomes the RPF route.

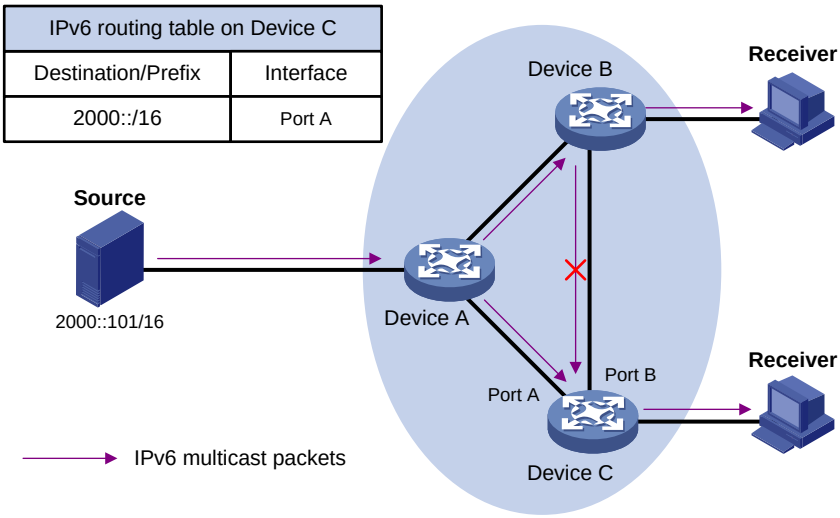
   In the RPF route, the outgoing interface is the RPF interface and the next hop is the RPF neighbor.
3. Determines whether the packet arrived at the RPF interface.
   - If the packet arrived at the RPF interface, the RPF check succeeds and the packet is forwarded.
   - If the packet arrived at the non-RPF interface, the RPF check fails and the packet is discarded.

RPF check implementation in IPv6 multicast

Implementing an RPF check on each received IPv6 multicast packet would heavily burden the device. The use of an IPv6 multicast forwarding table is the solution to this issue. When the device creates an IPv6 multicast forwarding entry for an IPv6 (S, G) packet, it sets the RPF interface of the packet as the incoming interface of the (S, G) entry. After the device receives another (S, G) packet, it looks up its IPv6 multicast forwarding table for a matching (S, G) entry.

- If no match is found, the device first determines the RPF route back to the packet source. Then, it creates a forwarding entry with the RPF interface as the incoming interface and performs one of the following tasks:
  - If the receiving interface is the RPF interface, the RPF check succeeds and the device forwards the packet out of all outgoing interfaces.
  - If the receiving interface is not the RPF interface, the RPF check fails and the device discards the packet.
- If a match is found and the matching forwarding entry contains the receiving interface, the device forwards the packet out of all outgoing interfaces.
- If a match is found but the matching forwarding entry does not contain the receiving interface, the device determines the RPF route back to the packet source. Then, the device performs one of the following tasks:
  - If the RPF interface is the incoming interface, it means that the forwarding entry is correct but the packet traveled along a wrong path. The packet fails the RPF check, and the device discards the packet.
  - If the RPF interface is not the incoming interface, it means that the forwarding entry has expired. The device replaces the incoming interface with the RPF interface and matches the receiving interface against the RPF interface. If the receiving interface is the RPF interface, the device forwards the packet out of all outgoing interfaces. Otherwise, it discards the packet.

Figure 1 RPF check process



As shown in Figure 1, assume that IPv6 unicast routes are available on the network. IPv6 MBGP is not configured. IPv6 multicast packets travel along the SPT from the multicast source to the receivers. The IPv6 multicast forwarding table on Device C contains the (S, G) entry, with Port A as the RPF interface.

- If an IPv6 multicast packet arrives at Device C on Port A, the receiving interface is the incoming interface of the (S, G) entry. Device C forwards the packet out of all outgoing interfaces.
- If an IPv6 multicast packet arrives at Device C on Port B, the receiving interface is not the incoming interface of the (S, G) entry. Device C searches its IPv6 unicast routing table and finds

that the outgoing interface to the source (the RPF interface) is Port A. This means that the (S, G) entry is correct but the packet traveled along a wrong path. The packet fails the RPF check, and Device C discards the packet.

IPv6 mtrace

IPv6 mtrace uses mtrace2 to trace the path along which IPv6 multicast group data travels from a source to a destination.

Device roles

IPv6 mtrace includes the following roles:

- **Last-hop router (LHR)**—An LHR is a router that has an IPv6 multicast-enabled interface on the same subnet as the destination and can forward specific IPv6 multicast data to the subnet.
- **First-hop router (FHR)**—An FHR is a router that is directly connected to the IPv6 multicast source.
- **Client**—A client is a router that initiates an mtrace2.

Process

The IPv6 mtrace process is as follows:

1. The client sends an mtrace2 Query message (with a **hops** field indicating the maximum number of hops to be traced) to the destination.
2. The LHR turns the received Query message to an mtrace2 Request message by adding local forwarding information and sends the Request message to the upstream neighbor.
3. Each router along the traced path adds its local forwarding information to the received Request message and sends the Request message to its upstream neighbor.
4. The FHR adds its local forwarding information to the received Request message. Then, it turns the Request message to an mtrace2 Reply message and sends the Reply message to the client.
5. The client interprets forwarding information in the Reply message and displays the information.

Restrictions and guidelines: IPv6 multicast routing and forwarding configuration

During ISSU on a multichassis IRF, if the incoming interface of Layer 3 multicast traffic contains non-local member ports, Layer 3 multicast traffic interruption might occur when the device is rebooting.

IPv6 multicast routing and forwarding tasks at a glance

To configure IPv6 multicast routing and forwarding, perform the following tasks:

1. Enabling IPv6 multicast routing
2. (Optional.) Configuring static IPv6 multicast routes
3. (Optional.) Specifying the longest prefix match principle
4. (Optional.) Configuring IPv6 multicast load splitting
5. (Optional.) Configuring an IPv6 multicast flow policy
6. (Optional.) Configuring an IPv6 multicast forwarding boundary

7. (Optional.) Using mtrace2 to trace an IPv6 multicast path
8. (Optional.) Setting the maximum number of cached unknown IPv6 multicast packets
9. (Optional.) Setting the maximum number of copied IPv6 multicast packets during software forwarding
10. (Optional.) Enabling IPv6 multicast replication resource reuse

Prerequisites for IPv6 multicast routing and forwarding

Before you configure multicast routing and forwarding, configure an IPv6 unicast routing protocol so that all devices in the domain can interoperate at the network layer.

Enabling IPv6 multicast routing

About this task

Enable IPv6 multicast routing before you configure any Layer 3 IPv6 multicast functionality in the public network or VPN instance.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 multicast routing and enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
   By default, IPv6 multicast routing is disabled.

Configuring static IPv6 multicast routes

About this task

When configuring a static IPv6 multicast route for an IPv6 multicast source, you can specify an RPF interface or an RPF neighbor for the IPv6 multicast traffic from that source.

Restrictions and guidelines

Static IPv6 multicast routes take effect only on the IPv6 multicast devices on which they are configured, and will not be broadcast or redistributed to other devices.

Procedure

1. Enter system view.
   system-view
2. Configure a static multicast route.
   ipv6 rpf-route-static [**vpn-instance** *vpn-instance-name*] *ipv6-source-address prefix-length* { *ipv6-gateway-address* | *interface-type interface-number* } [**preference** *preference*]
3. (Optional.) Delete all static IPv6 multicast routes.
   delete ipv6 rpf-route-static [**vpn-instance** *vpn-instance-name*]
   To delete a single static IPv6 multicast route, use the **undo ip rpf-route-static** command.

Specifying the longest prefix match principle

About this task

You can enable the device to use the longest prefix match principle for RPF route selection. For more information about RPF route selection, see "RPF check process."

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
3. Specify the longest prefix match principle for RPF route selection.
   longest-match
   By default, the route preference principle is used.

Configuring IPv6 multicast load splitting

About this task

You can enable the device to split multiple IPv6 multicast data flows on a per-source basis or on a per-source-and-group basis.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
3. Configure IPv6 multicast load splitting.
   load-splitting { **balance-ecmp** | **balance-ucmp** | **ecmp** | **flow-ucmp** | **source** | **source-group** | **ucmp** }
   By default, IPv6 multicast load splitting is disabled.

Configuring an IPv6 multicast flow policy

About this task

You can configure the estimated bandwidth for matching IPv6 multicast flows through an IPv6 multicast flow policy on a downstream device if the following conditions exist:

- ECMP routes between the downstream device and an upstream device.
- The multicast load-splitting mode is **flow-ucmp**.

When the downstream device selects an incoming interface on the upstream device, it selects the link with the smallest bandwidth usage calculated based on the estimated bandwidth. If two links have the same bandwidth usage, the link with the higher next-hop IPv6 address is selected.

The bandwidth usage is calculated according to the following formula: Bandwidth usage = (Used bandwidth+Estimated bandwidth)/(Total interface bandwidth x (1–Unicast reserved bandwidth)).

- The estimated bandwidth is configured by the **bandwidth** command in IPv6 multicast flow policy view.

- The total interface bandwidth is configured by the **bandwidth** command in interface view.
- The unicast reserved bandwidth is configured by the **flow-ucmp unicast reserve-bandwidth** or **ipv6 multicast flow-ucmp unicast reserve-bandwidth** command.

Restrictions and guidelines for configuring an IPv6 multicast flow policy

For this feature to take effect, you must configure the multicast load-splitting mode as **flow-ucmp**.

Adding, deleting, or modifying an IPv6 multicast flow policy does not affect the link selection results of existing multicast flows and affects link selection for only new multicast flows. Before configuring the multicast load-splitting mode as **flow-ucmp**, you must plan the configuration of an IPv6 multicast flow policy and the unicast reserved bandwidth configuration.

The change of link bandwidth and the change in the number of ECMP routes do not affect the route selection of existing multicast flows.

This feature does not affect multicast source-side link selection, link selection in IPv6 PIM-DM, RPT link selection in IPv6 PIM-SM, or local RP link selection.

If both IPv4 and IPv6 multicast flows exist in the network, increase the unicast reserved bandwidth as needed to avoid congestion.

Creating an IPv6 multicast flow policy

1. Enter system view.
   system-view
2. Enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
3. Create an IPv6 multicast flow policy.
   - Create a custom IPv6 multicast flow policy.
     flow-policy name *policy-name*
   - Create the default IPv6 multicast flow policy.
     flow-policy default

Configuring the multicast group range

About this task

Perform this task to configure a multicast group range in which an IPv6 multicast flow policy takes effect. If a multicast group matches the ACL specified in an IPv6 multicast flow policy, the estimated bandwidth configured in the policy is used for the multicast flows of the multicast group.

Restrictions and guidelines

This feature is not supported by the default IPv6 multicast flow policy.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]

3. Enter IPv6 multicast flow policy view.
   flow-policy { **default** | **name** *policy-name* }
4. Configure the multicast group range.
   acl { *ipv6-acl-number* | **name** *ipv6-acl-name* }
   By default, an IPv6 multicast flow policy does not take effect on any multicast groups.

Configuring the estimated bandwidth

About this task

The estimated bandwidth is selected for IPv6 multicast flows as follows:

1. If the multicast group of an IPv6 multicast flow matches the ACL specified in a custom IPv6 multicast flow policy, the estimated bandwidth configured in the policy is selected for the multicast flow.
2. If the multicast group of an IPv6 multicast flow does not match an ACL in any custom IPv6 multicast flow policy, the estimated bandwidth configured in the default policy is selected for the multicast flow.
3. If the default policy is not created or the estimated bandwidth is not configured in the default policy, the estimated bandwidth for the multicast flow is 0.

Procedure

1. Enter system view.
   system-view
2. Enter IPv6 MRIB view.
   ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]
3. Enter IPv6 multicast flow policy view.
   flow-policy { **default** | **name** *policy-name* }
4. Configure the estimated bandwidth.
   bandwidth *bandwidth* { **gbps** | **kbps** | **mbps** }
   By default, the estimated bandwidth is 0.

Configuring the IPv6 unicast reserved bandwidth percentage

About this task

Perform this task when IPv6 unicast traffic and IPv6 multicast traffic coexist in the network. The available bandwidth for multicast traffic is calculated based on the total interface bandwidth and the configured IPv6 unicast reserved bandwidth percentage. For example, if the total interface bandwidth is 100 kbps and the unicast reserved bandwidth percentage is 20% (20 kbps), the available bandwidth for multicast traffic is 80 kbps.

Restrictions and guidelines

For this feature to take effect, you must configure the multicast load-splitting mode as **flow-ucmp**.

The **flow-ucmp unicast reserve-bandwidth** command takes effect on all interfaces, and the **ipv6 multicast flow-ucmp unicast reserve-bandwidth** command takes effect only on the current interface. If you execute both commands, the latter has higher priority.

Setting the global IPv6 unicast reserved bandwidth percentage

1. Enter system view.
   system-view
2. Enter IPv6 MRIB view.

ipv6 multicast routing [**vpn-instance** *vpn-instance-name*]

3. Set the global IPv6 unicast reserved bandwidth percentage.
   flow-ucmp unicast reserve-bandwidth *percentage*
   By default, the global IPv6 unicast reserved bandwidth percentage is not set.

Setting the IPv6 unicast reserved bandwidth percentage on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the IPv6 unicast reserved bandwidth percentage on the interface.
   ipv6 multicast flow-ucmp unicast reserve-bandwidth *percentage*
   By default, the IPv6 unicast reserved bandwidth percentage is not set on an interface.

Configuring an IPv6 multicast forwarding boundary

About this task

You can configure an interface as an IPv6 multicast forwarding boundary for an IPv6 multicast group range. The interface cannot receive or forward IPv6 multicast packets for the groups in the range.

Restrictions and guidelines

You do not need to enable IPv6 multicast routing before this configuration.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure an IPv6 multicast forwarding boundary.
   ipv6 multicast boundary { *ipv6-group-address prefix-length* | **scope** { *scope-id* | **admin-local** | **global** | **organization-local** | **site-local** } }
   By default, an interface is not an IPv6 multicast forwarding boundary for any IPv6 multicast groups.

Using mtrace2 to trace an IPv6 multicast path

Restrictions and guidelines

For successful IPv6 mtrace, do not use a UDP port number used by other modules.

You must specify the same UDP port number on all devices on the traced path.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Specify the UDP port number used by IPv6 mtrace.
   ipv6 mtrace-service port *number*

By default, IPv6 mtrace uses UDP port number 10240.

3. Use mtrace2 to trace an IPv6 multicast path.

   mtrace v2 ipv6 [**vpn-instance** *vpn-instance-name*] { *source-address* | *group-address* } * [**destination** *address* | **port** *number* | **wait-time** *time* | **max-hop** *count*] * [**verbose**]

   The UDP port number specified in this command must be the same as that specified in the **mtrace-service port** command.

Setting the maximum number of cached unknown IPv6 multicast packets

About this task

The device caches an IPv6 multicast packet for a period of time if no matching multicast forwarding entry is found for the packet. If a multicast forwarding entry is established for the packet within the time period, the device forwards the packet. This mechanism prevents the device from mistakenly dropping IPv6 multicast packets when the multicast forwarding entries for these packets are to be established.

You can set the maximum number of unknown IPv6 multicast packets that can be cached for an (S, G) entry, in total, or both.

Restrictions and guidelines

As a best practice, set the value in the **ipv6 multicast forwarding-table cache-unknown total** command to be far greater than the value in the **ipv6 multicast forwarding-table cache-unknown per-entry** command.

Procedure

1. Enter system view.

   system-view

2. Set the maximum number of unknown IPv6 multicast packets that can be cached for an (S, G) entry.

   ipv6 multicast forwarding-table cache-unknown per-entry *per-entry-limit*

   By default, the device can cache only one unknown IPv6 multicast packet for an (S, G) entry.

3. Set the maximum number of unknown IPv6 multicast packets that can be cached in total.

   ipv6 multicast forwarding-table cache-unknown total *total-limit*

   By default, the device can cache 1024 unknown IPv6 multicast packets in total.

Setting the maximum number of copied IPv6 multicast packets during software forwarding

About this task

Setting a greater limit will cause high CPU usage and degrade forwarding performance. Setting a small limit will cause IPv6 multicast packets to be dropped.

CAUTION:

Perform this task under the guidance of H3C Support.

Procedure

1. Enter system view.
   system-view
2. Set the maximum number of copied IPv6 multicast packets during software forwarding.
   ipv6 multicast cpu-forwarding max-copy-count *count*
   The default setting is 5.

Enabling IPv6 multicast replication resource reuse

About this task

IPv6 multicast replication resources are hardware resources used to replicate IPv6 multicast traffic. By default, multiple IPv6 (S, G) entries with the same outgoing interface use multiple IPv6 multicast replication resources.

This feature allows IPv6 (S, G) entries with the same outgoing interface to use the same IPv6 multicast replication resource.

Restrictions and guidelines

For this feature to take effect ,you must use the **reboot** command to reboot the device.

Procedure

1. Enter system view.
   system-view
2. Enable IPv6 multicast replication resource reuse.
   ipv6 multicast replication-share enable
   By default, IPv6 multicast replication resource reuse is disabled.

Display and maintenance commands for IPv6 multicast routing and forwarding

CAUTION:
The **reset** commands might cause IPv6 multicast data transmission failures.

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about the interfaces maintained by the IPv6 MRIB.	**display ipv6 mrib** [**vpn-instance** *vpn-instance-name*] **interface** [*interface-type interface-number*]
Display IPv6 multicast boundary information.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **boundary** { **group** [*ipv6-group-address* [*prefix-length*]] | **scope** [*scope-id*] } [**interface** *interface-type interface-number*]
Display IPv6 multicast fast forwarding entries.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **fast-forwarding cache** [*ipv6-source-address* |

Task	Command
	ipv6-group-address] * [**slot** *slot-number*]
Display statistics for IPv6 multicast forwarding events.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **forwarding event** [**slot** *slot-number*]
Display IPv6 multicast forwarding entries.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **forwarding-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** *interface-type interface-number* | **outgoing-interface** { **exclude** | **include** | **match** } *interface-type interface-number* | **replicate** *replicate-index* | **slot** *slot-number* | **statistics**] *
Display IPv6 multicast forwarding capability information.	**display ipv6 multicast forwarding-table capability** [**slot slot-number**]
Display information about the DF list in the IPv6 multicast forwarding table.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **forwarding-table df-list** [*ipv6-group-address*] [**verbose**] [**slot** *slot-number*]
Display IPv6 multicast routing entries.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **routing-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** *interface-type interface-number* | **outgoing-interface** { **exclude** | **include** | **match** } *interface-type interface-number*] *
Display static IPv6 multicast routing entries.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **routing-table static** [*ipv6-source-address* [*prefix-length*]]
Display RPF information for an IPv6 multicast source.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **rpf-info** *ipv6-source-address* [*ipv6-group-address*]
Display IPv6 multicast replication resource information.	**display ipv6 multicast replication-share info** [**outgoing-interface** { **exclude** | **include** | **match** } *interface-type interface-number* | **replicate** *replicate-index*] * [**slot** *slot-number*]
Display link selection information for IPv6 multicast flow policies.	**display ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **flow-policy info** [**interface** *interface-type interface-number* | **source** *source-address* | **group** *group-address* | **policy** { **default** | **name** *policy-name* }] *
Clear IPv6 multicast fast forwarding entries.	**reset ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **fast-forwarding cache** { { *ipv6-source-address* |

Task	Command
	ipv6-group-address } * | **all** } [**slot** *slot-number*]
Clear statistics for IPv6 multicast forwarding events.	**reset ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **forwarding event**
Clear IPv6 multicast forwarding entries.	**reset ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **forwarding-table** { { *ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* } } * | **all** }
Clear IPv6 multicast routing entries.	**reset ipv6 multicast** [**vpn-instance** *vpn-instance-name*] **routing-table** { { *ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** *interface-type interface-number* } * | **all** }

NOTE:

- When you clear an IPv6 multicast routing entry, the associated IPv6 multicast forwarding entry is also cleared.
- When you clear an IPv6 multicast forwarding entry, the associated IPv6 multicast routing entry is also cleared.

IPv6 multicast routing and forwarding configuration examples

Example: Changing an RPF route

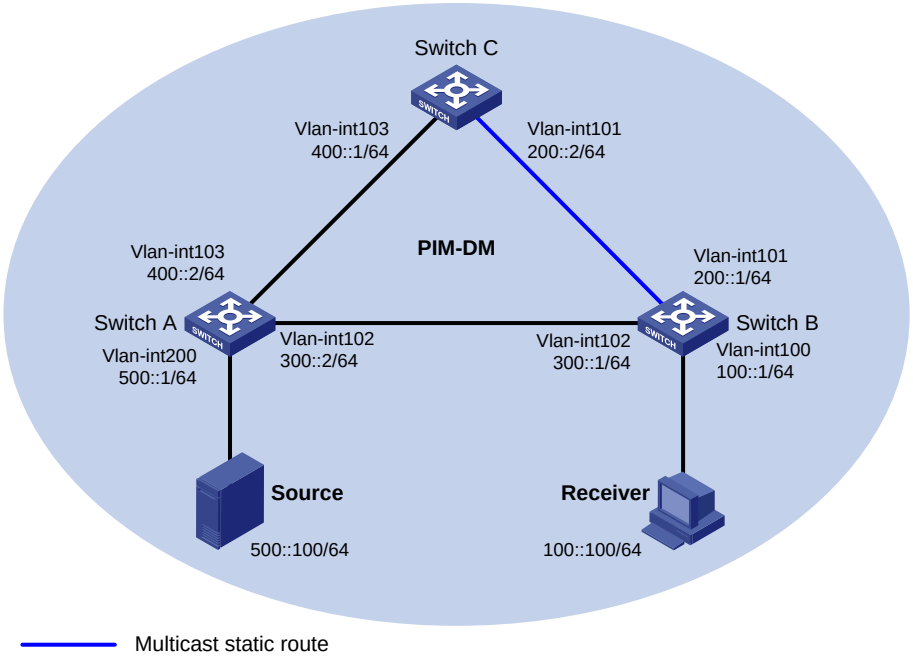
Network configuration

As shown in Figure 2:

- IPv6 PIM-DM runs on the network.
- All switches on the network support IPv6 multicast.
- Switch A, Switch B, and Switch C run OSPFv3.
- Typically, the receiver host can receive the IPv6 multicast data from the source through the path: Switch A to Switch B, which is the same as the unicast route.

Configure the switches so that the IPv6 multicast data from the source travels to the receiver along the following path: Switch A to Switch C to Switch B. This path is different from the unicast route.

Figure 2 Network diagram



Prerequisites

1. Assign an IP address and prefix length for each interface, as shown in Figure 2.
2. Configure OSPFv3 on the switches in the IPv6 PIM-DM domain.

Procedure

1. Enable IPv6 multicast routing, and enable MLD and IPv6 PIM-DM:

   # On Switch B, enable IPv6 multicast routing.

   ```
   <SwitchB> system-view
   [SwitchB] ipv6 multicast routing
   [SwitchB-mrib6] quit
   ```

   # Enable MLD on the receiver-side interface VLAN-interface 100.

   ```
   [SwitchB] interface vlan-interface 100
   [SwitchB-Vlan-interface100] mld enable
   [SwitchB-Vlan-interface100] quit
   ```

   # Enable IPv6 PIM-DM on the other interfaces.

   ```
   [SwitchB] interface vlan-interface 101
   [SwitchB-Vlan-interface101] ipv6 pim dm
   [SwitchB-Vlan-interface101] quit
   [SwitchB] interface vlan-interface 102
   [SwitchB-Vlan-interface102] ipv6 pim dm
   [SwitchB-Vlan-interface102] quit
   ```

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Enable IPv6 PIM-DM on each interface.

   ```
   [SwitchA] interface vlan-interface 200
   [SwitchA-Vlan-interface200] ipv6 pim dm
   ```

```
[SwitchA-Vlan-interface200] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] ipv6 pim dm
[SwitchA-Vlan-interface102] quit
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] ipv6 pim dm
[SwitchA-Vlan-interface103] quit
```

Enable IPv6 multicast routing and IPv6 PIM-DM on Switch C in the same way Switch A is configured. (Details not shown.)

2. Display RPF information for the source on Switch B.

```
[SwitchB] display ipv6 multicast rpf-info 500::100
 RPF information about source 500::100:
     RPF interface: Vlan-interface102, RPF neighbor: 300::2
     Referenced prefix/prefix length: 500::/64
     Referenced route type: igp
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

The output shows that the current RPF route on Switch B is contributed by a unicast routing protocol and the RPF neighbor is Switch A.

3. On Switch B, configure a static IPv6 multicast route to the source and specify Switch C as the RPF neighbor.

```
[SwitchB] ipv6 rpf-route-static 500::100 64 200::2
```

Verifying the configuration

Display RPF information for the source on Switch B.

```
[SwitchB] display ipv6 multicast rpf-info 500::100
 RPF information about source 500::100:
     RPF interface: Vlan-interface101, RPF neighbor: 200::2
     Referenced prefix/prefix length: 500::/64
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

The output shows the following information:

- The RPF route on Switch B is the configured static IPv6 multicast route.
- The RPF neighbor of Switch B is Switch C.

Example: Creating an RPF route

Network configuration

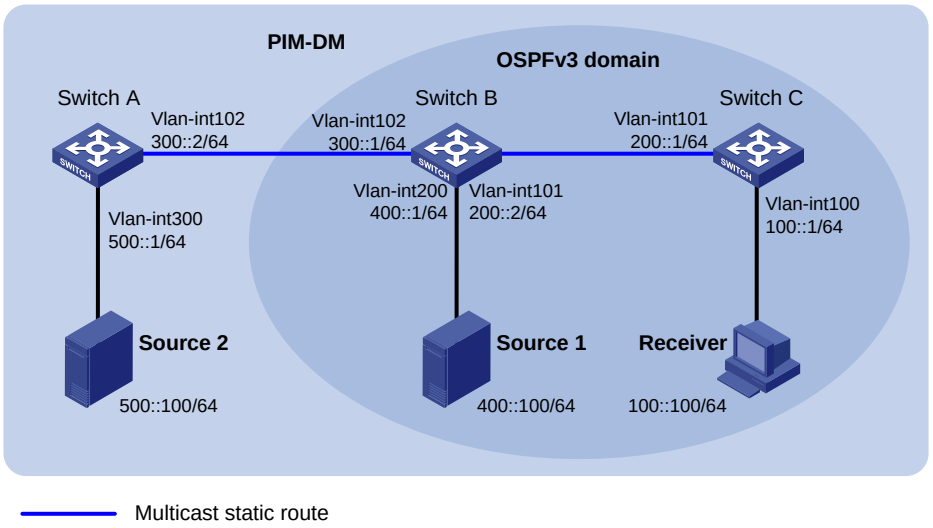
As shown in Figure 3:

- IPv6 PIM-DM runs on the network.
- All switches on the network support IP multicast.
- Switch B and Switch C run OSPFv3, and have no unicast routes to Switch A.

- Typically, the receiver host receives the IPv6 multicast data from the source 1 in the OSPFv3 domain.

Configure the switches so that the receiver host can receive IPv6 multicast data from Source 2, which is outside the OSPFv3 domain.

Figure 3 Network diagram



Prerequisites

1. Assign an IP address and subnet mask for each interface, as shown in Figure 3.
2. Configure OSPFv3 on Switch B and Switch C.

Procedure

1. Enable IPv6 multicast routing, and enable MLD and IPv6 PIM-DM:

   # On Switch C, enable IPv6 multicast routing.

   ```
   <SwitchC> system-view
   [SwitchC] ipv6 multicast routing
   [SwitchC-mrib6] quit
   ```

   # Enable MLD on the receiver-side interface VLAN-interface 100.

   ```
   [SwitchC] interface vlan-interface 100
   [SwitchC-Vlan-interface100] mld enable
   [SwitchC-Vlan-interface100] quit
   ```

   # Enable IPv6 PIM-DM on VLAN-interface 101.

   ```
   [SwitchC] interface vlan-interface 101
   [SwitchC-Vlan-interface101] ipv6 pim dm
   [SwitchC-Vlan-interface101] quit
   ```

   # On Switch A, enable IPv6 multicast routing.

   ```
   <SwitchA> system-view
   [SwitchA] ipv6 multicast routing
   [SwitchA-mrib6] quit
   ```

   # Enable IPv6 PIM-DM on each interface.

   ```
   [SwitchA] interface vlan-interface 300
   [SwitchA-Vlan-interface300] ipv6 pim dm
   [SwitchA-Vlan-interface300] quit
   [SwitchA] interface vlan-interface 102
   ```

```
[SwitchA-Vlan-interface102] ipv6 pim dm
[SwitchA-Vlan-interface102] quit
```

Enable IPv6 multicast routing and IPv6 PIM-DM on Switch B in the same way Switch A is configured. (Details not shown.)

2. Display RPF information for Source 2 on Switch B and Switch C.

```
[SwitchB] display ipv6 multicast rpf-info 500::100
[SwitchC] display ipv6 multicast rpf-info 500::100
```

No output is displayed because no RPF routes to Source 2 exist on Switch B and Switch C.

3. Configure a static IPv6 multicast route:

Configure a static IPv6 multicast route on Switch B and specify Switch A as its RPF neighbor to Source 2.

```
[SwitchB] ipv6 rpf-route-static 500::100 64 300::2
```

Configure a static IPv6 multicast route on Switch C and specify Switch B as its RPF neighbor to Source 2.

```
[SwitchC] ipv6 rpf-route-static 500::100 64 200::2
```

Verifying the configuration

Display RPF information for Source 2 on Switch B.

```
[SwitchB] display ipv6 multicast rpf-info 500::100
 RPF information about source 50::100:
     RPF interface: Vlan-interface102, RPF neighbor: 300::2
     Referenced prefix/prefix length: 500::/64
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

Display RPF information for Source 2 on Switch C.

```
[SwitchC] display ipv6 multicast rpf-info 500::100
 RPF information about source 500::100:
     RPF interface: Vlan-interface101, RPF neighbor: 200::2
     Referenced prefix/prefix length: 500::/64
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
     Source AS: 0
     C-multicast route target: 0x0000000000000000
```

The output shows that the RPF routes to Source 2 exist on Switch B and Switch C. These RPF routes are the configured static IPv6 multicast routes.

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

MPLS Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the MPLS fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring basic MPLS 1

About MPLS 1
MPLS features 1
Basic concepts 1
MPLS network architecture 2
LSP establishment 3
MPLS forwarding 4
PHP 5
Protocols and standards 5
Restrictions and guidelines: basic MPLS configuration 6
MPLS tasks at a glance 6
Enabling MPLS 6
Specifying the label type advertised by the egress node to the penultimate hop 7
Configuring TTL propagation 7
Enabling sending MPLS TTL-expired messages 9
Enabling SNMP notifications for MPLS 9
Enabling private SNMP notifications for MPLS 9
Display and maintenance commands for MPLS 10

Configuring basic MPLS

About MPLS

Multiprotocol Label Switching (MPLS) provides connection-oriented label switching over connectionless IP backbone networks. It integrates both the flexibility of IP routing and the simplicity of Layer 2 switching.

MPLS features

MPLS has the following features:

- **High speed and efficiency**—MPLS uses short- and fixed-length labels to forward packets, avoiding complicated routing table lookups.
- **Multiprotocol support**—MPLS resides between the link layer and the network layer. It can work over various link layer protocols (for example, Ethernet) to provide connection-oriented services for various network layer protocols (for example, IPv4 and IPX).
- **Good scalability**—The connection-oriented switching and multilayer label stack features enable MPLS to deliver various extended services, such as VPN, traffic engineering, and QoS.

Basic concepts

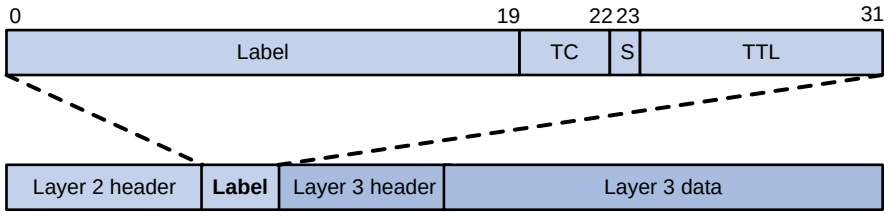
FEC

MPLS groups packets with the same characteristics (such as packets with the same destination or service class) into a forwarding equivalence class (FEC). Packets of the same FEC are handled in the same way on an MPLS network.

Label

A label uniquely identifies an FEC and has local significance.

Figure 1 Format of a label



A label is encapsulated between the Layer 2 header and Layer 3 header of a packet. It is four bytes long and consists of the following fields:

- **Label**—20-bit label value.
- **TC**—3-bit traffic class, used for QoS. It is also called Exp.
- **S**—1-bit bottom of stack flag. A label stack can contain multiple labels. The label nearest to the Layer 2 header is called the top label, and the label nearest to the Layer 3 header is called the bottom label. The **S** field is set to 1 if the label is the bottom label and set to 0 if not.
- **TTL**—8-bit time to live field used for MPLS loop prevention.

LSR

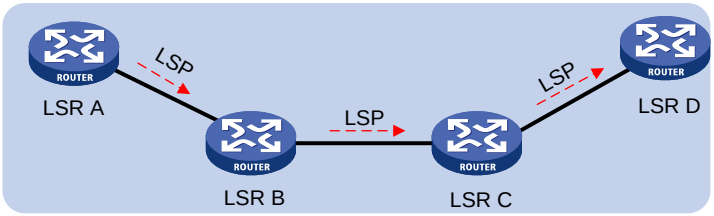
A router that performs MPLS forwarding is a label switching router (LSR).

LSP

A label switched path (LSP) is the path along which packets of an FEC travel through an MPLS network.

An LSP is a unidirectional packet forwarding path. Two neighboring LSRs are called the upstream LSR and downstream LSR along the direction of an LSP. As shown in Figure 2, LSR B is the downstream LSR of LSR A, and LSR A is the upstream LSR of LSR B.

Figure 2 Label switched path



LFIB

The Label Forwarding Information Base (LFIB) on an MPLS network functions like the Forwarding Information Base (FIB) on an IP network. When an LSR receives a labeled packet, it searches the LFIB to obtain information for forwarding the packet. The information includes the label operation type, the outgoing label value, and the next hop.

Control plane and forwarding plane

An MPLS node consists of a control plane and a forwarding plane.

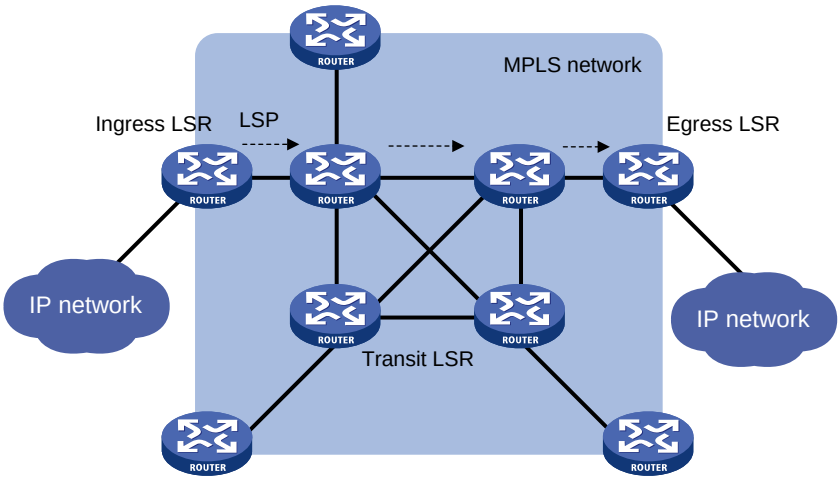
- **Control plane**—Assigns labels, distributes FEC-label mappings to neighbor LSRs, creates the LFIB, and establishes and removes LSPs.
- **Forwarding plane**—Forwards packets according to the LFIB.

MPLS network architecture

An MPLS network has the following types of LSRs:

- **Ingress LSR**—Ingress LSR of packets. It labels packets entering into the MPLS network.
- **Transit LSR**—Intermediate LSRs in the MPLS network. The transit LSRs on an LSP forward packets to the egress LSR according to labels.
- **Egress LSR**—Egress LSR of packets. It removes labels from packets and forwards the packets to their destination networks.

Figure 3 MPLS network architecture



LSP establishment

LSPs include static and dynamic LSPs.

Static LSP

To establish a static LSP, you must configure an LFIB entry on each LSR along the LSP. Establishing static LSPs consumes fewer resources than establishing dynamic LSPs, but static LSPs cannot automatically adapt to network topology changes. Therefore, static LSPs are suitable for small-scale networks with simple, stable topologies.

Dynamic LSP

A dynamic LSP is established by a label distribution protocol (also called an MPLS signaling protocol). A label distribution protocol classifies FECs, distributes FEC-label mappings, and establishes and maintains LSPs. Label distribution protocols include protocols designed specifically for label distribution, such as the Label Distribution Protocol (LDP), and protocols extended to support label distribution, such as MP-BGP.

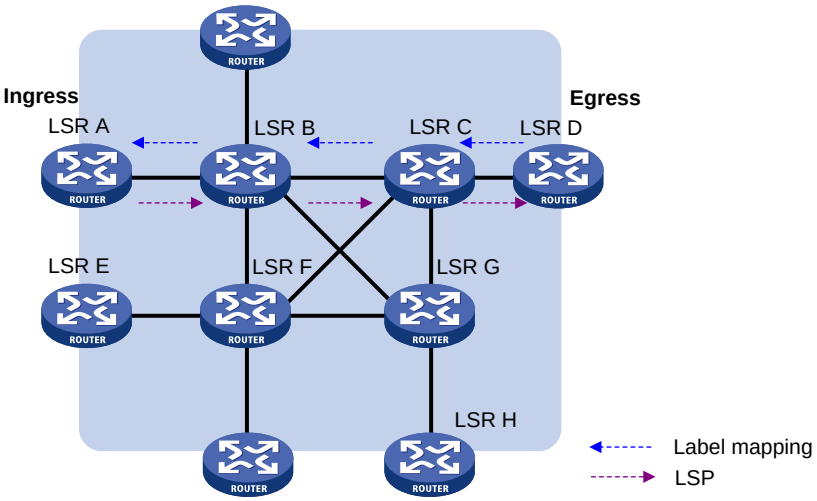
In this document, the term "label distribution protocols" refers to all protocols for label distribution. The term "LDP" refers to the RFC 5036 LDP.

A dynamic LSP is established in the following steps:

1. A downstream LSR classifies FECs according to destination addresses.
2. The downstream LSR assigns a label for each FEC, and distributes the FEC-label binding to its upstream LSR.
3. The upstream LSR establishes an LFIB entry for the FEC according to the binding information.

After all LSRs along the LSP establish an LFIB entry for the FEC, a dynamic LSP is established for the packets of this FEC.

Figure 4 Dynamic LSP establishment



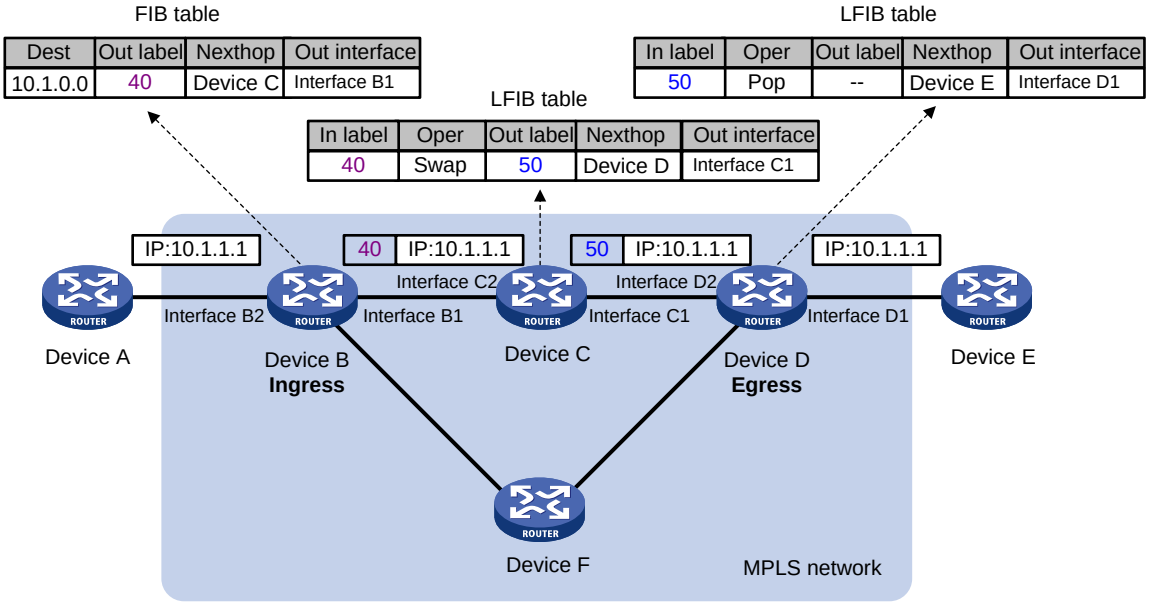
MPLS forwarding

As shown in Figure 5, a packet is forwarded over the MPLS network as follows:

1. Device B (the ingress LSR) receives a packet with no label. Then, it performs the following operations:
   a. Identifies the FIB entry that matches the destination address of the packet.
   b. Adds the outgoing label (40, in this example) to the packet.
   c. Forwards the labeled packet out of Interface B1 to the next hop LSR Device C.
2. When receiving the labeled packet, Device C processes the packet as follows:
   a. Identifies the LFIB entry that has an incoming label of 40.
   b. Uses the outgoing label 50 of the entry to replace label 40 in the packet.
   c. Forwards the labeled packet out of Interface C1 to the next hop LSR Device D.
3. When receiving the labeled packet, Device D (the egress LSR) processes the packet as follows:
   a. Identifies the LFIB entry that has an incoming label of 50.
   b. Removes the label from the packet.
   c. Forwards the packet out of Interface D1 to the next hop LSR Device E.

   If the LFIB entry records no outgoing interface or next hop information, Device D performs the following operations:
   d. Identifies the FIB entry by the IP header.
   e. Forwards the packet according to the FIB entry.

Figure 5 MPLS forwarding



PHP

An egress node must perform two forwarding table lookups to forward a packet:

- Two LFIB lookups (if the packet has more than one label).
- One LFIB lookup and one FIB lookup (if the packet has only one label).

The penultimate hop popping (PHP) feature can pop the label at the penultimate node, so the egress node only performs one table lookup.

A PHP-capable egress node can send either of the following null labels to the penultimate node:

- **Implicit null label**—The value of an implicit null label is 3. If an incoming packet matches an LFIB entry containing the implicit null label, the penultimate node pops the top label and forwards the packet to the egress node. The egress node forwards the packet based on the header of its next layer.
- **Explicit null label**—The value of an explicit null label is 0 for IPv4 networks and 2 for IPv6 networks (value 2 is not supported in the current software version) . Sometimes, the egress node must use the TC field in the label to perform QoS. To keep the TC information, you can configure the egress node to send the penultimate node an explicit null label. If an incoming packet matches an LFIB entry containing the explicit null label, the penultimate hop replaces the top label value with value 0, and forwards the packet to the egress node. When the egress node receives the packet, it does not look up the LFIB table for the label of 0 or 2. The egress node gets the TC information from the label, pops the label, and then forwards the packet based on the header of its next layer.

Protocols and standards

- RFC 3031, *Multiprotocol Label Switching Architecture*
- RFC 3032, *MPLS Label Stack Encoding*
- RFC 5462, *Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field*

Restrictions and guidelines: basic MPLS configuration

Only advanced mode and expert mode support basic MPLS configuration. For more information about device operating modes, see device management configuration in *Fundamentals Configuration Guide*.

The device does not support MPLS forwarding for tunneled packets.

MPLS tasks at a glance

To configure basic MPLS, perform the following configuration tasks:

1. Enabling MPLS
2. (Optional.) Specifying the label type advertised by the egress node to the penultimate hop
3. (Optional.) Configuring TTL propagation

   When TTL propagation is enabled, the ingress node copies the TTL value of an IP packet to the TTL field of the label. The IP tracert facility can show the real path along which the packet has traveled.
4. (Optional.) Enabling sending MPLS TTL-expired messages
5. (Optional.) Enabling SNMP notifications for MPLS
6. (Optional.) Enabling private SNMP notifications for MPLS

Enabling MPLS

Prerequisites

Before you enable MPLS, configure static routes or an IGP protocol to ensure IP connectivity among LSRs.

Procedure

1. Enter system view.

   system-view
2. Configure an LSR ID for the local node.

   mpls lsr-id *lsr-id*

   By default, no LSR ID is configured.

   An LSR ID must be unique in an MPLS network and in IP address format. As a best practice, use the IP address of a loopback interface as an LSR ID.
3. Enter the view of the interface that needs to perform MPLS forwarding.

   interface *interface-type interface-number*
4. Enable MPLS on the interface.

   mpls enable

   By default, MPLS is disabled on the interface.

Specifying the label type advertised by the egress node to the penultimate hop

About this task

- Select the label type advertised by the egress node to the penultimate hop as follows:
- As a best practice, configure the egress node to advertise an implicit null label to the penultimate hop if the penultimate hop supports PHP.
- If you want to simplify packet forwarding on egress but keep labels to determine QoS policies, configure the egress node to advertise an explicit null label to the penultimate hop.
- Use non-null labels only in particular scenarios. For example, when OAM is configured on the egress node, the egress node can get the OAM function entity status only through non-null labels.

Restrictions and guidelines

For LDP LSPs, the **mpls label advertise** command triggers LDP to delete the LSPs established before the command is executed and then re-establish the LSPs.

For BGP LSPs, the **mpls label advertise** command takes effect only on the BGP LSPs established after the command is executed. To apply the new setting to BGP LSPs established before the command is executed, delete the routes corresponding to the BGP LSPs, and then redistribute the routes.

Procedure

1. Enter system view.

   system-view
2. Specify the label type advertised by the egress node to the penultimate hop.

   mpls label advertise { **explicit-null** | **implicit-null** | **non-null** }

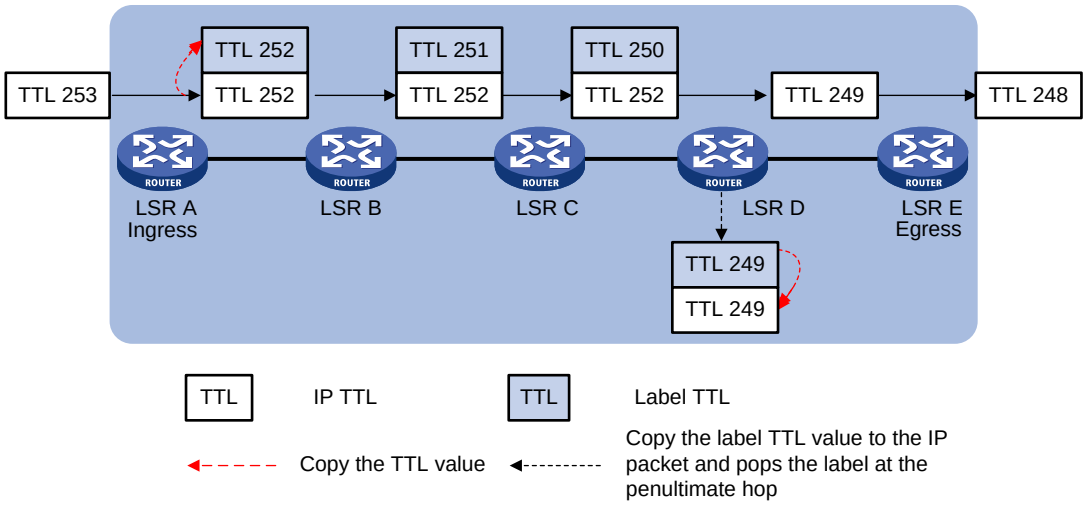
   By default, an egress node advertises an implicit null label to the penultimate hop.

Configuring TTL propagation

About this task

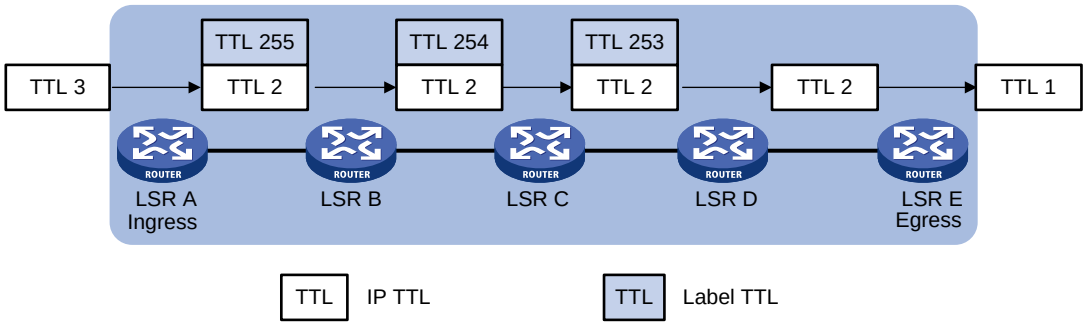
When TTL propagation is enabled, the ingress node copies the TTL value of an IP packet to the TTL field of the label. Each LSR on the LSP decreases the label TTL value by 1. The LSR that pops the label copies the remaining label TTL value back to the IP TTL of the packet. The IP TTL value can reflect how many hops the packet has traversed in the MPLS network. The IP tracert facility can show the real path along which the packet has traveled.

Figure 6 TTL propagation



When TTL propagation is disabled, the ingress node sets the label TTL to 255. Each LSR on the LSP decreases the label TTL value by 1. The LSR that pops the label does not change the IP TTL value when popping the label. Therefore, the MPLS backbone nodes are invisible to user networks, and the IP tracert facility cannot show the real path in the MPLS network.

Figure 7 Without TTL propagation



Restrictions and guidelines

The TTL propagation configuration does not take effect on existing LSPs. To make the TTL propagation configuration take effect on existing LSPs, perform the following operation after you execute the **mpls ttl propagate** command:

- For static LSPs, re-configure the static LSPs.
- For dynamically esetablished LSPs, reset LDP sessions by using the **reset mpls ldp** command.

This feature affects only the propagation between IP TTL and label TTL. Within an MPLS network, TTL is always copied between the labels of an MPLS packet.

As a best practice, set the same TTL processing mode on all LSRs of an LSP.

To enable TTL propagation for a VPN, you must enable it on all PE devices in the VPN. Then, you can get the same traceroute result (hop count) from those PEs.

Procedure

1. Enter system view.
   system-view
2. Enable TTL propagation.

mpls ttl propagate { **public** | **vpn** }

By default, TTL propagation is enabled for public network packets and is disabled for VPN packets.

Enabling sending MPLS TTL-expired messages

About this task

This feature enables an LSR to generate an ICMP TTL-expired message upon receiving an MPLS packet with a TTL of 1. If the MPLS packet has only one label, the LSR sends the ICMP TTL-expired message back to the source through IP routing. If the MPLS packet has multiple labels, the LSR sends it along the LSP to the egress node, which then sends the message back to the source.

Procedure

1. Enter system view.

   system-view

2. Enable sending MPLS TTL-expired messages.

   mpls ttl expiration enable

   By default, this feature is enabled.

Enabling SNMP notifications for MPLS

About this task

To report critical MPLS events to an NMS, enable SNMP notifications for MPLS. For MPLS event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for MPLS.

   snmp-agent trap enable mpls

   By default, SNMP notifications for MPLS are disabled.

Enabling private SNMP notifications for MPLS

About this task

This feature enables generating SNMP notifications for specific MPLS events, such as the following:

- MPLS resources are exhausted.
- MPLS label usage exceeds the upper threshold or drops below the lower threshold.
- A static LSP goes up or down.

These MPLS event notifications are not defined in an RFC and are private MPLS notifications supported by Comware. For SNMP to output the private MPLS notifications correctly, you must also configure parameters for sending SNMP notifications. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Enable private SNMP notifications for MPLS.

   snmp-agent trap enable mpls-private [**entryres–exhausted** | **hard-resource** | **label-layer-exceeded** | **label-threshold** | **label-total** | **resources-exceeded** | **slsp-label-dup** | **slsp-up-down** | **sr-label-conflict** | **tunnel-to-vpn**] *

   By default, private SNMP notifications for MPLS are disabled.

3. (Optional.) Configure alarm thresholds for MPLS label usage.

   mpls label-usage threshold-alarm percent { **dynamic** | **egress** | **egress6** | **l3vpn** | **srgb** | **srlb** | **static** | **vpls** | **vpws** } **upper-limit** *upper-percent* **lower-limit** *lower-percent*

   By default, the usage alarm upper threshold is 80% and lower threshold is 70% for all types of MPLS labels.

Display and maintenance commands for MPLS

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ILM entries.	**display mpls forwarding ilm** [*label*] [**slot** *slot-number*]
Display NHLFE entries.	**display mpls forwarding nhlfe** [*nid*] [**slot** *slot-number*]
Display MPLS interface information.	**display mpls interface** [*interface-type interface-number*]
Display usage information for MPLS labels.	**display mpls label** { *label-value1* [**to** *label-value2*] | **all** }
Display LSP information.	**display mpls lsp** [**egress** | **in-label** *label-value* | **ingress** | **outgoing-interface** *interface-type interface-number* | **protocol** { **bgp** | **ldp** | **local** | **static** } | **transit**] [**vpn-instance** *vpn-instance-name*] [*ipv4-address mask-length*] [**verbose**]
Display MPLS Nexthop Information Base (NIB) information.	**display mpls nib** [*nib-id*]
Display usage information for NIDs.	**display mpls nid** [*nid-value1* [**to** *nid-value2*]]
Display MPLS summary information.	**display mpls summary**

Contents

Configuring a static LSP 1
About static LSP 1
Restrictions and guidelines: Static LSP configuration 1
Prerequisites for static LSP 1
Configuring the ingress node of a static LSP 1
Configuring a transit node of a static LSP 2
Configuring the egress node of a static LSP 2
Display and maintenance commands for static LSP 3
Static LSP configuration examples 3
Example: Configuring static LSPs 3

Configuring a static LSP

About static LSP

A static label switched path (LSP) is established by manually specifying the incoming label and outgoing label on each node (ingress, transit, or egress node) of the forwarding path.

Static LSPs consume fewer resources, but they cannot automatically adapt to network topology changes. Therefore, static LSPs are suitable for small and stable networks with simple topologies.

Restrictions and guidelines: Static LSP configuration

The outgoing label specified on an LSR must be the same as the incoming label specified on the directly connected downstream LSR.

LSPs are unidirectional. You must configure an LSP for each direction of the data forwarding path.

Only advanced mode and expert mode support static LSP. For more information about system operating modes, see device management in *Fundamentals Configuration Guide*.

The device does not support MPLS forwarding for tunneled packets.

Prerequisites for static LSP

Before you configure a static LSP, perform the following tasks:

- Identify the ingress node, transit nodes, and egress node of the static LSP.
- Enable MPLS on all interfaces that participate in MPLS forwarding. For more information, see "Configuring basic MPLS."
- Make sure the ingress node has a route to the destination address of the LSP.

Configuring the ingress node of a static LSP

About this task

The ingress node of a static LSP performs the following operations:

1. Determines a Forwarding Equivalence Class (FEC) for a packet according to the destination address.
2. Adds the label for that FEC into the packet.
3. Forwards the packet to the next hop or out of the outgoing interface.

On the ingress node, you must specify the outgoing label for the destination address (the FEC) and the next hop or the outgoing interface.

Restrictions and guidelines

On the ingress node, if you specify a next hop, make sure the ingress node has an active route to the next hop address.

Procedure

1. Enter system view.

system-view

2. Configure the ingress node of a static LSP.

   static-lsp ingress *lsp-name* **destination** *ip-address* { *mask* | *mask-length* } { **nexthop** *next-hop-ip-address* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label*

Configuring a transit node of a static LSP

About this task

A transit node swaps the label carried in a received packet with an outgoing label, and forwards the packet to the next hop or out of the outgoing interface.

On a transit node, you must specify the incoming label, the outgoing label, and the next hop or the outgoing interface.

Restrictions and guidelines

On a transit node, if you specify a next hop, make sure the transit node has an active route to the next hop address.

Procedure

1. Enter system view.

   system-view

2. Configure a transit node of a static LSP.

   static-lsp transit *lsp-name* **in-label** *in-label* { **nexthop** *next-hop-ip-address* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label*

Configuring the egress node of a static LSP

About this task

If PHP is not configured, the egress node processes a received packet as follows:

1. Pops the incoming label of the packet.
2. Performs label forwarding according to the inner label of the packet or performs IP forwarding.

Therefore, you must perform this task to specify the incoming label on the egress node.

You do not need to perform this task if PHP is configured. To configure PHP for a static LSP, set the outgoing label to 0 or 3 on the penultimate hop of the static LSP.

You can associate a static LSP with an LDP LSP to simplify packet processing when the following conditions are met:

- Packets are forwarded over the static LSP and the LDP LSP to the destination.
- The egress node of the static LSP is also the ingress node of the LDP LSP.

To associate a static LSP with an LDP LSP, specify the incoming label and destination address on the egress node of the static LSP. After receiving a packet with the specified incoming label, the egress node of the static LSP swaps the label with the outgoing label for the LDP LSP. Then, the node forwards the packet to the next hop for the LDP LSP.

Restrictions and guidelines

On the egress node, if you specify a destination address, make sure the egress node has a route to the destination address.

Procedure

1. Enter system view.

   system-view

2. Configure the egress node of a static LSP.

   static-lsp egress *lsp-name* **in-label** *in-label* [**destination** *ip-address* { *mask* | *mask-length* }]

Display and maintenance commands for static LSP

Execute **display** commands in any view.

Task	Command
Display static LSP information.	**display mpls static-lsp** [**lsp-name** *lsp-name* | **openflow**]

Static LSP configuration examples

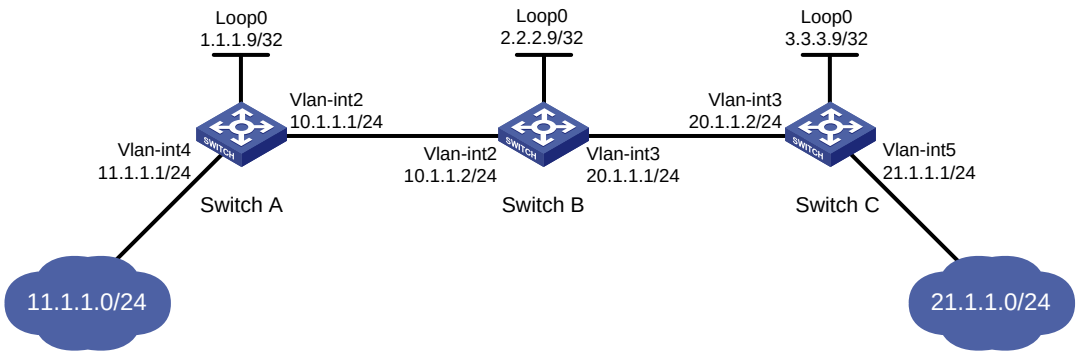
Example: Configuring static LSPs

Network configuration

Switch A, Switch B, and Switch C all support MPLS.

Establish static LSPs between Switch A and Switch C, so that subnets 11.1.1.0/24 and 21.1.1.0/24 can access each other over MPLS.

Figure 1 Network diagram



Procedure

1. Create VLANs and assign IP addresses to interfaces:

   # Configure Switch A:

   ```
   <Sysname> system-view
   [Sysname] sysname SwitchA
   [SwitchA] vlan 2
   [SwitchA-vlan2] port gigabitethernet 1/0/1
   [SwitchA-vlan2] quit
   ```

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 10.1.1.1 24
[SwitchA-Vlan-interface2] undo shutdown
[SwitchA-Vlan-interface2] quit
[SwitchA] vlan 4
[SwitchA-vlan4] port gigabitethernet 1/0/2
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ip address 11.1.1.1 24
[SwitchA-Vlan-interface4] undo shutdown
[SwitchA-Vlan-interface4] quit
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] undo shutdown
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] undo shutdown
[SwitchA-GigabitEthernet1/0/2] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.9 32
[SwitchA-LoopBack0] quit
```

Configure Switch B.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 10.1.1.2 24
[SwitchB-Vlan-interface2] undo shutdown
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ip address 20.1.1.1 24
[SwitchB-Vlan-interface3] undo shutdown
[SwitchB-Vlan-interface3] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] undo shutdown
[SwitchB-GigabitEthernet1/0/1] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] undo shutdown
[SwitchB-GigabitEthernet1/0/2] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.9 32
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
<Sysname> system-view
[Sysname] sysname SwitchC
```

```
[SwitchC] vlan 3
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ip address 20.1.1.2 24
[SwitchC-Vlan-interface3] undo shutdown
[SwitchC-Vlan-interface3] quit
[SwitchC] vlan 5
[SwitchC-vlan5] port gigabitethernet 1/0/2
[SwitchC-vlan5] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ip address 21.1.1.1 24
[SwitchC-Vlan-interface5] undo shutdown
[SwitchC-Vlan-interface5] quit
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] undo shutdown
[SwitchC-GigabitEthernet1/0/1] quit
[SwitchC] interface gigabitethernet 1/0/2
[SwitchC-GigabitEthernet1/0/2] undo shutdown
[SwitchC-GigabitEthernet1/0/2] quit
[SwitchC]interface loopback 0
[SwitchC-LoopBack0] ip address 3.3.3.9 32
[SwitchC-LoopBack0] quit
```

2. Configure a static route to the destination address of each static LSP:

On Switch A, configure a static route to network 21.1.1.0/24.

```
[SwitchA] ip route-static 21.1.1.0 24 10.1.1.2
```

On Switch C, configure a static route to network 11.1.1.0/24.

```
[SwitchC] ip route-static 11.1.1.0 255.255.255.0 20.1.1.1
```

3. Configure basic MPLS on the switches:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.9
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] mpls enable
[SwitchA-Vlan-interface2] quit
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.9
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] mpls enable
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mpls enable
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.9
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] mpls enable
[SwitchC-Vlan-interface3] quit
```

4. Configure a static LSP from Switch A to Switch C:

Configure the LSP ingress node, Switch A.

```
[SwitchA] static-lsp ingress AtoC destination 21.1.1.0 24 nexthop 10.1.1.2 out-label
30
```

Configure the LSP transit node, Switch B.

```
[SwitchB] static-lsp transit AtoC in-label 30 nexthop 20.1.1.2 out-label 50
```

Configure the LSP egress node, Switch C.

```
[SwitchC] static-lsp egress AtoC in-label 50
```

5. Configure a static LSP from Switch C to Switch A:

Configure the LSP ingress node, Switch C.

```
[SwitchC] static-lsp ingress CtoA destination 11.1.1.0 24 nexthop 20.1.1.1 out-label
40
```

Configure the LSP transit node, Switch B.

```
[SwitchB] static-lsp transit CtoA in-label 40 nexthop 10.1.1.1 out-label 70
```

Configure the LSP egress node, Switch A.

```
[SwitchA] static-lsp egress CtoA in-label 70
```

Verifying the configuration

Display static LSP information on switches, for example, on Switch A.

```
[SwitchA] display mpls static-lsp
Total: 2
Name            FEC               In/Out Label Nexthop/Out Interface    State
AtoC            21.1.1.0/24       NULL/30      10.1.1.2                 Up
CtoA            -/-               70/NULL      -                        Up
```

Test the connectivity of the LSP from Switch A to Switch C.

```
[SwitchA] ping mpls -a 11.1.1.1 ipv4 21.1.1.0 24
MPLS ping FEC 21.1.1.0/24 with 100 bytes of data:
100 bytes from 20.1.1.2: Sequence=1 time=4 ms
100 bytes from 20.1.1.2: Sequence=2 time=1 ms
100 bytes from 20.1.1.2: Sequence=3 time=1 ms
100 bytes from 20.1.1.2: Sequence=4 time=1 ms
100 bytes from 20.1.1.2: Sequence=5 time=1 ms

--- Ping statistics for FEC 21.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/1/4 ms
```

Test the connectivity of the LSP from Switch C to Switch A.

```
[SwitchC] ping mpls -a 21.1.1.1 ipv4 11.1.1.0 24
MPLS ping FEC 11.1.1.0/24 with 100 bytes of data:
100 bytes from 10.1.1.1: Sequence=1 time=5 ms
100 bytes from 10.1.1.1: Sequence=2 time=1 ms
100 bytes from 10.1.1.1: Sequence=3 time=1 ms
100 bytes from 10.1.1.1: Sequence=4 time=1 ms
100 bytes from 10.1.1.1: Sequence=5 time=1 ms

--- Ping statistics for FEC 11.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/1/5 ms
```

Configuration files

- Switch A

```
#
 sysname SwitchA
#
 mpls lsr-id 1.1.1.9
#
vlan 2
#
vlan 4
#
interface LoopBack0
 ip address 1.1.1.9 255.255.255.255
#
interface Vlan-interface2
 ip address 10.1.1.1 255.255.255.0
 mpls enable
#
interface Vlan-interface4
 ip address 11.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 4
#
 ip route-static 21.1.1.0 24 10.1.1.2
#
 static-lsp ingress AtoC destination 21.1.1.0 24 nexthop 10.1.1.2 out-label 30
 static-lsp egress CtoA in-label 70
#
return
```

- Switch B

```
#
 sysname SwitchB
#
 mpls lsr-id 2.2.2.9
#
vlan 2 to 3
#
interface LoopBack0
 ip address 2.2.2.9 255.255.255.255
#
interface Vlan-interface2
 ip address 10.1.1.2 255.255.255.0
```

```
 mpls enable
#
interface Vlan-interface3
 ip address 20.1.1.1 255.255.255.0
 mpls enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
 static-lsp transit AtoC in-label 30 nexthop 20.1.1.2 out-label 50
 static-lsp transit CtoA in-label 40 nexthop 10.1.1.1 out-label 70
#
return
```

- Switch C

```
#
 sysname SwitchC
#
 mpls lsr-id 3.3.3.9
#
vlan 3
#
vlan 5
#
interface LoopBack0
 ip address 3.3.3.9 255.255.255.255
#
interface Vlan-interface3
 ip address 20.1.1.2 255.255.255.0
 mpls enable
#
interface Vlan-interface5
 ip address 21.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 5
#
 ip route-static 11.1.1.0 24 20.1.1.1
#
```

```
 static-lsp egress AtoC in-label 50
 static-lsp ingress CtoA destination 11.1.1.0 24 nexthop 20.1.1.1 out-label 40
#
return
```

Contents

Configuring LDP 1

About LDP 1
Terminology 1
LDP messages 1
LDP operation 2
Label distribution and control 3
LDP GR 4
LDP NSR 6
LDP-IGP synchronization 6
LDP FRR 7
LDP diagnostic troubleshooting 8
Protocols 9
Restrictions and guidelines: LDP configuration 9
LDP tasks at a glance 9
Enabling LDP 10
Restrictions and guidelines for enabling LDP 10
Enabling LDP globally 10
Enabling LDP on an interface 10
Configuring Hello parameters 10
About hello timers 10
Restrictions and guidelines for hello parameter configuration 11
Setting Link Hello timers 11
Setting Targeted Hello timers for an LDP peer 11
Configuring LDP session parameters 11
About LDP session parameters 11
Restrictions and guidelines for configuring LDP session parameters 12
Configuring LDP sessions parameters for Basic Discovery mechanism 12
Configuring LDP IPv4 sessions parameters for Extended Discovery mechanism 13
Configuring LDP IPv6 sessions parameters for Extended Discovery mechanism 13
Configuring LDP backoff 14
Setting a DSCP value for outgoing LDP packets 14
Configuring LDP to redistribute BGP unicast routes 14
Configuring an LSP generation policy 15
Configuring a label mapping propagation policy 16
Configuring the LDP label distribution control mode 16
Configuring a label advertisement policy 17
Configuring a label acceptance policy 18
Disabling distribution of public label mappings to remote LDP peers 19
Configuring LDP MD5 authentication 20
Configuring LDP loop detection parameters 20
Configuring LDP session protection 21
Configuring LDP GR 22
Configuring LDP NSR 22
Configuring LDP-IGP synchronization 23
Restrictions and guidelines for LDP-IGP synchronization 23
Configuring LDP-OSPF synchronization 23
Configuring LDP IS-IS synchronization 24
Configuring LDP FRR 25
Configuring LDP LFA FRR 25
Configuring LDP remote LFA FRR 25
Enabling SNMP notifications for LDP 25
Enabling private SNMP notifications for LDP 26
Display and maintenance commands for LDP 26
IPv4 LDP configuration examples 27
Example: Configuring LDP LSPs 27
Example: Configuring label acceptance control 35
Example: Configuring label advertisement control 47

Example: Configuring LDP FRR ····60
IPv6 LDP configuration examples ····71
Example: Configuring IPv6 LDP LSP ····71
Example: Configuring IPv6 label acceptance control ····81
Example: Configuring IPv6 label advertisement control ····94

Configuring LDP

About LDP

The Label Distribution Protocol (LDP) dynamically distributes FEC-label mapping information between LSRs to establish LSPs.

Terminology

LDP session

Two LSRs establish a TCP-based LDP session to exchange FEC-label mappings.

LDP peer

Two LSRs that use LDP to exchange FEC-label mappings are LSR peers.

Label spaces and LDP identifiers

Label spaces include the following types:

- **Per-interface label space**—Each interface uses a single, independent label space. Different interfaces can use the same label values.
- **Per-platform label space**—Each LSR uses a single label space. The device only supports the per-platform label space.

A six-byte LDP Identifier (LDP ID) identifies a label space on an LSR. It is in the format of <LSR ID>:<*label space number*>, where:

- The LSR ID takes four bytes to identity the LSR.
- The label space number takes two bytes to identify a label space within the LSR.

A label space number of 0 indicates that the label space is a per-platform label space. A label space number other than 0 indicates a per-interface label space.

LDP uses the same LDP ID format on IPv4 and IPv6 networks. An LDP ID must be globally unique.

FECs and FEC-label mappings

MPLS groups packets with the same characteristics (such as the same destination or service class) into a class, called an FEC. The packets of the same FEC are handled in the same way on an MPLS network.

An LSR assigns a label for an FEC and advertises the FEC-label mapping, or FEC-label binding, to its peers in a Label Mapping message.

LDP messages

LDP mainly uses the following types of messages:

- **Discovery messages**—Declare and maintain the presence of LSRs, such as Hello messages.
- **Session messages**—Establish, maintain, and terminate sessions between LDP peers, such as Initialization messages used for parameter negotiation and Keepalive messages used to maintain sessions.
- **Advertisement messages**—Create, alter, and remove FEC-label mappings, such as Label Mapping messages used to advertise FEC-label mappings.
- **Notification messages**—Provide advisory information and notify errors, such as Notification messages.

LDP uses UDP to transport discovery messages for efficiency, and uses TCP to transport session, advertisement, and notification messages for reliability.

LDP operation

LDP can operate on an IPv4 or IPv6 network, or a network where IPv4 coexists with IPv6. LDP operates similarly on IPv4 and IPv6 networks.

LDP operates in the following phases: discovering and maintaining LDP peers, establishing and maintaining LDP sessions, and establishing LSPs.

Discovering and maintaining LDP peers

LDP discovers peers in the following ways:

- **Basic Discovery**—Discovers directly connected LSRs.
  - On an IPv4 network, an LSR sends IPv4 Link Hello messages to multicast address 224.0.0.2. All directly connected LSRs can discover the LSR and establish an IPv4 Link Hello adjacency.
  - On an IPv6 network, an LSR sends IPv6 Link Hello messages to FF02:0:0:0:0:0:0:2. All directly connected LSRs can discover the LSR and establish an IPv6 Link Hello adjacency.
  - On a network where IPv4 and IPv6 coexist, an LSR sends both IPv4 and IPv6 Link Hello messages to each directly connected LSR and keeps both the IPv4 and IPv6 Link Hello adjacencies with a neighbor.
- **Extended Discovery**—Sends LDP IPv4 Targeted Hello messages to an IPv4 address or LDP IPv6 Targeted Hello messages to an IPv6 address. The destination LSR can discover the LSR and establish a hello adjacency. This mechanism is typically used in LDP session protection.

LDP can establish two hello adjacencies with a directly connected neighbor through both discovery mechanisms. It sends Hello messages at the hello interval to maintain a hello adjacency. If LDP receives no Hello message from a hello adjacency before the hello hold timer expires, it removes the hello adjacency.

Establishing and maintaining LDP sessions

LDP establishes a session to a peer in the following steps:

1. Establishes a TCP connection to the neighbor.

   On a network where IPv4 and IPv6 coexist, LDP establishes an IPv6 TCP connection. If LDP does not receive new IPv6 Hello message in a long period of time, LDP tries to establish an IPv4 TCP connection.
2. Negotiates session parameters such as LDP version, label distribution method, and Keepalive timer, and establishes an LDP session to the neighbor if the negotiation succeeds.

After a session is established, LDP sends LDP PDUs (an LDP PDU carries one or more LDP messages) to maintain the session. If no information is exchanged between the LDP peers within the Keepalive interval, LDP sends Keepalive messages at the Keepalive interval to maintain the session. If LDP receives no LDP PDU from a neighbor before the keepalive hold timer expires, or the last hello adjacency with the neighbor is removed, LDP terminates the session.

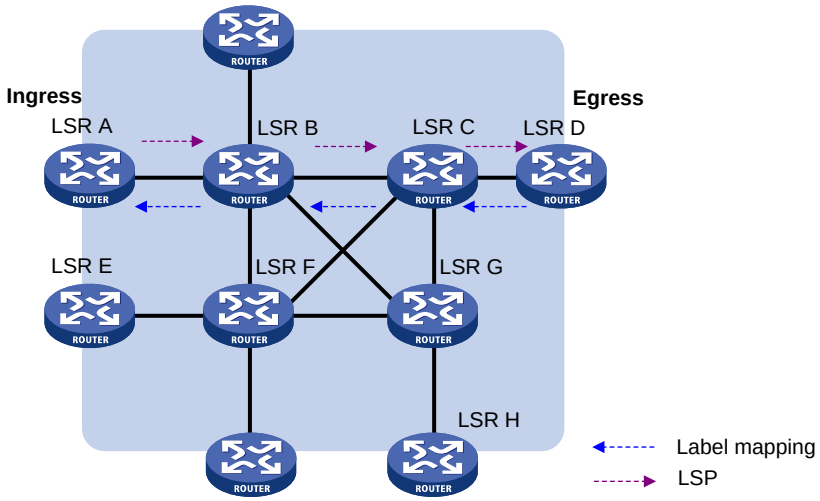
LDP can also send a Shutdown message to a neighbor to terminate the LDP session.

An LSR can establish only one LDP session to a neighbor. The session can be used to exchange IPv4 and IPv6 FEC-label mappings at the same time.

Establishing LSPs

LDP classifies FECs according to destination IP addresses in IP routing entries, creates FEC-label mappings, and advertises the mappings to LDP peers through LDP sessions. After an LDP peer receives an FEC-label mapping, it uses the received label and the label locally assigned to that FEC to create an LFIB entry for that FEC. When all LSRs (from the Ingress to the Egress) establish an LFIB entry for the FEC, an LSP is established exclusively for the FEC.

Figure 1 Dynamically establishing an LSP



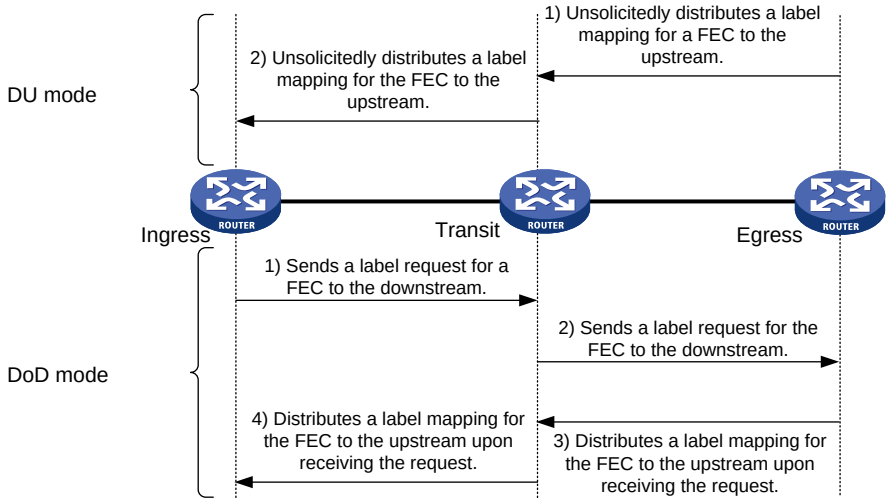
Label distribution and control

Label advertisement modes

LDP advertises label-FEC mappings in one of the following ways:

- **Downstream Unsolicited (DU) mode**—Distributes FEC-label mappings to the upstream LSR, without waiting for label requests. The device supports only the DU mode.
- **Downstream on Demand (DoD) mode**—Sends a label request for an FEC to the downstream LSR. After receiving the label request, the downstream LSR distributes the FEC-label mapping for that FEC to the upstream LSR.

Figure 2 Label advertisement modes



NOTE:

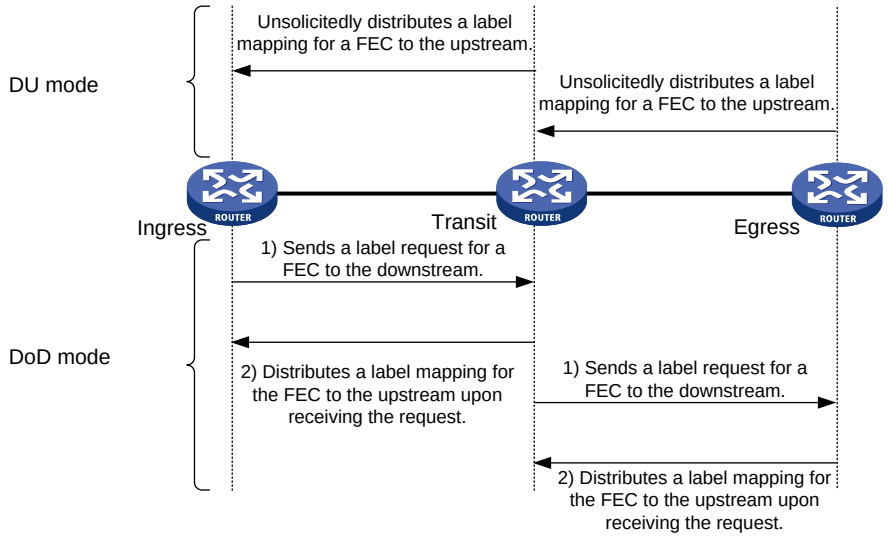
To successfully establish an LSP, a pair of upstream and downstream LSRs must use the same label advertisement mode.

Label distribution control

LDP controls label distribution in one of the following ways:

- **Independent label distribution**—Distributes an FEC-label mapping to an upstream LSR at any time. An LSR might distribute a mapping for an FEC to its upstream LSR before it receives a label mapping for that FEC from its downstream LSR. As shown in Figure 3, in DU mode, each LSR distributes a label mapping for an FEC to its upstream LSR whenever it is ready to label-switch the FEC. The LSRs do not need to wait for a label mapping for the FEC from its downstream LSR. In DoD mode, an LSR distributes a label mapping for an FEC to its upstream LSR after it receives a label request for the FEC. The LSR does not need to wait for a label mapping for the FEC from its downstream LSR.

  Figure 3 Independent label distribution control mode



- **Ordered label distribution**—Distributes a label mapping for an FEC to its upstream LSR only after it receives a label mapping for that FEC from its downstream LSR unless the local node is the egress node of the FEC. As shown in Figure 2, in DU mode, an LSR distributes a label mapping for an FEC to its upstream LSR only if it receives a label mapping for the FEC from its downstream LSR. In DoD mode, when an LSR (Transit) receives a label request for an FEC from its upstream LSR (Ingress), it continues to send a label request for the FEC to its downstream LSR (Egress). After the transit LSR receives a label mapping for the FEC from the egress LSR, it distributes a label mapping for the FEC to the ingress LSR.

Label retention mode

The label retention mode specifies whether an LSR maintains a label mapping for an FEC learned from a neighbor that is not its next hop.

- **Liberal label retention**—Retains a received label mapping for an FEC regardless of whether the advertising LSR is the next hop of the FEC. This mechanism allows for quicker adaptation to topology changes, but it wastes system resources because LDP has to keep useless labels. The device only supports liberal label retention.
- **Conservative label retention**—Retains a received label mapping for an FEC only when the advertising LSR is the next hop of the FEC. This mechanism saves label resources, but it cannot quickly adapt to topology changes.

LDP GR

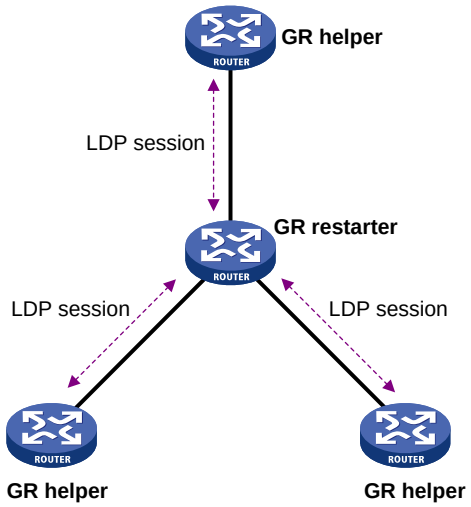
LDP Graceful Restart (GR) preserves label forwarding information when the signaling protocol or control plane fails, so that LSRs can still forward packets according to forwarding entries.

As shown in Figure 4, GR defines the following roles:

- **GR restarter**—An LSR that performs GR. It must be GR-capable.
- **GR helper**—A neighbor LSR that helps the GR restarter to complete GR.

The device can act as a GR restarter or a GR helper.

Figure 4 LDP GR



As shown in Figure 4, LDP GR operates as follows:

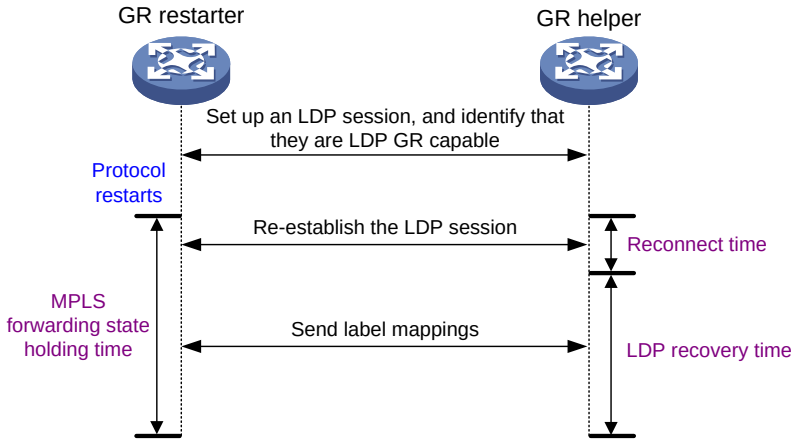
2. LSRs establish an LDP session. The L flag of the Fault Tolerance TLV in their Initialization messages is set to 1 to indicate that they support LDP GR.
3. When LDP restarts, the GR restarter starts the MPLS Forwarding State Holding timer, and marks the MPLS forwarding entries as stale. When the GR helper detects that the LDP session to the GR restarter goes down, it performs the following operations:
   a. Marks the FEC-label mappings learned from the session as stale.
   b. Starts the Reconnect timer received from the GR restarter.
4. After LDP completes restart, the GR restarter re-establishes an LDP session to the GR helper.
   - If the LDP session is not set up before the Reconnect timer expires, the GR helper deletes the stale FEC-label mappings and the corresponding MPLS forwarding entries.
   - If the LDP session is successfully set up before the Reconnect timer expires, the GR restarter sends the remaining time of the MPLS Forwarding State Holding timer to the GR helper.

     The remaining time is sent as the LDP Recovery time.
5. After the LDP session is re-established, the GR helper starts the LDP Recovery timer.
6. The GR restarter and the GR helper exchange label mappings and update their MPLS forwarding tables.

   The GR restarter compares each received label mapping against stale MPLS forwarding entries. If a match is found, the restarter deletes the stale mark for the matching entry. Otherwise, it adds a new entry for the label mapping.

   The GR helper compares each received label mapping against stale FEC-label mappings. If a match is found, the helper deletes the stale mark for the matching mapping. Otherwise, it adds the received FEC-label mapping and a new MPLS forwarding entry for the mapping.
7. When the MPLS Forwarding State Holding timer expires, the GR restarter deletes all stale MPLS forwarding entries.
8. When the LDP Recovery timer expires, the GR helper deletes all stale FEC-label mappings.

Figure 5 LDP GR operation



LDP NSR

To use LDP nonstop routing (NSR), the device must have a minimum of two MPUs.

LDP NSR ensures nonstop services of LDP when LDP has redundant processes on multiple MPUs. In contrast to Graceful Restart, NSR does not require an LDP peer to recover MPLS forwarding information.

LDP NSR backs up protocol states and data (including LDP session and LSP information) from the active process to the standby process. When the LDP active process fails, the standby process becomes active and takes over processing seamlessly. The LDP peers are not notified of the LDP interruption. The LDP session stays in Operational state, and the forwarding is not interrupted.

The LDP active process fails when one of the following situations occurs:

- The active process restarts.
- The MPU where the active process resides fails.
- The MPU where the active process resides performs an ISSU.
- The LDP process is operating at a different location than the one determined by the process placement feature.

LDP-IGP synchronization

Basic operating mechanism

LDP establishes LSPs based on the IGP optimal route. If LDP is not synchronized with IGP, MPLS traffic forwarding might be interrupted.

LDP is not synchronized with IGP when one of the following situations occurs:

- A link is up, and IGP advertises and uses this link. However, LDP LSPs on this link have not been established.
- An LDP session on a link is down, and LDP LSPs on the link have been removed. However, IGP still uses this link.
- The **Ordered** label distribution control mode is used. IGP used the link before the local device received the label mappings from the downstream LSR to establish LDP LSPs.

After LDP-IGP synchronization is enabled, IGP advertises the actual cost of a link only when LDP convergence on the link is completed. Before LDP convergence is completed, IGP advertises the maximum cost of the link. In this way, the link is visible on the IGP topology, but IGP does not select

this link as the optimal route when other links are available. Therefore, the device can avoid discarding MPLS packets when there is not an LDP LSP established on the optimal route.

LDP convergence on a link is completed when both the following situations occur:

- The local device establishes an LDP session to a minimum of one peer, and the LDP session is already in **Operational** state.
- The local device has distributed the label mappings to a minimum of one peer.

Notification delay for LDP convergence completion

By default, LDP immediately sends a notification to IGP that LDP convergence has completed. However, immediate notifications might cause MPLS traffic forwarding interruptions in one of the following scenarios:

- LDP peers use the **Ordered** label distribution control mode. The device has not received a label mapping from downstream at the time LDP notifies IGP that LDP convergence has completed.
- A large number of label mappings are distributed from downstream. Label advertisement is not completed when LDP notifies IGP that LDP convergence has completed.

To avoid traffic forwarding interruptions in these scenarios, configure the notification delay. When LDP convergence on a link is completed, LDP waits before notifying IGP.

Notification delay for LDP restart or active/standby switchover

When an LDP restart or an active/standby switchover occurs, LDP takes time to converge, and LDP notifies IGP of the LDP-IGP synchronization status as follows:

- If a notification delay is not configured, LDP immediately notifies IGP of the current synchronization states during convergence, and then updates the states after LDP convergence. This could impact IGP processing.
- If a notification delay is configured, LDP notifies IGP of the LDP-IGP synchronization states in bulk when one of the following events occurs:
  - LDP recovers to the state before the restart or switchover.
  - The maximum delay timer expires.

LDP FRR

LDP LFA FRR

A link or router failure on a path can cause packet loss until LDP establishes a new LSP on the new path. LDP Loop Free Alternate (LFA) FRR enables fast rerouting to minimize the failover time. LDP FRR is based on IP FRR and is enabled automatically after IP FRR is enabled.

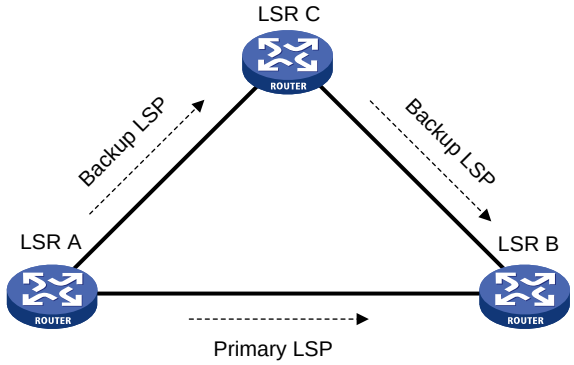
You can use one of the following methods to enable IP FRR:

- Configure an IGP to automatically calculate a backup next hop.
- Configure an IGP to specify a backup next hop by using a routing policy.

As shown in Figure 6, configure IP FRR on LSR A. The IGP automatically calculates a backup next hop or it specifies a backup next hop through a routing policy. LDP creates a primary LSP and a backup LSP according to the primary route and the backup route calculated by IGP. When the primary LSP operates correctly, it forwards the MPLS packets. When the primary LSP fails, LDP directs packets to the backup LSP.

When packets are forwarded through the backup LSP, IGP calculates the optimal path based on the new network topology. When IGP route convergence occurs, LDP establishes a new LSP according to the optimal path. If a new LSP is not established after IGP route convergence, traffic forwarding might be interrupted. As a best practice, enable LDP-IGP synchronization to work with LDP LFA FRR to reduce traffic interruption.

Figure 6 Network diagram for LDP LFA FRR



LDP remote LFA FRR

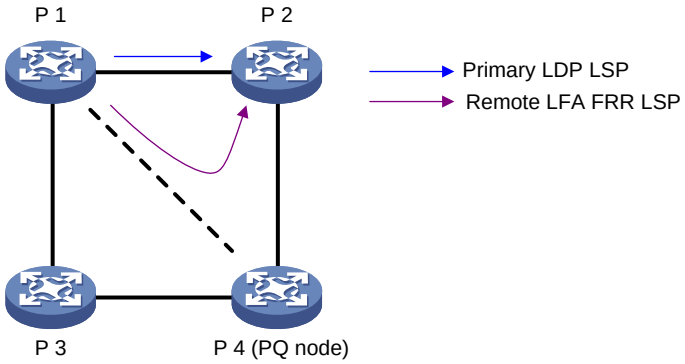
LDP LFA FRR might not be able to compute backup paths for full backup coverage on a large-scale network. To address this issue, you can deploy LDP remote LFA FRR.

As shown in Figure 7, P 1 is the source node. P 2 is the destination node. The primary LDP LSP is P 1—P 2. LDP remote LFA FRR can establish a backup LDP LSP (P 1—P 4—P 2) to protect the primary LDP LSP, as follows:

1. The IGP calculates the PQ node (P 4) using the remote LFA algorithm.
2. LDP does the following:
   a. Automatically creates the remote peer according to the PQ node address.
   b. Establishes the remote LDP session between the source node and the PQ node.
   c. Assigns a label to the destination address on the session so as to establish the remote LFA FRR LSP (P 1—P 4—P 2).

When the primary LDP LSP fails, P 1 instantly switches traffic to the remote LFA FRR LSP to minimize traffic loss. For more information about remote LFA, see IS-IS configuration in *Layer 3—IP Routing Configuration Guide*.

Figure 7 LDP remote LFA FRR



LDP diagnostic troubleshooting

Use LDP diagnostic troubleshooting to verify whether the LDP module is running correctly. For example, whether LDP sessions can be established successfully and whether LDP peers run correctly without flapping. For more information about LDP diagnostic troubleshooting, see diagnostic troubleshooting configuration in *Network Management and Monitoring Configuration Guide*.

Protocols

- RFC 5036, *LDP Specification*
- *draft-ietf-mpls-ldp-ipv6-09.txt*

Restrictions and guidelines: LDP configuration

Only advanced mode and expert mode support LDP. For more information about system operating modes, see device management in *Fundamentals Configuration Guide*.

The device does not support MPLS forwarding for tunneled packets.

LDP tasks at a glance

To configure LDP, perform the following tasks:

1. Enabling LDP
2. (Optional.) Tuning and optimizing LDP
   - Configuring Hello parameters
   - Configuring LDP session parameters
   - Configuring LDP backoff
   - Setting a DSCP value for outgoing LDP packets
3. (Optional.) Tuning and controlling LSP generation
   - Configuring LDP to redistribute BGP unicast routes
   - Configuring an LSP generation policy
   - Configuring a label mapping propagation policy
4. (Optional.) Distributing and managing LDP labels
   - Configuring the LDP label distribution control mode
   - Configuring a label advertisement policy
   - Configuring a label acceptance policy
   - Disabling distribution of public label mappings to remote LDP peers
5. (Optional.) Configuring LDP MD5 authentication
6. (Optional.) Configuring LDP loop detection parameters

   This feature is applicable to an MPLS network where most of the devices do not support the TTL mechanism, such as ATM switches.
7. (Optional.) Enhancing LDP availability
   - Configuring LDP session protection
   - Configuring LDP GR
   - Configuring LDP NSR
   - Configuring LDP-IGP synchronization
   - Configuring LDP FRR
8. (Optional.) Enabling SNMP notifications for LDP
9. (Optional.) Enabling private SNMP notifications for LDP

Enabling LDP

Restrictions and guidelines for enabling LDP

To enable LDP, you must first enable LDP globally and then enable LDP on relevant interfaces.

Enabling LDP globally

1. Enter system view.
   system-view
2. Enable LDP for the local node or for a VPN.
   - Enable LDP for the local node and enter LDP view.
     mpls ldp
   - Execute the following commands in sequence to enable LDP for a VPN, and then enter LDP-VPN instance view:
     mpls ldp
     vpn-instance *vpn-instance-name*

   By default, LDP is globally disabled.
3. Configure an LDP LSR ID.
   lsr-id *lsr-id*
   By default, the LDP LSR ID is the same as the MPLS LSR ID.

Enabling LDP on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   If the interface is bound to a VPN instance, you must enable LDP for the VPN instance by using the **vpn-instance** command in LDP view.
3. Enable IPv4 LDP on the interface.
   mpls ldp enable
   By default, IPv4 LDP is disabled on an interface.
4. Enable IPv6 LDP on the interface.
   mpls ldp ipv6 enable
   By default, IPv6 LDP is disabled on an interface.

Configuring Hello parameters

About hello timers

LDP has the following hello timers:

- Link Hello hold time and Link Hello interval.

If an interface is enabled with both IPv4 LDP and IPv6 LDP, the parameters configured on the interface can be used for both IPv4 and IPv6 Link Hello messages.

- Targeted Hello hold time and Targeted Hello interval for a peer.

Restrictions and guidelines for hello parameter configuration

Changes to hello parameters do not take effect on established LDP sessions. To apply the changes to existing LDP sessions on a network, you must use the **reset mpls ldp** command to reset all LDP sessions on that network.

Setting Link Hello timers

1. Enter system view.
   system-view
2. Enter the view of the interface where you want to establish an LDP session.
   interface *interface-type interface-number*
3. Set the Link Hello hold time.
   mpls ldp timer hello-hold *timeout*
   By default, the Link Hello hold time is 15 seconds.
4. Set the Link Hello interval.
   mpls ldp timer hello-interval *interval*
   By default, the Link Hello interval is 5 seconds.

Setting Targeted Hello timers for an LDP peer

1. Enter system view.
   system-view
2. Enter LDP view.
   mpls ldp
3. Specify an LDP peer and enter LDP peer view. The device will send unsolicited Targeted Hellos to the peer and can respond to the Targeted Hellos received from the peer.
   targeted-peer { *ipv4-address* | *ipv6-address* } [*peer-name*]
   By default, the device does not send Targeted Hellos to a peer, or respond to Targeted Hellos received from a peer.
4. Set the Targeted Hello hold time.
   mpls ldp timer hello-hold *timeout*
   By default, the Targeted Hello hold time is 45 seconds.
5. Set the Target Hello interval.
   mpls ldp timer hello-interval *interval*
   By default, the Targeted Hello interval is 15 seconds.

Configuring LDP session parameters

About LDP session parameters

An LDP session has the following parameters:

- Keepalive hold time and Keepalive interval.
- LDP transport address—IP address for establishing TCP connections.

LDP uses Basic Discovery and Extended Discovery mechanisms to discovery LDP peers and establish LDP sessions with them.

Restrictions and guidelines for configuring LDP session parameters

When you configure LDP session parameters, follow these guidelines:

- The configured LDP transport address must be the IP address of an up interface on the device. Otherwise, no LDP session can be established.
- Make sure the LDP transport addresses of the local and peer LSRs can reach each other. Otherwise, no TCP connection can be established.
- Changes to LDP session parameters do not take effect on established LDP sessions. To apply the changes to existing LDP sessions on a network, you must use the **reset mpls ldp** command to reset all LDP sessions on that network.

Configuring LDP sessions parameters for Basic Discovery mechanism

1. Enter system view.

   system-view
2. Enter the view of the LDP session interface (the interface where you want to establish an LDP session).

   interface *interface-type interface-number*
3. Set the Keepalive hold time.

   mpls ldp timer keepalive-hold *timeout*

   By default, the Keepalive hold time is 45 seconds.
4. Set the Keepalive interval.

   mpls ldp timer keepalive-interval *interval*

   By default, the Keepalive interval is 15 seconds.
5. Configure the IPv4 LDP transport address.

   mpls ldp transport-address { *ipv4-address* | **interface** }

   The default LDP IPv4 transport address is the local LSR ID if the interface where you want to establish an LDP session belongs to the public network. If the interface belongs to a VPN, the default LDP IPv4 transport address is the primary IP address of the interface.

   If the LDP session interface belongs to a VPN instance, the interface with the IP address specified by this command must belong to the same VPN instance.
6. Configure the IPv6 LDP transport address.

   mpls ldp transport-address *ipv6-address*

   By default, no IPv6 LDP transport address is specified.

Configuring LDP IPv4 sessions parameters for Extended Discovery mechanism

1. Enter system view.

   system-view

2. Enter LDP view.

   mpls ldp

3. Specify an IPv4 LDP peer and enter LDP peer view. The device will send unsolicited IPv4 Targeted Hellos to the peer and can respond to IPv4 Targeted Hellos received from the targeted peer.

   targeted-peer *ipv4-address* [*peer-name*]

   By default, the device does not send IPv4 Targeted Hellos to a peer, or respond to IPv4 Targeted Hellos received from a peer.

4. Set the Keepalive hold time.

   mpls ldp timer keepalive-hold *timeout*

   By default, the Keepalive hold time is 45 seconds.

5. Set the Keepalive interval.

   mpls ldp timer keepalive-interval *interval*

   By default, the Keepalive interval is 15 seconds.

6. Configure the LDP transport address.

   mpls ldp transport-address *ipv4-address*

   By default, the LDP transport address is the LSR ID of the local device.

Configuring LDP IPv6 sessions parameters for Extended Discovery mechanism

1. Enter system view.

   system-view

2. Enter LDP view.

   mpls ldp

3. Specify an IPv6 LDP peer and enter LDP peer view. The device will send unsolicited IPv6 Targeted Hellos to the peer and can respond to IPv6 Targeted Hellos received from the targeted IPv6 peer.

   targeted-peer *ipv6-address* [*peer-name*]

   By default, the device does not send IPv6 Targeted Hellos to a peer, or respond to IPv6 Targeted Hellos received from a peer.

4. Set the Keepalive hold time.

   mpls ldp timer keepalive-hold *timeout*

   By default, the Keepalive hold time is 45 seconds.

5. Set the Keepalive interval.

   mpls ldp timer keepalive-interval *interval*

   By default, the Keepalive interval is 15 seconds.

6. Configure the LDP transport address.

   mpls ldp transport-address *ipv6-address*

   By default, the LDP IPv6 transport address is not configured.

Configuring LDP backoff

About this task

If LDP session parameters (for example, the label advertisement mode) are incompatible, two LDP peers cannot establish a session, and they will keep negotiating with each other.

The LDP backoff mechanism can mitigate this problem by using an initial delay timer and a maximum delay timer. After failing to establish a session to a peer LSR for the first time, LDP does not start an attempt until the initial delay timer expires. If the session setup fails again, LDP waits for two times the initial delay before the next attempt, and so forth until the maximum delay time is reached. After that, the maximum delay time will always take effect.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view or enter LDP-VPN instance view.
   - Enter LDP view:
     mpls ldp
   - Execute the following commands in sequence to enter LDP-VPN instance view:
     mpls ldp
     vpn-instance *vpn-instance-name*
3. Set the initial delay time and maximum delay time.
   backoff initial *initial-time* **maximum** *maximum-time*
   By default, the initial delay time is 15 seconds, and the maximum delay time is 120 seconds.

Setting a DSCP value for outgoing LDP packets

About this task

To control the transmission preference of outgoing LDP packets, set a DSCP value for outgoing LDP packets.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view.
   mpls ldp
3. Set a DSCP value for outgoing LDP packets.
   dscp *dscp-value*
   By default, the DSCP value for outgoing LDP packets is 48.

Configuring LDP to redistribute BGP unicast routes

About this task

By default, LDP automatically redistributes IGP routes, including the BGP routes that have been redistributed into IGP. Then, LDP assigns labels to the IGP routes and labeled BGP routes, if these

routes are permitted by an LSP generation policy. LDP does not automatically redistribute BGP unicast routes if the routes are not redistributed into the IGP.

For example, on a carrier's carrier network where IGP is not configured between a PE of a Level 1 carrier and a CE of a Level 2 carrier, LDP cannot redistribute BGP unicast routes to assign labels to them. For this network to operate correctly, you can enable LDP to redistribute BGP unicast routes. If the routes are permitted by an LSP generation policy, LDP assigns labels to them to establish LSPs. For more information about carrier's carrier, see "Configuring MPLS L3VPN."

Procedure

1. Enter system view.
   system-view
2. Enter LDP view or enter LDP-VPN instance view.
   - Enter LDP view.
     mpls ldp
   - Execute the following commands in sequence to enter LDP-VPN instance view:
     mpls ldp
     vpn-instance *vpn-instance-name*
3. Enable LDP to redistribute BGP IPv4 unicast routes.
   import bgp [*as-number*]
   By default, LDP does not redistribute BGP IPv4 unicast routes.
4. Enable LDP to redistribute BGP IPv6 unicast routes.
   ipv6 import bgp [*as-number*]
   By default, LDP does not redistribute BGP IPv6 unicast routes.

Configuring an LSP generation policy

About this task

LDP assigns labels to the routes that have been redistributed into LDP to generate LSPs. An LSP generation policy specifies which redistributed routes can be used by LDP to generate LSPs to control the number of LSPs, as follows:

- Use all routes to establish LSPs.
  On loopback interfaces, only routes with a 32-bit mask can trigger LDP to establish LSPs.
- Use the routes permitted by an IP prefix list to establish LSPs. For information about IP prefix list configuration, see *Layer 3—IP Routing Configuration Guide*.
- Use only IPv4 host routes with a 32-bit mask or IPv6 host routes with a 128-bit prefix to establish LSPs.

Restrictions and guidelines

By default, LDP uses only IPv4 host routes with a 32-bit mask or IPv6 host routes with a 128-bit prefix to establish LSPs. The other two methods can result in more LSPs than the default policy. To change the policy, make sure the system resources and bandwidth resources are sufficient.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view or enter LDP-VPN instance view.
   - Enter LDP view.

mpls ldp

- Execute the following commands in sequence to enter LDP-VPN instance view:

  mpls ldp

  vpn-instance *vpn-instance-name*

3. Configure an IPv4 LSP generation policy.

   lsp-trigger { **all** | **prefix-list** *prefix-list-name* }

   By default, LDP uses only the redistributed IPv4 routes with a 32-bit mask to establish LSPs.

4. Configure an IPv6 LSP generation policy.

   ipv6 lsp-trigger { **all** | **prefix-list** *prefix-list-name* }

   By default, LDP uses only the redistributed IPv6 routes with a 128-bit prefix to establish LSPs.

Configuring a label mapping propagation policy

About this task

In an LDP network, LSRs establish LSPs by advertising label mapping messages. When a PE acts as both an ingress LSR and a transit LSR, there might be a large number of transit LSPs on the PE. You can configure a label mapping propagation policy on the PE to reduce the number of transit LSPs and thus save network resources. A label mapping propagation policy allows the PE to advertise only permitted label mappings to the peers to establish transit LSPs.

Restrictions and guidelines

If you execute both the **lsp-trigger** and **propagate mapping** commands, establishment of transit LSPs is controlled by the **propagate mapping** command.

Procedure

1. Enter system view.

   system-view

2. Enter LDP view or LDP-VPN instance view.

   - Enter LDP view.

     mpls ldp

   - Execute the following commands in sequence to enter LDP-VPN instance view:

     mpls ldp

     vpn-instance *vpn-instance-name*

3. Configure a label mapping propagation policy.

   propagate mapping { **none** | **prefix-list** *prefix-list-name* }

   By default, no label mapping propagation policy is configured. The device advertises all IPv4 prefix label mappings permitted by the LSP generation policy.

Configuring the LDP label distribution control mode

1. Enter system view.

   system-view

2. Enter LDP view or enter LDP-VPN instance view.

- Enter LDP view.
  mpls ldp
- Execute the following commands in sequence to enter LDP-VPN instance view:
  mpls ldp
  vpn-instance *vpn-instance-name*

3. Configure the label distribution control mode.
   label-distribution { **independent** | **ordered** }
   By default, the Ordered label distribution mode is used.

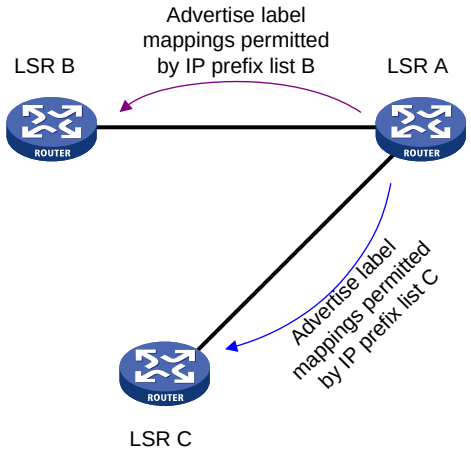
Configuring a label advertisement policy

About this task

A label advertisement policy uses IP prefix lists to control the FEC-label mappings advertised to peers.

As shown in Figure 8, LSR A advertises label mappings for FECs permitted by IP prefix list B to LSR B. It advertises label mappings for FECs permitted by IP prefix list C to LSR C.

Figure 8 Label advertisement control diagram



Restrictions and guidelines

A label advertisement policy on an LSR and a label acceptance policy on its upstream LSR can achieve the same purpose. As a best practice, use label advertisement policies to reduce network load if downstream LSRs support label advertisement control.

Before you configure an LDP label advertisement policy, create an IP prefix list. For information about IP prefix list configuration, see routing policy configuration in *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view or enter LDP-VPN instance view.
   - Enter LDP view.
     mpls ldp
   - Execute the following commands in sequence to enter LDP-VPN instance view:
     mpls ldp

vpn-instance *vpn-instance-name*

3. Configure an IPv4 label advertisement policy.

   advertise-label prefix-list *prefix-list-name* [**peer** *peer-prefix-list-name*]

   By default, LDP advertises all IPv4 FEC-label mappings permitted by the LSP generation policy to all peers.

4. Configure an IPv6 label advertisement policy.

   ipv6 advertise-label prefix-list *prefix-list-name* [**peer** *peer-prefix-list-name*]

   By default, LDP advertises all IPv6 FEC-label mappings permitted by the LSP generation policy to all peers.

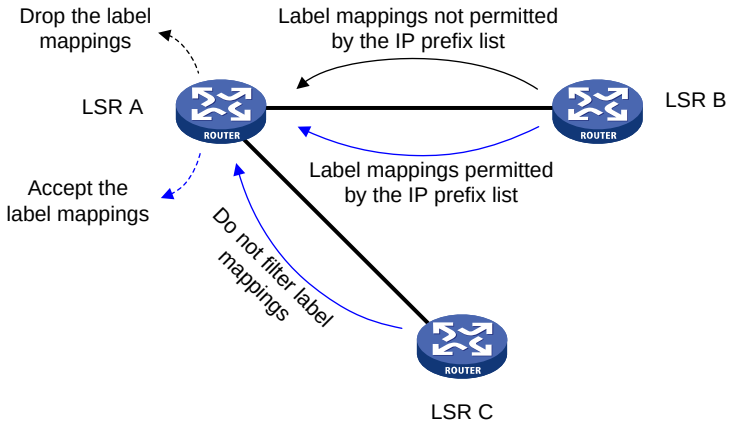
Configuring a label acceptance policy

About this task

A label acceptance policy uses an IP prefix list to control the label mappings received from a peer.

As shown in Figure 9, LSR A uses an IP prefix list to filter label mappings from LSR B, and it does not filter label mappings from LSR C.

Figure 9 Label acceptance control diagram



Restrictions and guidelines

A label advertisement policy on an LSR and a label acceptance policy on its upstream LSR can achieve the same purpose. As a best practice, use the label advertisement policy to reduce network load.

You must create an IP prefix list before you configure a label acceptance policy. For information about IP prefix list configuration, see routing policy configuration in *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter LDP view or enter LDP-VPN instance view.

   - Enter LDP view.

     mpls ldp

- Execute the following commands in sequence to enter LDP-VPN instance view:
  mpls ldp
  vpn-instance *vpn-instance-name*

3. Configure an IPv4 label acceptance policy.
   accept-label peer *peer-lsr-id* **prefix-list** *prefix-list-name*
   By default, LDP accepts all IPv4 FEC-label mappings.
4. Configure an IPv6 label acceptance policy.
   ipv6 accept-label peer *peer-lsr-id* **prefix-list** *prefix-list-name*
   By default, LDP accepts all IPv6 FEC-label mappings.

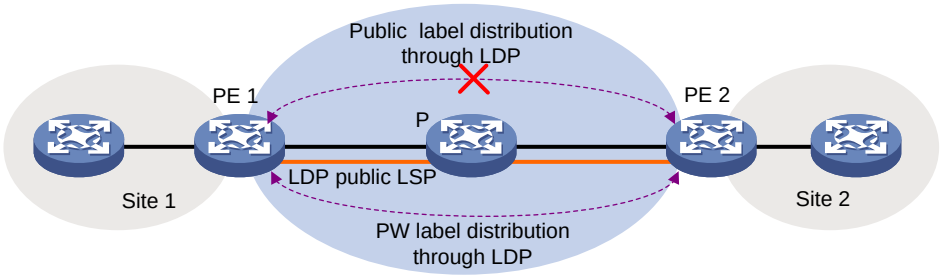
Disabling distribution of public label mappings to remote LDP peers

About this task

As shown in Figure 10, PEs on an MPLS L2VPN or VPLS network that uses LDP as the signaling protocol must establish remote LDP sessions to distribute the label mappings for PW label assignment. Typically, the public tunnel that transports PWs is established hop by hop with P devices in between, rather than directly between PEs. It is not necessary to distribute public labels on the remote LDP session between PEs.

By default, LDP distributes public label mappings to remote peers through remote LDP sessions. This causes PEs to store public labels unnecessarily. To conserve system resources, disable LDP from distributing public label mappings to its remote peers.

Figure 10 Disabling distribution of public label mappings to remote LDP peers



Restrictions and guidelines

This feature does not affect PW label distribution via LDP for L2VPN.

To prevent LDP from distributing public label mappings to any remote peers, use the **targeted-peer label-mapping restricted** command in LDP view. To prevent LDP from distributing public label mappings to a remote peer, use the **label-mapping disable** command in LDP peer view. For a remote LDP peer, the configuration in its peer view takes priority over the configuration in LDP view. The configuration in LDP view takes effect only if no setting is configured for this feature in LDP peer view.

This feature does not apply to the remote LDP sessions established by using the **accept target-hello** command or by the remote LFA feature. These types of remote LDP sessions can distribute public label mappings, whether this feature is enabled or not.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view.
   mpls ldp
3. Disable distribution of public label mappings to any remote peers
   targeted-peer label-mapping restricted
   By default, the system allows distribution of public label mappings to remote peers.
4. Enter LDP peer view.
   targeted-peer *ipv4-address*
5. Enable or disable distribution of public label mappings to the specified remote LDP peer.
   label-mapping { **enable** | **disable** }
   By default, distribution of public label mappings to remote peers depends on the setting for the **targeted-peer label-mapping restricted** command in LDP view.
   To have this command take effect, make sure the address specified in the **targeted-peer** command is the LSR ID of a remote device.

Configuring LDP MD5 authentication

About this task

To improve security for LDP sessions, you can configure MD5 authentication for the underlying TCP connections to check the integrity of LDP messages.

Restrictions and guidelines

For two LDP peers to establish an LDP session successfully, make sure the LDP MD5 authentication configuration on the LDP peers are consistent.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view or enter LDP-VPN instance view.
   - Enter LDP view.
     mpls ldp
   - Execute the following commands in sequence to enter LDP-VPN instance view:
     mpls ldp
     vpn-instance *vpn-instance-name*
3. Enable LDP MD5 authentication.
   md5-authentication *peer-lsr-id* { **cipher** | **plain** } *string*
   By default, LDP MD5 authentication is disabled.

Configuring LDP loop detection parameters

About this task

The device does not support LDP loop detection. However, you must configure LDP loop detection parameters on the device so the device can establish an LDP session with a neighbor in the following situation:

- The neighbor device supports LDP loop detection.
- The neighbor device requires both devices of the LDP session to advertise the same loop detection settings during LDP session negotiation.

Restrictions and guidelines

Configuring LDP loop detection parameters enables the device to have the loop detection negotiation capability during LDP session establishment. The device still cannot perform LDP loop detection.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view or enter LDP-VPN instance view.
   - Enter LDP view.
     mpls ldp
   - Execute the following commands in sequence to enter LDP-VPN instance view:
     mpls ldp
     vpn-instance *vpn-instance-name*
3. Enable loop detection.
   loop-detect
   By default, loop detection is disabled.
4. (Optional.) Set the maximum hop count.
   maxhops *hop-number*
   By default, the maximum hop count is 32.
5. (Optional.) Set the path vector limit.
   pv-limit *pv-number*
   By default, the path vector limit is 32.

Configuring LDP session protection

About this task

If two LDP peers have both a direct link and an indirect link in between, you can configure this feature to protect their LDP session when the direct link fails.

LDP establishes both a Link Hello adjacency over the direct link and a Targeted Hello adjacency over the indirect link with the peer. When the direct link fails, LDP deletes the Link Hello adjacency but still maintains the Targeted Hello adjacency. In this way, the LDP session between the two peers is kept available, and the FEC-label mappings based on this session are not deleted. When the direct link recovers, the LDP peers do not need to re-establish the LDP session or re-learn the FEC-label mappings.

When you enable the session protection feature, you can also specify the session protection duration. If the Link Hello adjacency does not recover within the duration, LDP deletes the Targeted Hello adjacency and the LDP session. If you do not specify the session protection duration, the two peers will always maintain the LDP session over the Targeted Hello adjacency.

Restrictions and guidelines

LDP session protection is applicable only to IPv4 networks.

Procedure

1. Enter system view.

system-view

2. Enter LDP view.
   mpls ldp
3. Enable the session protection feature.
   session protection [**duration** *time*] [**peer** *peer-prefix-list-name*]
   By default, session protection is disabled.

Configuring LDP GR

Prerequisites

Before you configure LDP GR, enable LDP on the GR restarter and GR helpers.

Restrictions and guidelines

The LDP GR configuration is required only on a GR restarter. Because the role (GR restarter or helper) of a device in a GR process is unpredictable, configure LDP GR on all involved devices as a best practice.

Procedure

1. Enter system view.
   system-view
2. Enter LDP view.
   mpls ldp
3. Enable LDP GR.
   graceful-restart
   By default, LDP GR is disabled.
4. (Optional.) Set the Reconnect timer for LDP GR.
   graceful-restart timer reconnect *reconnect-time*
   By default, the Reconnect time is 120 seconds.
5. (Optional.) Set the MPLS Forwarding State Holding timer for LDP GR.
   graceful-restart timer forwarding-hold *hold-time*
   By default, the MPLS Forwarding State Holding time is 180 seconds.

Configuring LDP NSR

1. Enter system view.
   system-view
2. Enter LDP view.
   mpls ldp
3. Enable LDP NSR.
   non-stop-routing
   By default, LDP NSR is disabled.

Configuring LDP-IGP synchronization

Restrictions and guidelines for LDP-IGP synchronization

After you enable LDP-IGP synchronization for an OSPF process, OSPF area, or an IS-IS process, LDP-IGP synchronization is enabled on the OSPF process interfaces or the IS-IS process interfaces.

You can execute the **mpls ldp igp sync disable** command to disable LDP-IGP synchronization on interfaces where LDP-IGP synchronization is not required.

LDP-IGP synchronization protection is only applicable to an IPv4 network.

LDP-IGP synchronization is not supported for an OSPF process and its OSPF areas if the OSPF process belongs to a VPN instance.

LDP-IGP synchronization is not supported for an IS-IS process that belongs to a VPN instance.

Configuring LDP-OSPF synchronization

Configuring LDP-OSPF synchronization for an OSPF process

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id*] *
3. Enable LDP-OSPF synchronization.
   mpls ldp sync
   By default, LDP-OSPF synchronization is disabled.
4. (Optional.) Disable LDP-OSPF synchronization on an interface:
   a. Return to system view.
      quit
   b. Enter interface view.
      interface *interface-type interface-number*
   c. Disable LDP-IGP synchronization on the interface.
      mpls ldp igp sync disable
      By default, LDP-IGP synchronization is enabled on an interface.
5. (Optional.) Configure LDP-OSPF synchronization parameters:
   a. Return to system view.
      quit
   b. Enter LDP view.
      mpls ldp
   c. Set the delay for LDP to notify IGP of the LDP convergence.
      igp sync delay *time*
      By default, LDP immediately notifies IGP of the LDP convergence completion.
   d. Set the maximum delay for LDP to notify IGP of the LDP-IGP synchronization status after an LDP restart or active/standby switchover.
      igp sync delay on-restart *time*
      By default, the maximum notification delay is 90 seconds.

Configuring LDP-OSPF synchronization for an OSPF area

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id*] *
3. Enter area view.
   area *area-id*
4. Enable LDP-OSPF synchronization.
   mpls ldp sync
   By default, LDP-OSPF synchronization is disabled.
5. (Optional.) Disable LDP-OSPF synchronization on an interface:
   a. Return to system view.
      quit
   b. Enter interface view.
      interface *interface-type interface-number*
   c. Disable LDP-IGP synchronization on the interface.
      mpls ldp igp sync disable
      By default, LDP-IGP synchronization is enabled on an interface.
6. (Optional.) Configure LDP-OSPF synchronization parameters:
   a. Return to system view.
      quit
   b. Enter LDP view.
      mpls ldp
   c. Set the delay for LDP to notify IGP of the LDP convergence.
      igp sync delay *time*
      By default, LDP immediately notifies IGP of the LDP convergence completion.
   d. Set the maximum delay for LDP to notify IGP of the LDP-IGP synchronization status after an LDP restart or active/standby switchover.
      igp sync delay on-restart *time*
      By default, the maximum notification delay is 90 seconds.

Configuring LDP IS-IS synchronization

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis [*process-id*]
3. Enable LDP-ISIS synchronization.
   mpls ldp sync [**level-1** | **level-2**]
   By default, LDP-ISIS synchronization is disabled.
4. (Optional.) Disable LDP-ISIS synchronization on an interface:
   a. Return to system view.
      quit
   b. Enter interface view.

interface *interface-type interface-number*

c. Disable LDP-IGP synchronization on the interface.

mpls ldp igp sync disable

By default, LDP-IGP synchronization is enabled on an interface.

5. (Optional.) Configure LDP-IGP synchronization parameters:

a. Return to system view.

quit

b. Enter LDP view.

mpls ldp

c. Set the delay for LDP to notify IGP of the LDP convergence completion.

igp sync delay *time*

By default, LDP immediately notifies IGP of the LDP convergence completion.

d. Set the maximum delay for LDP to notify IGP of the LDP-IGP synchronization status after an LDP restart or an active/standby switchover occurs.

igp sync delay on-restart *time*

By default, the maximum notification delay is 90 seconds.

Configuring LDP FRR

Configuring LDP LFA FRR

LDP FRR is based on IP FRR, and is enabled automatically after IP FRR is enabled. For information about configuring IP FRR, see *Layer 3—IP Routing Configuration Guide*.

Configuring LDP remote LFA FRR

Restrictions and guidelines

In a remote LFA network, configure this feature on the PQ nodes.

Procedure

1. Enter system view.

system-view

2. Enter LDP view.

mpls ldp

3. Enable the device to automatically set up targeted LDP sessions on receiving Targeted Hellos.

accept target-hello { **all** | **prefix-list** *prefix-list-name* }

By default, the device does not automatically establish targeted LDP sessions when it receives Targeted Hellos.

Enabling SNMP notifications for LDP

About this task

This command enables generating SNMP notifications for LDP upon LDP session changes, as defined in RFC 3815. For LDP event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for LDP.
   snmp-agent trap enable ldp
   By default, SNMP notifications for LDP are enabled.

Enabling private SNMP notifications for LDP

About this task

This feature enables generating SNMP notifications for the following events:

- LDP configuration consistency checks on multiple links between LDP peers.
- The number of LDP sessions exceeds the upper threshold.
- The number of LDP sessions drops below the lower threshold from the upper threshold.

These notifications are not defined in an RFC and are private LDP notifications supported by Comware. For SNMP to output the private LDP notifications correctly, you must also configure parameters for sending SNMP notifications. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable private SNMP notifications for LDP.
   snmp-agent trap enable ldp-private
   By default, private SNMP notifications for LDP are disabled.

Display and maintenance commands for LDP

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display LDP discovery information.	**display mpls ldp discovery** [**vpn-instance** *vpn-instance-name*] [[**interface** *interface-type interface-number* | **peer** *peer-lsr-id*] [**ipv6**] | **targeted-peer** { *ipv4-address* | *ipv6-address* }] [**verbose**]
Display error TCP packets received by the LDP session.	**display mpls ldp error-packet**
Display LDP FEC-label mapping information.	**display mpls ldp fec** [**vpn-instance** *vpn-instance-name*] [*ipv4-address mask-length* | *ipv6-address prefix-length* | [**ipv6**] [**summary**]]
Display LDP-IGP synchronization information.	**display mpls ldp igp sync** [**interface** *interface-type interface-number*]

Task	Command
Display LDP interface information.	**display mpls ldp interface** [**vpn-instance** *vpn-instance-name*] [*interface-type interface-number*] [**ipv6**]
Display LDP LSP information.	**display mpls ldp lsp** [**vpn-instance** *vpn-instance-name*] [*ipv4-address mask-length* | *ipv6-address prefix-length* | **ipv6**]
Display LDP running parameters.	**display mpls ldp parameter** [**vpn-instance** *vpn-instance-name*]
Display LDP peer and session information.	**display mpls ldp peer** [**vpn-instance** *vpn-instance-name*] [*peer-lsr-id*] [**verbose**]
Display LDP summary information.	**display mpls ldp summary** [**all** | **vpn-instance** *vpn-instance-name*]
Display information about LDP session interruptions.	**display mpls ldp troubleshooting** [[**vpn-instance** *vpn-instance-name*] **peer** *peer-lsr-id*]
Reset LDP sessions.	**reset mpls ldp** [**vpn-instance** *vpn-instance-name*] [**peer** *peer-id*]

IPv4 LDP configuration examples

Example: Configuring LDP LSPs

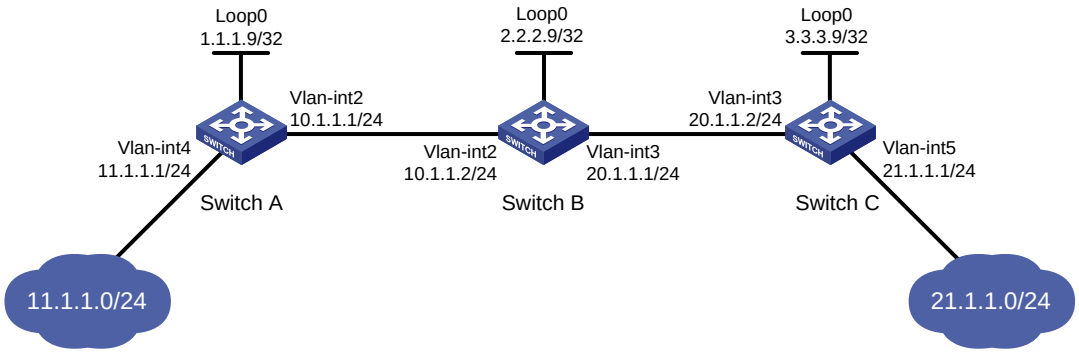
Network configuration

Switch A, Switch B, and Switch C all support MPLS.

Configure LDP to establish LSPs between Switch A and Switch C, so subnets 11.1.1.0/24 and 21.1.1.0/24 can reach each other over MPLS.

Configure LDP to establish LSPs only for destinations 1.1.1.9/32, 2.2.2.9/32, 3.3.3.9/32, 11.1.1.0/24, and 21.1.1.0/24 on Switch A, Switch B, and Switch C.

Figure 11 Network diagram



Requirements analysis

- To ensure that the LSRs establish IPv4 LSPs automatically, enable IPv4 LDP on each LSR.
- To establish IPv4 LDP LSPs, configure an IPv4 routing protocol to ensure IP connectivity between the LSRs. This example uses OSPF.
- To control the number of IPv4 LSPs, configure an IPv4 LSP generation policy on each LSR.

Procedure

1. Create VLANs and assign IP addresses to interfaces:

Configure Switch A.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/1
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 10.1.1.1 24
[SwitchA-Vlan-interface2] quit
[SwitchA] vlan 4
[SwitchA-vlan4] port gigabitethernet 1/0/2
[SwitchA-vlan4] quit
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ip address 11.1.1.1 24
[SwitchA-Vlan-interface4] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.9 32
[SwitchA-LoopBack0] quit
```

Configure Switch B.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 10.1.1.2 24
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ip address 20.1.1.1 24
[SwitchB-Vlan-interface3] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.9 32
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] vlan 3
```

```
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ip address 20.1.1.2 24
[SwitchC-Vlan-interface3] quit
[SwitchC] vlan 5
[SwitchC-vlan5] port gigabitethernet 1/0/2
[SwitchC-vlan5] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ip address 21.1.1.1 24
[SwitchC-Vlan-interface5] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 3.3.3.9 32
[SwitchC-LoopBack0] quit
```

2. Configure OSPF on each switch to ensure IP connectivity between them:

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 11.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[SwitchC-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 21.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Verify that the switches have learned the routes to each other. This example uses Switch A.

```
[SwitchA] display ip routing-table

Destinations : 18       Routes : 18

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
1.1.1.9/32         Direct  0   0           127.0.0.1       InLoop0
2.2.2.9/32         O_INTRA 10  1           10.1.1.2        Vlan2
```

```
3.3.3.9/32          O_INTRA 10  2           10.1.1.2        Vlan2
10.1.1.0/24         Direct  0   0           10.1.1.1        Vlan2
10.1.1.1/32         Direct  0   0           127.0.0.1       InLoop0
10.1.1.255/32       Direct  0   0           10.1.1.1        Vlan2
11.1.1.0/24         Direct  0   0           11.1.1.1        Vlan4
11.1.1.1/32         Direct  0   0           127.0.0.1       InLoop0
11.1.1.255/32       Direct  0   0           11.1.1.1        Vlan4
20.1.1.0/24         O_INTRA 10  2           10.1.1.2        Vlan2
21.1.1.0/24         O_INTRA 10  3           10.1.1.2        Vlan2
127.0.0.0/8         Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4         Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24        Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

3. Enable MPLS and IPv4 LDP:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.9
[SwitchA] mpls ldp
[SwitchA-ldp] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] mpls enable
[SwitchA-Vlan-interface2] mpls ldp enable
[SwitchA-Vlan-interface2] quit
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.9
[SwitchB] mpls ldp
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] mpls enable
[SwitchB-Vlan-interface2] mpls ldp enable
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mpls enable
[SwitchB-Vlan-interface3] mpls ldp enable
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.9
[SwitchC] mpls ldp
[SwitchC-ldp] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] mpls enable
[SwitchC-Vlan-interface3] mpls ldp enable
[SwitchC-Vlan-interface3] quit
```

4. Configure IPv4 LSP generation policies:

On Switch A, create IP prefix list **switcha**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchA] ip prefix-list switcha index 10 permit 1.1.1.9 32
```

```
[SwitchA] ip prefix-list switcha index 20 permit 2.2.2.9 32
[SwitchA] ip prefix-list switcha index 30 permit 3.3.3.9 32
[SwitchA] ip prefix-list switcha index 40 permit 11.1.1.0 24
[SwitchA] ip prefix-list switcha index 50 permit 21.1.1.0 24
[SwitchA] mpls ldp
[SwitchA-ldp] lsp-trigger prefix-list switcha
[SwitchA-ldp] quit
```

On Switch B, create IP prefix list **switchb**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchB] ip prefix-list switchb index 10 permit 1.1.1.9 32
[SwitchB] ip prefix-list switchb index 20 permit 2.2.2.9 32
[SwitchB] ip prefix-list switchb index 30 permit 3.3.3.9 32
[SwitchB] ip prefix-list switchb index 40 permit 11.1.1.0 24
[SwitchB] ip prefix-list switchb index 50 permit 21.1.1.0 24
[SwitchB] mpls ldp
[SwitchB-ldp] lsp-trigger prefix-list switchb
[SwitchB-ldp] quit
```

On Switch C, create IP prefix list **switchc**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchC] ip prefix-list switchc index 10 permit 1.1.1.9 32
[SwitchC] ip prefix-list switchc index 20 permit 2.2.2.9 32
[SwitchC] ip prefix-list switchc index 30 permit 3.3.3.9 32
[SwitchC] ip prefix-list switchc index 40 permit 11.1.1.0 24
[SwitchC] ip prefix-list switchc index 50 permit 21.1.1.0 24
[SwitchC] mpls ldp
[SwitchC-ldp] lsp-trigger prefix-list switchc
[SwitchC-ldp] quit
```

Verifying the configuration

Display LDP LSP information on the switches, for example, on Switch A.

```
[SwitchA] display mpls ldp lsp
Status Flags: * - stale, L - liberal, B - backup, N/A – unavailable
FECs: 5           Ingress: 3          Transit: 3      Egress: 2

FEC               In/Out Label        Nexthop         OutInterface/LSINDEX
1.1.1.9/32        3/-
                  -/1279(L)
2.2.2.9/32        -/3                 10.1.1.2        Vlan2
                  1279/3              10.1.1.2        Vlan2
3.3.3.9/32        -/1278              10.1.1.2        Vlan2
                  1278/1278           10.1.1.2        Vlan2
11.1.1.0/24       1277/-
                  -/1277(L)
21.1.1.0/24       -/1276              10.1.1.2        Vlan2
                  1276/1276           10.1.1.2        Vlan2
```

Test the connectivity of the LDP LSP from Switch A to Switch C.

```
[SwitchA] ping mpls -a 11.1.1.1 ipv4 21.1.1.0 24
MPLS ping FEC 21.1.1.0/24 with 100 bytes of data:
100 bytes from 20.1.1.2: Sequence=1 time=1 ms
```

```
100 bytes from 20.1.1.2: Sequence=2 time=1 ms
100 bytes from 20.1.1.2: Sequence=3 time=8 ms
100 bytes from 20.1.1.2: Sequence=4 time=2 ms
100 bytes from 20.1.1.2: Sequence=5 time=1 ms

--- Ping statistics for FEC 21.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/2/8 ms
```

Test the connectivity of the LDP LSP from Switch C to Switch A.

```
[SwitchC] ping mpls -a 21.1.1.1 ipv4 11.1.1.0 24
MPLS ping FEC 11.1.1.0/24 with 100 bytes of data:
100 bytes from 10.1.1.1: Sequence=1 time=1 ms
100 bytes from 10.1.1.1: Sequence=2 time=1 ms
100 bytes from 10.1.1.1: Sequence=3 time=1 ms
100 bytes from 10.1.1.1: Sequence=4 time=1 ms
100 bytes from 10.1.1.1: Sequence=5 time=1 ms

--- Ping statistics for FEC 11.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/1/1 ms
```

Configuration files

- Switch A

```
#
 sysname SwitchA
#
ospf 1
 area 0.0.0.0
  network 1.1.1.9 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 11.1.1.0 0.0.0.255
#
 mpls lsr-id 1.1.1.9
#
vlan 2
#
vlan 4
#
mpls ldp
 lsp-trigger prefix-list switcha
#
interface LoopBack0
 ip address 1.1.1.9 255.255.255.255
#
interface Vlan-interface2
 ip address 10.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
```

```
#
interface Vlan-interface4
 ip address 11.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 4
#
 ip prefix-list switcha index 10 permit 1.1.1.9 32
 ip prefix-list switcha index 20 permit 2.2.2.9 32
 ip prefix-list switcha index 30 permit 3.3.3.9 32
 ip prefix-list switcha index 40 permit 11.1.1.0 24
 ip prefix-list switcha index 50 permit 21.1.1.0 24
#
return
```

- Switch B

```
#
 sysname SwitchB
#
ospf 1
 area 0.0.0.0
  network 2.2.2.9 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 20.1.1.0 0.0.0.255
#
 mpls lsr-id 2.2.2.9
#
vlan 2 to 3
#
mpls ldp
 lsp-trigger prefix-list switchb
#
interface LoopBack0
 ip address 2.2.2.9 255.255.255.255
#
interface Vlan-interface2
 ip address 10.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface3
 ip address 20.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
```

```
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
 ip prefix-list switchb index 10 permit 1.1.1.9 32
 ip prefix-list switchb index 20 permit 2.2.2.9 32
 ip prefix-list switchb index 30 permit 3.3.3.9 32
 ip prefix-list switchb index 40 permit 11.1.1.0 24
 ip prefix-list switchb index 50 permit 21.1.1.0 24
#
return
```

- Switch C

```
#
 sysname SwitchC
#
ospf 1
 area 0.0.0.0
  network 3.3.3.9 0.0.0.0
  network 20.1.1.0 0.0.0.255
  network 21.1.1.0 0.0.0.255
#
 mpls lsr-id 3.3.3.9
#
vlan 3
#
vlan 5
#
mpls ldp
 lsp-trigger prefix-list switchc
#
interface LoopBack0
 ip address 3.3.3.9 255.255.255.255
#
interface Vlan-interface3
 ip address 20.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface5
 ip address 21.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
```

```
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 5
#
 ip prefix-list switchc index 10 permit 1.1.1.9 32
 ip prefix-list switchc index 20 permit 2.2.2.9 32
 ip prefix-list switchc index 30 permit 3.3.3.9 32
 ip prefix-list switchc index 40 permit 11.1.1.0 24
 ip prefix-list switchc index 50 permit 21.1.1.0 24
#
return
```

Example: Configuring label acceptance control

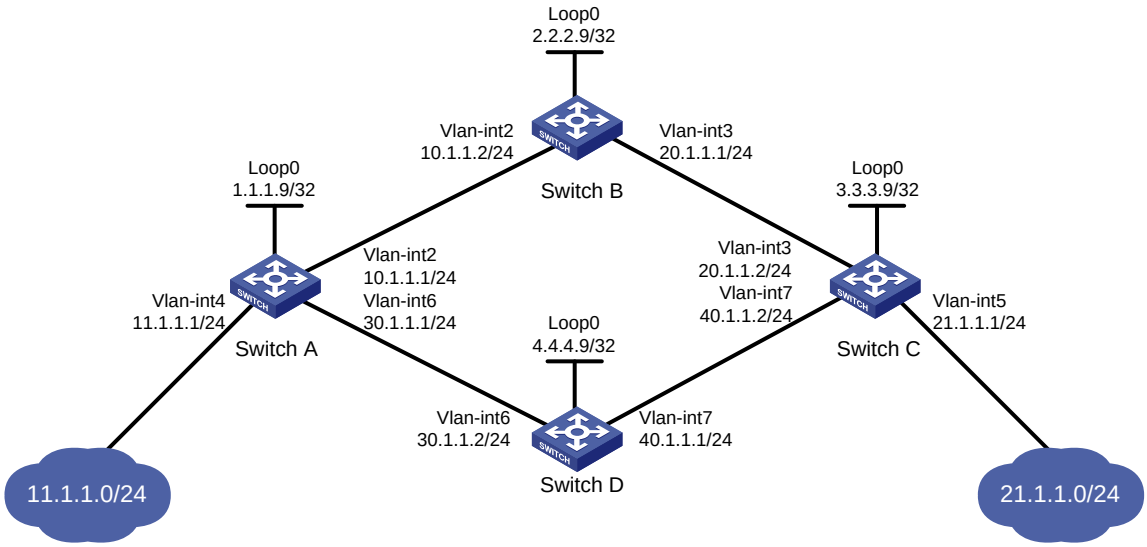
Network configuration

Two links, Switch A—Switch B—Switch C and Switch A—Switch D—Switch C, exist between subnets 11.1.1.0/24 and 21.1.1.0/24.

Configure LDP to establish LSPs only for routes to subnets 11.1.1.0/24 and 21.1.1.0/24.

Configure LDP to establish LSPs only on the link Switch A—Switch B—Switch C to forward traffic between subnets 11.1.1.0/24 and 21.1.1.0/24.

Figure 12 Network diagram



Requirements analysis

- To ensure that the LSRs establish IPv4 LSPs automatically, enable IPv4 LDP on each LSR.
- To establish IPv4 LDP LSPs, configure an IPv4 routing protocol to ensure IP connectivity between the LSRs. This example uses OSPF.
- To ensure that LDP establishes IPv4 LSPs only for the routes 11.1.1.0/24 and 21.1.1.0/24, configure IPv4 LSP generation policies on each LSR.
- To ensure that LDP establishes IPv4 LSPs only over the link Switch A—Switch B—Switch C, configure IPv4 label acceptance policies as follows:

- Switch A accepts only the label mapping for FEC 21.1.1.0/24 received from Switch B. Switch A denies the label mapping for FEC 21.1.1.0/24 received from Switch D.
- Switch C accepts only the label mapping for FEC 11.1.1.0/24 received from Switch B. Switch C denies the label mapping for FEC 11.1.1.0/24 received from Switch D.

Prerequisites

Before configuration, disable the spanning tree feature globally or map each VLAN to an MSTI. For more information, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Create VLANs and assign IP addresses to interfaces:

Configure Switch A.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/1
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 10.1.1.1 24
[SwitchA-Vlan-interface2] quit
[SwitchA] vlan 4
[SwitchA-vlan4] port gigabitethernet 1/0/2
[SwitchA-vlan4] quit
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ip address 11.1.1.1 24
[SwitchA-Vlan-interface4] quit
[SwitchA] vlan 6
[SwitchA-vlan6] port gigabitethernet 1/0/3
[SwitchA-vlan6] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] ip address 30.1.1.1 24
[SwitchA-Vlan-interface6] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.9 32
[SwitchA-LoopBack0] quit
```

Configure Switch B.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 10.1.1.2 24
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ip address 20.1.1.1 24
[SwitchB-Vlan-interface3] quit
```

```
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.9 32
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] vlan 3
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ip address 20.1.1.2 24
[SwitchC-Vlan-interface3] quit
[SwitchC] vlan 5
[SwitchC-vlan5] port gigabitethernet 1/0/2
[SwitchC-vlan5] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ip address 21.1.1.1 24
[SwitchC-Vlan-interface5] quit
[SwitchC] vlan 7
[SwitchC-vlan7] port gigabitethernet 1/0/3
[SwitchC-vlan7] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] ip address 40.1.1.2 24
[SwitchC-Vlan-interface7] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 3.3.3.9 32
[SwitchC-LoopBack0] quit
```

Configure Switch D.

```
<Sysname> system-view
[Sysname] sysname SwitchD
[SwitchD] vlan 6
[SwitchD-vlan6] port gigabitethernet 1/0/1
[SwitchD-vlan6] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] ip address 30.1.1.2 24
[SwitchD-Vlan-interface6] quit
[SwitchD] vlan 7
[SwitchD-vlan7] port gigabitethernet 1/0/2
[SwitchD-vlan7] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] ip address 40.1.1.1 24
[SwitchD-Vlan-interface7] quit
[SwitchD] interface loopback 0
[SwitchD-LoopBack0] ip address 4.4.4.9 32
[SwitchD-LoopBack0] quit
```

2. Configure MSTP:

Configure Switch A.

```
[SwitchA] stp region-configuration
```

```
[SwitchA-mst-region] region-name example
[SwitchA-mst-region] instance 2 vlan 2
[SwitchA-mst-region] instance 4 vlan 4
[SwitchA-mst-region] instance 6 vlan 6
[SwitchA-mst-region] revision-level 0
[SwitchA-mst-region] active region-configuration
[SwitchA-mst-region] quit
[SwitchA] stp global enable
```

Configure Switch B.

```
[SwitchB] stp region-configuration
[SwitchB-mst-region] region-name example
[SwitchB-mst-region] instance 2 vlan 2
[SwitchB-mst-region] instance 3 vlan 3
[SwitchB-mst-region] revision-level 0
[SwitchB-mst-region] active region-configuration
[SwitchB-mst-region] quit
[SwitchB] stp global enable
```

Configure Switch C.

```
[SwitchC] stp region-configuration
[SwitchC-mst-region] region-name example
[SwitchC-mst-region] instance 3 vlan 3
[SwitchC-mst-region] instance 5 vlan 5
[SwitchC-mst-region] instance 7 vlan 7
[SwitchC-mst-region] revision-level 0
[SwitchC-mst-region] active region-configuration
[SwitchC-mst-region] quit
[SwitchC] stp global enable
```

Configure Switch D.

```
[SwitchD] stp region-configuration
[SwitchD-mst-region] region-name example
[SwitchD-mst-region] instance 6 vlan 6
[SwitchD-mst-region] instance 7 vlan 7
[SwitchD-mst-region] revision-level 0
[SwitchD-mst-region] active region-configuration
[SwitchD-mst-region] quit
[SwitchD] stp global enable
```

3. Configure OSPF:

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 11.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
[SwitchB] ospf
```

```
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[SwitchC-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 40.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 21.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
[SwitchD] ospf
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 4.4.4.9 0.0.0.0
[SwitchD-ospf-1-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 40.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

4. Enable MPLS and IPv4 LDP:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.9
[SwitchA] mpls ldp
[SwitchA-ldp] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] mpls enable
[SwitchA-Vlan-interface2] mpls ldp enable
[SwitchA-Vlan-interface2] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] mpls enable
[SwitchA-Vlan-interface6] mpls ldp enable
[SwitchA-Vlan-interface6] quit
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.9
[SwitchB] mpls ldp
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] mpls enable
[SwitchB-Vlan-interface2] mpls ldp enable
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mpls enable
[SwitchB-Vlan-interface3] mpls ldp enable
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.9
[SwitchC] mpls ldp
[SwitchC-ldp] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] mpls enable
[SwitchC-Vlan-interface3] mpls ldp enable
[SwitchC-Vlan-interface3] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] mpls enable
[SwitchC-Vlan-interface7] mpls ldp enable
[SwitchC-Vlan-interface7] quit
```

Configure Switch D.

```
[SwitchD] mpls lsr-id 4.4.4.9
[SwitchD] mpls ldp
[SwitchD-ldp] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] mpls enable
[SwitchD-Vlan-interface6] mpls ldp enable
[SwitchD-Vlan-interface6] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] mpls enable
[SwitchD-Vlan-interface7] mpls ldp enable
[SwitchD-Vlan-interface7] quit
```

5. Configure IPv4 LSP generation policies:

On Switch A, create IP prefix list **switcha**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchA] ip prefix-list switcha index 10 permit 11.1.1.0 24
[SwitchA] ip prefix-list switcha index 20 permit 21.1.1.0 24
[SwitchA] mpls ldp
[SwitchA-ldp] lsp-trigger prefix-list switcha
[SwitchA-ldp] quit
```

On Switch B, create IP prefix list **switchb**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchB] ip prefix-list switchb index 10 permit 11.1.1.0 24
[SwitchB] ip prefix-list switchb index 20 permit 21.1.1.0 24
[SwitchB] mpls ldp
[SwitchB-ldp] lsp-trigger prefix-list switchb
[SwitchB-ldp] quit
```

On Switch C, create IP prefix list **switchc**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchC] ip prefix-list switchc index 10 permit 11.1.1.0 24
[SwitchC] ip prefix-list switchc index 20 permit 21.1.1.0 24
[SwitchC] mpls ldp
[SwitchC-ldp] lsp-trigger prefix-list switchc
[SwitchC-ldp] quit
```

On Switch D, create IP prefix list **switchd**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchD] ip prefix-list switchd index 10 permit 11.1.1.0 24
```

```
[SwitchD] ip prefix-list switchd index 20 permit 21.1.1.0 24
[SwitchD] mpls ldp
[SwitchD-ldp] lsp-trigger prefix-list switchd
[SwitchD-ldp] quit
```

6. Configure IPv4 label acceptance policies:

On Switch A, create IP prefix list **prefix-from-b** to permit subnet 21.1.1.0/24. Switch A uses this list to filter FEC-label mappings received from Switch B.

```
[SwitchA] ip prefix-list prefix-from-b index 10 permit 21.1.1.0 24
```

On Switch A, create IP prefix list **prefix-from-d** to deny subnet 21.1.1.0/24. Switch A uses this list to filter FEC-label mappings received from Switch D.

```
[SwitchA] ip prefix-list prefix-from-d index 10 deny 21.1.1.0 24
```

On Switch A, configure label acceptance policies to filter FEC-label mappings received from Switch B and Switch D.

```
[SwitchA] mpls ldp
[SwitchA-ldp] accept-label peer 2.2.2.9 prefix-list prefix-from-b
[SwitchA-ldp] accept-label peer 4.4.4.9 prefix-list prefix-from-d
[SwitchA-ldp] quit
```

On Switch C, create IP prefix list **prefix-from-b** to permit subnet 11.1.1.0/24. Switch C uses this list to filter FEC-label mappings received from Switch B.

```
[SwitchC] ip prefix-list prefix-from-b index 10 permit 11.1.1.0 24
```

On Switch C, create IP prefix list **prefix-from-d** to deny subnet 11.1.1.0/24. Switch A uses this list to filter FEC-label mappings received from Switch D.

```
[SwitchC] ip prefix-list prefix-from-d index 10 deny 11.1.1.0 24
```

On Switch C, configure label acceptance policies to filter FEC-label mappings received from Switch B and Switch D.

```
[SwitchC] mpls ldp
[SwitchC-ldp] accept-label peer 2.2.2.9 prefix-list prefix-from-b
[SwitchC-ldp] accept-label peer 4.4.4.9 prefix-list prefix-from-d
[SwitchC-ldp] quit
```

Verifying the configuration

Display LDP LSP information on the switches, for example, on Switch A. Verify that the next hop of the LSP for FEC 21.1.1.0/24 is Switch B (10.1.1.2). The LSP has been established over the link Switch A—Switch B—Switch C, not over the link Switch A—Switch D—Switch C.

```
[SwitchA] display mpls ldp lsp
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 1           Transit: 1           Egress: 1

FEC                In/Out Label        Nexthop         OutInterface/LSINDEX
11.1.1.0/24        24128/-
21.1.1.0/24        -/24127             10.1.1.2        Vlan2
                   24127/24127         10.1.1.2        Vlan2
```

Test the connectivity of the LDP LSP from Switch A to Switch C.

```
[SwitchA] ping mpls -a 11.1.1.1 ipv4 21.1.1.0 24
MPLS ping FEC 21.1.1.0/24 with 100 bytes of data:
100 bytes from 20.1.1.2: Sequence=1 time=1 ms
100 bytes from 20.1.1.2: Sequence=2 time=1 ms
100 bytes from 20.1.1.2: Sequence=3 time=8 ms
100 bytes from 20.1.1.2: Sequence=4 time=2 ms
```

```
100 bytes from 20.1.1.2: Sequence=5 time=1 ms

--- Ping statistics for FEC 21.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/2/8 ms
```

Test the connectivity of the LDP LSP from Switch C to Switch A.

```
[SwitchC] ping mpls -a 21.1.1.1 ipv4 11.1.1.0 24
MPLS ping FEC 11.1.1.0/24 with 100 bytes of data:
100 bytes from 10.1.1.1: Sequence=1 time=1 ms
100 bytes from 10.1.1.1: Sequence=2 time=1 ms
100 bytes from 10.1.1.1: Sequence=3 time=1 ms
100 bytes from 10.1.1.1: Sequence=4 time=1 ms
100 bytes from 10.1.1.1: Sequence=5 time=1 ms

--- Ping statistics for FEC 11.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/1/1 ms
```

Configuration files

- Switch A

```
#
 sysname SwitchA
#
ospf 1
 area 0.0.0.0
  network 1.1.1.9 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 11.1.1.0 0.0.0.255
  network 30.1.1.0 0.0.0.255
#
 mpls lsr-id 1.1.1.9
#
vlan 2
#
vlan 4
#
vlan 6
#
stp region-configuration
 region-name example
 instance 2 vlan 2
 instance 4 vlan 4
 instance 6 vlan 6
 active region-configuration
#
 stp global enable
#
mpls ldp
```

```
 lsp-trigger prefix-list switcha
 accept-label peer 2.2.2.9 prefix-list prefix-from-b
 accept-label peer 4.4.4.9 prefix-list prefix-from-d
#
interface LoopBack0
 ip address 1.1.1.9 255.255.255.255
#
interface Vlan-interface2
 ip address 10.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface4
 ip address 11.1.1.1 255.255.255.0
#
interface Vlan-interface6
 ip address 30.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 4
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 6
#
 ip prefix-list prefix-from-b index 10 permit 21.1.1.0 24
 ip prefix-list prefix-from-d index 10 deny 21.1.1.0 24
 ip prefix-list switcha index 10 permit 11.1.1.0 24
 ip prefix-list switcha index 20 permit 21.1.1.0 24
#
return
```

- Switch B

```
#
 sysname SwitchB
#
ospf 1
 area 0.0.0.0
  network 2.2.2.9 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 20.1.1.0 0.0.0.255
#
```

```
 mpls lsr-id 2.2.2.9
#
vlan 2 to 3
#
stp region-configuration
 region-name example
 instance 2 vlan 2
 instance 3 vlan 3
 active region-configuration
#
 stp global enable
#
mpls ldp
 lsp-trigger prefix-list switchb
#
interface LoopBack0
 ip address 2.2.2.9 255.255.255.255
#
interface Vlan-interface2
 ip address 10.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface3
 ip address 20.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
 ip prefix-list switchb index 10 permit 11.1.1.0 24
 ip prefix-list switchb index 20 permit 21.1.1.0 24
#
return
```

- Switch C

```
#
 sysname SwitchC
#
ospf 1
 area 0.0.0.0
  network 3.3.3.9 0.0.0.0
  network 20.1.1.0 0.0.0.255
```

```
  network 21.1.1.0 0.0.0.255
  network 40.1.1.0 0.0.0.255
#
 mpls lsr-id 3.3.3.9
#
vlan 3
#
vlan 5
#
vlan 7
#
stp region-configuration
 region-name example
 instance 3 vlan 3
 instance 5 vlan 5
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
 lsp-trigger prefix-list switchc
 accept-label peer 2.2.2.9 prefix-list prefix-from-b
 accept-label peer 4.4.4.9 prefix-list prefix-from-d
#
interface LoopBack0
 ip address 3.3.3.9 255.255.255.255
#
interface Vlan-interface3
 ip address 20.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface5
 ip address 21.1.1.1 255.255.255.0
#
interface Vlan-interface7
 ip address 40.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 5
```

```
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 7
#
 ip prefix-list prefix-from-b index 10 permit 11.1.1.0 24
 ip prefix-list prefix-from-d index 10 deny 11.1.1.0 24
 ip prefix-list switchc index 10 permit 11.1.1.0 24
 ip prefix-list switchc index 20 permit 21.1.1.0 24
#
return
```

- Switch D

```
#
 sysname SwitchD
#
ospf 1
 area 0.0.0.0
  network 4.4.4.9 0.0.0.0
  network 30.1.1.0 0.0.0.255
  network 40.1.1.0 0.0.0.255
#
 mpls lsr-id 4.4.4.9
#
vlan 6 to 7
#
stp region-configuration
 region-name example
 instance 6 vlan 6
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
 lsp-trigger prefix-list switchd
#
interface LoopBack0
 ip address 4.4.4.9 255.255.255.255
#
interface Vlan-interface6
 ip address 30.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface7
 ip address 40.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
```

```
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 6
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 7
#
 ip prefix-list switchd index 10 permit 11.1.1.0 24
 ip prefix-list switchd index 20 permit 21.1.1.0 24
#
return
```

Example: Configuring label advertisement control

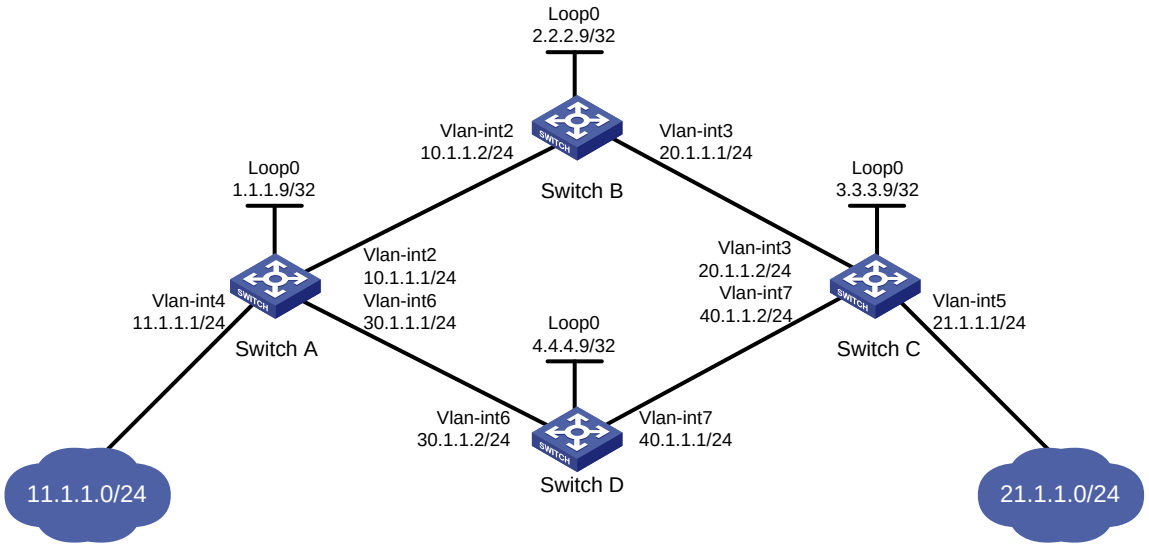
Network configuration

Two links, Switch A—Switch B—Switch C and Switch A—Switch D—Switch C, exist between subnets 11.1.1.0/24 and 21.1.1.0/24.

Configure LDP to establish LSPs only for routes to subnets 11.1.1.0/24 and 21.1.1.0/24.

Configure LDP to establish LSPs only on the link Switch A—Switch B—Switch C to forward traffic between subnets 11.1.1.0/24 and 21.1.1.0/24.

Figure 13 Network diagram



Requirements analysis

- To ensure that the LSRs establish IPv4 LSPs automatically, enable Ipv4 LDP on each LSR.
- To establish Ipv4 LDP LSPs, configure an Ipv4 routing protocol to ensure IP connectivity between the LSRs. This example uses OSPF.
- To ensure that LDP establishes Ipv4 LSPs only for the routes 11.1.1.0/24 and 21.1.1.0/24, configure Ipv4 LSP generation policies on each LSR.
- To ensure that LDP establishes Ipv4 LSPs only over the link Switch A—Switch B—Switch C, configure Ipv4 label advertisement policies as follows:

- Switch A advertises only the label mapping for FEC 11.1.1.0/24 to Switch B.
- Switch C advertises only the label mapping for FEC 21.1.1.0/24 to Switch B.
- Switch D does not advertise label mapping for FEC 21.1.1.0/24 to Switch A. Switch D does not advertise label mapping for FEC 11.1.1.0/24 to Switch C.

Procedure

1. Create VLANs and assign IP addresses to interfaces:

Configure Switch A.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/1
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 10.1.1.1 24
[SwitchA-Vlan-interface2] quit
[SwitchA] vlan 4
[SwitchA-vlan4] port gigabitethernet 1/0/2
[SwitchA-vlan4] quit
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ip address 11.1.1.1 24
[SwitchA-Vlan-interface4] quit
[SwitchA] vlan 6
[SwitchA-vlan6] port gigabitethernet 1/0/3
[SwitchA-vlan6] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] ip address 30.1.1.1 24
[SwitchA-Vlan-interface6] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.9 32
[SwitchA-LoopBack0] quit
```

Configure Switch B.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 10.1.1.2 24
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ip address 20.1.1.1 24
[SwitchB-Vlan-interface3] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.9 32
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] vlan 3
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ip address 20.1.1.2 24
[SwitchC-Vlan-interface3] quit
[SwitchC] vlan 5
[SwitchC-vlan5] port gigabitethernet 1/0/2
[SwitchC-vlan5] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ip address 21.1.1.1 24
[SwitchC-Vlan-interface5] quit
[SwitchC] vlan 7
[SwitchC-vlan7] port gigabitethernet 1/0/3
[SwitchC-vlan7] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] ip address 40.1.1.2 24
[SwitchC-Vlan-interface7] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 3.3.3.9 32
[SwitchC-LoopBack0] quit
```

Configure Switch D.

```
<Sysname> system-view
[Sysname] sysname SwitchD
[SwitchD] vlan 6
[SwitchD-vlan6] port gigabitethernet 1/0/1
[SwitchD-vlan6] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] ip address 30.1.1.2 24
[SwitchD-Vlan-interface6] quit
[SwitchD] vlan 7
[SwitchD-vlan7] port gigabitethernet 1/0/2
[SwitchD-vlan7] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] ip address 40.1.1.1 24
[SwitchD-Vlan-interface7] quit
[SwitchD] interface loopback 0
[SwitchD-LoopBack0] ip address 4.4.4.9 32
[SwitchD-LoopBack0] quit
```

2. Configure MSTP:

Configure Switch A.

```
[SwitchA] stp region-configuration
[SwitchA-mst-region] region-name example
[SwitchA-mst-region] instance 2 vlan 2
[SwitchA-mst-region] instance 4 vlan 4
```

```
[SwitchA-mst-region] instance 6 vlan 6
[SwitchA-mst-region] revision-level 0
[SwitchA-mst-region] active region-configuration
[SwitchA-mst-region] quit
[SwitchA] stp global enable
```

Configure Switch B.

```
[SwitchB] stp region-configuration
[SwitchB-mst-region] region-name example
[SwitchB-mst-region] instance 2 vlan 2
[SwitchB-mst-region] instance 3 vlan 3
[SwitchB-mst-region] revision-level 0
[SwitchB-mst-region] active region-configuration
[SwitchB-mst-region] quit
[SwitchB] stp global enable
```

Configure Switch C.

```
[SwitchC] stp region-configuration
[SwitchC-mst-region] region-name example
[SwitchC-mst-region] instance 3 vlan 3
[SwitchC-mst-region] instance 5 vlan 5
[SwitchC-mst-region] instance 7 vlan 7
[SwitchC-mst-region] revision-level 0
[SwitchC-mst-region] active region-configuration
[SwitchC-mst-region] quit
[SwitchC] stp global enable
```

Configure Switch D.

```
[SwitchD] stp region-configuration
[SwitchD-mst-region] region-name example
[SwitchD-mst-region] instance 6 vlan 6
[SwitchD-mst-region] instance 7 vlan 7
[SwitchD-mst-region] revision-level 0
[SwitchD-mst-region] active region-configuration
[SwitchD-mst-region] quit
[SwitchD] stp global enable
```

3. Configure OSPF:

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 11.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
```

```
[SwitchB-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[SwitchC-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 40.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 21.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
[SwitchD] ospf
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 4.4.4.9 0.0.0.0
[SwitchD-ospf-1-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 40.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

4. Enable MPLS and Ipv4 LDP:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.9
[SwitchA] mpls ldp
[SwitchA-ldp] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] mpls enable
[SwitchA-Vlan-interface2] mpls ldp enable
[SwitchA-Vlan-interface2] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] mpls enable
[SwitchA-Vlan-interface6] mpls ldp enable
[SwitchA-Vlan-interface6] quit
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.9
[SwitchB] mpls ldp
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] mpls enable
[SwitchB-Vlan-interface2] mpls ldp enable
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mpls enable
[SwitchB-Vlan-interface3] mpls ldp enable
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.9
[SwitchC] mpls ldp
```

```
[SwitchC-ldp] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] mpls enable
[SwitchC-Vlan-interface3] mpls ldp enable
[SwitchC-Vlan-interface3] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] mpls enable
[SwitchC-Vlan-interface7] mpls ldp enable
[SwitchC-Vlan-interface7] quit
```

Configure Switch D.

```
[SwitchD] mpls lsr-id 4.4.4.9
[SwitchD] mpls ldp
[SwitchD-ldp] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] mpls enable
[SwitchD-Vlan-interface6] mpls ldp enable
[SwitchD-Vlan-interface6] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] mpls enable
[SwitchD-Vlan-interface7] mpls ldp enable
[SwitchD-Vlan-interface7] quit
```

5. Configure Ipv4 LSP generation policies:

On Switch A, create IP prefix list **switcha**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchA] ip prefix-list switcha index 10 permit 11.1.1.0 24
[SwitchA] ip prefix-list switcha index 20 permit 21.1.1.0 24
[SwitchA] mpls ldp
[SwitchA-ldp] lsp-trigger prefix-list switcha
[SwitchA-ldp] quit
```

On Switch B, create IP prefix list **switchb**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchB] ip prefix-list switchb index 10 permit 11.1.1.0 24
[SwitchB] ip prefix-list switchb index 20 permit 21.1.1.0 24
[SwitchB] mpls ldp
[SwitchB-ldp] lsp-trigger prefix-list switchb
[SwitchB-ldp] quit
```

On Switch C, create IP prefix list **switchc**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchC] ip prefix-list switchc index 10 permit 11.1.1.0 24
[SwitchC] ip prefix-list switchc index 20 permit 21.1.1.0 24
[SwitchC] mpls ldp
[SwitchC-ldp] lsp-trigger prefix-list switchc
[SwitchC-ldp] quit
```

On Switch D, create IP prefix list **switchd**, and configure LDP to use only the routes permitted by the prefix list to establish LSPs.

```
[SwitchD] ip prefix-list switchd index 10 permit 11.1.1.0 24
[SwitchD] ip prefix-list switchd index 20 permit 21.1.1.0 24
[SwitchD] mpls ldp
[SwitchD-ldp] lsp-trigger prefix-list switchd
```

```
[SwitchD-ldp] quit
```

6. Configure Ipv4 label advertisement policies:

On Switch A, create IP prefix list **prefix-to-b** to permit subnet 11.1.1.0/24. Switch A uses this list to filter FEC-label mappings advertised to Switch B.

```
[SwitchA] ip prefix-list prefix-to-b index 10 permit 11.1.1.0 24
```

On Switch A, create IP prefix list **peer-b** to permit 2.2.2.9/32. Switch A uses this list to filter peers.

```
[SwitchA] ip prefix-list peer-b index 10 permit 2.2.2.9 32
```

On Switch A, configure a label advertisement policy to advertise only the label mapping for FEC 11.1.1.0/24 to Switch B.

```
[SwitchA] mpls ldp
[SwitchA-ldp] advertise-label prefix-list prefix-to-b peer peer-b
[SwitchA-ldp] quit
```

On Switch C, create IP prefix list **prefix-to-b** to permit subnet 21.1.1.0/24. Switch C uses this list to filter FEC-label mappings advertised to Switch B.

```
[SwitchC] ip prefix-list prefix-to-b index 10 permit 21.1.1.0 24
```

On Switch C, create IP prefix list **peer-b** to permit 2.2.2.9/32. Switch C uses this list to filter peers.

```
[SwitchC] ip prefix-list peer-b index 10 permit 2.2.2.9 32
```

On Switch C, configure a label advertisement policy to advertise only the label mapping for FEC 21.1.1.0/24 to Switch B.

```
[SwitchC] mpls ldp
[SwitchC-ldp] advertise-label prefix-list prefix-to-b peer peer-b
[SwitchC-ldp] quit
```

On Switch D, create IP prefix list **prefix-to-a** to deny subnet 21.1.1.0/24. Switch D uses this list to filter FEC-label mappings to be advertised to Switch A.

```
[SwitchD] ip prefix-list prefix-to-a index 10 deny 21.1.1.0 24
[SwitchD] ip prefix-list prefix-to-a index 20 permit 0.0.0.0 0 less-equal 32
```

On Switch D, create IP prefix list **peer-a** to permit 1.1.1.9/32. Switch D uses this list to filter peers.

```
[SwitchD] ip prefix-list peer-a index 10 permit 1.1.1.9 32
```

On Switch D, create IP prefix list **prefix-to-c** to deny subnet 11.1.1.0/24. Switch D uses this list to filter FEC-label mappings to be advertised to Switch C.

```
[SwitchD] ip prefix-list prefix-to-c index 10 deny 11.1.1.0 24
[SwitchD] ip prefix-list prefix-to-c index 20 permit 0.0.0.0 0 less-equal 32
```

On Switch D, create IP prefix list **peer-c** to permit subnet 3.3.3.9/32. Switch D uses this list to filter peers.

```
[SwitchD] ip prefix-list peer-c index 10 permit 3.3.3.9 32
```

On Switch D, configure a label advertisement policy. This policy ensures that Switch D does not advertise label mappings for FEC 21.1.1.0/24 to Switch A, and does not advertise label mappings for FEC 11.1.1.0/24 to Switch C.

```
[SwitchD] mpls ldp
[SwitchD-ldp] advertise-label prefix-list prefix-to-a peer peer-a
[SwitchD-ldp] advertise-label prefix-list prefix-to-c peer peer-c
[SwitchD-ldp] quit
```

Verifying the configuration

Display LDP LSP information on each switch. Verify that Switch A and Switch C have received FEC-label mappings only from Switch B. Switch B has received FEC-label mappings from both

Switch A and Switch C. Switch D does not receive FEC-label mappings from Switch A or Switch C. LDP has established an LSP only over the link Switch A—Switch B—Switch C.

```
[SwitchA] display mpls ldp lsp
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 1           Transit: 1           Egress: 1

FEC               In/Out Label        Nexthop         OutInterface/LSINDEX
11.1.1.0/24       24128/-
                  -/24128(L)
                  -/24128(L)
21.1.1.0/24       -/24127             10.1.1.2        Vlan2
                  24127/24127         10.1.1.2        Vlan2
[SwitchB] display mpls ldp lsp
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 2           Transit: 2           Egress: 0
FEC               In/Out Label        Nexthop         OutInterface/LSINDEX
11.1.1.0/24       -/24128             10.1.1.1        Vlan2
                  24128/24128         10.1.1.1        Vlan2
21.1.1.0/24       -/24128             20.1.1.2        Vlan3
                  24127/24128         20.1.1.2        Vlan3
[SwitchC] display mpls ldp lsp
Status Flags: * - stale, L - liberal, B - backup, N/A – unavailable
FECs: 2           Ingress: 1          Transit: 1      Egress: 1
FEC               In/Out Label        Nexthop         OutInterface/LSINDEX
11.1.1.0/24       -/1277              20.1.1.1        Vlan3
                  1148/1277           20.1.1.1        Vlan3
21.1.1.0/24       1149/-
                  -/1276(L)
                  -/1150(L)
[SwitchD] display mpls ldp lsp
Status Flags: * - stale, L - liberal, B - backup, N/A – unavailable
FECs: 2           Ingress: 0          Transit: 0      Egress: 2
FEC               In/Out Label        Nexthop         OutInterface/LSINDEX
11.1.1.0/24       1151/-
                  -/1277(L)
21.1.1.0/24       1150/-
```

Test the connectivity of the LDP LSP from Switch A to Switch C.

```
[SwitchA] ping mpls -a 11.1.1.1 ipv4 21.1.1.0 24
MPLS ping FEC 21.1.1.0/24 with 100 bytes of data:
100 bytes from 20.1.1.2: Sequence=1 time=1 ms
100 bytes from 20.1.1.2: Sequence=2 time=1 ms
100 bytes from 20.1.1.2: Sequence=3 time=8 ms
100 bytes from 20.1.1.2: Sequence=4 time=2 ms
100 bytes from 20.1.1.2: Sequence=5 time=1 ms

--- Ping statistics for FEC 21.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/2/8 ms
```

Test the connectivity of the LDP LSP from Switch C to Switch A.

```
[SwitchC] ping mpls -a 21.1.1.1 ipv4 11.1.1.0 24
MPLS ping FEC 11.1.1.0/24 with 100 bytes of data:
100 bytes from 10.1.1.1: Sequence=1 time=1 ms
100 bytes from 10.1.1.1: Sequence=2 time=1 ms
100 bytes from 10.1.1.1: Sequence=3 time=1 ms
100 bytes from 10.1.1.1: Sequence=4 time=1 ms
100 bytes from 10.1.1.1: Sequence=5 time=1 ms

--- Ping statistics for FEC 11.1.1.0/24 ---
5 packets transmitted, 5 packets received, 0.0% packet loss
Round-trip min/avg/max = 1/1/1 ms
```

Configuration files

- Switch A

```
#
 sysname SwitchA
#
ospf 1
 area 0.0.0.0
  network 1.1.1.9 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 11.1.1.0 0.0.0.255
  network 30.1.1.0 0.0.0.255
#
 mpls lsr-id 1.1.1.9
#
vlan 2
#
vlan 4
#
vlan 6
#
stp region-configuration
 region-name example
 instance 2 vlan 2
 instance 4 vlan 4
 instance 6 vlan 6
 active region-configuration
#
 stp global enable
#
mpls ldp
 lsp-trigger prefix-list switcha
 advertise-label prefix-list prefix-to-b peer peer-b
#
interface LoopBack0
 ip address 1.1.1.9 255.255.255.255
#
```

```
interface Vlan-interface2
 ip address 10.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface4
 ip address 11.1.1.1 255.255.255.0
#
interface Vlan-interface6
 ip address 30.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 4
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 6
#
 ip prefix-list peer-b index 10 permit 2.2.2.9 32
 ip prefix-list prefix-to-b index 10 permit 11.1.1.0 24
 ip prefix-list switcha index 10 permit 11.1.1.0 24
 ip prefix-list switcha index 20 permit 21.1.1.0 24
#
return
```

- Switch B

```
#
 sysname SwitchB
#
ospf 1
 area 0.0.0.0
  network 2.2.2.9 0.0.0.0
  network 10.1.1.0 0.0.0.255
  network 20.1.1.0 0.0.0.255
#
 mpls lsr-id 2.2.2.9
#
vlan 2 to 3
#
stp region-configuration
 region-name example
 instance 2 vlan 2
```

```
 instance 3 vlan 3
 active region-configuration
#
 stp global enable
#
mpls ldp
 lsp-trigger prefix-list switchb
#
interface LoopBack0
 ip address 2.2.2.9 255.255.255.255
#
interface Vlan-interface2
 ip address 10.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface3
 ip address 20.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
 ip prefix-list switchb index 10 permit 11.1.1.0 24
 ip prefix-list switchb index 20 permit 21.1.1.0 24
#
return
```

- Switch C

```
#
 sysname SwitchC
#
ospf 1
 area 0.0.0.0
  network 3.3.3.9 0.0.0.0
  network 20.1.1.0 0.0.0.255
  network 21.1.1.0 0.0.0.255
  network 40.1.1.0 0.0.0.255
#
 mpls lsr-id 3.3.3.9
#
vlan 3
#
```

```
vlan 5
#
vlan 7
#
stp region-configuration
 region-name example
 instance 3 vlan 3
 instance 5 vlan 5
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
 lsp-trigger prefix-list switchc
 advertise-label prefix-list prefix-to-b peer peer-b
#
interface LoopBack0
 ip address 3.3.3.9 255.255.255.255
#
interface Vlan-interface3
 ip address 20.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface5
 ip address 21.1.1.1 255.255.255.0
#
interface Vlan-interface7
 ip address 40.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 5
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 7
#
 ip prefix-list peer-b index 10 permit 2.2.2.9 32
 ip prefix-list prefix-to-b index 10 permit 21.1.1.0 24
 ip prefix-list switchc index 10 permit 11.1.1.0 24
```

```
 ip prefix-list switchc index 20 permit 21.1.1.0 24
#
return
```

- Switch D

```
#
 sysname SwitchD
#
ospf 1
 area 0.0.0.0
  network 4.4.4.9 0.0.0.0
  network 30.1.1.0 0.0.0.255
  network 40.1.1.0 0.0.0.255
#
 mpls lsr-id 4.4.4.9
#
vlan 6 to 7
#
stp region-configuration
 region-name example
 instance 6 vlan 6
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
 lsp-trigger prefix-list switchd
 advertise-label prefix-list prefix-to-a peer peer-a
 advertise-label prefix-list prefix-to-c peer peer-c
#
interface LoopBack0
 ip address 4.4.4.9 255.255.255.255
#
interface Vlan-interface6
 ip address 30.1.1.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface7
 ip address 40.1.1.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 6
#
interface GigabitEthernet1/0/2
```

```
 port link-mode bridge
 port access vlan 7
#
 ip prefix-list peer-a index 10 permit 1.1.1.9 32
 ip prefix-list peer-c index 10 permit 3.3.3.9 32
 ip prefix-list prefix-to-a index 10 deny 21.1.1.0 24
 ip prefix-list prefix-to-a index 20 permit 0.0.0.0 0 less-equal 32
 ip prefix-list prefix-to-c index 10 deny 11.1.1.0 24
 ip prefix-list prefix-to-c index 20 permit 0.0.0.0 0 less-equal 32
 ip prefix-list switchd index 10 permit 11.1.1.0 24
 ip prefix-list switchd index 20 permit 21.1.1.0 24
#
return
```

Example: Configuring LDP FRR

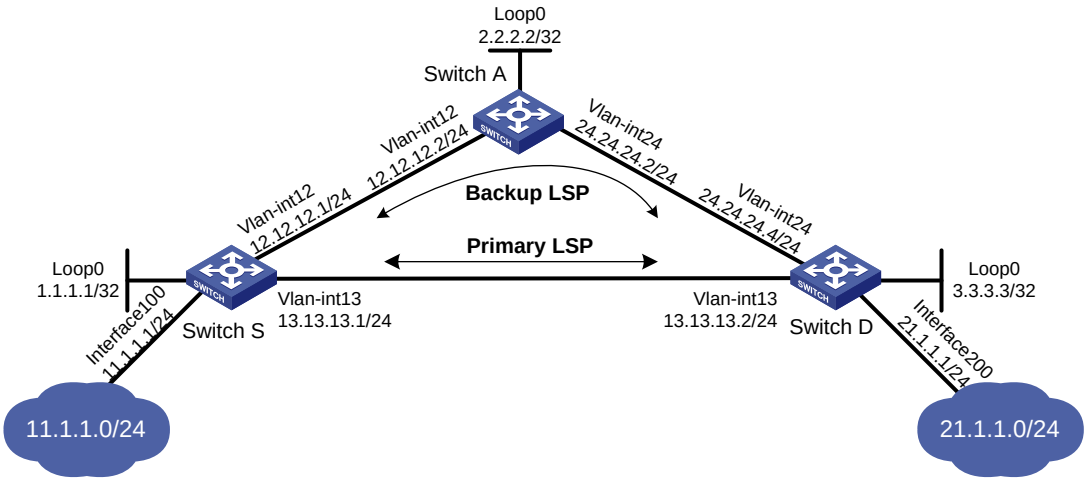
Network configuration

Switch S, Switch A, and Switch D reside in the same OSPF domain. Configure OSPF FRR so LDP can establish a primary LSP and a backup LSP on the Switch S—Switch D and the Switch S—Switch A—Switch D links, respectively.

When the primary LSP operates correctly, traffic between subnets 11.1.1.0/24 and 21.1.1.0/24 is forwarded through the LSP.

When the primary LSP fails, traffic between the two subnets is immediately switched to the backup LSP.

Figure 14 Network diagram



Requirements analysis

- To ensure that the LSRs establish IPv4 LSPs automatically, enable IPv4 LDP on each LSR.
- To establish IPv4 LDP LSPs, configure an IPv4 routing protocol to ensure IP connectivity between the LSRs. This example uses OSPF.
- To ensure that LDP establishes IPv4 LSPs only for the routes 11.1.1.0/24 and 21.1.1.0/24, configure IPv4 LSP generation policies on each LSR.
- To allow LDP to establish backup LSRs, configure OSPF FRR on Switch S and Switch D.

Prerequisites

Before configuration, disable the spanning tree feature globally or map each VLAN to an MSTI. For more information, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Create VLANs and assign IP addresses to interfaces:

Configure Switch A.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 12
[SwitchA-vlan12] port gigabitethernet 1/0/1
[SwitchA-vlan12] quit
[SwitchA] interface vlan-interface 12
[SwitchA-Vlan-interface12] ip address 12.12.12.2 24
[SwitchA-Vlan-interface12] quit
[SwitchA] vlan 24
[SwitchA-vlan24] port gigabitethernet 1/0/2
[SwitchA-vlan24] quit
[SwitchA] interface vlan-interface 24
[SwitchA-Vlan-interface24] ip address 24.24.24.2 24
[SwitchA-Vlan-interface24] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 2.2.2.2 32
[SwitchA-LoopBack0] quit
```

Configure Switch S.

```
<Sysname> system-view
[Sysname] sysname SwitchS
[SwitchS] vlan 12
[SwitchS-vlan12] port gigabitethernet 1/0/1
[SwitchS-vlan12] quit
[SwitchS] interface vlan-interface 12
[SwitchS-Vlan-interface12] ip address 12.12.12.1 24
[SwitchS-Vlan-interface12] quit
[SwitchS] vlan 13
[SwitchS-vlan13] port gigabitethernet 1/0/2
[SwitchS-vlan13] quit
[SwitchS] interface vlan-interface 13
[SwitchS-Vlan-interface13] ip address 13.13.13.1 24
[SwitchS-Vlan-interface13] quit
[SwitchS] vlan 100
[SwitchS-vlan100] port gigabitethernet 1/0/3
[SwitchS-vlan100] quit
[SwitchS] interface vlan-interface 100
[SwitchS-Vlan-interface100] ip address 11.1.1.1 24
[SwitchS-Vlan-interface100] quit
[SwitchS] interface loopback 0
[SwitchS-LoopBack0] ip address 1.1.1.1 32
[SwitchS-LoopBack0] quit
```

Configure Switch D.

```
<Sysname> system-view
[Sysname] sysname SwitchD
[SwitchD] vlan 13
[SwitchD-vlan13] port gigabitethernet 1/0/1
[SwitchD-vlan13] quit
[SwitchD] interface vlan-interface 13
[SwitchD-Vlan-interface13] ip address 13.13.13.2 24
[SwitchD-Vlan-interface13] quit
[SwitchD] vlan 24
[SwitchD-vlan24] port gigabitethernet 1/0/2
[SwitchD-vlan24] quit
[SwitchD] interface vlan-interface 24
[SwitchD-Vlan-interface24] ip address 24.24.24.4 24
[SwitchD-Vlan-interface24] quit
[SwitchD] vlan 200
[SwitchD-vlan200] port gigabitethernet 1/0/3
[SwitchD-vlan200] quit
[SwitchD] interface vlan-interface 200
[SwitchD-Vlan-interface200] ip address 21.1.1.1 24
[SwitchD-Vlan-interface200] quit
[SwitchD] interface loopback 0
[SwitchD-LoopBack0] ip address 3.3.3.3 32
[SwitchD-LoopBack0] quit
```

2. Configure OSPF:

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0
[SwitchA-ospf-1-area-0.0.0.0] network 12.12.12.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 24.24.24.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch S.

```
[SwitchS] ospf
[SwitchS-ospf-1] area 0
[SwitchS-ospf-1-area-0.0.0.0] network 1.1.1.1 0.0.0.0
[SwitchS-ospf-1-area-0.0.0.0] network 12.12.12.0 0.0.0.255
[SwitchS-ospf-1-area-0.0.0.0] network 13.13.13.0 0.0.0.255
[SwitchS-ospf-1-area-0.0.0.0] network 11.1.1.0 0.0.0.255
[SwitchS-ospf-1-area-0.0.0.0] quit
[SwitchS-ospf-1] quit
```

Configure Switch D.

```
[SwitchD] ospf
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 3.3.3.3 0.0.0.0
[SwitchD-ospf-1-area-0.0.0.0] network 13.13.13.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 24.24.24.0 0.0.0.255
```

```
[SwitchC-ospf-1-area-0.0.0.0] network 21.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

3. Configure OSPF FRR by using one of the following methods:

- (Method 1.) Enable OSPF FRR to calculate a backup next hop by using the LFA algorithm:

  # Configure Switch S.

  ```
  [SwitchS] bfd echo-source-ip 10.10.10.10
  [SwitchS] ospf 1
  [SwitchS-ospf-1] fast-reroute lfa
  [SwitchS-ospf-1] quit
  ```

  # Configure Switch D.

  ```
  [SwitchD] bfd echo-source-ip 11.11.11.11
  [SwitchD] ospf 1
  [SwitchD-ospf-1] fast-reroute lfa
  [SwitchD-ospf-1] quit
  ```

- (Method 2.) Enable OSPF FRR to specify a backup next hop by using a routing policy:

  # Configure Switch S.

  ```
  [SwitchS] bfd echo-source-ip 10.10.10.10
  [SwitchS] ip prefix-list abc index 10 permit 21.1.1.0 24
  [SwitchS] route-policy frr permit node 10
  [SwitchS-route-policy-frr-10] if-match ip address prefix-list abc
  [SwitchS-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
  12 backup-nexthop 12.12.12.2
  [SwitchS-route-policy-frr-10] quit
  [SwitchS] ospf 1
  [SwitchS-ospf-1] fast-reroute route-policy frr
  [SwitchS-ospf-1] quit
  ```

  # Configure Switch D.

  ```
  [SwitchD] bfd echo-source-ip 10.10.10.10
  [SwitchD] ip prefix-list abc index 10 permit 11.1.1.0 24
  [SwitchD] route-policy frr permit node 10
  [SwitchD-route-policy-frr-10] if-match ip address prefix-list abc
  [SwitchD-route-policy-frr-10] apply fast-reroute backup-interface vlan-interface
  24 backup-nexthop 24.24.24.2
  [SwitchD-route-policy-frr-10] quit
  [SwitchD] ospf 1
  [SwitchD-ospf-1] fast-reroute route-policy frr
  [SwitchD-ospf-1] quit
  ```

4. Enable MPLS and IPv4 LDP:

Configure Switch S.

```
[SwitchS] mpls lsr-id 1.1.1.1
[SwitchS] mpls ldp
[SwitchS-mpls-ldp] quit
[SwitchS] interface vlan-interface 12
[SwitchS-Vlan-interface12] mpls enable
[SwitchS-Vlan-interface12] mpls ldp enable
[SwitchS-Vlan-interface12] quit
[SwitchS] interface vlan-interface 13
```

```
[SwitchS-Vlan-interface13] mpls enable
[SwitchS-Vlan-interface13] mpls ldp enable
[SwitchS-Vlan-interface13] quit
```

Configure Switch D.

```
[SwitchD] mpls lsr-id 3.3.3.3
[SwitchD] mpls ldp
[SwitchD-mpls-ldp] quit
[SwitchD] interface vlan-interface 13
[SwitchD-Vlan-interface13] mpls enable
[SwitchD-Vlan-interface13] mpls ldp enable
[SwitchD-Vlan-interface13] quit
[SwitchD] interface vlan-interface 24
[SwitchD-Vlan-interface24] mpls enable
[SwitchD-Vlan-interface24] mpls ldp enable
[SwitchD-Vlan-interface24] quit
```

Configure Switch A.

```
[SwitchA] mpls lsr-id 2.2.2.2
[SwitchA] mpls ldp
[SwitchA-mpls-ldp] quit
[SwitchA] interface vlan-interface 12
[SwitchA-Vlan-interface12] mpls enable
[SwitchA-Vlan-interface12] mpls ldp enable
[SwitchA-Vlan-interface12] quit
[SwitchA] interface vlan-interface 24
[SwitchA-Vlan-interface24] mpls enable
[SwitchA-Vlan-interface24] mpls ldp enable
[SwitchA-Vlan-interface24] quit
```

5. Configure IPv4 LSP generation policies so LDP uses all static routes and IGP routes to establish LSPs:

Configure Switch S.

```
[SwitchS] mpls ldp
[SwitchS-ldp] lsp-trigger all
[SwitchS-ldp] quit
```

Configure Switch D.

```
[SwitchD] mpls ldp
[SwitchD-ldp] lsp-trigger all
[SwitchD-ldp] quit
```

Configure Switch A.

```
[SwitchA] mpls ldp
[SwitchA-ldp] lsp-trigger all
[SwitchA-ldp] quit
```

Verifying the configuration

Execute the **display mpls ldp lsp** command on Switch S and Switch D to verify that primary and backup LSPs have been established between them. An LSP is backup if a letter B in brackets is displayed next to its outgoing label. Take Switch S for example.

```
[SwitchS] display mpls ldp lsp 21.1.1.0 24
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 1           Ingress: 2           Transit: 2           Egress: 0
```

```
FEC                 In/Out Label        Nexthop         OutInterface/LSINDEX
21.1.1.0/24         -/24126             13.13.13.2      Vlan13
                    24122/24126         13.13.13.2      Vlan13
                    -/24122(B)          12.12.12.2      Vlan12
                    24122/24122(B)      12.12.12.2      Vlan12
```

Configuration files (OSPF FRR method 1)

- Switch A

```
#
 sysname SwitchA
#
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 12.12.12.0 0.0.0.255
  network 24.24.24.0 0.0.0.255
#
 mpls lsr-id 2.2.2.2
#
vlan 12
#
vlan 24
#
mpls ldp
 lsp-trigger all
#
interface LoopBack0
 ip address 2.2.2.2 255.255.255.255
#
interface Vlan-interface12
 ip address 12.12.12.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface24
 ip address 24.24.24.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 12
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 24
#
return
```

- Switch S

```
#
 sysname SwitchS
#
 bfd echo-source-ip 10.10.10.10
#
ospf 1
 fast-reroute lfa
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 11.1.1.0 0.0.0.255
  network 12.12.12.0 0.0.0.255
  network 13.13.13.0 0.0.0.255
#
 mpls lsr-id 1.1.1.1
#
vlan 12 to 13
#
vlan 100
#
mpls ldp
 lsp-trigger all
#
interface LoopBack0
 ip address 1.1.1.1 255.255.255.255
#
interface Vlan-interface12
 ip address 12.12.12.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface13
 ip address 13.13.13.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface100
 ip address 11.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 12
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 13
#
interface GigabitEthernet1/0/3
```

```
 port link-mode bridge
 port access vlan 100
#
return
```

- Switch D

```
#
 sysname SwitchD
#
 bfd echo-source-ip 11.11.11.11
#
ospf 1
 fast-reroute lfa
 area 0.0.0.0
  network 3.3.3.3 0.0.0.0
  network 13.13.13.0 0.0.0.255
  network 21.1.1.0 0.0.0.255
  network 24.24.24.0 0.0.0.255
#
 mpls lsr-id 3.3.3.3
#
vlan 13
#
vlan 24
#
vlan 200
#
mpls ldp
 lsp-trigger all
#
interface LoopBack0
 ip address 3.3.3.3 255.255.255.255
#
interface Vlan-interface13
 ip address 13.13.13.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface24
 ip address 24.24.24.4 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface200
 ip address 21.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 13
```

```
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 24
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 200
#
Return
```

Configuration files (OSPF FRR method 2)

- Switch A

```
#
 sysname SwitchA
#
ospf 1
 area 0.0.0.0
  network 2.2.2.2 0.0.0.0
  network 12.12.12.0 0.0.0.255
  network 24.24.24.0 0.0.0.255
#
 mpls lsr-id 2.2.2.2
#
vlan 12
#
vlan 24
#
mpls ldp
 lsp-trigger all
#
interface LoopBack0
 ip address 2.2.2.2 255.255.255.255
#
interface Vlan-interface12
 ip address 12.12.12.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface24
 ip address 24.24.24.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 12
#
interface GigabitEthernet1/0/2
```

```
 port link-mode bridge
 port access vlan 24
#
return
```

- Switch S

```
#
 sysname SwitchS
#
 bfd echo-source-ip 10.10.10.10
#
ospf 1
 fast-reroute route-policy frr
 area 0.0.0.0
  network 1.1.1.1 0.0.0.0
  network 11.1.1.0 0.0.0.255
  network 12.12.12.0 0.0.0.255
  network 13.13.13.0 0.0.0.255
#
 mpls lsr-id 1.1.1.1
#
vlan 12 to 13
#
vlan 100
#
mpls ldp
 lsp-trigger all
#
interface LoopBack0
 ip address 1.1.1.1 255.255.255.255
#
interface Vlan-interface12
 ip address 12.12.12.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface13
 ip address 13.13.13.1 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface100
 ip address 11.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 12
#
interface GigabitEthernet1/0/2
```

```
 port link-mode bridge
 port access vlan 13
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 100
#
route-policy frr permit node 10
 if-match ip address prefix-list abc
 apply fast-reroute backup-interface Vlan-interface12 backup-nexthop 12.12.12.2
#
 ip prefix-list abc index 10 permit 21.1.1.0 24
#
return
```

- Switch D

```
#
 sysname SwitchD
#
 bfd echo-source-ip 10.10.10.10
#
ospf 1
 fast-reroute route-policy frr
 area 0.0.0.0
  network 3.3.3.3 0.0.0.0
  network 13.13.13.0 0.0.0.255
  network 21.1.1.0 0.0.0.255
  network 24.24.24.0 0.0.0.255
#
 mpls lsr-id 3.3.3.3
#
vlan 13
#
vlan 24
#
vlan 200
#
mpls ldp
 lsp-trigger all
#
interface LoopBack0
 ip address 3.3.3.3 255.255.255.255
#
interface Vlan-interface13
 ip address 13.13.13.2 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface24
```

```
 ip address 24.24.24.4 255.255.255.0
 mpls enable
 mpls ldp enable
#
interface Vlan-interface200
 ip address 21.1.1.1 255.255.255.0
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 13
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 24
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 200
#
route-policy frr permit node 10
 if-match ip address prefix-list abc
 apply fast-reroute backup-interface Vlan-interface24 backup-nexthop 24.24.24.2
#
 ip prefix-list abc index 10 permit 11.1.1.0 24
#
return
```

IPv6 LDP configuration examples

Example: Configuring IPv6 LDP LSP

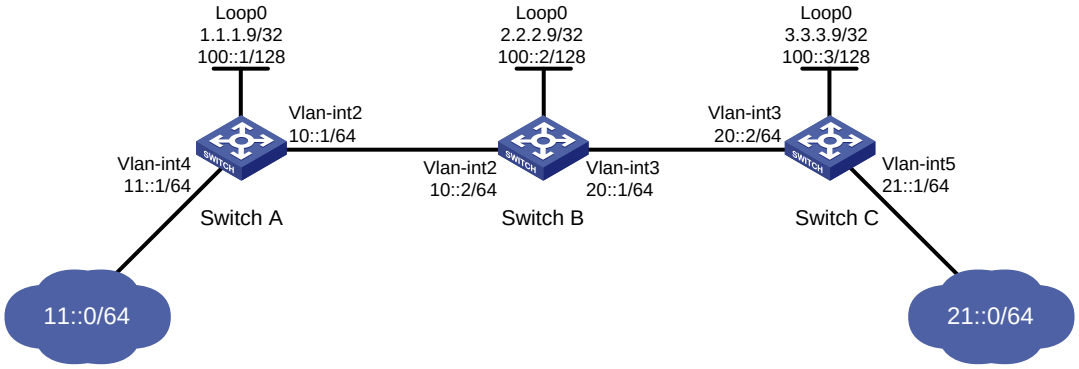
Network configuration

Switch A, Switch B, and Switch C all support MPLS.

Configure LDP to establish IPv6 LSPs between Switch A and Switch C, so subnets 11::0/64 and 21::0/64 can reach each other over MPLS.

Configure LDP to establish IPv6 LSPs only for destinations 100::1/128, 100::2/128, 100::3/128, 11::0/64, and 21::0/64 on Switch A, Switch B, and Switch C.

Figure 15 Network diagram



Requirements analysis

- To ensure that the LSRs establish IPv6 LSPs automatically, enable IPv6 LDP on each LSR.
- To establish IPv6 LDP LSPs, configure an IPv6 routing protocol to ensure IP connectivity between the LSRs. This example uses OSPFv3.
- To control the number of IPv6 LSPs, configure an IPv6 LSP generation policy on each LSR.

Procedure

1. Create VLANs and assign IP addresses to interfaces:

   # Configure Switch A.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/1
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address 10::1 64
[SwitchA-Vlan-interface2] quit
[SwitchA] vlan 4
[SwitchA-vlan4] port gigabitethernet 1/0/2
[SwitchA-vlan4] quit
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ipv6 address 11::1 64
[SwitchA-Vlan-interface4] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.9 32
[SwitchA-LoopBack0] ipv6 address 100::1 128
[SwitchA-LoopBack0] quit
```

   # Configure Switch B.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address 10::2 64
[SwitchB-Vlan-interface2] quit
```

```
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ipv6 address 20::1 64
[SwitchB-Vlan-interface3] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.9 32
[SwitchB-LoopBack0] ipv6 address 100::2 128
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] vlan 3
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ipv6 address 20::2 64
[SwitchC-Vlan-interface3] quit
[SwitchC] vlan 5
[SwitchC-vlan5] port gigabitethernet 1/0/2
[SwitchC-vlan5] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ipv6 address 21::1 64
[SwitchC-Vlan-interface5] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 3.3.3.9 32
[SwitchC-LoopBack0] ipv6 address 100::3 128
[SwitchC-LoopBack0] quit
```

2. Configure OSPFv3 on each switch to ensure IP connectivity between them:

Configure Switch A.

```
[SwitchA] ospfv3
[SwitchA-ospfv3-1] router-id 1.1.1.9
[SwitchA-ospfv3-1] area 0
[SwitchA-ospfv3-1-area-0.0.0.0] quit
[SwitchA-ospfv3-1] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchA-LoopBack0] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface2] quit
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface4] quit
```

Configure Switch B.

```
[SwitchB] ospfv3
[SwitchB-ospfv3-1] router-id 2.2.2.9
```

```
[SwitchB-ospfv3-1] area 0
[SwitchB-ospfv3-1-area-0.0.0.0] quit
[SwitchB-ospfv3-1] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchB-LoopBack0] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ospfv3 1 area 0.0.0.0
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ospfv3 1 area 0.0.0.0
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.9
[SwitchC-ospfv3-1] area 0
[SwitchC-ospfv3-1-area-0.0.0.0] quit
[SwitchC-ospfv3-1] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchC-LoopBack0] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface3] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface5] quit
```

Verify that the switches have learned the routes to each other. This example uses Switch A.

```
[SwitchA] display ipv6 routing-table

Destinations : 13       Routes : 13

Destination: ::1/128                                    Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: ::FFFF:127.0.0.0/104                       Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 10::/64                                    Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan2                                      Cost      : 0

Destination: 10::1/128                                  Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0
```

```
Destination: 11::/64                                    Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : Vlan4                                      Cost      : 0

Destination: 11::1/128                                  Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 20::/64                                    Protocol  : O_INTRA
NextHop    : FE80::7AAF:ABFF:FE27:302                   Preference: 10
Interface  : Vlan2                                      Cost      : 2

Destination: 21::/64                                    Protocol  : O_INTRA
NextHop    : FE80::7AAF:ABFF:FE27:302                   Preference: 10
Interface  : Vlan2                                      Cost      : 3

Destination: 100::1/128                                 Protocol  : Direct
NextHop    : ::1                                        Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: 100::2/128                                 Protocol  : O_INTRA
NextHop    : FE80::20C:29FF:FE9D:EAC0                   Preference: 10
Interface  : Vlan2                                      Cost      : 1

Destination: 100::3/128                                 Protocol  : O_INTRA
NextHop    : FE80::20C:29FF:FE9D:EAC0                   Preference: 10
Interface  : Vlan2                                      Cost      : 2

Destination: FE80::/10                                  Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : InLoop0                                    Cost      : 0

Destination: FF00::/8                                   Protocol  : Direct
NextHop    : ::                                         Preference: 0
Interface  : NULL0                                      Cost      : 0
```

3. Enable MPLS and IPv6 LDP:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.9
[SwitchA] mpls ldp
[SwitchA-ldp] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] mpls enable
[SwitchA-Vlan-interface2] mpls ldp ipv6 enable
[SwitchA-Vlan-interface2] mpls ldp transport-address 10::1
[SwitchA-Vlan-interface2] quit
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.9
[SwitchB] mpls ldp
```

```
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] mpls enable
[SwitchB-Vlan-interface2] mpls ldp ipv6 enable
[SwitchB-Vlan-interface2] mpls ldp transport-address 10::2
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mpls enable
[SwitchB-Vlan-interface3] mpls ldp ipv6 enable
[SwitchB-Vlan-interface3] mpls ldp transport-address 20::1
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.9
[SwitchC] mpls ldp
[SwitchC-ldp] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] mpls enable
[SwitchC-Vlan-interface3] mpls ldp ipv6 enable
[SwitchC-Vlan-interface3] mpls ldp transport-address 20::2
[SwitchC-Vlan-interface3] quit
```

4. Configure IPv6 LSP generation policies:

On Switch A, create IPv6 prefix list **switcha**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchA] ipv6 prefix-list switcha index 10 permit 100::1 128
[SwitchA] ipv6 prefix-list switcha index 20 permit 100::2 128
[SwitchA] ipv6 prefix-list switcha index 30 permit 100::3 128
[SwitchA] ipv6 prefix-list switcha index 40 permit 11::0 64
[SwitchA] ipv6 prefix-list switcha index 50 permit 21::0 64
[SwitchA] mpls ldp
[SwitchA-ldp] ipv6 lsp-trigger prefix-list switcha
[SwitchA-ldp] quit
```

On Switch B, create IPv6 prefix list **switchb**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchB] ipv6 prefix-list switchb index 10 permit 100::1 128
[SwitchB] ipv6 prefix-list switchb index 20 permit 100::2 128
[SwitchB] ipv6 prefix-list switchb index 30 permit 100::3 128
[SwitchB] ipv6 prefix-list switchb index 40 permit 11::0 64
[SwitchB] ipv6 prefix-list switchb index 50 permit 21::0 64
[SwitchB] mpls ldp
[SwitchB-ldp] ipv6 lsp-trigger prefix-list switchb
[SwitchB-ldp] quit
```

On Switch C, create IPv6 prefix list **switchc**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchC] ipv6 prefix-list switchc index 10 permit 100::1 128
[SwitchC] ipv6 prefix-list switchc index 20 permit 100::2 128
[SwitchC] ipv6 prefix-list switchc index 30 permit 100::3 128
[SwitchC] ipv6 prefix-list switchc index 40 permit 11::0 64
[SwitchC] ipv6 prefix-list switchc index 50 permit 21::0 64
```

```
[SwitchC] mpls ldp
[SwitchC-ldp] ipv6 lsp-trigger prefix-list switchc
[SwitchC-ldp] quit
```

Verifying the configuration

Display IPv6 LDP LSP information on the switches, for example, on Switch A.

```
[SwitchA] display mpls ldp lsp ipv6
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 5           Ingress: 3           Transit: 3           Egress: 2

FEC: 11::/64
In/Out Label: 24128/-                         OutInterface : -
Nexthop     : -
In/Out Label: -/24127(L)                      OutInterface : -
Nexthop     : -

FEC: 21::/64
In/Out Label: -/24126                         OutInterface : Vlan2
Nexthop     : FE80::7AAF:ABFF:FE27:302
In/Out Label: 24126/24126                     OutInterface : Vlan2
Nexthop     : FE80::7AAF:ABFF:FE27:302

FEC: 100::1/128
In/Out Label: 3/-                             OutInterface : -
Nexthop     : -
In/Out Label: -/24128(L)                      OutInterface : -
Nexthop     : -

FEC: 100::2/128
In/Out Label: -/3                             OutInterface : Vlan2
Nexthop     : FE80::7AAF:ABFF:FE27:302
In/Out Label: 24127/3                         OutInterface : Vlan2
Nexthop     : FE80::7AAF:ABFF:FE27:302

FEC: 100::3/128
In/Out Label: -/24125                         OutInterface : Vlan2
Nexthop     : FE80::7AAF:ABFF:FE27:302
In/Out Label: 24125/24125                     OutInterface : Vlan2
Nexthop     : FE80::7AAF:ABFF:FE27:302
```

Test the connectivity of the IPv6 LDP LSP from Switch A to Switch C.

```
[SwitchA] ping ipv6 -a 11::1 21::1
Ping6(56 data bytes) 11::1 --> 21::1, press CTRL+C to break
56 bytes from 21::1, icmp_seq=0 hlim=63 time=1.000 ms
56 bytes from 21::1, icmp_seq=1 hlim=63 time=2.000 ms
56 bytes from 21::1, icmp_seq=2 hlim=63 time=1.000 ms
56 bytes from 21::1, icmp_seq=3 hlim=63 time=2.000 ms
56 bytes from 21::1, icmp_seq=4 hlim=63 time=2.000 ms

--- Ping6 statistics for 21::1 ---
```

```
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/1.600/2.000/0.490 ms
```

Test the connectivity of the IPv6 LDP LSP from Switch C to Switch A.

```
[SwitchC] ping ipv6 -a 21::1 11::1
Ping6(56 data bytes) 21::1 --> 11::1, press CTRL+C to break
56 bytes from 11::1, icmp_seq=0 hlim=63 time=1.000 ms
56 bytes from 11::1, icmp_seq=1 hlim=63 time=3.000 ms
56 bytes from 11::1, icmp_seq=2 hlim=63 time=0.000 ms
56 bytes from 11::1, icmp_seq=3 hlim=63 time=2.000 ms
56 bytes from 11::1, icmp_seq=4 hlim=63 time=2.000 ms

--- Ping6 statistics for 11::1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/1.600/3.000/1.020 ms
```

Configuration files

- Switch A

```
#
 sysname SwitchA
#
ospfv3 1
 router-id 1.1.1.9
 area 0.0.0.0
#
 mpls lsr-id 1.1.1.9
#
vlan 2
#
vlan 4
#
mpls ldp
 ipv6 lsp-trigger prefix-list switcha
#
interface LoopBack0
 ip address 1.1.1.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::1/128
#
interface Vlan-interface2
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 10::1
 ipv6 address 10::1/64
#
interface Vlan-interface4
 ospfv3 1 area 0.0.0.0
 ipv6 address 11::1/64
```

```
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 4
#
 ipv6 prefix-list switcha index 10 permit 100::1 128
 ipv6 prefix-list switcha index 20 permit 100::2 128
 ipv6 prefix-list switcha index 30 permit 100::3 128
 ipv6 prefix-list switcha index 40 permit 11:: 64
 ipv6 prefix-list switcha index 50 permit 21:: 64
#
return
```

- Switch B

```
#
 sysname SwitchB
#
ospfv3 1
 router-id 2.2.2.9
 area 0.0.0.0
#
 mpls lsr-id 2.2.2.9
#
vlan 2 to 3
#
mpls ldp
 ipv6 lsp-trigger prefix-list switchb
#
interface LoopBack0
 ip address 2.2.2.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::2/128
#
interface Vlan-interface2
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 10::2
 ipv6 address 10::2/64
#
interface Vlan-interface3
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 20::1
```

```
 ipv6 address 20::1/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
 ipv6 prefix-list switchb index 10 permit 100::1 128
 ipv6 prefix-list switchb index 20 permit 100::2 128
 ipv6 prefix-list switchb index 30 permit 100::3 128
 ipv6 prefix-list switchb index 40 permit 11:: 64
 ipv6 prefix-list switchb index 50 permit 21:: 64
#
return
```

- Switch C

```
#
 sysname SwitchC
#
ospfv3 1
 router-id 3.3.3.9
 area 0.0.0.0
#
 mpls lsr-id 3.3.3.9
#
vlan 3
#
vlan 5
#
mpls ldp
 ipv6 lsp-trigger prefix-list switchc
#
interface LoopBack0
 ip address 3.3.3.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::3/128
#
interface Vlan-interface3
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 20::2
 ipv6 address 20::2/64
#
interface Vlan-interface5
 ospfv3 1 area 0.0.0.0
```

```
 ipv6 address 21::1/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 5
#
 ipv6 prefix-list switchc index 10 permit 100::1 128
 ipv6 prefix-list switchc index 20 permit 100::2 128
 ipv6 prefix-list switchc index 30 permit 100::3 128
 ipv6 prefix-list switchc index 40 permit 11:: 64
 ipv6 prefix-list switchc index 50 permit 21:: 64
#
return
```

Example: Configuring IPv6 label acceptance control

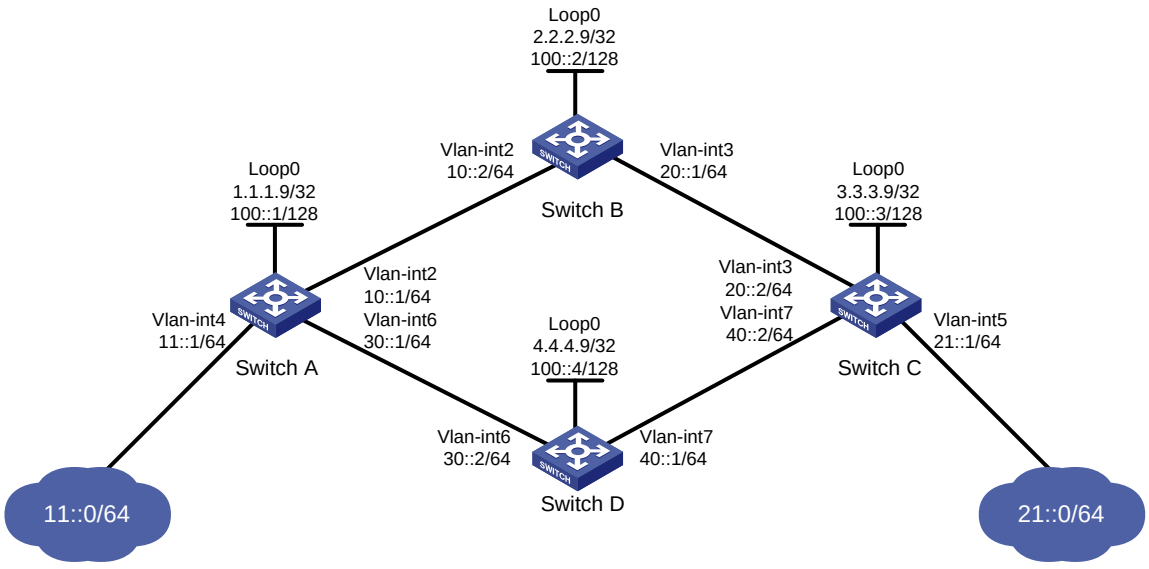
Network configuration

Two links, Switch A—Switch B—Switch C and Switch A—Switch D—Switch C, exist between subnets 11::0/64 and 21::0/64.

Configure LDP to establish LSPs only for routes to subnets 11::0/64 and 21::0/64.

Configure LDP to establish LSPs only on the link Switch A—Switch B—Switch C to forward traffic between subnets 11::0/64 and 21::0/64.

Figure 16 Network diagram



Requirements analysis

- To ensure that the LSRs establish IPv6 LSPs automatically, enable IPv6 LDP on each LSR.
- To establish IPv6 LDP LSPs, configure an IPv6 routing protocol to ensure IP connectivity between the LSRs. This example uses OSPFv3.

- To ensure that LDP establishes IPv6 LSPs only for the routes 11::0/64 and 21::0/64, configure IPv6 LSP generation policies on each LSR.
- To ensure that LDP establishes IPv6 LSPs only over the link Switch A—Switch B—Switch C, configure IPv6 label acceptance policies as follows:
  - Switch A accepts only the label mapping for FEC 21::0/64 received from Switch B. Switch A denies the label mapping for FEC 21::0/64 received from Switch D.
  - Switch C accepts only the label mapping for FEC 11::0/64 received from Switch B. Switch C denies the label mapping for FEC 11::0/64 received from Switch D.

Procedure

1. Create VLANs, assign IP addresses to the VLAN interfaces, and configure OSPFv3 on each switch to ensure IP connectivity between them:

Configure Switch A.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[RouterA] ospfv3
[RouterA-ospfv3-1] router-id 1.1.1.9
[RouterA-ospfv3-1] area 0
[RouterA-ospfv3-1-area-0.0.0.0] quit
[RouterA-ospfv3-1] quit
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/1
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address 10::1 64
[SwitchA-Vlan-interface2] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface2] quit
[SwitchA] vlan 4
[SwitchA-vlan4] port gigabitethernet 1/0/2
[SwitchA-vlan4] quit
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ipv6 address 11::1 64
[SwitchA-Vlan-interface4] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface4] quit
[SwitchA] vlan 6
[SwitchA-vlan6] port gigabitethernet 1/0/3
[SwitchA-vlan6] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] ip address 30::1 64
[SwitchA-Vlan-interface6] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface6] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.9 32
[SwitchA-LoopBack0] ipv6 address 100::1 128
[SwitchA-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchA-LoopBack0] quit
```

Configure Switch B.

```
<Sysname> system-view
[Sysname] sysname SwitchB
```

```
[RouterB] ospfv3
[RouterB-ospfv3-1] router-id 2.2.2.9
[RouterB-ospfv3-1] area 0
[RouterB-ospfv3-1-area-0.0.0.0] quit
[RouterB-ospfv3-1] quit
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address 10::2 64
[SwitchB-Vlan-interface2] ospfv3 1 area 0.0.0.0
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ipv6 address 20::1 64
[SwitchB-Vlan-interface3] ospfv3 1 area 0.0.0.0
[SwitchB-Vlan-interface3] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.9 32
[RouterB-LoopBack0] ipv6 address 100::2 128
[RouterB-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[RouterC] ospfv3
[RouterC-ospfv3-1] router-id 3.3.3.9
[RouterC-ospfv3-1] area 0
[RouterC-ospfv3-1-area-0.0.0.0] quit
[RouterC-ospfv3-1] quit
[SwitchC] vlan 3
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ipv6 address 20::2 64
[SwitchC-Vlan-interface3] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface3] quit
[SwitchC] vlan 5
[SwitchC-vlan5] port gigabitethernet 1/0/2
[SwitchC-vlan5] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ipv6 address 21::1 64
[SwitchC-Vlan-interface5] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface5] quit
[SwitchC] vlan 7
[SwitchC-vlan7] port gigabitethernet 1/0/3
```

```
[SwitchC-vlan7] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] ipv6 address 40::2 64
[SwitchC-Vlan-interface7] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface7] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 3.3.3.9 32
[SwitchC-LoopBack0] ipv6 address 100::3 128
[SwitchC-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchC-LoopBack0] quit
```

Configure Switch D.

```
<Sysname> system-view
[Sysname] sysname SwitchD
[RouterD] ospfv3
[RouterD-ospfv3-1] router-id 4.4.4.9
[RouterD-ospfv3-1] area 0
[RouterD-ospfv3-1-area-0.0.0.0] quit
[RouterD-ospfv3-1] quit
[SwitchD] vlan 6
[SwitchD-vlan6] port gigabitethernet 1/0/1
[SwitchD-vlan6] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] ipv6 address 30::2 64
[SwitchD-Vlan-interface6] ospfv3 1 area 0.0.0.0
[SwitchD-Vlan-interface6] quit
[SwitchD] vlan 7
[SwitchD-vlan7] port gigabitethernet 1/0/2
[SwitchD-vlan7] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] ipv6 address 40::1 24
[SwitchD-Vlan-interface7] ospfv3 1 area 0.0.0.0
[SwitchD-Vlan-interface7] quit
[SwitchD] interface loopback 0
[SwitchD-LoopBack0] ip address 4.4.4.9 32
[SwitchD-LoopBack0] ipv6 address 100::4 128
[SwitchD-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchD-LoopBack0] quit
```

2. Configure MSTP:

Configure Switch A.

```
[SwitchA] stp region-configuration
[SwitchA-mst-region] region-name example
[SwitchA-mst-region] instance 2 vlan 2
[SwitchA-mst-region] instance 4 vlan 4
[SwitchA-mst-region] instance 6 vlan 6
[SwitchA-mst-region] revision-level 0
[SwitchA-mst-region] active region-configuration
[SwitchA-mst-region] quit
[SwitchA] stp global enable
```

Configure Switch B.

```
[SwitchB] stp region-configuration
[SwitchB-mst-region] region-name example
[SwitchB-mst-region] instance 2 vlan 2
[SwitchB-mst-region] instance 3 vlan 3
[SwitchB-mst-region] revision-level 0
[SwitchB-mst-region] active region-configuration
[SwitchB-mst-region] quit
[SwitchB] stp global enable
```

Configure Switch C.

```
[SwitchC] stp region-configuration
[SwitchC-mst-region] region-name example
[SwitchC-mst-region] instance 3 vlan 3
[SwitchC-mst-region] instance 5 vlan 5
[SwitchC-mst-region] instance 7 vlan 7
[SwitchC-mst-region] revision-level 0
[SwitchC-mst-region] active region-configuration
[SwitchC-mst-region] quit
[SwitchC] stp global enable
```

Configure Switch D.

```
[SwitchD] stp region-configuration
[SwitchD-mst-region] region-name example
[SwitchD-mst-region] instance 6 vlan 6
[SwitchD-mst-region] instance 7 vlan 7
[SwitchD-mst-region] revision-level 0
[SwitchD-mst-region] active region-configuration
[SwitchD-mst-region] quit
[SwitchD] stp global enable
```

3. Enable MPLS and IPv6 LDP:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.9
[SwitchA] mpls ldp
[SwitchA-ldp] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] mpls enable
[SwitchA-Vlan-interface2] mpls ldp ipv6 enable
[SwitchA-Vlan-interface2] mpls ldp transport-address 10::1
[SwitchA-Vlan-interface2] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] mpls enable
[SwitchA-Vlan-interface6] mpls ldp ipv6 enable
[SwitchA-Vlan-interface6] mpls ldp transport-address 30::1
[SwitchA-Vlan-interface6] quit
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.9
[SwitchB] mpls ldp
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 2
```

```
[SwitchB-Vlan-interface2] mpls enable
[SwitchB-Vlan-interface2] mpls ldp ipv6 enable
[SwitchB-Vlan-interface2] mpls ldp transport-address 10::2
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mpls enable
[SwitchB-Vlan-interface3] mpls ldp ipv6 enable
[SwitchB-Vlan-interface3] mpls ldp transport-address 20::1
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.9
[SwitchC] mpls ldp
[SwitchC-ldp] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] mpls enable
[SwitchC-Vlan-interface3] mpls ldp ipv6 enable
[SwitchC-Vlan-interface3] mpls ldp transport-address 20::2
[SwitchC-Vlan-interface3] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] mpls enable
[SwitchC-Vlan-interface7] mpls ldp ipv6 enable
[SwitchC-Vlan-interface7] mpls ldp transport-address 40::2
[SwitchC-Vlan-interface7] quit
```

Configure Switch D.

```
[SwitchD] mpls lsr-id 4.4.4.9
[SwitchD] mpls ldp
[SwitchD-ldp] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] mpls enable
[SwitchD-Vlan-interface6] mpls ldp ipv6 enable
[SwitchD-Vlan-interface6] mpls ldp transport-address 30::2
[SwitchD-Vlan-interface6] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] mpls enable
[SwitchD-Vlan-interface7] mpls ldp ipv6 enable
[SwitchD-Vlan-interface7] mpls ldp transport-address 40::1
[SwitchD-Vlan-interface7] quit
```

4. Configure IPv6 LSP generation policies:

On Switch A, create IPv6 prefix list **switcha**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchA] ipv6 prefix-list switcha index 10 permit 11::0 64
[SwitchA] ipv6 prefix-list switcha index 20 permit 21::0 64
[SwitchA] mpls ldp
[SwitchA-ldp] ipv6 lsp-trigger prefix-list switcha
[SwitchA-ldp] quit
```

On Switch B, create IPv6 prefix list **switchb**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchB] ipv6 prefix-list switchb index 10 permit 11::0 64
```

```
[SwitchB] ipv6 prefix-list switchb index 20 permit 21::0 64
[SwitchB] mpls ldp
[SwitchB-ldp] ipv6 lsp-trigger prefix-list switchb
[SwitchB-ldp] quit
```

On Switch C, create IPv6 prefix list **switchc**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchC] ipv6 prefix-list switchc index 10 permit 11::0 64
[SwitchC] ipv6 prefix-list switchc index 20 permit 21::0 64
[SwitchC] mpls ldp
[SwitchC-ldp] ipv6 lsp-trigger prefix-list switchc
[SwitchC-ldp] quit
```

On Switch D, create IPv6 prefix list **switchd**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchD] ipv6 prefix-list switchd index 10 permit 11::0 64
[SwitchD] ipv6 prefix-list switchd index 20 permit 21::0 64
[SwitchD] mpls ldp
[SwitchD-ldp] ipv6 lsp-trigger prefix-list switchd
[SwitchD-ldp] quit
```

5. Configure IPv6 label acceptance policies:

On Switch A, create IPv6 prefix list **prefix-from-b** to permit subnet 21::0/64. Switch A uses this list to filter FEC-label mappings received from Switch B.

```
[SwitchA] ipv6 prefix-list prefix-from-b index 10 permit 21::0 64
```

On Switch A, create IPv6 prefix list **prefix-from-d** to deny subnet 21::0/64. Switch A uses this list to filter FEC-label mappings received from Switch D.

```
[SwitchA] ipv6 prefix-list prefix-from-d index 10 deny 21::0 64
```

On Switch A, configure IPv6 label acceptance policies to filter FEC-label mappings received from Switch B and Switch D.

```
[SwitchA] mpls ldp
[SwitchA-ldp] ipv6 accept-label peer 2.2.2.9 prefix-list prefix-from-b
[SwitchA-ldp] ipv6 accept-label peer 4.4.4.9 prefix-list prefix-from-d
[SwitchA-ldp] quit
```

On Switch C, create IPv6 prefix list **prefix-from-b** to permit subnet 11::0/64. Switch C uses this list to filter FEC-label mappings received from Switch B.

```
[SwitchC] ipv6 prefix-list prefix-from-b index 10 permit 11::0 64
```

On Switch C, create IPv6 prefix list **prefix-from-d** to deny subnet 11::0/64. Switch A uses this list to filter FEC-label mappings received from Switch D.

```
[SwitchC] ipv6 prefix-list prefix-from-d index 10 deny 11::0 64
```

On Switch C, configure IPv6 label acceptance policies to filter FEC-label mappings received from Switch B and Switch D.

```
[SwitchC] mpls ldp
[SwitchC-ldp] ipv6 accept-label peer 2.2.2.9 prefix-list prefix-from-b
[SwitchC-ldp] ipv6 accept-label peer 4.4.4.9 prefix-list prefix-from-d
[SwitchC-ldp] quit
```

Verifying the configuration

Display IPv6 LDP LSP information on the switches, for example, on Switch A. Verify that the next hop of the IPv6 LSP for FEC 21::0/64 is Switch B (FE80::20C:29FF:FE9D:EAC0). The IPv6 LSP has been established over the link Switch A—Switch B—Switch C, not over the link Switch A—Switch D—Switch C.

```
[SwitchA] display mpls ldp lsp ipv6
```

```
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 1           Transit: 1           Egress: 1

FEC: 11::/64
In/Out Label: 24128/-                          OutInterface : -
Nexthop     : -

FEC: 21::/64
In/Out Label: -/24127                          OutInterface : Vlan2
Nexthop     : FE80::7ABF:12FF:FE6C:202
In/Out Label: 24127/24127                      OutInterface : Vlan2
Nexthop     : FE80::7ABF:12FF:FE6C:202
```

Test the connectivity of the IPv6 LDP LSP from Switch A to Switch C.

```
[SwitchA] ping ipv6 -a 11::1 21::1
Ping6(56 data bytes) 11::1 --> 21::1, press CTRL+C to break
56 bytes from 21::1, icmp_seq=0 hlim=63 time=1.000 ms
56 bytes from 21::1, icmp_seq=1 hlim=63 time=1.000 ms
56 bytes from 21::1, icmp_seq=2 hlim=63 time=1.000 ms
56 bytes from 21::1, icmp_seq=3 hlim=63 time=0.000 ms
56 bytes from 21::1, icmp_seq=4 hlim=63 time=1.000 ms

--- Ping6 statistics for 21::1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/0.800/1.000/0.400 ms
```

Test the connectivity of the IPv6 LDP LSP from Switch C to Switch A.

```
[SwitchC] ping ipv6 -a 21::1 11::1
Ping6(56 data bytes) 21::1 --> 11::1, press CTRL+C to break
56 bytes from 11::1, icmp_seq=0 hlim=63 time=1.000 ms
56 bytes from 11::1, icmp_seq=1 hlim=63 time=1.000 ms
56 bytes from 11::1, icmp_seq=2 hlim=63 time=2.000 ms
56 bytes from 11::1, icmp_seq=3 hlim=63 time=1.000 ms
56 bytes from 11::1, icmp_seq=4 hlim=63 time=1.000 ms

--- Ping6 statistics for 11::1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/1.200/2.000/0.400 ms
```

Configuration files

- Switch A

```
#
 sysname SwitchA
#
ospfv3 1
 router-id 1.1.1.9
 area 0.0.0.0
#
 mpls lsr-id 1.1.1.9
#
```

```
vlan 2
#
vlan 4
#
vlan 6
#
stp region-configuration
 region-name example
 instance 2 vlan 2
 instance 4 vlan 4
 instance 6 vlan 6
 active region-configuration
#
 stp global enable
#
mpls ldp
 ipv6 lsp-trigger prefix-list switcha
 ipv6 accept-label peer 2.2.2.9 prefix-list prefix-from-b
 ipv6 accept-label peer 4.4.4.9 prefix-list prefix-from-d
#
interface LoopBack0
 ip address 1.1.1.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::1/128
#
interface Vlan-interface2
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 10::1
 ipv6 address 10::1/64
#
interface Vlan-interface4
 ospfv3 1 area 0.0.0.0
 ipv6 address 11::1/64
#
interface Vlan-interface6
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 30::1
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
```

```
 port access vlan 4
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 6
#
 ipv6 prefix-list prefix-from-b index 10 permit 21:: 64
 ipv6 prefix-list prefix-from-d index 10 deny 21:: 64
 ipv6 prefix-list switcha index 10 permit 11:: 64
 ipv6 prefix-list switcha index 20 permit 21:: 64
#
return
```

- Switch B

```
#
 sysname SwitchB
#
ospfv3 1
 router-id 2.2.2.9
 area 0.0.0.0
#
 mpls lsr-id 2.2.2.9
#
vlan 2 to 3
#
stp region-configuration
 region-name example
 instance 2 vlan 2
 instance 3 vlan 3
 active region-configuration
#
 stp global enable
#
mpls ldp
 ipv6 lsp-trigger prefix-list switchb
#
interface LoopBack0
 ip address 2.2.2.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::2/128
#
interface Vlan-interface2
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 10::2
 ipv6 address 10::2/64
#
interface Vlan-interface3
```

```
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 20::1
 ipv6 address 20::1/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
 ipv6 prefix-list switchb index 10 permit 11:: 64
 ipv6 prefix-list switchb index 20 permit 21:: 64
#
return
```

- Switch C

```
#
 sysname SwitchC
#
ospfv3 1
 router-id 3.3.3.9
 area 0.0.0.0
#
 mpls lsr-id 3.3.3.9
#
vlan 3
#
vlan 5
#
vlan 7
#
stp region-configuration
 region-name example
 instance 3 vlan 3
 instance 5 vlan 5
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
 ipv6 lsp-trigger prefix-list switchc
 ipv6 accept-label peer 2.2.2.9 prefix-list prefix-from-b
 ipv6 accept-label peer 4.4.4.9 prefix-list prefix-from-d
#
```

```
interface LoopBack0
 ip address 3.3.3.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::3/128
#
interface Vlan-interface3
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 20::2
 ipv6 address 20::2/64
#
interface Vlan-interface5
 ospfv3 1 area 0.0.0.0
 ipv6 address 21::1/64
#
interface Vlan-interface7
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 40::2
 ipv6 address 40::2/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 5
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 7
#
 ipv6 prefix-list prefix-from-b index 10 permit 11:: 64
 ipv6 prefix-list prefix-from-d index 10 deny 11:: 64
 ipv6 prefix-list switchc index 10 permit 11:: 64
 ipv6 prefix-list switchc index 20 permit 21:: 64
#
return
```

- Switch D

```
#
 sysname SwitchD
#
ospfv3 1
 router-id 4.4.4.9
 area 0.0.0.0
```

```
#
 mpls lsr-id 4.4.4.9
#
vlan 6 to 7
#
stp region-configuration
 region-name example
 instance 6 vlan 6
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
 ipv6 lsp-trigger prefix-list switchd
#
interface LoopBack0
 ip address 4.4.4.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::4/128
#
interface Vlan-interface6
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 30::2
 ipv6 address 30::2/64
#
interface Vlan-interface7
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 40::1
 ipv6 address 40::1/24
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 6
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 7
#
 ipv6 prefix-list switchd index 10 permit 11:: 64
 ipv6 prefix-list switchd index 20 permit 21:: 64
#
return
```

Example: Configuring IPv6 label advertisement control

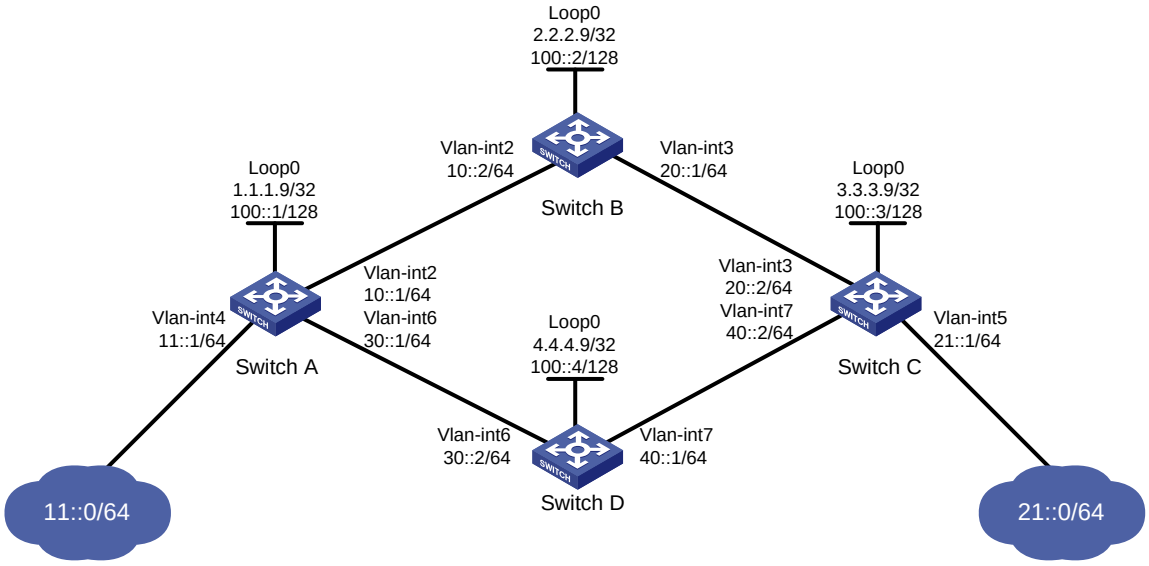
Network configuration

Two links, Switch A—Switch B—Switch C and Switch A—Switch D—Switch C, exist between subnets 11::0/64 and 21::0/64.

Configure LDP to establish LSPs only for routes to subnets 11::0/64 and 21::0/64.

Configure LDP to establish LSPs only on the link Switch A—Switch B—Switch C to forward traffic between subnets 11::0/64 and 21::0/64.

Figure 17 Network diagram



Requirements analysis

- To ensure that the LSRs establish IPv6 LSPs automatically, enable IPv6 LDP on each LSR.
- To establish IPv6 LDP LSPs, configure an IPv6 routing protocol to ensure IP connectivity between the LSRs. This example uses OSPFv3.
- To ensure that LDP establishes IPv6 LSPs only for the routes 11::0/64 and 21::0/64, configure IPv6 LSP generation policies on each LSR.
- To ensure that LDP establishes IPv6 LSPs only over the link Switch A—Switch B—Switch C, configure IPv6 label advertisement policies as follows:
  - Switch A advertises only the label mapping for FEC 11::0/64 to Switch B.
  - Switch C advertises only the label mapping for FEC 21::0/64 to Switch B.
  - Switch D does not advertise label mapping for FEC 21::0/64 to Switch A. Switch D does not advertise label mapping for FEC 11::0/64 to Switch C.

Prerequisites

Before configuration, disable the spanning tree feature globally or map each VLAN to an MSTI. For more information, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Create VLANs, assign IP addresses to the VLAN interfaces, and configure OSPFv3 on each switch to ensure IP connectivity between them:

   # Configure Switch A.

   ```
   <Sysname> system-view
   ```

```
[Sysname] sysname SwitchA
[RouterA] ospfv3
[RouterA-ospfv3-1] router-id 1.1.1.9
[RouterA-ospfv3-1] area 0
[RouterA-ospfv3-1-area-0.0.0.0] quit
[RouterA-ospfv3-1] quit
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/1
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address 10::1 64
[SwitchA-Vlan-interface2] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface2] quit
[SwitchA] vlan 4
[SwitchA-vlan4] port gigabitethernet 1/0/2
[SwitchA-vlan4] quit
[SwitchA] interface vlan-interface 4
[SwitchA-Vlan-interface4] ipv6 address 11::1 64
[SwitchA-Vlan-interface4] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface4] quit
[SwitchA] vlan 6
[SwitchA-vlan6] port gigabitethernet 1/0/3
[SwitchA-vlan6] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] ipv6 address 30::1 64
[SwitchA-Vlan-interface6] ospfv3 1 area 0.0.0.0
[SwitchA-Vlan-interface6] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.9 32
[SwitchA-LoopBack0] ipv6 address 100::1 128
[SwitchA-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchA-LoopBack0] quit
```

Configure Switch B.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[RouterB] ospfv3
[RouterB-ospfv3-1] router-id 2.2.2.9
[RouterB-ospfv3-1] area 0
[RouterB-ospfv3-1-area-0.0.0.0] quit
[RouterB-ospfv3-1] quit
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address 10::2 64
[SwitchB-Vlan-interface2] ospfv3 1 area 0.0.0.0
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
```

```
[SwitchB-vlan3] port gigabitethernet 1/0/2
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ipv6 address 20::1 64
[SwitchB-Vlan-interface3] ospfv3 1 area 0.0.0.0
[SwitchB-Vlan-interface3] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.9 32
[RouterB-LoopBack0] ipv6 address 100::2 128
[RouterB-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchB-LoopBack0] quit
```

Configure Switch C.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[RouterC] ospfv3
[RouterC-ospfv3-1] router-id 3.3.3.9
[RouterC-ospfv3-1] area 0
[RouterC-ospfv3-1-area-0.0.0.0] quit
[RouterC-ospfv3-1] quit
[SwitchC] vlan 3
[SwitchC-vlan3] port gigabitethernet 1/0/1
[SwitchC-vlan3] quit
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] ipv6 address 20::2 64
[SwitchC-Vlan-interface3] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface3] quit
[SwitchC] vlan 5
[SwitchC-vlan5] port gigabitethernet 1/0/2
[SwitchC-vlan5] quit
[SwitchC] interface vlan-interface 5
[SwitchC-Vlan-interface5] ipv6 address 21::1 64
[SwitchC-Vlan-interface5] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface5] quit
[SwitchC] vlan 7
[SwitchC-vlan7] port gigabitethernet 1/0/3
[SwitchC-vlan7] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] ipv6 address 40::2 64
[SwitchC-Vlan-interface7] ospfv3 1 area 0.0.0.0
[SwitchC-Vlan-interface7] quit
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 3.3.3.9 32
[SwitchC-LoopBack0] ipv6 address 100::3 128
[SwitchC-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchC-LoopBack0] quit
```

Configure Switch D.

```
<Sysname> system-view
[Sysname] sysname SwitchD
```

```
[RouterD] ospfv3
[RouterD-ospfv3-1] router-id 4.4.4.9
[RouterD-ospfv3-1] area 0
[RouterD-ospfv3-1-area-0.0.0.0] quit
[RouterD-ospfv3-1] quit
[SwitchD] vlan 6
[SwitchD-vlan6] port gigabitethernet 1/0/1
[SwitchD-vlan6] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] ipv6 address 30::2 64
[SwitchD-Vlan-interface6] ospfv3 1 area 0.0.0.0
[SwitchD-Vlan-interface6] quit
[SwitchD] vlan 7
[SwitchD-vlan7] port gigabitethernet 1/0/2
[SwitchD-vlan7] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] ipv6 address 40::1 24
[SwitchD-Vlan-interface7] ospfv3 1 area 0.0.0.0
[SwitchD-Vlan-interface7] quit
[SwitchD] interface loopback 0
[SwitchD-LoopBack0] ip address 4.4.4.9 32
[SwitchD-LoopBack0] ipv6 address 100::4 128
[SwitchD-LoopBack0] ospfv3 1 area 0.0.0.0
[SwitchD-LoopBack0] quit
```

2. Configure MSTP:

Configure Switch A.

```
[SwitchA] stp region-configuration
[SwitchA-mst-region] region-name example
[SwitchA-mst-region] instance 2 vlan 2
[SwitchA-mst-region] instance 4 vlan 4
[SwitchA-mst-region] instance 6 vlan 6
[SwitchA-mst-region] revision-level 0
[SwitchA-mst-region] active region-configuration
[SwitchA-mst-region] quit
[SwitchA] stp global enable
```

Configure Switch B.

```
[SwitchB] stp region-configuration
[SwitchB-mst-region] region-name example
[SwitchB-mst-region] instance 2 vlan 2
[SwitchB-mst-region] instance 3 vlan 3
[SwitchB-mst-region] revision-level 0
[SwitchB-mst-region] active region-configuration
[SwitchB-mst-region] quit
[SwitchB] stp global enable
```

Configure Switch C.

```
[SwitchC] stp region-configuration
[SwitchC-mst-region] region-name example
[SwitchC-mst-region] instance 3 vlan 3
```

```
[SwitchC-mst-region] instance 5 vlan 5
[SwitchC-mst-region] instance 7 vlan 7
[SwitchC-mst-region] revision-level 0
[SwitchC-mst-region] active region-configuration
[SwitchC-mst-region] quit
[SwitchC] stp global enable
```

Configure Switch D.

```
[SwitchD] stp region-configuration
[SwitchD-mst-region] region-name example
[SwitchD-mst-region] instance 6 vlan 6
[SwitchD-mst-region] instance 7 vlan 7
[SwitchD-mst-region] revision-level 0
[SwitchD-mst-region] active region-configuration
[SwitchD-mst-region] quit
[SwitchD] stp global enable
```

3. Enable MPLS and IPv6 LDP:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.9
[SwitchA] mpls ldp
[SwitchA-ldp] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] mpls enable
[SwitchA-Vlan-interface2] mpls ldp ipv6 enable
[SwitchA-Vlan-interface2] mpls ldp transport-address 10::1
[SwitchA-Vlan-interface2] quit
[SwitchA] interface vlan-interface 6
[SwitchA-Vlan-interface6] mpls enable
[SwitchA-Vlan-interface6] mpls ldp ipv6 enable
[SwitchA-Vlan-interface6] mpls ldp transport-address 30::1
[SwitchA-Vlan-interface6] quit
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.9
[SwitchB] mpls ldp
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] mpls enable
[SwitchB-Vlan-interface2] mpls ldp ipv6 enable
[SwitchB-Vlan-interface2] mpls ldp transport-address 10::2
[SwitchB-Vlan-interface2] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mpls enable
[SwitchB-Vlan-interface3] mpls ldp ipv6 enable
[SwitchB-Vlan-interface3] mpls ldp transport-address 20::1
[SwitchB-Vlan-interface3] quit
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.9
[SwitchC] mpls ldp
[SwitchC-ldp] quit
```

```
[SwitchC] interface vlan-interface 3
[SwitchC-Vlan-interface3] mpls enable
[SwitchC-Vlan-interface3] mpls ldp ipv6 enable
[SwitchC-Vlan-interface3] mpls ldp transport-address 20::2
[SwitchC-Vlan-interface3] quit
[SwitchC] interface vlan-interface 7
[SwitchC-Vlan-interface7] mpls enable
[SwitchC-Vlan-interface7] mpls ldp ipv6 enable
[SwitchC-Vlan-interface7] mpls ldp transport-address 40::2
[SwitchC-Vlan-interface7] quit
```

Configure Switch D.

```
[SwitchD] mpls lsr-id 4.4.4.9
[SwitchD] mpls ldp
[SwitchD-ldp] quit
[SwitchD] interface vlan-interface 6
[SwitchD-Vlan-interface6] mpls enable
[SwitchD-Vlan-interface6] mpls ldp ipv6 enable
[SwitchD-Vlan-interface6] mpls ldp transport-address 30::2
[SwitchD-Vlan-interface6] quit
[SwitchD] interface vlan-interface 7
[SwitchD-Vlan-interface7] mpls enable
[SwitchD-Vlan-interface7] mpls ldp ipv6 enable
[SwitchD-Vlan-interface7] mpls ldp transport-address 40::1
[SwitchD-Vlan-interface7] quit
```

4. Configure IPv6 LSP generation policies:

On Switch A, create IPv6 prefix list **switcha**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchA] ipv6 prefix-list switcha index 10 permit 11::0 64
[SwitchA] ipv6 prefix-list switcha index 20 permit 21::0 64
[SwitchA] mpls ldp
[SwitchA-ldp] ipv6 lsp-trigger prefix-list switcha
[SwitchA-ldp] quit
```

On Switch B, create IPv6 prefix list **switchb**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchB] ipv6 prefix-list switchb index 10 permit 11::0 64
[SwitchB] ipv6 prefix-list switchb index 20 permit 21::0 64
[SwitchB] mpls ldp
[SwitchB-ldp] ipv6 lsp-trigger prefix-list switchb
[SwitchB-ldp] quit
```

On Switch C, create IPv6 prefix list **switchc**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchC] ipv6 prefix-list switchc index 10 permit 11::0 64
[SwitchC] ipv6 prefix-list switchc index 20 permit 21::0 64
[SwitchC] mpls ldp
[SwitchC-ldp] ipv6 lsp-trigger prefix-list switchc
[SwitchC-ldp] quit
```

On Switch D, create IPv6 prefix list **switchd**, and configure LDP to use only the routes permitted by the prefix list to establish IPv6 LSPs.

```
[SwitchD] ipv6 prefix-list switchd index 10 permit 11::0 64
```

```
[SwitchD] ipv6 prefix-list switchd index 20 permit 21::0 64
[SwitchD] mpls ldp
[SwitchD-ldp] ipv6 lsp-trigger prefix-list switchd
[SwitchD-ldp] quit
```

5. Configure IPv6 label advertisement policies:

On Switch A, create IPv6 prefix list **prefix-to-b** to permit subnet 11::0/64. Switch A uses this list to filter FEC-label mappings advertised to Switch B.

```
[SwitchA] ipv6 prefix-list prefix-to-b index 10 permit 11::0 64
```

On Switch A, create IP prefix list **peer-b** to permit 2.2.2.9/32. Switch A uses this list to filter peers.

```
[SwitchA] ip prefix-list peer-b index 10 permit 2.2.2.9 32
```

On Switch A, configure an IPv6 label advertisement policy to advertise only the label mapping for FEC 11::0/64 to Switch B.

```
[SwitchA] mpls ldp
[SwitchA-ldp] ipv6 advertise-label prefix-list prefix-to-b peer peer-b
[SwitchA-ldp] quit
```

On Switch C, create IPv6 prefix list **prefix-to-b** to permit subnet 21::0/64. Switch C uses this list to filter FEC-label mappings advertised to Switch B.

```
[SwitchC] ipv6 prefix-list prefix-to-b index 10 permit 21::0 64
```

On Switch C, create IP prefix list **peer-b** to permit 2.2.2.9/32. Switch C uses this list to filter peers.

```
[SwitchC] ip prefix-list peer-b index 10 permit 2.2.2.9 32
```

On Switch C, configure an IPv6 label advertisement policy to advertise only the label mapping for FEC 21::0/64 to Switch B.

```
[SwitchC] mpls ldp
[SwitchC-ldp] ipv6 advertise-label prefix-list prefix-to-b peer peer-b
[SwitchC-ldp] quit
```

On Switch D, create IPv6 prefix list **prefix-to-a** to deny subnet 21::0/64. Switch D uses this list to filter FEC-label mappings to be advertised to Switch A.

```
[SwitchD] ipv6 prefix-list prefix-to-a index 10 deny 21::0 64
[SwitchD] ipv6 prefix-list prefix-to-a index 20 permit 0::0 0 less-equal 128
```

On Switch D, create IP prefix list **peer-a** to permit 1.1.1.9/32. Switch D uses this list to filter peers.

```
[SwitchD] ip prefix-list peer-a index 10 permit 1.1.1.9 32
```

On Switch D, create IPv6 prefix list **prefix-to-c** to deny subnet 11::0/64. Switch D uses this list to filter FEC-label mappings to be advertised to Switch C.

```
[SwitchD] ipv6 prefix-list prefix-to-c index 10 deny 11::0 64
[SwitchD] ipv6 prefix-list prefix-to-c index 20 permit 0::0 0 less-equal 128
```

On Switch D, create IP prefix list **peer-c** to permit subnet 3.3.3.9/32. Switch D uses this list to filter peers.

```
[SwitchD] ip prefix-list peer-c index 10 permit 3.3.3.9 32
```

On Switch D, configure an IPv6 label advertisement policy. This policy ensures that Switch D does not advertise label mappings for FEC 21::0/64 to Switch A, and does not advertise label mappings for FEC 11::0/64 to Switch C.

```
[SwitchD] mpls ldp
[SwitchD-ldp] ipv6 advertise-label prefix-list prefix-to-a peer peer-a
[SwitchD-ldp] ipv6 advertise-label prefix-list prefix-to-c peer peer-c
[SwitchD-ldp] quit
```

Verifying the configuration

Display LDP LSP information on the switches, for example, on Switch A. Verify that Switch A and Switch C have received FEC-label mappings only from Switch B. Switch B has received FEC-label mappings from both Switch A and Switch C. Switch D does not receive FEC-label mappings from Switch A or Switch C. LDP has established an IPv6 LSP only over the link Switch A—Switch B—Switch C.

```
[SwitchA] display mpls ldp lsp ipv6
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 1           Transit: 1           Egress: 1

FEC: 11::/64
In/Out Label: 24128/-                        OutInterface : -
Nexthop     : -
In/Out Label: -/24128(L)                     OutInterface : -
Nexthop     : -
In/Out Label: -/24128(L)                     OutInterface : -
Nexthop     : -

FEC: 21::/64
In/Out Label: -/24127                        OutInterface : Vlan2
Nexthop     : FE80::7AEB:BFF:FEF8:202
In/Out Label: 24127/24127                    OutInterface : Vlan2
Nexthop     : FE80::7AEB:BFF:FEF8:202
[SwitchB] display mpls ldp lsp ipv6
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 2           Transit: 2           Egress: 0

FEC: 11::/64
In/Out Label: -/24128                        OutInterface : Vlan2
Nexthop     : FE80::7AEB:8FF:FEF9:102
In/Out Label: 24128/24128                    OutInterface : Vlan2
Nexthop     : FE80::7AEB:8FF:FEF9:102

FEC: 21::/64
In/Out Label: -/24128                        OutInterface : Vlan3
Nexthop     : FE80::7AEB:EFF:FEB1:302
In/Out Label: 24127/24128                    OutInterface : Vlan3
Nexthop     : FE80::7AEB:EFF:FEB1:302
[SwitchC] display mpls ldp lsp ipv6
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 1           Transit: 1           Egress: 1

FEC: 11::/64
In/Out Label: -/24128                        OutInterface : Vlan3
Nexthop     : FE80::7AEB:BFF:FEF8:202
In/Out Label: 24127/24128                    OutInterface : Vlan3
Nexthop     : FE80::7AEB:BFF:FEF8:202

FEC: 21::/64
```

```
In/Out Label: 24128/-                          OutInterface : -
Nexthop     : -
In/Out Label: -/24127(L)                       OutInterface : -
Nexthop     : -
[SwitchD] display mpls ldp lsp ipv6
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 2           Ingress: 0           Transit: 0           Egress: 2

FEC: 11::/64
In/Out Label: 24128/-                          OutInterface : -
Nexthop     : -

FEC: 21::/64
In/Out Label: 24127/-                          OutInterface : -
Nexthop     : -
```

Test the connectivity of the IPv6 LDP LSP from Switch A to Switch C.

```
[SwitchA] ping ipv6 -a 11::1 21::1
Ping6(56 data bytes) 11::1 --> 21::1, press CTRL+C to break
56 bytes from 21::1, icmp_seq=0 hlim=63 time=0.000 ms
56 bytes from 21::1, icmp_seq=1 hlim=63 time=2.000 ms
56 bytes from 21::1, icmp_seq=2 hlim=63 time=2.000 ms
56 bytes from 21::1, icmp_seq=3 hlim=63 time=2.000 ms
56 bytes from 21::1, icmp_seq=4 hlim=63 time=1.000 ms

--- Ping6 statistics for 21::1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/1.400/2.000/0.800 ms
```

Test the connectivity of the IPv6 LDP LSP from Switch C to Switch A.

```
[SwitchC] ping ipv6 -a 21::1 11::1
Ping6(56 data bytes) 21::1 --> 11::1, press CTRL+C to break
56 bytes from 11::1, icmp_seq=0 hlim=63 time=1.000 ms
56 bytes from 11::1, icmp_seq=1 hlim=63 time=1.000 ms
56 bytes from 11::1, icmp_seq=2 hlim=63 time=2.000 ms
56 bytes from 11::1, icmp_seq=3 hlim=63 time=1.000 ms
56 bytes from 11::1, icmp_seq=4 hlim=63 time=2.000 ms

--- Ping6 statistics for 11::1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.000/1.400/2.000/0.490 ms
```

Configuration files

- Switch A

```
#
 sysname SwitchA
#
ospfv3 1
 router-id 1.1.1.9
 area 0.0.0.0
```

```
#
 mpls lsr-id 1.1.1.9
#
vlan 2
#
vlan 4
#
vlan 6
#
stp region-configuration
 region-name example
 instance 2 vlan 2
 instance 4 vlan 4
 instance 6 vlan 6
 active region-configuration
#
 stp global enable
#
mpls ldp
 ipv6 lsp-trigger prefix-list switcha
 ipv6 advertise-label prefix-list prefix-to-b peer peer-b
#
interface LoopBack0
 ip address 1.1.1.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::1/128
#
interface Vlan-interface2
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 10::1
 ipv6 address 10::1/64
#
interface Vlan-interface4
 ospfv3 1 area 0.0.0.0
 ipv6 address 11::1/64
#
interface Vlan-interface6
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 30::1
 ipv6 address 30::1/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
```

```
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 4
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 6
#
 ip prefix-list peer-b index 10 permit 2.2.2.9 32
#
 ipv6 prefix-list prefix-to-b index 10 permit 11:: 64
 ipv6 prefix-list switcha index 10 permit 11:: 64
 ipv6 prefix-list switcha index 20 permit 21:: 64
#
return
```

- Switch B

```
#
 sysname SwitchB
#
ospfv3 1
 router-id 2.2.2.9
 area 0.0.0.0
#
 mpls lsr-id 2.2.2.9
#
vlan 2 to 3
#
stp region-configuration
 region-name example
 instance 2 vlan 2
 instance 3 vlan 3
 active region-configuration
#
 stp global enable
#
mpls ldp
 ipv6 lsp-trigger prefix-list switchb
#
interface LoopBack0
 ip address 2.2.2.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::2/128
#
interface Vlan-interface2
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
```

```
 mpls ldp transport-address 10::2
 ipv6 address 10::2/64
#
interface Vlan-interface3
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 20::1
 ipv6 address 20::1/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 2
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 3
#
 ipv6 prefix-list switchb index 10 permit 11:: 64
 ipv6 prefix-list switchb index 20 permit 21:: 64
#
return
```

- Switch C

```
#
 sysname SwitchC
#
ospfv3 1
 router-id 3.3.3.9
 area 0.0.0.0
#
 mpls lsr-id 3.3.3.9
#
vlan 3
#
vlan 5
#
vlan 7
#
stp region-configuration
 region-name example
 instance 3 vlan 3
 instance 5 vlan 5
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
```

```
 ipv6 lsp-trigger prefix-list switchc
 ipv6 advertise-label prefix-list prefix-to-b peer peer-b
#
interface LoopBack0
 ip address 3.3.3.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::3/128
#
interface Vlan-interface3
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 20::2
 ipv6 address 20::2/64
#
interface Vlan-interface5
 ospfv3 1 area 0.0.0.0
 ipv6 address 21::1/64
#
interface Vlan-interface7
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 40::2
 ipv6 address 40::2/64
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 3
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 5
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 port access vlan 7
#
 ip prefix-list peer-b index 10 permit 2.2.2.9 32
#
 ipv6 prefix-list prefix-to-b index 10 permit 21:: 64
 ipv6 prefix-list switchc index 10 permit 11:: 64
 ipv6 prefix-list switchc index 20 permit 21:: 64
#
return
```

- Switch D

```
#
 sysname SwitchD
```

```
#
ospfv3 1
 router-id 4.4.4.9
 area 0.0.0.0
#
 mpls lsr-id 4.4.4.9
#
vlan 6 to 7
#
stp region-configuration
 region-name example
 instance 6 vlan 6
 instance 7 vlan 7
 active region-configuration
#
 stp global enable
#
mpls ldp
 ipv6 lsp-trigger prefix-list switchd
 ipv6 advertise-label prefix-list prefix-to-a peer peer-a
 ipv6 advertise-label prefix-list prefix-to-c peer peer-c
#
interface LoopBack0
 ip address 4.4.4.9 255.255.255.255
 ospfv3 1 area 0.0.0.0
 ipv6 address 100::4/128
#
interface Vlan-interface6
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 30::2
 ipv6 address 30::2/64
#
interface Vlan-interface7
 ospfv3 1 area 0.0.0.0
 mpls enable
 mpls ldp ipv6 enable
 mpls ldp transport-address 40::1
 ipv6 address 40::1/24
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port access vlan 6
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port access vlan 7
```

```
#
 ip prefix-list peer-a index 10 permit 1.1.1.9 32
 ip prefix-list peer-c index 10 permit 3.3.3.9 32
#
 ipv6 prefix-list prefix-to-a index 10 deny 21:: 64
 ipv6 prefix-list prefix-to-a index 20 permit :: 0 less-equal 128
 ipv6 prefix-list prefix-to-c index 10 deny 11:: 64
 ipv6 prefix-list prefix-to-c index 20 permit :: 0 less-equal 128
 ipv6 prefix-list switchd index 10 permit 11:: 64
 ipv6 prefix-list switchd index 20 permit 21:: 64
#
return
```

Contents

Configuring tunnel policies 1
About tunnel policies 1
Restrictions and guidelines: Tunnel policy configuration 1
Configuring a tunnel policy 1
Display and maintenance commands for tunnel policies 1

Configuring tunnel policies

About tunnel policies

Tunnel policies enable a PE to forward traffic for each MPLS VPN over a preferred tunnel or load share the traffic over multiple tunnels. Using tunnel policies can facilitate network planning and management and reduce processing overhead on PEs.

For more information about MPLS VPN, see "Configuring MPLS L3VPN."

Restrictions and guidelines: Tunnel policy configuration

Only advanced mode and expert mode support tunnel policy. For more information about system operating modes, see device management in *Fundamentals Configuration Guide*.

The device does not support MPLS forwarding for tunneled packets.

To configure a VPN to exclusively use a tunnel, perform the following operations:

1. Use the **preferred-path** command to specify the tunnel as the preferred tunnel in a tunnel policy.
2. Apply the policy only to that VPN.

Configuring a tunnel policy

1. Enter system view.
   system-view
2. Create a tunnel policy, and enter tunnel policy view.
   tunnel-policy *tunnel-policy-name*
3. Configure the tunnel policy.
   Choose one option as needed:
   select-seq [**strict**] **lsp load-balance-number** *number*
   By default, the policy selects only one tunnel.

Display and maintenance commands for tunnel policies

Execute **display** commands in any view.

Task	Command
Display tunnel information.	**display mpls tunnel** { **all** | **statistics** | [**vpn-instance** *vpn-instance-name*] **destination** *ipv4-address* }
Display tunnel policy information.	**display tunnel-policy** [*tunnel-policy-name*]

Contents

Configuring MPLS L3VPN 3

About MPLS L3VPN 3
Basic MPLS L3VPN architecture 3
MPLS L3VPN concepts 3
MPLS L3VPN route advertisement 5
MPLS L3VPN packet forwarding 6
MPLS L3VPN networking schemes 7
Inter-AS VPN 9
Carrier's carrier 14
Nested VPN 15
HoVPN 17
OSPF VPN extension 18
BGP AS number substitution and SoO attribute 21
MPLS L3VPN FRR 22
ECMP VPN route redistribution 24
Protocols and standards 24
MPLS L3VPN tasks at a glance 24
Prerequisites for MPLS L3VPN 25
Configuring VPN instances 26
Creating a VPN instance 26
Associating a VPN instance with a Layer 3 interface 26
Configuring route related attributes for a VPN instance 27
Configuring routing between a PE and a CE 28
Configuring static routing between a PE and a CE 28
Configuring RIP between a PE and a CE 28
Configuring OSPF between a PE and a CE 28
Configuring IS-IS between a PE and a CE 30
Configuring EBGP between a PE and a CE 30
Configuring IBGP between a PE and a CE 32
Configuring routing between PEs 33
Configuring BGP VPNv4 route control 33
About BGP VPNv4 route control 33
Controlling BGP VPNv4 route advertisement, reception, and saving 33
Setting a preferred value for received routes 34
Configuring BGP VPNv4 route reflection 34
Configuring BGP VPNv4 route attributes 35
Configuring BGP VPNv4 route filtering 36
Configuring BGP VPNv4 route dampening 37
Configuring BGP VPNv4 optimal route selection delay 37
Advertising BGP RPKI validation state to a peer or peer group 38
Configuring inter-AS VPN 38
Configuring inter-AS option A 38
Configuring inter-AS option B 38
Configuring inter-AS option C (method 1) 40
Configuring inter-AS option C (method 2) 42
Configuring nested VPN 44
Configuring HoVPN 46
Configuring the UPE 46
Configuring the SPE 46
Configuring route re-origination 47
Specifying the VPN label processing mode on the egress PE 48
Configuring MPLS L3VPN FRR 48
About MPLS L3VPN FRR 48
Restrictions and guidelines 49
Configuring FRR by using a routing policy 49
Enabling MPLS L3VPN FRR for BGP-VPN IPv4 unicast address family 50
Configuring an OSPF sham link 50

About OSPF sham links ····· 50
Prerequisites ····· 50
Redistributing the loopback interface address ····· 51
Creating a sham link ····· 51
Configuring BGP AS number substitution and SoO attribute ····· 52
Configuring BGP RT filtering ····· 52
Configuring the BGP additional path feature ····· 53
Configuring link bandwidth attributes for BGP VPNv4 routes ····· 54
Minimizing the priority of BGP routes sent to peers ····· 57
Minimizing the priority of BGP routes sent to a peer after the peer state changes from down to up ····· 57
Minimizing the priority of BGP routes sent to peers after the device restarts ····· 60
Restoring the priority of BGP routes sent to peers ····· 61
Specifying a local network to be advertised in the public instance or a VPN instance ····· 61
Configuring BGP VPNv4 routes to use private network next hops ····· 62
Configuring route replication ····· 63
Configuring the public instance ····· 63
Configuring route replication for public and VPN instances ····· 64
Configuring BGP route replication between public and VPN instances ····· 65
Configuring route re-origination ····· 65
Enabling ECMP VPN route redistribution ····· 66
Enabling prioritized withdrawal of specific routes ····· 66
Enabling SNMP notifications for MPLS L3VPN ····· 67
Enabling logging for BGP route flapping ····· 67
Display and maintenance commands for MPLS L3VPN ····· 68
Resetting BGP connections ····· 68
Displaying MPLS L3VPN information ····· 68
MPLS L3VPN configuration examples ····· 71
Example: Configuring basic MPLS L3VPN ····· 71
Example: Configuring a hub-spoke network ····· 76
Example: Configuring MPLS L3VPN inter-AS option A ····· 82
Example: Configuring MPLS L3VPN inter-AS option B ····· 87
Example: Configuring MPLS L3VPN inter-AS option C (method 1) ····· 92
Example: Configuring MPLS L3VPN inter-AS option C (method 2) ····· 99
Example: Configuring MPLS L3VPN carrier's carrier in the same AS ····· 105
Example: Configuring MPLS L3VPN carrier's carrier in different ASs ····· 112
Example: Configuring nested VPN ····· 119
Example: Configuring HoVPN ····· 129
Example: Configuring an OSPF sham link ····· 136
Example: Configuring BGP AS number substitution ····· 140
Example: Configuring BGP AS number substitution and SoO attribute ····· 144
Example: Configuring MPLS L3VPN FRR through VPNv4 route backup for an IPv4 route ····· 147

Configuring MPLS L3VPN

About MPLS L3VPN

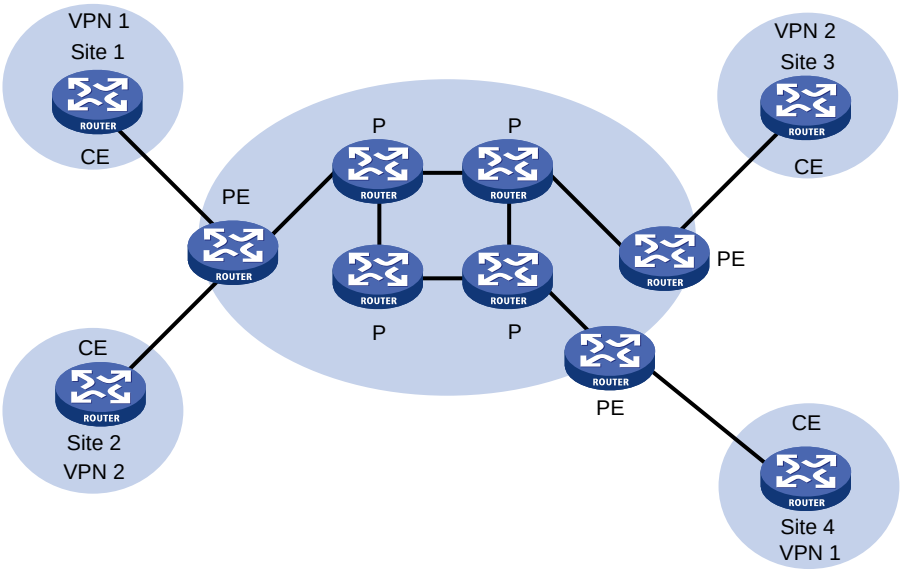
MPLS L3VPN is a L3VPN technology used to interconnect geographically dispersed VPN sites. MPLS L3VPN uses BGP to advertise VPN routes and uses MPLS to forward VPN packets over a service provider backbone.

Basic MPLS L3VPN architecture

As shown in Figure 1, a basic MPLS L3VPN architecture has the following types of devices:

- **Customer edge device**—A CE device resides on a customer network and has one or more interfaces directly connected to a service provider network. It does not support MPLS.
- **Provider edge device**—A PE device resides at the edge of a service provider network and is connected to one or more CEs. All MPLS VPN services are processed on PEs.
- **Provider device**—A P device is a core device on a service provider network. It is not directly connected to any CEs. A P device has only basic MPLS forwarding capability and does not handle VPN routing information.

Figure 1 Basic MPLS L3VPN architecture



MPLS L3VPN concepts

Site

A site has the following features:

- A site is a group of IP systems with IP connectivity that does not rely on any service provider networks.
- The classification of a site depends on the topology relationship of the devices, rather than the geographical positions. However, the devices at a site are, in most cases, adjacent to each other geographically.

- The devices at a site can belong to multiple VPNs, which means that a site can belong to multiple VPNs.
- A site is connected to a provider network through one or more CEs. A site can contain multiple CEs, but a CE can belong to only one site.

Sites connected to the same provider network can be classified into different sets by policies. Only the sites in the same set can access each other through the provider network. Such a set is called a VPN.

VPN instance

VPN instances implement route isolation, data independence, and data security for VPNs.

A VPN instance has the following components:

- A separate Label Forwarding Information Base (LFIB).
- An IP routing table.
- Interfaces bound to the VPN instance.
- VPN instance administration information, including route distinguishers (RDs), route targets (RTs), and route filtering policies.

To associate a site with a VPN instance, bind the VPN instance to the PE's interface connected to the site. A site can be associated with only one VPN instance, and different sites can be associated with the same VPN instance. A VPN instance contains the VPN membership and routing rules of associated sites.

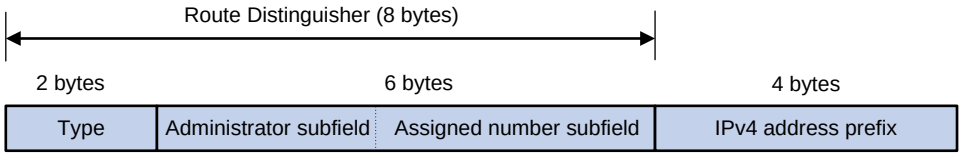
VPN-IPv4 address

Each VPN independently manages its address space. The address spaces of VPNs might overlap. For example, if both VPN 1 and VPN 2 use the addresses on subnet 10.110.10.0/24, address space overlapping occurs.

BGP cannot process overlapping VPN address spaces. For example, if both VPN 1 and VPN 2 use the subnet 10.110.10.0/24 and each advertise a route destined for the subnet, BGP selects only one of them. This results in the loss of the other route.

Multiprotocol BGP (MP-BGP) can solve this problem by advertising VPN-IPv4 addresses (also called VPNv4 addresses).

Figure 2 VPN-IPv4 address structure



As shown in Figure 2, a VPN-IPv4 address consists of 12 bytes. The first eight bytes represent the RD, followed by a four-byte IPv4 prefix. The RD and the IPv4 prefix form a unique VPN-IPv4 prefix.

An RD can be in one of the following formats:

- When the Type field is 0, the Administrator subfield occupies two bytes, the Assigned number subfield occupies four bytes, and the RD format is *16-bit AS number*:*32-bit user-defined number*. For example, 100:1.
- When the Type field is 1, the Administrator subfield occupies four bytes, the Assigned number subfield occupies two bytes, and the RD format is *32-bit IPv4 address*:*16-bit user-defined number*. For example, 172.1.1.1:1.
- When the Type field is 2, the Administrator subfield occupies four bytes, the Assigned number subfield occupies two bytes, and the RD format is *32-bit AS number*:*16-bit user-defined number*, where the minimum value of the AS number is 65536. For example, 65536:1.

NOTE:

When you configure an RD through CLI, you can specify an AS number in integer format or dotted decimal notation. For example, 65536:1 and 1.0:1 specify the same RD.

To guarantee global uniqueness for a VPN-IPv4 address, do not set the Administrator subfield to any private AS number or private IP address.

Route target attribute

MPLS L3VPN uses route target (also called VPN target) community attributes to control the advertisement of VPN routing information. A VPN instance on a PE supports the following types of route target attributes:

- **Export target attribute**—A PE sets the export target attribute for VPN-IPv4 routes learned from directly connected sites before advertising them to other PEs.
- **Import target attribute**—A PE checks the export target attribute of VPN-IPv4 routes received from other PEs. If the export target attribute matches the import target attribute of a VPN instance, the PE adds the routes to the routing table of the VPN instance.

Route target attributes define which sites can receive VPN-IPv4 routes, and from which sites a PE can receive routes.

Like RDs, route target attributes can be one of the following formats:

- *16-bit AS number*:*32-bit user-defined number*. For example, 100:1.
- *32-bit IPv4 address:16-bit user-defined number*. For example, 172.1.1.1:1.
- *32-bit AS number*:*16-bit user-defined number*, where the minimum value of the AS number is 65536. For example, 65536:1.

NOTE:

When you configure an RD from the CLI, you can specify an AS number in integer format or dotted decimal notation. For example, values 65536:1 and 1.0:1 specify the same RD.

-

MP-BGP

MP-BGP supports multiple address families, including IPv4 multicast and VPN-IPv4 address families.

In MPLS L3VPN, MP-BGP advertises VPN-IPv4 routes for VPN sites between PEs.

MPLS L3VPN route advertisement

In a basic MPLS L3VPN, CEs and PEs are responsible for advertising VPN routing information. P routers maintain only the routes within the backbone. A PE maintains only routing information for directly connected VPNs, rather than for all VPNs.

VPN routing information is advertised through the path local CE—ingress PE—egress PE—remote CE.

Route advertisement from the local CE to the ingress PE

The CE advertises standard IPv4 routing information to the ingress PE over a static route, RIP route, OSPF route, IS-IS route, EBGP route, or IBGP route.

Route advertisement from the ingress PE to the egress PE

The ingress PE performs the following operations:

1. Adds RDs and route target attributes to these standard IPv4 routes and assign these routes MPLS labels to create VPN-IPv4 routes.

2. Saves the VPN-IPv4 routes to the routing table of the VPN instance created for the CE.
3. Advertises the VPN-IPv4 routes (with export targets and MPLS labels) to the egress PE through MP-BGP.

Route advertisement from the egress PE to the remote CE

After receiving the VPN-IPv4 routes, the egress PE performs the following operations:

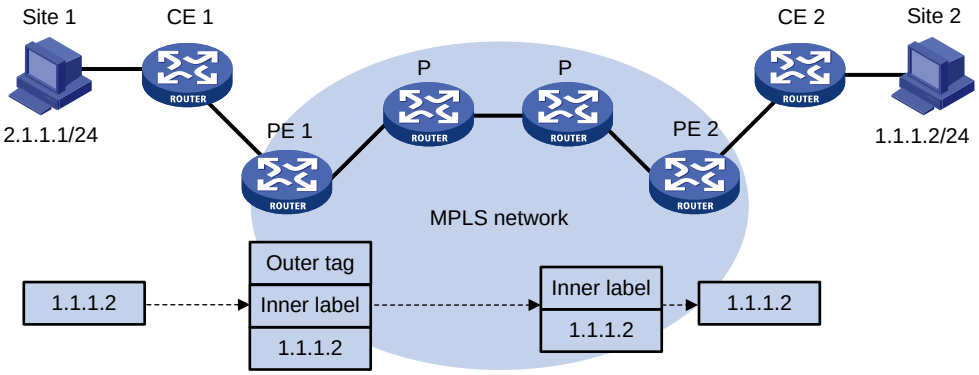
1. Compares the routes' export target attributes with the local import target attributes.
2. Adds the routes to the routing table of the VPN instance if the export and local import target attributes match each other.
3. Restores the VPN-IPv4 routes to the original IPv4 routes.
4. Advertises those routes to the connected CE over a static route, RIP route, OSPF route, IS-IS route, EBGP route, or IBGP route.

MPLS L3VPN packet forwarding

In a basic MPLS L3VPN (within a single AS), a PE adds the following information into VPN packets:

- **Outer tag**—Identifies the public tunnel from the local PE to the remote PE. The public tunnel can be an LSP or a GRE tunnel. Based on the outer tag, a VPN packet can be forwarded along the public tunnel to the remote PE. For a GRE public tunnel, the outer tag is the GRE encapsulation. For an LSP tunnel, the outer tag is an MPLS label.
- **Inner label**—Identifies the remote VPN site. The remote PE uses the inner label to forward packets to the target VPN site. MP-BGP advertises inner labels for VPN-IPv4 routes among PEs.

Figure 3 VPN packet forwarding



As shown in Figure 3, a VPN packet is forwarded from Site 1 to Site 2 by using the following process:

1. Site 1 sends an IP packet with the destination address 1.1.1.2. CE 1 transmits the packet to PE 1.
2. PE 1 performs the following operations:
   a. Finds the matching VPN route based on the inbound interface and destination address of the packet.
   b. Labels the packet with both the inner label and the outer tag.
   c. Forwards the packet to the public tunnel.
3. P devices forward the packet to PE 2 by the outer tag.
   - If the outer tag is an MPLS label, the label is removed from the packet at the penultimate hop.
   - If the outer tag is GRE encapsulation, PE 2 removes the GRE encapsulation.
4. PE 2 performs the following operations:

a. Uses the inner label to find the matching VPN instance to which the destination address of the packet belongs.
b. Looks up the routing table of the VPN instance for the output interface.
c. Removes the inner label and forwards the packet out of the interface to CE 2.

5. CE 2 transmits the packet to the destination through IP forwarding.

When two sites of a VPN are connected to the same PE, the PE directly forwards packets between the two sites through the VPN routing table without adding any tag or label.

MPLS L3VPN networking schemes

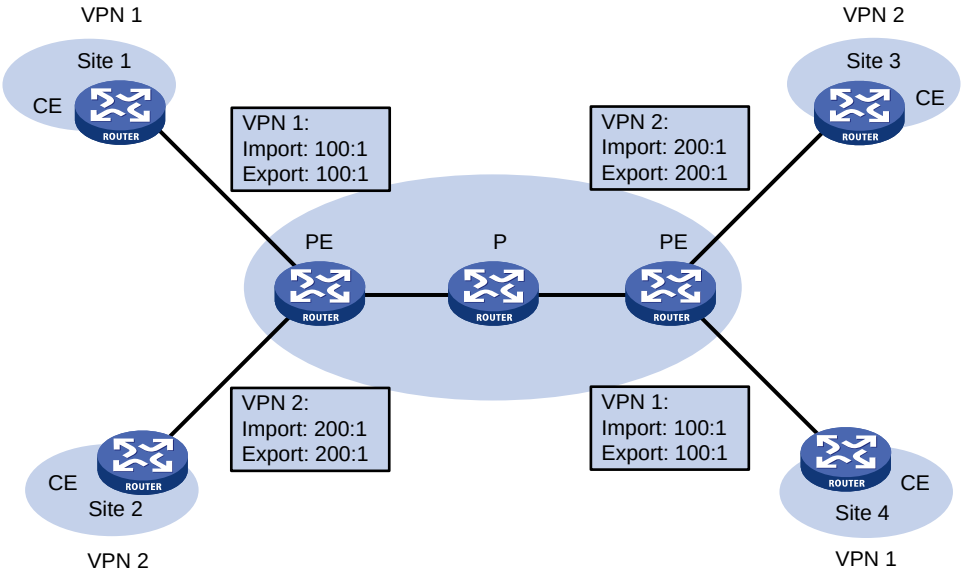
In MPLS L3VPNs, route target attributes are used to control the advertisement and reception of VPN routes between sites. They work independently and can be configured with multiple values to support flexible VPN access control and implement multiple types of VPN networking schemes.

Basic VPN networking scheme

In the simplest case, all users in a VPN form a closed user group. They can forward traffic to each other but cannot communicate with any user outside the VPN.

For the basic VPN networking scheme, you must assign a route target to each VPN for identifying the export target attribute and import target attribute of the VPN. Moreover, this route target cannot be used by any other VPNs.

Figure 4 Network diagram for basic VPN networking scheme



As shown in Figure 4, the route target for VPN 1 is 100:1, while that for VPN 2 is 200:1. The two VPN 1 sites can communicate with each other, and the two VPN 2 sites can communicate with each other. However, the VPN 1 sites cannot communicate with the VPN 2 sites.

Hub and spoke networking scheme

The hub and spoke networking scheme is suitable for a VPN where all users must communicate with each other through an access control device.

In a hub and spoke network as shown in Figure 5, configure route targets as follows:

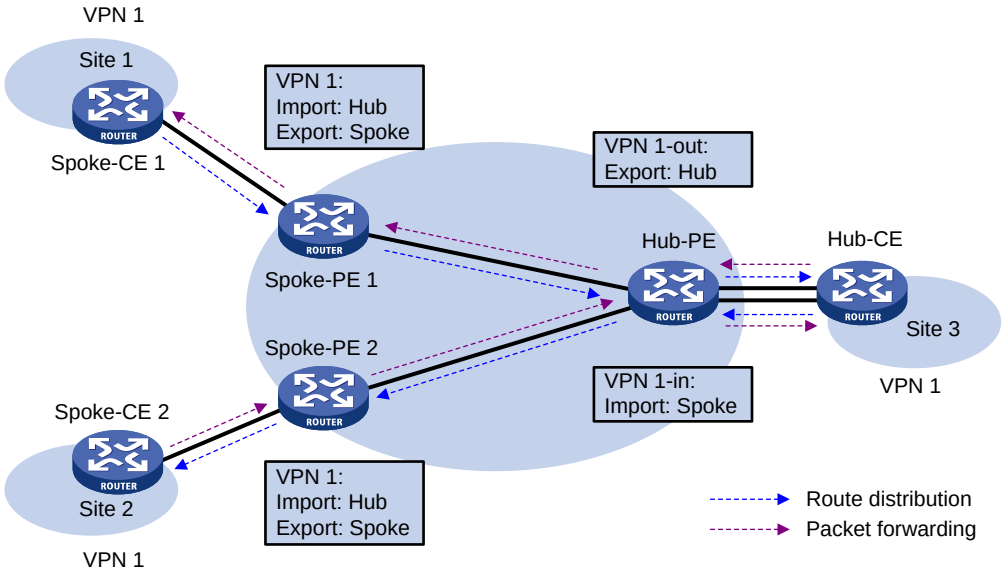
- On spoke PEs (PEs connected to spoke sites), set the export target to Spoke and the import target to Hub.
- On the hub PE (PE connected to the hub site), use two interfaces that each belong to a different VPN instance to connect the hub CE. One VPN instance receives routes from spoke PEs and

has the import target set to Spoke. The other VPN instance advertises routes to spoke PEs and has the export target set to Hub.

These route targets rules produce the following results:

- The hub PE can receive all VPN-IPv4 routes from spoke PEs.
- All spoke PEs can receive VPN-IPv4 routes advertised by the hub PE.
- The hub PE advertises the routes learned from a spoke PE to the other spoke PEs so the spoke sites can communicate with each other through the hub site.
- The import target attribute of a spoke PE is different from the export target attribute of any other spoke PE. Any two spoke PEs do not directly advertise VPN-IPv4 routes to each other. Therefore, they cannot directly access each other.

Figure 5 Network diagram for hub and spoke network



A route in Site 1 is advertised to Site 2 by using the following process:

Table 1 Spoke-CE 1 advertises a route in Site 1 to Spoke-PE 1.

2. Spoke-PE 1 changes the route to a VPN-IPv4 route and advertises the VPN-IPv4 route to Hub-PE through MP-BGP.
3. Hub-PE adds the VPN-IPv4 route into the routing table of VPN 1-in, changes it to the original IPv4 route, and advertises the IPv4 route to Hub-CE.
4. Hub-CE advertises the IPv4 route back to Hub-PE.
5. Hub-PE adds the IPv4 route to the routing table of VPN 1-out, changes it to a VPN-IPv4 route, and advertises the VPN-IPv4 route to Spoke-PE 2 through MP-BGP.
6. Spoke-PE 2 changes the VPN-IPv4 route to the original IPv4 route, and advertises the IPv4 route to Site 2.

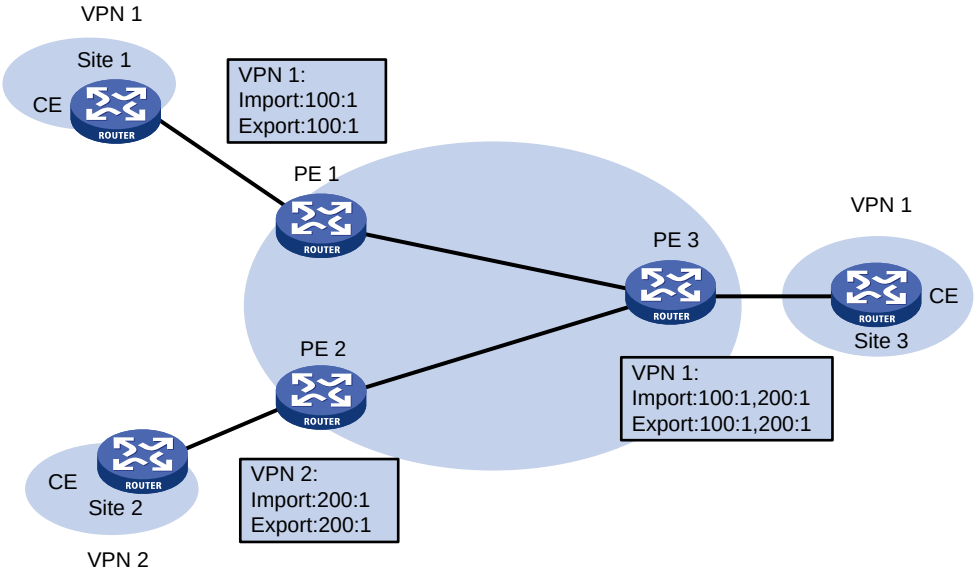
After spoke sites exchange routes through the hub site, they can communicate with each other through the hub site.

Extranet networking scheme

The extranet networking scheme allows specific resources in a VPN to be accessed by users not in the VPN.

In this networking scheme, if a VPN instance needs to access a shared site, the export target attribute and the import target attribute of the VPN instance must be contained in the import target attribute and the export target attribute of the VPN instance of the shared site, respectively.

Figure 6 Network diagram for extranet networking scheme



As shown in Figure 6, route targets configured on PEs produce the following results:

- PE 3 can receive VPN-IPv4 routes from PE 1 and PE 2.
- PE 1 and PE 2 can receive VPN-IPv4 routes advertised by PE 3.
- Site 1 and Site 3 of VPN 1 can communicate with each other, and Site 2 of VPN 2 and Site 3 of VPN 1 can communicate with each other.
- PE 3 advertises neither the VPN-IPv4 routes received from PE 1 to PE 2 nor the VPN-IPv4 routes received from PE 2 to PE 1 (routes learned from an IBGP neighbor are not advertised to any other IBGP neighbor). Therefore, Site 1 of VPN 1 and Site 2 of VPN 2 cannot communicate with each other.

Inter-AS VPN

In an inter-AS VPN networking scenario, multiple sites of a VPN are connected to multiple ISPs in different ASs, or to multiple ASs of an ISP.

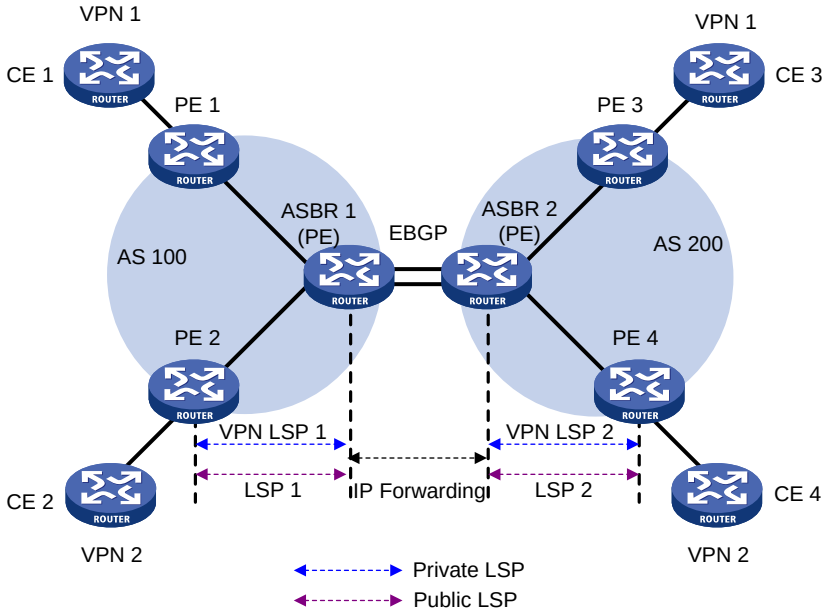
Inter AS-VPN provides the following solutions:

- **VRF-to-VRF connections between ASBRs**—This solution is also called inter-AS option A.
- **EBGP redistribution of labeled VPN-IPv4 routes between ASBRs**—ASBRs advertise VPN-IPv4 routes to each other through MP-EBGP. This solution is also called inter-AS option B.
- **Multihop EBGP redistribution of labeled VPN-IPv4 routes between PE routers**—PEs advertise VPN-IPv4 routes to each other through MP-EBGP. This solution is also called inter-AS option C.

Inter-AS option A

In this solution, PEs of two ASs are directly connected, and each PE is also the ASBR of its AS. Each PE treats the other as a CE and advertises unlabeled IPv4 unicast routes through EBGP. The PEs associate a VPN instance with a minimum of one interface.

Figure 7 Network diagram for inter-AS option A



As shown in Figure 7, in VPN 1, routes are advertised from CE 1 to CE 3 by using the following process:

Table 2 PE 1 advertises the VPN routes learned from CE 1 to ASBR 1 through MP-IBGP.

2. ASBR 1 performs the following operations:
   a. Adds the routes to the routing table of the VPN instance whose import target attribute matches the export target attribute of the routes.
   b. Advertises the routes as IPv4 unicast routes to its CE (ASBR 2) through EBGP.
3. ASBR 2 adds the IPv4 unicast routes to the routing table of the VPN instance that is bound to the receiving interface, and advertises the routes to PE 3 through MP-IBGP.
4. PE 3 advertises the received routes to CE 3.

Packets forwarded within an AS are VPN packets that carry two labels. Packets forwarded between ASBRs are common IP packets.

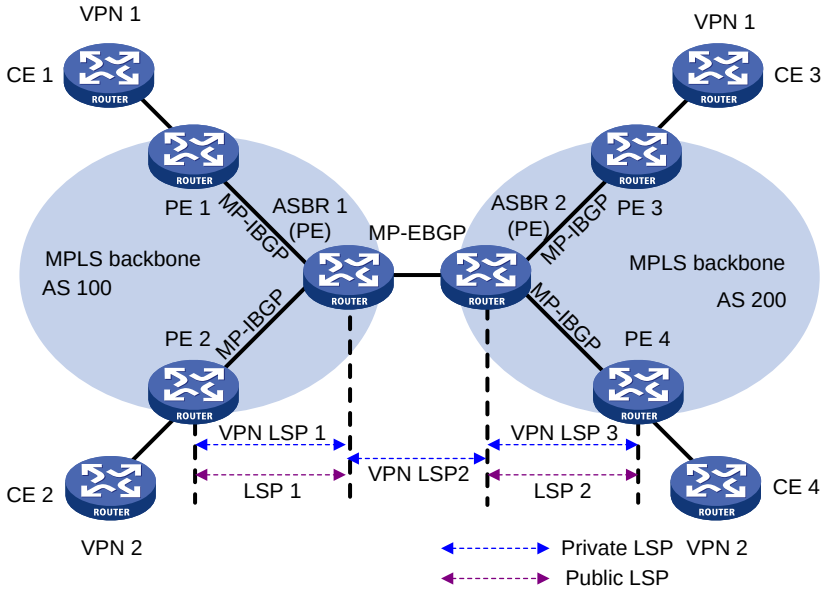
Inter-AS option A is easy to carry out because no special configuration is required on the PEs acting as the ASBRs.

However, it has limited scalability because the PEs acting as the ASBRs must manage all the VPN routes and create VPN instances on a per-VPN basis. This leads to excessive VPN-IPv4 routes on the PEs. Associating a separate interface with each VPN also requires additional system resources.

Inter-AS option B

In this solution, two ASBRs use MP-EBGP to exchange VPN-IPv4 routes that they obtain from the PEs in their respective ASs.

Figure 8 Network diagram for inter-AS option B



As shown in Figure 8, in VPN 1, routes are advertised from CE 1 to CE 3 by using the following process:

Table 3 PE 1 advertises the VPN routes learned from CE 1 to ASBR 1 through MP-IBGP.

Assume that the inner label assigned by PE 1 for the routes is L1.

2. ASBR 1 advertises the VPN-IPv4 routes to ASBR 2 through MP-EBGP.

   Before advertising the routes, ASBR 1 modifies the next hop as its own address, assigns a new inner label (L2) to the routes, and associates L1 with L2.

3. ASBR 2 advertises the VPN-IPv4 routes to PE 3 through MP-IBGP.

   Before advertising the routes, ASBR 2 modifies the next hop as its own address, assigns a new inner label (L3) to the routes, and associates L2 with L3.

4. PE 3 advertises the received routes to CE 3.

A packet is forwarded from CE 3 to CE 1 by using the following process:

Table 4 PE 3 encapsulates the received packet with two labels, and forwards the encapsulated packet to ASBR 2.

One of the labels is L3, and the other is the outer tag for the public tunnel from PE 3 to ASBR 2.

5. ASBR 2 removes the outer tag, replaces L3 with L2, and forwards the packet to ASBR 1.

   Packets between ASBR 1 and ASBR 2 carry only one inner label.

6. ASBR 1 replaces L2 with L1, adds the outer tag of the public tunnel from ASBR 1 to PE 1, and forwards the packet to PE 1.

7. PE 1 removes the inner label and outer tag and forwards the packet to CE 1.

In this solution, ASBRs must receive all inter-AS VPN routes. Therefore, ASBRs cannot filter incoming VPN-IPv4 routes by route targets.

Inter-AS option B has better scalability than option A. However, it requires that ASBRs maintain and advertise VPN routes.

Inter-AS option C

The Inter-AS option A and option B solutions require that the ASBRs maintain and advertise VPN-IPv4 routes. When every AS needs to exchange a great amount of VPN routes, the ASBRs

might become bottlenecks, which hinders network extension. Inter-AS option C has better scalability because it makes PEs directly exchange VPN-IPv4 routes.

In this solution, PEs exchange VPN-IPv4 routes over a multihop MP-EBGP session. Each PE must have a route to the peer PE and a label for the route so that the inter-AS public tunnel between the PEs can be set up. Inter-AS option C sets up a public tunnel by using the following methods:

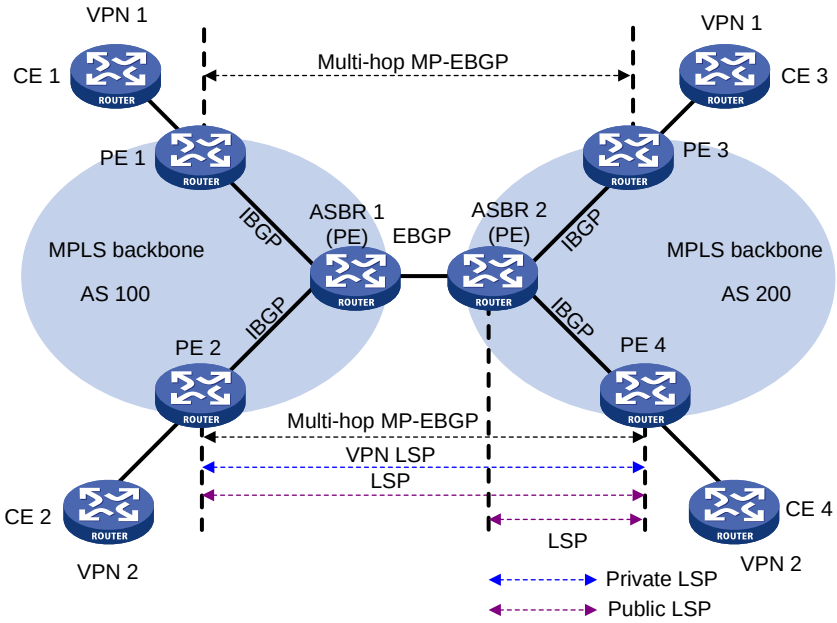
- Method 1:
  - **a.** The PE and the ASBR within an AS establish a public tunnel by using a label distribution protocol such as LDP.
  - **b.** The local and remote ASBRs advertise labeled IPv4 unicast routes through BGP to establish an inter-AS public tunnel.

    Labeled IPv4 unicast route advertisement refers to the process of assigning MPLS labels to IPv4 unicast routes and advertising IPv4 unicast routes and their labels.
- Method 2:

  In method 2, the PE and ASBR within an AS do not need to establish an IBGP peer relationship.
  - **c.** The local ASBR redistributes IGP routes to the BGP routing table, assigns labels to the routes, and advertises the labeled routes to the remote ASBR.
  - **d.** The remote ASBR redistributes BGP routes to the IGP routing table.
  - **e.** The local and remote PEs then can learn the route s to reach each other. After the PEs learn the routes to each other, they establish a public tunnel by using a label distribution protocol such as LDP.

The following is an example of configuring inter-AS option C by using method 1.

Figure 9 Network diagram for inter-AS option C (method 1)



As shown in Figure 9, in VPN 1, routes are advertised from CE 1 to CE 3 by using the following process:

2. PE 1 advertises the VPN routes learned from CE 1 as VPN-IPv4 routes to PE 3 through multihop MP-EBGP.

   Assume that the inner label assigned by PE 1 for the routes is Lx.
3. PE 3 advertises the received routes to CE 3.

Setting up an inter-AS public tunnel is difficult in this solution. A public tunnel, for example, the one from PE 3 to PE 1, is set up by using the following process:

Table 5 Within AS 100, the public tunnel from ASBR 1 to PE 1 is set up by using a label distribution protocol, for example, LDP.

Assume that the outgoing label for the public tunnel on ASBR 1 is L1.

4. ASBR 1 advertises labeled IPv4 unicast routes to ASBR 2 through EBGP.

   The route destined for PE 1 and the label (L2) assigned by ASBR 1 for the route are advertised from ASBR 1 to ASBR 2. The next hop of the route is ASBR 1. The public tunnel from ASBR 2 to ASBR 1 is set up. The incoming label for the public tunnel on ASBR 1 is L2.

5. ASBR 2 advertises labeled IPv4 unicast routes to PE 3 through IBGP.

   The route destined for PE 1 and the label (L3) assigned by ASBR 2 for the route are advertised from ASBR 2 to PE 3. The next hop for the route is ASBR 2. The public tunnel from PE 3 to ASBR 2 is set up. The incoming label for the public tunnel on ASBR 2 is L3, and the outgoing label is L2.

6. MPLS packets cannot be forwarded directly from PE 3 to ASBR 2. Within AS 200, the public tunnel from PE 3 to ASBR 2 is required to be set up hop by hop through a label distribution protocol, for example, LDP.

   Assume that the outgoing label for the public tunnel on PE 3 is Lv.

After route advertisement and public tunnel setup, a packet is forwarded from CE 3 to CE 1 by using the following process:

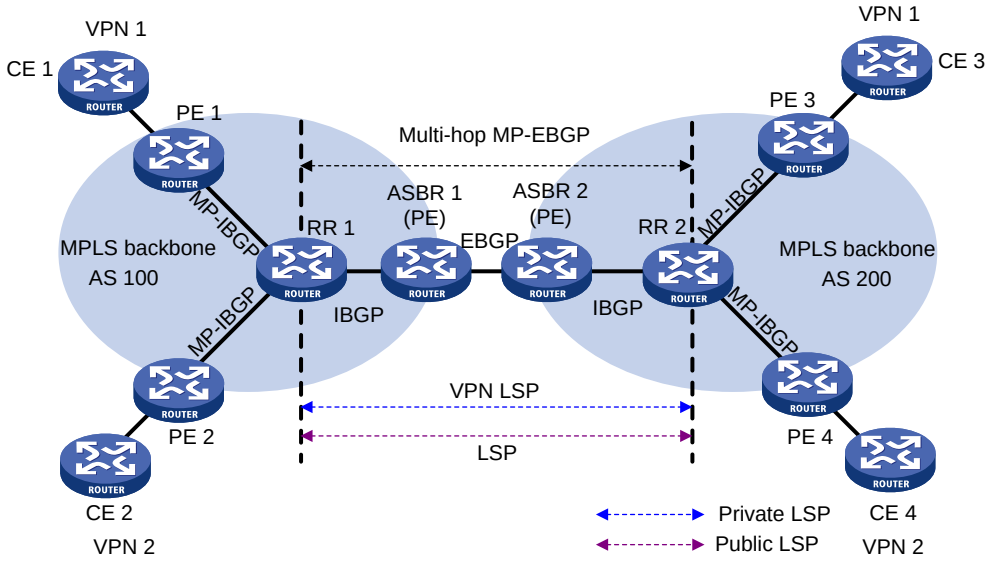
Table 6 PE 3 performs the following routing table lookups for the packet:

   a. Finds a matching route with next hop PE 1 and inner label Lx, and encapsulates the packet with label Lx.
   b. Finds the route to PE 1 with next hop ASBR 2 and label L3, and encapsulates the packet with label L3 as the outer label.
   c. Finds the route to ASBR 2 with outgoing label Lv, and encapsulates the packet with label Lv as the outmost label.

7. AS 200 transmits the packet to ASBR 2 by the outmost label.
8. ASBR 2 removes the outmost label, replaces L3 with L2, and forwards the packet to ASBR 1.
9. ASBR 1 replaces L2 with L1, and forwards the packet.
10. AS 100 transmits the packet to PE 1 by the outer label.
11. PE 1 removes the outer label, and forwards the packet to CE 1 according to the inner label Lx.

As shown in Figure 10, to improve scalability, you can specify a route reflector (RR) in each AS to exchange VPN-IPv4 routes with PEs in the same AS. The RR in each AS maintains all VPN-IPv4 routes. The RRs in two ASs establish a multihop MP-EBGP session to advertise VPN-IPv4 routes.

Figure 10 Network diagram for inter-AS option C using RRs



Carrier's carrier

If a customer of an MPLS L3VPN service provider is also a service provider:

- The MPLS L3VPN service provider is called the provider carrier or the Level 1 carrier.
- The customer is called the customer carrier or the Level 2 carrier.

This networking model is referred to as carrier's carrier.

The PEs of the Level 2 carrier directly exchange customer networks over a BGP session. The Level 1 carrier only learns the backbone networks of the Level 2 carrier, without learning customer networks.

For packets between customer networks to travel through the Level 1 carrier, the PE of the Level 1 carrier and the CE of the Level 2 carrier must assign labels to the backbone networks of the Level 2 carrier. The CE of the Level 2 carrier is a PE within the Level 2 carrier network.

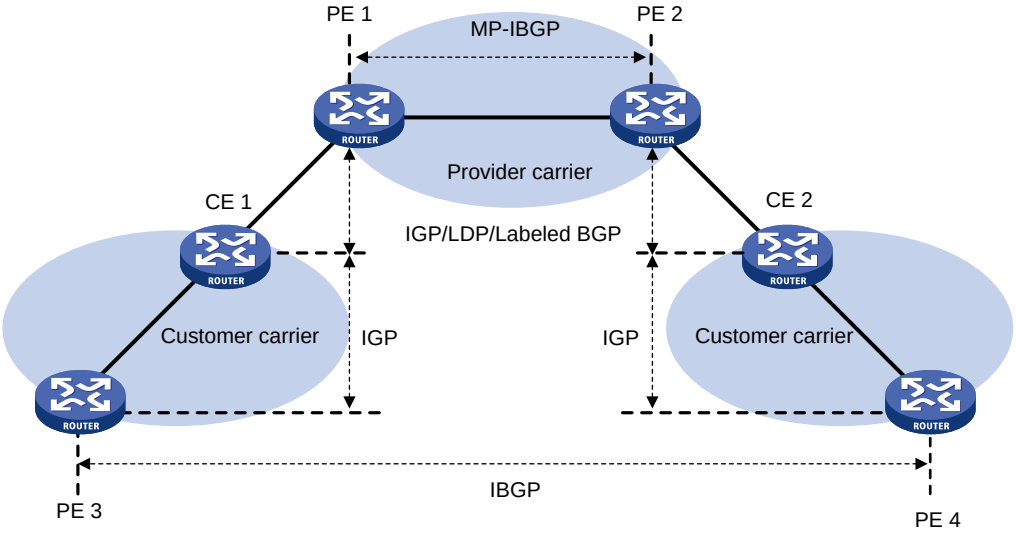
Follow these guidelines to assign labels:

- If the PE and the CE are in the same AS, you must configure IGP and LDP between them. If they are in different ASs, you must configure MP-EBGP to assign labels to IPv4 unicast routes exchanged between them.
- You must enable MPLS on the CE of the Level 2 carrier regardless of whether the PE and CE are in the same AS.

A Level 2 carrier can be an ordinary ISP or an MPLS L3VPN service provider.

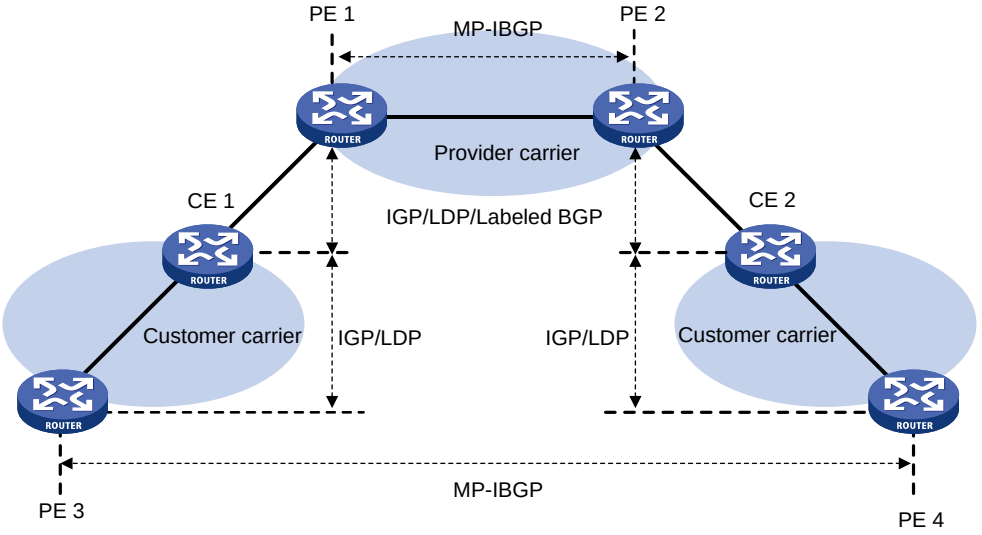
As shown in Figure 11, when the customer carrier is an ordinary ISP, its PEs and CEs run IGP to communicate with each other. The PEs do not need to run MPLS. PE 3 and PE 4 exchange customer network routes (IPv4 unicast routes) through an IBGP session.

Figure 11 Scenario where the Level 2 carrier is an ISP



As shown in Figure 12, when the customer carrier is an MPLS L3VPN service provider, its PEs and CEs must run IGP and LDP to communicate with each other. PE 3 and PE 4 exchange customer network routes (VPN-IPv4 routes) through an MP-IBGP session.

Figure 12 Scenario where the Level 2 carrier is an MPLS L3VPN service provider



NOTE:

As a best practice, establish equal cost LSPs between the Level 1 carrier and the Level 2 carrier if equal cost routes exist between them.

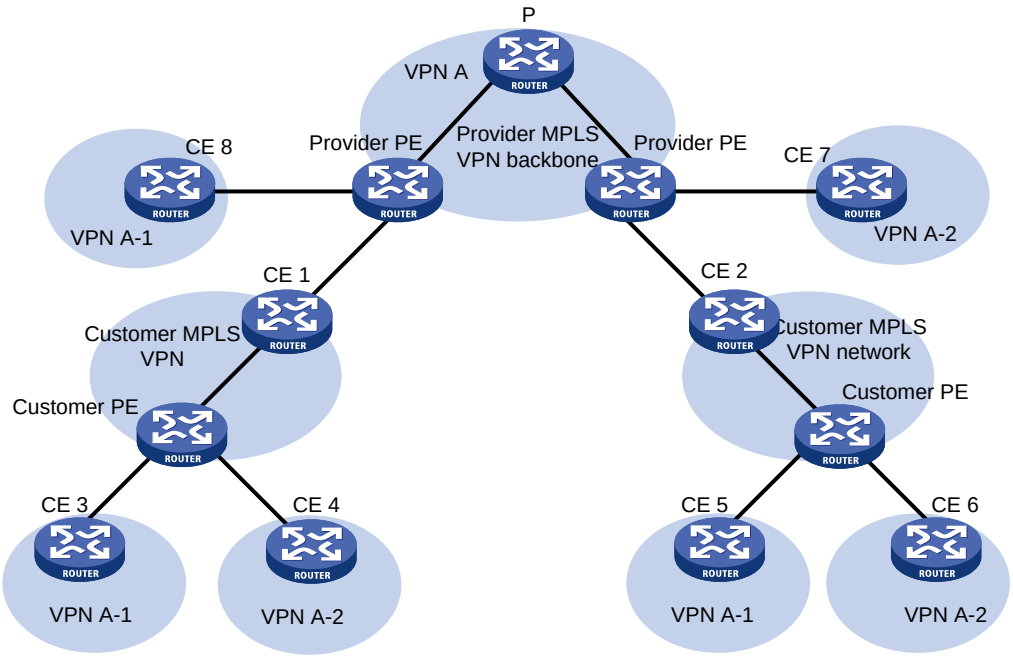
Nested VPN

The nested VPN technology exchanges VPNv4 routes between PEs and CEs of the ISP MPLS L3VPN and allows a customer to manage its own internal VPNs. Figure 13 shows a nested VPN network. On the service provider's MPLS VPN network, there is a customer VPN named VPN A. The customer VPN contains two sub-VPNs, VPN A-1 and VPN A-2.

The service provider PEs consider the customer's network as a common VPN user and do not join any sub-VPNs. The service provider CE devices (CE 1 and CE 2) exchange VPNv4 routes including sub-VPN routing information with the service provider PEs, which implements the propagation of the sub-VPN routing information throughout the customer network.

The nested VPN technology supports both symmetric networking and asymmetric networking. Sites of the same VPN can have the same number or different numbers of internal VPNs. Nested VPN also supports multiple-level nesting of internal VPNs.

Figure 13 Network diagram for nested VPN



In a nested VPN network, routing information is propagated by using the following process:

Table 7 After receiving VPN routes from customer CEs, a customer PE advertises VPN-IPv4 routes to the provider CEs through MP-BGP.

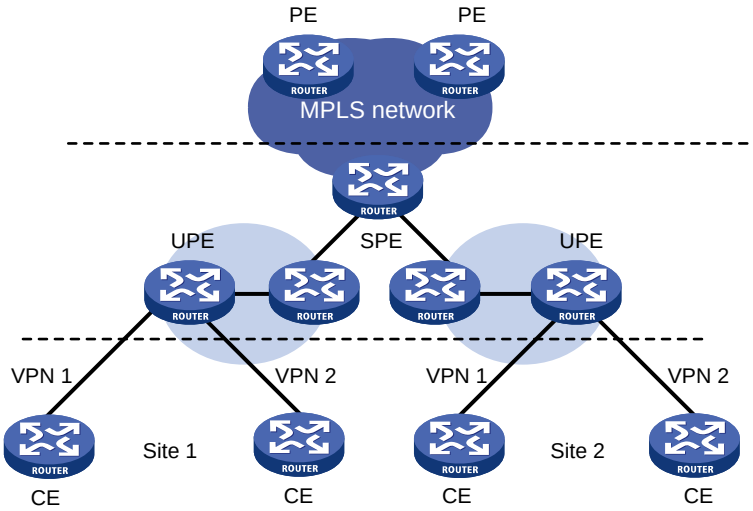
2. The provider CEs advertise the VPN-IPv4 routes to a provider PE through MP-BGP.
3. After receiving a VPN-IPv4 route, the provider PE keeps the customer's internal VPN information, and appends the customer's MPLS VPN attributes on the service provider network. It replaces the RD of the VPN-IPv4 route with the RD of the customer's MPLS VPN on the service provider network. It also adds the export route-target (ERT) attribute of the customer's MPLS VPN on the service provider network to the extended community attribute list of the route. The internal VPN information for the customer is maintained on the provider PE.
4. The provider PE advertises VPN-IPv4 routes carrying the comprehensive VPN information to the other PEs of the service provider.
5. After another provider PE receives the VPN-IPv4 routes, it matches the VPN-IPv4 routes to the import targets of its local VPNs. Each local VPN accepts routes of its own and advertises them to provider CEs. If a provider CE (such as CE 7 and CE 8 in Figure 13) is connected to a provider PE through an IPv4 connection, the PE advertises IPv4 routes to the CE. If it is a VPN-IPv4 connection (a customer MPLS VPN network), the PE advertises VPN-IPv4 routes to the CE.
6. After receiving VPN-IPv4 routes from the provider CE, a customer PE matches those routes to local import targets. Each customer VPN accepts only its own routes and advertises them to connected customer CEs (such as CE 3, CE 4, CE 5, and CE 6 in Figure 13).

HoVPN

Hierarchy of VPN (HoVPN), also called Hierarchy of PE (HoPE), prevents PEs from being bottlenecks and is applicable to large-scale VPN deployment.

HoVPN divides PEs into underlayer PEs (UPEs) or user-end PEs, and superstratum PEs (SPEs) or service provider-end PEs. UPEs and SPEs have different functions and comprise a hierarchical PE. The HoPE and common PEs can coexist in an MPLS network.

Figure 14 Basic architecture of HoVPN



As shown in Figure 14, UPEs and SPEs play the following different roles:

- A UPE is directly connected to CEs. It provides user access. It maintains the routes of directly connected VPN sites. It does not maintain the routes of the remote sites in the VPN, or it only maintains their summary routes. A UPE assigns inner labels to the routes of its directly connected sites, and advertises the labels along with VPN routes to the SPE through MP-BGP. A UPE features high access capability, small routing table capacity, and low forwarding performance.
- An SPE is connected to UPEs and resides inside the service provider network. It manages and advertises VPN routes. It maintains all the routes of the VPNs connected through UPEs, including the routes of both the local and remote sites. An SPE advertises routes along with labels to UPEs, including the default routes of VPN instances or summary routes and the routes permitted by the routing policy. By using routing policies, you can control which sites in a VPN can communicate with each other. An SPE features large routing table capacity, high forwarding performance, and fewer interface resources.

Either MP-IBGP or MP-EBGP can run between SPE and UPE. When MP-IBGP runs between SPE and UPEs, the SPE acts as the RR of multiple UPEs and reflects routes between UPEs.

HoVPN supports HoPE recursion:

- An HoPE can act as a UPE to form a new HoPE with an SPE.
- An HoPE can act as an SPE to form a new HoPE with multiple UPEs.

HoVPN supports multilevel recursion. In HoPE recursion, the concepts of SPE and UPE are relative. A PE might be the SPE of its underlayer PEs and a UPE of its SPE at the same time.

Figure 15 Recursion of HoPEs

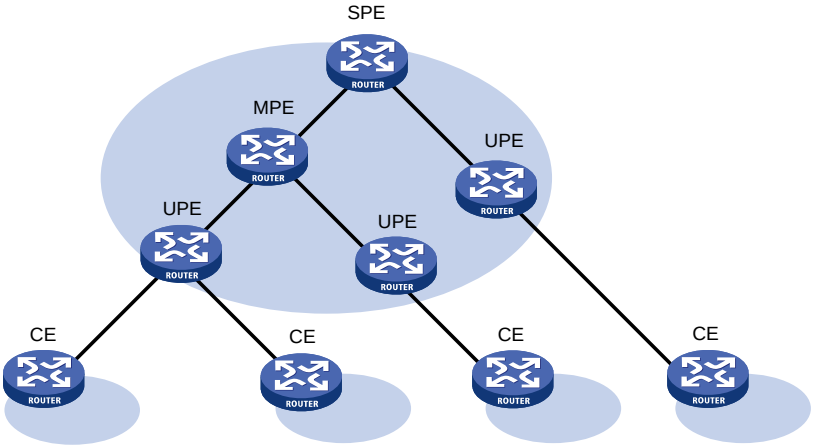


Figure 15 shows a three-level HoPE. The PE in the middle is called the middle-level PE (MPE). MP-BGP runs between SPE and MPE, and between MPE and UPE.

MP-BGP advertises the following routes:

- All the VPN routes of UPEs to the SPEs.
- The default routes of the VPN instance of the SPEs or the VPN routes permitted by the routing policies to the UPEs.

The SPE maintains the VPN routes of all sites in the HoVPN. Each UPE maintains only VPN routes of its directly connected sites. An MPE has fewer routes than the SPE but has more routes than a UPE.

OSPF VPN extension

This section describes the OSPF VPN extension. For more information about OSPF, see *Layer 3—IP Routing Configuration Guide*.

OSPF for VPNs on a PE

If OSPF runs between a CE and a PE to exchange VPN routes, the PE must support multiple OSPF instances to create independent routing tables for VPN instances. Each OSPF process is bound to a VPN instance. Routes learned by an OSPF process are added into the routing table of the bound VPN instance.

OSPF area configuration between a PE and a CE

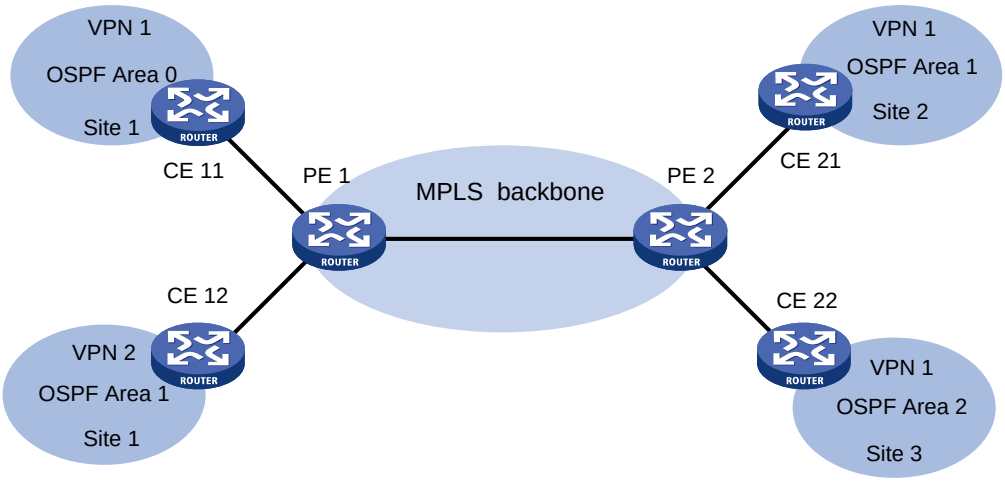
The OSPF area between a PE and a CE can be either a non-backbone area or a backbone area.

In the OSPF VPN extension application, the MPLS VPN backbone is considered the backbone area (area 0). The area 0 of each site must be connected to the MPLS VPN backbone (physically connected or logically connected through a virtual link) because OSPF requires that the backbone area be contiguous.

BGP/OSPF interaction

If OSPF runs between PEs and CEs, each PE redistributes BGP routes to OSPF and advertises the routes to CEs through OSPF. OSPF considers the routes redistributed from BGP as external routes but the OSPF routes actually belong to the same OSPF domain. This problem can be resolved by configuring the same domain ID for sites in an OSPF domain.

Figure 16 Network diagram for BGP/OSPF interaction



As shown in Figure 16, CE 11, CE 21, and CE 22 belong to the same VPN and the same OSPF domain.

Before domain ID configuration, VPN 1 routes are advertised from CE 11 to CE 21 and CE 22 by using the following process:

Table 8 PE 1 redistributes OSPF routes from CE 11 into BGP, and advertises the VPN routes to PE 2 through BGP.

2. PE 2 redistributes the BGP routes to OSPF, and advertises them to CE 21 and CE 22 in AS External LSAs (Type 5) or NSSA External LSAs (Type 7).

After domain ID configuration, VPN 1 routes are advertised from CE 11 to CE 21 and CE 22 by using the following process:

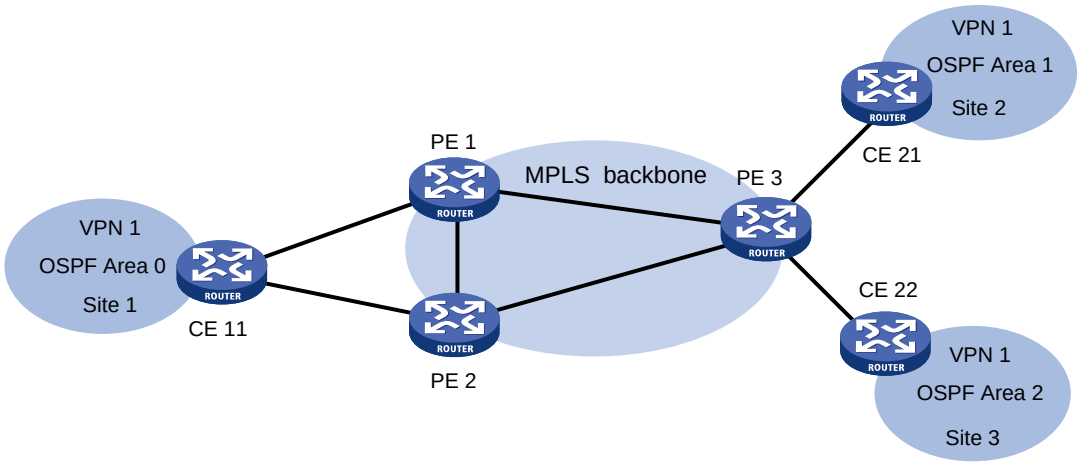
Table 9 PE 1 redistributes OSPF routes into BGP, adds the domain ID to the redistributed BGP VPNv4 routes as a BGP extended community attribute, and advertises the routes to PE 2.

3. PE 2 compares the domain ID in the received routes with the locally configured domain ID. If they are the same and the received routes are intra-area or inter-area routes, OSPF advertises these routes in Network Summary LSAs (Type 3). Otherwise, OSPF advertises these routes in AS External LSAs (Type 5) or NSSA External LSAs (Type 7).

Routing loop avoidance

As shown in Figure 17, Site 1 is connected to two PEs. When a PE advertises VPN routes learned from MP-BGP to Site 1 through OSPF, the routes might be received by the other PE. This results in a routing loop.

Figure 17 Network diagram for routing loop avoidance



OSPF VPN extension uses the following tags to avoid routing loops:

- Down bit, or DN bit. An OSPF process for a VPN instance uses the DN bit to avoid routing loops as follows:

  When a PE redistributes BGP routes into OSPF and creates Type 3 LSAs, Type 5 LSAs, or Type 7 LSAs, it sets the DN bit for the LSAs. When receiving the LSAs advertised by CE 11, the other PE ignores the LSAs whose DN bit is set to avoid routing loops.

- VPN route tag, the external route tag in redistributed routes of a VPN instance. An OSPF process for a VPN instance uses the route tag to avoid routing loops as follows:

  The two PEs connected to the same site use the same route tag. When a PE redistributes BGP routes into OSPF and creates Type 5 or 7 LSAs, it adds the route tag to the LSAs. When receiving the Type 5 or 7 LSAs advertised by CE 11, the other PE compares the route tag in the LSAs against the local route tag. If they are the same, the PE ignores the LSAs to avoid routing loops.

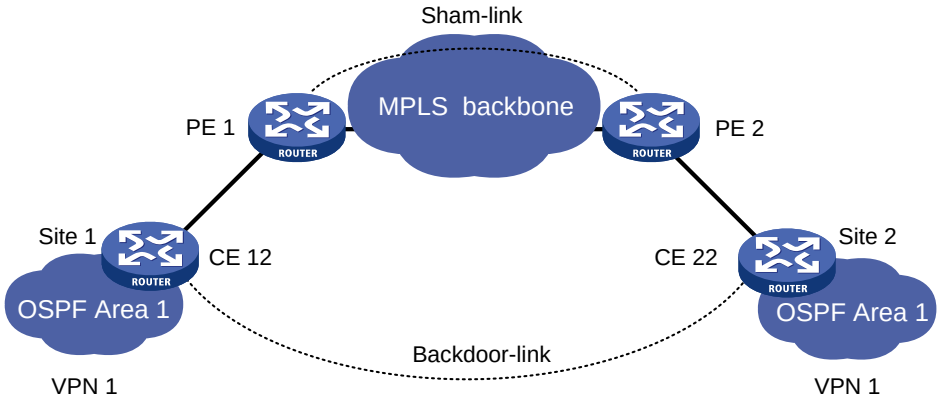
OSPF sham link

As shown in Figure 18, two routes exist between Site 1 and Site 2 of VPN 1:

- **A route over MPLS backbone**—It is an inter-area route if PE 1 and PE 2 have the same domain ID, or is an external route if PE 1 and PE 2 are configured with no domain ID or with different domain IDs.
- **A direct route between CEs**—It is an intra-area route that is called a backdoor link.

VPN traffic is always forwarded through the backdoor link because it has a higher priority than the inter-area route. To forward VPN traffic over the inter-area route, you can establish a sham link between the two PEs to change the inter-area route to an intra-area route.

Figure 18 Network diagram for sham link



A sham link is considered a virtual point-to-point link within a VPN and is advertised in a Type 1 LSA. It is identified by the source IP address and destination IP address that are the local PE address and the remote PE address in the VPN address space. Typically, the source and destination addresses are loopback interface addresses with a 32-bit mask.

To add a route to the destination IP address of a sham link to a VPN instance, the remote PE must advertise the source IP address of the sham link as a VPN-IPv4 address through MP-BGP. To avoid routing loops, a PE does not advertise the sham link's destination address.

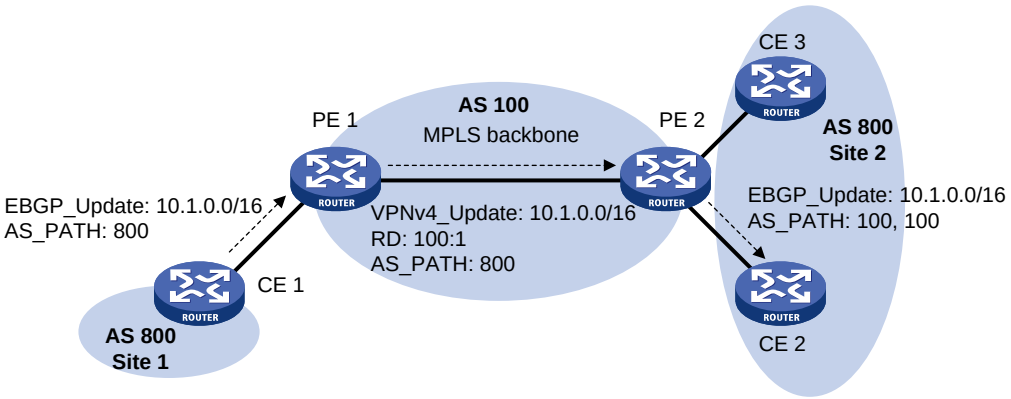
BGP AS number substitution and SoO attribute

BGP detects routing loops by examining AS numbers. If EBGP runs between PE and CE, you must assign different AS numbers to geographically different sites or configure the BGP AS number substitution feature to ensure correct transmission of routing information.

The BGP AS number substitution feature allows geographically different CEs to use the same AS number. If the AS_PATH of a route contains the AS number of a CE, the PE replaces the AS number with its own AS number before advertising the route to that CE.

After you enable the BGP AS number substitution feature, the PE performs BGP AS number substitution for all routes and re-advertises them to connected CEs in the peer group.

Figure 19 Application of BGP AS number substitution and SoO attribute



As shown in Figure 19, both Site 1 and Site 2 use the AS number 800. AS number substitution is enabled on PE 2 for CE 2. Before advertising updates received from CE 1 to CE 2, PE 2 substitutes its own AS number 100 for the AS number 800. In this way, CE 2 can correctly receive the routing information from CE 1.

However, the AS number substitution feature also introduces a routing loop in Site 2 because route updates originated from CE 3 can be advertised back to Site 2 through PE 2 and CE 2. To remove the routing loop, you can configure the same SoO attribute on PE 2 for CE 2 and CE 3. PE 2 adds the SoO attribute to route updates received from CE 2 or CE 3, and checks the SoO attribute of route updates to be advertised to CE 2 or CE 3. The SoO attribute of the route updates from CE 3 is the same as the SoO attribute for CE 2, and PE 2 does not advertise route updates to CE 2.

For more information about the SoO attribute, see *Layer 3—IP Routing Configuration Guide*.

MPLS L3VPN FRR

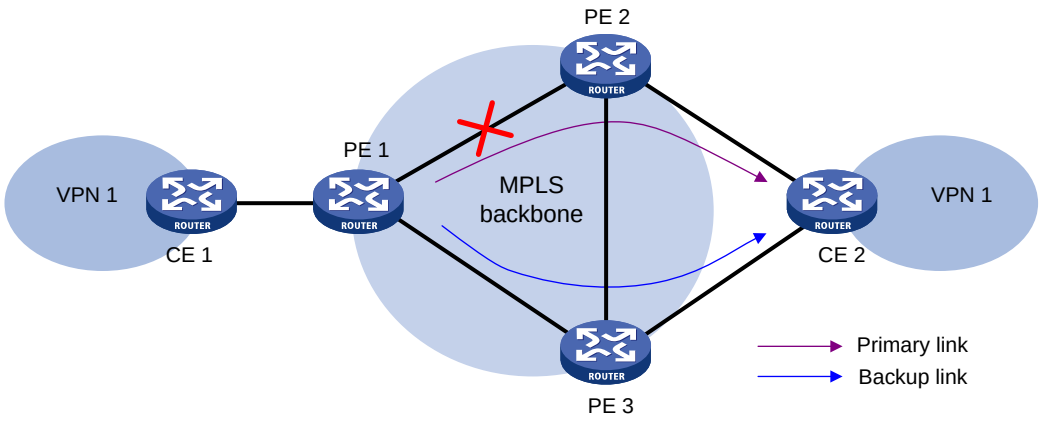
MPLS L3VPN Fast Reroute (FRR) is applicable to a dual-homed scenario. By using BFD to detect the primary link, FRR enables a PE to use the backup link when the primary link fails. The PE then selects a new optimal route, and uses the new optimal route to forward traffic.

MPLS L3VPN FRR supports the following types of backup:

- VPNv4 route backup for a VPNv4 route.
- VPNv4 route backup for an IPv4 route.
- IPv4 route backup for a VPNv4 route.

VPNv4 route backup for a VPNv4 route

Figure 20 Network diagram



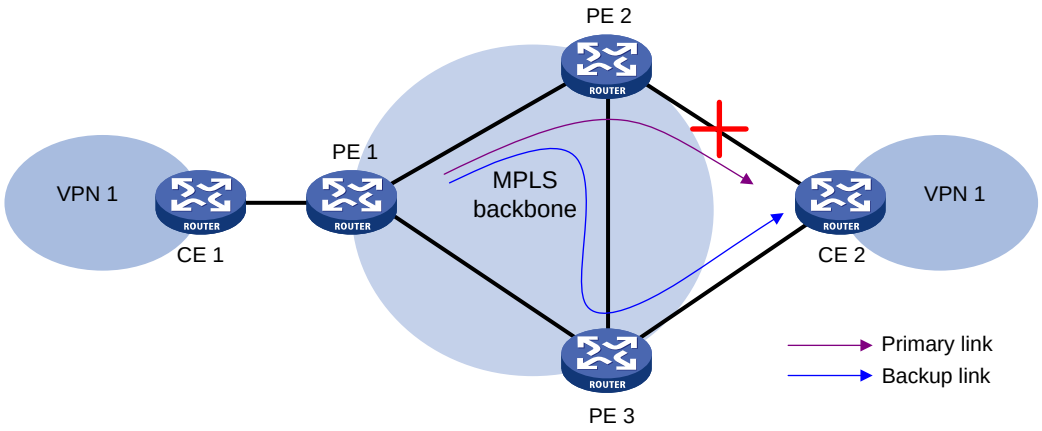
As shown in Figure 20, configure FRR on the ingress node PE 1, and specify the backup next hop for VPN 1 as PE 3. When PE 1 receives a VPNv4 route to CE 2 from both PE 2 and PE 3, it uses the route from PE 2 as the primary link, and the route from PE 3 as the backup link.

When the tunnel PE 1—PE 2 operates correctly, traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—CE 2. When the tunnel fails, the traffic goes through the path CE 1—PE 1—PE 3—CE 2.

In this scenario, PE 1 is responsible for primary link detection and traffic switchover.

VPNv4 route backup for an IPv4 route

Figure 21 Network diagram



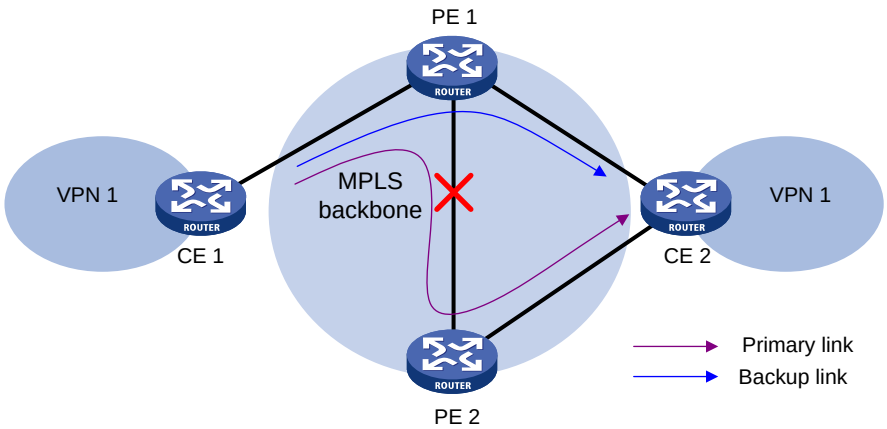
As shown in Figure 21, configure FRR on the egress node PE 2, and specify the backup next hop for VPN 1 as PE 3. When PE 2 receives an IPv4 route from CE 2 and a VPNv4 route from PE 3 (both routes are destined for VPN 1 connected to CE 2), PE 2 uses the IPv4 route as the primary link, and the VPNv4 route as the backup link.

PE 2 uses ARP or echo-mode BFD to detect the connectivity of the link from PE 2 to CE 2. When the link operates correctly, traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—CE 2. When the link fails, PE 2 switches traffic to the link PE 2—PE 3—CE 2, and traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—PE 3—CE 2. This avoids traffic interruption before route convergence completes (switching to the link CE 1—PE 1—PE 3—CE 2).

In this scenario, PE 2 is responsible for primary link detection and traffic switchover.

IPv4 route backup for a VPNv4 route

Figure 22 Network diagram



As shown in Figure 22, configure FRR on PE 1, and specify the backup next hop for VPN 1 as CE 2. When PE 1 receives an IPv4 route from CE 2 and a VPNv4 route from PE 2 (both routes are destined for VPN 1 connected to CE 2), PE 1 uses the VPNv4 route as the primary link, and the IPv4 route as the backup link.

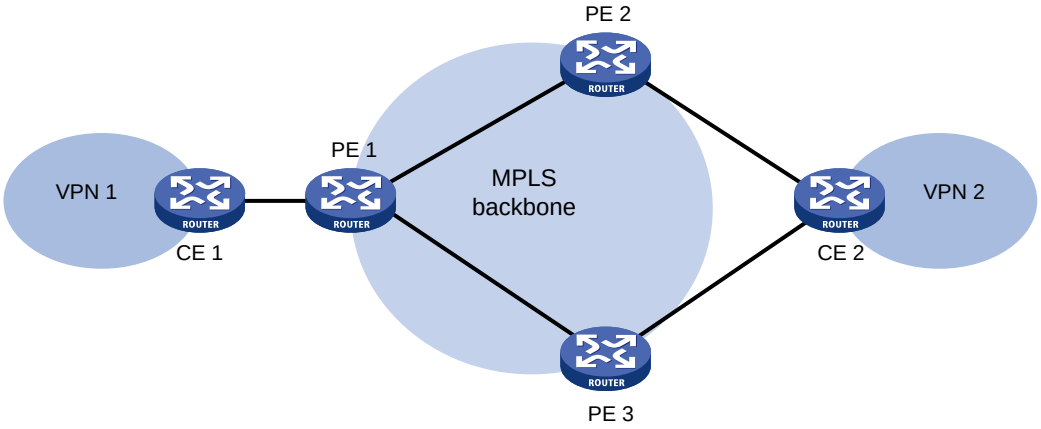
When the tunnel operates correctly, traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—CE 2. When the tunnel fails, the traffic goes through the path CE 1—PE 1—CE 2.

In this scenario, PE 1 is responsible for primary link detection and traffic switchover.

ECMP VPN route redistribution

This feature enables a VPN instance to redistribute all routes that have the same prefix and RD into its routing table. Based on the ECMP routes, the device can perform load sharing (as configured by the **balance** command) or MPLS L3VPN FRR. For more information about the **balance** command, see BGP in *Layer 3—IP Routing Command Reference*.

Figure 23 Network diagram



As shown in Figure 23, CE 1 accesses the backbone network through VPN instance **VPN1** created on PE 1. The RD of VPN instance **VPN1** is 1:1. CE 2 accesses the backbone network through VPN instances created on PE 2 and PE 3. The VPN instances created on PE 2 and PE 3 have the same name **VPN2** and the same RD 1:2. VPN instances **VPN1** and **VPN2** can communicate with each other.

Both PE 2 and PE 3 can advertise routes from CE 2 to PE 1, and the advertised routes have the same RD 1:2. By default, BGP redistributes only the optimal route into the routing table of VPN instance **VPN1**. After you enable this feature on VPN instance **VPN1**, BGP redistributes routes from both PE 2 and PE 3 to the routing table of VPN instance **VPN1**. The two routes can be used for load balancing or MPLS L3VPN FRR.

Protocols and standards

- RFC 3107, *Carrying Label Information in BGP-4*
- RFC 4360, *BGP Extended Communities Attribute*
- RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*
- RFC 4577, *OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)*

MPLS L3VPN tasks at a glance

Unless otherwise indicated, configure MPLS L3VPN on PEs.

To configure MPLS L3VPN, perform the following tasks:

1. Configuring MPLS L3VPN basics
   a. Configuring VPN instances
   b. Configuring routing between a PE and a CE
   c. Configuring routing between PEs
   d. (Optional.) Configuring BGP VPNv4 route control

2. Configuring advanced MPLS L3VPN networks

   Choose the following tasks as needed:

   - Configuring inter-AS VPN

     Perform this task when sites of a VPN are connected to different ASs of an ISP.
   - Configuring nested VPN

     Deploy the nest VPN when a large number of VPNs exist and you want to hide from the outside of the customer internal VPNs.
   - Configuring HoVPN

     HoVPN prevents PEs from being bottlenecks and is applicable to large-scale VPN deployment.
3. (Optional.) Specifying the VPN label processing mode on the egress PE
4. (Optional.) Configuring MPLS L3VPN FRR
5. (Optional.) Controlling route advertisement and reception in an MPLS L3VPN network
   - Configuring an OSPF sham link
   - Configuring BGP AS number substitution and SoO attribute
   - Configuring BGP RT filtering

     Perform this task to reduce the number of routes advertised in an MPLS L3VPN.
   - Configuring the BGP additional path feature

     Perform this task to enable BGP to advertise multiple routes with the same prefix and different next hops to a peer or peer group.
   - Configuring link bandwidth attributes for BGP VPNv4 routes
   - Minimizing the priority of BGP routes sent to peers
   - Specifying a local network to be advertised in the public instance or a VPN instance
   - Configuring BGP VPNv4 routes to use private network next hops
   - Configuring route replication

     Perform this task to enable mutual replication of routes between public and VPN instances or between different VPN instances so that the public and VPN instances or different VPN instances can communicate with each other.
   - Enabling ECMP VPN route redistribution

     Perform this task to enable a VPN instance to redistribute all routes that have the same prefix and RD into its routing table to perform load sharing or MPLS L3VPN FRR.
   - Enabling prioritized withdrawal of specific routes

     Perform this task to configure BGP to send the withdrawal messages of specific routes prior to other routes to achieve fast switchover of traffic and reduce the traffic interruption time.
6. (Optional.) Maintaining MPLS L3VPN networks
   - Enabling SNMP notifications for MPLS L3VPN
   - Enabling logging for BGP route flapping

Prerequisites for MPLS L3VPN

Before you configure basic MPLS L3VPN, perform the following tasks:

1. Configure an IGP on the PEs and P devices to ensure IP connectivity within the MPLS backbone.
2. Configure basic MPLS for the MPLS backbone.
3. Configure MPLS LDP on the PEs and P devices to establish LDP LSPs.

Configuring VPN instances

All VPN instance configurations are performed on PEs.

Creating a VPN instance

About this task

A VPN instance is a collection of the VPN membership and routing rules of its associated site. A VPN instance might correspond to more than one VPN.

Restrctions and guidelines

Deleting the RD for an VPN instance will cause the BGP-VPN instance configuration for the VPN instance to be deleted. When you perform this operation, make sure you fully understand its impact on the service.

Procedure

1. Enter system view.
   system-view
2. Create a VPN instance and enter VPN instance view.
   ip vpn-instance *vpn-instance-name*
3. Configure an RD for the VPN instance.
   route-distinguisher *route-distinguisher*
   By default, no RD is configured for a VPN instance.
4. (Optional.) Configure a description for the VPN instance.
   description *text*
   By default, no description is configured for a VPN instance.
5. (Optional.) Configure a VPN ID for the VPN instance.
   vpn-id *vpn-id*
   By default, no VPN ID is configured for a VPN instance.
6. (Optional.) Configure an SNMP context for the VPN instance.
   snmp context-name *context-name*
   By default, no SNMP context is configured.

Associating a VPN instance with a Layer 3 interface

Table 10 Enter system view.

   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Associate a VPN instance with the interface.
   ip binding vpn-instance *vpn-instance-name*
   By default, an interface is not associated with a VPN instance and belongs to the public network.

CAUTION:
Associating a VPN instance with an interface or disassociating a VPN instance from an

interface will clear the IP address and routing protocol settings of the interface.

The **ip binding vpn-instance** command deletes the IP address of the current interface. You must reconfigure an IP address for the interface after configuring the command.

Configuring route related attributes for a VPN instance

Restrictions and guidelines

IPv4 VPN prefers the configurations in VPN instance IPv4 address family view over the configurations in VPN instance view.

Prerequisites

Before you perform this task, create the routing policies to be used by this task. For information about routing policies, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter VPN instance view or VPN instance IPv4 address family view.
   - Enter VPN instance view.

     ip vpn-instance *vpn-instance-name*
   - Execute the following commands in sequence to enter VPN instance IPv4 address family view:

     ip vpn-instance *vpn-instance-name*

     address-family ipv4

3. Configure route targets.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, no route targets are configured.

4. Set the maximum number of active routes.

   routing-table limit *number* { *warn-threshold* | **simply-alert** }

   By default, the number of active routes in a VPN instance is not limited.

   Setting the maximum number of active routes for a VPN instance can prevent the device from learning too many routes.

5. Apply an import routing policy.

   import route-policy *route-policy*

   By default, all routes matching the import target attribute are accepted.

6. Apply an export routing policy.

   export route-policy *route-policy*

   By default, routes to be advertised are not filtered.

7. Apply a tunnel policy to the VPN instance.

   tnl-policy *tunnel-policy-name*

   By default, only one LSP tunnel is selected (no load balancing).

   If the specified tunnel policy does not exist, the default tunnel policy is used.

   For information about tunnel policies, see "Configuring tunnel policies."

Configuring routing between a PE and a CE

Configuring static routing between a PE and a CE

About this task

Perform this configuration on the PE. On the CE, configure a common static route.

For more information about static routing, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Configure a static route for a VPN instance.

   ip route-static vpn-instance *s-vpn-instance-name dest-address* { *mask-length* | *mask* } { *interface-type interface-number* [*next-hop-address*] | *next-hop-address* [**public**] | **vpn-instance** *d-vpn-instance-name next-hop-address* }

Configuring RIP between a PE and a CE

About this task

Perform this configuration on the PE. On the CE, create a common RIP process.

For more information about RIP, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Create a RIP process for a VPN instance and enter RIP view.

   rip [*process-id*] **vpn-instance** *vpn-instance-name*

   A RIP process can belong to only one VPN instance.

3. Enable RIP on the interface attached to the specified network.

   network *network-address* [*wildcard-mask*]

   By default, RIP is disabled on an interface.

Configuring OSPF between a PE and a CE

About this task

Perform this configuration on the PE. On the CE, create a common OSPF process.

For more information about OSPF, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Create an OSPF process for a VPN instance and enter the OSPF view.

   ospf [*process-id* | **router-id** *router-id*] * **vpn-instance** *vpn-instance-name*

Parameter	Usage guidelines
router-id *router-id*	An OSPF process bound to a VPN instance does not use the public network router ID configured in system view. Therefore, you must specify a router ID when creating a process or configure an IP address for a minimum of one interface in the bound VPN instance.
vpn-instance *vpn-instance-name*	An OSPF process can belong to only one VPN instance. If you delete a VPN instance, all OSPF processes of the VPN instance are also deleted.

3. Redistribute BGP routes.

   import-route bgp [*as-number*] [**allow-ibgp**] [**cost** *cost-value* | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   By default, OSPF does not redistribute routes from other routing protocols.

   If the **vpn-instance-capability simple** command is not configured for the OSPF process, the **allow-ibgp** keyword is optional to redistribute VPNv4 routes learned from MP-IBGP peers. In any other cases, if you do not specify the **allow-ibgp** keyword, the OSPF process does not redistribute VPNv4 routes learned from MP-IBGP peers.

4. (Optional.) Confiure OSPF route attributes:

   a. Set an OSPF domain ID.

      domain-id *domain-id* [**secondary**]

      The default domain ID is 0.

Description	Restrictions and guidelines
The domain ID is carried in the routes of the OSPF process. When redistributing routes from the OSPF process, BGP adds the domain ID as an extended community attribute into BGP route.	An OSPF process can be configured with only one primary domain ID. Domain IDs of different OSPF processes can be the same. All OSPF processes of a VPN must be configured with the same domain ID.

   b. (Optional.) Configure the type codes of OSPF extended community attributes.

      ext-community-type { **domain-id** *type-code1* | **router-id** *type-code2* | **route-type** *type-code3* }

      By default, the type codes for Domain ID, Router ID, and Route Type are 0x0005, 0x0107, 0x0306, respectively.

   c. Configure the external route tag value for VPN redistributed routes on a PE.

      route-tag *tag-value*

      By default, if BGP runs within an MPLS backbone, and the BGP AS number is not greater than 65535, the first two octets of the external route tag are 0xD000 and the last two octets are the local BGP AS number. If the AS number is greater than 65535, the external route tag is 0.

   d. Set the DN bit in OSPF LSAs.

      dn-bit-set { **ase** | **nssa** | **summary** }

      By default, when a PE redistributes BGP routes into OSPF and creates OSPF LSAs, it sets the DN bit for the Network Summary LSA (Type 3 LSA), AS External LSA (Type 5 LSA), and NSSA External LSA (Type 7 LSA).

   e. Configure OSPF to check the DN bit in OSPF LSAs.

      dn-bit-check { **ase** | **nssa** | **summary** }

By default, OSPF on a PE checks the DN bit of Network Summary LSA (Type 3 LSA), AS External LSA (Type 5 LSA), and NSSA External LSA (Type 7 LSA).

f. Enable external route check for OSPF LSAs.

route-tag-check enable

By default, the external route check feature is enabled for OSPF LSAs.

5. Create an OSPF area and enter area view.

area *area-id*

6. Enable OSPF on the interface attached to the specified network in the area.

network *ip-address wildcard-mask*

By default, an interface neither belongs to any area nor runs OSPF.

Configuring IS-IS between a PE and a CE

About this task

Perform this configuration on the PE. On the CE, configure common IS-IS.

For more information about IS-IS, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Create an IS-IS process for a VPN instance and enter IS-IS view.

isis [*process-id*] **vpn-instance** *vpn-instance-name*

An IS-IS process can belong to only one VPN instance.

3. Configure a network entity title for the IS-IS process.

network-entity *net*

By default, no NET is configured.

4. Return to system view.

quit

5. Enter interface view.

interface *interface-type interface-number*

6. Enable the IS-IS process on the interface.

isis enable [*process-id*]

By default, no IS-IS process is enabled on the interface.

Configuring EBGP between a PE and a CE

Configuring the PE

1. Enter system view.

system-view

2. Enable a BGP instance and enter BGP instance view.

bgp *as-number* [**instance** *instance-name*]

By default, BGP is not enabled.

3. Enter BGP-VPN instance view.

ip vpn-instance *vpn-instance-name*

Configuration commands in BGP-VPN instance view are the same as those in BGP instance view. For more information, see BGP in *Layer 3—IP Routing Configuration Guide*.

4. Configure the CE as the VPN EBGP peer.

   peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*

   For more information about this command, see BGP in *Layer 3—IP Routing Configuration Guide*.

5. Create the BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

6. Enable IPv4 unicast route exchange with the specified peer.

   peer { *group-name* | *ip-address* [*mask-length*] } **enable**

   By default, BGP does not exchange IPv4 unicast routes with a peer.

7. Redistribute the routes of the local CE.

   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]

   A PE must redistribute the routes of the local CE into its VPN routing table so it can advertise them to the peer PE.

8. Allow the local AS number to appear in the AS_PATH attribute of a received route, and set the maximum number of repetitions.

   peer { *group-name* | *ip-address* [*mask-length*] } **allow-as-loop** [*number*]

   By default, BGP discards incoming route updates that contain the local AS number.

   Execute this command in a hub-spoke network where EBGP is running between a PE and a CE to enable the PE to receive the route updates from the CE.

Configuring the CE

1. Enter system view.

   system-view

2. Enable a BGP instance and enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is not enabled.

3. Configure the PE as a BGP peer.

   peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*

4. Create the BGP IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

5. Enable IPv4 unicast route exchange with the specified peer or peer group.

   peer { *group-name* | *ip-address* [*mask-length*] } **enable**

   By default, BGP does not exchange IPv4 unicast routes with any peer.

6. Configure route redistribution.

   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]

   A CE must redistribute its routes to the PE so the PE can advertise them to the peer CE.

Configuring IBGP between a PE and a CE

Restrictions and guidelines

Use IBGP between PE and CE only in a basic MPLS L3VPN network. In networks such as Hub&Spoke, Extranet, inter-AS VPN, carrier's carrier, nested VPN, and HoVPN, you cannot use IBGP between PE and CE.

Configuring the PE

1. Enter system view.

   system-view

2. Enable a BGP instance and enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is not enabled.

3. Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*

   Configuration commands in BGP-VPN instance view are the same as those in BGP instance view. For more information, see *Layer 3—IP Routing Configuration Guide*.

4. Configure the CE as the VPN IBGP peer.

   peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*

5. Create the BGP-VPN IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

6. Enable IPv4 unicast route exchange with the specified peer.

   peer { *group-name* | *ip-address* [*mask-length*] } **enable**

   By default, BGP does not exchange IPv4 unicast routes with any peer.

7. Configure the CE as a client of the RR to enable the PE to advertise routes learned from the IBGP peer CE to other IBGP peers.

   peer { *group-name* | *ip-address* [*mask-length*] } **reflect-client**

   By default, no RR or RR client is configured.

   Configuring an RR does not change the next hop of a route. To change the next hop of a route, configure an inbound policy on the receiving side.

8. (Optional.) Enable route reflection between clients.

   reflect between-clients

   By default, route reflection between clients is enabled.

9. (Optional.) Configure the cluster ID for the RR.

   reflector cluster-id { *cluster-id* | *ip-address* }

   By default, the RR uses its own router ID as the cluster ID.

   If multiple RRs exist in a cluster, use this command to configure the same cluster ID for all RRs in the cluster to avoid routing loops.

Configuring the CE

1. Enter system view.

   system-view

2. Enable a BGP instance and enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is not enabled.

3. Configure the PE as an IBGP peer.

peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*

4. Create the BGP IPv4 unicast address family and enter its view.
   address-family ipv4 [**unicast**]
5. Enable IPv4 unicast route exchange with the specified peer.
   peer { *group-name* | *ip-address* [*mask-length*] } **enable**
   By default, BGP does not exchange IPv4 unicast routes with a peer.
6. Configure route redistribution.
   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]
   A CE must redistribute its routes to the PE so the PE can advertise them to the peer CE.

Configuring routing between PEs

1. Enter system view.
   system-view
2. Enable a BGP instance and enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
   By default, BGP is not enabled.
3. Configure the remote PE as a BGP peer.
   peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*
4. (Optional.) Specify the source interface for TCP connections.
   peer { *group-name* | *ip-address* [*mask-length*] } **connect-interface** *interface-type interface-number*
   By default, BGP uses the output interface of the optimal route destined for the peer as the source interface.
5. Create the BGP VPNv4 address family and enter its view.
   address-family vpnv4
6. Enable BGP VPNv4 route exchange with the specified peer.
   peer { *group-name* | *ip-address* [*mask-length*] } **enable**
   By default, BGP does not exchange BGP VPNv4 routes with a peer.

Configuring BGP VPNv4 route control

About BGP VPNv4 route control

BGP VPNv4 route control is configured similarly with BGP route control, except that it is configured in BGP VPNv4 address family view. For more information about BGP route control, see *Layer 3—IP Routing Configuration Guide*.

Controlling BGP VPNv4 route advertisement, reception, and saving

1. Enter system view.
   system-view
2. Enter BGP instance view.

bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Advertise a default VPN route to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **default-route-advertise vpn-instance** *vpn-instance-name*

   By default, no default VPN route is advertised to a peer or peer group.

5. Set the maximum number of routes BGP can receive from a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **route-limit** *prefix-number* [{ **alert-only** | **discard** | **reconnect** *reconnect-time* } | *percentage-value*] *

   By default, the number of routes that BGP can receive from a peer or peer group is not limited.

6. Save all route updates from a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **keep-all-routes**

   By default, BGP does not save route updates from a peer.

Setting a preferred value for received routes

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Set a preferred value for routes received from a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **preferred-value** *value*

   By default, the preferred value for routes received from a peer or peer group is 0.

Configuring BGP VPNv4 route reflection

About this task

To ensure the connectivity of IBGP peers, you must establish full-mesh IBGP connections, which costs massive network and CPU resources.

To reduce IBGP connections in the network, you can configure a router as a route reflector (RR) and configure other routers as its clients. You only need to establish IBGP connections between the RR and its clients to enable the RR to forward routes to the clients.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Configure the device as a route reflector and specify a peer or peer group as its client.

peer { *group-name* | *ipv4-address* [*mask-length*] } **reflect-client**

By default, no route reflector or client is configured.

5. (Optional.) Enable route reflection between clients.

   reflect between-clients

   By default, route reflection between clients is enabled.

6. (Optional.) Configure a cluster ID for the RR.

   reflector cluster-id { *cluster-id* | *ip-address* }

   By default, the RR uses its own router ID as the cluster ID.

7. (Optional.) Configure a filtering policy for reflected routes.

   rr-filter *ext-comm-list-number*

   By default, the RR does not filter reflected routes.

8. (Optional.) Enable the route reflector to change the attributes of routes to be reflected.

   reflect change-path-attribute

   By default, the route reflector cannot change the attributes of routes to be reflected.

Configuring BGP VPNv4 route attributes

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Configure the NEXT_HOP attribute.
   - Set the device as the next hop for routes sent to a peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **next-hop-local**

     By default, the device uses its address as the next hop for routes advertised to peers.
   - Configure the device to not change the next hop of routes advertised to a peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **next-hop-invariable**

   By default, the device uses its address as the next hop for routes advertised to peers.

   On an RR in an inter-AS option C scenario, you must configure this command to not change the next hop of VPNv4 routes advertised to BGP peers and RR clients.

5. Configure the AS_PATH attribute.
   - Allow the local AS number to appear in the AS_PATH attribute of a received route and set the maximum number of repetitions.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **allow-as-loop** [*number*]

     By default, BGP discards incoming routes that contain the local AS number.
   - Remove private AS numbers in BGP updates sent to an EBGP peer or peer group.

     peer { *group-name* | *ipv4-address* [*mask-length*] } **public-as-only** [{ **force** | **limited** } [**replace**] [**include-peer-as**]]

     By default, BGP updates sent to an EBGP peer or peer group can carry both public and private AS numbers.

6. Advertise the COMMUNITY attribute to a peer or peer group.

peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise-community**

By default, BGP does not advertise the COMMUNITY attribute to a peer or peer group.

7. Configure the SoO attribute for a peer for peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **soo** *site-of-origin*

   By default, the SoO attribute is not configured.

8. Advertise the Tag community in the extended community attribute to the specified peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise tag**

   By default, BGP does not advertise the Tag extended community to a peer or peer group.

Configuring BGP VPNv4 route filtering

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Filter advertised routes.

   filter-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* | **prefix-list** *prefix-list-name* } **export** [**direct** | { **isis** | **ospf** | **rip** } *process-id* | **static**]

   By default, BGP does not filter advertised routes.

5. Filter received routes.

   filter-policy { *ipv4-acl-number* | **name** *ipv4-acl-name* | **prefix-list** *prefix-list-name* } **import**

   By default, BGP does not filter received routes.

6. Configure AS_PATH list-based route filtering for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-path-acl** *as-path-acl-number* { **export** | **import** }

   By default, AS_PATH list-based route filtering is not configured.

7. Configure ACL-based route filtering for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **filter-policy** { *ipv4-acl-number* | **name** *ipv4-acl-name* } { **export** | **import** }

   By default, ACL-based route filtering is not configured.

8. Configure IP prefix list-based route filtering for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **prefix-list** *prefix-list-name* { **export** | **import** }

   By default, no IP prefix list-based route filtering is configured.

9. Apply a routing policy to routes advertised to or received from a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **route-policy** *route-policy-name* { **export** | **import** }

   By default, no routing policy is applied for a peer.

10. Enable route target filtering for received BGP VPNv4 routes.

    policy vpn-target

By default, route target filtering is enabled for received VPNv4 routes. Only VPNv4 routes whose export route target attribute matches the local import route target attribute are added to the routing table.

11. Enable the first AS number check for EBGP route advertisement to EBGP peers.

    peer-as-check enable

    By default, BGP does not check the first AS number of a received route and advertises the route to other peers.

    After you execute this command, BGP checks the first AS number of a received route. If the AS number of an EBGP peer is the first AS number of the EBGP route to be advertised, BGP will not advertise the route to that EBGP peer.

Configuring BGP VPNv4 route dampening

About this task

This feature enables BGP to not select unstable routes as optimal routes.

Restrictions and guidelines

This feature applies only to IBGP routes.

If an IBGP peer goes down after you configure this feature, VPNv4 routes coming from the peer are dampened but not deleted.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Configure BGP VPNv4 route dampening.

   dampening ibgp[*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *

   By default, BGP VPNv4 route dampening is not configured.

Configuring BGP VPNv4 optimal route selection delay

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Set the BGP VPNv4 optimal route selection delay timer. Choose one or more of the following tasks:

   - Configure optimal route selection delay for all BGP routes in the address family.

     route-select delay *delay-value*

   - Configure optimal route selection delay only for BGP routes from a peer within the specified time after the peer state changes from down to up.

     route-select suppress on-peer-up *milliseconds*

By default, the optimal route selection delay timer is 0 seconds, which means optimal route selection is not delayed.

Advertising BGP RPKI validation state to a peer or peer group

Restrictions and guidelines

BGP advertises the BGP RPKI validation state to a peer or peer group through the extended community attribute. For more information about BGP RPKI, see BGP configuration in *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP VPNv4 address family view.
   address-family vpnv4
4. Advertise the BGP RPKI validation state to the specified peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise origin-as-validation**

   By default, BGP does not advertise the BGP RPKI validation state.

Configuring inter-AS VPN

Configuring inter-AS option A

Inter-AS option A applies to scenarios with a few VPNs.

To configure inter-AS option A, create VPN instances on PEs and ASBRs. The VPN instances on PEs are used to allow CEs to access the network. The VPN instances on ASBRs are used to access the peer ASBRs. An ASBR considers the peer ASBR as a CE.

The route targets configured on the PEs must match those configured on the ASBRs in the same AS to make sure VPN routes sent by the PEs (or ASBRs) can be received by the ASBRs (or PEs). Route targets configured on the PEs in different ASs do not have such requirements.

Configuring inter-AS option B

Restrictions and guidelines

An ASBR always sets itself as the next hop of VPNv4 routes advertised to an MP-IBGP peer regardless of the **peer next-hop-local** command.

Configuring a PE

Configure basic MPLS L3VPN, and specify the ASBR in the same AS as an MP-IBGP peer. The route targets for the VPN instances on the PEs in different ASs must match for the same VPN.

Configuring an ASBR

1. Enter system view.

system-view

2. Enable MPLS and LDP on the interface connected to an internal router of the AS:
   a. Configure an LSR ID for the local LSR.
      mpls lsr-id *lsr-id*
      By default, no LSR ID is configured.
   b. Enable LDP on the local LSR and enter LDP view.
      mpls ldp
      By default, LDP is disabled.
   c. Return to system view.
      quit
   d. Enter interface view of the interface connected to an internal router of the AS.
      interface *interface-type interface-number*
   e. Enable MPLS on the interface.
      mpls enable
      By default, MPLS is disabled on the interface.
   f. Enable MPLS LDP on the interface.
      mpls ldp enable
      By default, MPLS LDP is disabled on the interface.
   g. Return to system view.
      quit
3. Enable MPLS on the interface connected to the remote ASBR:
   a. Enter interface view of the interface connected to the remote ASBR.
      interface *interface-type interface-number*
   b. Enable MPLS on the interface.
      mpls enable
      By default, MPLS is disabled on the interface.
   c. Return to system view.
      quit
4. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
5. Configure PEs in the same AS as IBGP peers and ASBRs in different ASs as EBGP peers.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
6. Enter BGP VPNv4 address family view.
   address-family vpnv4
7. Enable BGP to exchange VPNv4 routes with the PE in the same AS and the ASBR in another AS.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP cannot exchange VPNv4 routes with a peer.
8. Disable route target filtering of received VPNv4 routes.
   undo policy vpn-target
   By default, route target filtering is enabled for received VPNv4 routes.

Configuring inter-AS option C (method 1)

Restrictions and guidelines

PEs are not directly connected. For the PEs to establish neighbor relationships, execute the **peer ebgp-max-hop** command to enable the local router to establish an EBGP session to an indirectly-connected peer.

Prerequisites

Before you configure inter-AS option C, perform the following tasks:

- Configure BGP to advertise routes destined for a PE on PEs or ASBRs. For more information, see *Layer 3—IP Routing Configuration Guide*.
- Configure VPN instances on PEs.
- Configure routing between PE and CE.

Configuring a PE

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Configure the ASBR in the same AS as an IBGP peer and configure the PE of another AS as an EBGP peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
4. Create the BGP IPv4 unicast address family and enter its view.
   address-family ipv4 [**unicast**]
5. Enable BGP to exchange IPv4 unicast routes with the ASBR in the same AS.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP does not exchange IPv4 unicast routes with any peer.
6. Enable BGP to exchange labeled IPv4 routes with the ASBR in the same AS.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **label-route-capability**
   By default, BGP cannot exchange labeled routes with any IPv4 peer or peer group.
7. Return to BGP instance view.
   quit
8. Enter BGP VPNv4 address family view.
   address-family vpnv4
9. Enable BGP to exchange VPNv4 routes with the PE in different ASs.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP cannot exchange VPNv4 routes with any peer.
10. (Optional.) Configure the PE to not change the next hop of routes advertised to the peer.
    peer { *group-name* | *ipv4-address* [*mask-length*] } **next-hop-invariable**
    By default, the device uses its address as the next hop of routes advertised to peers.
    Configure this command on the RR so the RR does not change the next hop of advertised VPNv4 routes.

Configuring an ASBR

1. Enter system view.
   system-view

2. Configure a routing policy:
   a. Create a routing policy, and enter routing policy view.
      route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*
   b. Match IPv4 routes carrying labels.
      if-match mpls-label

      By default, no MPLS label match criterion is configured.

      You can configure **if-match** clauses in the routing policy to filter routes. Routes surviving the filtering are assigned labels, and all others are advertised as common IPv4 routes.
   c. Set labels for IPv4 routes.
      apply mpls-label

      By default, no MPLS label is set for IPv4 routes.
   d. Return to system view.
      quit
3. Enable MPLS and LDP on the interface connected to an internal router of the AS:
   a. Configure an LSR ID for the local LSR.
      mpls lsr-id *lsr-id*

      By default, no LSR ID is configured.
   b. Enable LDP on the local LSR and enter LDP view.
      mpls ldp

      By default, LDP is disabled.
   c. Return to system view.
      quit
   d. Enter interface view of the interface connected to an internal router of the AS.
      interface *interface-type interface-number*
   e. Enable MPLS on the interface.
      mpls enable

      By default, MPLS is disabled on the interface.
   f. Enable MPLS LDP on the interface.
      mpls ldp enable

      By default, MPLS LDP is disabled on the interface.
   g. Return to system view.
      quit
4. Enable MPLS on the interface connected to the remote ASBR.
   a. Enter interface view of the interface connected to the remote ASBR.
      interface *interface-type interface-number*
   b. Enable MPLS on the interface.
      mpls enable

      By default, MPLS is disabled on the interface.
   c. Return to system view.
      quit
5. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
6. Configure PEs in the same AS as IBGP peers and ASBRs in different ASs as EBGP peers.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*

7. Create BGP IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

8. Enable BGP to exchange IPv4 unicast routes with the PE in the same AS and the ASBR in another AS.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**

   By default, BGP cannot exchange IPv4 unicast routes with a peer.

9. Enable BGP to exchange labeled IPv4 routes with the PE in the same AS and the ASBR in another AS.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **label-route-capability**

   By default, BGP cannot exchange labeled IPv4 routes with a peer.

10. Specify the device as the next hop of routes sent to a peer or peer group.

    peer { *group-name* | *ipv4-address* [*mask-length*] } **next-hop-local**

    By default, the device does not set itself as the next hop for routes sent to a peer or peer group.

11. Apply a routing policy to routes advertised to or received from a peer or peer group.

    peer { *group-name* | *ipv4-address* [*mask-length*] } **route-policy** *route-policy-name* { **export** | **import** }

    By default, no routing policy is applied.

Configuring inter-AS option C (method 2)

Restrictions and guidelines

PEs are not directly connected. For the PEs to establish neighbor relationships, execute the **peer ebgp-max-hop** command to enable the local router to establish an EBGP session to an indirectly-connected peer.

Prerequisites

Before you configure inter-AS option C (method 2), perform the following tasks:

- Configure BGP on the PE or ASBR to advertise the route destined for the PE. For more information, see BGP configuration in *Layer 3—IP Routing Configuration Guide*.
- Configure VPN instances on PEs.
- Configure routing between PE and CE.

Configuring a PE

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Configure the PE in another AS as an EBGP peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*

4. Enter BGP VPNv4 address family view.

   address-family vpnv4

5. Enable BGP to exchange VPNv4 routes with the PE in another AS.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**

   By default, BGP cannot exchange VPNv4 routes with any peer.

6. (Optional.) Configure the PE to not change the next hop of routes advertised to the peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **next-hop-invariable**

By default, the device uses its address as the next hop of routes advertised to peers.

Configure this command on the RR so the RR does not change the next hop of advertised VPNv4 routes.

Configuring an ASBR

1. Enter system view.

   system-view

2. Configure a routing policy:

   a. Create a routing policy, and enter routing policy view.

      route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*

   b. Match IPv4 routes carrying labels.

      if-match mpls-label

      By default, no MPLS label match criterion is configured.

      You can configure **if-match** clauses in the routing policy to filter routes. Routes surviving the filtering are assigned labels, and all others are advertised as common IPv4 routes.

   c. Set labels for IPv4 routes.

      apply mpls-label

      By default, no MPLS label is set for IPv4 routes.

   d. Return to system view.

      quit

3. Enable MPLS and LDP on the interface connected to an internal router of the AS:

   a. Configure an LSR ID for the local LSR.

      mpls lsr-id *lsr-id*

      By default, no LSR ID is configured.

   b. Enable LDP on the local LSR and enter LDP view.

      mpls ldp

      By default, LDP is disabled.

   c. Return to system view.

      quit

   d. Enter interface view of the interface connected to an internal router of the AS.

      interface *interface-type interface-number*

   e. Enable MPLS on the interface.

      mpls enable

      By default, MPLS is disabled on the interface.

   f. Enable MPLS LDP on the interface.

      mpls ldp enable

      By default, MPLS LDP is disabled on the interface.

   g. Return to system view.

      quit

4. Enable MPLS on the interface connected to the remote ASBR.

   a. Enter interface view of the interface connected to the remote ASBR.

      interface *interface-type interface-number*

   b. Enable MPLS on the interface.

      mpls enable

      By default, MPLS is disabled on the interface.

c. Return to system view.

quit

5. Enter BGP instance view.

bgp *as-number* [**instance** *instance-name*]

6. Configure the ASBR in another AS as an EBGP peer.

peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*

7. Create BGP IPv4 unicast address family and enter its view.

address-family ipv4 [**unicast**]

8. Enable BGP to exchange IPv4 unicast routes with the ASBR in another AS.

peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**

By default, BGP cannot exchange IPv4 unicast routes with a peer.

9. Enable BGP to exchange labeled IPv4 routes with the ASBR in another AS.

peer { *group-name* | *ipv4-address* [*mask-length*] } **label-route-capability**

By default, BGP cannot exchange labeled IPv4 routes with a peer.

10. Specify the device as the next hop of routes sent to a peer or peer group.

peer { *group-name* | *ipv4-address* [*mask-length*] } **next-hop-local**

By default, the device does not set itself as the next hop for routes sent to an IBGP peer or peer group.

11. Apply a routing policy to routes advertised to or received from a peer or peer group.

peer { *group-name* | *ipv4-address* [*mask-length*] } **route-policy** *route-policy-name* { **export** | **import** }

By default, no routing policy is applied.

Configuring nested VPN

Restrictions and guidelines

When you configure nested VPN, follow these guidelines:

- The address spaces of sub-VPNs of a VPN cannot overlap.
- Do not assign nested VPN peers addresses that public network peers use.
- Nested VPN does not support multihop EBGP. A provider PE and a provider CE must use the addresses of the directly connected interfaces to establish a neighbor relationship.

Procedure

1. Connect the customer CE to the customer PE:
   a. Configure VPN instances on the customer PE.

   See "Configuring VPN instances."

   b. Configure route exchange between customer PE and customer CE.

   See "Configuring routing between a PE and a CE."

2. Configure the customer PE to exchange sub-VPN routing information with the provider CE:
   a. Configure BGP VPNv4 route exchange between customer PE and provider CE.

   See "Configuring routing between PEs."

   b. Execute the following commands in sequence to configure the provider CE to receive all VPNv4 routes (not filter VPNv4 routes by route targets).

   system-view

bgp *as-number* [**instance** *instance-name*]

address-family vpnv4

undo policy vpn-target

By default, route target filtering is enabled for received VPNv4 routes. Only VPNv4 routes whose export route target attribute matches the local import route target attribute are added to the routing table.

This step is not required in a nested VPN network when no provider CE is deployed and the customer PE is connected directly to the provider PE.

3. Connect the provider CE to the provider PE:
   a. Configure VPN instances on the provider PE.

      See "Configuring VPN instances."
   b. Configure route exchange between the provider CE and provider PE.

      See "Configuring routing between a PE and a CE."
4. Configure BGP VPNv4 route exchange between the provider PE and provider CE:

   This step contains only configurations on the provider PE. For information about configuring the provider CE, see "Configuring routing between PEs." If the customer PE is directly connected to the provider PE, you must configure provider CE settings on the customer PE because the PE also functions as a provider CE.
   a. Execute the following commands in sequence to enter BGP VPNv4 address family view:

      system-view

      bgp *as-number* [**instance** *instance-name*]

      address-family vpnv4
   b. Enable nested VPN.

      nesting-vpn

      By default, nested VPN is disabled.
   c. Return to BGP instance view.

      quit
   d. Enter BGP-VPN instance view.

      ip vpn-instance *vpn-instance-name*
   e. Configure the provider CE as a BGP peer.

      peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
   f. Create BGP-VPN VPNv4 address family and enter BGP-VPN VPNv4 address family view.

      address-family vpnv4
   g. Enable BGP VPNv4 route exchange with the peer CE or the peer group of the peer CE.

      peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**

      By default, BGP does not exchange VPNv4 routes with any peer.
   h. (Optional.) Configure the SoO attribute for the BGP peer or peer group.

      peer { *group-name* | *ipv4-address* [*mask-length*] } **soo** *site-of-origin*

      By default, the SoO attribute is not configured.
5. Configure BGP VPNv4 route exchange between provider PEs.

   See "Configuring routing between PEs."

Configuring HoVPN

Configuring the UPE

Configure basic MPLS L3VPN settings on the UPE.

Configuring the SPE

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Specify a BGP peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
4. Enter BGP VPNv4 address family view.
   address-family vpnv4
5. Enable BGP VPNv4 route exchange with the peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP does not exchange VPNv4 routes with any peer.
6. Specify the BGP peer or peer group as a UPE.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **upe**
   By default, no peer is a UPE.
7. Advertise routes to the UPE.
   - Advertise a default VPN route to the UPE.
     peer { *group-name* | *ipv4-address* [*mask-length*] } **default-route-advertise vpn-instance** *vpn-instance-name*
     The device advertises a default route using the local address as the next hop to the UPE, regardless of whether the default route exists in the local routing table.
   - Advertise routes permitted by a routing policy to the UPE.
     peer { *group-name* | *ipv4-address* [*mask-length*] } **upe route-policy** *route-policy-name* **export**

   By default, no route is advertised to the UPE.

   Do not configure both commands.
8. Return to BGP instance view.
   quit
9. Create a BGP-VPN instance and enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
   You do not need to associate the VPN instance to an interface on the SPE.
   This step adds the learned VPNv4 routes to the BGP routing table of the VPN instance.

Configuring route re-origination

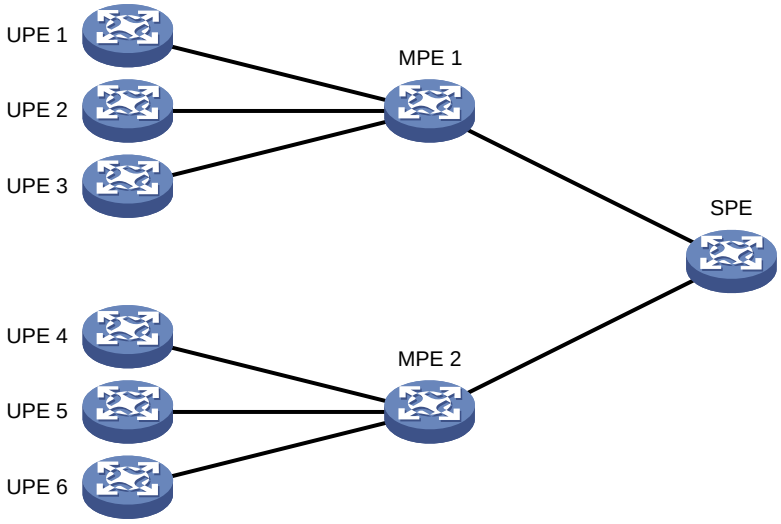
About this task

In an HoVPN network, different UPEs communicate with each other through MPEs and SPEs. You can configure route re-origination on MPEs to reduce the number of private network labels that a UPE receives for VPNv4 routes.

As shown in Figure 24, if a network contains many UPEs that use the per-VPN-instance label allocation mode, and the MPEs in the network use the per-next-hop label allocation mode, the SPE will receive a large number of labels, which might cause traffic forwarding errors.

To resolve this issue, you can configure route re-origination on MPEs. The MPEs then can redistribute the BGP routes received from UPEs into local VPN instances and reoriginate these routes. The MPEs can modify the information of the reoriginated routes. After setting the per-VPN instance label allocation mode, the MPEs only need to allocate the number of VPN labels equal to the number of local VPN instances, regardless of the number of UPEs. The SPE only needs to receive the VPN labels allocated by the MPEs, significantly reducing the resource load on the SPE.

Figure 24 Route re-origination



Restrictions and guidelines

This feature can reoriginate the BGP routes that are imported into a local VPN instance and have a different RD from that of the local VPN instance. It cannot reoriginate the BGP routes that are received from remote peers and have the same RD as that of the local VPN instance.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
4. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
5. Configure the device to re-originate the optimal routes in the VPN instance and advertise the re-originated routes to VPNv4 peers.

advertise route-reoriginate [**route-policy** *route-policy-name*] [**replace-rt**]

By default, the device does not re-originate the optimal routes in a VPN instance, and it sends the original VPNv4 routes to VPNv4 peers.

Specifying the VPN label processing mode on the egress PE

About this task

An egress PE can process VPN labels in either POPGO or POP mode:

- **POPGO forwarding**—Pops the label and forwards the packet out of the output interface corresponding to the label.
- **POP forwarding**—Pops the label and forwards the packet through the FIB table.

Restrictions and guidelines

The POPGO forwarding mode (**vpn popgo**) and per-VPN instance label allocation mode (**label-allocation-mode per-vrf**) are mutually exclusive. Do not configure both modes in a BGP instance. For more information about the **label-allocation-mode** command, see *Layer 3—IP Routing Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Specify the VPN label processing mode as POPGO forwarding.
   vpn popgo
   The default is POP forwarding.

Configuring MPLS L3VPN FRR

About MPLS L3VPN FRR

You can use the following methods to configure MPLS L3VPN FRR:

- **Method 1**—Execute the **pic** command in BGP-VPN IPv4 unicast address family view. The device calculates a backup next hop for each BGP route in the VPN instance if there are two or more unequal-cost routes to reach the destination.
- **Method 2**—Execute the **fast-reroute route-policy** command in BGP-VPN IPv4 unicast address family view to use a routing policy. In the routing policy, specify a backup next hop by using the **apply fast-reroute backup-nexthop** command. The backup next hop calculated by the device must be the same as the specified backup next hop. Otherwise, the device does not generate a backup next hop for the primary route. You can also configure **if-match** clauses in the routing policy to identify the routes protected by FRR.

If both methods are configured, Method 2 takes precedence over Method 1.

Restrictions and guidelines

Executing the **pic** command in BGP-VPN IPv4 unicast address family view might cause routing loops. Use it with caution.

Configuring FRR by using a routing policy

1. Enter system view.

   system-view
2. Configure BFD.
   - Configure the source IP address for BFD echo packets.

     bfd echo-source-ip *ip-address*

     By default, the source IP address for BFD echo packets is not configured.

     This command is required when echo-mode BFD is used to detect primary route connectivity in VPNv4 route backup for an IPv4 route. For more information about this command, see *High Availability Command Reference*.
3. Configure a routing policy:
   a. Create a routing policy and enter routing policy view.

      route-policy *route-policy-name* **permit node** *node-number*

   b. Set the backup next hop for FRR.

      apply fast-reroute backup-nexthop *ip-address*

      By default, no backup next hop address is set for FRR.

   c. Return to system view.

      quit

   For more information about the commands, see *Layer 3—IP Routing Command Reference*.
4. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
5. (Optional.) Use echo-mode BFD to detect the connectivity to the next hop of the primary route.

   primary-path-detect bfd echo

   By default, ARP is used to detect the connectivity to the next hop.

   Use this command if necessary in VPNv4 route backup an IPv4 route.

   For more information about this command, see *Layer 3—IP Routing Command Reference*.
6. Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*
7. Enter BGP-VPN IPv4 unicast address family view.

   address-family ipv4 [**unicast**]
8. Apply a routing policy to FRR.

   fast-reroute route-policy *route-policy-name*

   By default, no routing policy is applied to FRR.

   The **apply fast-reroute backup-nexthop** command can take effect in the routing policy that is being used. Other **apply** commands do not take effect.

   For more information about the command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Enabling MPLS L3VPN FRR for BGP-VPN IPv4 unicast address family

1. Enter system view.
   system-view
2. Configure BFD.
   - Configure the source IP address for BFD echo packets.
     bfd echo-source-ip *ip-address*

     By default, the source IP address for BFD echo packets is not configured.

     This command is required when echo-mode BFD is used to detect primary route connectivity in VPNv4 route backup for an IPv4 route. For more information about this command, see BFD commands in *High Availability Command Reference*.
3. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
4. (Optional.) Use echo-mode BFD to detect the connectivity to the next hop of the primary route.
   primary-path-detect bfd echo

   By default, ARP is used to detect the connectivity to the next hop.

   Use this command if necessary in VPNv4 route backup an IPv4 route.

   For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.
5. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
6. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
7. Enable MPLS L3VPN FRR for the address family.
   pic

   By default, MPLS L3VPN FRR is disabled.

   For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Configuring an OSPF sham link

About OSPF sham links

When a backdoor link exists between the two sites of a VPN, you can create a sham link between PEs to forward VPN traffic through the sham link on the backbone rather than the backdoor link. A sham link is considered an OSPF intra-area route.

The source and destination addresses of the sham link must be loopback interface addresses with 32-bit masks. The loopback interfaces must be bound to VPN instances, and their addresses are advertised through BGP.

Prerequisites

Before you configure an OSPF sham link, perform the following tasks:

- Configure basic MPLS L3VPN (OSPF is used between PE and CE).

- Configure OSPF in the LAN where customer CEs reside.

Redistributing the loopback interface address

1. Enter system view.
   system-view
2. Create a loopback interface and enter loopback interface view.
   interface loopback *interface-number*
3. Associate the loopback interface with a VPN instance.
   ip binding vpn-instance *vpn-instance-name*

   By default, the interface is not associated with any VPN instances and belongs to the public network.
4. Configure an IP address for the loopback interface.
   ip address *ip-address* { *mask-length* | *mask* }

   By default, no IP address is configured for the loopback interface.
5. Return to system view.
   quit
6. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
7. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
8. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
9. Redistribute direct routes into BGP (including the loopback interface route).
   import-route direct

   By default, no direct routes are redistributed into BGP.

Creating a sham link

1. Enter system view.
   system-view
2. Enter OSPF view.
   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *

   As a best practice, specify a router ID.
3. Set the external route tag for imported VPN routes.
   route-tag *tag-value*

   By default, if BGP runs within an MPLS backbone, and the BGP AS number is not greater than 65535, the first two octets of the external route tag are 0xD000 and the last two octets are the local BGP AS number. If the AS number is greater than 65535, the external route tag is 0.
4. Enter OSPF area view.
   area *area-id*
5. Configure a sham link.
   sham-link *source-ip-address destination-ip-address* [**cost** *cost-value* | **dead** *dead-interval* | **hello** *hello-interval* | { **authentication-none** | { **hmac-md5** | **hmac-sha-256** | **hmac-sm3** | **md5** } [*key-id* { **cipher** | **plain** }

string] | **keychain** *keychain-name* | **simple** [{ **cipher** | **plain** } *string*] } | **retransmit** *retrans-interval* | **trans-delay** *delay* | **ttl-security hops** *hop-count*] *

Configuring BGP AS number substitution and SoO attribute

About this task

When CEs at different sites have the same AS number, configure the BGP AS number substitution feature to avoid route loss.

When a PE uses different interfaces to connect different CEs in a site, the BGP AS number substitution feature introduces a routing loop. To remove the routing loop, configure the SoO attribute on the PE.

For more information about the BGP AS number substitution feature and the SoO attribute, see "BGP AS number substitution and SoO attribute."

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
4. Enable the BGP AS number substitution feature.
   peer { *ipv4-address* [*mask-length*] | *group-name* } **substitute-as**
   By default, BGP AS number substitution is disabled.
5. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
6. (Optional.) Configure the SoO attribute for a BGP peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **soo** *site-of-origin*
   By default, the SoO attribute is not configured.

Configuring BGP RT filtering

About this task

The BGP RT filtering feature reduces the number of routes advertised in an MPLS L3VPN.

After RT filtering is configured, a PE advertises its import target attribute to the peer PEs in the RT filter address family. The peer PEs use the received import target attribute to filter routes and advertise only the routes that match the attribute to the PE.

When a large number of IBGP peers exist, the BGP RT filtering and the route reflection features are used together as a best practice. Route reflection reduces the number of IBGP connections. BGP RT filtering reduces the number of routes advertised in the network.

For more information about the BGP RT filtering commands, see *Layer 3—IP Routing Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP IPv4 RT filter address family view.
   address-family ipv4 rtfilter
4. Enable the device to exchange routing information with a peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, the device cannot exchange routing information with a peer or peer group.
5. (Optional.) Advertise a default route to a peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **default-route-advertise** [**route-policy** *route-policy-name*]
   By default, no default route is advertised.
6. (Optional.) Set a preferred value for routes received from a peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **preferred-value** *value*
   By default, the preferred value for routes received from a peer or peer group is 0.
7. (Optional.) Configure the device as a route reflector and specify a peer or peer group as its client.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **reflect-client**
   By default, no route reflector or client is configured.
8. (Optional.) Enable route reflection between clients.
   reflect between-clients
   By default, route reflection between clients is enabled.
9. (Optional.) Configure the cluster ID of the route reflector.
   reflector cluster-id { *cluster-id* | *ipv4-address* }
   By default, a route reflector uses its own router ID as the cluster ID.
10. (Optional.) Set the optimal route selection delay timer. Choose one or more of the following tasks:
    - Configure optimal route selection delay for all BGP routes in the address family.
      route-select delay *delay-value*
    - Configure optimal route selection delay only for BGP routes from a peer within the specified time after the peer state changes from down to up.
      route-select suppress on-peer-up *milliseconds*

    By default, the optimal route selection delay timer is 0 seconds, which means optimal route selection is not delayed.

Configuring the BGP additional path feature

About this task

By default, BGP advertises only one optimal route. When the optimal route fails, traffic forwarding will be interrupted until route convergence completes.

The BGP additional path (Add-Path) feature enables BGP to advertise multiple routes with the same prefix and different next hops to a peer or peer group. When the optimal route fails, the suboptimal route becomes the optimal route, which shortens the traffic interruption time.

You can enable the BGP additional path sending, receiving, or both sending and receiving capabilities on a BGP peer. For two BGP peers to successfully negotiate the additional path capabilities, make sure one end has the sending capability and the other end has the receiving capability.

For more information about the BGP additional path configuration commands, see BGP commands in *Layer 3—IP Routing Command Reference*.

Procedure

1. Enter system view.

   system-view
2. Enter BGP VPNv4 address family view or BGP-VPN VPNv4 address family view.
   - Execute the following commands in sequence to enter BGP VPNv4 address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family vpnv4
   - Execute the following commands in sequence to enter BGP-VPN VPNv4 address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family vpnv4
3. Configure the BGP additional path capabilities.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **additional-paths** { **receive** | **send** } *

   By default, no BGP additional path capabilities are configured.
4. Set the maximum number of Add-Path optimal routes that can be advertised to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise additional-paths best** *number*

   By default, a maximum of one Add-Path optimal route can be advertised to a peer or peer group.
5. Set the maximum total number of Add-Path optimal routes that can be advertised to all peers.

   additional-paths select-best *best-number*

   By default, a maximum total number of one Add-Path optimal route can be advertised to all peers.

   This command is not supported in BGP-VPN VPNv4 address family view.

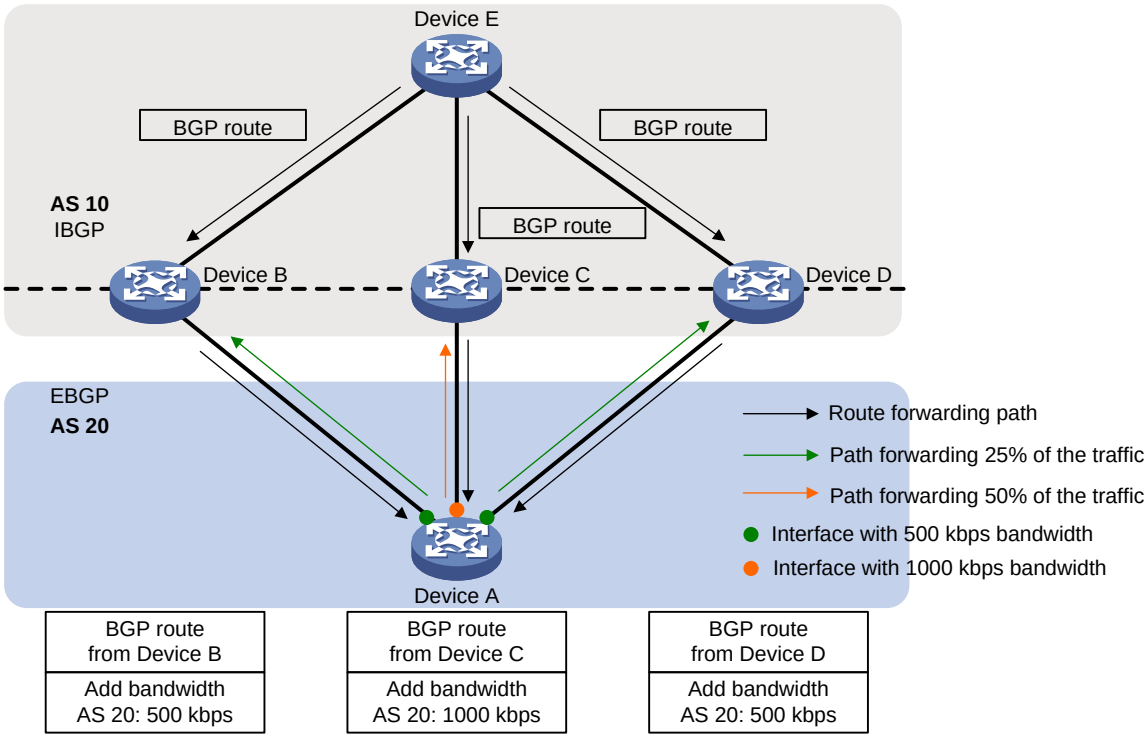
Configuring link bandwidth attributes for BGP VPNv4 routes

About this task

ECMP load balancing equally distributes traffic to the equal-cost links without considering link bandwidth differences. This might cause insufficient usage of high-speed links and traffic congestion on low-speed links. To resolve this issue, you can configure Unequal Cost Multiple Path (UCMP) load balancing based on the link bandwidth attributes of BGP routes.

The link bandwidth attribute of a route indicates the link bandwidth of the outgoing interface of the route. After UCMP load balancing is enabled for an EBGP peer on the device, the device adds the link bandwidth attribute into the route received from the EBGP peer. This attribute carries the AS number of the current BGP process and the interface bandwidth for establishing a direct EBGP session. UCMP BGP routes are formed in the following situation: BGP route load balancing is configured by using the **balance** command in BGP-VPN IPv4 address family view. All the BGP routes for load balancing carry a link bandwidth attribute. In traffic forwarding, the proportion of the traffic carried on each BGP route is that of the route's link bandwidth to the total link bandwidth of all UCMP routes. As shown in Figure 25, Device A receives routes for the same destination address from three EBGP peers. Assume Device A adds link bandwidth attribute values 500 kbps, 1000 kbps, and 500 kbps for the three BGP routes. The traffic load proportions carried on the three BGP routes are 25%, 50%, and 25%, respectively. For more information the **balance** command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Figure 25 UCMP load balancing based on BGP routes with link bandwidth attributes



By default, the BGP routes with link bandwidth attributes are advertised only to IBGP peers and the added link bandwidth attributes are optional non-transitive attributes. You can change the advertisement rule for BGP routes with link bandwidth attributes as needed:

- To advertise link bandwidth attributes to EBGP peers, execute the **peer advertise ebgp bandwidth** command.
- To make BGP peers that cannot identify link bandwidth attributes to carry link bandwidth attributes when they forward BGP routes, execute the **peer advertise bandwidth transitive** command. This command can convert the link bandwidth attributes carried in BGP routes into optional transitive attributes.

Restrictions and guidelines

When you configure this feature to add link bandwidth attributes to BGP routes, follow these restrictions and guidelines:

- This feature applies only to direct EBGP peers or peer groups. If the specified peer or peer group becomes an indirectly connected neighbor, the device no longer adds link bandwidth attributes to the routes received from that neighbor.

- Only if all BGP routes for the same destination address carry the link bandwidth attributes, can these routes provide UCMP load balancing. Otherwise, the routes can only provide ECMP load balancing.
- A BGP route can carry a maximum of one link bandwidth attribute. If a received route already has a link bandwidth attribute, the device cannot add a link bandwidth attribute to the route any more.

In advertisement of BGP routes with link bandwidth attributes, the following restrictions apply:

- If you use the **peer route-policy export** command to apply a routing policy to a BGP peer/peer group, and the routing policy uses the **apply extcommunity** command to configure a link bandwidth attribute for BGP routes, that link bandwidth attribute will be carried in the BGP routes advertised to the BGP peer/peer group.
- When the device receives a BGP route with a link bandwidth attribute, the device deletes the link bandwidth attribute before advertising the route if the next hop address of the route needs to be changed before advertisement. This is because the forwarding link instructed by this route has changed and thus the carried link bandwidth no longer indicates the bandwidth of the actual forwarding link. This restriction does not exist if the device receives a route without a link bandwidth attribute and then adds a link bandwidth attribute to the route.
- For UCMP routes, if the Add-Path feature is configured to advertise only one optimal route, the optimal route advertised will carry a link bandwidth attribute that is the total link bandwidth of all UCMP routes. If the Add-Path feature is configured to advertise more than one optimal route, the advertised routes carry their respective link bandwidth attributes.

Procedure

1. Enter system view.

   system-view

2. Enter BGP VPNv4 address family view or BGP-VPN VPNv4 address family view.
   - Execute the following commands in sequence to enter BGP VPNv4 address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family vpnv4
   - Execute the following commands in sequence to enter BGP-VPN VPNv4 address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family vpnv4

3. Configure a link bandwidth attribute for a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **bandwidth** [*bandwidth-value*]

   By default, no link bandwidth attribute is configured for a peer or peer group.

4. (Optional.) Configure the device to advertise link bandwidth attributes to EBGP peers.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise ebgp bandwidth**

   By default, the device does not advertise link bandwidth attributes to EBGP peers.

5. (Optional.) Convert the link bandwidth attributes to optional transitive attributes in the BGP routes to be advertised to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **advertise bandwidth transitive**

   By default, the link bandwidth attributes in BGP routes are optional non-transitive attributes.

Minimizing the priority of BGP routes sent to peers

Minimizing the priority of BGP routes sent to a peer after the peer state changes from down to up

About this task

By default, a device immediately sends BGP routes to a peer after the peer comes up. If ARP entry convergence is not completed on the device, traffic matching the BGP routes might fail to be forwarded, causing packet loss. You can configure this feature to resolve this issue.

After you configure this feature on a device, the device works as follows:

- After a peer comes up, the device minimizes the priority (sets the local preference to 0 and the MED value to 4294967295) of BGP routes sent to the peer within the specified time. Therefore, the peer will not select BGP routes sent by the device as the optimal routes.
- After the specified timer expires, the device will immediately send BGP routes with the original priority to the peer. All subsequent BGP routes that the device sends to the peer carry the original priority.

If ARP entry convergence is completed on the device before the specified timer expires, you can manually restore the priority of BGP routes by executing either of the following commands:

- **`reset bgp advertise lowest-priority on-peer-up`**: After you execute this command, the device restores the priority of BGP routes sent to peers whose timers have not expired and immediately sends the routes to those peers. However, the device still minimizes the priority of BGP routes sent to a peer after the peer comes up.
- **`undo advertise lowest-priority on-peer-up duration`**: After you execute this command, the device immediately sends BGP routes with the original priority to peers whose timers have not expired. All subsequent BGP routes that the device sends to a peer carry the original priority even if the peer state changes.

Restrictions and guidelines

You can execute the **`advertise lowest-priority on-peer-up duration`** command multiple times to edit the duration for the device to minimize the priority of BGP routes. This command takes effect as follows:

- If the timer for a peer has not expired, the command takes effect immediately and the timer is reset. For example, if you execute the **`advertise lowest-priority on-peer-up duration`** *`10`* command, and then execute the **`advertise lowest-priority on-peer-up duration`** *`6`* command six seconds after the state of a peer changes from down to up, the device minimizes the priority of BGP routes sent to the peer for the next six seconds.
- If the timer has not started or has expired for all peers, the command takes effect when the state of a peer changes from down to up.

If you execute the **`advertise lowest-priority on-peer-up duration`** or **`advertise lowest-priority on-startup duration`** command together with the **`bgp update-delay on-startup`** or **`route-update-delay`** command or execute the **`advertise lowest-priority on-peer-up duration`** and **`advertise lowest-priority on-startup duration`** commands together, the configuration result is as shown in Table 11.

Table 11 Configuration results

Command A	Command B	Configuration result
`bgp update-delay`	**`advertise lowest-priority`**	In an address family where command B is executed, command

on-startup	**on-startup duration** (in address family view)	B takes effect and command A does not take effect. In an address family where command B is not executed, command A takes effect.
	advertise lowest-priority on-peer-up duration (in address family view)	In an address family where command B is not executed, command A takes effect. Suppose that the timer specified by command A is TA and the timer specified by command B is TB. In an address family where command B is executed, the configuration result is as follows: • If TA ≥ TB, only command A takes effect. • If TA < TB, the device does not send BGP routes to peers within the time period specified by TA, and then sends BGP routes with the minimized priority to peers within the time period obtained by TB - TA. After that, the device immediately sends BGP routes with the original priority to peers and stops minimizing the priority of BGP routes.
route-update-delay	**advertise lowest-priority on-startup duration** (in BGP VPNv4 or BGP-VPN VPNv4 address family view)	Suppose that the timer specified by command B is TB, and the longer one between the timer specified by the **peer route-update-interval** command and the timer specified by command A is TS. In BGP VPNv4 or BGP-VPN VPNv4 address family view, the configuration result is as follows: • If TS ≥ TB, command A takes effect. • If TS < TB, the device does not send BGP routes to peers within the time period specified by TS, and then sends BGP routes with the minimized priority to peers within the time period obtained by TB - TS. After that, the device immediately sends BGP routes with the original priority to peers and stops minimizing the priority of BGP routes. For a peer, if the **peer route-update-interval** command does not specify the peer, TS is the longer one between the default value of the **peer route-update-interval** command and the timer specified by command A.
	advertise	Suppose that the timer specified by

	lowest-priority on-peer-up duration (in BGP VPNv4 or BGP-VPN VPNv4 address family view)	command B is TB, and the longer one between the timer specified by the **peer route-update-interval** command and the timer specified by command A is TS. In BGP VPNv4 or BGP-VPN VPNv4 address family view, the configuration result is as follows: • If TS ≥ TB, command A takes effect. • If TS < TB, the device does not send BGP routes to peers within the time period specified by TS, and then sends BGP routes with the minimized priority to peers within the time period obtained by TB - TS. After that, the device immediately sends BGP routes with the original priority to peers and stops minimizing the priority of BGP routes. For a peer, if the **peer route-update-interval** command does not specify the peer, TS is the longer one between the default value of the **peer route-update-interval** command and the timer specified by command A.
advertise lowest-priority on-peer-up duration (in address family view)	**advertise lowest-priority on-startup duration** (in address family view)	Command A and command B are mutually exclusive in an address family.

For more information about the **advertise lowest-priority on-peer-up duration**, **advertise lowest-priority on-startup duration**, **peer route-update-interval**, **route-update-delay**, and **bgp update-delay on-startup** commands, see BGP commands in *Layer 3—IP Routing Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter BGP VPNv4 address family view or BGP-VPN VPNv4 address family view.
   - Execute the following commands in sequence to enter BGP VPNv4 address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family vpnv4
   - Execute the following commands in sequence to enter BGP-VPN VPNv4 address family view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
     address-family vpnv4

3. Minimize the priority of BGP routes sent to a peer within the specified time after the peer state changes from down to up.

   advertise lowest-priority on-peer-up duration *seconds*

   By default, the device does not minimize the priority of BGP routes sent to peers.

Minimizing the priority of BGP routes sent to peers after the device restarts

About this task

By default, a device immediately sends BGP routes to its peers after the device restarts and its BGP process restores. If ARP entry convergence is not completed on the device, traffic matching the BGP routes might fail to be forwarded, causing packet loss. You can configure this feature to resolve this issue.

After you configure this feature on a device, the device works as follows:

- After the device restarts and its BGP process restores, the device minimizes the priority (sets the local preference to 0 and the MED value to 4294967295) of BGP routes sent to its peers within the specified time. Therefore, the peers will not select BGP routes sent by the device as the optimal routes.
- After the specified timer expires, the device will immediately send BGP routes with the original priority to its peers. All subsequent BGP routes that the device sends to its peers carry the original priority.

If ARP entry convergence is completed on the device before the specified timer expires, you can manually restore the priority of BGP routes by executing either of the following commands:

- **reset bgp advertise lowest-priority on-startup**: After you execute this command, the device restores the priority of BGP routes sent to peers and immediately sends the routes to peers. However, the device still minimizes the priority of BGP routes sent to its peers after the device restarts.
- **undo advertise lowest-priority on-startup duration**: After you execute this command, the device immediately sends BGP routes with the original priority to its peers. All subsequent BGP routes that the device sends to its peers carry the original priority even if the device restarts.

Restrictions and guidelines

This feature takes effect after you save the configuration and restart the device.

The confjiguration results differ if you execute the **advertise lowest-priority on-peer-up duration** or **advertise lowest-priority on-startup duration** command together with the **bgp update-delay on-startup** or **route-update-delay** command, or if you execute the **advertise lowest-priority on-peer-up duration** and **advertise lowest-priority on-startup duration** commands together. For more information about the configuration results, see "Minimizing the priority of BGP routes sent to a peer after the peer state changes from down to up."

For more information about the **advertise lowest-priority on-peer-up duration**, **advertise lowest-priority on-startup duration**, **peer route-update-interval**, **route-update-delay**, and **bgp update-delay on-startup** commands, see BGP commands in *Layer 3—IP Routing Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Enter BGP VPNv4 address family view or BGP-VPN VPNv4 address family view.
   - Execute the following commands in sequence to enter BGP VPNv4 address family view:
     bgp *as-number* [**instance** *instance-name*]
     address-family vpnv4
   - Execute the following commands in sequence to enter BGP-VPN VPNv4 address family view:
     bgp *as-number* [**instance** *instance-name*]
     ip vpn-instance *vpn-instance-name*
     address-family vpnv4
3. Minimize the priority of BGP routes sent to peers within the specified time after the device restarts and its BGP process restores.
   advertise lowest-priority on-startup duration *seconds*
   By default, the device does not minimize the priority of BGP routes sent to peers.

Restoring the priority of BGP routes sent to peers

About this task

After configuring the device to minimize the priority of BGP routes sent to a peer within the specified time after the device restarts or after the peer state changes from down to up, you can perform this task to manually restore the priority of BGP routes. After you configure this feature, the device restores the priority of BGP routes sent to peers whose timers have not expired and immediately sends the routes to those peers. However, the device still minimizes the priority of BGP routes sent to a peer after the peer comes up or after the device restarts.

Procedure

Execute the following command in user view to restore the priority of BGP routes sent to peers:

reset bgp [**instance** *instance-name*] **advertise lowest-priority** { **on-peer-up** | **on-startup** }

Specifying a local network to be advertised in the public instance or a VPN instance

About this task

This feature specifies a local network of the public instance or a VPN instance. Then, you can configure BGP to redistribute the network (by using the **import-route local-aggregate** command) and advertise the network. For more information about the **import-route local-aggregate** command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Restrictions and guidelines

The specified local network route must exist and be active in the routing table of the public instance or VPN instance.

Procedure

1. Enter system view.
   system-view
2. Enter public instance IPv4 address family view or VPN instance IPv4 address family view.

- Execute the following commands in sequence to enter public instance IPv4 address family view:

  ip public-instance

  address-family ipv4

- Execute the following commands in sequence to enter VPN instance IPv4 address family view:

  ip vpn-instance *vpn-instance-name*

  address-family ipv4

3. Specify a local network to be advertised.

   network *ipv4-address* [*mask-length* | *mask*]

   By default, no local network in the public instance or a VPN instance will be advertised.

Configuring BGP VPNv4 routes to use private network next hops

About this task

By default, the device does not change the next hop attribute of a received BGP VPNv4 route. The next hop address of a BGP VPNv4 route is a public address. This feature changes the next hop of a BGP VPNv4 route received from a peer or peer group to an IP address in the VPN instance. The outgoing label of the VPNv4 route is also changed to an invalid value. For example, the device received a VPNv4 route and its next hop address is 10.1.1.1, which is a public address by default. After this feature is configured, the next hop address changes to private address 10.1.1.1.

Restrictions and guidelines

After you configure this feature, the following applies:

- The device re-establishes the BGP sessions to the specified peer or to all peers in the specified peer group.
- The device receives a BGP VPNv4 route only when its RD is the same as a local RD.
- When advertising a BGP VPNv4 route received from the specified peer or peer group, the device does not change the route target attribute of the route.
- If you delete a VPN instance or its RD, BGP VPNv4 routes received from the specified peer or peer group and in the VPN instance will be deleted.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view:

   address-family vpnv4

4. Change the next hop of a BGP VPNv4 route received from a peer or peer group to a VPN instance address.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **next-hop-vpn**

   By default, the device does not change the next hop attribute of a received BGP VPNv4 route, and the next hop belongs to the public network.

Configuring route replication

Configuring the public instance

About this task

Configure the public instance to enable the mutual access between public network and private network users.

Restrictions and guidelines

In an MPLS L3VPN network, for the public network and the VPN network to communicate with each other through route target matching, perform the following tasks:

- Configure matching route targets for the public instance and VPN instance.
- Use the **route-replicate enable** command in BGP instance view to enable mutual BGP route replication between the public and VPN instances.

Procedure

1. Enter system view.

   system-view
2. Enter public instance view.

   ip public-instance
3. Configure an RD for the public instance.

   route-distinguisher *route-distinguisher*

   By default, no RD is configured for the public instance.
4. Configure a route target for the public instance.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, no route target is configured for the public instance.
5. Set the maximum number of active route prefixes supported by the public instance.

   routing-table limit *number* { *warn-threshold* | **simply-alert** }

   By default, no limit is set for the number of active route prefixes supported by the public instance.
6. Enter public instance IPv4 address family view.

   address-family ipv4
7. Configure a route target for the IPv4 address family of the public instance.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, no route target is configured for the IPv4 address family of the public instance.
8. Apply an import routing policy to the public instance.

   import route-policy *route-policy*

   By default, all routes matching the import target attribute are accepted.
9. Apply an export routing policy to the public instance.

   export route-policy *route-policy*

   By default, routes to be advertised are not filtered.
10. Set the maximum number of active route prefixes supported by the public instance.

    routing-table limit *number* { *warn-threshold* | **simply-alert** }

By default, no limit is set for the number of active route prefixes supported by the public instance.

The configuration in public instance IPv4 address family view takes precedence over the configuration in public instance view.

Configuring route replication for public and VPN instances

About this task

In a BGP/MPLS L3VPN network, only VPN instances that have matching route targets can communicate with each other.

The route replication feature provides the following functions:

- Enables a VPN instance to communicate with the public network or other VPN instances by replicating routes from the public instance or other VPN instances.
- Enables the public network to communicate with a VPN instance by replicating routes from the VPN instance to the public instance.

In an intelligent traffic control network, traffic of different tenants is assigned to different VPNs. To enable the tenants to communicate with the public network, configure this feature to replicate routes from the public instance to the VPN instances.

VLINK direct routes are generated based on ARP entries learned by interfaces. The **route-replicate from vpn-instance protocol direct** and **route-replicate from public protocol direct** commands replicate VLINK direct routes, but the VLINK direct routes cannot be added to the FIB, causing traffic forwarding failures. To address this issue, you can specify the **vlink-direct** keyword to replicate VLINK direct routes and add the VLINK direct routes to the FIB.

Configuring a VPN instance to replicate routes from the public instance or another VPN instance

1. Enter system view.

   system-view
2. Enter VPN instance view.

   ip vpn-instance *vpn-instance-name*
3. Enter VPN instance IPv4 address family view.

   address-family ipv4
4. Replicate routes from the public instance or other VPN instances.

   route-replicate from { **public** | **vpn-instance** *vpn-instance-name* } **protocol** { **bgp** *as-number* | **direct** | **static** | { **isis** | **ospf** | **rip** } *process-id* | **vlink-direct** } [**advertise**] [**route-policy** *route-policy-name*]

   By default, a VPN instance cannot replicate routes from the public instance or other VPN instances.

Configuing the public instance to replicate routes from a VPN instance

1. Enter system view.

   system-view
2. Enter public instance view.

   ip public-instance
3. Enter public instance IPv4 address family view.

   address-family ipv4
4. Replicate routes from a VPN instance to the public instance.

route-replicate from vpn-instance *vpn-instance-name* **protocol** { **bgp** *as-number* | **direct** | **static** | { **isis** | **ospf** | **rip** } *process-id* | **vlink-direct** } [**advertise**] [**route-policy** *route-policy-name*]

By default, the public instance cannot replicate routes from VPN instances.

Configuring BGP route replication between public and VPN instances

About this task

In traffic cleaning scenarios, traffic between the public and private networks are filtered by firewalls and traffic of different tenants is assigned to different VPNs. To enable the tenants to communicate with the public network under the protection of firewalls, BGP route replication between public and VPN instances is required.

By default, only VPN instances that have matching route targets can redistribute BGP routes from each other, while the public instance and VPN instances cannot. After you configure this feature, the public instance and VPN instances that have matching route targets can replicate BGP routes from each other, enabling communication between the public network and VPN users.

This feature also replicates the BGP route attributes, so that the device can select proper forwarding paths according to the route attributes.

Restrictions and guidelines

After this feature is enabled, the public network and VPNs cannot be isolated. Configure this feature only in specific scenarios, for example, the traffic cleaning scenario.

To use this feature to implement IPv4 route replication between the public instance and a VPN instance, make sure the VPN instance and the BGP IPv4 unicast address family have been created.

Procedure

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enable BGP route replication between public and VPN instances.

   route-replicate enable

   By default, BGP route replication between public and VPN instances is disabled.

Configuring route re-origination

About this task

By default, BGP routes in different VPN instances are isolated. In some networks, a device might need to advertise routes across VPN instances or advertise the routes in a VPN instance through other VPN instances to hide the routing information of the VPN instance.

After you configure this feature, the current VPN instance will re-originate the BGP unicast routes from other VPN instances that have the same route target as the current VPN instance. Locally redistributed routes (such as the IGP routes redistributed by using the **import-route** command) will not be re-originated. The re-originated routes can be advertised to BGP peers.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*

4. Enter BGP-VPN IPv4 unicast address family view.

   address-family ipv4 [**unicast**]

5. Re-originate BGP unicast routes from other VPN instances.

   advertise route-reoriginate [**route-policy** *route-policy-name*] [**replace-rt**]

   By default, a VPN instance cannot re-originate BGP unicast routes from other VPN instances.

Enabling ECMP VPN route redistribution

About this task

For multiple routes that have the same prefix and RD, a VPN instance redistributes only the optimal route into its routing table by default. This feature enables a VPN instance to redistribute all routes that have the same prefix and RD into its routing table to perform load sharing or MPLS L3VPN FRR.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter a BGP address family view.
   - Enter BGP IPv4 unicast address family view.

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view:

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]

4. Enable ECMP VPN route redistribution.

   vpn-route cross multipath

   By default, ECMP VPN route redistribution is disabled. If multiple routes have the same prefix and RD, a VPN instance redistributes only the optimal route into its routing table.

In BGP IPv4 unicast address family view,this command redistributes ECMP routes to the routing table of the public instance.

Enabling prioritized withdrawal of specific routes

About this task

This feature enables BGP to send the withdrawal messages of specific routes prior to other routes. This can achieve fast switchover of traffic on the specified routes to available routes to reduce the traffic interruption time.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP VPNv4 address family view.

   address-family vpnv4

4. Enable prioritized withdrawal of the routes that match the specified routing policy.

   update-first route-policy *route-policy-name*

   By default, BGP does not send the withdrawal messages of specific routes prior to other routes.

Enabling SNMP notifications for MPLS L3VPN

About this task

To report critical MPLS L3VPN events to an NMS, enable SNMP notifications for MPLS L3VPN. For MPLS L3VPN event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for MPLS L3VPN.

   snmp-agent trap enable l3vpn [**vrf-down** | **vrf-up**] *

   By default, SNMP notifications for MPLS L3VPN are enabled.

Enabling logging for BGP route flapping

About this task

This feature enables BGP to generate logs for BGP route flappings that trigger log generation. The generated logs are sent to the information center. For the logs to be output correctly, you must also configure information center on the device. For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter BGP VPNv4 address family view or BGP-VPN VPNv4 address family view.

   - Execute the following commands in sequence to enter BGP VPNv4 address family view:

     bgp *as-number* [**instance** *instance-name*]

     address-family vpnv4

   - Execute the following commands in sequence to enter BGP-VPN VPNv4 address family view:

     bgp *as-number* [**instance** *instance-name*]

     ip vpn-instance *vpn-instance-name*

     address-family vpnv4

3. Enable logging for BGP route flapping.

   log-route-flap *monitor-time monitor-count* [*log-count-limit* | **route-policy** *route-policy-name*] *

   By default, logging for BGP route flapping is disabled.

Display and maintenance commands for MPLS L3VPN

Resetting BGP connections

You can soft-reset or reset BGP sessions to apply new BGP configurations. A soft reset operation updates BGP routing information without tearing down BGP connections. A reset operation updates BGP routing information by tearing down, and then re-establishing BGP connections. Soft reset requires that BGP peers have route refresh capability.

Execute the following commands in user view to soft reset or reset BGP connections:

Task	Command
Soft-reset BGP sessions for the VPNv4 address family.	**refresh bgp** [**instance** *instance-name*] { *ipv4-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **vpnv4** [**vpn-instance** *vpn-instance-name*]
Soft-reset BGP sessions for the BGP IPv4 RT filter family.	**refresh bgp** [**instance** *instance-name*] { *ipv4-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **ipv4 rtfilter**
Reset BGP sessions for the VPNv4 address family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | **all** | **external** | **internal** | **group** *group-name* } **vpnv4** [**vpn-instance** *vpn-instance-name*]
Reset BGP sessions for the BGP IPv4 RT filter family.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | **all** | **external** | **internal** | **group** *group-name* } **ipv4 rtfilter**

For more information about the **refresh bgp vpnv4** and **reset bgp vpnv4** commands, see *Layer 3—IP Routing Command Reference*.

Displaying MPLS L3VPN information

Execute the following commands in any view to display MPLS L3VPN:

Task	Command
Display BGP VPNv4 route dampening parameters.	**display bgp** [**instance** *instance-name*] **dampening parameter vpnv4**
Display BGP RT filter peer group information.	**display bgp** [**instance** *instance-name*] **group ipv4 rtfilter** [**group-name** *group-name*]
Display BGP VPNv4 peer group information.	**display bgp** [**instance** *instance-name*] **group vpnv4** [**vpn-instance** *vpn-instance-name*] [**group-name** *group-name*]
Display BGP RT filter information.	**display bgp** [**instance** *instance-name*] **ipv4 rtfilter** [**peer** *ipv4-address* [**statistics**] |

Task	Command
	statistics]
Display BGP RT filter peer information.	**display bgp** [**instance** *instance-name*] **peer ipv4 rtfilter** [*ipv4-address mask-length* | { *ipv4-address* | **group-name** *group-name* } **log-info** | [*ipv4-address*] **verbose**]
Display BGP VPNv4 peer information.	**display bgp** [**instance** *instance-name*] **peer vpnv4** [**vpn-instance** *vpn-instance-name*] [*ipv4-address mask-length* | { *ipv4-address* | **group-name** *group-name* } **log-info** | [*ipv4-address*] **verbose**]
Display information about dampened BGP VPNv4 routes.	**display bgp** [**instance** *instance-name*] **routing-table dampened vpnv4**
Display BGP VPNv4 route flapping information.	**display bgp** [**instance** *instance-name*] **routing-table flap-info vpnv4** [*ipv4-address* [{ *mask* | *mask-length* } [**longest-match**]] | **as-path-acl** *as-path-acl-number*]
Display incoming labels for BGP IPv4 unicast routes.	**display bgp** [**instance** *instance-name*] **routing-table ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] **inlabel**
Display outgoing labels for BGP IPv4 unicast routes.	**display bgp** [**instance** *instance-name*] **routing-table ipv4** [**unicast**] [**vpn-instance** *vpn-instance-name*] **outlabel**
Display BGP RT filter routing information.	**display bgp** [**instance** *instance-name*] **routing-table ipv4 rtfilter** [**default-rt** [**advertise-info** | **as-path** | **cluster-list**] | [**origin-as** *as-number*] [*route-target* [**advertise-info** | **as-path** | **cluster-list**]] | **peer** *ipv4-address* { **advertised-routes** | **received-routes** } [**default-rt** | [**origin-as** *as-number*] [*route-target*] | **statistics**] | **statistics**]
Display BGP VPNv4 routes.	**display bgp** [**instance** *instance-name*] **routing-table vpnv4** [[**route-distinguisher** *route-distinguisher*] [*ipv4-address* [{ *mask-length* | *mask* } [**longest-match**]] | *ipv4-address* [*mask-length* | *mask*] **advertise-info** | *ipv4-address* [*mask-length* | *mask*] { **as-path** | **cluster-list** | **community** | **ext-community** } | [**vpn-instance** *vpn-instance-name*] **peer** *ipv4-address* { **advertised-routes** | **received-routes** } [*ipv4-address* [*mask-length* | *mask*] | **statistics**] | **statistics**] **display bgp** [**instance** *instance-name*] **routing-table vpnv4** [**route-distinguisher** *route-distinguisher*] **as-path-acl** { *as-path-acl-number* | *as-path-acl-name* } **display bgp** [**instance** *instance-name*] **routing-table vpnv4** [**route-distinguisher**

Task	Command
	route-distinguisher] [**statistics**] **community** [*community-number*&<1-32> | *aa:nn*&<1-32>] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table vpnv4** [**route-distinguisher** *route-distinguisher*] [**statistics**] **community-list** *basic-community-list-number* | *adv-community-list-number* | *comm-list-name* } [**whole-match**] **display bgp** [**instance** *instance-name*] **routing-table vpnv4** [**route-distinguisher** *route-distinguisher*] [**statistics**] **ext-community** [**bandwidth** *link-bandwidth-value* | **rt** *route-target* | **soo** *site-of-origin*]&<1-32> [**whole-match**]
Display incoming labels for BGP VPNv4 routes.	**display bgp** [**instance** *instance-name*] **routing-table vpnv4 inlabel**
Display outgoing labels for BGP VPNv4 routes.	**display bgp** [**instance** *instance-name*] **routing-table vpnv4 outlabel**
Display BGP IPv4 RT filter address family update group information.	**display bgp** [**instance** *instance-name*] **update-group ipv4 rtfilter** [*ipv4-address*]
Display BGP VPNv4 address family update group information.	**display bgp** [**instance** *instance-name*] **update-group vpnv4** [**vpn-instance** *vpn-instance-name*] [*ipv4-address*]
Display the FIB of a VPN instance.	**display fib vpn-instance** *vpn-instance-name*
Display FIB entries that match the specified destination IP address in the specified VPN instance.	**display fib vpn-instance** *vpn-instance-name ip-address* [*mask-length* | *mask*]
Display the routing table for a VPN instance.	**display ip routing-table vpn-instance** *vpn-instance-name* [**statistics** | **verbose**]
Display information about a specific or all VPN instances.	**display ip vpn-instance** [**instance-name** *vpn-instance-name*]
Display OSPF sham link information.	**display ospf** [*process-id*] **sham-link** [**area** *area-id*]
Clear BGP VPNv4 route dampening information and release dampened routes.	**reset bgp** [**instance** *instance-name*] **dampening vpnv4** [*ipv4-address* [*mask* | *mask-length*]]
Clear BGP VPNv4 route flapping statistics.	**reset bgp** [**instance** *instance-name*] **flap-info vpnv4** [*ipv4-address* [*mask* | *mask-length*] | **as-path-acl** *as-path-acl-number* | **peer** *ipv4-address* [*mask-length*]]

For more information about the **display ip routing-table**, **display bgp group vpnv4**, **display bgp peer vpnv4**, and **display bgp update-group vpnv4** commands, see *Layer 3—IP Routing Command Reference*.

MPLS L3VPN configuration examples

Example: Configuring basic MPLS L3VPN

Network configuration

CE 1 and CE 3 belong to VPN 1. CE 2 and CE 4 belong to VPN 2.

VPN 1 uses route target attribute 111:1. VPN 2 uses route target attribute 222:2. Users of different VPNs cannot access each other.

EBGP is used to exchange VPN routing information between CE and PE.

PEs use OSPF to communicate with each other and use MP-IBGP to exchange VPN routing information.

Figure 26 Network diagram

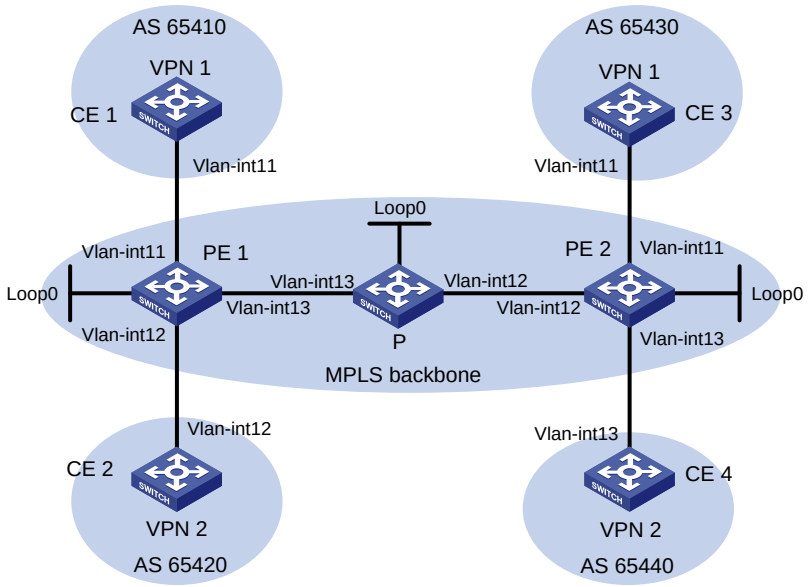


Table 12 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Vlan-int11	10.1.1.1/24	P	Loop0	2.2.2.9/32
PE 1	Loop0	1.1.1.9/32		Vlan-int12	172.2.1.1/24
	Vlan-int11	10.1.1.2/24		Vlan-int13	172.1.1.2/24
	Vlan-int13	172.1.1.1/24	PE 2	Loop0	3.3.3.9/32
	Vlan-int12	10.2.1.2/24		Vlan-int12	172.2.1.2/24
CE 2	Vlan-int12	10.2.1.1/24		Vlan-int11	10.3.1.2/24
CE 3	Vlan-int11	10.3.1.1/24		Vlan-int13	10.4.1.2/24
CE 4	Vlan-int13	10.4.1.1/24			

Procedure

1. Configure an IGP on the MPLS backbone to ensure IP connectivity within the backbone:

   # Configure PE 1.

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 1.1.1.9 32
[PE1-LoopBack0] quit
[PE1] interface vlan-interface 13
[PE1-Vlan-interface13] ip address 172.1.1.1 24
[PE1-Vlan-interface13] quit
[PE1] ospf
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

   # Configure the P device.

```
<P> system-view
[P] interface loopback 0
[P-LoopBack0] ip address 2.2.2.9 32
[P-LoopBack0] quit
[P] interface vlan-interface 13
[P-Vlan-interface13] ip address 172.1.1.2 24
[P-Vlan-interface13] quit
[P] interface vlan-interface 12
[P-Vlan-interface12] ip address 172.2.1.1 24
[P-Vlan-interface12] quit
[P] ospf
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

   # Configure PE 2.

```
<PE2> system-view
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 3.3.3.9 32
[PE2-LoopBack0] quit
[PE2] interface vlan-interface 12
[PE2-Vlan-interface12] ip address 172.2.1.2 24
[PE2-Vlan-interface12] quit
[PE2] ospf
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

Execute the **display ospf peer** command to verify that OSPF adjacencies in Full state have been established between PE 1, P, and PE 2. Execute the **display ip routing-table** command to verify that the PEs have learned the routes to the loopback interfaces of each other. (Details not shown.)

2. Configure basic MPLS and MPLS LDP on the MPLS backbone to establish LDP LSPs:

Configure PE 1.

```
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls ldp
[PE1-ldp] quit
[PE1] interface vlan-interface 13
[PE1-Vlan-interface13] mpls enable
[PE1-Vlan-interface13] mpls ldp enable
[PE1-Vlan-interface13] quit
```

Configure the P device.

```
[P] mpls lsr-id 2.2.2.9
[P] mpls ldp
[P-ldp] quit
[P] interface vlan-interface 13
[P-Vlan-interface13] mpls enable
[P-Vlan-interface13] mpls ldp enable
[P-Vlan-interface13] quit
[P] interface vlan-interface 12
[P-Vlan-interface12] mpls enable
[P-Vlan-interface12] mpls ldp enable
[P-Vlan-interface12] quit
```

Configure PE 2.

```
[PE2] mpls lsr-id 3.3.3.9
[PE2] mpls ldp
[PE2-ldp] quit
[PE2] interface vlan-interface 12
[PE2-Vlan-interface12] mpls enable
[PE2-Vlan-interface12] mpls ldp enable
[PE2-Vlan-interface12] quit
```

Execute the **display mpls ldp peer** command to verify that LDP sessions in Operational state have been established between PE 1, P, and PE 2. Execute the **display mpls ldp lsp** command to verify that the LSPs have been established by LDP. (Details not shown.)

3. Configure VPN instances on PEs:

Configure PE 1.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 111:1
[PE1-vpn-instance-vpn1] quit
[PE1] ip vpn-instance vpn2
[PE1-vpn-instance-vpn2] route-distinguisher 100:2
[PE1-vpn-instance-vpn2] vpn-target 222:2
[PE1-vpn-instance-vpn2] quit
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip binding vpn-instance vpn1
```

```
[PE1-Vlan-interface11] ip address 10.1.1.2 24
[PE1-Vlan-interface11] quit
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance vpn2
[PE1-Vlan-interface12] ip address 10.2.1.2 24
[PE1-Vlan-interface12] quit
```

Configure PE 2.

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 200:1
[PE2-vpn-instance-vpn1] vpn-target 111:1
[PE2-vpn-instance-vpn1] quit
[PE2] ip vpn-instance vpn2
[PE2-vpn-instance-vpn2] route-distinguisher 200:2
[PE2-vpn-instance-vpn2] vpn-target 222:2
[PE2-vpn-instance-vpn2] quit
[PE2] interface vlan-interface 11
[PE2-Vlan-interface11] ip binding vpn-instance vpn1
[PE2-Vlan-interface11] ip address 10.3.1.2 24
[PE2-Vlan-interface11] quit
[PE2] interface vlan-interface 13
[PE2-Vlan-interface13] ip binding vpn-instance vpn2
[PE2-Vlan-interface13] ip address 10.4.1.2 24
[PE2-Vlan-interface13] quit
```

Configure IP addresses for the CEs according to Figure 26. (Details not shown.)

Execute the **display ip vpn-instance** command on the PEs to display the configuration of the VPN instance.

Use the **ping** command on the PEs to verify that the PEs can ping their attached CEs.

4. Establish EBGP peer relationships between PEs and CEs, and redistribute VPN routes into BGP:

Configure CE 1.

```
<CE1> system-view
[CE1] bgp 65410
[CE1-bgp-default] peer 10.1.1.2 as-number 100
[CE1-bgp-default] address-family ipv4 unicast
[CE1-bgp-default-ipv4] peer 10.1.1.2 enable
[CE1-bgp-default-ipv4] import-route direct
[CE1-bgp-default-ipv4] quit
[CE1-bgp-default] quit
```

Configure the other three CEs in the same way that CE 1 is configured. (Details not shown.)

Configure PE 1.

```
[PE1] bgp 100
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 10.1.1.1 as-number 65410
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] peer 10.1.1.1 enable
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] ip vpn-instance vpn2
```

```
[PE1-bgp-default-vpn2] peer 10.2.1.1 as-number 65420
[PE1-bgp-default-vpn2] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn2] peer 10.2.1.1 enable
[PE1-bgp-default-ipv4-vpn2] quit
[PE1-bgp-default-vpn2] quit
[PE1-bgp-default] quit
```

Configure PE 2 in the same way that PE 1 is configured. (Details not shown.)

Execute the **display bgp peer ipv4 vpn-instance** command on the PEs to verify that a BGP peer relationship in Established state has been established between a PE and a CE. (Details not shown.)

5. Establish an MP-IBGP peer relationship between PEs:

Configure PE 1.

```
[PE1] bgp 100
[PE1-bgp-default] peer 3.3.3.9 as-number 100
[PE1-bgp-default] peer 3.3.3.9 connect-interface loopback 0
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 3.3.3.9 enable
[PE1-bgp-default-vpnv4] quit
[PE1-bgp-default] quit
```

Configure PE 2.

```
[PE2] bgp 100
[PE2-bgp-default] peer 1.1.1.9 as-number 100
[PE2-bgp-default] peer 1.1.1.9 connect-interface loopback 0
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] peer 1.1.1.9 enable
[PE2-bgp-default-vpnv4] quit
[PE2-bgp-default] quit
```

Execute the **display bgp peer vpnv4** command on the PEs to verify that a BGP peer relationship in Established state has been established between the PEs. (Details not shown.)

Verifying the configuration

Execute the **display ip routing-table vpn-instance** command on the PEs.

```
[PE1] display ip routing-table vpn-instance vpn1

Destinations : 13        Routes : 13

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.1.1.0/24         Direct 0    0            10.1.1.2        Vlan11
10.1.1.0/32         Direct 0    0            10.1.1.2        Vlan11
10.1.1.2/32         Direct 0    0            127.0.0.1       InLoop0
10.1.1.255/32       Direct 0    0            10.1.1.2        Vlan11
10.3.1.0/24         BGP    255  0            3.3.3.9         Vlan13
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
```

```
255.255.255.255/32  Direct 0    0           127.0.0.1       InLoop0
```

The output shows that PE 1 has a route to the remote CE. Output on PE 2 is similar.

Verify that CEs of the same VPN can ping each other, whereas those of different VPNs cannot. For example, CE 1 can ping CE 3 (10.3.1.1) but cannot ping CE 4 (10.4.1.1). (Details not shown.)

Example: Configuring a hub-spoke network

Network configuration

The Spoke-CEs cannot communicate directly. They can communicate only through Hub-CE.

Configure EBGP between the Spoke-CEs and Spoke-PEs and between Hub-CE and Hub-PE to exchange VPN routing information.

Configure OSPF between the Spoke-PEs and Hub-PE to implement communication between the PEs, and configure MP-IBGP between them to exchange VPN routing information.

Figure 27 Network diagram

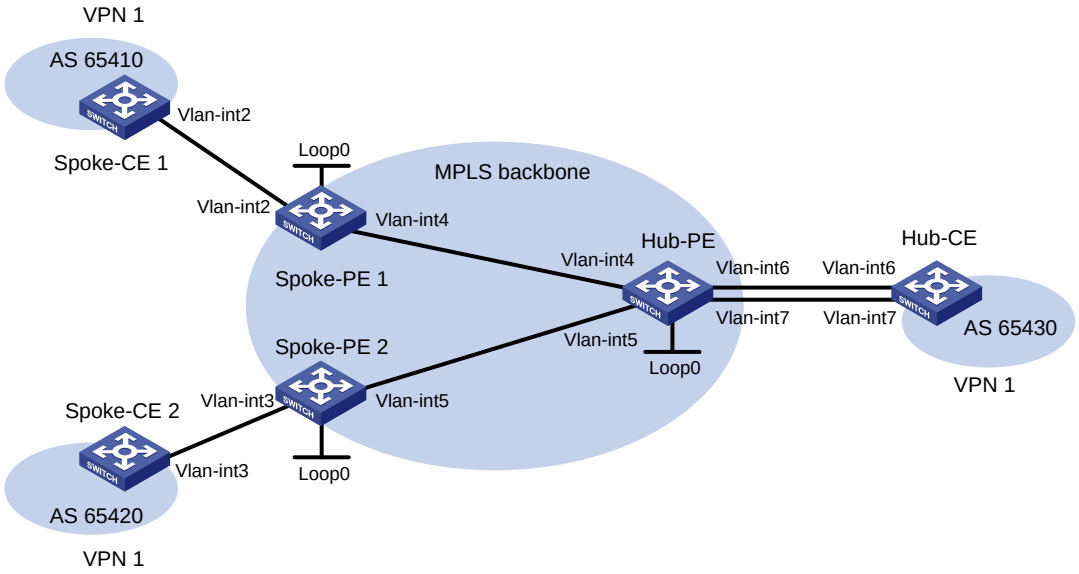


Table 13 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Spoke-CE 1	Vlan-int2	10.1.1.1/24	Hub-CE	Vlan-int6	10.3.1.1/24
Spoke-PE 1	Loop0	1.1.1.9/32		Vlan-int7	10.4.1.1/24
	Vlan-int2	10.1.1.2/24	Hub-PE	Loop0	2.2.2.9/32
	Vlan-int4	172.1.1.1/24		Vlan-int4	172.1.1.2/24
Spoke-CE 2	Vlan-int3	10.2.1.1/24		Vlan-int5	172.2.1.2/24
Spoke-PE 2	Loop0	3.3.3.9/32		Vlan-int6	10.3.1.2/24
	Vlan-int3	10.2.1.2/24		Vlan-int7	10.4.1.2/24
	Vlan-int5	172.2.1.1/24			

Procedure

1. Configure an IGP on the MPLS backbone to ensure IP connectivity within the backbone:

Configure Spoke-PE 1.

```
<Spoke-PE1> system-view
[Spoke-PE1] interface loopback 0
[Spoke-PE1-LoopBack0] ip address 1.1.1.9 32
[Spoke-PE1-LoopBack0] quit
[Spoke-PE1] interface vlan-interface 4
[Spoke-PE1-Vlan-interface4] ip address 172.1.1.1 24
[Spoke-PE1-Vlan-interface4] quit
[Spoke-PE1] ospf
[Spoke-PE1-ospf-1] area 0
[Spoke-PE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[Spoke-PE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[Spoke-PE1-ospf-1-area-0.0.0.0] quit
[Spoke-PE1-ospf-1] quit
```

Configure Spoke-PE 2.

```
<Spoke-PE2> system-view
[Spoke-PE2] interface loopback 0
[Spoke-PE2-LoopBack0] ip address 3.3.3.9 32
[Spoke-PE2-LoopBack0] quit
[Spoke-PE2] interface vlan-interface 5
[Spoke-PE2-Vlan-interface5] ip address 172.2.1.1 24
[Spoke-PE2-Vlan-interface5] quit
[Spoke-PE2] ospf
[Spoke-PE2-ospf-1] area 0
[Spoke-PE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[Spoke-PE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[Spoke-PE2-ospf-1-area-0.0.0.0] quit
[Spoke-PE2-ospf-1] quit
```

Configure Hub-PE.

```
<Hub-PE> system-view
[Hub-PE] interface loopback 0
[Hub-PE-LoopBack0] ip address 2.2.2.9 32
[Hub-PE-LoopBack0] quit
[Hub-PE] interface vlan-interface 4
[Hub-PE-Vlan-interface4] ip address 172.1.1.2 24
[Hub-PE-Vlan-interface4] quit
[Hub-PE] interface vlan-interface 5
[Hub-PE-Vlan-interface5] ip address 172.2.1.2 24
[Hub-PE-Vlan-interface5] quit
[Hub-PE] ospf
[Hub-PE-ospf-1] area 0
[Hub-PE-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[Hub-PE-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[Hub-PE-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[Hub-PE-ospf-1-area-0.0.0.0] quit
[Hub-PE-ospf-1] quit
```

Execute the **`display ospf peer`** command on the devices to verify that OSPF adjacencies in Full state have been established between Spoke-PE 1, Spoke-PE 2, and Hub-PE. Execute

the **display ip routing-table** command on the devices to verify that the PEs have learned the routes to the loopback interfaces of each other. (Details not shown.)

2. Configure basic MPLS and MPLS LDP on the MPLS backbone to establish LDP LSPs:

Configure Spoke-PE 1.

```
[Spoke-PE1] mpls lsr-id 1.1.1.9
[Spoke-PE1] mpls ldp
[Spoke-PE1-ldp] quit
[Spoke-PE1] interface vlan-interface 4
[Spoke-PE1-Vlan-interface4] mpls enable
[Spoke-PE1-Vlan-interface4] mpls ldp enable
[Spoke-PE1-Vlan-interface4] quit
```

Configure Spoke-PE 2.

```
[Spoke-PE2] mpls lsr-id 3.3.3.9
[Spoke-PE2] mpls ldp
[Spoke-PE2-ldp] quit
[Spoke-PE2] interface vlan-interface 5
[Spoke-PE2-Vlan-interface5] mpls enable
[Spoke-PE2-Vlan-interface5] mpls ldp enable
[Spoke-PE2-Vlan-interface5] quit
```

Configure Hub-PE.

```
[Hub-PE] mpls lsr-id 2.2.2.9
[Hub-PE] mpls ldp
[Hub-PE-ldp] quit
[Hub-PE] interface vlan-interface 4
[Hub-PE-Vlan-interface4] mpls enable
[Hub-PE-Vlan-interface4] mpls ldp enable
[Hub-PE-Vlan-interface4] quit
[Hub-PE] interface vlan-interface 5
[Hub-PE-Vlan-interface5] mpls enable
[Hub-PE-Vlan-interface5] mpls ldp enable
[Hub-PE-Vlan-interface5] quit
```

Execute the **display mpls ldp peer** command on the devices to verify that LDP sessions in Operational state have been established between Spoke-PE 1, Spoke-PE 2, and Hub-PE. Execute the **display mpls ldp lsp** command on the devices to verify that the LSPs have been established by LDP. (Details not shown.)

3. Configure VPN instances on the Spoke-PEs and Hub-PE:

Configure Spoke-PE 1.

```
[Spoke-PE1] ip vpn-instance vpn1
[Spoke-PE1-vpn-instance-vpn1] route-distinguisher 100:1
[Spoke-PE1-vpn-instance-vpn1] vpn-target 111:1 import-extcommunity
[Spoke-PE1-vpn-instance-vpn1] vpn-target 222:2 export-extcommunity
[Spoke-PE1-vpn-instance-vpn1] quit
[Spoke-PE1] interface vlan-interface 2
[Spoke-PE1-Vlan-interface2] ip binding vpn-instance vpn1
[Spoke-PE1-Vlan-interface2] ip address 10.1.1.2 24
[Spoke-PE1-Vlan-interface2] quit
```

Configure Spoke-PE 2.

```
[Spoke-PE2] ip vpn-instance vpn1
```

```
[Spoke-PE2-vpn-instance-vpn1] route-distinguisher 100:2
[Spoke-PE2-vpn-instance-vpn1] vpn-target 111:1 import-extcommunity
[Spoke-PE2-vpn-instance-vpn1] vpn-target 222:2 export-extcommunity
[Spoke-PE2-vpn-instance-vpn1] quit
[Spoke-PE2] interface vlan-interface 3
[Spoke-PE2-Vlan-interface3] ip binding vpn-instance vpn1
[Spoke-PE2-Vlan-interface3] ip address 10.2.1.2 24
[Spoke-PE2-Vlan-interface3] quit
```

Configure Hub-PE.

```
[Hub-PE] ip vpn-instance vpn1-in
[Hub-PE-vpn-instance-vpn1-in] route-distinguisher 100:3
[Hub-PE-vpn-instance-vpn1-in] vpn-target 222:2 import-extcommunity
[Hub-PE-vpn-instance-vpn1-in] quit
[Hub-PE] ip vpn-instance vpn1-out
[Hub-PE-vpn-instance-vpn1-out] route-distinguisher 100:4
[Hub-PE-vpn-instance-vpn1-out] vpn-target 111:1 export-extcommunity
[Hub-PE-vpn-instance-vpn1-out] quit
[Hub-PE] interface vlan-interface 6
[Hub-PE-Vlan-interface6] ip binding vpn-instance vpn1-in
[Hub-PE-Vlan-interface6] ip address 10.3.1.2 24
[Hub-PE-Vlan-interface6] quit
[Hub-PE] interface vlan-interface 7
[Hub-PE-Vlan-interface7] ip binding vpn-instance vpn1-out
[Hub-PE-Vlan-interface7] ip address 10.4.1.2 24
[Hub-PE-Vlan-interface7] quit
```

Configure IP addresses for the CEs according to Figure 27Table 13. (Details not shown.)

Execute the **display ip vpn-instance** command on the PEs to display the configuration of the VPN instance.

Use the **ping** command on the PEs to verify that the PEs can ping their attached CEs.

4. Establish EBGP peer relationships between the PEs and CEs, and redistribute VPN routes into BGP:

Configure Spoke-CE 1.

```
<Spoke-CE1> system-view
[Spoke-CE1] bgp 65410
[Spoke-CE1-bgp-default] peer 10.1.1.2 as-number 100
[Spoke-CE1-bgp-default] address-family ipv4
[Spoke-CE1-bgp-default-ipv4] peer 10.1.1.2 enable
[Spoke-CE1-bgp-default-ipv4] import-route direct
[Spoke-CE1-bgp-default-ipv4] quit
[Spoke-CE1-bgp-default] quit
```

Configure Spoke-CE 2.

```
<Spoke-CE2> system-view
[Spoke-CE2] bgp 65420
[Spoke-CE2-bgp-default] peer 10.2.1.2 as-number 100
[Spoke-CE2-bgp-default] address-family ipv4
[Spoke-CE2-bgp-default-ipv4] peer 10.2.1.2 enable
[Spoke-CE2-bgp-default-ipv4] import-route direct
[Spoke-CE2-bgp-default-ipv4] quit
```

```
[Spoke-CE2-bgp-default] quit
```

Configure Hub-CE.

```
<Hub-CE> system-view
[Hub-CE] bgp 65430
[Hub-CE-bgp-default] peer 10.3.1.2 as-number 100
[Hub-CE-bgp-default] peer 10.4.1.2 as-number 100
[Hub-CE-bgp-default] address-family ipv4
[Hub-CE-bgp-default-ipv4] peer 10.3.1.2 enable
[Hub-CE-bgp-default-ipv4] peer 10.4.1.2 enable
[Hub-CE-bgp-default-ipv4] import-route direct
[Hub-CE-bgp-default-ipv4] quit
[Hub-CE-bgp-default] quit
```

Configure Spoke-PE 1.

```
[Spoke-PE1] bgp 100
[Spoke-PE1-bgp-default] ip vpn-instance vpn1
[Spoke-PE1-bgp-default-vpn1] peer 10.1.1.1 as-number 65410
[Spoke-PE1-bgp-default-vpn1] address-family ipv4
[Spoke-PE1-bgp-default-ipv4-vpn1] peer 10.1.1.1 enable
[Spoke-PE1-bgp-default-ipv4-vpn1] quit
[Spoke-PE1-bgp-default-vpn1] quit
[Spoke-PE1-bgp-default] quit
```

Configure Spoke-PE 2.

```
[Spoke-PE2] bgp 100
[Spoke-PE2-bgp-default] ip vpn-instance vpn1
[Spoke-PE2-bgp-default-vpn1] peer 10.2.1.1 as-number 65420
[Spoke-PE2-bgp-default-vpn1] address-family ipv4
[Spoke-PE2-bgp-default-ipv4-vpn1] peer 10.2.1.1 enable
[Spoke-PE2-bgp-default-ipv4-vpn1] quit
[Spoke-PE2-bgp-default-vpn1] quit
[Spoke-PE2-bgp-default] quit
```

Configure Hub-PE.

```
[Hub-PE] bgp 100
[Hub-PE-bgp-default] ip vpn-instance vpn1-in
[Hub-PE-bgp-default-vpn1-in] peer 10.3.1.1 as-number 65430
[Hub-PE-bgp-default-vpn1-in] address-family ipv4
[Hub-PE-bgp-default-ipv4-vpn1-in] peer 10.3.1.1 enable
[Hub-PE-bgp-default-ipv4-vpn1-in] quit
[Hub-PE-bgp-default-vpn1-in] quit
[Hub-PE-bgp-default] ip vpn-instance vpn1-out
[Hub-PE-bgp-default-vpn1-out] peer 10.4.1.1 as-number 65430
[Hub-PE-bgp-default-vpn1-out] address-family ipv4
[Hub-PE-bgp-default-ipv4-vpn1-out] peer 10.4.1.1 enable
[Hub-PE-bgp-default-ipv4-vpn1-out] peer 10.4.1.1 allow-as-loop 2
[Hub-PE-bgp-default-ipv4-vpn1-out] quit
[Hub-PE-bgp-default-vpn1-out] quit
[Hub-PE-bgp-default] quit
```

Execute the **display bgp peer ipv4 vpn-instance** command on the PEs to verify that a BGP peer relationship in Established state has been established between a PE and a CE. (Details not shown.)

5. Establish an MP-IBGP peer relationship between the Spoke-PEs and Hub-PE:

Configure Spoke-PE 1.

```
[Spoke-PE1] bgp 100
[Spoke-PE1-bgp-default] peer 2.2.2.9 as-number 100
[Spoke-PE1-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[Spoke-PE1-bgp-default] address-family vpnv4
[Spoke-PE1-bgp-default-vpnv4] peer 2.2.2.9 enable
[Spoke-PE1-bgp-default-vpnv4] quit
[Spoke-PE1-bgp-default] quit
```

Configure Spoke-PE 2.

```
[Spoke-PE2] bgp 100
[Spoke-PE2-bgp-default] peer 2.2.2.9 as-number 100
[Spoke-PE2-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[Spoke-PE2-bgp-default] address-family vpnv4
[Spoke-PE2-bgp-default-vpnv4] peer 2.2.2.9 enable
[Spoke-PE2-bgp-default-vpnv4] quit
[Spoke-PE2-bgp-default] quit
```

Configure Hub-PE.

```
[Hub-PE] bgp 100
[Hub-PE-bgp-default] peer 1.1.1.9 as-number 100
[Hub-PE-bgp-default] peer 1.1.1.9 connect-interface loopback 0
[Hub-PE-bgp-default] peer 3.3.3.9 as-number 100
[Hub-PE-bgp-default] peer 3.3.3.9 connect-interface loopback 0
[Hub-PE-bgp-default] address-family vpnv4
[Hub-PE-bgp-default-vpnv4] peer 1.1.1.9 enable
[Hub-PE-bgp-default-vpnv4] peer 3.3.3.9 enable
[Hub-PE-bgp-default-vpnv4] quit
[Hub-PE-bgp-default] quit
```

Execute the **display bgp peer vpnv4** command on the PEs to verify that a BGP peer relationship in Established state has been established between the PEs. (Details not shown.)

Verifying the configuration

Execute the **display ip routing-table vpn-instance** command on the PEs to display the routes to the CEs. This example uses Spoke-PE 1 to verify that the next hop of the route from a Spoke-PE to its connected Spoke-CE is Hub-PE.

```
[Spoke-PE1] display ip routing-table vpn-instance vpn1

Destinations : 15        Routes : 15

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
10.1.1.0/24        Direct  0   0           10.1.1.2        Vlan2
10.1.1.0/32        Direct  0   0           10.1.1.2        Vlan2
10.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
10.1.1.255/32      Direct  0   0           10.1.1.2        Vlan2
10.2.1.0/24        BGP     255 0           2.2.2.9         Vlan4
```

```
10.3.1.0/24         BGP     255 0           2.2.2.9         Vlan4
10.4.1.0/24         BGP     255 0           2.2.2.9         Vlan4
127.0.0.0/8         Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32        Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4         Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24        Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
```

Verify that Spoke-CE 1 and Spoke-CE 2 can ping each other. The TTL value indicates that traffic from Spoke-CE 1 to Spoke-CE 2 passes six hops (255-250+1) and is forwarded through Hub-CE. This example uses Spoke-CE 1 to verify their connectivity.

```
[Spoke-CE1] ping 10.2.1.1
Ping 10.2.1.1 (10.2.1.1): 56 data bytes, press CTRL+C to break
56 bytes from 10.2.1.1: icmp_seq=0 ttl=250 time=1.000 ms
56 bytes from 10.2.1.1: icmp_seq=1 ttl=250 time=2.000 ms
56 bytes from 10.2.1.1: icmp_seq=2 ttl=250 time=0.000 ms
56 bytes from 10.2.1.1: icmp_seq=3 ttl=250 time=1.000 ms
56 bytes from 10.2.1.1: icmp_seq=4 ttl=250 time=0.000 ms

--- Ping statistics for 10.2.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.000/0.800/2.000/0.748 ms
```

Example: Configuring MPLS L3VPN inter-AS option A

Network configuration

CE 1 and CE 2 belong to the same VPN. CE 1 accesses the network through PE 1 in AS 100, and CE 2 accesses the network through PE 2 in AS 200.

Configure MPLS L3VPN inter-AS option A, and use the VRF-to-VRF method to manage VPN routes.

Run OSPF on the MPLS backbone in each AS.

Figure 28 Network diagram

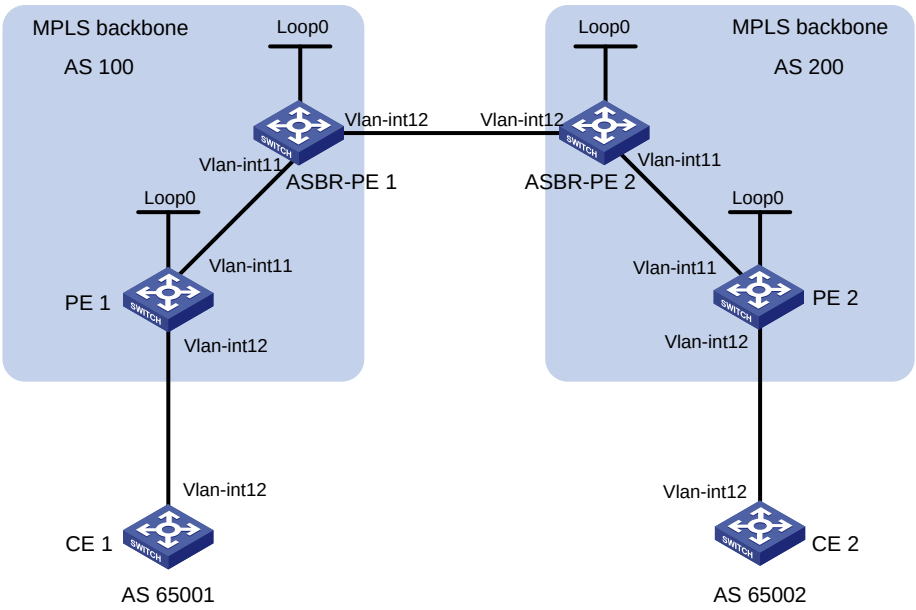


Table 14 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Vlan-int12	10.1.1.1/24	CE 2	Vlan-int12	10.2.1.1/24
PE 1	Loop0	1.1.1.9/32	PE 2	Loop0	4.4.4.9/32
	Vlan-int12	10.1.1.2/24		Vlan-int12	10.2.1.2/24
	Vlan-int11	172.1.1.2/24		Vlan-int11	162.1.1.2/24
ASBR-PE 1	Loop0	2.2.2.9/32	ASBR-PE 2	Loop0	3.3.3.9/32
	Vlan-int11	172.1.1.1/24		Vlan-int11	162.1.1.1/24
	Vlan-int12	192.1.1.1/24		Vlan-int12	192.1.1.2/24

Procedure

1. Configure IGP on the MPLS backbone to implement the connectivity in the backbone.

   This example uses OSPF. (Details not shown.)

   # Execute the **`display ospf peer`** command to verify that each ASBR-PE has established an OSPF adjacency in Full state with the PE in the same AS, and that PEs and ASBR-PEs in the same AS have learned the routes to the loopback interfaces of each other. Verify that each ASBR-PE and the PE in the same AS can ping each other. (Details not shown.)

2. Configure basic MPLS and MPLS LDP on the MPLS backbone to establish LDP LSPs:

   # Configure basic MPLS on PE 1, and enable MPLS LDP on the interface connected to ASBR-PE 1.

   ```
   <PE1> system-view
   [PE1] mpls lsr-id 1.1.1.9
   [PE1] mpls ldp
   [PE1-ldp] quit
   [PE1] interface vlan-interface 11
   [PE1-Vlan-interface11] mpls enable
   [PE1-Vlan-interface11] mpls ldp enable
   ```

```
[PE1-Vlan-interface11] quit
```

Configure basic MPLS on ASBR-PE 1, and enable MPLS LDP on the interface connected to PE 1.

```
<ASBR-PE1> system-view
[ASBR-PE1] mpls lsr-id 2.2.2.9
[ASBR-PE1] mpls ldp
[ASBR-PE1-ldp] quit
[ASBR-PE1] interface vlan-interface 11
[ASBR-PE1-Vlan-interface11] mpls enable
[ASBR-PE1-Vlan-interface11] mpls ldp enable
[ASBR-PE1-Vlan-interface11] quit
```

Configure basic MPLS on ASBR-PE 2, and enable MPLS LDP on the interface connected to PE 2.

```
<ASBR-PE2> system-view
[ASBR-PE2] mpls lsr-id 3.3.3.9
[ASBR-PE2] mpls ldp
[ASBR-PE2-ldp] quit
[ASBR-PE2] interface vlan-interface 11
[ASBR-PE2-Vlan-interface11] mpls enable
[ASBR-PE2-Vlan-interface11] mpls ldp enable
[ASBR-PE2-Vlan-interface11] quit
```

Configure basic MPLS on PE 2, and enable MPLS LDP on the interface connected to ASBR-PE 2.

```
<PE2> system-view
[PE2] mpls lsr-id 4.4.4.9
[PE2] mpls ldp
[PE2-ldp] quit
[PE2] interface vlan-interface 11
[PE2-Vlan-interface11] mpls enable
[PE2-Vlan-interface11] mpls ldp enable
[PE2-Vlan-interface11] quit
```

Execute the **display mpls ldp peer** command on the devices to verify that the session status is Operational, and that each PE and the ASBR-PE in the same AS have established a neighbor relationship. (Details not shown.)

3. Configure VPN instances on PEs:

For the same VPN, the route targets for the VPN instance on the PE must match those for the VPN instance on the ASBR-PE in the same AS. This is not required for PEs in different ASs.

Configure CE 1.

```
<CE1> system-view
[CE1] interface vlan-interface 12
[CE1-Vlan-interface12] ip address 10.1.1.1 24
[CE1-Vlan-interface12] quit
```

Configure PE 1.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:2
[PE1-vpn-instance-vpn1] vpn-target 100:1 both
[PE1-vpn-instance-vpn1] quit
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance vpn1
```

```
[PE1-Vlan-interface12] ip address 10.1.1.2 24
[PE1-Vlan-interface12] quit
```

Configure CE 2.

```
<CE2> system-view
[CE2] interface vlan-interface 12
[CE2-Vlan-interface12] ip address 10.2.1.1 24
[CE2-Vlan-interface12] quit
```

Configure PE 2.

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance] route-distinguisher 200:2
[PE2-vpn-instance] vpn-target 200:1 both
[PE2-vpn-instance] quit
[PE2] interface vlan-interface 12
[PE2-Vlan-interface12] ip binding vpn-instance vpn1
[PE2-Vlan-interface12] ip address 10.2.1.2 24
[PE2-Vlan-interface12] quit
```

On ASBR-PE 1, create a VPN instance, and bind the instance to the interface connected to ASBR-PE 2. ASBR-PE 1 considers ASBR-PE 2 to be its CE.

```
[ASBR-PE1] ip vpn-instance vpn1
[ASBR-PE1-vpn-instance-vpn1] route-distinguisher 100:1
[ASBR-PE1-vpn-instance-vpn1] vpn-target 100:1 both
[ASBR-PE1-vpn-instance-vpn1] quit
[ASBR-PE1] interface vlan-interface 12
[ASBR-PE1-Vlan-interface12] ip binding vpn-instance vpn1
[ASBR-PE1-Vlan-interface12] ip address 192.1.1.1 24
[ASBR-PE1-Vlan-interface12] quit
```

On ASBR-PE 2, create a VPN instance, and bind the instance to the interface connected to ASBR-PE 1. ASBR-PE 2 considers ASBR-PE 1 to be its CE.

```
[ASBR-PE2] ip vpn-instance vpn1
[ASBR-PE2-vpn-instance-vpn1] route-distinguisher 200:1
[ASBR-PE2-vpn-instance-vpn1] vpn-target 200:1 both
[ASBR-PE2-vpn-instance-vpn1] quit
[ASBR-PE2] interface vlan-interface 12
[ASBR-PE2-Vlan-interface12] ip binding vpn-instance vpn1
[ASBR-PE2-Vlan-interface12] ip address 192.1.1.2 24
[ASBR-PE2-Vlan-interface12] quit
```

Execute the **`display ip vpn-instance`** command to display VPN instance configurations. Verify that the PEs can ping the CEs, and the ASBR-PEs can ping each other. (Details not shown.)

4. Establish EBGP peer relationships between PEs and CEs, and redistribute VPN routes into BGP:

Configure CE 1.

```
[CE1] bgp 65001
[CE1-bgp-default] peer 10.1.1.2 as-number 100
[CE1-bgp-default] address-family ipv4 unicast
[CE1-bgp-default-ipv4] peer 10.1.1.2 enable
[CE1-bgp-default-ipv4] import-route direct
[CE1-bgp-default-ipv4] quit
[CE1-bgp-default] quit
```

Configure PE 1.

```
[PE1] bgp 100
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 10.1.1.1 as-number 65001
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] peer 10.1.1.1 enable
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

Configure CE 2.

```
[CE2] bgp 65002
[CE2-bgp-default] peer 10.2.1.2 as-number 200
[CE2-bgp-default] address-family ipv4 unicast
[CE2-bgp-default-ipv4] peer 10.2.1.2 enable
[CE2-bgp-default-ipv4] import-route direct
[CE2-bgp-default-ipv4] quit
[CE2-bgp-default] quit
```

Configure PE 2.

```
[PE2] bgp 200
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] peer 10.2.1.1 as-number 65002
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] peer 10.2.1.1 enable
[PE2-bgp-default-ipv4-vpn1] quit
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

5. Establish an MP-IBGP peer relationship between each PE and the ASBR-PE in the same AS, and an EBGP peer relationship between the ASBR-PEs:

Configure PE 1.

```
[PE1] bgp 100
[PE1-bgp-default] peer 2.2.2.9 as-number 100
[PE1-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 2.2.2.9 enable
[PE1-bgp-default-vpnv4] peer 2.2.2.9 next-hop-local
[PE1-bgp-default-vpnv4] quit
[PE1-bgp-default] quit
```

Configure ASBR-PE 1.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp-default] ip vpn-instance vpn1
[ASBR-PE1-bgp-default-vpn1] peer 192.1.1.2 as-number 200
[ASBR-PE1-bgp-default-vpn1] address-family ipv4 unicast
[ASBR-PE1-bgp-default-ipv4-vpn1] peer 192.1.1.2 enable
[ASBR-PE1-bgp-default-ipv4-vpn1] quit
[ASBR-PE1-bgp-default-vpn1] quit
[ASBR-PE1-bgp-default] peer 1.1.1.9 as-number 100
[ASBR-PE1-bgp-default] peer 1.1.1.9 connect-interface loopback 0
[ASBR-PE1-bgp-default] address-family vpnv4
```

```
[ASBR-PE1-bgp-default-vpnv4] peer 1.1.1.9 enable
[ASBR-PE1-bgp-default-vpnv4] peer 1.1.1.9 next-hop-local
[ASBR-PE1-bgp-default-vpnv4] quit
[ASBR-PE1-bgp-default] quit
```

Configure ASBR-PE 2.

```
[ASBR-PE2] bgp 200
[ASBR-PE2-bgp-default] ip vpn-instance vpn1
[ASBR-PE2-bgp-default-vpn1] peer 192.1.1.1 as-number 100
[ASBR-PE2-bgp-default-vpn1] address-family ipv4 unicast
[ASBR-PE2-bgp-default-ipv4-vpn1] peer 192.1.1.1 enable
[ASBR-PE2-bgp-default-ipv4-vpn1] quit
[ASBR-PE2-bgp-default-vpn1] quit
[ASBR-PE2-bgp-default] peer 4.4.4.9 as-number 200
[ASBR-PE2-bgp-default] peer 4.4.4.9 connect-interface loopback 0
[ASBR-PE2-bgp-default] address-family vpnv4
[ASBR-PE2-bgp-default-vpnv4] peer 4.4.4.9 enable
[ASBR-PE2-bgp-default-vpnv4] peer 4.4.4.9 next-hop-local
[ASBR-PE2-bgp-default-vpnv4] quit
[ASBR-PE2-bgp-default] quit
```

Configure PE 2.

```
[PE2] bgp 200
[PE2-bgp-default] peer 3.3.3.9 as-number 200
[PE2-bgp-default] peer 3.3.3.9 connect-interface loopback 0
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] peer 3.3.3.9 enable
[PE2-bgp-default-vpnv4] peer 3.3.3.9 next-hop-local
[PE2-bgp-default-vpnv4] quit
[PE2-bgp-default] quit
```

Verifying the configuration

Verify that the CEs can learn the interface routes from each other and ping each other. (Details not shown.)

Example: Configuring MPLS L3VPN inter-AS option B

Network configuration

Site 1 and Site 2 belong to the same VPN. CE 1 of Site 1 accesses the network through PE 1 in AS 100. CE 2 of Site 2 accesses the network through PE 2 in AS 600. PEs in the same AS run IS-IS.

PE 1 and ASBR-PE 1 exchange VPNv4 routes through MP-IBGP. PE 2 and ASBR-PE 2 exchange VPNv4 routes through MP-IBGP. ASBR-PE 1 and ASBR-PE 2 exchange VPNv4 routes through MP-EBGP.

ASBRs do not perform route target filtering of received VPN-IPv4 routes.

Figure 29 Network diagram

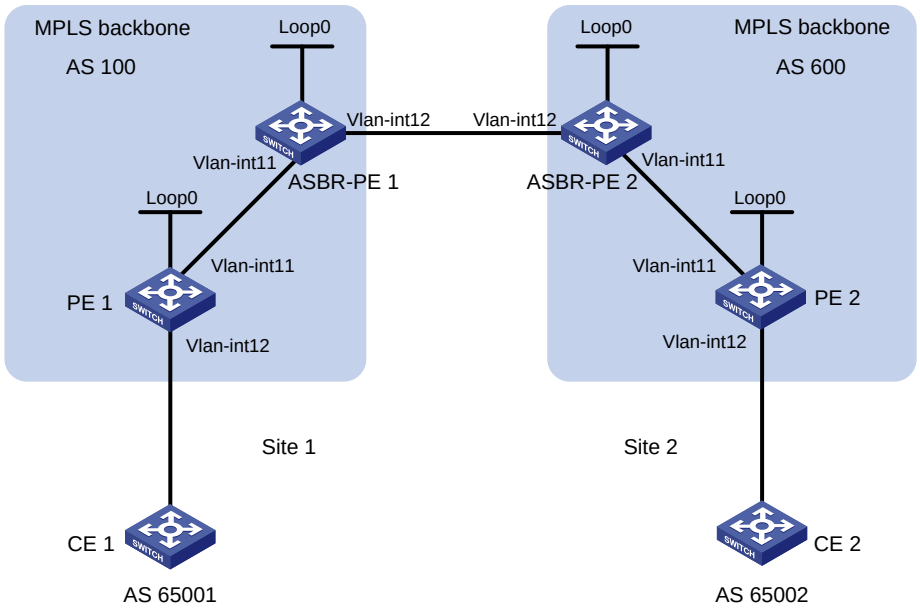


Table 15 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
PE 1	Loop0	2.2.2.9/32	PE 2	Loop0	5.5.5.9/32
	Vlan-int12	30.0.0.1/8		Vlan-int12	20.0.0.1/8
	Vlan-int11	1.1.1.2/8		Vlan-int11	9.1.1.2/8
ASBR-PE 1	Loop0	3.3.3.9/32	ASBR-PE 2	Loop0	4.4.4.9/32
	Vlan-int11	1.1.1.1/8		Vlan-int11	9.1.1.1/8
	Vlan-int12	11.0.0.2/8		Vlan-int12	11.0.0.1/8

Procedure

1. Configure PE 1:

   # Configure IS-IS on PE 1.

```
<PE1> system-view
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0001.00
[PE1-isis-1] quit
```

   # Configure the LSR ID, and enable MPLS and LDP.

```
[PE1] mpls lsr-id 2.2.2.9
[PE1] mpls ldp
[PE1-ldp] quit
```

   # Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip address 1.1.1.2 255.0.0.0
[PE1-Vlan-interface11] isis enable 1
[PE1-Vlan-interface11] mpls enable
[PE1-Vlan-interface11] mpls ldp enable
[PE1-Vlan-interface11] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 2.2.2.9 32
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 11:11
[PE1-vpn-instance-vpn1] vpn-target 1:1 2:2 import-extcommunity
[PE1-vpn-instance-vpn1] vpn-target 1:1 export-extcommunity
[PE1-vpn-instance-vpn1] quit
```

Bind the interface connected to CE 1 to the created VPN instance.

```
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance vpn1
[PE1-Vlan-interface12] ip address 30.0.0.1 8
[PE1-Vlan-interface12] quit
```

Enable BGP on PE 1.

```
[PE1] bgp 100
```

Configure IBGP peer 3.3.3.9 as a VPNv4 peer.

```
[PE1-bgp-default] peer 3.3.3.9 as-number 100
[PE1-bgp-default] peer 3.3.3.9 connect-interface loopback 0
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 3.3.3.9 enable
[PE1-bgp-default-vpnv4] quit
```

Redistribute direct routes to the VPN routing table of **vpn1**.

```
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] import-route direct
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

2. Configure ASBR-PE 1:

Enable IS-IS on ASBR-PE 1.

```
<ASBR-PE1> system-view
[ASBR-PE1] isis 1
[ASBR-PE1-isis-1] network-entity 10.0000.0000.0000.0002.00
[ASBR-PE1-isis-1] quit
```

Configure the LSR ID, and enable MPLS and LDP.

```
[ASBR-PE1] mpls lsr-id 3.3.3.9
[ASBR-PE1] mpls ldp
[ASBR-PE1-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[ASBR-PE1] interface vlan-interface11
[ASBR-PE1-Vlan-interface11] ip address 1.1.1.1 255.0.0.0
[ASBR-PE1-Vlan-interface11] isis enable 1
[ASBR-PE1-Vlan-interface11] mpls enable
[ASBR-PE1-Vlan-interface11] mpls ldp enable
[ASBR-PE1-Vlan-interface11] quit
```

Configure VLAN-interface 12, and enable MPLS on it.

```
[ASBR-PE1] interface vlan-interface 12
[ASBR-PE1-Vlan-interface12] ip address 11.0.0.2 255.0.0.0
[ASBR-PE1-Vlan-interface12] mpls enable
[ASBR-PE1-Vlan-interface12] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[ASBR-PE1] interface loopback 0
[ASBR-PE1-LoopBack0] ip address 3.3.3.9 32
[ASBR-PE1-LoopBack0] isis enable 1
[ASBR-PE1-LoopBack0] quit
```

Enable BGP on ASBR-PE 1.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp-default] peer 2.2.2.9 as-number 100
[ASBR-PE1-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[ASBR-PE1-bgp-default] peer 11.0.0.1 as-number 600
[ASBR-PE1-bgp-default] peer 11.0.0.1 connect-interface vlan-interface 12
```

Disable route target based filtering of received VPNv4 routes.

```
[ASBR-PE1-bgp-default] address-family vpnv4
[ASBR-PE1-bgp-default-vpnv4] undo policy vpn-target
```

Configure both IBGP peer 2.2.2.0 and EBGP peer 11.0.0.1 as VPNv4 peers.

```
[ASBR-PE1-bgp-default-vpnv4] peer 11.0.0.1 enable
[ASBR-PE1-bgp-default-vpnv4] peer 2.2.2.9 enable
[ASBR-PE1-bgp-default-vpnv4] quit
[ASBR-PE1-bgp-default] quit
[ASBR-PE1-bgp-default] quit
```

3. Configure ASBR-PE 2:

Enable IS-IS on ASBR-PE 2.

```
<ASBR-PE2> system-view
[ASBR-PE2] isis 1
[ASBR-PE2-isis-1] network-entity 10.0000.0000.0000.0003.00
[ASBR-PE2-isis-1] quit
```

Configure the LSR ID, and enable MPLS and LDP.

```
[ASBR-PE2] mpls lsr-id 4.4.4.9
[ASBR-PE2] mpls ldp
[ASBR-PE2-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[ASBR-PE2] interface vlan-interface 11
[ASBR-PE2-Vlan-interface11] ip address 9.1.1.1 255.0.0.0
[ASBR-PE2-Vlan-interface11] isis enable 1
[ASBR-PE2-Vlan-interface11] mpls enable
[ASBR-PE2-Vlan-interface11] mpls ldp enable
[ASBR-PE2-Vlan-interface11] quit
```

Configure VLAN-interface 12, and enable MPLS on it.

```
[ASBR-PE2] interface vlan-interface 12
[ASBR-PE2-Vlan-interface12] ip address 11.0.0.1 255.0.0.0
[ASBR-PE2-Vlan-interface12] mpls enable
[ASBR-PE2-Vlan-interface12] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[ASBR-PE2] interface loopback 0
[ASBR-PE2-LoopBack0] ip address 4.4.4.9 32
[ASBR-PE2-LoopBack0] isis enable 1
[ASBR-PE2-LoopBack0] quit
```

Enable BGP on ASBR-PE 2.

```
[ASBR-PE2] bgp 600
[ASBR-PE2-bgp-default] peer 11.0.0.2 as-number 100
[ASBR-PE2-bgp-default] peer 11.0.0.2 connect-interface vlan-interface 12
[ASBR-PE2-bgp-default] peer 5.5.5.9 as-number 600
[ASBR-PE2-bgp-default] peer 5.5.5.9 connect-interface loopback 0
```

Disable route target based filtering of received VPNv4 routes.

```
[ASBR-PE2-bgp-default] address-family vpnv4
[ASBR-PE2-bgp-default-vpnv4] undo policy vpn-target
```

Configure both IBGP peer 5.5.5.9 and EBGP peer 11.0.0.2 as VPNv4 peers.

```
[ASBR-PE2-bgp-default-vpnv4] peer 11.0.0.2 enable
[ASBR-PE2-bgp-default-vpnv4] peer 5.5.5.9 enable
[ASBR-PE2-bgp-default-vpnv4] quit
[ASBR-PE2-bgp-default] quit
```

4. Configure PE 2:

Enable IS-IS on PE 2.

```
<PE2> system-view
[PE2] isis 1
[PE2-isis-1] network-entity 10.0000.0000.0000.0004.00
[PE2-isis-1] quit
```

Configure the LSR ID, and enable MPLS and LDP.

```
[PE2] mpls lsr-id 5.5.5.9
[PE2] mpls ldp
[PE2-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[PE2] interface vlan-interface 11
[PE2-Vlan-interface11] ip address 9.1.1.2 255.0.0.0
[PE2-Vlan-interface11] isis enable 1
[PE2-Vlan-interface11] mpls enable
[PE2-Vlan-interface11] mpls ldp enable
[PE2-Vlan-interface11] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 5.5.5.9 32
[PE2-LoopBack0] isis enable 1
[PE2-LoopBack0] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 12:12
[PE2-vpn-instance-vpn1] vpn-target 1:1 2:2 import-extcommunity
[PE2-vpn-instance-vpn1] vpn-target 2:2 export-extcommunity
[PE2-vpn-instance-vpn1] quit
```

Bind the interface connected to CE 2 to the created VPN instance.

```
[PE2] interface vlan-interface12
[PE2-Vlan-interface12] ip binding vpn-instance vpn1
[PE2-Vlan-interface12] ip address 20.0.0.1 8
[PE2-Vlan-interface12] quit
```

Enable BGP on PE 2.

```
[PE2] bgp 600
```

Configure IBGP peer 4.4.4.9 as a VPNv4 peer.

```
[PE2-bgp-default] peer 4.4.4.9 as-number 600
[PE2-bgp-default] peer 4.4.4.9 connect-interface loopback 0
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] peer 4.4.4.9 enable
[PE2-bgp-default-vpnv4] quit
```

Redistribute direct routes to the VPN routing table of **vpn1**.

```
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] import-route direct
[PE2-bgp-default-ipv4-vpn1] quit
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

Verifying the configuration

Use the following command on PE 1 to verify its connectivity to PE 2.

```
[PE1] ping -a 30.0.0.1 -vpn-instance vpn1 20.0.0.1
Ping 20.0.0.1 (20.0.0.1) from 30.0.0.1: 56 data bytes, press CTRL+C to break
56 bytes from 20.0.0.1: icmp_seq=0 ttl=255 time=1.208 ms
56 bytes from 20.0.0.1: icmp_seq=1 ttl=255 time=0.867 ms
56 bytes from 20.0.0.1: icmp_seq=2 ttl=255 time=0.551 ms
56 bytes from 20.0.0.1: icmp_seq=3 ttl=255 time=0.566 ms
56 bytes from 20.0.0.1: icmp_seq=4 ttl=255 time=0.570 ms

--- Ping statistics for 20.0.0.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.551/0.752/1.208/0.257 ms
```

Example: Configuring MPLS L3VPN inter-AS option C (method 1)

Network configuration

Site 1 and Site 2 belong to the same VPN. Site 1 accesses the network through PE 1 in AS 100. Site 2 accesses the network through PE 2 in AS 600. PEs in the same AS run IS-IS.

PE 1 and ASBR-PE 1 exchange labeled IPv4 routes through IBGP. PE 2 and ASBR-PE 2 exchange labeled IPv4 routes through IBGP. PE 1 and PE 2 exchange VPNv4 routes through MP-EBGP.

ASBR-PE 1 and ASBR-PE 2 use their respective routing policies and label routes received from each other.

ASBR-PE 1 and ASBR-PE 2 use EBGP to exchange labeled IPv4 routes.

Figure 30 Network diagram

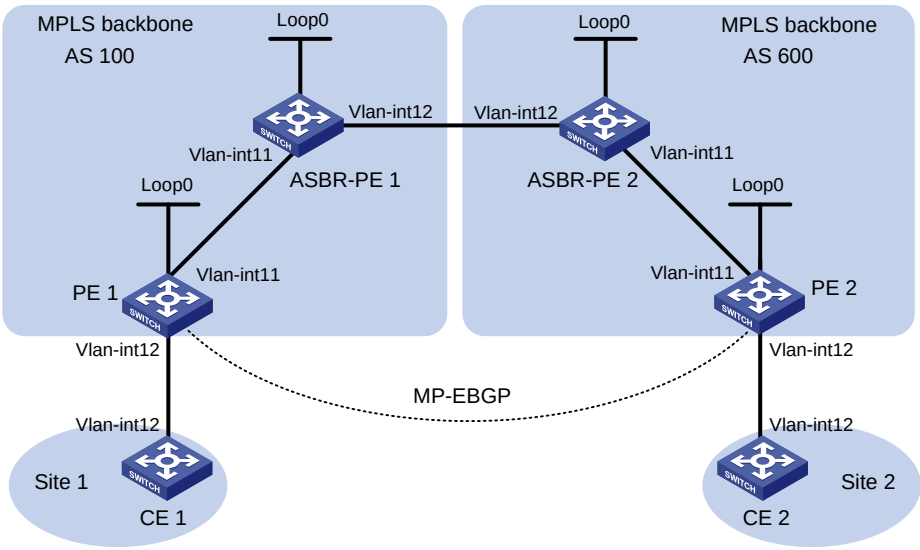


Table 16 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
PE 1	Loop0	2.2.2.9/32	PE 2	Loop0	5.5.5.9/32
	Vlan-int11	1.1.1.2/8		Vlan-int11	9.1.1.2/8
	Vlan-int12	30.0.0.1/24		Vlan-int12	20.0.0.1/24
ASBR-PE 1	Loop0	3.3.3.9/32	ASBR-PE 2	Loop0	4.4.4.9/32
	Vlan-int11	1.1.1.1/8		Vlan-int11	9.1.1.1/8
	Vlan-int12	11.0.0.2/8		Vlan-int12	11.0.0.1/8
CE 1	Vlan-int12	30.0.0.2/24	CE 2	Vlan-int12	20.0.0.2/24

Procedure

1. Configure CE 1:

   # Configure an IP address for VLAN-interface 12.

   ```
   <CE1> system-view
   [CE1] interface vlan-interface 12
   [CE1-Vlan-interface12] ip address 30.0.0.2 24
   [CE1-Vlan-interface12] quit
   ```

   # Establish an EBGP peer relationship with PE 1, and redistribute VPN routes.

   ```
   [CE1] bgp 65001
   [CE1-bgp-default] peer 30.0.0.1 as-number 100
   [CE1-bgp-default] address-family ipv4 unicast
   [CE1-bgp-default-ipv4] peer 30.0.0.1 enable
   [CE1-bgp-default-ipv4] import-route direct
   [CE1-bgp-default-ipv4] quit
   [CE1-bgp-default] quit
   ```

2. Configure PE 1:

   # Configure IS-IS on PE 1.

   ```
   <PE1> system-view
   ```

```
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0001.00
[PE1-isis-1] quit
```

Configure the LSR ID, and enable MPLS and LDP.

```
[PE1] mpls lsr-id 2.2.2.9
[PE1] mpls ldp
[PE1-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip address 1.1.1.2 255.0.0.0
[PE1-Vlan-interface11] isis enable 1
[PE1-Vlan-interface11] mpls enable
[PE1-Vlan-interface11] mpls ldp enable
[PE1-Vlan-interface11] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 2.2.2.9 32
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 11:11
[PE1-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE1-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE1-vpn-instance-vpn1] quit
```

Associate VLAN-interface 12 with VPN instance **vpn1**, and specify the IP address for the interface.

```
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance vpn1
[PE1-Vlan-interface12] ip address 30.0.0.1 24
[PE1-Vlan-interface12] quit
```

Enable BGP on PE 1.

```
[PE1] bgp 100
```

Enable the capability to advertise labeled routes to IBGP peer 3.3.3.9 and to receive labeled routes from the peer.

```
[PE1-bgp-default] peer 3.3.3.9 as-number 100
[PE1-bgp-default] peer 3.3.3.9 connect-interface loopback 0
[PE1-bgp-default] address-family ipv4 unicast
[PE1-bgp-default-ipv4] peer 3.3.3.9 enable
[PE1-bgp-default-ipv4] peer 3.3.3.9 label-route-capability
[PE1-bgp-default-ipv4] quit
```

Configure the maximum hop count from PE 1 to EBGP peer 5.5.5.9 as 10.

```
[PE1-bgp-default] peer 5.5.5.9 as-number 600
[PE1-bgp-default] peer 5.5.5.9 connect-interface loopback 0
[PE1-bgp-default] peer 5.5.5.9 ebgp-max-hop 10
```

Configure peer 5.5.5.9 as a VPNv4 peer.

```
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 5.5.5.9 enable
```

```
[PE1-bgp-default-vpnv4] quit
```

Establish an EBGP peer relationship with CE 1, and add the learned BGP routes to the routing table of VPN instance **vpn1**.

```
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 30.0.0.2 as-number 65001
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] peer 30.0.0.2 enable
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

3. Configure ASBR-PE 1:

Enable IS-IS on ASBR-PE 1.

```
<ASBR-PE1> system-view
[ASBR-PE1] isis 1
[ASBR-PE1-isis-1] network-entity 10.0000.0000.0000.0002.00
[ASBR-PE1-isis-1] quit
```

Configure the LSR ID, and enable MPLS and LDP.

```
[ASBR-PE1] mpls lsr-id 3.3.3.9
[ASBR-PE1] mpls ldp
[ASBR-PE1-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[ASBR-PE1] interface vlan-interface 11
[ASBR-PE1-Vlan-interface11] ip address 1.1.1.1 255.0.0.0
[ASBR-PE1-Vlan-interface11] isis enable 1
[ASBR-PE1-Vlan-interface11] mpls enable
[ASBR-PE1-Vlan-interface11] mpls ldp enable
[ASBR-PE1-Vlan-interface11] quit
```

Configure VLAN-interface 12, and enable MPLS on it.

```
[ASBR-PE1] interface vlan-interface 12
[ASBR-PE1-Vlan-interface12] ip address 11.0.0.2 255.0.0.0
[ASBR-PE1-Vlan-interface12] mpls enable
[ASBR-PE1-Vlan-interface12] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[ASBR-PE1] interface loopback 0
[ASBR-PE1-LoopBack0] ip address 3.3.3.9 32
[ASBR-PE1-LoopBack0] isis enable 1
[ASBR-PE1-LoopBack0] quit
```

Create routing policies.

```
[ASBR-PE1] route-policy policy1 permit node 1
[ASBR-PE1-route-policy-policy1-1] apply mpls-label
[ASBR-PE1-route-policy-policy1-1] quit
[ASBR-PE1] route-policy policy2 permit node 1
[ASBR-PE1-route-policy-policy2-1] if-match mpls-label
[ASBR-PE1-route-policy-policy2-1] apply mpls-label
[ASBR-PE1-route-policy-policy2-1] quit
```

Enable BGP on ASBR-PE 1, and apply the routing policy **policy2** to routes advertised to IBGP peer 2.2.2.9.

```
[ASBR-PE1] bgp 100
```

```
[ASBR-PE1-bgp-default] peer 2.2.2.9 as-number 100
[ASBR-PE1-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[ASBR-PE1-bgp-default] address-family ipv4 unicast
[ASBR-PE1-bgp-default-ipv4] peer 2.2.2.9 enable
[ASBR-PE1-bgp-default-ipv4] peer 2.2.2.9 route-policy policy2 export
```

Enable the capability to advertise labeled routes to IBGP peer 2.2.2.9 and to receive labeled routes from the peer.

```
[ASBR-PE1-bgp-default-ipv4] peer 2.2.2.9 label-route-capability
```

Redistribute routes from IS-IS process 1 to BGP.

```
[ASBR-PE1-bgp-default-ipv4] import-route isis 1
[ASBR-PE1-bgp-default-ipv4] quit
```

Apply the routing policy **policy1** to routes advertised to EBGP peer 11.0.0.1.

```
[ASBR-PE1-bgp-default] peer 11.0.0.1 as-number 600
[ASBR-PE1-bgp-default] address-family ipv4 unicast
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 enable
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 route-policy policy1 export
```

Enable the capability to advertise labeled routes to EBGP peer 11.0.0.1 and to receive labeled routes from the peer.

```
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 label-route-capability
[ASBR-PE1-bgp-default-ipv4] quit
[ASBR-PE1-bgp-default] quit
```

4. Configure ASBR-PE 2:

Enable IS-IS on ASBR-PE 2.

```
<ASBR-PE2> system-view
[ASBR-PE2] isis 1
[ASBR-PE2-isis-1] network-entity 10.0000.0000.0000.0003.00
[ASBR-PE2-isis-1] quit
```

Configure the LSR ID, and enable MPLS and LDP.

```
[ASBR-PE2] mpls lsr-id 4.4.4.9
[ASBR-PE2] mpls ldp
[ASBR-PE2-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[ASBR-PE2] interface vlan-interface 11
[ASBR-PE2-Vlan-interface11] ip address 9.1.1.1 255.0.0.0
[ASBR-PE2-Vlan-interface11] isis enable 1
[ASBR-PE2-Vlan-interface11] mpls enable
[ASBR-PE2-Vlan-interface11] mpls ldp enable
[ASBR-PE2-Vlan-interface11] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[ASBR-PE2] interface loopback 0
[ASBR-PE2-LoopBack0] ip address 4.4.4.9 32
[ASBR-PE2-LoopBack0] isis enable 1
[ASBR-PE2-LoopBack0] quit
```

Configure VLAN-interface 12, and enable MPLS on it.

```
[ASBR-PE2] interface vlan-interface 12
[ASBR-PE2-Vlan-interface12] ip address 11.0.0.1 255.0.0.0
[ASBR-PE2-Vlan-interface12] mpls enable
[ASBR-PE2-Vlan-interface12] quit
```

Create routing policies.

```
[ASBR-PE2] route-policy policy1 permit node 1
[ASBR-PE2-route-policy-policy1-1] apply mpls-label
[ASBR-PE2-route-policy-policy1-1] quit
[ASBR-PE2] route-policy policy2 permit node 1
[ASBR-PE2-route-policy-policy2-1] if-match mpls-label
[ASBR-PE2-route-policy-policy2-1] apply mpls-label
[ASBR-PE2-route-policy-policy2-1] quit
```

Enable BGP on ASBR-PE 2, and enable the capability to advertise labeled routes to IBGP peer 5.5.5.9 and to receive labeled routes from the peer.

```
[ASBR-PE2] bgp 600
[ASBR-PE2-bgp-default] peer 5.5.5.9 as-number 600
[ASBR-PE2-bgp-default] peer 5.5.5.9 connect-interface loopback 0
[ASBR-PE2-bgp-default] address-family ipv4 unicast
[ASBR-PE2-bgp-default-ipv4] peer 5.5.5.9 enable
[ASBR-PE2-bgp-default-ipv4] peer 5.5.5.9 label-route-capability
```

Apply routing policy **policy2** to routes advertised to IBGP peer 5.5.5.9.

```
[ASBR-PE2-bgp-default-ipv4] peer 5.5.5.9 route-policy policy2 export
```

Redistribute routes from IS-IS process 1 into BGP.

```
[ASBR-PE2-bgp-default-ipv4] import-route isis 1
[ASBR-PE2-bgp-default-ipv4] quit
```

Apply routing policy **policy1** to routes advertised to EBGP peer 11.0.0.2.

```
[ASBR-PE2-bgp-default] peer 11.0.0.2 as-number 100
[ASBR-PE2-bgp-default] address-family ipv4 unicast
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 enable
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 route-policy policy1 export
```

Enable the capability to advertise labeled routes to EBGP peer 11.0.0.2 and to receive labeled routes from the peer.

```
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 label-route-capability
[ASBR-PE2-bgp-default-ipv4] quit
[ASBR-PE2-bgp-default] quit
```

5. Configure PE 2:

Enable IS-IS on PE 2.

```
<PE2> system-view
[PE2] isis 1
[PE2-isis-1] network-entity 10.0000.0000.0000.0004.00
[PE2-isis-1] quit
```

Configure the LSR ID, and enable MPLS and LDP.

```
[PE2] mpls lsr-id 5.5.5.9
[PE2] mpls ldp
[PE2-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[PE2] interface vlan-interface 11
[PE2-Vlan-interface11] ip address 9.1.1.2 255.0.0.0
[PE2-Vlan-interface11] isis enable 1
[PE2-Vlan-interface11] mpls enable
[PE2-Vlan-interface11] mpls ldp enable
[PE2-Vlan-interface11] quit
```

Configure Loopback 0, and enable IS-IS on it.

```
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 5.5.5.9 32
[PE2-LoopBack0] isis enable 1
[PE2-LoopBack0] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 11:11
[PE2-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE2-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE2-vpn-instance-vpn1] quit
```

Associate VLAN-interface 12 with VPN instance **vpn1**, and specify the IP address for the interface.

```
[PE2] interface vlan-interface 12
[PE2-Vlan-interface12] ip binding vpn-instance vpn1
[PE2-Vlan-interface12] ip address 20.0.0.1 24
[PE2-Vlan-interface12] quit
```

Enable BGP on PE 2.

```
[PE2] bgp 600
```

Enable the capability to advertise labeled routes to IBGP peer 4.4.4.9 and to receive labeled routes from the peer.

```
[PE2-bgp-default] peer 4.4.4.9 as-number 600
[PE2-bgp-default] peer 4.4.4.9 connect-interface loopback 0
[PE2-bgp-default] address-family ipv4 unicast
[PE2-bgp-default-ipv4] peer 4.4.4.9 enable
[PE2-bgp-default-ipv4] peer 4.4.4.9 label-route-capability
[PE2-bgp-default-ipv4] quit
```

Configure the maximum hop count from PE 2 to EBGP peer 2.2.2.9 as 10.

```
[PE2-bgp-default] peer 2.2.2.9 as-number 100
[PE2-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[PE2-bgp-default] peer 2.2.2.9 ebgp-max-hop 10
```

Configure peer 2.2.2.9 as a VPNv4 peer.

```
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] peer 2.2.2.9 enable
[PE2-bgp-default-vpnv4] quit
```

Establish an EBGP peer relationship with CE 2, and add the learned BGP routes to the routing table of VPN instance **vpn1**.

```
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] peer 20.0.0.2 as-number 65002
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] peer 20.0.0.2 enable
[PE2-bgp-default-ipv4-vpn1] quit
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

6. Configure CE 2:

Configure an IP address for VLAN-interface 12.

```
<CE2> system-view
[CE2] interface vlan-interface 12
```

```
[CE2-Vlan-interface12] ip address 20.0.0.2 24
[CE2-Vlan-interface12] quit
```

Establish an EBGP peer relationship with PE 2, and redistribute VPN routes.

```
[CE2] bgp 65002
[CE2-bgp-default] peer 20.0.0.1 as-number 600
[CE2-bgp-default] address-family ipv4 unicast
[CE2-bgp-default-ipv4] peer 20.0.0.1 enable
[CE2-bgp-default-ipv4] import-route direct
[CE2-bgp-default-ipv4] quit
[CE2-bgp-default] quit
```

Verifying the configuration

Execute the **`display ip routing table`** command on CE 1 and CE 2 to verify that CE 1 and CE 2 have a route to each other. Verify that CE 1 and CE 2 can ping each other. (Details not shown.)

Example: Configuring MPLS L3VPN inter-AS option C (method 2)

Network configuration

Site 1 and Site 2 belong to the same VPN. Site 1 accesses the network through PE 1 in AS 100, and Site 2 accesses the network through PE 2 in AS 600. PEs in the same AS run IS-IS.

PE 1 and PE 2 are MP-EBGP peers and exchange VPNv4 routes.

ASBR-PE 1 and ASBR-PE 2 label the routes received from each other, use EBGP to exchange labeled IPv4 routes, and redistribute IGP and BGP routes from each other.

Figure 31 Network diagram

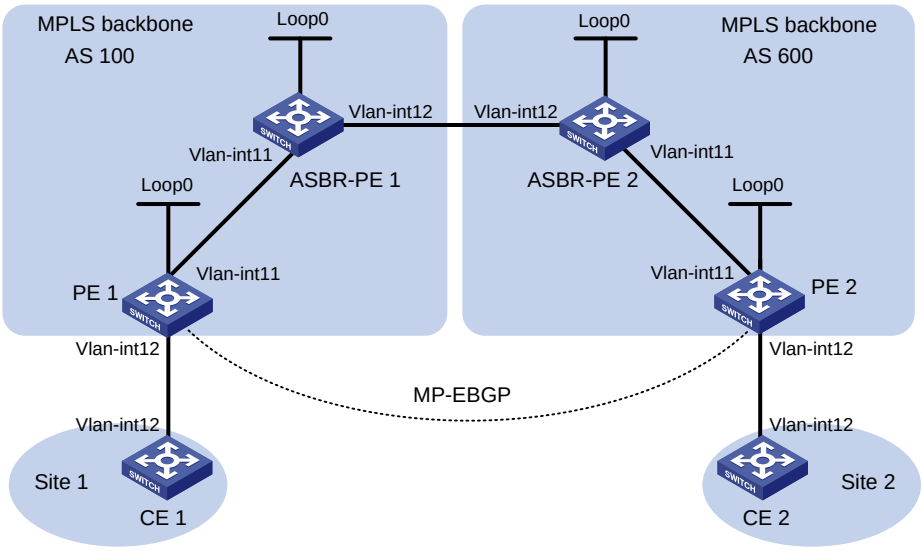


Table 17 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
PE 1	Loop0	2.2.2.9/32	PE 2	Loop0	5.5.5.9/32
	Vlan-int11	1.1.1.2/8		Vlan-int11	9.1.1.2/8
	Vlan-int12	30.0.0.1/24		Vlan-int12	20.0.0.1/24

ASBR-PE 1	Loop0	3.3.3.9/32	ASBR-PE 2	Loop0	4.4.4.9/32
	Vlan-int11	1.1.1.1/8		Vlan-int11	9.1.1.1/8
	Vlan-int12	11.0.0.2/8		Vlan-int12	11.0.0.1/8
CE 1	Vlan-int12	30.0.0.2/24	CE 2	Vlan-int12	20.0.0.2/24

Procedure

1. Configure CE 1:

Configure an IP address for VLAN-interface 12.

```
<CE1> system-view
[CE1] interface vlan-interface 12
[CE1-Vlan-interface12] ip address 30.0.0.2 24
[CE1-Vlan-interface12] quit
```

Establish an EBGP peer relationship with PE 1, and redistribute VPN routes.

```
[CE1] bgp 65001
[CE1-bgp-default] peer 30.0.0.1 as-number 100
[CE1-bgp-default] address-family ipv4 unicast
[CE1-bgp-default-ipv4] peer 30.0.0.1 enable
[CE1-bgp-default-ipv4] import-route direct
[CE1-bgp-default-ipv4] quit
[CE1-bgp-default] quit
```

2. Configure PE 1:

Configure IS-IS on PE 1.

```
<PE1> system-view
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0001.00
[PE1-isis-1] quit
```

Configure an LSR ID, and enable MPLS and LDP.

```
[PE1] mpls lsr-id 2.2.2.9
[PE1] mpls ldp
[PE1-ldp] quit
```

Configure interface VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip address 1.1.1.2 255.0.0.0
[PE1-Vlan-interface11] isis enable 1
[PE1-Vlan-interface11] mpls enable
[PE1-Vlan-interface11] mpls ldp enable
[PE1-Vlan-interface11] quit
```

Configure interface Loopback 0, and enable IS-IS on it.

```
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 2.2.2.9 32
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 11:11
[PE1-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE1-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
```

```
[PE1-vpn-instance-vpn1] quit
```

Associate VLAN-interface 12 with VPN instance **vpn1**, and specify an IP address for the interface.

```
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance vpn1
[PE1-Vlan-interface12] ip address 30.0.0.1 24
[PE1-Vlan-interface12] quit
```

Enable BGP on PE 1.

```
[PE1] bgp 100
```

Configure the maximum hop count from PE 1 to EBGP peer 5.5.5.9 as 10.

```
[PE1-bgp-default] peer 5.5.5.9 as-number 600
[PE1-bgp-default] peer 5.5.5.9 connect-interface loopback 0
[PE1-bgp-default] peer 5.5.5.9 ebgp-max-hop 10
```

Configure peer 5.5.5.9 as a VPNv4 peer.

```
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 5.5.5.9 enable
[PE1-bgp-default-vpnv4] quit
```

Establish an EBGP peer relationship with CE 1, and add the learned BGP routes to the routing table of VPN instance **vpn1**.

```
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 30.0.0.2 as-number 65001
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] peer 30.0.0.2 enable
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

3. Configure ASBR-PE 1:

Enable IS-IS on ASBR-PE 1.

```
<ASBR-PE1> system-view
[ASBR-PE1] isis 1
[ASBR-PE1-isis-1] network-entity 10.0000.0000.0000.0002.00
```

Redistribute BGP routes.

```
[ASBR-PE1-isis-1] address-family ipv4 unicast
[ASBR-PE1-isis-1-ipv4] import-route bgp
[ASBR-PE1-isis-1-ipv4] quit
[ASBR-PE1-isis-1] quit
```

Configure an LSR ID, and enable MPLS and LDP.

```
[ASBR-PE1] mpls lsr-id 3.3.3.9
[ASBR-PE1] mpls ldp
[ASBR-PE1-ldp] quit
```

Configure VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[ASBR-PE1] interface vlan-interface 11
[ASBR-PE1-Vlan-interface11] ip address 1.1.1.1 255.0.0.0
[ASBR-PE1-Vlan-interface11] isis enable 1
[ASBR-PE1-Vlan-interface11] mpls enable
[ASBR-PE1-Vlan-interface11] mpls ldp enable
[ASBR-PE1-Vlan-interface11] quit
```

Configure interface VLAN-interface 12, and enable MPLS on it.

```
[ASBR-PE1] interface vlan-interface 12
[ASBR-PE1-Vlan-interface12] ip address 11.0.0.2 255.0.0.0
[ASBR-PE1-Vlan-interface12] mpls enable
[ASBR-PE1-Vlan-interface12] quit
```

Configure interface Loopback 0, and enable IS-IS on it.

```
[ASBR-PE1] interface loopback 0
[ASBR-PE1-LoopBack0] ip address 3.3.3.9 32
[ASBR-PE1-LoopBack0] isis enable 1
[ASBR-PE1-LoopBack0] quit
```

Create routing policy **policy**1.

```
[ASBR-PE1] route-policy policy1 permit node 1
[ASBR-PE1-route-policy-policy1-1] apply mpls-label
[ASBR-PE1-route-policy-policy1-1] quit
```

Enable BGP on ASBR-PE 1, and redistribute routes from IS-IS process 1.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp-default] address-family ipv4 unicast
[ASBR-PE1-bgp-default-ipv4] import-route isis 1
[ASBR-PE1-bgp-default-ipv4] quit
```

Apply the routing policy **policy1** to routes advertised to EBGP peer 11.0.0.1.

```
[ASBR-PE1-bgp-default] peer 11.0.0.1 as-number 600
[ASBR-PE1-bgp-default] address-family ipv4 unicast
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 enable
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 route-policy policy1 export
```

Enable the capability to advertise labeled routes to EBGP peer 11.0.0.1 and to receive labeled routes from the peer.

```
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 label-route-capability
[ASBR-PE1-bgp-default-ipv4] quit
[ASBR-PE1-bgp-default] quit
```

4. Configure ASBR-PE 2:

Enable IS-IS on ASBR-PE 2.

```
<ASBR-PE2> system-view
[ASBR-PE2] isis 1
[ASBR-PE2-isis-1] network-entity 10.0000.0000.0000.0003.00
```

Redistribute BGP routes.

```
[ASBR-PE2-isis-1] address-family ipv4 unicast
[ASBR-PE2-isis-1-ipv4] import-route bgp
[ASBR-PE2-isis-1-ipv4] quit
[ASBR-PE2-isis-1] quit
```

Configure an LSR ID, and enable MPLS and LDP.

```
[ASBR-PE2] mpls lsr-id 4.4.4.9
[ASBR-PE2] mpls ldp
[ASBR-PE2-ldp] quit
```

Configure interface VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[ASBR-PE2] interface vlan-interface 11
[ASBR-PE2-Vlan-interface11] ip address 9.1.1.1 255.0.0.0
[ASBR-PE2-Vlan-interface11] isis enable 1
[ASBR-PE2-Vlan-interface11] mpls enable
[ASBR-PE2-Vlan-interface11] mpls ldp enable
```

```
[ASBR-PE2-Vlan-interface11] quit
```

Configure interface Loopback 0, and enable IS-IS on it.

```
[ASBR-PE2] interface loopback 0
[ASBR-PE2-LoopBack0] ip address 4.4.4.9 32
[ASBR-PE2-LoopBack0] isis enable 1
[ASBR-PE2-LoopBack0] quit
```

Configure interface VLAN-interface 12, and enable MPLS on it.

```
[ASBR-PE2] interface vlan-interface 12
[ASBR-PE2-Vlan-interface12] ip address 11.0.0.1 255.0.0.0
[ASBR-PE2-Vlan-interface12] mpls enable
[ASBR-PE2-Vlan-interface12] quit
```

Create routing policy **policy**1.

```
[ASBR-PE2] route-policy policy1 permit node 1
[ASBR-PE2-route-policy-policy1-1] apply mpls-label
[ASBR-PE2-route-policy-policy1-1] quit
```

Enable BGP on ASBR-PE 2, and redistribute routes from IS-IS process 1.

```
[ASBR-PE2] bgp 600
[ASBR-PE2-bgp-default] address-family ipv4 unicast
[ASBR-PE2-bgp-default-ipv4] import-route isis 1
[ASBR-PE2-bgp-default-ipv4] quit
```

Apply routing policy **policy1** to routes advertised to EBGP peer 11.0.0.2.

```
[ASBR-PE2-bgp-default] peer 11.0.0.2 as-number 100
[ASBR-PE2-bgp-default] address-family ipv4 unicast
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 enable
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 route-policy policy1 export
```

Enable the capability to advertise labeled routes to EBGP peer 11.0.0.2 and to receive labeled routes from the peer.

```
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 label-route-capability
[ASBR-PE2-bgp-default-ipv4] quit
[ASBR-PE2-bgp-default] quit
```

5. Configure PE 2:

Enable IS-IS on PE 2.

```
<PE2> system-view
[PE2] isis 1
[PE2-isis-1] network-entity 10.0000.0000.0000.0004.00
[PE2-isis-1] quit
```

Configure an LSR ID, and enable MPLS and LDP.

```
[PE2] mpls lsr-id 5.5.5.9
[PE2] mpls ldp
[PE2-ldp] quit
```

Configure interface VLAN-interface 11, and enable IS-IS, MPLS, and LDP on the interface.

```
[PE2] interface vlan-interface 11
[PE2-Vlan-interface11] ip address 9.1.1.2 255.0.0.0
[PE2-Vlan-interface11] isis enable 1
[PE2-Vlan-interface11] mpls enable
[PE2-Vlan-interface11] mpls ldp enable
[PE2-Vlan-interface11] quit
```

Configure interface Loopback 0, and enable IS-IS on it.

```
[PE2] interface loopback 0
[PE2-LoopBack0] ip address 5.5.5.9 32
[PE2-LoopBack0] isis enable 1
[PE2-LoopBack0] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 11:11
[PE2-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE2-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE2-vpn-instance-vpn1] quit
```

Associate interface VLAN-interface 12 with VPN instance **vpn1**, and specify an IP address for the interface.

```
[PE2] interface vlan-interface 12
[PE2-Vlan-interface12] ip binding vpn-instance vpn1
[PE2-Vlan-interface12] ip address 20.0.0.1 24
[PE2-Vlan-interface12] quit
```

Enable BGP on PE 2.

```
[PE2] bgp 600
```

Configure the maximum hop count from PE 2 to EBGP peer 2.2.2.9 as 10.

```
[PE2-bgp-default] peer 2.2.2.9 as-number 100
[PE2-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[PE2-bgp-default] peer 2.2.2.9 ebgp-max-hop 10
```

Configure peer 2.2.2.9 as a VPNv4 peer.

```
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] peer 2.2.2.9 enable
[PE2-bgp-default-vpnv4] quit
```

Establish an EBGP peer relationship with CE 2, and add the learned BGP routes to the routing table of VPN instance **vpn1**.

```
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] peer 20.0.0.2 as-number 65002
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] peer 20.0.0.2 enable
[PE2-bgp-default-ipv4-vpn1] quit
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

6. Configure CE 2:

Configure an IP address for interface VLAN-interface 12.

```
<CE2> system-view
[CE2] interface vlan-interface 12
[CE2-Vlan-interface12] ip address 20.0.0.2 24
[CE2-Vlan-interface12] quit
```

Establish an EBGP peer relationship with PE 2, and redistribute VPN routes.

```
[CE2] bgp 65002
[CE2-bgp-default] peer 20.0.0.1 as-number 600
[CE2-bgp-default] address-family ipv4 unicast
[CE2-bgp-default-ipv4] peer 20.0.0.1 enable
[CE2-bgp-default-ipv4] import-route direct
```

```
[CE2-bgp-default-ipv4] quit
[CE2-bgp-default] quit
```

Verifying the configuration

Execute the **display ip routing table** command on CE 1 and CE 2 to verify that CE 1 and CE 2 have a route to each other. Verify that CE 1 and CE 2 can ping each other. (Details not shown.)

Example: Configuring MPLS L3VPN carrier's carrier in the same AS

Network configuration

Configure carrier's carrier for the scenario shown in Figure 32. In this scenario:

- PE 1 and PE 2 are the provider carrier's PE switches. They provide VPN services for the customer carrier.
- CE 1 and CE 2 are the customer carrier's switches. They are connected to the provider carrier's backbone as CE switches.
- PE 3 and PE 4 are the customer carrier's PE switches. They provide MPLS L3VPN services for the end customers.
- CE 3 and CE 4 are customers of the customer carrier.
- The customer carrier and the provider carrier reside in the same AS.

The key to carrier's carrier deployment is to configure exchange of two kinds of routes:

- Exchange of the customer carrier's internal routes on the provider carrier's backbone.
- Exchange of the end customers' VPN routes between PE 3 and PE 4, the PEs of the customer carrier. In this process, an MP-IBGP peer relationship must be established between PE 3 and PE 4.

Figure 32 Network diagram

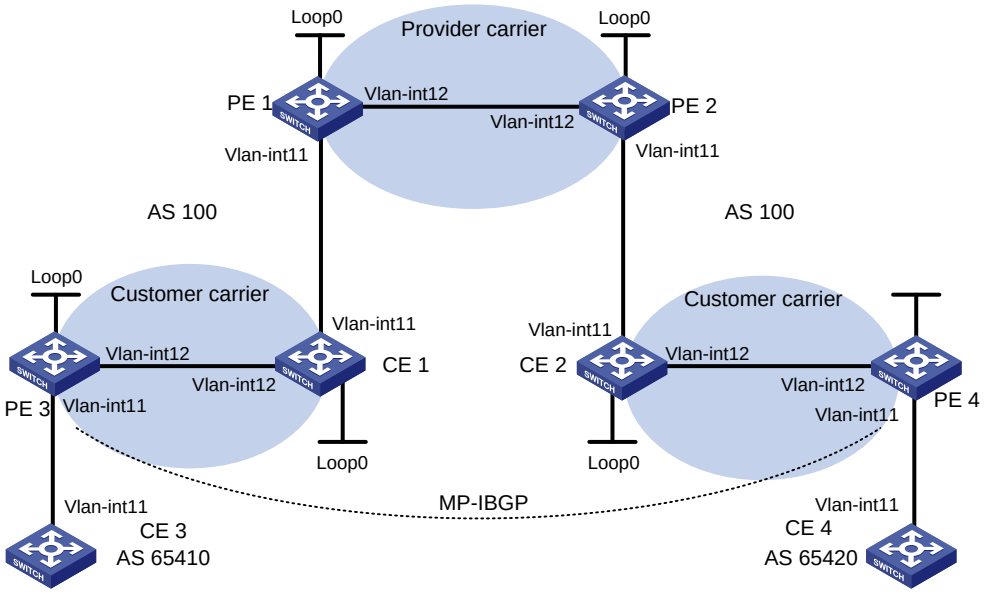


Table 18 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 3	Vlan-int11	100.1.1.1/24	CE 4	Vlan-int11	120.1.1.1/24

Device	Interface	IP address	Device	Interface	IP address
PE 3	Loop0	1.1.1.9/32	PE 4	Loop0	6.6.6.9/32
	Vlan-int11	100.1.1.2/24		Vlan-int11	120.1.1.2/24
	Vlan-int12	10.1.1.1/24		Vlan-int12	20.1.1.2/24
CE 1	Loop0	2.2.2.9/32	CE 2	Loop0	5.5.5.9/32
	Vlan-int12	10.1.1.2/24		Vlan-int11	21.1.1.2/24
	Vlan-int11	11.1.1.1/24		Vlan-int12	20.1.1.1/24
PE 1	Loop0	3.3.3.9/32	PE 2	Loop0	4.4.4.9/32
	Vlan-int11	11.1.1.2/24		Vlan-int12	30.1.1.2/24
	Vlan-int12	30.1.1.1/24		Vlan-int11	21.1.1.1/24

Procedure

1. Configure MPLS L3VPN on the provider carrier backbone. Enable IS-IS as the IGP, enable LDP between PE 1 and PE 2, and establish an MP-IBGP peer relationship between the PEs:

Configure PE 1.

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 3.3.3.9 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 3.3.3.9
[PE1] mpls ldp
[PE1-ldp] quit
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0004.00
[PE1-isis-1] quit
[PE1] interface loopback 0
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip address 30.1.1.1 24
[PE1-Vlan-interface12] isis enable 1
[PE1-Vlan-interface12] mpls enable
[PE1-Vlan-interface12] mpls ldp enable
[PE1-Vlan-interface12] mpls ldp transport-address interface
[PE1-Vlan-interface12] quit
[PE1] bgp 100
[PE1-bgp-default] peer 4.4.4.9 as-number 100
[PE1-bgp-default] peer 4.4.4.9 connect-interface loopback 0
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 4.4.4.9 enable
[PE1-bgp-default-vpnv4] quit
[PE1-bgp-default] quit
```

Configure PE 2 in the same way that PE 1 is configured. (Details not shown.)

On PE 1 or PE 2, execute the following commands:

- Execute the **display mpls ldp peer** command to verify that an LDP session in Operational state has been established between PE 1 and PE 2. (Details not shown.)
- Execute the **display bgp peer vpnv4** command to verify that a BGP peer relationship in Established state has been established between PE 1 and PE 2. (Details not shown.)
- Execute the **display isis peer** command to verify that the IS-IS neighbor relationship has been established between PE 1 and PE 2. (Details not shown.)

2. Configure the customer carrier network. Enable IS-IS as the IGP, and enable LDP between PE 3 and CE 1, and between PE 4 and CE 2:

Configure PE 3.

```
<PE3> system-view
[PE3] interface loopback 0
[PE3-LoopBack0] ip address 1.1.1.9 32
[PE3-LoopBack0] quit
[PE3] mpls lsr-id 1.1.1.9
[PE3] mpls ldp
[PE3-ldp] quit
[PE3] isis 2
[PE3-isis-2] network-entity 10.0000.0000.0000.0001.00
[PE3-isis-2] quit
[PE3] interface loopback 0
[PE3-LoopBack0] isis enable 2
[PE3-LoopBack0] quit
[PE3] interface vlan-interface 12
[PE3-Vlan-interface12] ip address 10.1.1.1 24
[PE3-Vlan-interface12] isis enable 2
[PE3-Vlan-interface12] mpls enable
[PE3-Vlan-interface12] mpls ldp enable
[PE3-Vlan-interface12] mpls ldp transport-address interface
[PE3-Vlan-interface12] quit
```

Configure CE 1.

```
<CE1> system-view
[CE1] interface loopback 0
[CE1-LoopBack0] ip address 2.2.2.9 32
[CE1-LoopBack0] quit
[CE1] mpls lsr-id 2.2.2.9
[CE1] mpls ldp
[CE1-ldp] quit
[CE1] isis 2
[CE1-isis-2] network-entity 10.0000.0000.0000.0002.00
[CE1-isis-2] quit
[CE1] interface loopback 0
[CE1-LoopBack0] isis enable 2
[CE1-LoopBack0] quit
[CE1] interface vlan-interface 12
[CE1-Vlan-interface12] ip address 10.1.1.2 24
[CE1-Vlan-interface12] isis enable 2
[CE1-Vlan-interface12] mpls enable
[CE1-Vlan-interface12] mpls ldp enable
```

```
[CE1-Vlan-interface12] mpls ldp transport-address interface
[CE1-Vlan-interface12] quit
```

PE 3 and CE 1 can establish an LDP session and IS-IS neighbor relationship between them.

Configure PE 4 and CE 2 in the same way that PE 3 and CE 1 are configured. (Details not shown.)

3. Allow CEs of the customer carrier to access PEs of the provider carrier, and redistribute IS-IS routes to BGP and BGP routes to IS-IS on the PEs:

Configure PE 1.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 200:1
[PE1-vpn-instance-vpn1] vpn-target 1:1
[PE1-vpn-instance-vpn1] quit
[PE1] mpls ldp
[PE1-ldp] vpn-instance vpn1
[PE1-ldp-vpn-instance-vpn1] quit
[PE1-ldp] quit
[PE1] isis 2 vpn-instance vpn1
[PE1-isis-2] network-entity 10.0000.0000.0000.0003.00
[PE1-isis-2] address-family ipv4
[PE1-isis-2-ipv4] import-route bgp
[PE1-isis-2-ipv4] quit
[PE1-isis-2] quit
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip binding vpn-instance vpn1
[PE1-Vlan-interface11] ip address 11.1.1.2 24
[PE1-Vlan-interface11] isis enable 2
[PE1-Vlan-interface11] mpls enable
[PE1-Vlan-interface11] mpls ldp enable
[PE1-Vlan-interface11] mpls ldp transport-address interface
[PE1-Vlan-interface11] quit
[PE1] bgp 100
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] import-route isis 2
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

Configure CE 1.

```
[CE1] interface vlan-interface 11
[CE1-Vlan-interface11] ip address 11.1.1.1 24
[CE1-Vlan-interface11] isis enable 2
[CE1-Vlan-interface11] mpls enable
[CE1-Vlan-interface11] mpls ldp enable
[CE1-Vlan-interface11] mpls ldp transport-address interface
[CE1-Vlan-interface11] quit
```

PE 1 and CE 1 can establish an LDP session and an IS-IS neighbor relationship between them.

Configure PE 2 and CE 2 in the same way that PE 1 and CE 1 are configured. (Details not shown.)

4. Connect the CEs of the end customers and the PEs of the customer carrier:

   # Configure CE 3.

```
<CE3> system-view
[CE3] interface vlan-interface11
[CE3-Vlan-interface11] ip address 100.1.1.1 24
[CE3-Vlan-interface11] quit
[CE3] bgp 65410
[CE3-bgp-default] peer 100.1.1.2 as-number 100
[CE3-bgp-default] address-family ipv4 unicast
[CE3-bgp-default-ipv4] peer 100.1.1.2 enable
[CE3-bgp-default-ipv4] import-route direct
[CE3-bgp-default-ipv4] quit
[CE3-bgp-default] quit
```

   # Configure PE 3.

```
[PE3] ip vpn-instance vpn1
[PE3-vpn-instance-vpn1] route-distinguisher 100:1
[PE3-vpn-instance-vpn1] vpn-target 1:1
[PE3-vpn-instance-vpn1] quit
[PE3] interface Vlan-interface 11
[PE3-Vlan-interface11] ip binding vpn-instance vpn1
[PE3-Vlan-interface11] ip address 100.1.1.2 24
[PE3-Vlan-interface11] quit
[PE3] bgp 100
[PE3-bgp-default] ip vpn-instance vpn1
[PE3-bgp-default-vpn1] peer 100.1.1.1 as-number 65410
[PE3-bgp-default-vpn1] address-family ipv4 unicast
[PE3-bgp-default-ipv4-vpn1] peer 100.1.1.1 enable
[PE3-bgp-default-ipv4-vpn1] quit
[PE3-bgp-default-vpn1] quit
[PE3-bgp-default] quit
```

   # Configure PE 4 and CE 4 in the same way that PE 3 and CE 3 are configured. (Details not shown.)

5. Configure MP-IBGP peer relationship between the PEs of the customer carrier to exchange the end customers' VPN routes:

   # Configure PE 3.

```
[PE3] bgp 100
[PE3-bgp-default] peer 6.6.6.9 as-number 100
[PE3-bgp-default] peer 6.6.6.9 connect-interface loopback 0
[PE3-bgp-default] address-family vpnv4
[PE3-bgp-default-vpnv4] peer 6.6.6.9 enable
[PE3-bgp-default-vpnv4] quit
[PE3-bgp-default] quit
```

   # Configure PE 4 in the same way that PE 3 is configured. (Details not shown.)

Verifying the configuration

1. Display the public network routing table and VPN routing table on the provider carrier PEs, for example, on PE 1:

   # Verify that the public network routing table contains only routes of the provider carrier network.

```
[PE1] display ip routing-table

Destinations : 14        Routes : 14

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
3.3.3.9/32         Direct  0   0           127.0.0.1       InLoop0
4.4.4.9/32         IS_L1   15  10          30.1.1.2        Vlan12
30.1.1.0/24        Direct  0   0           30.1.1.1        Vlan12
30.1.1.0/32        Direct  0   0           30.1.1.1        Vlan12
30.1.1.1/32        Direct  0   0           127.0.0.1       InLoop0
30.1.1.255/32      Direct  0   0           30.1.1.1        Vlan12
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that the VPN routing table contains the internal routes of the customer carrier network, but it does not contain the VPN routes that the customer carrier maintains.

```
[PE1] display ip routing-table vpn-instance vpn1

Destinations : 18        Routes : 18

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
1.1.1.9/32         IS_L1   15  20          11.1.1.1        Vlan11
2.2.2.9/32         IS_L1   15  10          11.1.1.1        Vlan11
5.5.5.9/32         BGP     255 10          4.4.4.9         Vlan12
6.6.6.9/32         BGP     255 20          4.4.4.9         Vlan12
10.1.1.0/24        IS_L1   15  20          11.1.1.1        Vlan11
11.1.1.0/24        Direct  0   0           11.1.1.2        Vlan11
11.1.1.0/32        Direct  0   0           11.1.1.2        Vlan11
11.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
11.1.1.255/32      Direct  0   0           11.1.1.2        Vlan11
20.1.1.0/24        BGP     255 20          4.4.4.9         Vlan12
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

2. Display the routing table on the customer carrier CEs, for example, on CE 1:

Verify that the routing table contains the internal routes of the customer carrier network, but it does not contain the VPN routes that the customer carrier maintains.

```
[CE1] display ip routing-table
```

```
Destinations : 21        Routes : 21

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
1.1.1.9/32         IS_L1   15  10          10.1.1.1        Vlan12
2.2.2.9/32         Direct  0   0           127.0.0.1       InLoop0
5.5.5.9/32         IS_L2   15  74          11.1.1.2        Vlan11
6.6.6.9/32         IS_L2   15  74          11.1.1.2        Vlan11
10.1.1.0/24        Direct  0   0           10.1.1.2        Vlan12
10.1.1.0/32        Direct  0   0           10.1.1.2        Vlan12
10.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
10.1.1.255/32      Direct  0   0           10.1.1.2        Vlan12
11.1.1.0/24        Direct  0   0           11.1.1.1        Vlan11
11.1.1.0/32        Direct  0   0           11.1.1.1        Vlan11
11.1.1.1/32        Direct  0   0           127.0.0.1       InLoop0
11.1.1.255/32      Direct  0   0           11.1.1.1        Vlan11
20.1.1.0/24        IS_L2   15  74          11.1.1.2        Vlan11
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

3. Display the public network routing table and VPN routing table on the customer carrier PEs, for example, on PE 3:

Verify that the public network routing table contains the internal routes of the customer carrier network.

```
[PE3] display ip routing-table

Destinations : 18        Routes : 18

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
1.1.1.9/32         Direct  0   0           127.0.0.1       InLoop0
2.2.2.9/32         IS_L1   15  10          10.1.1.2        Vlan12
5.5.5.9/32         IS_L2   15  84          10.1.1.2        Vlan12
6.6.6.9/32         IS_L2   15  84          10.1.1.2        Vlan12
10.1.1.0/24        Direct  0   0           10.1.1.1        Vlan12
10.1.1.0/32        Direct  0   0           10.1.1.1        Vlan12
10.1.1.1/32        Direct  0   0           127.0.0.1       InLoop0
10.1.1.255/32      Direct  0   0           10.1.1.1        Vlan12
11.1.1.0/24        IS_L1   15  20          10.1.1.2        Vlan12
20.1.1.0/24        IS_L2   15  84          10.1.1.2        Vlan12
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

```
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that the VPN routing table contains the route to the remote VPN customer.

```
[PE3] display ip routing-table vpn-instance vpn1

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
100.1.1.0/24       Direct  0   0           100.1.1.2       Vlan11
100.1.1.0/32       Direct  0   0           100.1.1.2       Vlan11
100.1.1.2/32       Direct  0   0           127.0.0.1       InLoop0
100.1.1.255/32     Direct  0   0           100.1.1.2       Vlan11
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
120.1.1.0/24       BGP     255 0           6.6.6.9         Vlan12
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

4. Verify that PE 3 and PE 4 can ping each other. (Details not shown.)
5. Verify that CE 3 and CE 4 can ping each other. (Details not shown.)

Example: Configuring MPLS L3VPN carrier's carrier in different ASs

Network configuration

Configure carrier's carrier for the scenario shown in Figure 33. In this scenario:

- PE 1 and PE 2 are the provider carrier's PE switches. They provide VPN services for the customer carrier.
- CE 1 and CE 2 are the customer carrier's switches. They are connected to the provider carrier's backbone as CE switches.
- PE 3 and PE 4 are the customer carrier's PE switches. They provide MPLS L3VPN services for the end customers.
- CE 3 and CE 4 are customers of the customer carrier.
- The customer carrier and the provider carrier reside in different ASs.

The key to carrier's carrier deployment is to configure exchange of two kinds of routes:

- Exchange of the customer carrier's internal routes on the provider carrier's backbone.
- Exchange of the end customers' VPN routes between PE 3 and PE 4, the PEs of the customer carrier. In this process, an MP-EBGP peer relationship must be established between PE 3 and PE 4.

Figure 33 Network diagram

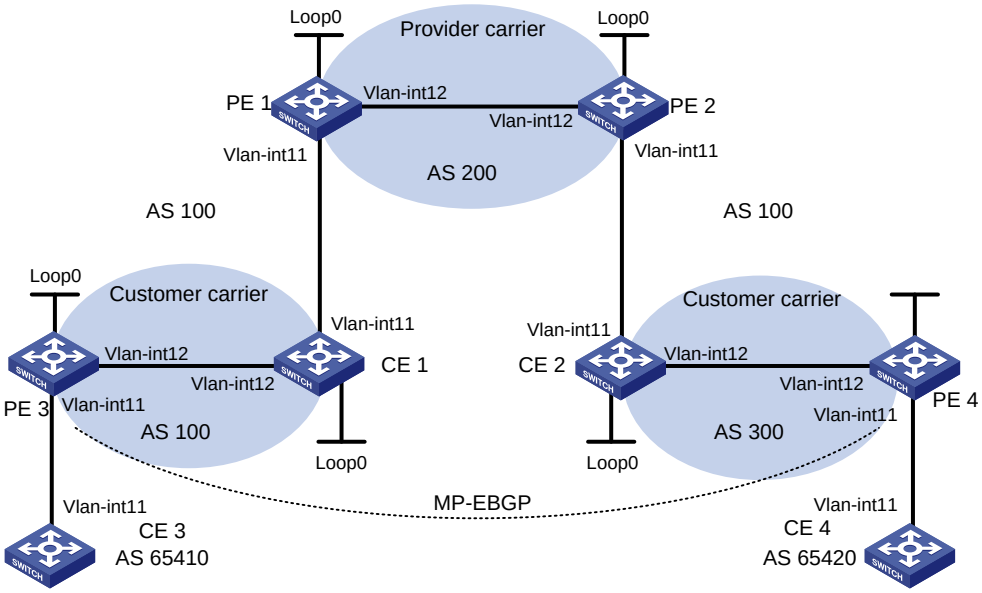


Table 19 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 3	Vlan-int11	100.1.1.1/24	CE 4	Vlan-int11	120.1.1.1/24
PE 3	Loop0	1.1.1.9/32	PE 4	Loop0	6.6.6.9/32
	Vlan-int11	100.1.1.2/24		Vlan-int11	120.1.1.2/24
	Vlan-int12	10.1.1.1/24		Vlan-int12	20.1.1.2/24
CE 1	Loop0	2.2.2.9/32	CE 2	Loop0	5.5.5.9/32
	Vlan-int12	10.1.1.2/24		Vlan-int11	21.1.1.2/24
	Vlan-int11	11.1.1.1/24		Vlan-int12	20.1.1.1/24
PE 1	Loop0	3.3.3.9/32	PE 2	Loop0	4.4.4.9/32
	Vlan-int11	11.1.1.2/24		Vlan-int12	30.1.1.2/24
	Vlan-int12	30.1.1.1/24		Vlan-int11	21.1.1.1/24

Procedure

1. Configure MPLS L3VPN on the provider carrier backbone. Enable IS-IS as the IGP, enable LDP between PE 1 and PE 2, and establish an MP-IBGP peer relationship between the PEs:

   # Configure PE 1.

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 3.3.3.9 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 3.3.3.9
[PE1] mpls ldp
[PE1-ldp] quit
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0004.00
```

```
[PE1-isis-1] quit
[PE1] interface loopback 0
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip address 30.1.1.1 24
[PE1-Vlan-interface12] isis enable 1
[PE1-Vlan-interface12] mpls enable
[PE1-Vlan-interface12] mpls ldp enable
[PE1-Vlan-interface12] mpls ldp transport-address interface
[PE1-Vlan-interface12] quit
[PE1] bgp 100
[PE1-bgp-default] peer 4.4.4.9 as-number 100
[PE1-bgp-default] peer 4.4.4.9 connect-interface loopback 0
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 4.4.4.9 enable
[PE1-bgp-default-vpnv4] quit
[PE1-bgp-default] quit
```

Configure PE 2 in the same way that PE 1 is configured. (Details not shown.)

On PE 1 or PE 2, execute the following commands:

- Execute the **display mpls ldp peer** command to verify that an LDP session in Operational state has been established between PE 1 and PE 2. (Details not shown.)
- Execute the **display bgp peer vpnv4** command to verify that a BGP peer relationship in Established state has been established between PE 1 and PE 2. (Details not shown.)
- Execute the **display isis peer** command to verify that the IS-IS neighbor relationship has been established between PE 1 and PE 2. (Details not shown.)

2. Configure the customer carrier network. Enable IS-IS as the IGP, and enable LDP between PE 3 and CE 1, and between PE 4 and CE 2:

Configure PE 3.

```
<PE3> system-view
[PE3] interface loopback 0
[PE3-LoopBack0] ip address 1.1.1.9 32
[PE3-LoopBack0] quit
[PE3] mpls lsr-id 1.1.1.9
[PE3] mpls ldp
[PE3-ldp] quit
[PE3] isis 2
[PE3-isis-2] network-entity 10.0000.0000.0000.0001.00
[PE3-isis-2] quit
[PE3] interface loopback 0
[PE3-LoopBack0] isis enable 2
[PE3-LoopBack0] quit
[PE3] interface vlan-interface 12
[PE3-Vlan-interface12] ip address 10.1.1.1 24
[PE3-Vlan-interface12] isis enable 2
[PE3-Vlan-interface12] mpls enable
[PE3-Vlan-interface12] mpls ldp enable
[PE3-Vlan-interface12] mpls ldp transport-address interface
```

```
[PE3-Vlan-interface12] quit
```

Configure CE 1.

```
<CE1> system-view
[CE1] interface loopback 0
[CE1-LoopBack0] ip address 2.2.2.9 32
[CE1-LoopBack0] quit
[CE1] mpls lsr-id 2.2.2.9
[CE1] mpls ldp
[CE1-ldp] import bgp
[CE1-ldp] quit
[CE1] isis 2
[CE1-isis-2] network-entity 10.0000.0000.0000.0002.00
[CE1-isis-2] address-family ipv4
[CE1-isis-2-ipv4] import-route bgp
[CE1-isis-2-ipv4] quit
[CE1-isis-2] quit
[CE1] interface loopback 0
[CE1-LoopBack0] isis enable 2
[CE1-LoopBack0] quit
[CE1] interface vlan-interface 12
[CE1-Vlan-interface12] ip address 10.1.1.2 24
[CE1-Vlan-interface12] isis enable 2
[CE1-Vlan-interface12] mpls enable
[CE1-Vlan-interface12] mpls ldp enable
[CE1-Vlan-interface12] mpls ldp transport-address interface
[CE1-Vlan-interface12] quit
```

PE 3 and CE 1 can establish an LDP session and IS-IS neighbor relationship between them.

Configure PE 4 and CE 2 in the same way that PE 3 and CE 1 are configured. (Details not shown.)

3. Allow CEs of the customer carrier to access PEs of the provider carrier:

Configure PE 1.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 200:1
[PE1-vpn-instance-vpn1] vpn-target 1:1
[PE1-vpn-instance-vpn1] quit
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip binding vpn-instance vpn1
[PE1-Vlan-interface11] ip address 11.1.1.2 24
[PE1-Vlan-interface11] mpls enable
[PE1-Vlan-interface11] quit
[PE1] bgp 200
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 11.1.1.1 as-number 100
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] peer 11.1.1.1 enable
[PE1-bgp-default-ipv4-vpn1] peer 11.1.1.1 label-route-capability
[PE1-bgp-default-ipv4-vpn1] peer 11.1.1.1 route-policy csc export
[PE1-bgp-default-ipv4-vpn1] quit
```

```
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
[PE1] route-policy csc permit node 0
[PE1-route-policy-csc-0] apply mpls-label
[PE1-route-policy-csc-0] quit
```

Configure CE 1.

```
[CE1] interface vlan-interface 11
[CE1-Vlan-interface11] ip address 11.1.1.1 24
[CE1-Vlan-interface11] mpls enable
[CE1-Vlan-interface11] quit
[CE1] bgp 100
[CE1-bgp-default] peer 11.1.1.2 as-number 200
[CE1-bgp-default] address-family ipv4 unicast
[CE1-bgp-default-ipv4] peer 11.1.1.2 enable
[CE1-bgp-default-ipv4] peer 11.1.1.2 label-route-capability
[CE1-bgp-default-ipv4] peer 11.1.1.2 route-policy csc export
[CE1-bgp-default-ipv4] import-route isis 2
[CE1-bgp-default-ipv4] quit
[CE1-bgp-default] quit
[CE1] route-policy csc permit node 0
[CE1-route-policy-csc-0] apply mpls-label
[CE1-route-policy-csc-0] quit
```

PE 1 and CE 1 can establish a BGP session and exchange labeled IPv4 unicast routes through BGP.

Configure PE 2 and CE 2 in the same way that PE 1 and CE 1 are configured. (Details not shown.)

4. Connect CEs of the end customers and the PEs of the customer carrier:

Configure CE 3.

```
<CE3> system-view
[CE3] interface vlan-interface 11
[CE3-Vlan-interface11] ip address 100.1.1.1 24
[CE3-Vlan-interface11] quit
[CE3] bgp 65410
[CE3-bgp-default] peer 100.1.1.2 as-number 100
[CE3-bgp-default] address-family ipv4 unicast
[CE3-bgp-default-ipv4] peer 100.1.1.2 enable
[CE3-bgp-default-ipv4] import-route direct
[CE3-bgp-default-ipv4] quit
[CE3-bgp-default] quit
```

Configure PE 3.

```
[PE3] ip vpn-instance vpn1
[PE3-vpn-instance-vpn1] route-distinguisher 100:1
[PE3-vpn-instance-vpn1] vpn-target 1:1
[PE3-vpn-instance-vpn1] quit
[PE3] interface vlan-interface 11
[PE3-Vlan-interface11] ip binding vpn-instance vpn1
[PE3-Vlan-interface11] ip address 100.1.1.2 24
[PE3-Vlan-interface11] quit
```

```
[PE3] bgp 100
[PE3-bgp-default] ip vpn-instance vpn1
[PE3-bgp-default-vpn1] peer 100.1.1.1 as-number 65410
[PE3-bgp-default-vpn1] address-family ipv4 unicast
[PE3-bgp-default-ipv4-vpn1] peer 100.1.1.1 enable
[PE3-bgp-default-ipv4-vpn1] quit
[PE3-bgp-default-vpn1] quit
[PE3-bgp-default] quit
```

Configure PE 4 and CE 4 in the same way that PE 3 and CE 3 are configured. (Details not shown.)

5. Configure an MP-EBGP peer relationship between the PEs of the customer carrier to exchange the VPN routes of the end customers:

Configure PE 3.

```
[PE3] bgp 100
[PE3-bgp-default] peer 6.6.6.9 as-number 300
[PE3-bgp-default] peer 6.6.6.9 connect-interface loopback 0
[PE3-bgp-default] peer 6.6.6.9 ebgp-max-hop 10
[PE3-bgp-default] address-family vpnv4
[PE3-bgp-default-vpnv4] peer 6.6.6.9 enable
[PE3-bgp-default-vpnv4] quit
[PE3-bgp-default] quit
```

Configure PE 4 in the same way that PE 3 is configured. (Details not shown.)

Verifying the configuration

1. Display the public network routing table and VPN routing table on the provider carrier PEs, for example, on PE 1:

Verify that the public network routing table contains only routes of the provider carrier network.

```
[PE1] display ip routing-table

Destinations : 14        Routes : 14

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
3.3.3.9/32         Direct  0   0           127.0.0.1       InLoop0
4.4.4.9/32         IS_L1   15  10          30.1.1.2        Vlan12
30.1.1.0/24        Direct  0   0           30.1.1.1        Vlan12
30.1.1.0/32        Direct  0   0           30.1.1.1        Vlan12
30.1.1.1/32        Direct  0   0           127.0.0.1       InLoop0
30.1.1.255/32      Direct  0   0           30.1.1.1        Vlan12
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that the VPN routing table contains the internal routes of the customer carrier, but it does not contain the VPN routes that the customer carrier maintains.

```
[PE1] display ip routing-table vpn-instance vpn1

Destinations : 14        Routes : 14

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
1.1.1.9/32         BGP     255 10          11.1.1.1        Vlan11
6.6.6.9/32         BGP     255 10          4.4.4.9         Vlan12
11.1.1.0/24        Direct  0   0           11.1.1.2        Vlan11
11.1.1.0/32        Direct  0   0           11.1.1.2        Vlan11
11.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
11.1.1.255/32      Direct  0   0           11.1.1.2        Vlan11
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

2. Display the routing table on the customer carrier CEs, for example, on CE 1.

Verify that the routing table contains the internal routes of the customer carrier network, but it does not contain the VPN routes that the customer carrier maintains.

```
[CE1] display ip routing-table

Destinations : 19        Routes : 19

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
1.1.1.9/32         IS_L1   15  10          10.1.1.1        Vlan12
2.2.2.9/32         Direct  0   0           127.0.0.1       InLoop0
6.6.6.9/32         BGP     255 0           11.1.1.2        Vlan11
10.1.1.0/24        Direct  0   0           10.1.1.2        Vlan12
10.1.1.0/32        Direct  0   0           10.1.1.2        Vlan12
10.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
10.1.1.255/32      Direct  0   0           10.1.1.2        Vlan12
11.1.1.0/24        Direct  0   0           11.1.1.1        Vlan11
11.1.1.0/32        Direct  0   0           11.1.1.1        Vlan11
11.1.1.1/32        Direct  0   0           127.0.0.1       InLoop0
11.1.1.255/32      Direct  0   0           11.1.1.1        Vlan11
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

3. Display the public network routing table and VPN routing table on the customer carrier PEs, for example, on PE 3:

Verify that the public network routing table contains the internal routes of the customer carrier network.

```
[PE3] display ip routing-table

Destinations : 15        Routes : 15

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
1.1.1.9/32         Direct  0   0           127.0.0.1       InLoop0
2.2.2.9/32         IS_L1   15  10          10.1.1.2        Vlan12
6.6.6.9/32         IS_L2   15  74          10.1.1.2        Vlan12
10.1.1.0/24        Direct  0   0           10.1.1.1        Vlan12
10.1.1.0/32        Direct  0   0           10.1.1.1        Vlan12
10.1.1.1/32        Direct  0   0           127.0.0.1       InLoop0
10.1.1.255/32      Direct  0   0           10.1.1.1        Vlan12
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that the VPN routing table contains the route to the remote VPN customer.

```
[PE3] display ip routing-table vpn-instance vpn1

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
100.1.1.0/24       Direct  0   0           100.1.1.2       Vlan11
100.1.1.0/32       Direct  0   0           100.1.1.2       Vlan11
100.1.1.2/32       Direct  0   0           127.0.0.1       InLoop0
100.1.1.255/32     Direct  0   0           100.1.1.2       Vlan11
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
120.1.1.0/24       BGP     255 0           6.6.6.9         Vlan12
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

4. Verify that PE 3 and PE 4 can ping each other. (Details not shown.)
5. Verify that CE 3 and CE 4 can ping each other. (Details not shown.)

Example: Configuring nested VPN

Network configuration

The service provider provides nested VPN services for users, as shown in Figure 34.

- PE 1 and PE 2 are PE devices on the service provider backbone. Both of them support the nested VPN feature.
- CE 1 and CE 2 are connected to the service provider backbone. Both of them support VPNv4 routes.
- PE 3 and PE 4 are PE devices of the customer VPN. Both of them support MPLS L3VPN.
- CE 3 through CE 6 are CE devices of the sub-VPNs for the customer VPN.

The key of nested VPN configuration is to understand the processing of routes of sub-VPNs on the service provider PEs:

- When receiving a VPNv4 route from a CE (CE 1 or CE 2 in this example), a service provider PE performs the following operations:
  a. Replaces the RD of the VPNv4 route with the RD of the MPLS VPN on the service provider network where the CE resides.
  b. Adds the export target attribute of the MPLS VPN on the service provider network to the extended community attribute list.
  c. Forwards the VPNv4 route.
- To implement exchange of sub-VPN routes between customer PEs and service provider PEs, MP-EBGP peers must be established between service provider PEs and customer CEs.

Figure 34 Network diagram

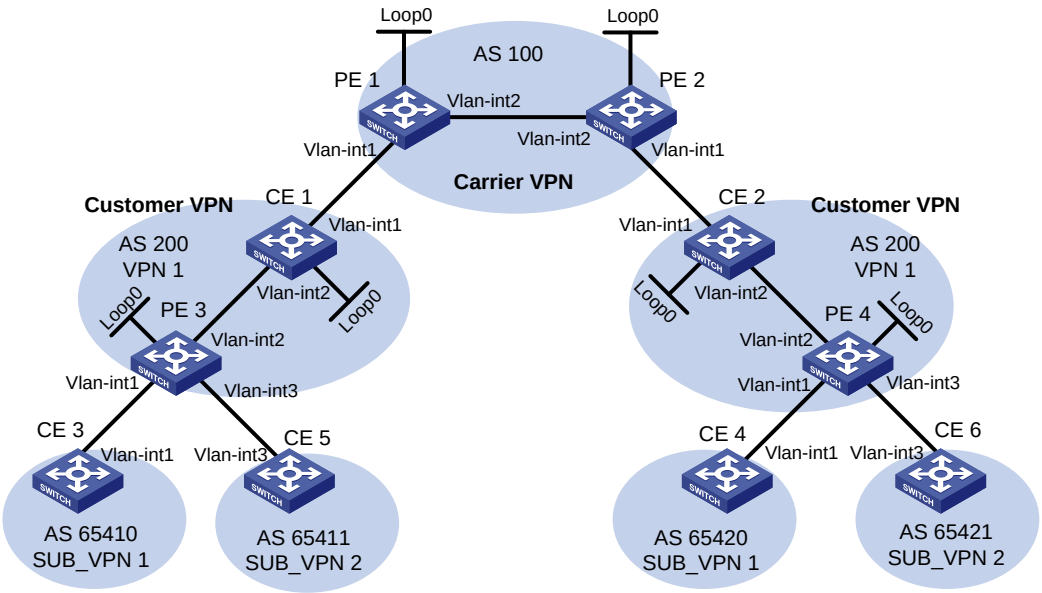


Table 20 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Loop0	2.2.2.9/32	CE 2	Loop0	5.5.5.9/32
	Vlan-int2	10.1.1.2/24		Vlan-int1	21.1.1.2/24
	Vlan-int1	11.1.1.1/24		Vlan-int2	20.1.1.1/24
CE 3	Vlan-int1	100.1.1.1/24	CE 4	Vlan-int1	120.1.1.1/24
CE 5	Vlan-int3	110.1.1.1/24	CE 6	Vlan-int3	130.1.1.1/24
PE 1	Loop0	3.3.3.9/32	PE 2	Loop0	4.4.4.9/32
	Vlan-int1	11.1.1.2/24		Vlan-int1	21.1.1.1/24

Device	Interface	IP address	Device	Interface	IP address
	Vlan-int2	30.1.1.1/24		Vlan-int2	30.1.1.2/24
PE 3	Loop0	1.1.1.9/32	PE 4	Loop0	6.6.6.9/32
	Vlan-int1	100.1.1.2/24		Vlan-int1	120.1.1.2/24
	Vlan-int2	10.1.1.1/24		Vlan-int2	20.1.1.2/24
	Vlan-int3	110.1.1.2/24		Vlan-int3	130.1.1.2/24

Procedure

1. Configure MPLS L3VPN on the service provider backbone. Use IS-IS as the IGP protocol, enable LDP, and establish an MP-IBGP peer relationship between PE 1 and PE 2:

Configure PE 1.

```
<PE1> system-view
[PE1] interface loopback 0
[PE1-LoopBack0] ip address 3.3.3.9 32
[PE1-LoopBack0] quit
[PE1] mpls lsr-id 3.3.3.9
[PE1] mpls ldp
[PE1-ldp] quit
[PE1] isis 1
[PE1-isis-1] network-entity 10.0000.0000.0000.0004.00
[PE1-isis-1] quit
[PE1] interface loopback 0
[PE1-LoopBack0] isis enable 1
[PE1-LoopBack0] quit
[PE1] interface vlan-interface 2
[PE1-Vlan-interface2] ip address 30.1.1.1 24
[PE1-Vlan-interface2] isis enable 1
[PE1-Vlan-interface2] mpls enable
[PE1-Vlan-interface2] mpls ldp enable
[PE1-Vlan-interface2] quit
[PE1] bgp 100
[PE1-bgp-default] peer 4.4.4.9 as-number 100
[PE1-bgp-default] peer 4.4.4.9 connect-interface loopback 0
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 4.4.4.9 enable
[PE1-bgp-default-vpnv4] quit
[PE1-bgp-default] quit
```

Configure PE 2 in the same way that PE 1 is configured. (Details not shown.)

On PE 1 or PE 2, execute the following commands:

- Execute the **`display mpls ldp peer`** command to verify that an LDP session in Operational state has been established between PE 1 and PE 2. (Details not shown.)
- Execute the **`display bgp peer vpnv4`** command to verify that a BGP peer relationship in Established state has been established between PE 1 and PE 2. (Details not shown.)
- Execute the **`display isis peer`** command to verify that the IS-IS neighbor relationship has been established between PE 1 and PE 2. (Details not shown.)

2. Configure the customer VPN. Use IS-IS as the IGP protocol, and enable LDP between PE 3 and CE 1, and between PE 4 and CE 2:

Configure PE 3.

```
<PE3> system-view
[PE3] interface loopback 0
[PE3-LoopBack0] ip address 1.1.1.9 32
[PE3-LoopBack0] quit
[PE3] mpls lsr-id 1.1.1.9
[PE3] mpls ldp
[PE3-ldp] quit
[PE3] isis 2
[PE3-isis-2] network-entity 10.0000.0000.0000.0001.00
[PE3-isis-2] quit
[PE3] interface loopback 0
[PE3-LoopBack0] isis enable 2
[PE3-LoopBack0] quit
[PE3] interface vlan-interface 2
[PE3-Vlan-interface2] ip address 10.1.1.1 24
[PE3-Vlan-interface2] isis enable 2
[PE3-Vlan-interface2] mpls enable
[PE3-Vlan-interface2] mpls ldp enable
[PE3-Vlan-interface2] quit
```

Configure CE 1.

```
<CE1> system-view
[CE1] interface loopback 0
[CE1-LoopBack0] ip address 2.2.2.9 32
[CE1-LoopBack0] quit
[CE1] mpls lsr-id 2.2.2.9
[CE1] mpls ldp
[CE1-ldp] quit
[CE1] isis 2
[CE1-isis-2] network-entity 10.0000.0000.0000.0002.00
[CE1-isis-2] quit
[CE1] interface loopback 0
[CE1-LoopBack0] isis enable 2
[CE1-LoopBack0] quit
[CE1] interface vlan-interface 2
[CE1-Vlan-interface2] ip address 10.1.1.2 24
[CE1-Vlan-interface2] isis enable 2
[CE1-Vlan-interface2] mpls enable
[CE1-Vlan-interface2] mpls ldp enable
[CE1-Vlan-interface2] quit
```

An LDP session and an IS-IS neighbor relationship can be established between PE 3 and CE 1.

Configure PE 4 and CE 2 in the same way that PE 3 and CE 1 are configured. (Details not shown.)

3. Connect CE 1 and CE 2 to service provider PEs:

Configure PE 1.

```
[PE1] ip vpn-instance vpn1
```

```
[PE1-vpn-instance-vpn1] route-distinguisher 200:1
[PE1-vpn-instance-vpn1] vpn-target 1:1
[PE1-vpn-instance-vpn1] quit
[PE1] interface vlan-interface 1
[PE1-Vlan-interface1] ip binding vpn-instance vpn1
[PE1-Vlan-interface1] ip address 11.1.1.2 24
[PE1-Vlan-interface1] mpls enable
[PE1-Vlan-interface1] quit
[PE1] bgp 100
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 11.1.1.1 as-number 200
[PE1-bgp-default-vpn1] address-family ipv4
[PE1-bgp-default-ipv4-vpn1] peer 11.1.1.1 enable
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

Configure CE 1.

```
[CE1] interface vlan-interface 1
[CE1-Vlan-interface1] ip address 11.1.1.1 24
[CE1-Vlan-interface1] mpls enable
[CE1-Vlan-interface1] quit
[CE1] bgp 200
[CE1-bgp-default] peer 11.1.1.2 as-number 100
[CE1-bgp-default] address-family ipv4
[CE1-bgp-default-ipv4] peer 11.1.1.2 enable
[CE1-bgp-default-ipv4] quit
[CE1-bgp-default] quit
```

Configure PE 2 and CE 2 in the same way that PE 1 and CE 1 are configured. (Details not shown.)

4. Connect sub-VPN CEs to the customer VPN PEs:

Configure CE 3.

```
<CE3> system-view
[CE3] interface vlan-interface 1
[CE3-Vlan-interface1] ip address 100.1.1.1 24
[CE3-Vlan-interface1] quit
[CE3] bgp 65410
[CE3-bgp-default] peer 100.1.1.2 as-number 200
[CE3-bgp-default] address-family ipv4 unicast
[CE3-bgp-default-ipv4] peer 100.1.1.2 enable
[CE3-bgp-default-ipv4] import-route direct
[CE3-bgp-default-ipv4] quit
[CE3-bgp-default] quit
```

Configure CE 5.

```
<CE5> system-view
[CE5] interface vlan-interface 3
[CE5-Vlan-interface3] ip address 110.1.1.1 24
[CE5-Vlan-interface3] quit
[CE5] bgp 65411
```

```
[CE5-bgp-default] peer 110.1.1.2 as-number 200
[CE5-bgp-default] address-family ipv4 unicast
[CE5-bgp-default-ipv4] peer 110.1.1.2 enable
[CE5-bgp-default-ipv4] import-route direct
[CE5-bgp-default-ipv4] quit
[CE5-bgp-default] quit
```

Configure PE 3.

```
[PE3] ip vpn-instance SUB_VPN1
[PE3-vpn-instance-SUB_VPN1] route-distinguisher 100:1
[PE3-vpn-instance-SUB_VPN1] vpn-target 2:1
[PE3-vpn-instance-SUB_VPN1] quit
[PE3] interface vlan-interface 1
[PE3-Vlan-interface1] ip binding vpn-instance SUB_VPN1
[PE3-Vlan-interface1] ip address 100.1.1.2 24
[PE3-Vlan-interface1] quit
[PE3] ip vpn-instance SUB_VPN2
[PE3-vpn-instance-SUB_VPN2] route-distinguisher 101:1
[PE3-vpn-instance-SUB_VPN2] vpn-target 2:2
[PE3-vpn-instance-SUB_VPN2] quit
[PE3] interface vlan-interface 3
[PE3-Vlan-interface3] ip binding vpn-instance SUB_VPN2
[PE3-Vlan-interface3] ip address 110.1.1.2 24
[PE3-Vlan-interface3] quit
[PE3] bgp 200
[PE3-bgp-default] ip vpn-instance SUB_VPN1
[PE3-bgp-default-SUB_VPN1] peer 100.1.1.1 as-number 65410
[PE3-bgp-default-SUB_VPN1] address-family ipv4 unicast
[PE3-bgp-default-ipv4-SUB_VPN1] peer 100.1.1.1 enable
[PE3-bgp-default-ipv4-SUB_VPN1] quit
[PE3-bgp-default-SUB_VPN1] quit
[PE3-bgp-default] ip vpn-instance SUB_VPN2
[PE3-bgp-default-SUB_VPN2] peer 100.1.1.1 as-number 65411
[PE3-bgp-default-SUB_VPN2] address-family ipv4 unicast
[PE3-bgp-default-ipv4-SUB_VPN2] peer 110.1.1.1 enable
[PE3-bgp-default-ipv4-SUB_VPN2] quit
[PE3-bgp-default-SUB_VPN2] quit
[PE3-bgp-default] quit
```

Configure PE 4, CE 4, and CE 6 in the same way that PE 3, CE 3, and CE 5 are configured. (Details not shown.)

5. Establish MP-EBGP peer relationships between service provider PEs and their CEs to exchange user VPNv4 routes:

On PE 1, enable nested VPN and VPNv4 route exchange with CE 1.

```
[PE1] bgp 100
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] nesting-vpn
[PE1-bgp-default-vpnv4] quit
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] address-family vpnv4
```

```
[PE1-bgp-default-vpnv4-vpn1] peer 11.1.1.1 enable
[PE1-bgp-default-vpnv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

Enable CE 1 to exchange VPNv4 routes with PE 1.

```
[CE1] bgp 200
[CE1-bgp-default] address-family vpnv4
[CE1-bgp-default-vpnv4] peer 11.1.1.2 enable
```

Allow the local AS number to appear in the AS-PATH attribute of the routes received.

```
[CE1-bgp-default-vpnv4] peer 11.1.1.2 allow-as-loop 2
```

Disable route target based filtering of received VPNv4 routes.

```
[CE1-bgp-default-vpnv4] undo policy vpn-target
[CE1-bgp-default-vpnv4] quit
[CE1-bgp-default] quit
```

Configure PE 2 and CE 2 in the same way that PE 1 and CE 1 are configured. (Details not shown.)

6. Establish MP-IBGP peer relationships between sub-VPN PEs and CEs of the customer VPN to exchange VPNv4 routes of sub-VPNs:

Configure PE 3.

```
[PE3] bgp 200
[PE3-bgp-default] peer 2.2.2.9 as-number 200
[PE3-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[PE3-bgp-default] address-family vpnv4
[PE3-bgp-default-vpnv4] peer 2.2.2.9 enable
```

Allow the local AS number to appear in the AS-PATH attribute of the routes received.

```
[PE3-bgp-default-vpnv4] peer 2.2.2.9 allow-as-loop 2
[PE3-bgp-default-vpnv4] quit
[PE3-bgp-default] quit
```

Configure CE 1.

```
[CE1] bgp 200
[CE1-bgp-default] peer 1.1.1.9 as-number 200
[CE1-bgp-default] peer 1.1.1.9 connect-interface loopback 0
[CE1-bgp-default] address-family vpnv4
[CE1-bgp-default-vpnv4] peer 1.1.1.9 enable
[CE1-bgp-default-vpnv4] undo policy vpn-target
[CE1-bgp-default-vpnv4] quit
[CE1-bgp-default] quit
```

Configure PE 4 and CE 2 in the same way that PE 3 and CE 1 are configured. (Details not shown.)

Verifying the configuration

1. Display the public routing table and VPN routing table on the provider PEs, for example, on PE 1:

Verify that the public routing table contains only routes on the service provider network.

```
[PE1] display ip routing-table

Destinations : 14        Routes : 14

Destination/Mask   Proto   Pre Cost        NextHop         Interface
```

```
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
3.3.3.9/32         Direct  0   0           127.0.0.1       InLoop0
4.4.4.9/32         IS_L1   15  10          30.1.1.2        Vlan2
30.1.1.0/24        Direct  0   0           30.1.1.1        Vlan2
30.1.1.0/32        Direct  0   0           30.1.1.1        Vlan2
30.1.1.1/32        Direct  0   0           127.0.0.1       InLoop0
30.1.1.255/32      Direct  0   0           30.1.1.1        Vlan2
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that the VPN routing table contains sub-VPN routes.

```
[PE1] display ip routing-table vpn-instance vpn1

Destinations : 17       Routes : 17

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
11.1.1.0/24        Direct  0   0           11.1.1.2        Vlan1
11.1.1.0/32        Direct  0   0           11.1.1.2        Vlan1
11.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
11.1.1.255/32      Direct  0   0           11.1.1.2        Vlan1
100.1.1.0/24       BGP     255 0           11.1.1.1        Vlan1
110.1.1.0/24       BGP     255 0           11.1.1.1        Vlan1
120.1.1.0/24       BGP     255 0           4.4.4.9         Vlan2
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
130.1.1.0/24       BGP     255 0           4.4.4.9         Vlan2
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

2. Display the VPNv4 routing table on the provider CEs, for example, on CE 1.

Verify that the VPNv4 routing table on the customer VPN contains internal sub-VPN routes.

```
[CE1] display bgp routing-table vpnv4

 BGP local router ID is 2.2.2.9
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
               Origin: i - IGP, e - EGP, ? - incomplete

 Total number of routes from all PEs: 4
```

```
 Route distinguisher: 100:1
 Total number of routes: 1

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >i 100.1.1.0/24       1.1.1.9         0          100        0       200 65410?

 Route distinguisher: 101:1
 Total number of routes: 1

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >i 110.1.1.0/24       1.1.1.9         0          100        0       200 65411?

 Route distinguisher: 200:1
 Total number of routes: 1

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 120.1.1.0/24       11.1.1.2                               0       100 200
                                                                       65420?

 Route Distinguisher: 201:1
 Total number of routes: 1

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 130.1.1.0/24       11.1.1.2                               0       100 200
                                                                       65421?
```

3. Display the VPN routing table on the customer PEs, for example, on PE 3:

Verify that the VPN routing table contains routes sent by the provider PE to the sub-VPN.

```
[PE3] display ip routing-table vpn-instance SUB_VPN1

Destinations : 11        Routes : 11

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
100.1.1.0/24       Direct  0   0           100.1.1.2       Vlan1
100.1.1.2/32       Direct  0   0           127.0.0.1       InLoop0
120.1.1.0/24       BGP     255 0           2.2.2.9         Vlan2
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

4. Display the routing table on the CEs of sub-VPNs in the customer VPN, for example, on CE 3 and CE 5:

Verify that the routing table contains the route to the remote sub-VPN on CE 3.

```
[CE3] display ip routing-table

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
100.1.1.0/24       Direct  0   0           100.1.1.1       Vlan1
100.1.1.0/32       Direct  0   0           100.1.1.1       Vlan1
100.1.1.1/32       Direct  0   0           127.0.0.1       InLoop0
100.1.1.255/32     Direct  0   0           100.1.1.1       Vlan1
120.1.1.0/24       BGP     255 0           100.1.1.2       Vlan1
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that the routing table contains the route to the remote sub-VPN on CE 5.

```
[CE5] display ip routing-table

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
110.1.1.0/24       Direct  0   0           110.1.1.1       Vlan1
110.1.1.0/32       Direct  0   0           110.1.1.1       Vlan1
110.1.1.1/32       Direct  0   0           127.0.0.1       InLoop0
110.1.1.255/32     Direct  0   0           110.1.1.1       Vlan1
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
130.1.1.0/24       BGP     255 0           110.1.1.2       Vlan1
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

5. Verify that CE 3 and CE 4 can ping each other. (Details not shown.)

6. Verify that CE5 and CE 6 can ping each other. (Details not shown.)

7. Verify that CE 3 and CE 6 cannot ping each other. (Details not shown.)

Example: Configuring HoVPN

Network configuration

There are two levels of networks, the backbone and the MPLS VPN networks, as shown in Figure 35.

- SPEs act as PEs to allow MPLS VPNs to access the backbone.
- UPEs act as PEs of the MPLS VPNs to allow end users to access the VPNs.
- Performance requirements for the UPEs are lower than those for the SPEs.
- SPEs advertise routes permitted by the routing policies to UPEs, permitting CE 1 and CE 3 in VPN 1 to communicate with each other, and forbidding CE 2 and CE 4 in VPN 2 from communicating with each other.

Figure 35 Network diagram

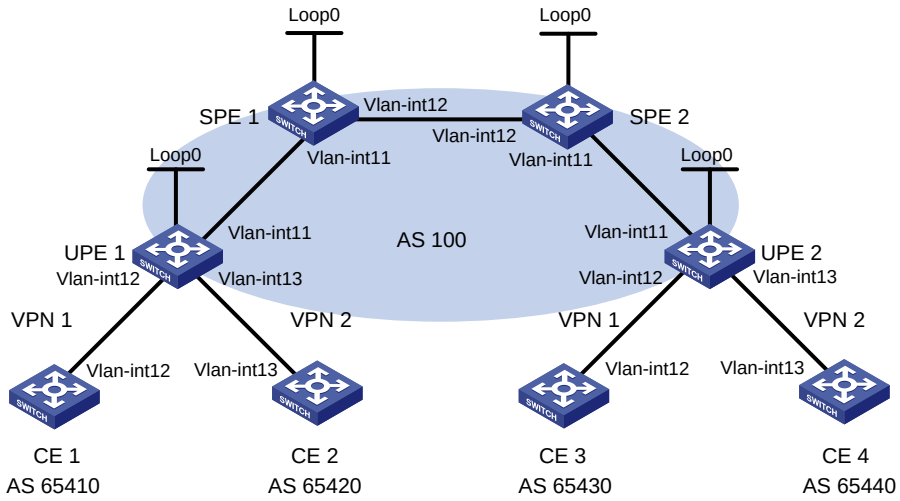


Table 21 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Vlan-int12	10.2.1.1/24	CE 3	Vlan-int12	10.1.1.1/24
CE 2	Vlan-int13	10.4.1.1/24	CE 4	Vlan-int13	10.3.1.1/24
UPE 1	Loop0	1.1.1.9/32	UPE 2	Loop0	4.4.4.9/32
	Vlan-int11	172.1.1.1/24		Vlan-int11	172.2.1.1/24
	Vlan-int12	10.2.1.2/24		Vlan-int12	10.1.1.2/24
	Vlan-int13	10.4.1.2/24		Vlan-int13	10.3.1.2/24
SPE 1	Loop0	2.2.2.9/32	SPE 2	Loop0	3.3.3.9/32
	Vlan-int11	172.1.1.2/24		Vlan-int11	172.2.1.2/24
	Vlan-int12	180.1.1.1/24		Vlan-int12	180.1.1.2/24

Procedure

1. Configure UPE 1:

   # Configure basic MPLS and MPLS LDP to establish LDP LSPs.

   ```
   <UPE1> system-view
   [UPE1] interface loopback 0
   ```

```
[UPE1-LoopBack0] ip address 1.1.1.9 32
[UPE1-LoopBack0] quit
[UPE1] mpls lsr-id 1.1.1.9
[UPE1] mpls ldp
[UPE1-ldp] quit
[UPE1] interface vlan-interface 11
[UPE1-Vlan-interface11] ip address 172.1.1.1 24
[UPE1-Vlan-interface11] mpls enable
[UPE1-Vlan-interface11] mpls ldp enable
[UPE1-Vlan-interface11] quit
```

Configure the IGP protocol (OSPF, in this example).

```
[UPE1] ospf
[UPE1-ospf-1] area 0
[UPE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[UPE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[UPE1-ospf-1-area-0.0.0.0] quit
[UPE1-ospf-1] quit
```

Configure VPN instances **vpn1** and **vpn2**, allowing CE 1 and CE 2 to access UPE 1.

```
[UPE1] ip vpn-instance vpn1
[UPE1-vpn-instance-vpn1] route-distinguisher 100:1
[UPE1-vpn-instance-vpn1] vpn-target 100:1 both
[UPE1-vpn-instance-vpn1] quit
[UPE1] ip vpn-instance vpn2
[UPE1-vpn-instance-vpn2] route-distinguisher 100:2
[UPE1-vpn-instance-vpn2] vpn-target 100:2 both
[UPE1-vpn-instance-vpn2] quit
[UPE1] interface vlan-interface 12
[UPE1-Vlan-interface12] ip binding vpn-instance vpn1
[UPE1-Vlan-interface12] ip address 10.2.1.2 24
[UPE1-Vlan-interface12] quit
[UPE1] interface vlan-interface 13
[UPE1-Vlan-interface13] ip binding vpn-instance vpn2
[UPE1-Vlan-interface13] ip address 10.4.1.2 24
[UPE1-Vlan-interface13] quit
```

Establish an MP-IBGP peer relationship with SPE 1.

```
[UPE1] bgp 100
[UPE1-bgp-default] peer 2.2.2.9 as-number 100
[UPE1-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[UPE1-bgp-default] address-family vpnv4
[UPE1-bgp-default-vpnv4] peer 2.2.2.9 enable
[UPE1-bgp-default-vpnv4] quit
```

Establish an EBGP peer relationship with CE 1.

```
[UPE1-bgp-default] ip vpn-instance vpn1
[UPE1-bgp-default-vpn1] peer 10.2.1.1 as-number 65410
[UPE1-bgp-default-vpn1] address-family ipv4 unicast
[UPE1-bgp-default-ipv4-vpn1] peer 10.2.1.1 enable
[UPE1-bgp-default-ipv4-vpn1] quit
[UPE1-bgp-default-vpn1] quit
```

Establish an EBGP peer relationship with CE 2.

```
[UPE1-bgp-default] ip vpn-instance vpn2
[UPE1-bgp-default-vpn2] peer 10.4.1.1 as-number 65420
[UPE1-bgp-default-vpn2] address-family ipv4 unicast
[UPE1-bgp-default-ipv4-vpn2] peer 10.4.1.1 enable
[UPE1-bgp-default-ipv4-vpn2] quit
[UPE1-bgp-default-vpn2] quit
[UPE1-bgp-default] quit
```

2. Configure CE 1.

```
<CE1> system-view
[CE1] interface vlan-interface 12
[CE1-Vlan-interface12] ip address 10.2.1.1 255.255.255.0
[CE1-Vlan-interface12] quit
[CE1] bgp 65410
[CE1-bgp-default] peer 10.2.1.2 as-number 100
[CE1-bgp-default] address-family ipv4 unicast
[CE1-bgp-default-ipv4] peer 10.2.1.2 enable
[CE1-bgp-default-ipv4] import-route direct
[CE1-bgp-default-ipv4] quit
[CE1-bgp-default] quit
```

3. Configure CE 2.

```
<CE2> system-view
[CE2] interface vlan-interface 13
[CE2-Vlan-interface13] ip address 10.4.1.1 255.255.255.0
[CE2-Vlan-interface13] quit
[CE2] bgp 65420
[CE2-bgp-default] peer 10.4.1.2 as-number 100
[CE2-bgp-default] address-family ipv4 unicast
[CE2-bgp-default-ipv4] peer 10.4.1.2 enable
[CE2-bgp-default-ipv4] import-route direct
[CE2-bgp-default-ipv4] quit
[CE2-bgp-default] quit
```

4. Configure UPE 2:

Configure basic MPLS and MPLS LDP to establish LDP LSPs.

```
<UPE2> system-view
[UPE2] interface loopback 0
[UPE2-Loopback0] ip address 4.4.4.9 32
[UPE2-Loopback0] quit
[UPE2] mpls lsr-id 4.4.4.9
[UPE2] mpls ldp
[UPE2-ldp] quit
[UPE2] interface vlan-interface 11
[UPE2-Vlan-interface11] ip address 172.2.1.1 24
[UPE2-Vlan-interface11] mpls enable
[UPE2-Vlan-interface11] mpls ldp enable
[UPE2-Vlan-interface11] quit
```

Configure the IGP protocol (OSPF, in this example).

```
[UPE2] ospf
```

```
[UPE2-ospf-1] area 0
[UPE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[UPE2-ospf-1-area-0.0.0.0] network 4.4.4.9 0.0.0.0
[UPE2-ospf-1-area-0.0.0.0] quit
[UPE2-ospf-1] quit
```

Configure VPN instances **vpn1** and **vpn2**, allowing CE 3 and CE 4 to access UPE 2.

```
[UPE2] ip vpn-instance vpn1
[UPE2-vpn-instance-vpn1] route-distinguisher 300:1
[UPE2-vpn-instance-vpn1] vpn-target 100:1 both
[UPE2-vpn-instance-vpn1] quit
[UPE2] ip vpn-instance vpn2
[UPE2-vpn-instance-vpn2] route-distinguisher 400:2
[UPE2-vpn-instance-vpn2] vpn-target 100:2 both
[UPE2-vpn-instance-vpn2] quit
[UPE2] interface vlan-interface 12
[UPE2-Vlan-interface12] ip binding vpn-instance vpn1
[UPE2-Vlan-interface12] ip address 10.1.1.2 24
[UPE2-Vlan-interface12] quit
[UPE2] interface vlan-interface 13
[UPE2-Vlan-interface13] ip binding vpn-instance vpn2
[UPE2-Vlan-interface13] ip address 10.3.1.2 24
[UPE2-Vlan-interface13] quit
```

Establish an MP-IBGP peer relationship with SPE 2.

```
[UPE2] bgp 100
[UPE2-bgp-default] peer 3.3.3.9 as-number 100
[UPE2-bgp-default] peer 3.3.3.9 connect-interface loopback 0
[UPE2-bgp-default] address-family vpnv4
[UPE2-bgp-default-vpnv4] peer 3.3.3.9 enable
[UPE2-bgp-default-vpnv4] quit
```

Establish an EBGP peer relationship with CE 3.

```
[UPE2-bgp-default] ip vpn-instance vpn1
[UPE2-bgp-default-vpn1] peer 10.1.1.1 as-number 65430
[UPE2-bgp-default-vpn1] address-family ipv4 unicast
[UPE2-bgp-default-ipv4-vpn1] peer 10.1.1.1 enable
[UPE2-bgp-default-ipv4-vpn1] quit
[UPE2-bgp-default-vpn1] quit
```

Establish an EBGP peer relationship with CE 4.

```
[UPE2-bgp-default] ip vpn-instance vpn2
[UPE2-bgp-default-vpn2] peer 10.3.1.1 as-number 65440
[UPE2-bgp-default-vpn2] address-family ipv4 unicast
[UPE2-bgp-default-ipv4-vpn2] peer 10.3.1.1 enable
[UPE2-bgp-default-ipv4-vpn2] quit
[UPE2-bgp-default-vpn2] quit
[UPE2-bgp-default] quit
```

5. Configure CE 3.

```
<CE3> system-view
[CE3] interface vlan-interface 12
[CE3-Vlan-interface12] ip address 10.1.1.1 255.255.255.0
```

```
[CE3-Vlan-interface12] quit
[CE3] bgp 65430
[CE3-bgp-default] peer 10.1.1.2 as-number 100
[CE3-bgp-default] address-family ipv4 unicast
[CE3-bgp-default-ipv4] peer 10.1.1.2 enable
[CE3-bgp-default-ipv4] import-route direct
[CE3-bgp-default-ipv4] quit
[CE3-bgp-default] quit
```

6. Configure CE 4.

```
<CE4> system-view
[CE4] interface vlan-interface 13
[CE4-Vlan-interface13] ip address 10.3.1.1 255.255.255.0
[CE4-Vlan-interface13] quit
[CE4] bgp 65440
[CE4-bgp-default] peer 10.3.1.2 as-number 100
[CE4-bgp-default] address-family ipv4 unicast
[CE4-bgp-default-ipv4] peer 10.3.1.2 enable
[CE4-bgp-default-ipv4] import-route direct
[CE4-bgp-default-ipv4] quit
[CE4-bgp-default] quit
```

7. Configure SPE 1:

Configure basic MPLS and MPLS LDP to establish LDP LSPs.

```
<SPE1> system-view
[SPE1] interface loopback 0
[SPE1-LoopBack0] ip address 2.2.2.9 32
[SPE1-LoopBack0] quit
[SPE1] mpls lsr-id 2.2.2.9
[SPE1] mpls ldp
[SPE1-ldp] quit
[SPE1] interface vlan-interface 11
[SPE1-Vlan-interface11] ip address 172.1.1.2 24
[SPE1-Vlan-interface11] mpls enable
[SPE1-Vlan-interface11] mpls ldp enable
[SPE1-Vlan-interface11] quit
[SPE1] interface vlan-interface 12
[SPE1-Vlan-interface12] ip address 180.1.1.1 24
[SPE1-Vlan-interface12] mpls enable
[SPE1-Vlan-interface12] mpls ldp enable
[SPE1-Vlan-interface12] quit
```

Configure the IGP protocol (OSPF, in this example).

```
[SPE1] ospf
[SPE1-ospf-1] area 0
[SPE1-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[SPE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[SPE1-ospf-1-area-0.0.0.0] network 180.1.1.0 0.0.0.255
[SPE1-ospf-1-area-0.0.0.0] quit
[SPE1-ospf-1] quit
```

Configure VPN instances **vpn1** and **vpn2**.

```
[SPE1] ip vpn-instance vpn1
[SPE1-vpn-instance-vpn1] route-distinguisher 500:1
[SPE1-vpn-instance-vpn1] vpn-target 100:1 both
[SPE1-vpn-instance-vpn1] quit
[SPE1] ip vpn-instance vpn2
[SPE1-vpn-instance-vpn2] route-distinguisher 700:1
[SPE1-vpn-instance-vpn2] vpn-target 100:2 both
[SPE1-vpn-instance-vpn2] quit
```

Establish MP-IBGP peer relationships with SPE 2 and UPE 1, and specify UPE 1 as a UPE.

```
[SPE1] bgp 100
[SPE1-bgp-default] peer 1.1.1.9 as-number 100
[SPE1-bgp-default] peer 1.1.1.9 connect-interface loopback 0
[SPE1-bgp-default] peer 3.3.3.9 as-number 100
[SPE1-bgp-default] peer 3.3.3.9 connect-interface loopback 0
[SPE1-bgp-default] address-family vpnv4
[SPE1-bgp-default-vpnv4] peer 3.3.3.9 enable
[SPE1-bgp-default-vpnv4] peer 1.1.1.9 enable
[SPE1-bgp-default-vpnv4] peer 1.1.1.9 upe
[SPE1-bgp-default-vpnv4] peer 1.1.1.9 next-hop-local
[SPE1-bgp-default-vpnv4] quit
```

Create BGP-VPN instances for VPN instances **vpn1** and **vpn2**, so the VPNv4 routes learned according to the RT attributes can be added into the BGP routing tables of the corresponding VPN instances.

```
[SPE1-bgp-default] ip vpn-instance vpn1
[SPE1-bgp-default-vpn1] quit
[SPE1-bgp-default] ip vpn-instance vpn2
[SPE1-bgp-default-vpn2] quit
[SPE1-bgp-default] quit
```

Advertise to UPE 1 the routes permitted by a routing policy (the routes of CE 3).

```
[SPE1] ip prefix-list hope index 10 permit 10.1.1.1 24
[SPE1] route-policy hope permit node 0
[SPE1-route-policy-hope-0] if-match ip address prefix-list hope
[SPE1-route-policy-hope-0] quit
[SPE1] bgp 100
[SPE1-bgp-default] address-family vpnv4
[SPE1-bgp-default-vpnv4] peer 1.1.1.9 upe route-policy hope export
[SPE1-bgp-default-vpnv4] quit
[SPE1-bgp-default] quit
```

8. Configure SPE 2:

Configure basic MPLS and MPLS LDP to establish LDP LSPs.

```
<SPE2> system-view
[SPE2] interface loopback 0
[SPE2-LoopBack0] ip address 3.3.3.9 32
[SPE2-LoopBack0] quit
[SPE2] mpls lsr-id 3.3.3.9
[SPE2] mpls ldp
[SPE2-ldp] quit
[SPE2] interface vlan-interface 12
```

```
[SPE2-Vlan-interface12] ip address 180.1.1.2 24
[SPE2-Vlan-interface12] mpls enable
[SPE2-Vlan-interface12] mpls ldp enable
[SPE2-Vlan-interface12] quit
[SPE2] interface vlan-interface 11
[SPE2-Vlan-interface11] ip address 172.2.1.2 24
[SPE2-Vlan-interface11] mpls enable
[SPE2-Vlan-interface11] mpls ldp enable
[SPE2-Vlan-interface11] quit
```

Configure the IGP protocol (OSPF, in this example).

```
[SPE2] ospf
[SPE2-ospf-1] area 0
[SPE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[SPE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[SPE2-ospf-1-area-0.0.0.0] network 180.1.1.0 0.0.0.255
[SPE2-ospf-1-area-0.0.0.0] quit
[SPE2-ospf-1] quit
```

Configure VPN instances **vpn1** and **vpn2**.

```
[SPE2] ip vpn-instance vpn1
[SPE2-vpn-instance-vpn1] route-distinguisher 600:1
[SPE2-vpn-instance-vpn1] vpn-target 100:1 both
[SPE2-vpn-instance-vpn1] quit
[SPE2] ip vpn-instance vpn2
[SPE2-vpn-instance-vpn2] route-distinguisher 800:1
[SPE2-vpn-instance-vpn2] vpn-target 100:2 both
[SPE2-vpn-instance-vpn2] quit
```

Establish MP-IBGP peer relationships with SPE 1 and UPE 2, and specify UPE 2 as a UPE.

```
[SPE2] bgp 100
[SPE2-bgp-default] peer 4.4.4.9 as-number 100
[SPE2-bgp-default] peer 4.4.4.9 connect-interface loopback 0
[SPE2-bgp-default] peer 2.2.2.9 as-number 100
[SPE2-bgp-default] peer 2.2.2.9 connect-interface loopback 0
[SPE2-bgp-default] address-family vpnv4
[SPE2-bgp-default-vpnv4] peer 2.2.2.9 enable
[SPE2-bgp-default-vpnv4] peer 4.4.4.9 enable
[SPE2-bgp-default-vpnv4] peer 4.4.4.9 upe
[SPE2-bgp-default-vpnv4] peer 4.4.4.9 next-hop-local
[SPE2-bgp-default-vpnv4] quit
```

Create BGP-VPN instances for VPN instances **vpn1** and **vpn2**, so the VPNv4 routes learned according to the RT attributes can be added into the BGP routing tables of the corresponding VPN instances.

```
[SPE2-bgp-default] ip vpn-instance vpn1
[SPE2-bgp-default-vpn1] quit
[SPE2-bgp-default] ip vpn-instance vpn2
[SPE2-bgp-default-vpn2] quit
[SPE2-bgp-default] quit
```

Advertise to UPE 2 the routes permitted by a routing policy (the routes of CE 1).

```
[SPE2] ip prefix-list hope index 10 permit  10.2.1.1 24
```

```
[SPE2] route-policy hope permit node 0
[SPE2-route-policy-hope-0] if-match ip address prefix-list hope
[SPE2-route-policy-hope-0] quit
[SPE2] bgp 100
[SPE2-bgp-default] address-family vpnv4
[SPE2-bgp-default-vpnv4] peer 4.4.4.9 upe route-policy hope export
[SPE2-bgp-default-vpnv4] quit
[SPE2-bgp-default] quit
```

Verifying the configuration

Verify that CE 1 and CE3 can learn each other's interface routes and can ping each other. CE 2 and CE 4 cannot learn each other's interface routes and cannot ping each other. (Details not shown.)

Example: Configuring an OSPF sham link

Network configuration

As shown in Figure 36, CE 1 and CE 2 belong to VPN 1. Configure an OSPF sham link between PE 1 and PE 2 so traffic between the CEs is forwarded through the MPLS backbone instead of the backdoor link.

Figure 36 Network diagram

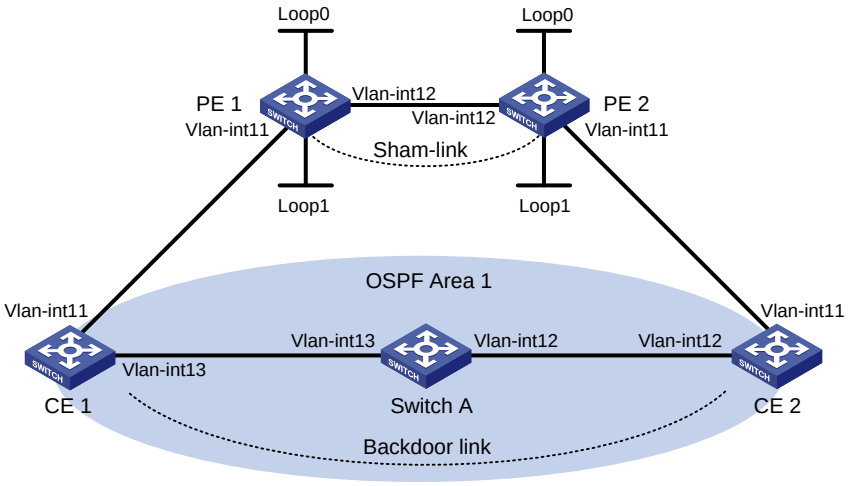


Table 22 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Vlan-int11	100.1.1.1/24	CE 2	Vlan-int11	120.1.1.1/24
	Vlan-int13	20.1.1.1/24		Vlan-int12	30.1.1.2/24
PE 1	Loop0	1.1.1.9/32	PE 2	Loop0	2.2.2.9/32
	Loop1	3.3.3.3/32		Loop1	5.5.5.5/32
	Vlan-int11	100.1.1.2/24		Vlan-int11	120.1.1.2/24
	Vlan-int12	10.1.1.1/24		Vlan-int12	10.1.1.2/24
Switch A	Vlan-int11	20.1.1.2/24			
	Vlan-int12	30.1.1.1/24			

Procedure

1. Configure OSPF on the customer networks:

   # Configure conventional OSPF on CE 1, Switch A, and CE 2 to advertise subnet addresses of the interfaces (see Figure 36Table 22). (Details not shown.)

   # Set the cost value to 2 for both the link between CE 1 and Switch A, and the link between CE 2 and Switch A. (Details not shown.)

   # Execute the **display ip routing-table** command to verify that CE 1 and CE 2 have learned the route to each other. (Details not shown.)

2. Configure MPLS L3VPN on the backbone:

   # Configure basic MPLS and MPLS LDP on PE 1 to establish LDP LSPs.

   ```
   <PE1> system-view
   [PE1] interface loopback 0
   [PE1-LoopBack0] ip address 1.1.1.9 32
   [PE1-LoopBack0] quit
   [PE1] mpls lsr-id 1.1.1.9
   [PE1] mpls ldp
   [PE1-ldp] quit
   [PE1] interface vlan-interface 12
   [PE1-Vlan-interface12] ip address 10.1.1.1 24
   [PE1-Vlan-interface12] mpls enable
   [PE1-Vlan-interface12] mpls ldp enable
   [PE1-Vlan-interface12] quit
   ```

   # Configure PE 1 to take PE 2 as an MP-IBGP peer.

   ```
   [PE1] bgp 100
   [PE1-bgp-default] peer 2.2.2.9 as-number 100
   [PE1-bgp-default] peer 2.2.2.9 connect-interface loopback 0
   [PE1-bgp-default] address-family vpnv4
   [PE1-bgp-default-vpnv4] peer 2.2.2.9 enable
   [PE1-bgp-default-vpnv4] quit
   [PE1-bgp-default] quit
   ```

   # Configure OSPF on PE 1.

   ```
   [PE1]ospf 1
   [PE1-ospf-1]area 0
   [PE1-ospf-1-area-0.0.0.0]network 1.1.1.9 0.0.0.0
   [PE1-ospf-1-area-0.0.0.0]network 10.1.1.0 0.0.0.255
   [PE1-ospf-1-area-0.0.0.0]quit
   [PE1-ospf-1]quit
   ```

   # Configure basic MPLS and MPLS LDP on PE 2 to establish LDP LSPs.

   ```
   <PE2> system-view
   [PE2] interface loopback 0
   [PE2-LoopBack0] ip address 2.2.2.9 32
   [PE2-LoopBack0] quit
   [PE2] mpls lsr-id 2.2.2.9
   [PE2] mpls ldp
   [PE2-ldp] quit
   [PE2] interface vlan-interface 12
   [PE2-Vlan-interface12] ip address 10.1.1.2 24
   [PE2-Vlan-interface12] mpls enable
   ```

```
[PE2-Vlan-interface12] mpls ldp enable
[PE2-Vlan-interface12] quit
```

Configure PE 2 to take PE 1 as an MP-IBGP peer.

```
[PE2] bgp 100
[PE2-bgp-default] peer 1.1.1.9 as-number 100
[PE2-bgp-default] peer 1.1.1.9 connect-interface loopback 0
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] peer 1.1.1.9 enable
[PE2-bgp-default-vpnv4] quit
[PE2-bgp-default] quit
```

Configure OSPF on PE 2.

```
[PE2]ospf 1
[PE2-ospf-1]area 0
[PE2-ospf-1-area-0.0.0.0]network 2.2.2.9 0.0.0.0
[PE2-ospf-1-area-0.0.0.0]network 10.1.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0]quit
[PE2-ospf-1]quit
```

3. Configure PEs to allow CE access:

Configure PE 1.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 100:1
[PE1-vpn-instance-vpn1] vpn-target 1:1
[PE1-vpn-instance-vpn1] quit
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] ip binding vpn-instance vpn1
[PE1-Vlan-interface11] ip address 100.1.1.2 24
[PE1-Vlan-interface11] quit
[PE1] ospf 100 vpn-instance vpn1
[PE1-ospf-100] domain-id 10
[PE1-ospf-100] area 1
[PE1-ospf-100-area-0.0.0.1] network 100.1.1.0 0.0.0.255
[PE1-ospf-100-area-0.0.0.1] quit
[PE1-ospf-100] quit
[PE2] bgp 100
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] import-route ospf 100
[PE1-bgp-default-ipv4-vpn1] import-route direct
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

Configure PE 2.

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 100:2
[PE2-vpn-instance-vpn1] vpn-target 1:1
[PE2-vpn-instance-vpn1] quit
[PE2] interface vlan-interface 11
[PE2-Vlan-interface11] ip binding vpn-instance vpn1
```

```
[PE2-Vlan-interface11] ip address 120.1.1.2 24
[PE2-Vlan-interface11] quit
[PE2] ospf 100 vpn-instance vpn1
[PE2-ospf-100] domain-id 10
[PE2-ospf-100] area 1
[PE2-ospf-100-area-0.0.0.1] network 120.1.1.0 0.0.0.255
[PE2-ospf-100-area-0.0.0.1] quit
[PE2-ospf-100] quit
[PE2] bgp 100
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] import-route ospf 100
[PE2-bgp-default-ipv4-vpn1] import-route direct
[PE2-bgp-default-ipv4-vpn1] quit
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

Execute the **display ip routing-table vpn-instance** command on the PEs. Verify that the path to the peer CE is along the OSPF route across the customer networks, instead of the BGP route across the backbone. (Details not shown.)

4. Configure a sham link:

Configure PE 1.

```
[PE1] interface loopback 1
[PE1-LoopBack1] ip binding vpn-instance vpn1
[PE1-LoopBack1] ip address 3.3.3.3 32
[PE1-LoopBack1] quit
[PE1] ospf 100
[PE1-ospf-100] area 1
[PE1-ospf-100-area-0.0.0.1] sham-link 3.3.3.3 5.5.5.5
[PE1-ospf-100-area-0.0.0.1] quit
[PE1-ospf-100] quit
```

Configure PE 2.

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip binding vpn-instance vpn1
[PE2-LoopBack1] ip address 5.5.5.5 32
[PE2-LoopBack1] quit
[PE2] ospf 100
[PE2-ospf-100] area 1
[PE2-ospf-100-area-0.0.0.1] sham-link 5.5.5.5 3.3.3.3
[PE2-ospf-100-area-0.0.0.1] quit
[PE2-ospf-100] quit
```

Verifying the configuration

Execute the **display ip routing-table vpn-instance** command on the PEs to verify the following results (details not shown):

- The path to the peer CE is now along the BGP route across the backbone.
- A route to the sham link destination address exists.

Execute the **display ip routing-table** command on the CEs. Verify that the next hop of the OSPF route to the peer CE is the interface connected to the PE (VLAN interface 11). This means that the VPN traffic to the peer CE is forwarded over the backbone. (Details not shown.)

Verify that a sham link has been established on PEs, for example, on PE 1.

```
[PE1] display ospf sham-link

          OSPF Process 100 with Router ID 100.1.1.2
                  Sham link
Area            Neighbor ID     Source IP       Destination IP  State  Cost
 0.0.0.1         120.1.1.2       3.3.3.3         5.5.5.5         P-2-P  1
```

Verify that the peer state is Full on PE 1.

```
[PE1] display ospf sham-link area 1

          OSPF Process 100 with Router ID 100.1.1.2

 Sham link: 3.3.3.3 --> 5.5.5.5
 Neighbor ID: 120.1.1.2        State: Full
 Area: 0.0.0.1
 Cost: 1  State: P-2-P  Type: Sham
 Timers: Hello 10, Dead 40, Retransmit 5, Transmit Delay 1
 Request list: 0  Retransmit list: 0
```

Example: Configuring BGP AS number substitution

Network configuration

As shown in Figure 37, CE 1 and CE 2 belong to VPN 1, and are connected to PE 1 and PE 2, respectively. The two CEs have the same AS number, 600.

Configure BGP AS number substitution on the PEs to enable the CEs to communicate with each other.

Figure 37 Network diagram

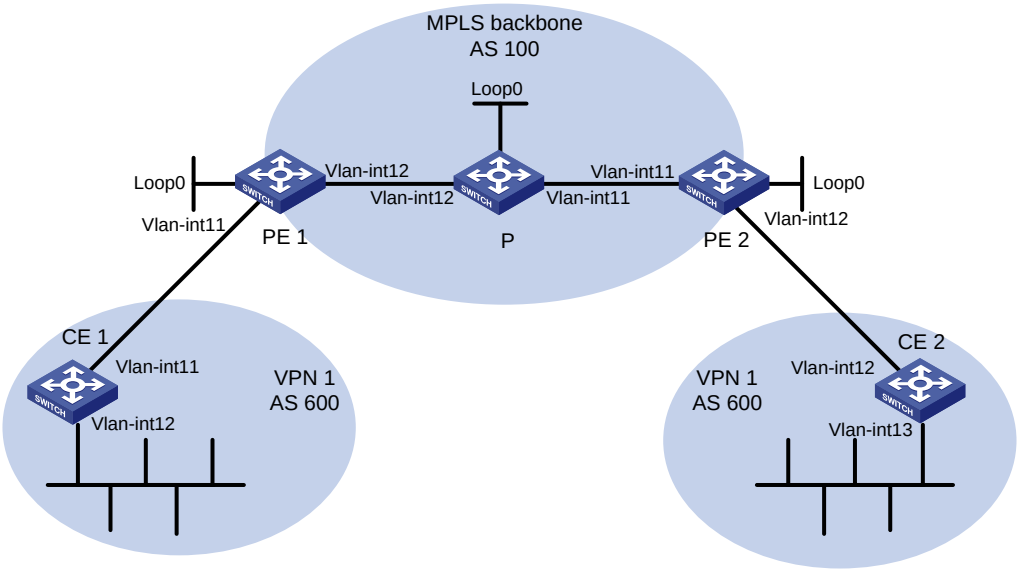


Table 23 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Vlan-int11	10.1.1.1/24	P	Loop0	2.2.2.9/32

Device	Interface	IP address	Device	Interface	IP address
	Vlan-int12	100.1.1.1/24		Vlan-int11	30.1.1.1/24
PE 1	Loop0	1.1.1.9/32		Vlan-int12	20.1.1.2/24
	Vlan-int11	10.1.1.2/24	PE 2	Loop0	3.3.3.9/32
	Vlan-int12	20.1.1.1/24		Vlan-int11	30.1.1.2/24
CE 2	Vlan-int12	10.2.1.1/24		Vlan-int12	10.2.1.2/24
	Vlan-int13	200.1.1.1/24			

Procedure

1. Configure basic MPLS L3VPN:
   - Configure OSPF on the MPLS backbone to allow the PEs and P device to learn the routes of the loopback interfaces from each other.
   - Configure basic MPLS and MPLS LDP on the MPLS backbone to establish LDP LSPs.
   - Establish an MP-IBGP peer relationship between the PEs to advertise VPNv4 routes.
   - Configure the VPN instance of VPN 1 on PE 2 to allow CE 2 to access the network.
   - Configure the VPN instance of VPN 1 on PE 1 to allow CE 1 to access the network.
   - Configure BGP as the PE-CE routing protocol, and redistribute routes of the CEs into the PEs.

   For more information about basic MPLS L3VPN configurations, see "Example: Configuring basic MPLS L3VPN."

   # Execute the **display ip routing-table** command on CE 2. The output shows that CE 2 has learned the route to network 10.1.1.0/24, where the interface used by CE 1 to access PE 1 resides. However, it has not learned the route to the VPN (100.1.1.0/24) behind CE 1.

```
<CE2> display ip routing-table

Destinations : 17        Routes : 17

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.1.1.0/24         BGP    255  0            10.2.1.2        Vlan12
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan12
10.2.1.0/32         Direct 0    0            10.2.1.1        Vlan12
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.255/32       Direct 0    0            10.2.1.1        Vlan12
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
200.1.1.0/24        Direct 0    0            200.1.1.1       Vlan13
200.1.1.0/32        Direct 0    0            200.1.1.1       Vlan13
200.1.1.1/32        Direct 0    0            127.0.0.1       InLoop0
200.1.1.255/24      Direct 0    0            200.1.1.1       Vlan13
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

Execute the **display ip routing-table** command on CE 1 to verify that CE 1 has not learned the route to the VPN behind CE 2. (Details not shown.)

Execute the **display ip routing-table vpn-instance** command on the PEs. The output shows the route to the VPN behind the peer CE. This example uses PE 2.

```
<PE2> display ip routing-table vpn-instance vpn1

Destinations : 15        Routes : 15

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.1.1.0/24         BGP    255  0            1.1.1.9         Vlan11
10.2.1.0/24         Direct 0    0            10.2.1.2        Vlan12
10.2.1.0/32         Direct 0    0            10.2.1.2        Vlan12
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.255/32       Direct 0    0            10.2.1.2        Vlan12
100.1.1.0/24        BGP    255  0            1.1.1.9         Vlan11
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
200.1.1.0/24        BGP    255  0            10.2.1.1        Vlan12
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

Enable BGP update packet debugging on PE 2. The output shows that PE 2 has advertised the route for 100.1.1.1/32, and the AS_PATH is 100 600.

```
<PE2> terminal monitor
<PE2> terminal logging level 7
<PE2> debugging bgp update vpn-instance vpn1 10.2.1.1 ipv4
<PE2> refresh bgp all export ipv4 vpn-instance vpn1
*Jun 13 16:12:52:096 2012 PE2 BGP/7/DEBUG:
        BGP.vpn1: Send UPDATE to peer 10.2.1.1 for following destinations:
        Origin       : Incomplete
        AS Path      : 100 600
        Next Hop     : 10.2.1.2
        100.1.1.0/24,
```

Execute the **display bgp routing-table ipv4 peer received-routes** command on CE 2 to verify that CE 2 has not received the route to 100.1.1.0/24.

```
<CE2> display bgp routing-table ipv4 peer 10.2.1.2 received-routes

 Total number of routes: 2

 BGP local router ID is 200.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
               Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
```

```
* >e 10.1.1.0/24        10.2.1.2                              0       100?
*  e 10.2.1.0/24        10.2.1.2        0                     0       100?
```

2. Configure BGP AS number substitution on PE 2.

```
<PE2> system-view
[PE2] bgp 100
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] peer 10.2.1.1 substitute-as
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] peer 10.2.1.1 enable
[PE2-bgp-default-ipv4-vpn1] quit
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

Verifying the configuration

The output shows that among the routes advertised by PE 2 to CE 2, the AS_PATH of 100.1.1.0/24 has changed from 100 600 to 100 100.

```
*Jun 13 16:15:59:456 2012 PE2 BGP/7/DEBUG:
        BGP.vpn1: Send UPDATE to peer 10.2.1.1 for following destinations:
        Origin       : Incomplete
        AS Path      : 100 100
        Next Hop     : 10.2.1.2
        100.1.1.0/24,
```

Display again the routing information that CE 2 has received, and the routing table.

```
<CE2> display bgp routing-table ipv4 peer 10.2.1.2 received-routes

 Total number of routes: 3

 BGP local router ID is 200.1.1.1
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
               Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 10.1.1.0/24        10.2.1.2                              0       100?
*  e 10.2.1.0/24        10.2.1.2        0                     0       100?
* >e 100.1.1.0/24       10.2.1.2                              0       100 100?
<CE2> display ip routing-table

Destinations : 18        Routes : 18

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.1.1.0/24         BGP    255  0            10.2.1.2        Vlan12
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan12
10.2.1.0/32         Direct 0    0            10.2.1.1        Vlan12
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
```

```
10.2.1.255/32       Direct 0    0           10.2.1.1        Vlan12
100.1.1.0/24        BGP    255  0           10.2.1.2        Vlan12
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0           127.0.0.1       InLoop0
200.1.1.0/24        Direct 0    0           200.1.1.1       Vlan13
200.1.1.0/32        Direct 0    0           200.1.1.1       Vlan13
200.1.1.1/32        Direct 0    0           127.0.0.1       InLoop0
200.1.1.255/32      Direct 0    0           200.1.1.1       Vlan13
224.0.0.0/4         Direct 0    0           0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0           0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0           127.0.0.1       InLoop0
```

Verify that the VLAN interfaces of CE 1 and CE 2 can ping each other. (Details not shown.)

Example: Configuring BGP AS number substitution and SoO attribute

Network configuration

CE 1, CE 2, and CE 3 belong to VPN 1, and are connected to PE1, PE 2, and PE 3, respectively.

CE 1 and CE 2 reside in the same site. CE1, CE2, and CE 3 all use AS number 600.

- To avoid route loss, configure BGP AS number substitution on PEs.
- To avoid routing loops, configure the same SoO attribute on PE 1 and PE 2 for CE 1 and CE 2.

Figure 38 Network diagram

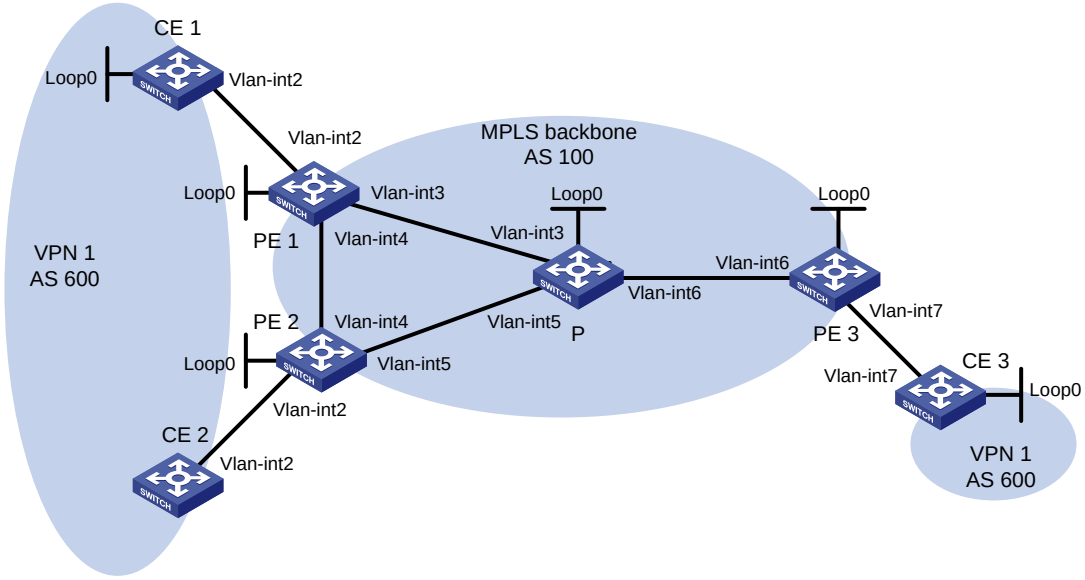


Table 24 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Loop0	100.1.1.1/32	CE 3	Loop0	200.1.1.1/32
	Vlan-int2	10.1.1.1/24		Vlan-int7	10.3.1.1/24

Device	Interface	IP address	Device	Interface	IP address
CE 2	Vlan-int2	10.2.1.1/24	PE 2	Loop0	2.2.2.9/32
PE 1	Loop0	1.1.1.9/32		Vlan-int2	10.2.1.2/24
	Vlan-int2	10.1.1.2/24		Vlan-int4	40.1.1.2/24
	Vlan-int3	30.1.1.1/24		Vlan-int5	50.1.1.1/24
	Vlan-int4	40.1.1.1/24	P	Loop0	3.3.3.9/32
PE 3	Loop0	4.4.4.9/32		Vlan-int3	30.1.1.2/24
	Vlan-int6	60.1.1.2/24		Vlan-int5	50.1.1.2/24
	Vlan-int7	10.3.1.2/24		Vlan-int6	60.1.1.1/24

Procedure

1. Configure basic MPLS L3VPN:
   - Configure OSPF on the MPLS backbone to allow the PEs and P device to learn the routes of the loopback interfaces from each other.
   - Configure basic MPLS and MPLS LDP on the MPLS backbone to establish LDP LSPs.
   - Establish an MP-IBGP peer relationship between the PEs to advertise VPN IPv4 routes.
   - Configure the VPN instance of VPN 1 on PE 1 to allow CE 1 to access the network.
   - Configure the VPN instance of VPN 1 on PE 2 to allow CE 2 to access the network.
   - Configure the VPN instance of VPN 1 on PE 3 to allow CE 3 to access the network.
   - Configure BGP as the PE-CE routing protocol, and redistribute routes of the CEs into the PEs.

   For more information about basic MPLS L3VPN configurations, see "Example: Configuring basic MPLS L3VPN."
2. Configure BGP AS number substitution:

   # Configure BGP AS number substitution on PE 1, PE 2, and PE 3. For more information about the configuration, see "Example: Configuring BGP AS number substitution."

   # Display routing information on CE 2. The output shows that CE 2 has learned the route for 100.1.1.1/32 from CE 1. A routing loop has occurred because CE1 and CE 2 reside in the same site.

```
<CE2> display bgp routing-table ipv4 peer 10.2.1.2 received-routes

 Total number of routes: 6

 BGP local router ID is 1.1.1.9
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
               Origin: i - IGP, e - EGP, ? - incomplete

     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

* >e 10.1.1.0/24        10.2.1.2                              0       100?
*    10.2.1.0/24        10.2.1.2        0                     0       100?
*    10.2.1.1/32        10.2.1.2        0                     0       100?
* >e 10.3.1.0/24        10.2.1.2                              0       100?
* >e 100.1.1.1/32       10.2.1.2                              0       100 100?
```

```
* >e 200.1.1.1/32       10.2.1.2                                0       100 100?
```

3. Configure BGP SoO attribute:

On PE 1, configure the SoO attribute as 1:100 for CE 1.

```
<PE1> system-view
[PE1] bgp 100
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] address-family ipv4
[PE1-bgp-default-ipv4-vpn1] peer 10.1.1.1 soo 1:100
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

On PE 2, configure the SoO attribute as 1:100 for CE 2.

```
<PE2> system-view
[PE2] bgp 100
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] address-family ipv4
[PE2-bgp-default-ipv4-vpn1] peer 10.2.1.1 soo 1:100
[PE2-bgp-default-ipv4-vpn1] quit
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

Verifying the configuration

PE 2 does not advertise routes received from CE 1 to CE 2 because the same SoO attribute has been configured for the CEs. Display the routing table of CE 2. The output shows that the route 100.1.1.1/32 has been removed.

```
<CE2> display ip routing-table

Destinations : 14        Routes : 14

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan2
10.2.1.0/32         Direct 0    0            10.2.1.1        Vlan2
10.2.1.1/32         Direct 0    0            127.0.0.1       Inloop0
10.2.1.255/32       Direct 0    0            10.2.1.1        Vlan2
10.3.1.0/24         BGP    255  0            10.2.1.2        Vlan2
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
200.1.1.1/32        BGP    255  0            10.2.1.2        Vlan2
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

Example: Configuring MPLS L3VPN FRR through VPNv4 route backup for an IPv4 route

Network configuration

CE 1 and CE 2 belong to VPN 1.

Configure EBGP between CEs and PEs to exchange VPN routes.

Configure OSPF to ensure connectivity between PEs, and configure MP-IBGP to exchange VPNv4 routing information between PEs.

Configure MPLS L3VPN FRR on PE 2 to achieve the following purposes:

- When the link PE 2—CE 2 operates correctly, traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—CE 2.
- When BFD detects that the link between PE 2 and CE 2 fails, traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—PE 3—CE 2.

Figure 39 Network diagram

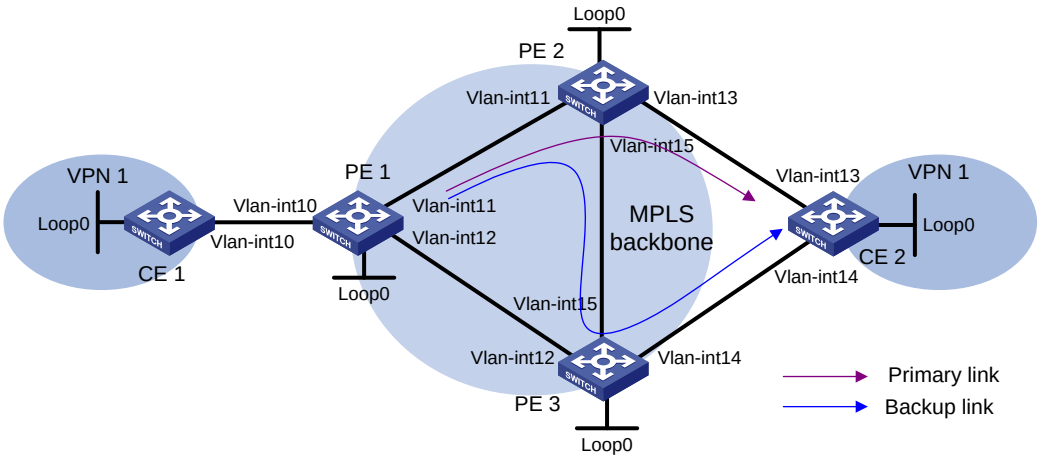


Table 25 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
CE 1	Loop0	5.5.5.5/32	PE 2	Loop0	2.2.2.2/32
	Vlan-int10	10.2.1.1/24		Vlan-int11	172.1.1.2/24
PE 1	Loop0	1.1.1.1/32		Vlan-int13	10.1.1.2/24
	Vlan-int10	10.2.1.2/24		Vlan-int15	172.3.1.2/24
	Vlan-int11	172.1.1.1/24	PE 3	Loop0	3.3.3.3/32
	Vlan-int12	172.2.1.1/24		Vlan-int12	172.2.1.3/24
CE 2	Loop0	4.4.4.4/32		Vlan-int14	10.3.1.2/24
	Vlan-int13	10.1.1.1/24		Vlan-int15	172.3.1.3/24
	Vlan-int14	10.3.1.1/24			

Procedure

1. Configure IP addresses and masks for interfaces as shown in Figure 39Table 25, and configure BGP and MPLS L3VPN. (Details not shown.)

For more information about configuring basic MPLS L3VPN, see "Example: Configuring basic MPLS L3VPN."

2. Configure MPLS L3VPN FRR on PE 2:

Configure the source IP address of BFD echo packets as 12.1.1.1.

```
<PE2> system-view
[PE2] bfd echo-source-ip 12.1.1.1
```

Create routing policy **frr**, and specify the backup next hop as 3.3.3.3 for the route to 4.4.4.4/32.

```
[PE2] ip prefix-list abc index 10 permit 4.4.4.4 32
[PE2] route-policy frr permit node 10
[PE2-route-policy-frr-10] if-match ip address prefix-list abc
[PE2-route-policy-frr-10] apply fast-reroute backup-nexthop 3.3.3.3
[PE2-route-policy-frr-10] quit
```

Use echo-mode BFD to detect the primary route connectivity.

```
[PE2] bgp 100
[PE2-bgp-default] primary-path-detect bfd echo
```

Configure FRR for VPN instance **vpn1** to use routing policy **frr**.

```
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] fast-reroute route-policy frr
```

Specify the preferred value as 200 for BGP routes received from CE 2. This value is greater than the preferred value (0) for routes from PE 3, so PE 2 prefers the routes from CE 2.

```
[PE2-bgp-default-ipv4-vpn1] peer 10.1.1.1 preferred-value 200
[PE2-bgp-default-vpn1] quit
[PE2-bgp-default] quit
```

Verifying the configuration

Display detailed information about the route to 4.4.4.4/32 on PE 2. The output shows the backup next hop for the route.

```
[PE2] display ip routing-table vpn-instance vpn1 4.4.4.4 32 verbose

Summary Count : 1

Destination: 4.4.4.4/32
   Protocol: BGP             Process ID: 0
  SubProtID: 0x2                    Age: 01h54m24s
       Cost: 0              Preference: 10
      IpPre: N/A            QosLocalID: N/A
        Tag: 0                   State: Active Adv
  OrigTblID: 0x0               OrigVrf: vpn1
    TableID: 0x102              OrigAs: 300
      NibID: 0x15000002         LastAs: 300
     AttrID: 0x0              Neighbor: 10.1.1.1
      Flags: 0x10060       OrigNextHop: 10.1.1.1
      Label: NULL          RealNextHop: 10.1.1.1
    BkLabel: 1275            BkNextHop: 172.3.1.3
  Tunnel ID: Invalid          Interface: Vlan-int13
BkTunnel ID: 0x409          BkInterface: Vlan-int15
   FtnIndex: 0x0          TrafficIndex: N/A
```

Connector: N/A

Contents

Configuring MPLS L2VPN 1
About MPLS L2VPN 1
Basic concepts of MPLS L2VPN 1
MPLS L2VPN network models 2
Remote connection establishment 3
Local connection establishment 4
TDM-CEM 4
PW data encapsulation types 5
Control word 7
MPLS L2VPN interworking 7
Bypass PW 8
MPLS L2VPN diagnostic troubleshooting 9
Restrictions and guidelines: MPLS L2VPN configuration 9
MPLS L2VPN tasks at a glance 9
Configuring a remote connection 9
Configuring a local connection 10
Prerequisites for MPLS L2VPN 10
Enabling L2VPN 10
Configuring an Ethernet service instance on an interface 11
Configuring a cross-connect 11
Configuring a PW 12
Configuring a PW class 12
Configuring a static PW 12
Configuring an LDP PW 13
Configuring a BGP PW 13
Configuring a remote CCC connection 15
Binding an AC to a cross-connect 16
About binding an AC to a cross-connect 16
Restrictions and guidelines for binding an AC to a cross-connect 16
Binding an Ethernet service instance to a non-BGP cross-connect 17
Binding an Ethernet service instance to a BGP cross-connect 17
Enabling SNMP notifications for L2VPN PW 17
Display and maintenance commands for MPLS L2VPN 18

Configuring MPLS L2VPN

MPLS L2VPN provides point-to-point and point-to-multipoint connections. This chapter describes only the MPLS L2VPN technologies that provide point-to-point connections. For information about the MPLS L2VPN technologies that provide point-to-multipoint connections, see "Configuring VPLS."

About MPLS L2VPN

MPLS L2VPN is an implementation of Pseudo Wire Emulation Edge-to-Edge (PWE3). It offers Layer 2 VPN services over an MPLS or IP backbone. MPLS L2VPN can transparently transmit Layer 2 data for different data link layer protocols such as Ethernet and ATM.

Basic concepts of MPLS L2VPN

CE

A customer edge (CE) is a customer device directly connected to the service provider network.

PE

A provider edge (PE) is a service provider device connected to one or more CEs. It provides VPN access by mapping and forwarding packets between user networks and public tunnels.

AC

An attachment circuit (AC) is a link between a CE and a PE, such as an FR DLCI, ATM VPI/VCI, Ethernet interface, VLAN, or PPP connection.

PW

A Pseudowire (PW) is a virtual bidirectional connection between two PEs. An MPLS PW comprises a pair of LSPs in opposite directions.

Public tunnel

A public tunnel is a connection that carries one or more PWs across the MPLS or IP backbone. It can be an LSP tunnel, a GRE tunnel,.

Cross-connect

A cross-connect connects two physical or virtual circuits such as ACs and PWs. It switches packets between the two physical or virtual circuits. Cross-connects include AC to AC cross-connect, AC to PW cross-connect, and PW to PW cross-connect.

Site ID

A site ID uniquely identifies a site in a VPN. Sites in different VPNs can have the same site ID.

RD

A route distinguisher (RD) is added before a site ID to distinguish the sites that have the same site ID but reside in different VPNs. An RD and a site ID uniquely identify a VPN site.

Label block

A label block is a set of labels. It includes the following parameters:

- **Label base**—The LB specifies the initial label value of the label block. A PE automatically selects an LB value that cannot be manually modified.

- **Label range**—The LR specifies the number of labels that the label block contains. The LB and LR determine the labels contained in the label block. For example, if the LB is 1000 and the LR is 5, the label block contains labels 1000 through 1004.
- **Label-block offset**—The LO specifies the offset of a label block. If the existing label block becomes insufficient as the VPN sites increase, you can add a new label block to enlarge the label range. A PE uses an LO to identify the position of the new label block. The LO value of a label block is the sum of the LRs of all previously assigned label blocks. For example, if the LR and LO of the first label block are 10 and 0, the LO of the second label block is 10. If the LR of the second label block is 20, the LO of the third label block is 30.

A label block with LB, LO, and LR as 1000, 10, and 5, respectively, is represented as 1000/10/5.

For example, a VPN has 10 sites, and a PE assigns the first label block LB1/0/10 to the VPN. When another 15 sites are added, the PE keeps the first label block and assigns the second label block LB2/10/15 to extend the network. LB1 and LB2 are the initial label values that are randomly selected by the PE.

Route target

PEs use the BGP route target attribute (also called VPN target attribute) to manage BGP L2VPN information advertisement. PEs support the following types of route target attributes:

- **Export target attribute**—When a PE sends L2VPN information to the peer PE in a BGP update message, it sets the route target attribute in the update message to an export target. L2VPN information includes the site ID, RD, and label block.
- **Import target attribute**—When a PE receives an update message from the peer PE, it checks the route target attribute in the update message. If the route target value matches an import target, the PE accepts the L2VPN information in the update message.

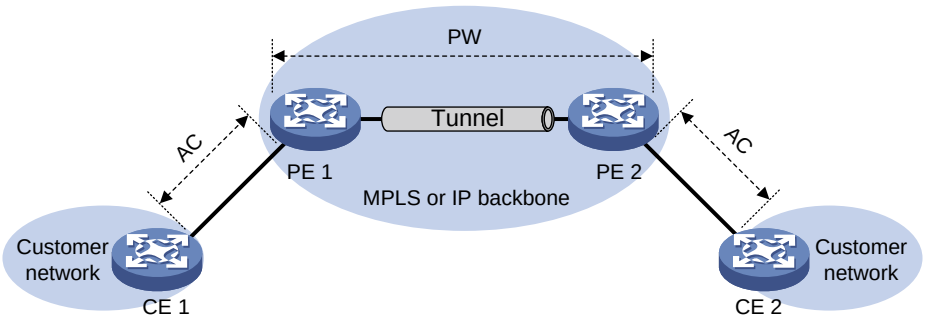
Route target attributes determine from which PEs a PE can receive L2VPN information.

MPLS L2VPN network models

MPLS L2VPN network models include the remote connection and local connection models.

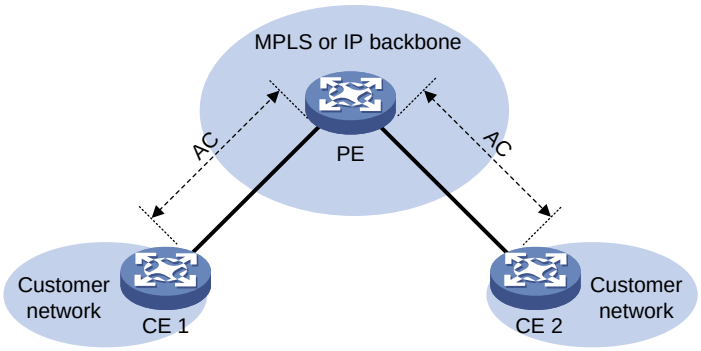
As shown in Figure 1, the remote connection model connects two CEs through a PW on an MPLS or IP backbone.

Figure 1 Remote connection model



As shown in Figure 2, the local connection model connects two CEs to the same PE so the CEs can communicate through the PE.

Figure 2 Local connection model



Remote connection establishment

To set up a remote MPLS L2VPN connection:

1. Set up a public tunnel to carry one or more PWs between PEs.
2. Set up a PW to connect customer networks.
3. Set up an AC between a PE and a CE.
4. Bind the AC to the PW.

After the PE receives packets from the AC, it adds the PW label into the packets and sends the packets to the peer PE through the public tunnel.

After the peer PE receives the packets from the public tunnel, it removes the PW label of the packets and forwards the packets to the AC bound to the PW.

Setting up a public tunnel

The public tunnel can be an LSP or GRE tunnel.

If multiple public tunnels exist between two PEs, you can configure a tunnel policy to control tunnel selection. For more information about tunnel policies, see "Configuring tunnel policies."

If a PW is established over an LSP tunnel, packets on the PW have two labels. The outer label is the public LSP tunnel label that MPLS uses to forward the packet to the peer PE. The inner label is the PW label that the peer PE uses to forward the packet to the destination CE.

Setting up a PW

PWs include static PWs, LDP PWs, BGP PWs, and Circuit Cross Connect (CCC) PWs.

- Static PW establishment

  To establish a static PW, configure the peer PE address, and the incoming and outgoing PW labels for the PW on the two PEs. Static PWs consume a small amount of resources but have complex configurations.

- LDP PW establishment

  To establish an LDP PW, configure LDP and specify the peer PE address on the two PEs. LDP defines a new FEC type named PW ID FEC for PEs to exchange PW-label bindings. The new FEC type uses a PW ID and a PW data encapsulation type to identify a PW. The PW ID is the ID of the PW between PEs. The PW data encapsulation type specifies the encapsulation type for data transmitted over the PW, such as ATM, FR, Ethernet, or VLAN. PEs advertise the PW label and PW ID FEC in label mapping messages to create a PW. Dynamic PWs have simple configurations but consume more resources than static PWs.

- BGP PW establishment

  To establish BGP PWs, BGP advertises label block information in an extended BGP update to PEs in the same VPN. Each PE uses the received label block information to calculate outgoing

labels and uses its own label block to calculate incoming labels. After two PEs complete label calculation, a BGP PW is established between them.

BGP PWs have the following features:

- **Simplified configuration**—There is no need to manually specify peer PEs. A PE automatically find peer PEs after receiving label block information from the peer PEs.
- **Reduced workload**—Label block advertisement enables assigning labels for multiple PWs at one time.

- CCC PW establishment

  To establish a CCC PW, manually specify the incoming and outgoing labels for the PW on the PEs, and create two static LSPs in opposite directions on P devices between PEs. There is no need to configure a public tunnel for the CCC PW. CCC employs only one level of label to transfer packets. The static LSPs on the P devices transfer data only for the CCC PW. They cannot be used by other connections or MPLS L3VPN.

Setting up an AC

Set up an AC by configuring a link layer connection between a PE and a CE.

An AC is an Ethernet service instance on a Layer 2 Ethernet interface or Layer 2 aggregate interface. It forwards packets that are received on the interface and meet the match criteria of the Ethernet service instance to the bound PW. If the match criterion is VLAN ID, the VLAN is unique on a per interface basis rather than on a global basis.

NOTE:

When VLANs are globally unique, packets with the same VLAN ID are forwarded over the PW bound with that VLAN ID regardless of the receiving interfaces. If VLANs are unique on a per interface basis, packets with the same VLAN ID from different interfaces can be forwarded over different PWs.

Binding the AC to the PW

Bind the Layer 3 physical interface, Layer 3 subinterface, or Ethernet service instance to the PW, so the PE forwards packets between the AC and the PW.

Local connection establishment

To set up a local MPLS L2VPN connection between two CEs:

1. Set up ACs:

   Configure the link layer protocol to set up an AC between the PE and each CE. For more information, see "Setting up an AC."

2. Bind the two ACs:

   Bind the PE's interfaces connected to the two CEs so the PE can forward packets between CEs.

TDM-CEM

Time Division Multiplexing Circuit Emulation (TDM-CEM) encapsulates TDM bit-streams as PWs over a Packet Switched Network (PSN).

PWE3 TDM-CEM has the following modes:

- **Structure Agnostic TDM over Packet (SAToP)**—Considers all TDM data as pure bit streams, and fragments and encapsulates the bit streams into packets. SAToP does not interpret TDM data or participate in the TDM signaling.
- **Circuit Emulation Service over PSN (CESoPSN)**—Identifies and processes the TDM frame format and signaling, and encapsulates the structured TDM data as packets.

TDM has the stringent clock requirements to deliver real-time synchronization services. PSN does not have the stringent clock requirements. To transport TDM services through PSN, use one of the following clock recovery modes on an egress PE to implement clock recovery:

- **Adaptive Clock Recovery (ACR)**—Based on packet arrival rate and jitter buffer fill level. ACR is used when the ingress and egress PEs do not have a common clock source.
- **Differential Clock Recovery (DCR)**—Based on time stamps received in RTP header. DCR is used when the ingress and egress PEs have a common clock source.

PW data encapsulation types

MPLS L2VPN transports Layer 2 data of different data link layer protocols through PWs. A PE encapsulates a Layer 2 packet received from an AC according to the PW data encapsulation type.

Relationship between AC types and PW data encapsulation types

The PW data encapsulation type is determined by the link type of the AC.

Table 1 Relationship between AC types and PW data encapsulation types

AC type	PW data encapsulation type
Ethernet	Ethernet
	VLAN

Ethernet over MPLS

Ethernet over MPLS uses MPLS L2VPN to connect Ethernets, and delivers Ethernet packets through a PW over the MPLS backbone.

The following PW data encapsulation types are available for Ethernet over MPLS:

- **Ethernet**—P-tag is not transferred on a PW.
  - For a packet from a CE:
    - If the packet contains a P-tag, the PE removes the P-tag, and adds a PW label and an outer tag into the packet before forwarding it.
    - If the packet contains no P-tag, the PE directly adds a PW label and an outer tag into the packet before forwarding it.
  - For a packet to a CE:
    - If the access mode is configured as VLAN by using the **ac interface** command, the PE adds a P-tag into the packet before sending it to the CE.
    - If the access mode is configured as Ethernet by using the **ac interface** command, the PE directly sends the packet to the CE.

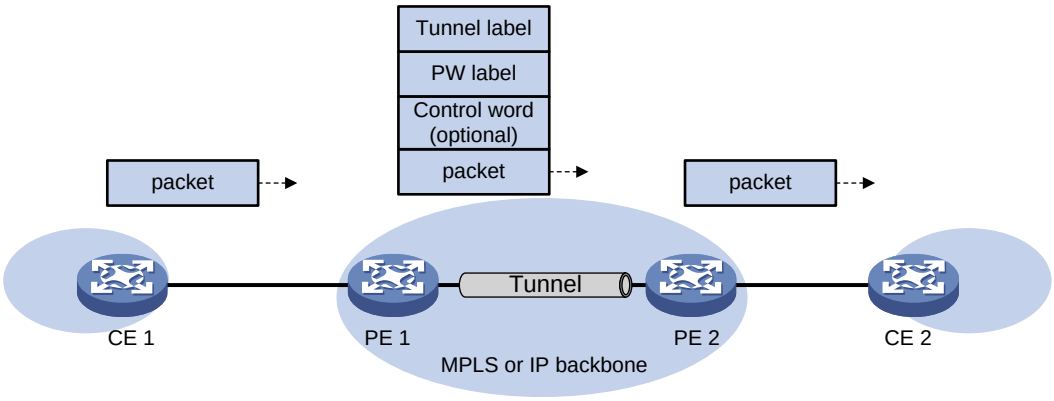
    You cannot rewrite or remove existing tags.
- **VLAN**—Packets transmitted over a PW must carry a P-tag.
  - For a packet from a CE:
    - If the peer PE does not require the ingress to rewrite the P-tag, the PE keeps the P-tag unchanged for the packet, and then encapsulates the packet. If the packet contains no P-tag, the PE adds a null label (the label value is 0) into the packet, and then encapsulates the packet.
    - If the peer PE requires the ingress to rewrite the P-tag, the PE changes the P-tag to the expected VLAN tag (the tag value might be 0), and then adds a PW label and an outer tag into the packet. If the packet contains no P-tag, the PE adds a VLAN tag expected by the peer PE (the tag value might be 0), and then adds a PW label and an outer tag into the packet.
  - For a packet to a CE:

- If the access mode is configured as VLAN by using the **ac interface** command, the PE rewrites or retains the P-tag before forwarding the packet.
- If the access mode is configured as Ethernet by using the **ac interface** command, the PE removes the P-tag before forwarding the packet.

Ethernet over MPLS supports the following modes:

- **Port mode**—A Layer 3 Ethernet interface is bound to a PW. Packets received from the Layer 3 Ethernet interface are forwarded through the bound PW. The default PW data encapsulation type for port mode is Ethernet.

  Figure 3 Packet encapsulation in port mode



- **VLAN mode**—A Layer 3 Ethernet subinterface is bound to a PW. Packets received from the VLAN are forwarded through the bound PW. The peer PE can modify the VLAN tag as needed. The default PW data encapsulation type for VLAN mode is VLAN.
- **Flexible mode**—An Ethernet service instance on a Layer 2 Ethernet interface or Layer 2 aggregate interface is bound to a PW. Packets that are received from the interface and meet the match criteria of the Ethernet service instance are forwarded through the bound PW. You can configure flexible match criteria for the Ethernet service instance. For example, configure the Ethernet service instance to match all packets, tagged packets, or untagged packets. The default PW data encapsulation type for flexible mode is VLAN. Flexible mode can also implement the port and VLAN modes through match criteria configuration.

PPP/HDLC over MPLS

PPP/HDLC over MPLS uses MPLS L2VPN to connect PPPs or HDLC networks, and delivers PPP or HDLC packets through a PW over the MPLS backbone.

If the link type of the AC is PPP, the PW data encapsulation type is PPP. If the link type of the AC is HDLC, the PW data encapsulation type is HDLC.

PPP/HDLC over MPLS supports only the port mode. You can associate a Layer 3 interface whose encapsulation type is PPP or HDLC with a PW.

In a PPP/HDLC over MPLS network, a PE processes a PPP or HDLC packet as follows:

1. After receiving a packet from a Layer 3 interface, the PE searches for the PW bound to the interface.

2. The PE encapsulates the packet and sends the packet to the peer PE through the PW.

3. The peer PE removes the outer encapsulation to get the original PPP or HDLC packet, and then forwards the packet to the user network.

TDM-CEM

TDM-CEM uses MPLS L2VPN to deliver TDM packets through a PW over the MPLS backbone. If the link type of the AC is E1 or T1, the following PW data encapsulation types are available:

- **SAToP**—Provides unstructured, transparent TDM transport of E1 or T1 frames through a PW over the MPLS backbone. All time slots on an E1 or T1 interface are forwarded through the same PW. SAToP includes structure-agnostic E1 over packet and structure-agnostic T1 (DS1) over packet.

  After you create an unstructured CEM interface and bind it to a PW, all bit streams received from the interface are forwarded through the PW.
- **CESoPSN basic mode**—Provides structured TDM transport of E1 or T1 frames through a PW.

  After you create a structured CEM interface and bind it to a PW, the bit streams of the specified time slots received from the interface are forwarded through the PW.

Control word

The control word field is between the MPLS label stack and the Layer 2 data. It carries control information for the Layer 2 frame, for example, the sequence number.

The control word feature has the following functions:

- Avoids fragment disorder. In multipath forwarding, fragments received might be disordered. The control word feature reorders the fragments according to the sequence number carried in the control word field.
- Transfers specific Layer 2 frame flags, such as the FECN bit and BECN bit of Frame Relay.
- Identifies the original payload length for packets that include padding.

When the PW data encapsulation type is FR DLCI or ATM AAL5 SDU VCC, packets on the PW always carry the control word field, and the control word feature cannot be disabled.

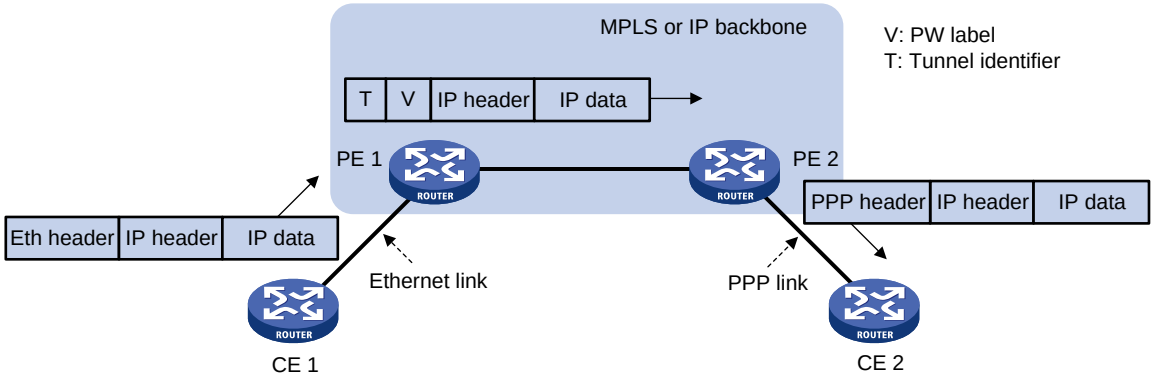
When the PW data encapsulation type is Ethernet or VLAN, the control word field is optional. You can configure whether to carry the control word field in packets sent on the PW. If you enable the control word feature on both PEs, packets transmitted on the PW carry the control word field. Otherwise, the packets do not carry the control word field.

MPLS L2VPN interworking

CEs might connect to PEs through various types of links, such as ATM, FR, HDLC, Ethernet, and PPP. MPLS L2VPN interworking connects such CEs and allow them to communicate.

MPLS L2VPN supports Ethernet interworking and IP interworking modes. The device only supports IP interworking. Only local MPLS L2VPN connections, static PWs, LDP PWs, and remote CCC connections support the interworking feature.

Figure 4 Ethernet to PPP interworking



As shown in Figure 4, a packet in an MPLS L2VPN interworking scenario is forwarded as follows:

2. CE 1 sends an Ethernet frame destined for CE 2 to PE 1.

3. PE 1 checks whether the packet encapsulated in the received Ethernet frame is an IP packet.
   - If yes, PE 1 removes the Ethernet header, adds PW label V and tunnel label T to the IP packet, and forwards the packet to PE 2 through the tunnel.
   - If not, PE 1 drops the frame.

4. PE 2 obtains the output interface according to the PW label V in the received packet, removes the PW label, adds an PPP header, and forwards the PPP frame through the output interface to CE 2.

In an MPLS L2VPN interworking scenario, link layer negotiation packets cannot be delivered on the backbone network. Therefore, Layer 2 connections cannot be established between CEs. CEs must establish Layer 2 connections with the PEs. For example, CE 2 and PE 2 must perform PPP negotiation to establish a PPP connection.

Bypass PW

To resolve the problem, you can create a single bypass PW or double bypass PWs between PE 2 and PE 3.

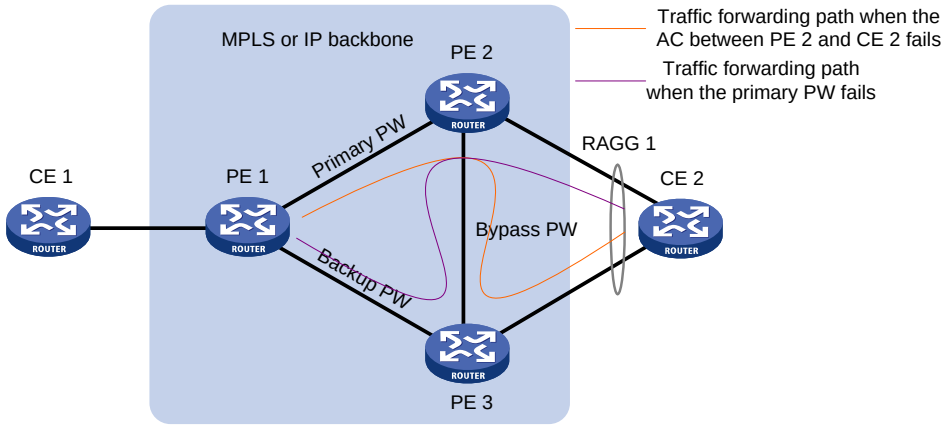
Single bypass PW

As shown in Figure 5, after you create a single bypass PW between PE 2 and PE 3, the bypass PW provides both PW bypass and AC bypass. Traffic is forwarded as follows:

- When the AC between PE 2 and CE 2 fails, traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—PE 3—CE 2.
- When the primary PW fails, traffic from CE 2 to CE 1 goes through the path CE 2—PE 2—PE 3—PE 1—CE 1.

In a single bypass PW network, traffic is switched to the bypass PW instead of the backup PW when the AC or primary PW fails.

Figure 5 Single bypass PW



Double bypass PWs

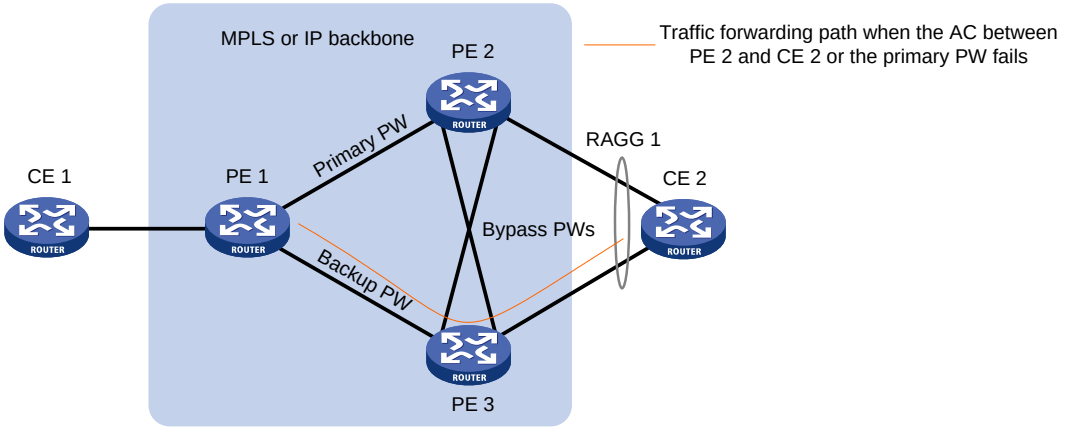
Double bypass PWs refer to an AC-bypass PW used to forward traffic when the AC fails and a PW-bypass PW used to forward traffic when the primary PW fails. Compared with single bypass PWs, double bypass PWs reduce the traffic switchover time and provide higher reliability.

As shown in Figure 6, after you create double bypass PWs between PE 2 and PE 3, traffic is forwarded as follows:

- When the AC between PE 2 and CE 2 fails, traffic from CE 1 to CE 2 goes through the path CE 1—PE 1—PE 2—PE 3—CE 2. The traffic is forwarded from PE 2 to PE 3 through the AC-bypass PW. After the primary/backup PW switchover completes, the traffic is switched to the backup PW, so the final traffic forwarding path is CE 1—PE 1—PE 3—CE 2.

- When the primary PW fails, traffic from CE 2 to CE 1 goes through the path CE 2—PE 2—PE 3—PE 1—CE 1. The traffic is forwarded from PE 2 to PE 3 through the PW-bypass PW. . After the primary/backup PW switchover completes, the traffic is switched to the backup PW, so the final traffic forwarding path is CE 2—PE 3—PE 1—CE 1.

Figure 6 Double bypass PWs



MPLS L2VPN diagnostic troubleshooting

This feature identifies whether the MPLS L2VPN module functions correctly, for exampel, whether PWs have been established successfully and whether PW flapping has occurred to help you locate MPLS L2VPN exceptions. For more information about MPLS L2VPN diagnostic troubleshooting, see diagnostic troubleshooting configuration in *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines: MPLS L2VPN configuration

For this feature to take effect, use the **system-working-mode** command to set the system operating mode to advanced or expert and restart the device before you configure this feature. For more information about the **system-working-mode** command, see "Managing the device" in *Fundamentals Configuration Guide*.

MPLS L2VPN tasks at a glance

Configuring a remote connection

1. Enabling L2VPN
2. Configuring an AC

   Configuring an Ethernet service instance on an interface
3. Configuring a cross-connect
4. Configuring a PW

   Configure a static PW, LDP PW, BGP PW, and remote CCC connection as needed.
   - (Optional.) Configuring a PW class
   - Configuring a static PW

- Configuring an LDP PW
- Configuring a BGP PW
- Configuring a remote CCC connection

5. Binding an AC to a cross-connect
6. (Optional.) Maintaining an MPLS L2VPN network

   Enabling SNMP notifications for L2VPN PW

Configuring a local connection

1. Enabling L2VPN
2. Configuring an AC

   To create a local connection, you must configure two ACs.

   Configuring an Ethernet service instance on an interface
3. Configuring a cross-connect
4. Binding an AC to a cross-connect

   Perform this task to bind the two ACs to the same cross-connect.

Prerequisites for MPLS L2VPN

To establish an MPLS L2VPN, you must perform the following tasks:

1. Configure an IGP to achieve IP connectivity within the backbone.
2. Configure basic MPLS, or LDP to set up public tunnels across the backbone.

Enabling L2VPN

Prerequisites

Before you enable L2VPN, perform the following tasks:

- Configure an LSR ID for the PE with the **mpls lsr-id** command.
- Enable MPLS with the **mpls enable** command on the core-facing interface of the PE.

For more information about the **mpls lsr-id** and **mpls enable** commands, see *MPLS Command Reference*.

Procedure

1. Enter system view.

   system-view
2. Enable L2VPN.

   l2vpn enable

   By default, L2VPN is disabled.

Configuring an Ethernet service instance on an interface

About this task

When the PE is connected to a CE through a Layer 2 Ethernet interface or Layer 2 aggregate interface, you can configure an Ethernet service instance on the interface to match packets for the AC.

Restrictions and guidelines

For information about configuring the match criterion of an Ethernet service instance by using the **encapsulation** command, see MPLS L2VPN commands in *MPLS Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Create an Ethernet service instance and enter Ethernet service instance view.
   service-instance *instance-id*
4. Configure a packet match criterion for the Ethernet service instance.
   - Match packets with the specified inner VLAN IDs.
     encapsulation c-vid { *vlan-id* | *vlan-id-list* }
   - Match packets with the specified outer VLAN IDs.
     encapsulation s-vid { *vlan-id* | *vlan-id-list* } [**only-tagged**]
   - Match packets with the specified outer VLAN IDs and the specified inner VLAN IDs.
     encapsulation s-vid *vlan-id* **c-vid** { *vlan-id-list* | **all** }
   - Match packets that do not have a VLAN tag.
     encapsulation untagged

   By default, no packet match criterion is configured.

Configuring a cross-connect

1. Enter system view.
   system-view
2. Create a cross-connect group and enter cross-connect group view.
   xconnect-group *group-name*
3. (Optional.) Configure a description for the cross-connect group.
   description *text*

   By default, no description is configured for a cross-connect group.
4. (Optional.) Enable the cross-connect group.
   undo shutdown

By default, the cross-connect group is enabled.

5. Create a cross-connect and enter cross-connect view.

   connection *connection-name*

6. (Optional.) Set an MTU for the PW.

   mtu *size*

   The default MTU is 1500 bytes.

   The two PEs on an LDP PW must have the same MTU configured for the PW. Otherwise, the PW cannot come up.

Configuring a PW

Configuring a PW class

About this task

You can configure PW attributes such as the PW data encapsulation type and enable control word in a PW class. PWs with the same attributes can use the same PW class.

Procedure

1. Enter system view.

   system-view

2. Create a PW class and enter PW class view.

   pw-class *class-name*

   By default, no PW classes exist.

3. Enable control word.

   control-word enable

   By default, control word is disabled.

4. Specify the PW data encapsulation type.

   pw-type { **ethernet** | **vlan** }

   By default, the PW data encapsulation type is VLAN.

Configuring a static PW

1. Enter system view.

   system-view

2. Enter cross-connect group view.

   xconnect-group *group-name*

3. Enter cross-connect view.

   connection *connection-name*

4. Configure a static PW, and enter cross-connect PW view.

   peer *ip-address* **pw-id** *pw-id* **in-label** *label-value* **out-label** *label-value* [**pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *

Configuring an LDP PW

About this task

After an LDP PW is created, the PE automatically sends a targeted hello to create an LDP session to the peer PE. Then, the PE exchanges the PW ID FEC and PW label mapping with the peer.

Prerequisites

Before you configure an LDP PW, enable global and interface MPLS LDP on the PE. For information about MPLS LDP configuration, see "Configuring LDP."

Procedure

1. Enter system view.
   system-view
2. Enter cross-connect group view.
   xconnect-group *group-name*
3. Enter cross-connect view.
   connection *connection-name*
4. Configure an LDP PW, and enter cross-connect PW view.
   peer *ip-address* **pw-id** *pw-id* [**pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *

Configuring a BGP PW

Configuring BGP to advertise MPLS L2VPN label block information

1. Enter system view.
   system-view
2. Enable BGP instance and enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
   By default, BGP is disabled.
3. Configure the remote PE as a BGP peer.
   peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*
   For more information about this command, see *Layer 3—IP Routing Command Reference*.
4. Create the BGP L2VPN address family and enter BGP L2VPN address family view.
   address-family l2vpn
5. Enable BGP to exchange BGP L2VPN information with the specified peer or peer group.
   peer { *group-name* | *ip-address* [*mask-length*] } **enable**
   By default, BGP cannot exchange BGP L2VPN information with any peer or peer group.
   For more information about this command, see *Layer 3—IP Routing Command Reference*.
6. Enable BGP to exchange label block information with the specified peer or peer group.
   peer { *group-name* | *ip-address* [*mask-length*] } **signaling** [**non-standard**]
   By default, BGP can exchange label block information with a BGP L2VPN peer or peer group by using RFC 4761 MP_REACH_NLRI.
7. Configure the BGP L2VPN address family.
   For more information, see "Configuring BGP L2VPN address family."
8. Reset BGP L2VPN sessions.

For more information, see "Resetting BGP sessions of L2VPN address family."

Creating a BGP PW

1. Enter system view.

   system-view

2. Enter cross-connect group view.

   xconnect-group *group-name*

3. Configure the cross-connect group to automatically discover neighbors and create PWs through BGP and enter auto-discovery cross-connect group view.

   auto-discovery bgp

   By default, a cross-connect group does not automatically discover neighbors or create PWs through BGP.

4. Configure an RD for the cross-connect group.

   route-distinguisher *route-distinguisher*

   By default, no RD is configured for the cross-connect group.

5. Configure route targets for the cross-connect group.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, no route targets are configured for the cross-connect group.

6. (Optional.) Specify a PW class for the auto-discovery cross-connect group.

   pw-class *class-name*

   By default, no PW class is specified.

7. (Optional.) Set an MTU for the PW.

   mtu *size*

   The default MTU is 1500 bytes.

8. Create a local site and enter site view.

   site *site-id* [**range** *range-value*] [**default-offset** *default-offset*]

9. Create a cross-connect and enter auto-discovery cross-connect view.

   connection remote-site-id *remote-site-id*

   After you execute this command, a PW to the specified remote site is created and is bound to the cross-connect.

10. (Optional.) Specify a tunnel policy for the auto-discovery cross-connect.

    tunnel-policy *tunnel-policy-name*

    By default, no tunnel policy is specified.

Configuring BGP L2VPN address family

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP L2VPN address family view.

   address-family l2vpn

4. Permit the local AS number to appear in routes from the specified peer or peer group and specify the appearance times.

   peer { *group-name* | *ip-address* [*mask-length*] } **allow-as-loop** [*number*]

By default, the local AS number is not allowed in routes from a peer or peer group.

For more information about this command, see *Layer 3—IP Routing Command Reference*.

5. Enable route target-based filtering of incoming BGP L2VPN information.

   policy vpn-target

   By default, route target-based filtering of incoming BGP L2VPN information is enabled.

6. Configure BGP route reflection:

   a. Configure the router as a route reflector and specify a peer or peer group as its client.

      peer { *group-name* | *ip-address* [*mask-length*] } **reflect-client**

      By default, no route reflector or client is configured.

   b. Enable L2VPN information reflection between clients.

      reflect between-clients

      By default, L2VPN information reflection between clients is enabled.

   c. Configure the cluster ID of the route reflector.

      reflector cluster-id { *cluster-id* | *ip-address* }

      By default, a route reflector uses its own router ID as the cluster ID.

   d. Configure a filtering policy for reflected L2VPN information.

      rr-filter *ext-comm-list-number*

      By default, the route reflector does not filter reflected L2VPN information.

   For more information about the commands, see *Layer 3—IP Routing Command Reference*.

7. Configure optimal route selection delay.

   - Set the optimal route selection delay timer for an address family.

     route-select delay *delay-value*

   - Enable optimal route selection delay based on DOWN-to-UP peer state changes and set a delay timer.

     route-select suppress on-peer-up *milliseconds*

   By default, optimal route selection is not delayed.

   For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Resetting BGP sessions of L2VPN address family

To reset BGP sessions of the L2VPN address family, execute one of the following command in user view:

- Perform manual soft-reset for BGP sessions of the L2VPN address family.

  refresh bgp [**instance** *instance-name*] { *ip-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **l2vpn**

- Reset BGP sessions of the L2VPN address family.

  reset bgp [**instance** *instance-name*] { *as-number* | *ip-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **l2vpn**

For more information about the commands, see *Layer 3—IP Routing Command Reference*.

Configuring a remote CCC connection

Restrictions and guidelines

The outgoing label specified on a device must be the same as the incoming label specified on the next-hop device.

CCC connection settings such as the encapsulation type and control word feature must be consistent on the two PEs. Otherwise, the PEs might fail to forward packets over the CCC connection.

Procedure

1. Configure the PE devices of the CCC connection:
   a. Enter system view.
      system-view
   b. Enter cross-connect group view.
      xconnect-group *group-name*
   c. Enter cross-connect view.
      connection *connection-name*
   d. Create a remote CCC connection.
      ccc in-label *in-label-value* **out-label** *out-label-value* { **nexthop** *nexthop* | **out-interface** *interface-type interface-number* } [**pw-class** *class-name*]
      Use the **out-interface** keyword to specify the outgoing interface only on a point-to-point link. On other types of interfaces such as Layer 3 Ethernet interfaces and VLAN interfaces, you must use the **nexthop** keyword to specify the IP address of the next hop.
2. Configure P devices of the CCC connection:
   a. Enter system view.
      system-view
   b. Configure a static LSP for each direction of the CCC connection.
      static-lsp transit *lsp-name* **in-label** *in-label* { **nexthop** *next-hop-ip-address* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label*
      For more information about this command, see *MPLS Command Reference*.

Binding an AC to a cross-connect

About binding an AC to a cross-connect

On a Layer 2 Ethernet or Layer 2 aggregate interface, you can create an Ethernet service instance and bind it to a cross-connect. The Ethernet service instance matches packets received on that interface. The matching packets are then forwarded to the bound PW or another AC. An Ethernet service instance can match all packets, tagged packets, or untagged packets.

When you bind an AC to a cross-connect, you can associate Track with the AC. Then, the AC is up only when one or more of the associated track entries are positive.

Restrictions and guidelines for binding an AC to a cross-connect

If a Layer 2 Ethernet interface has been added to a link aggregation group, you cannot bind an Ethernet service instance on the interface to a cross-connect, and vice versa.

Binding an Ethernet service instance to a non-BGP cross-connect

1. Enter system view.
   system-view
2. Enter cross-connect group view.
   xconnect-group *group-name*
3. Enter cross-connect view.
   connection *connection-name*
4. Bind the Ethernet service instance on the interface to the cross-connect and enter cross-connect AC view.
   ac interface *interface-type interface-number*[**track** *track-entry-number*&<1-3>]
   By default, no Ethernet service instance is bound to the cross-connect.

Binding an Ethernet service instance to a BGP cross-connect

1. Enter system view.
   system-view
2. Enter cross-connect group view.
   xconnect-group *group-name*
3. Enter auto-discovery cross-connect group view.
   auto-discovery bgp
4. Enter site view.
   site *site-id* [**range** *range-value*] [**default-offset** *default-offset-value*]
5. Enter auto-discovery cross-connect view.
   connection remote-site-id *remote-site-id*
6. Bind the Ethernet service instance on the interface to the BGP cross-connect.
   ac interface *interface-type interface-number*[**track** *track-entry-number*&<1-3>]
   By default, no Ethernet service instance is bound to the BGP cross-connect.

Enabling SNMP notifications for L2VPN PW

About this task

This feature enables L2VPN to generate SNMP notifications when the device does not support L2VPN flow labels, or PW deletions, CCC up and down, PW switchovers, or PW status changes occur. For L2VPN event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for L2VPN PW.

snmp-agent trap enable l2vpn [**ccc-up-down** | **flow-label-nonsupport** | **pw-delete** | **pw-switch** | **pw-up-down**] *

By default, SNMP notifications for L2VPN PW are disabled.

Display and maintenance commands for MPLS L2VPN

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display BGP L2VPN peer group information.	**display bgp** [**instance** *instance-name*] **group l2vpn** [**group-name** *group-name*]
Display L2VPN label block information discovered by BGP.	**display bgp** [**instance** *instance-name*] **l2vpn signaling** [**peer** *ip-address* { **advertised** | **received** } | **route-distinguisher** *route-distinguisher* [**site-id** *site-id* [**label-offset** *label-offset* [**advertise-info**]]]]
Display BGP L2VPN peer information.	**display bgp** [**instance** *instance-name*] **peer l2vpn** [*ip-address mask-length* | **group-name** *group-name* **log-info** | *ip-address* { **log-info** | **verbose** } | **verbose**]
Display BGP L2VPN update group information.	**display bgp** [**instance** *instance-name*] **update-group l2vpn** [*ip-address*]
Display L2VPN label block information.	**display l2vpn bgp** [**peer** *ip-address* | **local**] [**xconnect-group** *group-name*] [**verbose**]
Display cross-connect forwarding information.	**display l2vpn forwarding** { **ac** | **pw** } [**xconnect-group** *group-name*] [**slot** *slot-number*] [**verbose]**
Display LDP PW label information.	**display l2vpn ldp** [**peer** *ip-address* [**pw-id** *pw-id*] | **xconnect-group** *group-name*] [**verbose**]
Display L2VPN PW information.	**display l2vpn pw** [**xconnect-group** *group-name*] [**protocol** { **bgp** | **ldp** | **static** }] [**verbose**]
Display L2VPN PW state machine information.	**display l2vpn pw state-machine** [**vsi** *vsi-name* | **xconnect-group** *group-name*]
Display PW class information.	**display l2vpn pw-class** [*class-name*]
Display Ethernet service instance information.	**display l2vpn service-instance** [**interface** *interface-type interface-number* [**service-instance** *instance-id*]] [**verbose**]
Display cross-connect group information.	**display l2vpn xconnect-group** [**name** *group-name*] [**verbose**]
Reset BGP sessions for L2VPN.	**reset bgp** [**instance** *instance-name*] { *as-number* | *ip-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **l2vpn**

For more information about the **display bgp group l2vpn**, **display bgp peer l2vpn**, **display bgp update-group l2vpn**, and **reset bgp l2vpn** commands, see *Layer 3—IP Routing Command Reference*.

Contents

Configuring VPLS 1

About VPLS 1
Basic VPLS architecture 1
Creating a PW 2
PW types 2
MAC address learning, aging, and withdrawal 3
Traffic forwarding and flooding 4
PW full mesh and split horizon 4
PW redundancy 4
Restrictions and guidelines: VPLS configuration 5
VPLS tasks at a glance 6
Prerequisites for VPLS 6
Enabling L2VPN 6
Configuring an AC 7
Configuring an Ethernet service instance 7
Configuring a VSI 7
Configuring a PW 8
Configuring a PW class 8
Configuring a static PW 8
Configuring an LDP PW 9
Configuring a BGP PW 9
Configuring a BGP auto-discovery LDP PW 11
Configuring the BGP L2VPN address family 12
Maintaining BGP sessions 13
Binding an AC to a VSI 13
About binding an AC to a VSI 13
Restrictions and guidelines for binding an AC to a VSI 14
Binding an Ethernet service instance to a VSI 14
Configuring PW redundancy 14
Configuring static PW redundancy 14
Configuring LDP PW redundancy 15
Configuring MAC address learning 16
Enabling VPLS statistics 16
Enabling packet statistics for an AC 16
Setting the L2VPN traffic statistics collection interval 16
Enabling SNMP notifications for L2VPN PW 17
Display and maintenance commands for VPLS 17

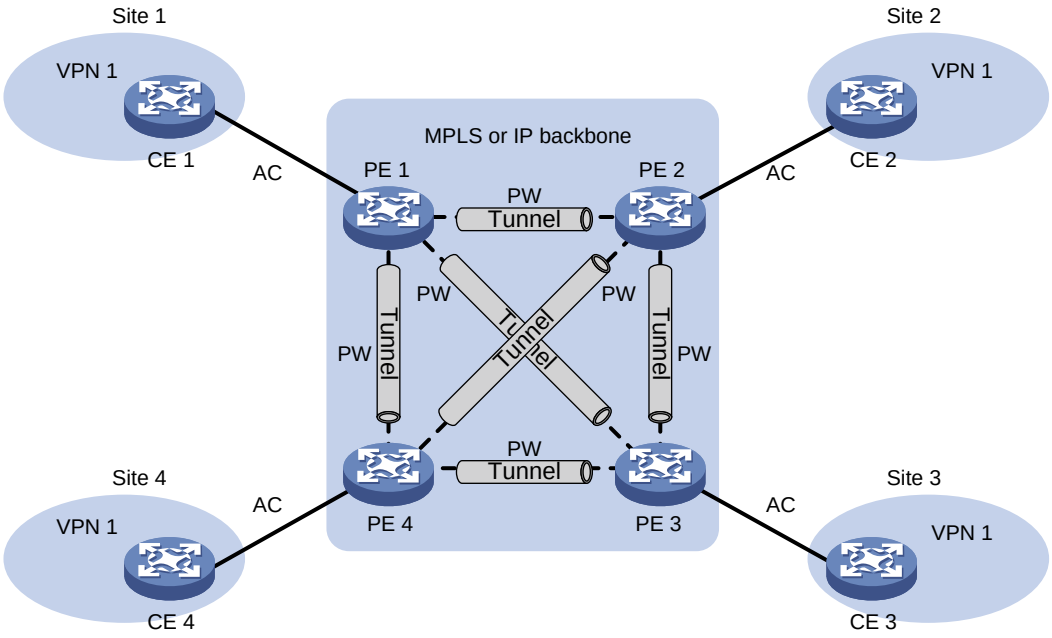
Configuring VPLS

About VPLS

Virtual Private LAN Service (VPLS) delivers a point-to-multipoint L2VPN service over an MPLS or IP backbone. The provider backbone emulates a switch to connect all geographically dispersed sites of each customer network. The backbone is transparent to the customer sites. The sites can communicate with each other as if they were on the same LAN.

Basic VPLS architecture

Figure 1 Basic VPLS architecture



As shown in Figure 1, the VPLS architecture mainly includes the following components:

- **CE**—A customer edge device is directly connected to the service provider network.
- **PE**—A provider edge device connects one or more CEs to the service provider network. A PE implements VPN access by mapping and forwarding packets between private networks and public network tunnels. A PE can be a UPE or NPE in a hierarchical VPLS.
- **AC**—An attachment circuit, physical or virtual, connects a CE and a PE, such as an Ethernet link or a VLAN.
- **PW**—A pseudowire is a bidirectional virtual connection between two PEs. An MPLS PW consists of two unidirectional MPLS LSPs in opposite directions.
- **Tunnel**—A tunnel can be an LSP tunnel, a GRE tunnel, . It carries one or more PWs over an IP/MPLS backbone. If a PW is carried on an LSP or MPLS TE tunnel, each packet on the PW contains two labels. The inner label is the PW label, which identifies the PW and ensures that the packet is forwarded to the correct VSI. The outer label is the public LSP or MPLS TE tunnel label, which ensures that the packet is correctly forwarded to the remote PE.
- **VPLS instance**—A customer network might contain multiple geographically dispersed sites (such as site 1 and site 3 in Figure 1.) The service provider uses VPLS to connect all the sites to

create a single Layer 2 VPN, which is referred to as a VPLS instance. Sites in different VPLS instances cannot communicate with each other at Layer 2.

- **VSI**—A virtual switch instance provides Layer 2 switching services for a VPLS instance on a PE. A VSI acts as a virtual switch that has all the functions of a conventional Ethernet switch, including source MAC address learning, MAC address aging, and flooding. VPLS uses VSIs to forward Layer 2 data packets in VPLS instances.

Creating a PW

In a VPLS network, PEs use PWs to forward packets among VPN sites.

A PW is created as follow:

1. Specify the address of the remote PE or configure an auto-discovery protocol to automatically find the remote PE. BGP is the most commonly used auto-discovery protocol.
2. Specify the incoming label and the outgoing label for the PW on the two PEs to create the PW. You can also use LDP or BGP to advertise the PW-label binding to the remote PE to establish the unidirectional LSPs. The PW is created once the two unidirectional LSPs are created.

PW types

PEs use PWs to forward packets among VPN sites. PWs include static PWs, LDP PWs, BGP PWs, and BGP auto-discovery LDP PWs.

Static PW

To create a static PW, specify the address of the remote PE, the incoming label, and the outgoing label.

LDP PW

To create an LDP PW, specify the address of the remote PE, and use LDP to advertise the PW-label binding to the remote PE. After the two PEs receive the PW-label binding from each other, they establish an LDP PW. The FEC type in the LDP message is PWid FEC Element that includes the PW ID field (FEC 128). The PW ID identifies the PW bound to the PW label.

BGP PW

To create a BGP PW, configure BGP to advertise label block information to the remote PE. After the two PEs receive label block information from each other, they use the label block information to calculate the incoming and outgoing labels and create the BGP PW. A PE also uses the received label block information to automatically find the remote PE.

BGP auto-discovery LDP PW

To create a BGP auto-discovery LDP PW, configure BGP to automatically find the remote PE, and use LDP to advertise the PW-label binding to the remote PE. After the two PEs receive the PW-label binding from each other, they establish a BGP auto-discovery LDP PW.

The information advertised by BGP includes the ID (for example, LSR ID) and VPLS ID of the advertising PE. The receiving PE compares the received VPLS ID with its own VPLS ID. If the two VPLS IDs are identical, the two PEs use LDP to establish a PW. If not, the PEs do not establish a PW. The FEC type in the LDP message is Generalized PWid FEC Element (FEC 129), which contains the VPLS ID, Source Attachment Individual Identifier (SAII), and Target Attachment Individual Identifier (TAII). The SAII is the LSR ID of the advertising PE. The TAII identifies the remote PE and is advertised by the remote PE. VPLS ID+SAII+TAII uniquely identifies a PW in a VPLS instance.

MAC address learning, aging, and withdrawal

Source MAC address learning

VPLS provides reachability through source MAC learning. A PE maintains a MAC address table for each VSI.

As shown in Figure 2, a PE learns source MAC addresses in the following ways:

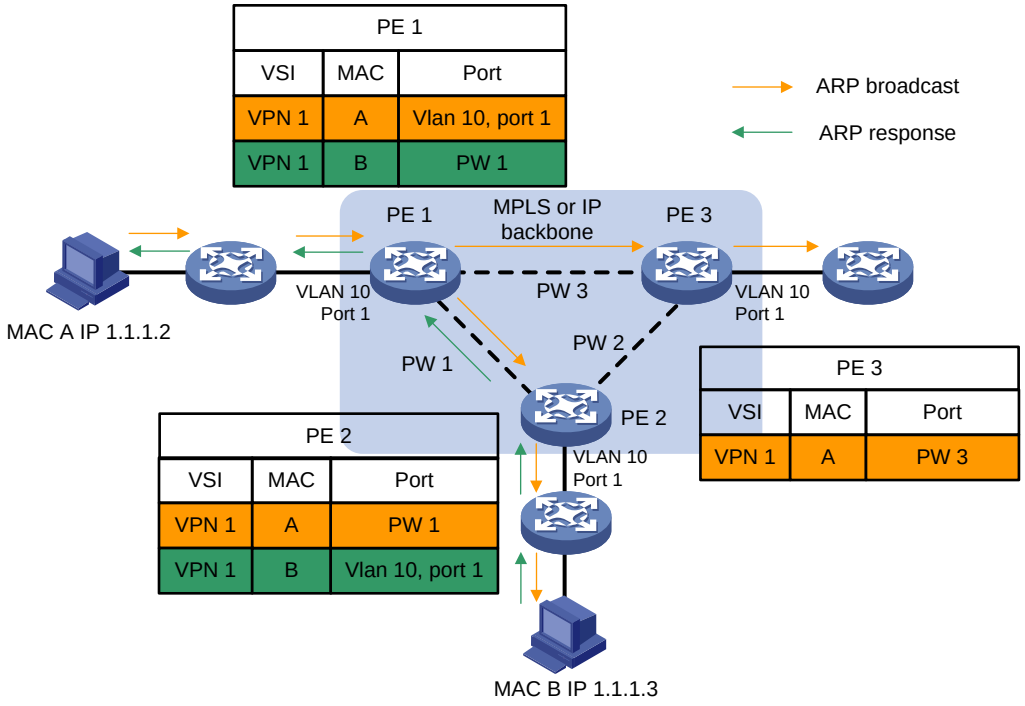
- Learning the source MAC addresses of directly connected sites.

  If the source MAC address of a packet from a CE does not exist in the MAC address table, the PE learns the source MAC address of the AC connected to the CE.

- Learning the source MAC addresses of remote sites connected through PWs.

  A VSI regards a PW as a logical Ethernet interface. If the source MAC address of a packet from a PW does not exist in the MAC address table, the PE learns the source MAC address on the PW of the VSI.

Figure 2 Source MAC address learning on a PE



MAC address aging

The MAC address table uses an aging timer for each dynamic MAC address entry. If no packet is received from a MAC address before the aging timer expires, VPLS deletes the MAC address.

MAC address withdrawal

When an AC or a PW goes down, the PE deletes MAC addresses on the AC or PW. Then it sends an LDP address withdrawal message to notify all other PEs in the VPLS instance to delete those MAC addresses.

Traffic forwarding and flooding

Unicast traffic forwarding and flooding

After a PE receives a unicast packet from an AC, the PE searches the MAC address table of the VSI bound to the AC for packet forwarding.

- If a match is found, the PE forwards the packet according to the matching entry.
  - If the outgoing interface in the entry is a PW, the PE inserts the PW label to the packet, and adds the public tunnel header to the packet. It then forwards the packet to the remote PE over the PW. If the PW is carried on an LSP tunnel, each packet on the PW contains two labels. The inner label is the PW label, which identifies the PW and ensures that the packet is forwarded to the correct VSI. The outer label is the public LSP tunnel label, which ensures that the packet is correctly forwarded to the remote PE.
  - If the outgoing interface in the entry is a local interface, the PE directly forwards the packet to the local interface.
- If no match is found, the PE floods the packet to all other ACs and PWs in the VSI.

After a PE receives a unicast packet from a PW, the PE searches the MAC address table of the VSI bound to the PW for packet forwarding.

- If a match is found, the PE forwards the packet through the outgoing interface in the matching entry.
- If no match is found, the PE floods the packet to all ACs in the VSI.

Multicast and broadcast traffic forwarding and flooding

After a PE receives a multicast or broadcast packet from an AC, the PE floods the packet to all other ACs and the PWs in the VSI bound to the AC.

After a PE receives a multicast or broadcast packet from a PW, the PE floods the packet to all ACs in the VSI bound to the PW.

PW full mesh and split horizon

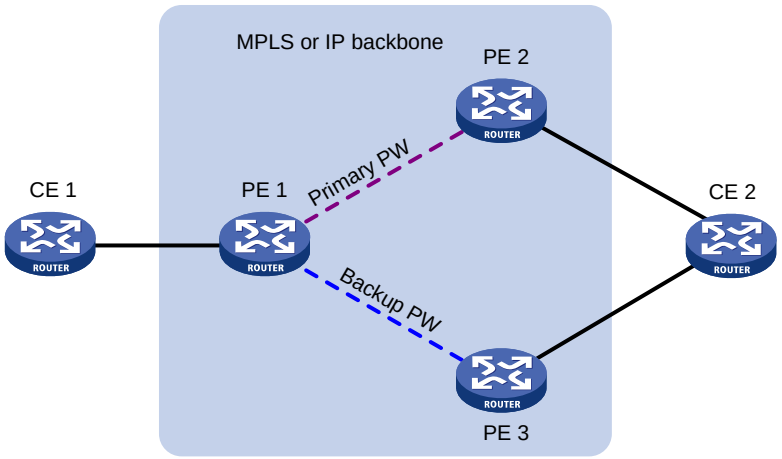
A Layer 2 network requires a loop prevention protocol such as STP to avoid loops. However, a loop prevention protocol on PEs brings management and maintenance difficulties. Therefore, VPLS uses the following methods to prevent loops:

- **Full mesh**—Every two PEs in a VPLS instance must establish a PW. The PWs form a full mesh among PEs in the VPLS instance.
- **Split horizon**—A PE does not forward packets received from a PW to any other PWs in the same VSI but only forwards those packets to ACs.

PW redundancy

PW redundancy provides redundant links between PEs so that the customer networks can communicate when the path over one PW fails. As shown in Figure 3, PE 1 establishes two PWs (one primary and one backup). The CEs communicate through the primary PW. When the primary PW fails, PE 1 brings up the backup PW and forwards packets from CE 1 to CE 2 through the backup PW. When CE 2 receives the packets, it updates its MAC address table, so that packets from CE 2 to CE 1 also travel through the backup PW. Only static PWs and LDP PWs support PW redundancy.

Figure 3 PW redundancy



VPLS determines whether the primary PW fails according to the LDP session status or the BFD result. The backup PW is used when one of the following conditions exists:

- The public tunnel of the primary PW is deleted, or BFD detects that the public tunnel has failed.
- The primary PW is deleted because the LDP session between PEs goes down, or BFD detects that the primary PW has failed.
- A manual PW switchover is performed.

A PW can be in either of the following states:

- **Active**—The PW is in active state and can forward packets.
- **Standby**—The PW is in standby state and cannot forward packets.

LDP PWs support the independent and master/slave PW redundancy operation modes.

- **Independent mode**—The two PEs of a PW use LDP to advertise their respective PW active/standby state to each other. A PW can forward traffic only when it is up and active at both ends of the PW. In this mode, make sure both PEs of a PW use the independent PW redundancy operation mode.
- **Master/slave mode**—One PE of a PW operates as the master node and the other PE operates as the slave node. The master PE determines the PW active/standby state and then uses LDP to advertise the PW state to the slave PE. The slave PE uses the same PW state as the master PE based on the information received from the master PE. In this way, the master and slave PEs for the set of redundant PWs can use the same active PW to forward user packets.

  A slave node does not advertise the PW active/standby state to the master node, and the master node ignores the PW active/standby state received from the slave nodes.

Restrictions and guidelines: VPLS configuration

A PE does not transmit LACP and LLDP packets through the VPLS network.

If a PE is enabled with the spanning tree feature, the PE does not transmit BPDUs through the VPLS network. To enable the PE to transmit BPDUs through the VPLS network, use the **`stp transparent enable`** command to enable BPDU transparent transmission for the PE. For more information about the **`stp transparent enable`** command, see STP commands in *Layer 2—LAN Switching Command Reference*.

The next-hop outgoing interface on the public network on the PW side cannot be a Layer 3 Ethernet subinterface.

VPLS tasks at a glance

Perform the following configuration tasks on a PE:

1. Enabling L2VPN
2. Configuring an AC
3. Configuring a VSI
4. Configuring a PW

   Choose the following tasks as needed:

   - (Optional.) Configuring a PW class
   - Configuring a static PW
   - Configuring an LDP PW
   - Configuring a BGP PW
   - Configuring a BGP auto-discovery LDP PW
   - (Optional.) Configuring the BGP L2VPN address family

     You can configure this task to control route advertisement and selection in BGP L2VPN address family when you configure a BGP PW or BGP auto-discovery LDP PW.
   - (Optional.) Maintaining BGP sessions

     If BGP configuration is changed, you can reset or soft-reset BGP sessions to validate the new configuration for BGP PWs and BGP auto-discovery LDP PWs.
5. Binding an AC to a VSI
6. (Optional.) Configuring PW redundancy
7. (Optional.) Configuring MAC address learning
8. (Optional.) Enabling VPLS statistics
9. (Optional.) Enabling SNMP notifications for L2VPN PW

Prerequisites for VPLS

To configure a VPLS network, perform the following tasks:

- Configure an IGP to ensure IP connectivity within the backbone.
- Configure basic MPLS, LDP, GRE, to establish public tunnels on the backbone network.

If the public tunnel is a GRE tunnel, perform the following tasks:

- Configure an LSR ID for the PE by using the **`mpls lsr-id`** command.
- Enable MPLS with the **`mpls enable`** command on the backbone interface of the PE.

For more information about the **`mpls lsr-id`** and **`mpls enable`** commands, see *MPLS Command Reference*.

Enabling L2VPN

1. Enter system view.

   `system-view`
2. Enable L2VPN.

   `l2vpn enable`

   By default, L2VPN is disabled.

Configuring an AC

Configuring an Ethernet service instance

About this task

When the PE is connected to a CE through a Layer 2 Ethernet or Layer 2 aggregate interface, configure an Ethernet service instance on the interface to match packets from the AC. The Ethernet service instance provides flexible methods for matching packets to be forwarded in the VSI.

Restrictions and guidelines

For information about configuring the match criterion of an Ethernet service instance by using the **encapsulation** command, see VPLS commands in *MPLS Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Create an Ethernet service instance and enter Ethernet service instance view.
   service-instance *instance-id*
4. Configure a packet match criterion for the Ethernet service instance.
   - Match packets with the specified inner VLAN IDs.
     encapsulation c-vid { *vlan-id* | *vlan-id-list* }
   - Match packets with the specified outer VLAN IDs.
     encapsulation s-vid { *vlan-id* | *vlan-id-list* } [**only-tagged**]
   - Match packets with the specified outer VLAN ID and inner VLAN IDs.
     encapsulation s-vid *vlan-id* **c-vid** { *vlan-id-list* | **all** }
   - Match packets that have a VLAN tag or packets that do not have a VLAN tag.
     encapsulation untagged

   By default, no packet match criterion is configured.

Configuring a VSI

1. Enter system view.
   system-view
2. Create a VSI and enter VSI view.
   vsi *vsi-name*
3. (Optional.) Configure a description for the VSI.
   description *text*

   By default, no description is configured for a VSI.
4. (Optional.) Set the default PW ID for the VSI.
   default-pw-id *default-pw-id*

By default, no default PW ID is configured for the VSI.

5. Set an MTU for the VSI.

   mtu *size*

   By default, the MTU of a VSI is 1500 bytes.

   This command takes effect only on protocol packets and does not take effect on forwarded data packets.

6. (Optional.) Enable the VSI.

   undo shutdown

   By default, a VSI is enabled.

Configuring a PW

Configuring a PW class

About this task

In a PW class, you can configure PW attributes such as the PW data encapsulation type. To simplify PW configuration, you can configure PWs with the same attributes by referencing the same PW class.

Procedure

1. Enter system view.

   system-view

2. Create a PW class and enter PW class view.

   pw-class *class-name*

   By default, no PW classes exist.

3. (Optional.) Specify the PW data encapsulation type.

   pw-type { **ethernet** | **vlan** } [**force-for-vpls**]

   By default, the PW data encapsulation type is VLAN.

Configuring a static PW

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Specify static signaling for PWs, and enter VSI static view.

   pwsignaling static

   By default, no PW signaling protocol is specified.

4. Configure a static PW, and enter VSI static PW view.

   peer *ip-address* [**pw-id** *pw-id*] **in-label** *label-value* **out-label** *label-value* [**no-split-horizon** | **pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *

Parameter	Description
pw-id *pw-id*	If both the default PW ID in the **default-pw-id** command and a PW ID in the **peer** command are configured, the PW ID in the **peer** command is used. If only the default PW ID is configured, the default PW ID is used. If no default PW ID is configured, you must provide a PW ID in the **peer** command.
no-split-horizon	You must specify the **no-split-horizon** keyword to disable split horizon when you configure an NPE to establish a U-PW with a UPE.

Configuring an LDP PW

Prerequisites

Before you configure an LDP PW, enable global and interface MPLS LDP on the PE. For information about MPLS LDP configuration, see "Configuring LDP."

Procedure

1. Enter system view.

   system-view
2. Enter VSI view.

   vsi *vsi-name*
3. Specify LDP signaling for PWs, and enter VSI LDP signaling view.

   pwsignaling ldp

   By default, no PW signaling protocol is specified.
4. Configure an LDP PW, and enter VSI LDP PW view.

   peer *ip-address* [**pw-id** *pw-id*] [**no-split-horizon** | **pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *

Parameter	Description
pw-id *pw-id*	If both the default PW ID in the **default-pw-id** command and a PW ID in the **peer** command are configured, the PW ID in the **peer** command is used. If only the default PW ID is configured, the default PW ID is used. If no default PW ID is configured, you must provide a PW ID in the **peer** command.
no-split-horizon	You must specify the **no-split-horizon** keyword to disable split horizon when you configure an NPE to establish a U-PW with a UPE.

Configuring a BGP PW

Configuring BGP to advertise VPLS label block information

1. Enter system view.

   system-view
2. Enable BGP instance and enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is disabled.

For more information about this command, see BGP in *Layer 3—IP Routing Command Reference*.

3. Configure the remote PE as a BGP peer.

   peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*

   For more information about this command, see BGP in *Layer 3—IP Routing Command Reference*.

4. Create the BGP L2VPN address family and enter BGP L2VPN address family view.

   address-family l2vpn

5. Enable BGP to exchange L2VPN information with the specified peer or peer group.

   peer { *group-name* | *ip-address* [*mask-length*] } **enable**

   By default, BGP cannot exchange L2VPN information with any peer or peer group.

   For more information about this command, see BGP in *Layer 3—IP Routing Command Reference*.

6. Enable BGP to exchange label block information with the specified peer or peer group.

   peer { *group-name* | *ip-address* [*mask-length*] } **signaling**

   By default, BGP can exchange label block information with a BGP L2VPN peer or peer group.

7. (Optional.) Configure the BGP L2VPN address family.

   For more information, see "Configuring the BGP L2VPN address family."

8. (Optional.) Maintain BGP sessions.

   For more information, see "Maintaining BGP sessions."

Creating a BGP PW

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Configure the VSI to automatically discover neighbors through BGP and enter auto-discovery VSI view.

   auto-discovery bgp

   By default, a VSI does not automatically discover neighbors through BGP.

4. Configure an RD for the auto-discovery VSI.

   route-distinguisher *route-distinguisher*

   By default, no RD is configured for the auto-discovery VSI.

5. Configure route targets for the auto-discovery VSI.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, no route targets are configured for the auto-discovery VSI.

6. (Optional.) Specify a PW class for the auto-discovery VSI.

   pw-class *class-name*

   By default, no PW class is specified.

7. (Optional.) Specify a tunnel policy for the auto-discovery VSI.

   tunnel-policy *tunnel-policy-name*

   By default, no tunnel policy is specified.

8. Use BGP to create a PW to an automatically discovered remote PE and enter auto-discovery VSI BGP signaling view.

   signaling-protocol bgp

By default, no signaling protocol is specified.

9. Create a local site.

   site *site-id* [**range** *range-value*] [**default-offset** *default-offset*]

Configuring a BGP auto-discovery LDP PW

Prerequisites for configuring a BGP auto-discovery LDP PW

Before you perform this configuration, enable MPLS LDP on interfaces and globally. For more information, see "Configuring LDP."

Configuring BGP to advertise VPLS PE information

1. Enter system view.

   system-view

2. Enable BGP instance and enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is disabled.

   For more information about this command, see BGP in *Layer 3—IP Routing Command Reference*.

3. Configure the remote PE as a BGP peer.

   peer { *group-name* | *ip-address* [*mask-length*] } **as-number** *as-number*

   For more information about this command, see BGP in *Layer 3—IP Routing Command Reference*.

4. Create the BGP L2VPN address family and enter BGP L2VPN address family view.

   address-family l2vpn

5. Enable BGP to exchange VPLS PE information with the specified peer or peer group.

   peer { *group-name* | *ip-address* [*mask-length*] } **enable**

   By default, BGP cannot exchange VPLS PE information with a peer or peer group.

   For more information about this command, see BGP in *Layer 3—IP Routing Command Reference*.

6. Enable BGP to exchange VPLS PE information with the specified peer or peer group.

   peer { *group-name* | *ip-address* [*mask-length*] } **auto-discovery** [**non-standard**]

   By default, BGP can exchange VPLS PE information with a BGP L2VPN peer or peer group by using RFC 6074 MP_REACH_NLRI.

7. (Optional.) Configure the BGP L2VPN address family.

   For more information, see "Configuring the BGP L2VPN address family."

8. (Optional.) Maintain BGP sessions.

   For more information, see "Maintaining BGP sessions."

Creating a BGP auto-discovery LDP PW

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Configure the VSI to automatically discover neighbors through BGP and enter auto-discovery VSI view.

   auto-discovery bgp

By default, a VSI does not automatically discover neighbors through BGP.

4. Configure an RD for the auto-discovery VSI.
   route-distinguisher *route-distinguisher*
   By default, no RD is configured for the auto-discovery VSI.
5. Configure route targets for the auto-discovery VSI.
   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]
   By default, no route targets are configured for the auto-discovery VSI.
6. (Optional.) Specify a PW class for the auto-discovery VSI.
   pw-class *class-name*
   By default, no PW class is specified.
7. (Optional.) Specify a tunnel policy for the auto-discovery VSI.
   tunnel-policy *tunnel-policy-name*
   By default, no tunnel policy is specified.
8. Use LDP to create a PW to an automatically discovered remote PE and enter auto-discovery VSI LDP signaling view.
   signaling-protocol ldp
   By default, no signaling protocol is specified.
9. Configure a VPLS ID for the VSI.
   vpls-id *vpls-id*
   By default, no VPLS ID is configured.

Configuring the BGP L2VPN address family

About this task

When you configure a BGP PW or a BGP auto-discovery LDP PW, you can perform this task to control route advertisement and selection in the BGP L2VPN address family.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP L2VPN address family view.
   address-family l2vpn
4. Permit the local AS number to appear in routes from the specified peer or peer group and specify the maximum number of the appearances.
   peer { *group-name* | *ip-address* [*mask-length*] } **allow-as-loop** [*number*]
   By default, the local AS number is not allowed in routes from a peer or peer group.
5. Enable route target-based filtering of incoming BGP L2VPN information.
   policy vpn-target
   By default, route target-based filtering of incoming BGP L2VPN information is enabled.
6. Configure BGP route reflection.
   - Configure the device as a route reflector and specify a peer or peer group as its client.
     peer { *group-name* | *ip-address* [*mask-length*] } **reflect-client**

By default, no route reflector or client is configured.

- Enable L2VPN information reflection between clients.

  reflect between-clients

  By default, L2VPN information reflection is enabled between clients.
- Configure the cluster ID of the route reflector.

  reflector cluster-id { *cluster-id* | *ip-address* }

  By default, a route reflector uses its own router ID as the cluster ID.
- Configure the route reflector to filter reflected L2VPN information.

  rr-filter *ext-comm-list-number*

  By default, a route reflector does not filter reflected L2VPN information.

7. Set the optimal route selection delay timer.

route-select delay *delay-value*

By default, the optimal route selection delay timer is 0 seconds, which means optimal route selection is not delayed. For more information about this command, see BGP commands in *Layer 3—IP Routing Command Reference*.

Maintaining BGP sessions

About this task

If BGP configuration is changed, you can reset or soft-reset BGP sessions to validate the new configuration for BGP PWs and BGP auto-discovery LDP PWs.

A soft reset updates BGP routing information without tearing down BGP connections. A reset operation updates BGP routing information by tearing down, and then re-establishing BGP connections. Soft reset requires that BGP peers have route refresh capability (support the ROUTE-REFRESH message).

For more information about the commands in this task, see BGP in *Layer 3—IP Routing Command Reference*.

Procedure

To maintain BGP sessions, execute one of the following commands in user view:

- Soft-reset L2VPN BGP sessions.

  refresh bgp [**instance** *instance-name*] { *ip-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **l2vpn**
- Reset L2VPN BGP sessions.

  reset bgp [**instance** *instance-name*] { *as-number* | *ip-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **l2vpn**

Binding an AC to a VSI

About binding an AC to a VSI

On a Layer 2 Ethernet or Layer 2 aggregate interface, you can create an Ethernet service instance and bind it to a VSI. The Ethernet service instance matches packets received on the interface. The matching packets are then forwarded based on the MAC address table of the VSI. An Ethernet service instance can match all packets, tagged packets, or untagged packets.

When you bind an AC to a VSI, you can associate Track with the AC. Then, the AC is up only when one or more of the associated track entries are positive.

Restrictions and guidelines for binding an AC to a VSI

If a Layer 2 Ethernet interface has been added to a link aggregation group, you cannot bind an Ethernet service instance on the interface to a VSI, and vice versa.

Binding an Ethernet service instance to a VSI

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view:
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view:
     interface bridge-aggregation *interface-number*
3. Create an Ethernet service instance and enter Ethernet service instance view.
   service-instance *instance-id*
   By default, no Ethernet service instances exist.
4. Bind the Ethernet service instance to a VSI.
   xconnect vsi *vsi-name* [**access-mode** { **ethernet** | **vlan** }] [**track** *track-entry-number*&<1-3>]
   By default, an Ethernet service instance is not bound to any VSI.

Configuring PW redundancy

Configuring static PW redundancy

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Specify static signaling for PWs, and enter VSI static view.
   pwsignaling static
   By default, no PW signaling protocol is specified for the VSI.
4. (Optional.) Specify the switchover mode and set the wait time for the switchover.
   revertive { **wtr** *wtr-time* | **never** }
   By default, the switchover mode is revertive and the switchover wait time is 0 seconds.
5. Configure a static PW and enter VSI static PW view.
   peer *ip-address* [**pw-id** *pw-id*] [**in-label** *label-value* **out-label** *label-value*] [**no-split-horizon** | **pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *
6. Configure a backup static PW and enter VSI static backup PW view.
   backup-peer *ip-address* [**pw-id** *pw-id*] **in-label** *label-value* **out-label** *label-value* [**pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *

If both the default PW ID in the **default-pw-id** command and a PW ID in the **backup-peer** command are configured, the PW ID in the **backup-peer** command is used. If only the default PW ID is configured, the default PW ID is used. If no default PW ID is configured, you must provide a PW ID in the **backup-peer** command.

7. (Optional.) Manually switch the traffic:
   a. Return to user view.
      return
   b. Manually switch traffic of the PW to its backup PW.
      l2vpn switchover peer *ip-address* **pw-id** *pw-id*

Configuring LDP PW redundancy

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Specify LDP signaling for PWs, and enter VSI LDP signaling view.
   pwsignaling ldp
   By default, no PW signaling protocol is specified for the VSI.
4. (Optional.) Use the master/slave PW redundancy mode and configure the local PE as the master node.
   pw-redundancy { **independent** | **master** }
   By default, the PW redundancy mode is master/slave and the local PE operates as the slave node.
   Do not configure this command on the local PE if the remote PE does not support the master/slave PW redundancy mode.
5. (Optional.) Specify the switchover mode and set the wait time for the switchover.
   revertive { **wtr** *wtr-time* | **never** }
   By default, the switchover mode is revertive and the switchover wait time is 0 seconds.
6. Configure an LDP PW and enter VSI LDP PW view.
   peer *ip-address* [**pw-id** *pw-id*] [**ignore-standby-state** | **no-split-horizon** | **pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *
   For the local PE to ignore the PW active/standby states received from the remote PE, specify the **ignore-standby-state** keyword.
7. Configure a backup LDP PW and enter VSI LDP backup PW view.
   backup-peer *ip-address* [**pw-id** *pw-id*] [**pw-class** *class-name* | **tunnel-policy** *tunnel-policy-name*] *
   If both the default PW ID in the **default-pw-id** command and a PW ID in the **backup-peer** command are configured, the PW ID in the **backup-peer** command is used. If only the default PW ID is configured, the default PW ID is used. If no default PW ID is configured, you must provide a PW ID in the **backup-peer** command.
8. (Optional.) Manually switch the traffic:
   a. Return to user view.
      return
   b. Manually switch traffic of the PW to its backup PW.
      l2vpn switchover peer *ip-address* **pw-id** *pw-id*

Configuring MAC address learning

Procedure

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable MAC address learning for the VSI.
   mac-learning enable
   By default, MAC address learning is enabled for a VSI.

Enabling VPLS statistics

Enabling packet statistics for an AC

Restrictions and guidelines for enabling packet statistics for an AC

If the AC is an Ethernet service instance, follow these restrictions and guidelines:

- To display packet statistics for the Ethernet service instance, use the **display l2vpn service-instance verbose** command. To clear packet statistics for the Ethernet service instance, use the **reset l2vpn statistics ac** command.
- Packet statistics for the Ethernet service instance takes effect only when you bind the Ethernet service instance to a VSI. If you change the bound VSI during the statistics collection, the packet statistics are re-collected.

Enabling packet statistics for an Ethernet service instance

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view:
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view:
     interface bridge-aggregation *interface-number*
3. Enter Ethernet service instance view.
   service-instance *instance-id*
4. Enable packet statistics for the Ethernet service instance.
   statistics enable
   By default, packet statistics is disabled for an Ethernet service instance.

Setting the L2VPN traffic statistics collection interval

About this task

Perform this task to set the L2VPN traffic statistics collection interval ACs.

Procedure

1. Enter system view.
   system-view
2. Set the L2VPN traffic statistics collection interval.
   l2vpn statistics interval *interval*
   By default, the L2VPN traffic statistics collection interval is 15 minutes.

Enabling SNMP notifications for L2VPN PW

About this task

This feature enables L2VPN to generate SNMP notifications when the device does not support L2VPN flow labels, or PW deletions, PW switchovers, or PW status changes occur. For L2VPN event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for L2VPN PW.
   snmp-agent trap enable l2vpn [**flow-label-nonsupport** | **pw-delete** | **pw-switch** | **pw-up-down**] *
   By default, SNMP notifications for L2VPN PW are disabled.

Display and maintenance commands for VPLS

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display BGP L2VPN peer group information.	**display bgp** [**instance** *instance-name*] **group l2vpn** [**group-name** *group-name*]
Display VPLS PE information discovered by BGP.	**display bgp** [**instance** *instance-name*] **l2vpn auto-discovery** [**peer** *ip-address* { **advertised** | **received** } | **route-distinguisher** *route-distinguisher* [**pe-address** *ip-address* [**advertise-info**]]]
Display VPLS label block information discovered by BGP.	**display bgp** [**instance** *instance-name*] **l2vpn signaling** [**peer** *ip-address* { **advertised** | **received** } | **route-distinguisher** *route-distinguisher* [**site-id** *site-id* [**label-offset** *label-offset* [**advertise-info**]]]]
Display BGP L2VPN peer information.	**display bgp** [**instance** *instance-name*] **peer l2vpn** [*ip-address mask-length* | **group-name** *group-name* **log-info** | *ip-address* { **log-info** | **verbose** } |

	verbose]
Display BGP L2VPN update group information.	**display bgp** [**instance** *instance-name*] **update-group l2vpn** [*ip-address*]
Display information about automatically discovered VPLS PEs.	**display l2vpn auto-discovery** [**peer** *ip-address*] [**vsi** *vsi-name*]
Display VPLS label block information.	**display l2vpn bgp** [**instance** *instance-name*] [**peer** *ip-address* | **local**] [**vsi** *vsi-name*] [**verbose**]
Display L2VPN forwarding information.	**display l2vpn forwarding** { **ac** | **pw** } [**vsi** *vsi-name*] [**slot** *slot-number*] [**verbose**]
Display MAC address table information for one or all VSIs.	**display l2vpn mac-address** [**vsi** *vsi-name*] [**dynamic**] [**count** | **verbose**]
Display the MAC address move records for all VSIs after the device is started.	**display l2vpn mac-address mac-move** [**slot** *slot-number*]
Display L2VPN PW information.	**display l2vpn mac-address** [**vsi** *vsi-name*] [**dynamic**] [**count** | **verbose**]
Display PW class information.	**display l2vpn pw-class** [*class-name*] [**verbose**]
Display Ethernet service instance information.	**display l2vpn service-instance** [**interface** *interface-type interface-number* [**service-instance** *instance-id*]] [**verbose**]
Display VSI information.	**display l2vpn vsi** [**name** *vsi-name*] [**verbose**]
Reset L2VPN BGP sessions.	**reset bgp** { *as-number* | *ip-address* [*mask-length*] | **all** | **external** | **group** *group-name* | **internal** } **l2vpn**
Clear MAC address entries for one or all VSIs.	**reset l2vpn mac-address** [**vsi** *vsi-name*]

Contents

Configuring MPLS OAM 1

About MPLS OAM 1
Fault management tools 1
MPLS ping 1
MPLS tracert 1
Protocols and standards 1
Restrictions and guidelines: MPLS OAM configuration 1
Verifying LSP connectivity 2
About LSP connectivity verification 2
Performing MPLS ping for LSPs 2
Performing MPLS tracert for LSPs 2

Configuring MPLS OAM

About MPLS OAM

MPLS Operation, Administration, and Maintenance (OAM) provides fault management tools for the following purposes:

- MPLS data plane connectivity verification.
- Data plane and control plane consistency verification.
- Fault locating.

Fault management tools

MPLS OAM provides on-demand fault management tools, which must be triggered manually, such as MPLS ping and MPLS tracert.

MPLS ping

MPLS ping tests the connectivity of a tunnel. At the ingress node, MPLS ping adds the label associated with a tunnel into an MPLS echo request and sends it to the egress node over the tunnel. The egress node processes the request and returns an MPLS echo reply to the ingress node. An MPLS echo reply with a success notification indicates that the tunnel is available for data forwarding. An MPLS echo reply with an error code indicates that the tunnel has failed.

MPLS tracert

MPLS tracert displays the path that a tunnel travels from the ingress to the egress to locate errors on the tunnel. MPLS tracert consecutively sends MPLS echo requests along the tunnel, with the TTL increasing from 1 to a specific value. Each hop along the tunnel returns an MPLS echo reply to the ingress due to TTL timeout so the ingress can collect information about each hop along the tunnel. This information allows you to locate the failed node or access information for each hop, for example, the label allocated by each downstream hop.

Protocols and standards

- RFC 4379, *Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures*

Restrictions and guidelines: MPLS OAM configuration

Only advanced mode and expert mode support MPLS OAM. For more information about system operating modes, see device management in *Fundamentals Configuration Guide*.

The device does not support MPLS forwarding for tunneled packets.

Verifying LSP connectivity

About LSP connectivity verification

To verify LSP connectivity, you can use the **ping mpls out-labels** command or the **tracert mpls out-labels** command to trigger LSP connectivity verification as needed.

Performing MPLS ping for LSPs

To verify MPLS LSP connectivity by specifying the outgoing labels, execute the following command in any view:

ping mpls [**-a** *source-ip* | **-c** *count* | **-exp** *exp-value* | **-h** *ttl-value* | **-m** *wait-time* | **-r** *reply-mode* | **-rtos** *tos-value* | **-s** *packet-size* | **-t** *time-out* | **-v**] * **out-labels** *out-label-value*&<1-5> **interface** *interface-type interface-number* [**nexthop** *nexthop-address*]

Performing MPLS tracert for LSPs

To trace the path that the LSPs of the specified outgoing labels take from the ingress node to the egress node, execute the following command in any view:

tracert mpls [**-a** *source-ip* | **-exp** *exp-value* | **-h** *ttl-value* | **-r** *reply-mode* | **-rtos** *tos-value* | **-t** *time-out* | **-v** | **fec-check**]* **out-labels** *out-label-value*&<1-5> **interface** *interface-type interface-number* [**nexthop** *nexthop-address*]

Contents

Configuring MCE 1

About MCE 1
MPLS L3VPN problems solved by MCE 1
Basic MPLS L3VPN architecture 1
MCE-related MPLS L3VPN concepts 2
MCE working mechanism 3
Restrictions and guidelines: MCE configuration 4
MCE tasks at a glance 4
Configuring VPN instances 5
Creating a VPN instance 5
Associating a VPN instance with a Layer 3 interface 5
Configuring route related attributes for a VPN instance 6
Configuring routing between an MCE and a VPN site 6
Configuring static routing between an MCE and a VPN site 6
Configuring RIP between an MCE and a VPN site 7
Configuring OSPF between an MCE and a VPN site 7
Configuring IS-IS between an MCE and a VPN site 8
Configuring EBGP between an MCE and a VPN site 9
Configuring IBGP between an MCE and a VPN site 10
Configuring routing between an MCE and a PE 11
About routing between an MCE and a PE 11
Configuring static routing between an MCE and a PE 11
Configuring RIP between an MCE and a PE 12
Configuring OSPF between an MCE and a PE 12
Configuring IS-IS between an MCE and a PE 12
Configuring EBGP between an MCE and a PE 13
Configuring IBGP between an MCE and a PE 13
Display and maintenance commands for MCE 14
MCE configuration examples 14
Example: Configuring the MCE that uses OSPF to advertise VPN routes to the PE 14
Example: Configuring the MCE that uses EBGP to advertise VPN routes to the PE 19

Configuring MCE

About MCE

The Multi-VPN Instance Customer Edge (MCE) feature allows multiple VPNs to share a CE with ensured data security in an MPLS L3VPN network. MCE provides traffic separation between VPNs by distinguishing routes of the VPNs.

MPLS L3VPN problems solved by MCE

MPLS L3VPN is a L3VPN technology used to interconnect geographically dispersed VPN sites. MPLS L3VPN uses BGP to advertise VPN routes and uses MPLS to forward VPN packets over a service provider backbone.

In conventional MPLS L3VPN, a VPN site must exclusively use a CE to connect a PE. With increasing business, a private network must be divided into multiple VPNs and the VPNs must be isolated from one another for data security. Using a CE exclusively for each VPN costs much in buying and maintaining the network devices. Using a common CE for multiple VPNs cannot ensure data security because the VPNs use the same routing table.

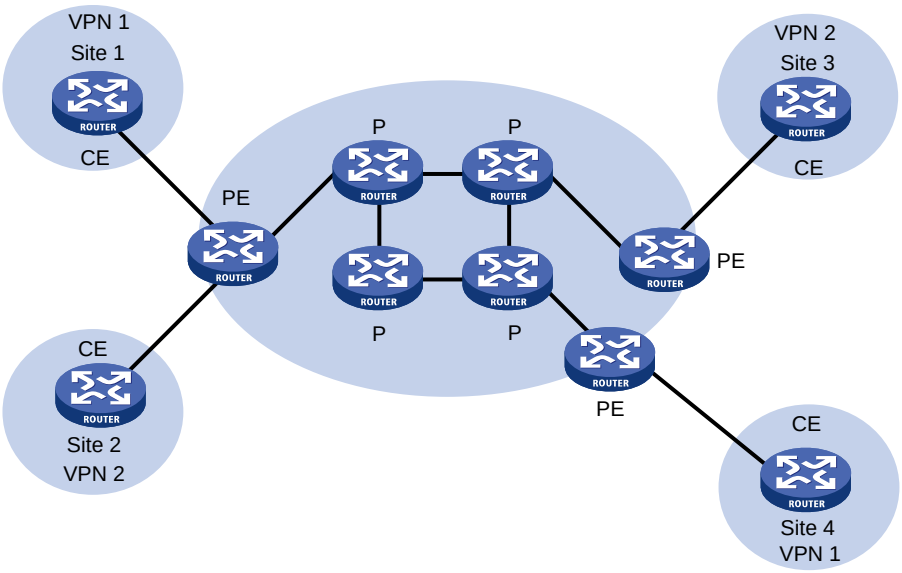
The MCE feature offers balance between cost and security. It creates multiple VPN instances on a CE to provide logically separated routing tables and address spaces for the VPNs, so the VPNs can share the CE. This CE is called the MCE device.

Basic MPLS L3VPN architecture

A basic MPLS L3VPN architecture has the following types of devices:

- **Customer edge device**—A CE device resides on a customer network and has one or more interfaces directly connected to a service provider network. It does not support MPLS.
- **Provider edge device**—A PE device resides at the edge of a service provider network and is connected to one or more CEs. All MPLS VPN services are processed on PEs.
- **Provider device**—A P device is a core device on a service provider network. It is not directly connected to any CEs. A P device has only basic MPLS forwarding capability and does not handle VPN routing information.

Figure 1 Basic MPLS L3VPN architecture



MCE-related MPLS L3VPN concepts

Site

A site has the following features:

- A site is a group of IP systems with IP connectivity that does not rely on any service provider networks.
- The classification of a site depends on the topology relationship of the devices, rather than the geographical positions. However, the devices at a site are, in most cases, adjacent to each other geographically.
- The devices at a site can belong to multiple VPNs, which means that a site can belong to multiple VPNs.
- A site is connected to a provider network through one or more CEs. A site can contain multiple CEs, but a CE can belong to only one site.

Sites connected to the same provider network can be classified into different sets by policies. Only the sites in the same set can access each other through the provider network. Such a set is called a VPN.

VPN instance

VPN instances implement route isolation, data independence, and data security for VPNs.

A VPN instance has the following components:

- A separate Label Forwarding Information Base (LFIB).
- An IP routing table.
- Interfaces bound to the VPN instance.
- VPN instance administration information, including route distinguishers (RDs), route targets (RTs), and route filtering policies.

To associate a site with a VPN instance, bind the VPN instance to the PE's interface connected to the site. A site can be associated with only one VPN instance, and different sites can be associated with the same VPN instance. A VPN instance contains the VPN membership and routing rules of associated sites.

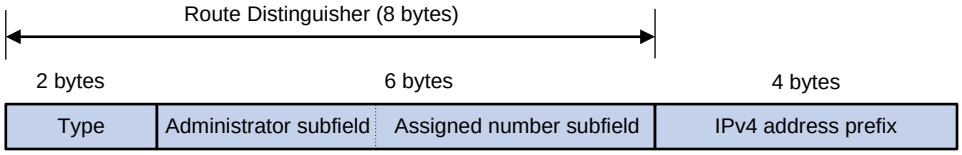
VPN-IPv4 address

Each VPN independently manages its address space. The address spaces of VPNs might overlap. For example, if both VPN 1 and VPN 2 use the addresses on subnet 10.110.10.0/24, address space overlapping occurs.

Multiprotocol BGP (MP-BGP) can solve this problem by advertising VPN-IPv4 addresses (also called VPNv4 addresses).

As shown in Figure 2, a VPN-IPv4 address consists of 12 bytes. The first eight bytes represent the RD, followed by a four-byte IPv4 prefix. The RD and the IPv4 prefix form a unique VPN-IPv4 prefix.

Figure 2 VPN-IPv4 address structure



An RD can be in one of the following formats:

- When the Type field is 0, the Administrator subfield occupies two bytes, the Assigned number subfield occupies four bytes, and the RD format is *16-bit AS number*:*32-bit user-defined number*. For example, 100:1.
- When the Type field is 1, the Administrator subfield occupies four bytes, the Assigned number subfield occupies two bytes, and the RD format is *32-bit IPv4 address*:*16-bit user-defined number*. For example, 172.1.1.1:1.
- When the Type field is 2, the Administrator subfield occupies four bytes, the Assigned number subfield occupies two bytes, and the RD format is *32-bit AS number*:*16-bit user-defined number*, where the minimum value of the AS number is 65536. For example, 65536:1.

To guarantee global uniqueness for a VPN-IPv4 address, do not set the Administrator subfield to any private AS number or private IP address.

Route target attribute

MPLS L3VPN uses route target community attributes to control the advertisement of VPN routing information. A VPN instance on a PE supports the following types of route target attributes:

- **Export target attribute**—A PE sets the export target attribute for VPN-IPv4 routes learned from directly connected sites before advertising them to other PEs.
- **Import target attribute**—A PE checks the export target attribute of VPN-IPv4 routes received from other PEs. If the export target attribute matches the import target attribute of a VPN instance, the PE adds the routes to the routing table of the VPN instance.

Route target attributes define which sites can receive VPN-IPv4 routes, and from which sites a PE can receive routes.

Like RDs, route target attributes can be one of the following formats:

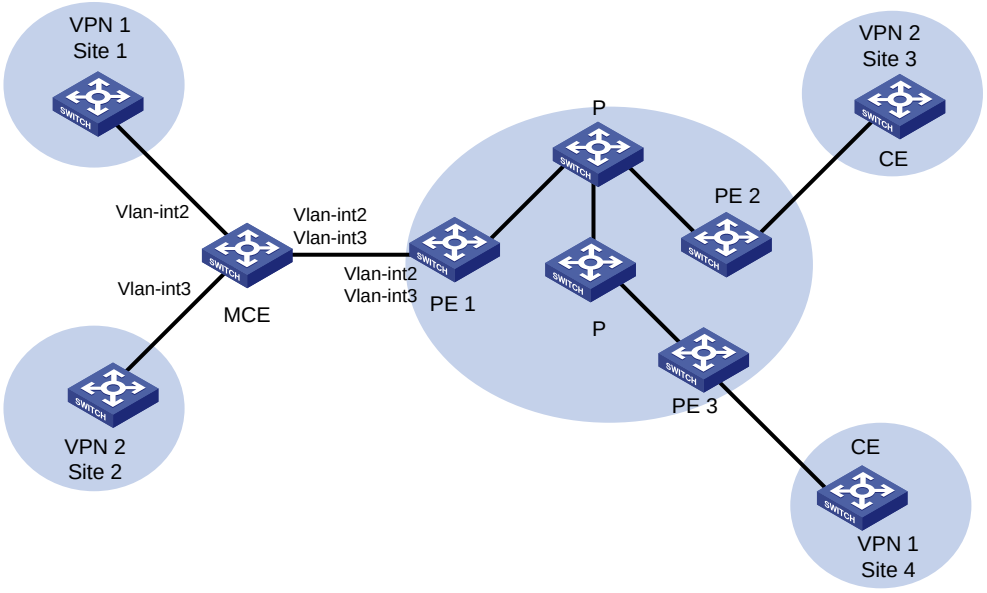
- *16-bit AS number*:*32-bit user-defined number*. For example, 100:1.
- *32-bit IPv4 address:16-bit user-defined number*. For example, 172.1.1.1:1.
- *32-bit AS number*:*16-bit user-defined number*, where the minimum value of the AS number is 65536. For example, 65536:1.

MCE working mechanism

As shown in Figure 3, the MCE exchanges private routes with VPN sites and PE 1, and adds the private routes to the routing tables of corresponding VPN instances.

- **Route exchange between MCE and VPN site**—Create VPN instances VPN 1 and VPN 2 on the MCE. Bind VLAN-interface 2 to VPN 1, and VLAN-interface 3 to VPN 2. The MCE adds a received route to the routing table of the VPN instance that is bound to the receiving interface.
- **Route exchange between MCE and PE**—The MCE connects to PE 1 through trunk links. On PE 1, create VPN instances for VPN 1 and VPN 2. Bind VLAN-interface 2 to VPN 1, and VLAN-interface 3 to VPN 2. The MCE and PE add a received route to the routing table of the VPN instance that is bound to the receiving interface.

Figure 3 Network diagram for the MCE feature



You can configure static routes, RIP, OSPF, IS-IS, EBGP, or IBGP between an MCE and a VPN site and between an MCE and a PE.

NOTE:

To implement dynamic IP assignment for DHCP clients in private networks, you can configure DHCP server or DHCP relay agent on the MCE. When the MCE functions as the DHCP server, the IP addresses assigned to different private networks cannot overlap.

Restrictions and guidelines: MCE configuration

On the MCE, disable routing loop detection to avoid route loss during route calculation, and disable route redistribution between routing protocols to save system resources.

MCE tasks at a glance

To configure MCE, perform the following tasks:

1. Configuring VPN instances

   Perform the following VPN instance tasks on PEs and MCEs:

   a. Creating a VPN instance
   b. Associating a VPN instance with a Layer 3 interface
   c. (Optional.) Configuring route related attributes for a VPN instance
2. Configuring routing between an MCE and a VPN site

3. Configuring routing between an MCE and a PE

Configuring VPN instances

Creating a VPN instance

About this task

A VPN instance is a collection of the VPN membership and routing rules of its associated site. A VPN instance might correspond to more than one VPN.

Procedure

1. Enter system view.
   system-view
2. Create a VPN instance and enter VPN instance view.
   ip vpn-instance *vpn-instance-name*
3. Configure an RD for the VPN instance.
   route-distinguisher *route-distinguisher*
   By default, no RD is configured for a VPN instance.
4. (Optional.) Configure a description for the VPN instance.
   description *text*
   By default, no description is configured for a VPN instance.
5. (Optional.) Configure a VPN ID for the VPN instance.
   vpn-id *vpn-id*
   By default, no VPN ID is configured for a VPN instance.

Associating a VPN instance with a Layer 3 interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   This interface connects the device to the CE.
3. Associate a VPN instance with the interface.
   ip binding vpn-instance *vpn-instance-name*
   By default, an interface is not associated with a VPN instance and it belongs to the public network.

CAUTION:
Associating a VPN instance with an interface or disassociating a VPN instance from an interface will clear the IP address and routing protocol settings of the interface.

Configuration of this command deletes the IP address of the current interface. You must reconfigure an IP address for the interface after configuring the command.

Configuring route related attributes for a VPN instance

Restrictions and guidelines

If you configure route related attributes in both VPN instance view and VPN instance IPv4 address family view, IPv4 VPN uses the configuration in VPN instance IPv4 address family view.

Prerequisites

Before you perform this task, create the routing policies to be used by this task. For information about routing policies, see *Layer 3—IP Routing Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter VPN instance view or VPN instance IPv4 address family view.
   - Enter VPN instance view.

     ip vpn-instance *vpn-instance-name*
   - Execute the following commands in sequence to enter VPN instance IPv4 address family view:

     ip vpn-instance *vpn-instance-name*

     address-family ipv4

3. Configure route targets.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, no route targets are configured.

4. Set the maximum number of active routes.

   routing-table limit *number* { *warn-threshold* | **simply-alert** }

   By default, the number of active routes in a VPN instance is not limited.

   Setting the maximum number of active routes for a VPN instance can prevent the device from learning too many routes.

5. Apply an import routing policy.

   import route-policy *route-policy*

   By default, all routes matching the import target attribute are accepted.

6. Apply an export routing policy.

   export route-policy *route-policy*

   By default, routes to be advertised are not filtered.

Configuring routing between an MCE and a VPN site

Configuring static routing between an MCE and a VPN site

About this task

An MCE can reach a VPN site through a static route. Static routing on a traditional CE is globally effective and does not support address overlapping among VPNs. An MCE supports binding a static route to a VPN instance, so that the static routes of different VPN instances can be isolated from each other.

Perform this configuration on the MCE. On the VPN site, configure a common static route.

Procedure

1. Enter system view.
   system-view
2. Configure a static route for a VPN instance.
   ip route-static vpn-instance *s-vpn-instance-name dest-address* { *mask-length* | *mask* } { *interface-type interface-number* [*next-hop-address*] | *next-hop-address* [**public**] | **vpn-instance** *d-vpn-instance-name next-hop-address* }
3. (Optional.) Configure the default preference for static routes.
   ip route-static default-preference *default-preference*
   The default preference is 60.

Configuring RIP between an MCE and a VPN site

About this task

Binding RIP processes to VPN instances can isolate routes of different VPNs. For more information about RIP, see *Layer 3—IP Routing Configuration Guide*.

Perform this configuration on the MCE. On a VPN site, create a common RIP process.

Procedure

1. Enter system view.
   system-view
2. Create a RIP process for a VPN instance and enter RIP view.
   rip [*process-id*] **vpn-instance** *vpn-instance-name*
   A RIP process can belong to only one VPN instance.
3. Enable RIP on the interface attached to the specified network.
   network *network-address* [*wildcard-mask*]
   By default, RIP is disabled on an interface.
4. Redistribute remote site routes advertised by the PE into RIP.
   import-route *protocol* [*as-number*] [*process-id* | **all-processes** | **allow-ibgp**] [**allow-direct** | **cost** *cost-value* | **route-policy** *route-policy-name* | **tag** *tag*] *
   By default, no route is redistributed into RIP.

Configuring OSPF between an MCE and a VPN site

About this task

Binding OSPF processes to VPN instances can isolate routes of different VPNs. For more information about OSPF, see *Layer 3—IP Routing Configuration Guide*.

Perform this configuration on the MCE. On a VPN site, create a common OSPF process.

Procedure

1. Enter system view.
   system-view
2. Create an OSPF process for a VPN instance and enter OSPF view.

ospf [*process-id*] **router-id** *router-id* **vpn-instance** *vpn-instance-name*

Keywords	Usage guidelines
router-id *router-id*	An OSPF process that is bound to a VPN instance does not use the public network router ID configured in system view. Therefore, you must specify a router ID when creating a process or configure an IP address for a minimum of one interface in the VPN instance.
vpn-instance *vpn-instance-name*	An OSPF process can belong to only one VPN instance. If you delete a VPN instance, all OSPF processes of the VPN instance are also deleted.

3. Redistribute remote site routes advertised by the PE into OSPF.

   import-route *protocol* [*as-number*] [*process-id* | **all-processes** | **allow-ibgp**] [**allow-direct** | **cost** *cost-value* | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   By default, no routes are redistributed into OSPF.
4. Create an OSPF area and enter OSPF area view.

   area *area-id*
5. Enable OSPF on the interface attached to the specified network in the area.

   network *ip-address wildcard-mask*

   By default, an interface neither belongs to any area nor runs OSPF.

Configuring IS-IS between an MCE and a VPN site

About this task

Binding IS-IS processes to VPN instances can isolate routes of different VPNs. For more information about IS-IS, see *Layer 3—IP Routing Configuration Guide*.

Perform this configuration on the MCE. On a VPN site, configure a common IS-IS process.

Procedure

1. Enter system view.

   system-view
2. Create an IS-IS process for a VPN instance and enter IS-IS view.

   isis [*process-id*] **vpn-instance** *vpn-instance-name*

   An IS-IS process can belong to only one VPN instance.
3. Configure a network entity title.

   network-entity *net*

   By default, no NET is configured.
4. Create the IS-IS IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]
5. Redistribute remote site routes advertised by the PE into IS-IS.

   import-route *protocol* [*as-number*] [*process-id* | **all-processes** | **allow-ibgp**] [**allow-direct** | **cost** *cost-value* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   By default, IS-IS does not redistribute routes from any other routing protocol.

   If you do not specify the route level in the command, the command redistributes routes to the level-2 routing table by default.

6. Return to system view.
   quit
7. Enter interface view.
   interface *interface-type interface-number*
8. Enable the IS-IS process on the interface.
   isis enable [*process-id*]
   By default, no IS-IS process is enabled on the interface.

Configuring EBGP between an MCE and a VPN site

About this task

To run EBGP between an MCE and a VPN site, you must configure a BGP peer for each VPN instance on the MCE, and redistribute the IGP routes of each VPN instance on the VPN site.

Restrictions and guidelines

You must configure a VPN site to advertise the VPN network addresses it can reach to the connected MCE.

Configuring the MCE

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
   By default, BGP is not enabled.
3. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
   Commands in BGP-VPN instance view are the same as those in BGP instance view. For more information, see *Layer 3—IP Routing Configuration Guide*.
4. Configure the VPN site as an EBGP peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
5. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
6. Enable BGP to exchange IPv4 unicast routes with the specified peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP does not exchange IPv4 unicast routes with peers.
7. (Optional.) Allow the local AS number to appear in the AS_PATH attribute of routes received from the peer, and set the maximum number of repetitions.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **allow-as-loop** [*number*]
   By default, BGP discards incoming route updates that contain the local AS number.
8. Redistribute remote site routes advertised by the PE into BGP.
   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]
   By default, no routes are redistributed into BGP.

Configuring the VPN site

1. Enter system view.

system-view

2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
   By default, BGP is not enabled.
3. Configure the MCE as an EBGP peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
4. Enter BGP IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
5. Enable BGP to exchange IPv4 unicast routes with the specified peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP does not exchange IPv4 unicast routes with peers or peer groups.
6. Redistribute the IGP routes of the VPN into BGP.
   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]
   By default, no routes are redistributed into BGP.

Configuring IBGP between an MCE and a VPN site

About this task

To run IBGP between an MCE and a VPN site, you must configure a BGP peer for each VPN instance on the MCE, and redistribute the IGP routes of each VPN instance on the VPN site.

Restrictions and guidelines

You must configure a VPN site to advertise VPN network addresses to the connected MCE.

Configuring the MCE

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
   By default, BGP is not enabled.
3. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
4. Configure the VPN site as an IBGP peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
5. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
6. Enable BGP to exchange IPv4 unicast routes with the specified peer or peer group.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP does not exchange IPv4 unicast routes with peers or peer groups.
7. (Optional.) Configure the local device as the RR, and specify the peer as the client of the RR.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **reflect-client**
   By default, no RR or RR client is configured.
   After you configure a VPN site as an IBGP peer, the MCE does not advertise the BGP routes learned from the VPN site to other IBGP peers, including VPNv4 peers. The MCE advertises

routes learned from a VPN site only when you configure the VPN site as a client of the RR (the MCE).

8. Redistribute remote site routes advertised by the PE into BGP.

   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]

   By default, no routes are redistributed into BGP.

Configuring the VPN site

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is not enabled.

3. Configure the MCE as an IBGP peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*

4. Enter BGP IPv4 unicast address family view.

   address-family ipv4 [**unicast**]

5. Enable BGP to exchange IPv4 unicast routes with the specified peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**

   By default, BGP does not exchange IPv4 unicast routes with peers.

6. Redistribute the IGP routes of the VPN into BGP.

   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]

   By default, no routes are redistributed into BGP.

Configuring routing between an MCE and a PE

About routing between an MCE and a PE

MCE-PE routing configuration includes these tasks:

- Binding the MCE-PE interfaces to VPN instances.
- Performing route configurations.
- Redistributing VPN routes into the routing protocol running between the MCE and the PE.

Perform the following configurations on the MCE. For information about configuring the PE, see the documentation for the PE.

Configuring static routing between an MCE and a PE

1. Enter system view.

   system-view

2. Configure a static route for a VPN instance.

   ip route-static vpn-instance *s-vpn-instance-name dest-address* { *mask-length* | *mask* } { *interface-type interface-number* [*next-hop-address*] | *next-hop-address* [**public**] | **vpn-instance** *d-vpn-instance-name next-hop-address* }

3. (Optional.) Set the default preference for static routes.

ip route-static default-preference *default-preference*

The default preference is 60.

Configuring RIP between an MCE and a PE

1. Enter system view.

   system-view
2. Create a RIP process for a VPN instance and enter RIP view.

   rip [*process-id*] **vpn-instance** *vpn-instance-name*
3. Enable RIP on the interface attached to the specified network.

   network *network-address* [*wildcard-mask*]

   By default, RIP is disabled on an interface.
4. Redistribute the VPN routes.

   import-route *protocol* [*as-number*] [*process-id* | **all-processes** | **allow-ibgp**] [**allow-direct** | **cost** *cost-value* | **route-policy** *route-policy-name* | **tag** *tag*] *

   By default, no routes are redistributed into RIP.

Configuring OSPF between an MCE and a PE

1. Enter system view.

   system-view
2. Create an OSPF process for a VPN instance and enter OSPF view.

   ospf [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *
3. Disable routing loop detection.

   vpn-instance-capability simple

   By default, routing loop detection is enabled and the MCE does not receive OSPF routes from the PE.
4. Redistribute the VPN routes.

   import-route *protocol* [*as-number*] [*process-id* | **all-processes** | **allow-ibgp**] [**allow-direct** | **cost** *cost-value* | **nssa-only** | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   By default, no routes are redistributed into OSPF.
5. Create an OSPF area and enter OSPF area view.

   area *area-id*
6. Enable OSPF on the interface attached to the specified network in the area.

   network *ip-address wildcard-mask*

   By default, an interface neither belongs to an area nor runs OSPF.

Configuring IS-IS between an MCE and a PE

1. Enter system view.

   system-view
2. Create an IS-IS process for a VPN instance and enter IS-IS view.

   isis [*process-id*] **vpn-instance** *vpn-instance-name*

3. Configure a network entity title.

   network-entity *net*

   By default, no NET is configured.

4. Create the IS-IS IPv4 unicast address family and enter its view.

   address-family ipv4 [**unicast**]

5. Redistribute VPN routes.

   import-route *protocol* [*as-number*] [*process-id* | **all-processes** | **allow-ibgp**] [**allow-direct** | **cost** *cost-value* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *

   By default, IS-IS does not redistribute routes from any other routing protocol.

   If you do not specify the route level in the command, the command redistributes routes to the level-2 routing table by default.

6. Return to system view.

   quit

7. Enter interface view.

   interface *interface-type interface-number*

8. Enable the IS-IS process on the interface.

   isis enable [*process-id*]

   By default, no IS-IS process is enabled on the interface.

Configuring EBGP between an MCE and a PE

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is not enabled.

3. Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*

4. Configure the PE as an EBGP peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*

5. Enter BGP-VPN IPv4 unicast address family view.

   address-family ipv4 [**unicast**]

6. Enable BGP to exchange IPv4 unicast routes with the specified peer.

   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**

   By default, BGP does not exchange IPv4 unicast routes with peers.

7. Redistribute the VPN routes of the VPN site.

   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]

   By default, no routes are redistributed into BGP.

Configuring IBGP between an MCE and a PE

1. Enter system view.

system-view

2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
   By default, BGP is not enabled.
3. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
4. Configure the PE as an IBGP peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **as-number** *as-number*
5. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
6. Enable BGP to exchange IPv4 unicast routes with the specified peer.
   peer { *group-name* | *ipv4-address* [*mask-length*] } **enable**
   By default, BGP does not exchange IPv4 unicast routes with peers.
7. Redistribute the VPN routes of the VPN site.
   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]
   By default, no routes are redistributed into BGP.

Display and maintenance commands for MCE

Execute **display** commands in any view.

Task	Command
Display VPN instance information.	**display ip vpn-instance** [**instance-name** *vpn-instance-name*]

For commands that display routing tables for VPN instances, see *Layer 3—IP Routing Command Reference*.

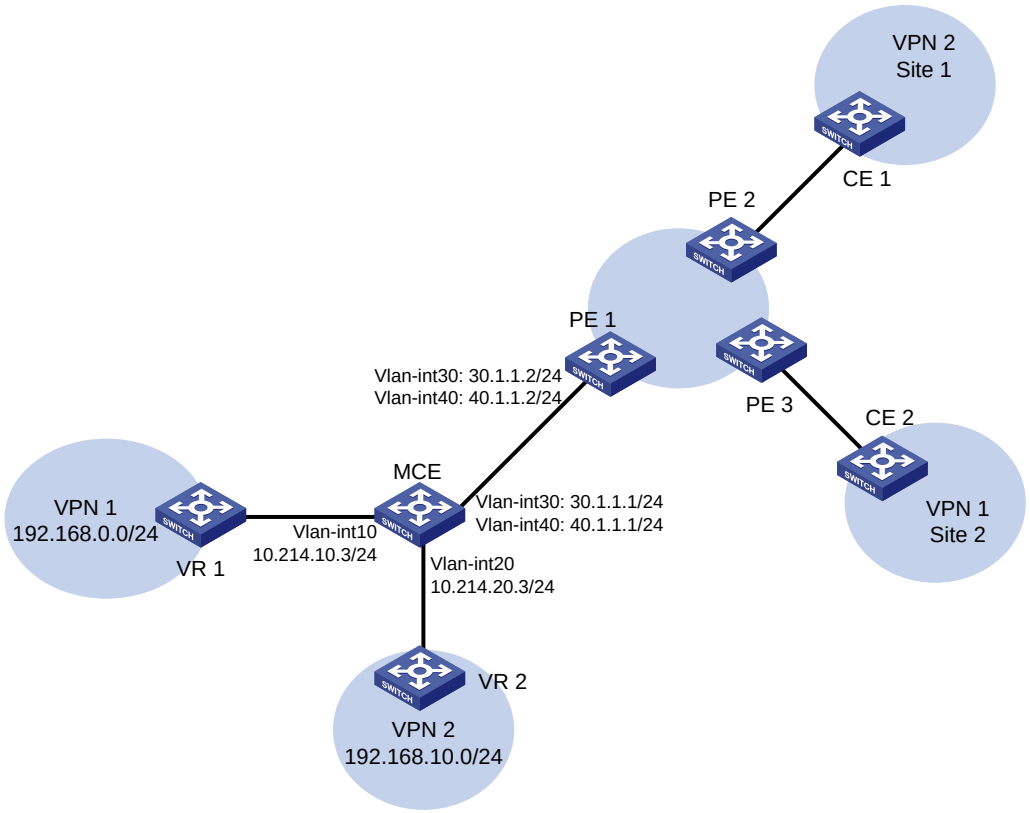
MCE configuration examples

Example: Configuring the MCE that uses OSPF to advertise VPN routes to the PE

Network configuration

As shown in Figure 4, VPN 2 runs OSPF. The edge devices of VPN 1 and VPN 2 are VR 1 and VR 2, respectively. Configure the MCE device to separate routes from different VPNs and to advertise the VPN routes to PE 1 through OSPF.

Figure 4 Network diagram



Procedure

1. Configure the VPN instances on the MCE and PE 1:

   # On the MCE, configure VPN instances **vpn1** and **vpn2**, and specify an RD and route targets for each VPN instance.

   ```
   <MCE> system-view
   [MCE] ip vpn-instance vpn1
   [MCE-vpn-instance-vpn1] route-distinguisher 10:1
   [MCE-vpn-instance-vpn1] vpn-target 10:1
   [MCE-vpn-instance-vpn1] quit
   [MCE] ip vpn-instance vpn2
   [MCE-vpn-instance-vpn2] route-distinguisher 20:1
   [MCE-vpn-instance-vpn2] vpn-target 20:1
   [MCE-vpn-instance-vpn2] quit
   ```

   # Bind VLAN-interface 10 to VPN instance **vpn1**, and configure an IP address for the VLAN interface.

   ```
   [MCE] interface vlan-interface 10
   [MCE-Vlan-interface10] ip binding vpn-instance vpn1
   [MCE-Vlan-interface10] ip address 10.214.10.3 24
   [MCE-Vlan-interface10] quit
   ```

   # Bind VLAN-interface 20 to VPN instance **vpn2**, and configure an IP address for the VLAN interface.

   ```
   [MCE] interface vlan-interface 20
   [MCE-Vlan-interface20] ip binding vpn-instance vpn2
   [MCE-Vlan-interface20] ip address 10.214.20.3 24
   ```

```
[MCE-Vlan-interface20] quit
```

On PE 1, configure VPN instances **vpn1** and **vpn2**, and specify an RD and route targets for each VPN instance.

```
<PE1> system-view
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 10:1
[PE1-vpn-instance-vpn1] vpn-target 10:1
[PE1-vpn-instance-vpn1] quit
[PE1] ip vpn-instance vpn2
[PE1-vpn-instance-vpn2] route-distinguisher 20:1
[PE1-vpn-instance-vpn2] vpn-target 20:1
[PE1-vpn-instance-vpn2] quit
```

2. Configure routing between the MCE and VPN sites:

The MCE is connected to VPN 1 directly, and no routing protocol is enabled in VPN 1. Therefore, you can configure static routes.

On VR 1, assign IP address 10.214.10.2/24 to the interface connected to MCE and 192.168.0.1/24 to the interface connected to VPN 1. Add ports to VLANs correctly. (Details not shown.)

On VR 1, configure a default route with the next hop being 10.214.10.3.

```
<VR1> system-view
[VR1] ip route-static 0.0.0.0 0.0.0.0 10.214.10.3
```

On the MCE, configure a static route to 192.168.0.0/24 with the next hop 10.214.10.2. Bind the static route to VPN instance **vpn1**.

```
[MCE] ip route-static vpn-instance vpn1 192.168.0.0 24 10.214.10.2
```

On the MCE, display the routing information maintained for VPN instance **vpn1**.

```
[MCE] display ip routing-table vpn-instance vpn1

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
10.214.10.0/24     Direct  0   0           10.214.10.3     Vlan10
10.214.10.0/32     Direct  0   0           10.214.10.3     Vlan10
10.214.10.3/32     Direct  0   0           127.0.0.1       InLoop0
10.214.10.255/32   Direct  0   0           10.214.10.3     Vlan10
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
192.168.0.0/24     Static  60  0           10.214.10.2     Vlan10
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

The output shows that the MCE has a static route for VPN instance **vpn1**.

Run OSPF in VPN 2. Create OSPF process 20 and bind it to VPN instance **vpn2** on the MCE, so that the MCE can learn the routes of VPN 2 and add them to the routing table of VPN instance **vpn2**.

```
[MCE] ospf 2 vpn-instance vpn2
```

Advertise subnet 10.214.20.0.

```
[MCE-ospf-2] area 0
[MCE-ospf-2-area-0.0.0.0] network 10.214.20.0 0.0.0.255
[MCE-ospf-2-area-0.0.0.0] quit
[MCE-ospf-2] quit
```

On VR 2, assign IP address 10.214.20.2/24 to the interface connected to MCE and 192.168.10.1/24 to the interface connected to VPN 2. (Details not shown.)

Configure OSPF process 2, and advertise subnets 192.168.10.0 and 10.214.20.0.

```
<VR2> system-view
[VR2] ospf 2
[VR2-ospf-2] area 0
[VR2-ospf-2-area-0.0.0.0] network 192.168.10.0 0.0.0.255
[VR2-ospf-2-area-0.0.0.0] network 10.214.20.0 0.0.0.255
[VR2-ospf-2-area-0.0.0.0] quit
[VR2-ospf-2] quit
```

On the MCE, display the routing information maintained for VPN instance **vpn2**.

```
[MCE] display ip routing-table vpn-instance vpn2

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
10.214.20.0/24     Direct  0   0           10.214.20.3     Vlan20
10.214.20.0/32     Direct  0   0           10.214.20.3     Vlan20
10.214.20.3/32     Direct  0   0           127.0.0.1       InLoop0
10.214.20.255/32   Direct  0   0           10.214.20.3     Vlan20
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
192.168.10.0/24    O_INTRA 10  2           10.214.20.2     Vlan20
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

The output shows that the MCE has learned the private routes of VPN 2. The MCE maintains the routes of VPN 1 and those of VPN2 in two different routing tables. In this way, routes from different VPNs are separated.

3. Configure routing between the MCE and PE 1:

On the MCE, bind VLAN-interface 30 to VPN instance **vpn1**, and configure an IP address for the VLAN interface.

```
[MCE] interface vlan-interface 30
[MCE-Vlan-interface30] ip binding vpn-instance vpn1
[MCE-Vlan-interface30] ip address 30.1.1.1 24
[MCE-Vlan-interface30] quit
```

Bind VLAN-interface 40 to VPN instance **vpn2**, and configure an IP address for the VLAN interface.

```
[MCE] interface vlan-interface 40
[MCE-Vlan-interface40] ip binding vpn-instance vpn2
[MCE-Vlan-interface40] ip address 40.1.1.1 24
[MCE-Vlan-interface40] quit
```

On PE 1, bind VLAN-interface 30 to VPN instance **vpn1**, and configure an IP address for the VLAN interface.

```
[PE1] interface vlan-interface 30
[PE1-Vlan-interface30] ip binding vpn-instance vpn1
[PE1-Vlan-interface30] ip address 30.1.1.2 24
[PE1-Vlan-interface30] quit
```

Bind VLAN-interface 40 to VPN instance **vpn2**, and configure an IP address for the VLAN interface.

```
[PE1] interface vlan-interface 40
[PE1-Vlan-interface40] ip binding vpn-instance vpn2
[PE1-Vlan-interface40] ip address 40.1.1.2 24
[PE1-Vlan-interface40] quit
```

Configure the IP address of the interface Loopback 0 as 101.101.10.1 for the MCE and as 100.100.10.1 for PE 1. Specify the loopback interface address as the router ID for the MCE and PE 1. (Details not shown.)

Enable OSPF process 10 on the MCE, and bind the process to VPN instance **vpn1**.

```
[MCE] ospf 10 router-id 101.101.10.1 vpn-instance vpn1
```

Disable OSPF routing loop detection for the VPN instance.

```
[MCE-ospf-10] vpn-instance-capability simple
```

Set the domain ID to 10.

```
[MCE-ospf-10] domain-id 10
```

On the MCE, advertise subnet 30.1.1.0 in area 0, and redistribute the static route of VPN 1.

```
[MCE-ospf-10] area 0
[MCE-ospf-10-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[MCE-ospf-10-area-0.0.0.0] quit
[MCE-ospf-10] import-route static
[MCE-ospf-10] quit
```

On PE 1, enable OSPF process 10, and bind the process to VPN instance **vpn1**.

```
[PE1] ospf 10 router-id 100.100.10.1 vpn-instance vpn1
```

Set the domain ID to 10.

```
[PE1-ospf-10] domain-id 10
```

Advertise subnet 30.1.1.0 in area 0.

```
[PE1-ospf-10] area 0
[PE1-ospf-10-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[PE1-ospf-10-area-0.0.0.0] quit
[PE1-ospf-10] quit
```

Use similar procedures to configure OSPF process 20 between MCE and PE 1 and redistribute VPN 2's routing information. (Details not shown.)

Verifying the configuration

Verify that PE 1 has learned the static route of VPN 1 through OSPF.

```
[PE1] display ip routing-table vpn-instance vpn1

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
30.1.1.0/24        Direct  0   0           30.1.1.2        Vlan30
30.1.1.0/32        Direct  0   0           30.1.1.2        Vlan30
```

```
30.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
30.1.1.255/32      Direct  0   0           30.1.1.2        Vlan30
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
192.168.0.0/24     O_ASE2  150 1           30.1.1.1        Vlan30
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that PE 1 has learned the routes of OSPF process 20 in VPN 2 through OSPF.

```
[PE1] display ip routing-table vpn-instance vpn2

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
40.1.1.0/24        Direct  0   0           40.1.1.2        Vlan40
40.1.1.0/32        Direct  0   0           40.1.1.2        Vlan40
40.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
40.1.1.255/32      Direct  0   0           40.1.1.2        Vlan40
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
192.168.10.0/24    O_ASE2  150 1           40.1.1.1        Vlan40
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

The routing information for the two VPNs has been redistributed into the routing tables on PE 1.

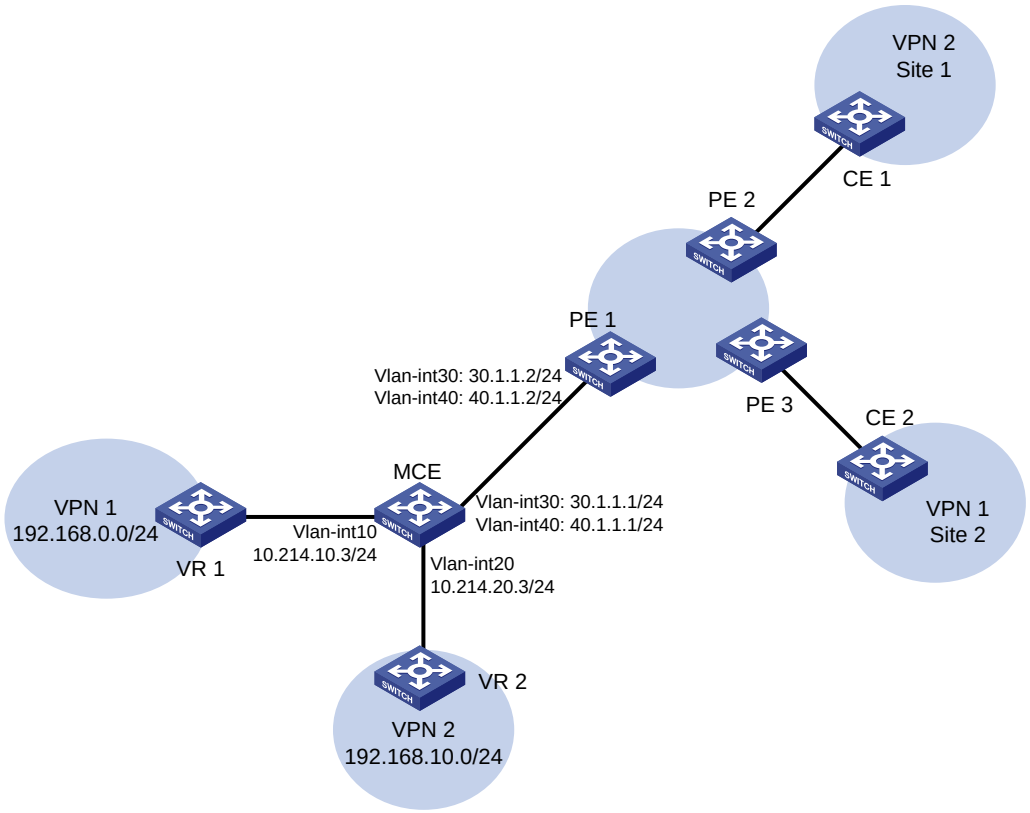
Example: Configuring the MCE that uses EBGP to advertise VPN routes to the PE

Network configuration

As shown in Figure 5, configure the MCE to advertise the routes of VPNs 1 and 2 to PE 1, so that the sites of each VPN can communicate with each other over the MPLS backbone.

Run OSPF in both VPN 1 and VPN 2. Run EBGP between the MCE and PE 1.

Figure 5 Network diagram



Procedure

1. Create VPN instances on the MCE and PE 1, and bind the VPN instances to VLAN interfaces. For the configuration procedure, see "Example: Configuring the MCE that uses OSPF to advertise VPN routes to the PE."
2. Configure routing between the MCE and VPN sites:

   # Enable an OSPF process on the devices in the two VPNs, and advertise the subnets. (Details not shown.)

   # Configure OSPF on the MCE, and bind OSPF process 10 to VPN instance **vpn1** to learn the routes of VPN 1.

```
<MCE> system-view
[MCE] ospf 10 router-id 10.10.10.1 vpn-instance vpn1
[MCE-ospf-10] area 0
[MCE-ospf-10-area-0.0.0.0] network 10.214.10.0 0.0.0.255
[MCE-ospf-10-area-0.0.0.0] quit
[MCE-ospf-10] quit
```

   # Display the routing table of VPN 1 on the MCE.

```
[MCE] display ip routing-table vpn-instance vpn1

Destinations : 13        Routes : 13

Destination/Mask    Proto   Pre Cost        NextHop         Interface
0.0.0.0/32          Direct  0   0           127.0.0.1       InLoop0
10.214.10.0/24      Direct  0   0           10.214.10.3     Vlan10
10.214.10.0/32      Direct  0   0           10.214.10.3     Vlan10
```

```
10.214.10.3/32      Direct  0   0           127.0.0.1       InLoop0
10.214.10.255/32    Direct  0   0           10.214.10.3     Vlan10
127.0.0.0/8         Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32        Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
192.168.0.0/24      O_INTRA 10  2           10.214.10.2     Vlan10
224.0.0.0/4         Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24        Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
```

The output shows that the MCE has learned the private route of VPN 1 through OSPF process 10.

On the MCE, bind OSPF process 20 to VPN instance **vpn2** to learn the routes of VPN 2. The configuration procedure is similar to that for OSPF process 10.

The output shows that the MCE has learned the private route of VPN 2 through OSPF.

```
[MCE] display ip routing-table vpn-instance vpn2

Destinations : 13        Routes : 13

Destination/Mask    Proto   Pre Cost        NextHop         Interface
0.0.0.0/32          Direct  0   0           127.0.0.1       InLoop0
10.214.20.0/24      Direct  0   0           10.214.20.3     Vlan20
10.214.20.0/32      Direct  0   0           10.214.20.3     Vlan20
10.214.20.3/32      Direct  0   0           127.0.0.1       InLoop0
10.214.20.255/32    Direct  0   0           10.214.20.3     Vlan20
127.0.0.0/8         Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32        Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32        Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
192.168.10.0/24     O_INTRA 10  2           10.214.20.2     Vlan20
224.0.0.0/4         Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24        Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32  Direct  0   0           127.0.0.1       InLoop0
```

3. Configure routing between the MCE and PE 1:

Configure the ports between the MCE and PE 1 as trunk ports. The configuration procedure is similar to that described in "Example: Configuring the MCE that uses OSPF to advertise VPN routes to the PE." (Details not shown.)

Enable BGP in AS 100 on the MCE.

```
[MCE] bgp 100
```

Enter the BGP-VPN instance view of VPN instance **vpn1**.

```
[MCE-bgp-default] ip vpn-instance vpn1
```

Specify the EBGP peer PE 1 in AS 200.

```
[MCE-bgp-default-vpn1] peer 30.1.1.2 as-number 200
```

Activate the EBGP VPNv4 peer PE 1, and redistribute routing information from OSPF process 10 to BGP.

```
[MCE-bgp-default-vpn1] address-family ipv4
[MCE-bgp-default-ipv4-vpn1] peer 30.1.1.2 enable
[MCE-bgp-default-ipv4-vpn1] import-route ospf 10
```

```
[MCE-bgp-default-ipv4-vpn1] quit
[MCE-bgp-default-vpn1] quit
[MCE-bgp-default] quit
```

On PE 1, enable BGP in AS 200, and specify the MCE as its EBGP peer.

```
[PE1] bgp 200
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 30.1.1.1 as-number 100
[PE1-bgp-default-vpn1] address-family ipv4
[PE1-bgp-default-ipv4-vpn1] peer 30.1.1.1 enable
[PE1-bgp-default-ipv4-vpn1] quit
[PE1-bgp-default-vpn1] quit
[PE1-bgp-default] quit
```

Use similar procedures to configure VPN 2 settings on MCE and PE 1. (Details not shown.)

Verifying the configuration

Verify that PE 1 has learned the OSPF route of VPN 1 through BGP.

```
[PE1] display ip routing-table vpn-instance vpn1

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
30.1.1.0/24        Direct  0   0           30.1.1.2        Vlan30
30.1.1.0/32        Direct  0   0           30.1.1.2        Vlan30
30.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
30.1.1.255/32      Direct  0   0           30.1.1.2        Vlan30
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
192.168.0.0/24     BGP     255 3           30.1.1.1        Vlan30
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

Verify that PE 1 has learned the OSPF route of VPN 2 through BGP.

```
[PE1] display ip routing-table vpn-instance vpn2

Destinations : 13        Routes : 13

Destination/Mask   Proto   Pre Cost        NextHop         Interface
0.0.0.0/32         Direct  0   0           127.0.0.1       InLoop0
40.1.1.0/24        Direct  0   0           40.1.1.2        Vlan40
40.1.1.0/32        Direct  0   0           40.1.1.2        Vlan40
40.1.1.2/32        Direct  0   0           127.0.0.1       InLoop0
40.1.1.255/32      Direct  0   0           40.1.1.2        Vlan40
127.0.0.0/8        Direct  0   0           127.0.0.1       InLoop0
127.0.0.0/32       Direct  0   0           127.0.0.1       InLoop0
127.0.0.1/32       Direct  0   0           127.0.0.1       InLoop0
```

```
127.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
192.168.10.0/24    BGP     255 3           40.1.1.1        Vlan40
224.0.0.0/4        Direct  0   0           0.0.0.0         NULL0
224.0.0.0/24       Direct  0   0           0.0.0.0         NULL0
255.255.255.255/32 Direct  0   0           127.0.0.1       InLoop0
```

The MCE has redistributed the OSPF routes of the two VPN instances into the EBGP routing tables of PE 1.

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Segment Routing Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the Segment Routing fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring SR-MPLS 1

About SR-MPLS 1
SR-MPLS characteristics 1
Basic concepts 1
How SR-MPLS works 1
Static segment configuration 2
Dynamic SID allocation through an IGP 2
Dynamic SID allocation through BGP 3
Label forwarding entry installation based on SIDs 3
SRLSP setup 5
Packet forwarding in SR-MPLS 5
MPLS SR and LDP interworking 8
SR to LDP 8
LDP to SR 9
SR over LDP 10
TI-LFA FRR 11
TI-LFA FRR background 11
TI-LFA FRR concepts 12
TI-LFA FRR path calculation 12
TI-LFA FRR forwarding process 13
Microloop avoidance after a network failure 14
Microloop avoidance after a failure recovery 15
Protocols and standards 16
Restrictions and guidelines: SR-MPLS configuration 16
SR-MPLS tasks at a glance 16
IP traffic forwarding over SRLSPs tasks at a glance 16
Configuring static segments 17
Prerequisites for static segment configuration 17
Configuring a static adjacency segment 17
Configuring a static prefix segment 18
Configuring IGP-based SID advertisement 18
Prerequisites for configuring IGP-based SID advertisement 18
Configuring the IGP to support SR-MPLS 19
Configuring prefix SIDs 19
Enabling SR-MPLS adjacency label allocation for the IGP 19
Assigning adjacency SIDs 20
Configuring the MPLS SRGB 21
Configuring BGP-based SID advertisement 21
BGP-based SID advertisement tasks at a glance 21
Prerequisites for configuring BGP-based SID advertisement 21
Configuring BGP to support SR-MPLS 21
Configuring BGP prefix SIDs 22
Configuring the device to prefer SRLSPs in traffic forwarding 22
Configuring SR-MPLS and LDP interworking 23
Restrictions and guidelines for SR-MPLS and LDP interworking 23
SR to LDP tasks at a glance 23
LDP to SR tasks at a glance 23
SR over LDP tasks at a glance 23
Prerequisites for SR-MPLS and LDP internetworking 24
Enabling advertisement of locally configured prefix-SID mappings 24
Configuring prefix-SID mappings 24
Enabling reception of advertised prefix-SID mappings 25
Configuring TI-LFA FRR 25
Restrictions and guidelines for TI-LFA FRR 25
TI-LFA FRR tasks at a glance 26
Enabling TI-LFA FRR 26
Disabling an interface from participating in TI-LFA calculation 26

Configuring microloop avoidance 26
Display and maintenance commands for SR-MPLS 28
SR-MPLS configuration examples 28
Example: Configuring SR-MPLS to LDP (IS-IS) 35
Example: Configuring SR-MPLS over LDP 38

Configuring SR-MPLS

About SR-MPLS

Segment Routing (SR) is a source routing technology. The source node selects a path for the packets, and then encodes the path in the packet header as an ordered list of segments. Each segment is identified by the segment identifier (SID). The SR nodes along the path forward the packets based on the SIDs in the packets. Only the source node needs to maintain the path status.

SR can operate with MPLS. In an MPLS network, SR uses MPLS labels as SIDs to forward packets on an LSP.

SR-MPLS characteristics

SR-MPLS has the following characteristics:

- SR-MPLS forwards packets based on the existing MPLS infrastructure. No infrastructure modifications are needed to implement SR on an MPLS network. For more information about the MPLS infrastructure, see MPLS basics configuration in *MPLS Configuration Guide*.
- SR-MPLS expands and optimizes existing IGPs and BGP, and uses the IGPs and BGP to distribute labels.

Basic concepts

- **SR node**—A node enabled with the SR-MPLS feature. The ingress node (source node) adds labels to packets. The transit nodes forward packets based on labels. The egress node removes labels and forwards packets to the destination networks. SR nodes form an SR domain.
- **Segment**—An instruction an SR node executes on the incoming packet.
- **SID**—Segment ID, which is MPLS label in SR-MPLS.
- **Segment type**—The following types of segments are available:
  - **Prefix segment**—SIDs are assigned to nodes based on destination address prefix. The nodes create prefix-specific forwarding entries.
  - **Adjacency segment**—SIDs are assigned to nodes based on adjacency.
- **SRLSP**—Segment routing label switched path, an LSP along which SR uses MPLS labels as SIDs to forward packets.
- **SRGB**—Segment routing global block, a range of global labels dedicated for SR-MPLS prefix SIDs. Different nodes can have different SRGBs. The minimum label value of an SRGB label range is referred to as the base value of the SRGB. Labels 16000 through 24000 are reserved for SRGBs.
- **SRLB**—Segment routing local block, a range of local labels dedicated for SR-MPLS adjacency SIDs. All nodes share the same SRLB. The minimum label value of the SRLB label range is referred to as the base value of the SRLB.

How SR-MPLS works

SR-MPLS involves the following steps:

1. Label allocation for all nodes and links along the packet forwarding paths.

The following methods are available:

- Static segment configuration
- Dynamic SID allocation

2. Label forwarding entry installation based on SIDs. All SR-MPLS devices in the SR domain use the allocated labels to create label forwarding entries.
3. SRLSP setup. You can manually configure SRLSPs, or use an IGP, BGP, or a controller to dynamically create SRLSPs.

When the ingress node of an SRLSP receives a packet, it adds labels to the packet and forwards the packet to the egress node through the SRLSP. After receiving a packet from the SRLSP, the egress node removes the label and forwards the packet based on the routing table.

Static segment configuration

Static SR-MPLS provides the following methods for configuring static segments for a destination:

- **Prefix segment method**—Manually configure the incoming label, outgoing label, and next hop for the destination address prefix on each SR node.
- **Adjacency segment method**—Manually configure the incoming label and next hop for the adjacency to the neighbor on each SR node.

Dynamic SID allocation through an IGP

SR-MPLS expands IGP protocols such as IS-IS to advertise SIDs in IGP protocol packets. SR-MPLS provides the following methods for dynamically allocating and advertising SIDs:

- **Prefix SID**—After you configure an SID for the loopback address on each SR node, the SIDs uniquely identify the SR nodes. All SR nodes in the SR domain use an IGP to advertise their own prefix SIDs. After receiving advertised prefix SIDs, each SR node calculates the prefix SIDs to the advertisers.

  The prefix SID advertisement can be one of the following types:

  - **Absolute value advertisement**—Each SR node advertises the prefix SID absolute value and the SRGB.
  - **Index value advertisement**—Each SR node advertises the prefix SID index and the SRGB.

    Each SR node is assigned a globally unique index value for the node's prefix. The prefix SID an SR node allocates to a prefix equals the SRGB base of the SR node plus the index for that prefix.

NOTE:

The device supports only the index value advertisement in the current software release.

- **Adjacency SID**—SR nodes use an IGP to advertise SIDs allocated to the IGP adjacencies (that is, the links to its IGP neighbors). The SIDs are used to identify the links.

  The adjacency SIDs can be allocated automatically or manually.

  - **Automatic adjacency SID allocation**—SR nodes allocate labels in SRLBs to the IGP adjacencies as SIDs.
  - **Manual adjacency SID allocation**—You can assign adjacency SID by using absolute values or index values. If you use index values, the adjacency SID of a link is the base value of the SRLB plus the index value for the link.

Dynamic SID allocation through BGP

SR-MPLS expands BGP to advertise prefix SIDs in BGP protocol packets. BGP uses routing policies to assign SIDs to prefixes when redistributing routes. The SIDs uniquely identify the SR nodes. All SR nodes in the SR domain use BGP to advertise their own prefix SIDs, including the SIDs and SID index values. From received advertisement packets, each SR node derives the prefix SIDs for the advertisers.

BGP prefix SIDs can be exchanged between BGP peers only through IPv4 unicast routes.

Dynamic SID allocation through BGP applies to MPLS VPN inter-AS option C networks.

Label forwarding entry installation based on SIDs

Label forwarding entry installation based on IGP prefix SIDs

Label forwarding entries based on prefix SIDs can be static or dynamic.

- **Static label forwarding entries**—The device creates local label forwarding entries based on manually configured incoming labels, outgoing labels, and next hops.
- **Dynamic label forwarding entries**—The device uses the IGP to flood in the SR domain the local SRGB and the index of the prefix SID for the local loopback interface. Based on the flooded information, the other devices in the domain calculates their local label forwarding entries by using the following rules:
  - Incoming label = Local SRGB base value + Index
  - Outgoing label = SRGB base value of the next hop + Index

Figure 1 Creating label forwarding entries based on IGP prefix SIDs

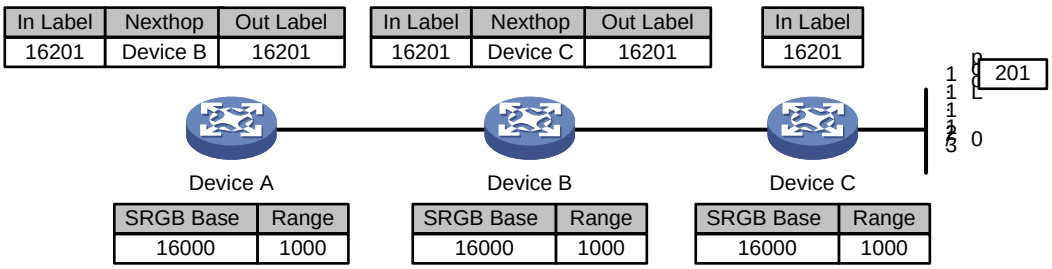


Figure 1 shows how dynamic label forwarding entries are created. After you assign index value 201 to loopback address 1.1.1.1/32 on Device C, Device C uses an IGP packet to advertise the index value and its local SRGB. Then, the devices calculate incoming and outgoing labels according to the previously mentioned label calculation rules.

- Devices C calculates its incoming label, which is 16201.
- Device B calculates its incoming label and outgoing label and creates a label forwarding entry. The incoming label is 16201. The outgoing label is 16201. The next hop is Device C.
- Device A calculates its incoming label and outgoing label and creates a label forwarding entry. The incoming label is 16201. The outgoing label is 16201. The next hop is Device B.

Label forwarding entry installation based on BGP prefix SIDs

The device redistributes routes to BGP and assigns prefix SIDs to redistributed routes based on routing policies. Then, the device uses BGP to flood in the SR domain the prefix SIDs and the indexes of the prefix SIDs. Based on the flooded information, the other devices in the domain calculates their local label forwarding entries by using the following rules:

- Incoming label = Local SRGB base value + Index
- Outgoing label = Received SID

Figure 2 Creating label forwarding entries based on BGP prefix SIDs

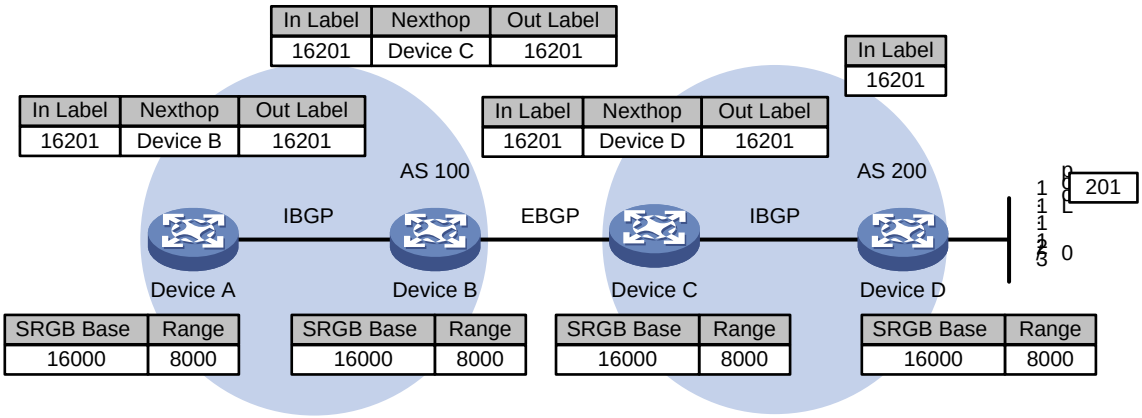


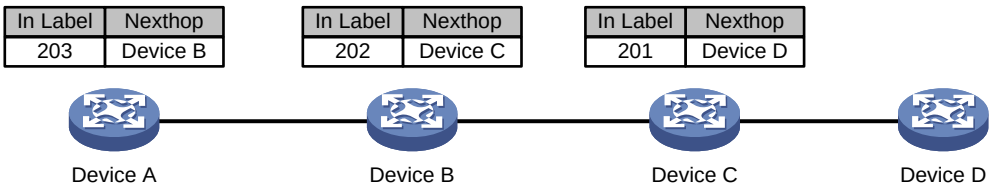
Figure 2 shows how label forwarding entries are created. After you assign index value 201 to loopback address 1.1.1.1/32 on Device D, Device D redistributes the route of the address and assigns it an SID (16201 in this example). Then, the device uses a BGP packet to advertise the SID and index value. The devices in the SR domain calculate incoming and outgoing labels according to the previously mentioned label calculation rules and create label forwarding entries.

- Devices D calculates its incoming label, which is 16201.
- Device C calculates its incoming label and outgoing label and creates a label forwarding entry. The incoming label is 16201. The outgoing label is 16201. The next hop is Device D.
- Device B calculates its incoming label and outgoing label and creates a label forwarding entry. The incoming label is 16201. The outgoing label is 16201. The next hop is Device C.
- Device A calculates its incoming label and outgoing label and creates a label forwarding entry. The incoming label is 16201. The outgoing label is 16201. The next hop is Device B.

Label forwarding entry installation based on adjacency SIDs

When using adjacency SIDs, each device allocates a static or dynamic incoming label to the link to its neighbor. The label has local significance. Multiple devices can use the same adjacency SID.

Figure 3 Creating label forwarding entries based on adjacency SIDs



As shown in Figure 3, the devices are running the same IGP. After IGP adjacencies are established between the devices, Device A, Device B, and Device C allocates labels and creates label forwarding entries as follows:

- Device A allocates label 203 to the link to its neighbor Device B.
- Device B allocates label 202 to the link to its neighbor Device C.
- Device C allocates label 201 to the link to its neighbor Device D.
- Device A creates a label forwarding entry with incoming label 203 and next hop Device B.
- Device B creates a label forwarding entry with incoming label 202 and next hop Device C.
- Device C creates a label forwarding entry with incoming label 201 and next hop Device D.

SRLSP setup

You can use the following methods to create SRLSPs:

- Manually configure an SRLSP.

  To configure an SRLSP, you must specify the label stack for packets to be forwarded along the SRLSP on the ingress node. Each label in the stack corresponds to a prefix SID or adjacency SID. The system can look for the outgoing label and next hop based on the prefix SID or adjacency SID.
- Configure SR nodes to use BGP or an IGP to dynamically establish an SRLSP.

  SR nodes follow these steps to establish SRLSPs:

  a. Use BGP or an IGP to collect prefix SID information from the SR-MPLS network.

  b. Calculate the shortest paths to other SR nodes based on the collected prefix SID information and the BGP or IGP network topology.

  c. Establish SRLSPs based on the shortest paths.
- Configure a controller to deploy SRLSP configuration to the device so the device creates SRLSPs accordingly.

  For more information about controller configuration, see the user guide for the controller.

Packet forwarding in SR-MPLS

Based on the SID allocation method, SR-MPLS uses one of the following packet forwarding methods:

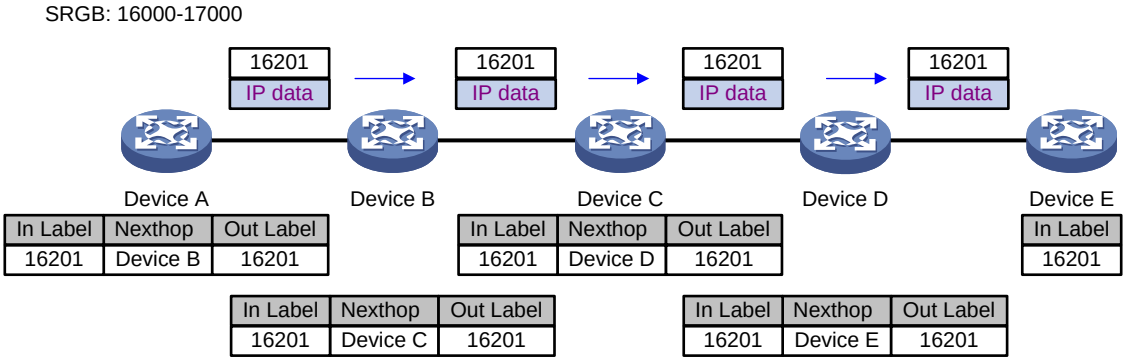
- **Prefix SID-based packet forwarding**—The ingress node encapsulates the prefix SID for the egress node to a packet. The transit nodes forward the packet based on label forwarding entries.
- **Adjacency SID-based packet forwarding**—The ingress node encapsulates the label stack that contains the adjacency SIDs of all links along the forwarding path to a packet. Each transit node uses the top label in the label stack to determine the next hop and pops the top label before forwarding the packet to the next hop.
- **Prefix and adjacency SID-based packet forwarding**—The nodes use prefix SID-based packet forwarding in combination with adjacency-based packet forwarding.

Prefix SID-based packet forwarding within the same AS

Figure 4 shows how a packet is forwarded along the SRLSP from Device A to Device E based on prefix SIDs. In this example, the outgoing label for the packet is 16201 on Device A.

1. Ingress node Device A searches for a forwarding entry for label 16201, adds outgoing label 16201 to the packet and sends the packet to the next hop (Device B).
2. When transit node Device B receives the packet, it searches for a label forwarding entry that matches the label in the packet. Then, Device B uses the outgoing label of the matched entry (16201) to replace the label in the packet and forwards the packet to the next hop (Device C).
3. Transit nodes Device C and Device D process the packet in the same way Device B does.
4. When egress node Device E receives the packet, it removes the label and forwards the packet by IP address.

Figure 4 Prefix SID-based packet forwarding within the same AS

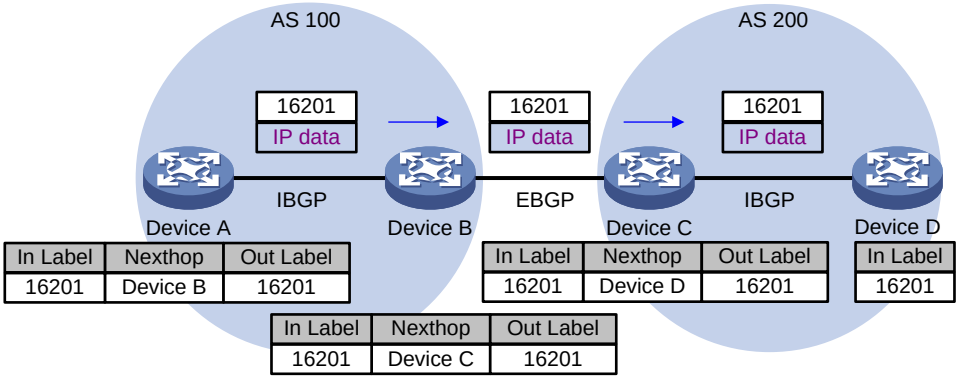


Prefix SID-based packet forwarding across ASs

Figure 5 shows how a packet is forwarded along the SRLSP from Device A to Device D based on prefix SIDs. In this example, the outgoing label for the packet is 16201 on Device A.

1. Ingress node Device A searches for a forwarding entry for label 16201, adds outgoing label 16201 to the packet, and sends the packet to the next hop (Device B).
2. When transit node Device B receives the packet, it searches for a label forwarding entry that matches the label in the packet (16201). Then, Device B uses the outgoing label of the matched entry (16201) to replace the label in the packet and forwards the packet to the next hop (Device C).
3. Transit nodes Device C processes the packet in the same way Device B does.
4. When egress node Device D receives the packet, it removes the label and forwards the packet by IP address.

Figure 5 Prefix SID-based packet forwarding across ASs



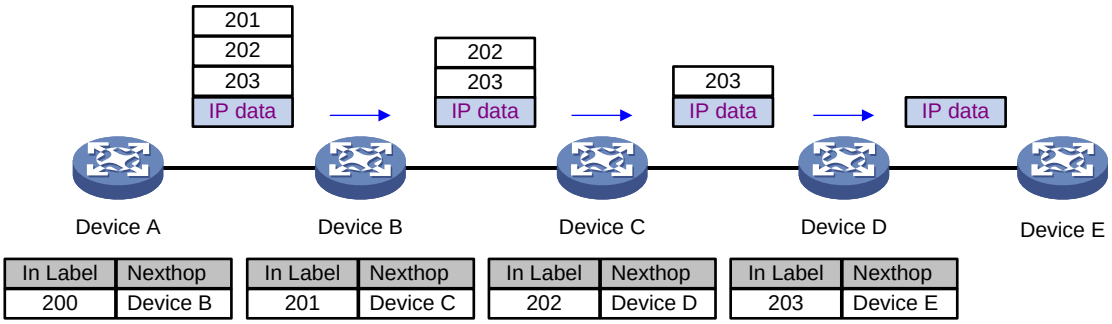
Adjacency SID-based packet forwarding

Figure 6 shows how a packet is forwarded along the SRLSP from Device A to Device E based on adjacency SIDs. In this example, the label stack for the packet is configured as (200, 201, 202, and 203) on Device A.

1. Ingress node Device A searches for a forwarding entry for the top label (200) to determine the next hop. Then, Device A adds label stack (201, 202, and 203) to the packet, and forwards the packet to the next hop (Device B).
2. When transit node Device B receives the packet, it searches for a forwarding entry for the top label (201) to determine the next hop. Then, Device B removes the top label from the stack and forwards the packet to the next hop (Device C).

3. When transit node Device C receives the packet, it searches for a forwarding entry for the top label (202) to determine the next hop. Then, Device C removes the top label from the stack and forwards the packet to the next hop (Device D).
4. When transit node Device D receives the packet, it searches for a forwarding entry for the label (203) to determine the next hop. Then, Device D removes the label stack from the packet and forwards the packet to the next hop (Device E).
5. When egress node Device E receives the packet, it forwards the packet by IP address.

Figure 6 Adjacency SID-based packet forwarding

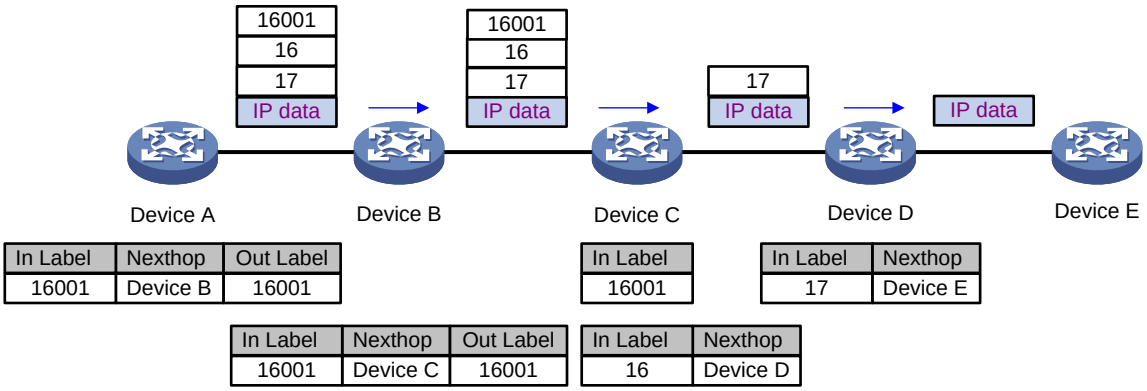


Prefix and adjacency SID-based packet forwarding

Figure 7 shows how a packet is forwarded along the SRLSP from Device A to Device E based on prefix SIDs and adjacency SIDs. In this example, the index value for the prefix SID of Device C is 1. The prefix SIDs for Device A, Device B, and Device C are all 16001. The Adjacency SIDs that Device C and Device D allocate to the adjacencies are 16 and 17, respectively. On Device A, the label stack for the packet is (16001, 16, 17).

1. Ingress node Device A searches for a forwarding entry for label 16001 to determine the outgoing label (16001) and next hop (Device B). Device adds label stack (16001, 16, 17) to the packet and sends the packet to the next hop (Device B).
2. When transit node Device B receives the packet, it searches for a label forwarding entry that matches the top label in the label stack (16001). Then, Device B uses the outgoing label of the matched entry (16001) to replace the top label and forwards the packet to the next hop (Device C).
3. When transit node Device C receives the packet, it removes the top label 16001 and searches for a forwarding entry for the next label (16) to determine the next hop. Then, Device C removes label 16 from the stack and forwards the packet to the next hop (Device D).
4. When transit node Device D receives the packet, it searches for a forwarding entry for the label (17) to determine the next hop. Then, Device D removes the label stack from the packet and forwards the packet to the next hop (Device E).
5. When egress node Device E receives the packet, it forwards the packet by IP address.

Figure 7 Prefix and adjacency SID-based packet forwarding



MPLS SR and LDP interworking

SR-MPLS and LDP interworking enables SR-MPLS networks and MPLS LDP networks to communicate and cooperate for MPLS forwarding.

SR-MPLS and LDP interworking supports the following modes:

- **SR to LDP**—Maps prefixes on an LDP network to SR-MPLS SIDs to forward traffic from an SR-MPLS network to an MPLS LDP network.
- **LDP to SR**—Uses an IGP protocol to advertise SIDs and associates SIDs with LDP labels to forward traffic from an MPLS LDP network to an SR-MPLS network.
- **SR over LDP**—Enables SR-MPLS networks to communicate through an LDP network.

SR to LDP

The SR to LDP mode defines the following roles:

- **Segment Routing Mapping Server (SRMS)**—Advertises prefix-SID mappings on behalf of SR-incapable devices (LDP devices). You need to configure prefix-SID mappings on the SRMS.
- **Segment Routing Mapping Client (SRMC)**—Receives prefix-SID mappings advertised by the SRMS and creates SR-MPLS label forwarding entries.

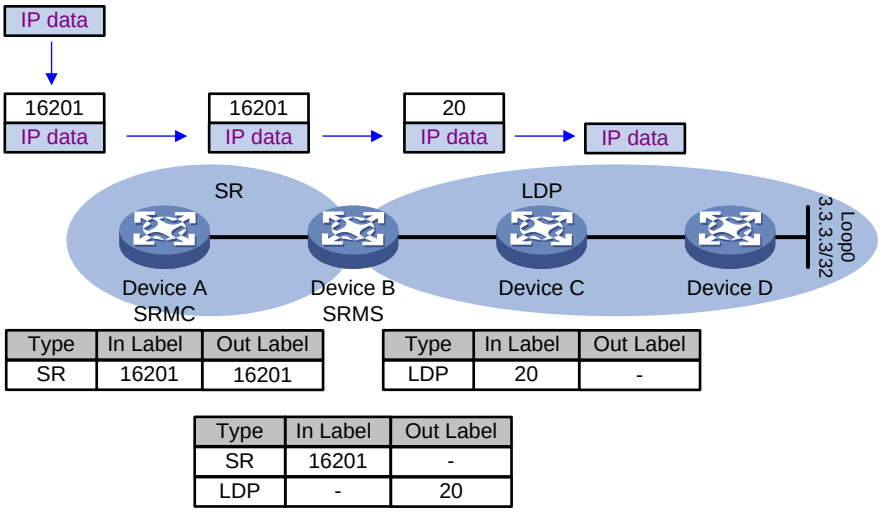
As shown in Figure 8, Device A and Device B are running SR-MPLS, and Device B, Device C, and Device D are running LDP. After Device D assigns a label for destination address 3.3.3.3/32, Device B and Device C each install an LDP label forwarding entry.

The following steps establish a mapping between the SRLSP and the LDP LSP:

1. Device B acts as an SRMS. It assigns SID index value 201 to the IP address of loopback interface 0 on Device D (3.3.3.3/32), and then advertises an IP prefix-SID mapping TLV to Device A.
2. Device A acts as an SRMC. After receiving the advertised mapping TLV, it creates an SR-MPLS label forwarding entry.

A mapping is established between the SRLSP and the LDP LSP.

Figure 8 SR to LDP packet forwarding



If Device A uses an SRLSP to send a packet to Device D, the packet is forwarded as follows:

3. Ingress node Device A adds label 16201 to the packet and then sends the packet to transit node Device B.
4. When transit node Device B receives the packet, it searches for a label forwarding entry that matches incoming label 16201. Because the entry does not have an outgoing label, Device B searches for a valid LDP outgoing label for destination address 3.3.3.3/32. Then, Device B adds the LDP label (20) to the packet as the SR outgoing label and forwards the packet to the next hop (Device C).
5. When transit node Device C receives the packet, it searches for an LDP label forwarding entry for incoming label 20. Then, it removes the label from the packet and forwards the packet to egress node Device D.
6. The egress node forwards the packet by IP address.

LDP to SR

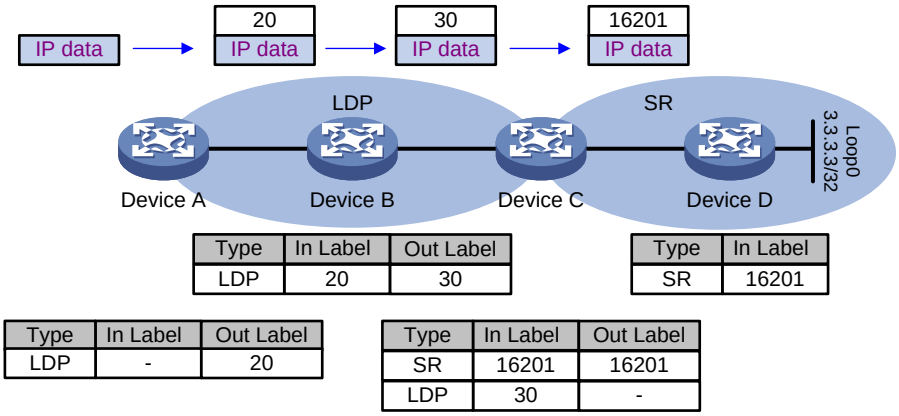
As shown in Figure 9, Device C and Device D are running SR-MPLS, and Device A, Device B, and Device C are running LDP. Each device running LDP assigns a label to destination address 3.3.3.3/32 and then installs an LDP label forwarding entry.

The following steps establish an association between the SR-MPLS label and the LDP label:

1. After you assign SID index value 201 to the IP address of loopback interface 0 on Device D (3.3.3.3/32), Device D sends an IGP protocol packet to advertise the index value and the local SRGB.
2. After receiving the packet, Device C creates an SR-MPLS label forwarding entry.

The SR-MPLS label and the LDP label are associated on Device C.

Figure 9 LDP to SR packet forwarding



A packet from Device A to Device D is forwarded as follows:

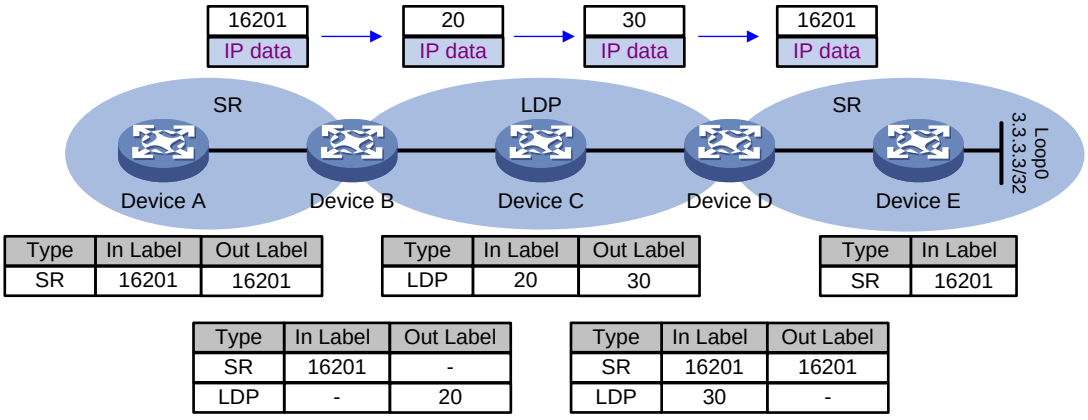
3. Ingress node Device A adds an LDP label (20) to the packet and forwards the packets to the next hop (Device B).
4. Device B searches for an LDP label forwarding entry, uses 30 as the outgoing label, and forwards the packet to the next hop (Device C).
5. Device C searches for an LDP label forwarding entry for incoming label 30 and finds no outgoing label. Because valid SR-MPLS outgoing label 16201 exists for the destination address 3.3.3.3.32, Device C uses the label (16201) as the outgoing label and forwards the packet to Device D.
6. Egress node Device D deletes label 16201 from the packet and forwards the packet by IP address.

SR over LDP

For SR-MPLS networks to communicate across an LDP network, the SR/LDP border devices must be able to connect the SR LSP and the LDP LSP as follows:

- **SR-MPLS to LDP interworking**—The border device installs SR-to-LDP label forwarding entries. For a packet from the SR-MPLS network to the LDP network, the SR-MPLS label forwarding entry on the border device does not have an outgoing label. The border device uses the outgoing label of the LDP label forwarding entry for the same destination address as the outgoing label of the packet.
- **LDP to SR-MPLS interworking**—The border device installs LDP-to-SR label forwarding entries. For a packet from the LDP network to the SR-MPLS network, the LDP forwarding entry on the border device does not have an outgoing label. The border device must use the outgoing label of the SR-MPLS forwarding entry for the same destination address as the outgoing label of the packet.

Figure 10 SR-MPLS over LDP



As shown in Figure 10, Device A, Device B, Device D, and Device E are running SR-MPLS. After you assign index value 201 to loopback interface address 3.3.3.3/32 on Device E, Device E will advertise the index value and the local SRGB. After Device A, Device B, and Device D receive the message, they will install their respective SR-MPLS label forwarding entries. Device B, Device C, and Device D are running LDP. They allocate labels to destination address 3.3.3.3/32 to generate the LDP label forwarding entries.

A packet that Device A sends to Device E will be forwarded as follows:

1. Ingress node Device A encapsulates label 16201 to the packet and forwards the packet to transit node Device B.
2. Transit node Device B searches for an SR-MPLS label forwarding entry for incoming label 16201 and finds that the entry does not have an outgoing label. Because an LDP label forwarding entry with outgoing label 20 exists for the destination address (3.3.3.3/32), Device B encapsulates outgoing label 20 in the packet. Then Device B sends the packet to transit node Device C.
3. Device C forwards the packet to Device D based on its LDP label forwarding entries. The outgoing label is 30.
4. Device D searches for an LDP label forwarding entry for incoming label 30 and finds that the entry does not have an outgoing label. Because an SR-MPLS label forwarding entry with outgoing label 16201 exists for the destination address (3.3.3.3/32), Device B encapsulates outgoing label 16201 in the packet. Then Device D sends the packet to egress node Device E.
5. Egress node Device E deletes the incoming label and forwards the packet by IP address.

TI-LFA FRR

Topology-Independent Loop-Free Alternate Fast Re-Route (TI-LFA FRR) provides link and node protection. When a link or node fails, TI-LFA FRR switches the traffic to the backup path to ensure continuous data forwarding.

TI-LFA FRR background

To minimize traffic loss during the route reconvergence process in SR-MPLS, you can enable the FRR feature on the device directly connected to the protected link or node. The device enabled with FRR is called the Point of Local Repair (PLR). The PLR calculates the shortest path to the destination and calculates an FRR backup path at the same time, and then writes the information into the FIB table. When a protected link or node fails, traffic is rerouted through the FRR backup path on the PLR node, without the need for the network topology to reconverge, significantly reducing traffic loss. FRR has the following mechanisms:

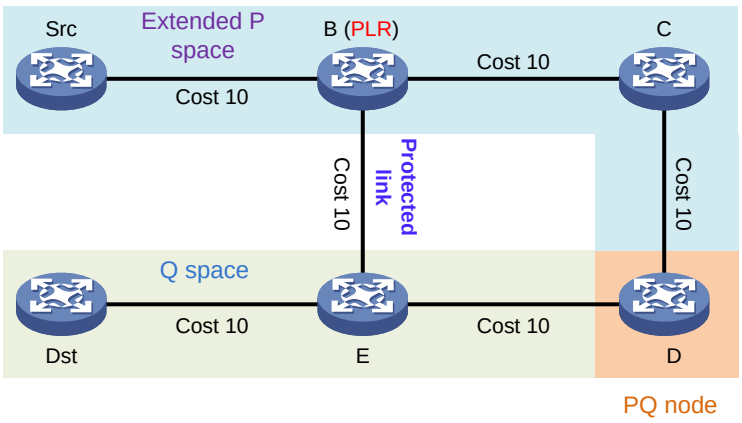
- **Loop-Free Alternate Fast Reroute (LFA FRR)**—To calculate the backup path, LFA FRR identifies a protected neighboring node (LFA node) of the PLR, enabling traffic to be forwarded to the destination node without passing through the protected link or node. In some scenarios, especially in a ring network, LFA FRR cannot calculate a backup path, making it topology dependent. According to RFC 6571, LFA FRR has a topology coverage of 80% to 90%.
- **Remote Loop-Free Alternate Fast Reroute (RLFA FRR)**—To improve the topology coverage of LFA FRR, RFC 7490 defines RLFA FRR, which enables traffic to be forwarded from the PLR to an RLFA FRR node and reach the destination node without passing through the protected link or node. Compared to LFA FRR, RLFA FRR does not restrict the protective node to being a neighbor of the PLR, providing more protection possibilities and increasing the topology coverage to 95% to 99%.
- **TI-LFA FRR suitable for SRv6 and SR-MPLS**—Compared to LFA FRR and RLFA FRR, TI-LFA FRR is topology independent, meaning FRR backup path calculation is not restricted by the network topology. PLR can automatically calculate a TI-LFA FRR backup path as long as a bypass forwarding path is available.

TI-LFA FRR concepts

TI-LFA FRR uses the concepts of RLFA FRR defined in RFC 7490:

- **P space**—A set of nodes reachable (using pre-convergence paths) from the PLR without using the protected link or node (including equal-cost path splits). Nodes in the P space are called P nodes.Calculation of P nodes typically involves building an SPF tree with the PLR as the root node, and then identifying nodes on the SPF tree that meet the loop-free requirement.
- **Extended P space**—A set of nodes reachable (using pre-convergence paths) from the neighbors of the PLR (except for the protected node) without using the protected link or node (including equal-cost path splits). The P space is a subset of the extended P space. Nodes in the extended P space are also called P nodes. Neighbor nodes of the PLR are N nodes. As shown in Figure 11, the expanded P space contains nodes Src, B, C, and D. The P nodes meet the following loop-free requirement: Distance (N, P) < Distance (N, PLR) + Distance (PLR, P).
- **Q space**—A set of nodes that can reach (using pre-convergence paths) the destination without using the protected link or node (including equal-cost path splits). Nodes in the Q space are called Q nodes.

Figure 11 TI-LFA FRR concepts



TI-LFA FRR path calculation

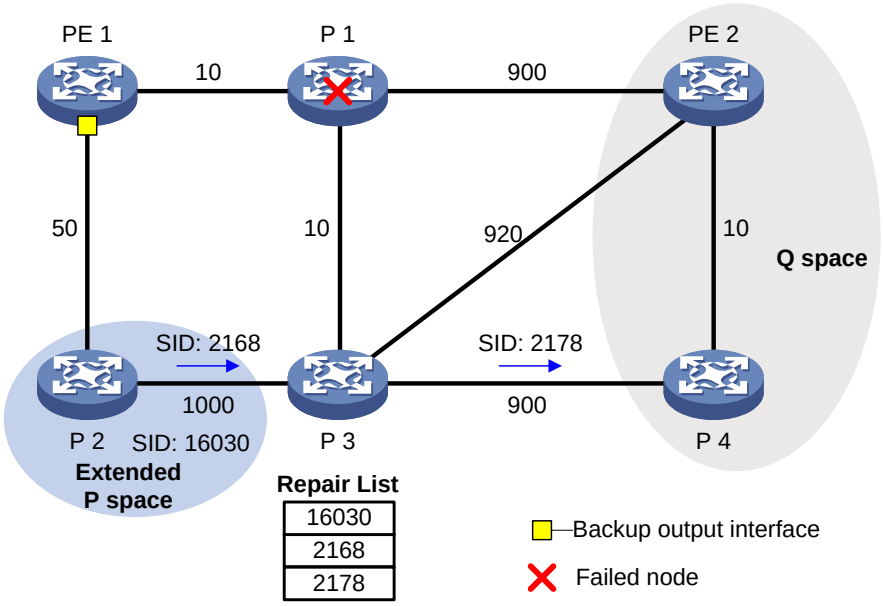
As shown in Figure 12, PE 1 is the source node. P 1 is the faulty node. PE 2 is the destination node. The numbers on links represent the link costs. A data flow traverses PE 1, P 1, and PE 2. To protect data against P 1 failure, TI-LFA FRR calculates the extended P space, Q space, shortest path tree

converged after P 1 fails, repair list, and backup outgoing interface, and creates the backup forwarding entry.

TI-LFA FRR calculates the backup path by using the following steps:

1. Calculates the extended P space: P 2.
2. Calculates the Q space: PE 2 and P 4.
3. Calculates the shortest path tree converged after P 1 fails: PE 1 --> P 2 --> P 4 --> PE 2.
4. Calculates the repair list: Node label of P 2 (16030), adjacency label of P 2 to P 3 (2168), and adjacency label of P 3 to P 4 (2178).
5. Calculates the backup outgoing interface, that is, the outgoing interface to the next hop after the link from PE 1 to P 1 fails.

Figure 12 TI-LFA FRR diagram



TI-LFA FRR forwarding process

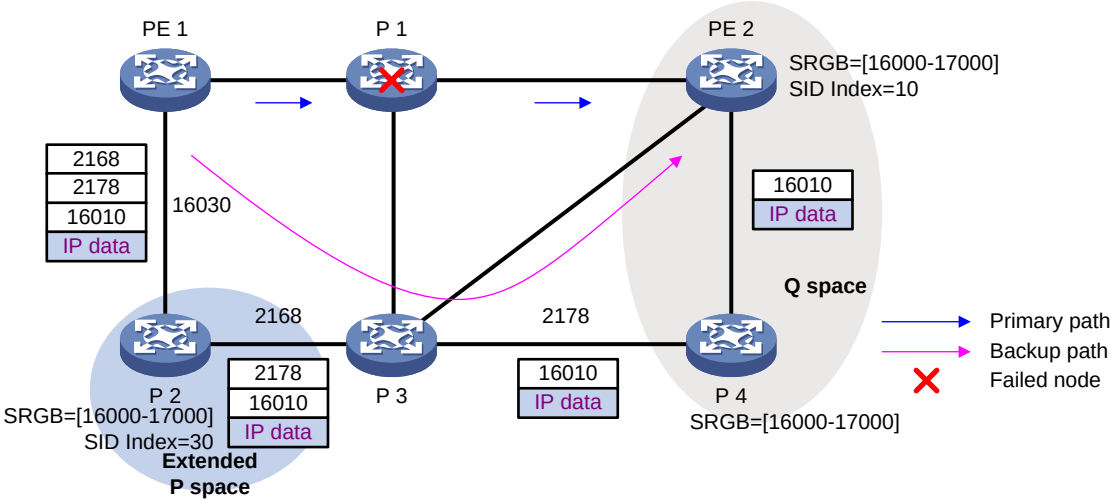
After TI-LFA FRR finishes backup path calculation, traffic will be switched to the backup path in response to a primary path failure.

As shown in Figure 13, P 2 is a P node and P 4 is a Q node. When the next hop on the primary path (P 1) fails, TI-LFA FRR switches the traffic to the backup path. The following are the detailed steps:

1. PE 1 encapsulates a label stack to a packet according to the repair list. The labels, from the outmost to inmost, are as follows:
   - Node label of P node P2 (16030), which equals the SRGB base value of PE 1 plus the SID index value of P 2.
   - Adjacency labels from P node P2 to Q node P 4, which are 2168 and 2178.
   - The destination's node label 16010, which equals the SRGB base value of Q node P 4 plus the SID index value of destination node PE 2.
2. P2 receives the packet, searches for a label forwarding entry based on the outmost label, pops label 2168, and forwards the packet to P 3.
3. P3 receives the packet, searches for a label forwarding entry based on the outmost label, pops label 2178, and forwards the packet to P 4.

4. P4 receives the packet, and searches for a label forwarding entry based on the outmost label. Because the outgoing label is 16010 and the next hop is PE 2, P 4 encapsulates 16010 as the outmost label and forwards the packet to PE 2.

Figure 13 Data forwarding over the TI-LFA FRR backup path

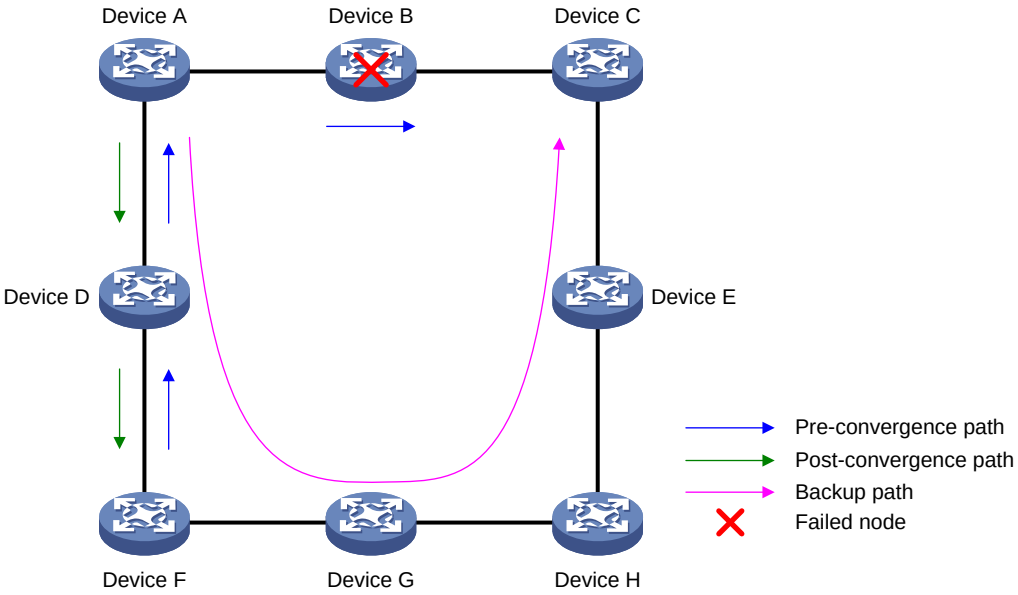


Microloop avoidance after a network failure

As shown in Figure 14, when Device B fails, traffic to Device C will be switched to the backup path calculated by TI-LFA. After Device A finishes route convergence, traffic will be switched to the post-convergence path. If Device D and Device F have not finished route convergence and still forward traffic along the pre-convergence path, a loop is formed between Device A and Device F. The loop exists until Device D and Device F finish route convergence.

FRR microloop avoidance and SR microloop avoidance can resolve this issue. After you configure TI-LFA, Device A first switches traffic to the backup path calculated by TI-LFA when Device B fails. Then, Device A waits for Device D and Device F to finish route convergence before starting route convergence. After Device A also finishes route convergence, Device A switches the traffic to the converged route.

Figure 14 Diagram for microloop avoidance after a network failure



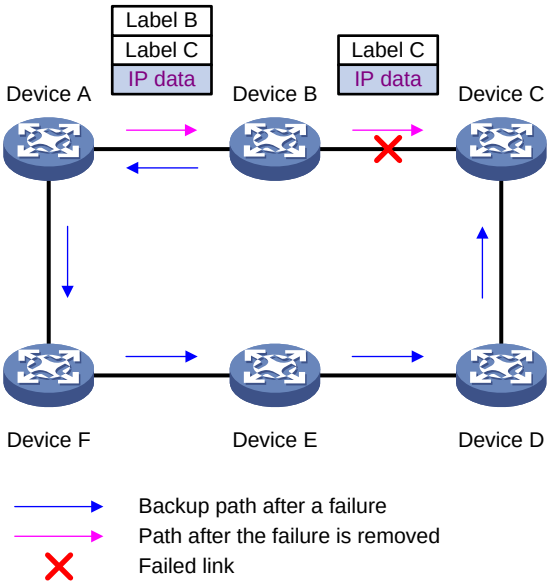
Microloop avoidance after a failure recovery

As shown in Figure 15, before the link between Device B and Device C recovers, traffic traverses along the backup path. After the link recovers, Device A will forward the traffic to Device B if Device A finishes route convergence before Device B. Before Device B also finishes route convergence, Device B still forwards the traffic along the backup path. A loop is formed between Device A and Device B.

SR microloop avoidance can resolve this issue. After the link recovers, SR microloop avoidance automatically calculates the optimal path from Device A to Device C and forwards traffic along the path. To forward a packet along the newly calculated path, Device A adds, for example, the adjacency label from Device B to Device C to the packet and then sends the packet to Device B. Then, Device B forwards the packet to Device C based on the path information.

When the microloop avoidance RIB-update-delay timer expires, Device B should have finished route convergence. Device A does not add path information to packets anymore, and it forwards packets to Device C as usual.

Figure 15 Diagram for microloop avoidance after a failure recovery



Protocols and standards

- *draft-bashandy-rtgwg-segment-routing-ti-lfa-04*
- *draft-ietf-spring-segment-routing-mpls-00*
- *draft-ietf-spring-segment-routing-02*
- *draft-ietf-isis-segment-routing-extensions-06*
- *draft-ietf-spring-segment-routing-11*
- *3draft-ietf-idr-bgpls-segment-routing-epe-15*
- *draft-ietf-idr-bgp-prefix-sid-19*
- RFC 7752, *North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP*
- RFC 8402, *Segment Routing Architecture*

Restrictions and guidelines: SR-MPLS configuration

This feature is supported only when the device operates in expert mode. For more inforamtion about device operating modes, see device management configuration in *Fundamentals Configuration Guide*.

For BGP-based SR, the SRGB range is fixed (16000 through 24000). To use both BGP and an IGP to advertise prefix SIDs, set the SRGB range for MPLS to 16000 through 24000.

SR-MPLS tasks at a glance

IP traffic forwarding over SRLSPs tasks at a glance

To forward IP traffic over SRLSPs, perform the following configuration tasks:

1. Configuring segments

   Select one of the following tasks:

   - Configuring static segments
   - Configuring IGP-based SID advertisement

2. (Optional.) Configuring the device to prefer SRLSPs in traffic forwarding
3. (Optional.) Configuring SR-MPLS and LDP interworking
4. (Optional.) Configuring TI-LFA FRR

Configuring static segments

Prerequisites for static segment configuration

Before you configure static segments for a static SRLSP, perform the following tasks:

- Determine the ingress node, transit nodes, and egress node of the static SRLSP.
- Enable MPLS on all nodes and interfaces that will participate in MPLS forwarding. For information about enabling MPLS, see basic MPLS configuration in *MPLS Configuration Guide*.

Configuring a static adjacency segment

Restrictions and guidelines

This task is required on all nodes of a static SRLSP.

Multiple static SRLSPs can share an adjacency segment.

If you specify the next hop address for a static adjacency segment, make sure the following requirements are met:

- The device has a route to the next hop address.
- MPLS is enabled on the outgoing interface of the route.

If you specify an outgoing interface for a static adjacency segment, make sure the following requirements are met:

- The interface is up.
- The interface can receive direct routes.
- MPLS is enabled on the interface.

A static adjacency segment must use a different incoming label than existing static LSPs, static PWs, and static CRLSPs. If not, the configured adjacency segment is unavailable. The adjacency segment cannot become available even if you change the incoming label of the static LSP, static PW, or static CRLSP. To resolve this problem, you must delete the existing adjacency segment and configure a new one with a different incoming label.

Procedure

1. Enter system view.

   system-view

2. Configure a static adjacency segment.

   static-sr-mpls adjacency *adjacency-path-name* **in-label** *label-value* { **nexthop** *ip-address* | **outgoing-interface** *interface-type interface-number* }

   The next hop address for a static adjacency segment cannot be a local public IP address.

Configuring a static prefix segment

Restrictions and guidelines

This task is required on all nodes of a static SRLSP.

Multiple static SRLSPs to the same destination can share a prefix segment.

A prefix segment must use the next hop or outgoing interface of the optimal route (non-BGP route) to the destination address of the prefix segment. You can configure multiple prefix segments to the destination address for load sharing if the optimal route has more than one next hop or outgoing interface. To avoid configuration failure, make sure all prefix segments use the same prefix segment name and incoming label.

Procedure

1. Enter system view.

   system-view

2. Configure a static prefix segment.

   static-sr-mpls prefix *prefix-path-name* **destination** *ip-address* { *mask-length* | *mask* } **in-label** *in-label-value* [{ **nexthop** *ip-address* | **outgoing-interface** *interface-type interface-number* } **out-label** *out-label-value*]

   The next hop address for a static prefix segment cannot be a local public IP address.

Configuring IGP-based SID advertisement

IGP-based prefix SID advertisement tasks at a glance

Perform the following tasks on each node along an SRLSP:

1. Configuring the IGP to support SR-MPLS
2. Configuring prefix SIDs
3. Configuring the MPLS SRGB

   Perform this task if you are configuring prefix SIDs.

IGP-based adjacency SID advertisement tasks at a glance

Perform the following tasks on each node along an SRLSP:

1. Configuring the IGP to support SR-MPLS
2. Enabling SR-MPLS adjacency label allocation for the IGP

   The device will allocate SIDs to adjacencies randomly.
3. (Optional.) Assigning adjacency SIDs

   You can assign adjacency SIDs to adjacencies manually.

Prerequisites for configuring IGP-based SID advertisement

Before you configure IGP-based SID advertisement, perform the following tasks:

- Determine the ingress node, transit nodes, and egress node of an SRLSP.
- Determine the following items for each node:
  - Index value for the prefix SID.
  - MPLS SRGB.
- Enable MPLS on all nodes and interfaces that will participate in MPLS forwarding. For information about enabling MPLS, see basic MPLS configuration in *MPLS Configuration Guide*.

Configuring the IGP to support SR-MPLS

Prerequisites

For SR-MPLS to take effect, set the IS-IS cost style to wide, compatible, or wide-compatible before configuring IS-IS to support SR-MPLS. For more information about the cost style, see IS-IS configuration in *Layer 3—IP Routing Configuration Guide*.

Configuring IS-IS to support SR-MPLS

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis *process-id*
3. Enter IS-IS IPv4 unicast address family view.
   address-family ipv4
4. Enable SR-MPLS.
   segment-routing mpls
   By default, SR-MPLS is disabled.

Configuring prefix SIDs

About this task

Configuring a prefix SID in loopback interface view binds the SID with the IP address of the loopback interface.

To configure a prefix SID, use one of the following methods:

- Specify an absolute value in the SRGB. The absolute value is used as the prefix SID.
- Specify an index value. The sum of the index value and the SRGB base value is used as the prefix SID. The prefix SID must be in the SRGB.

Restrictions and guidelines

To use a prefix SID for a group of SR nodes in anycast scenarios, specify the **n-flag-clear** keyword to set the Node-SID flag bit of the prefix SID to 0.

To configure an IS-IS prefix SID, you must enable an IS-IS process on the loopback interface.

Configuring an IS-IS prefix SID

1. Enter system view.
   system-view
2. Enter loopback interface view.
   interface loopback *interface-number*
3. Configure an IS-IS prefix SID.
   isis [**process-id** *process-id*] **prefix-sid** { **absolute** *absolute-value* | **index** *index-value* } [**n-flag-clear**] [**explicit-null**]
   By default, no IS-IS prefix SID is configured.

Enabling SR-MPLS adjacency label allocation for the IGP

Restrictions and guidelines

For this feature to take effect, you must enable SR-MPLS.

Enabling SR-MPLS adjacency label allocation for IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis *process-id*
3. Enter IS-IS IPv4 unicast address family view.
   address-family ipv4
4. Enable SR-MPLS adjacency label allocation.
   segment-routing adjacency enable
   By default, SR-MPLS adjacency label allocation is disabled.

Assigning adjacency SIDs

About this task

After you enable SR-MPLS adjacency label allocation, the device randomly allocates adjacency SIDs to the links to its IGP neighbors. If the link to an IGP neighbor flaps, the adjacency SID of the link keeps changing. For a link to always use the same adjacency SID, use this feature to assign a specific adjacency SID to the link.

Restrictions and guidelines

To assign adjacency SIDs by using absolute values, make sure the SIDs are in the label range of the SRLB. To assign adjacency SIDs by using index values, make sure the index values plus the base value of the SRLB are in the label range of the SRLB.

The assigned adjacency SIDs take effect after you enable SR-MPLS and SR-MPLS adjacency label allocation.

Prerequisites

Before assigning adjacency SIDs, execute the **display mpls label** command to display the usage status of the labels that you want to assign as adjacency SIDs. Make sure the labels are in **Idle** state.

A label that is not in **Idle** state is being used by another protocol. If you assign it to a link as an adjacency SID, the adjacency SID is not available even if the status of the label changes to **Idle** later. To use the adjacency SID, you must remove the adjacency SID assignment and assign the adjacency SID again.

Assigning adjacency SIDs to IS-IS adjacencies

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Assign an adjacency SID to an IS-IS adjacency link.
   isis adjacency-sid { **absolute** *absolute-value* | **index** *index-value* } [**nexthop** *nexthop-address*]
   By default, an IS-IS adjacency link does not have an adjacency SID.

Configuring the MPLS SRGB

Restrictions and guidelines

To configure the SRGB of a node successfully, make sure the SRGB contains the prefix SID configured for the node.

MPLS reserves labels in the range of 16000 to 24000 for SRGB. The minimum and maximum label values of the SRGB must be in the range.

Configuring the MPLS SRGB for IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis *process-id*
3. Configure the MPLS SRGB.
   segment-routing global-block *minivalue maxivalue*
   By default, the minimum label value is 16000, and the maximum label value is 24000.

Configuring BGP-based SID advertisement

BGP-based SID advertisement tasks at a glance

Perform the following tasks on each node along an SRLSP:

1. Configuring BGP to support SR-MPLS
2. Configuring BGP prefix SIDs

Prerequisites for configuring BGP-based SID advertisement

Before you configure BGP-based SID advertisement, perform the following tasks:

- Determine the ingress node, transit nodes, and egress node of an SRLSP.
- Determine the index value for the prefix SID of each node:
- Enable MPLS on all nodes and interfaces that will participate in MPLS forwarding. For information about enabling MPLS, see basic MPLS configuration in *MPLS Configuration Guide*.

Configuring BGP to support SR-MPLS

1. Enter system view.
   system-view
2. Enable BGP and enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Create the BGP IPv4 unicast address family and enter its view.
   address-family ipv4 [**unicast**]
4. Enable SR-MPLS.
   segment-routing mpls
   By default, SR-MPLS is disabled.

Configuring BGP prefix SIDs

About this task

By specifying a prefix SID for a routing policy and applying the policy to redistributed routes, you can define the binding between the SID and the prefix.

Restrictions and guidelines

To configure BGP prefix SIDs, you must enable the capability to exchange labeled routes with peers.

Configuring a BGP prefix SID

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
4. Redistribute routes from IS-IS and apply a routing policy to the redistributed routes.
   import-route isis *process-id* **route-policy** *route-policy-name*
5. Return to BGP instance view.
   quit
6. Return to system view.
   quit
7. Enter the routing policy view.
   route-policy *route-policy-name* { **deny** | **permit** } **node** *node-number*
8. Specify a label index value.
   apply label-index *index-value*

   By default, a routing policy does not have a label index value.

Configuring the device to prefer SRLSPs in traffic forwarding

About this task

This feature enables the device to preferentially use SRLSPs to forward traffic when both SRLSPs and LDP LSPs are available for traffic forwarding. If you do not configure this feature, the device prefers to use LDP LSPs for traffic forwarding.

Restrictions and guidelines

This feature takes effect only when SR-MPLS is enabled and SRLSPs use prefix SIDs.

Configuring the device to prefer SRLSPs established by IS-IS in traffic forwarding

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis *process-id*
3. Enter IS-IS IPv4 unicast address family view.
   address-family ipv4

4. Configure the device to prefer SRLSPs in traffic forwarding.
   segment-routing sr-prefer [**prefix-list** *prefix-list-name*]
   By default, the device prefers LDP LSPs to SRLSPs.

Configuring SR-MPLS and LDP interworking

Restrictions and guidelines for SR-MPLS and LDP interworking

To configure SR-MPLS and LDP interworking, make sure the SRLSP uses prefix SIDs.

In the current software version, the SRMS and SRMC devices support only IS-IS-based SR-MPLS.

SR to LDP tasks at a glance

SRMS tasks at a glance

Perform the following tasks on an SRMS:

1. Enabling SR-MPLS on the SRMS:
   a. Configuring the IGP to support SR-MPLS
   b. Configuring prefix SIDs
   c. Configuring the MPLS SRGB
2. Enabling advertisement of locally configured prefix-SID mappings
3. Configuring prefix-SID mappings

SRMC tasks at a glance

Perform the following tasks on each SRMC:

1. Enabling SR-MPLS on the SRMC:
   a. Configuring the IGP to support SR-MPLS
   b. Configuring prefix SIDs
   c. Configuring the MPLS SRGB
2. Enabling reception of advertised prefix-SID mappings

LDP to SR tasks at a glance

1. Configuring the IGP to support SR-MPLS
2. Configuring prefix SIDs
3. Configuring the MPLS SRGB

SR over LDP tasks at a glance

1. Configuring the IGP to support SR-MPLS
2. Configuring prefix SIDs
3. Configuring the MPLS SRGB

Prerequisites for SR-MPLS and LDP internetworking

Before you configure SR-MPLS and LDP internetworking, perform the following tasks:

- Determine the ingress node, transit nodes, and egress node of an SRLSP.
- Determine the index value for the prefix SID of each node.
- Enable MPLS on all nodes and interfaces that will participate in MPLS forwarding. For information about enabling MPLS, see basic MPLS configuration in *MPLS Configuration Guide*.

Enabling advertisement of locally configured prefix-SID mappings

Restrictions and guidelines

Configure this feature on the device that acts as the SRMS.

Configuring IS-IS to advertise locally configured prefix-SID mappings

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis *process-id*
3. Enter IS-IS IPv4 unicast address family view.
   address-family ipv4
4. Enable advertisement of locally configured prefix-SID mappings.
   segment-routing mapping-server advertise-local
   By default, advertisement is disabled for locally configured prefix-SID mappings.

Configuring prefix-SID mappings

Restrictions and guidelines

Configure this feature on the device that acts as the SRMS.

The specified prefixes must not belong to any existing prefix-SID mappings.

Prerequisites

Before you configure prefix-SID mappings in bulk, plan the number of mappings carefully.

Procedure

1. Enter system view.
   system-view
2. Enable SR and enter segment routing view.
   segment-routing
3. Configure a prefix-SID mapping.
   mapping-server prefix-sid-map *ip-address mask-length start-value* [**range** *range-value*] [**attached**]
   By default, no prefix-SID mappings exist.
   If you specify a value greater than 65535 for the *range-value* argument, the prefix-SID mappings cannot be advertised through IS-IS or OSPF.

Enabling reception of advertised prefix-SID mappings

Restrictions and guidelines

Configure this feature on the devices that act as the SRMCs.

Configuring IS-IS to receive advertised prefix-SID mappings.

1. Enter system view.
   `system-view`
2. Enter IS-IS view.
   `isis` *`process-id`*
3. Enter IS-IS IPv4 unicast address family view.
   `address-family ipv4`
4. Enable reception of prefix-SID mappings.
   `segment-routing mapping-server receive`
   By default, reception of prefix-SID mappings is enabled.

Configuring TI-LFA FRR

Restrictions and guidelines for TI-LFA FRR

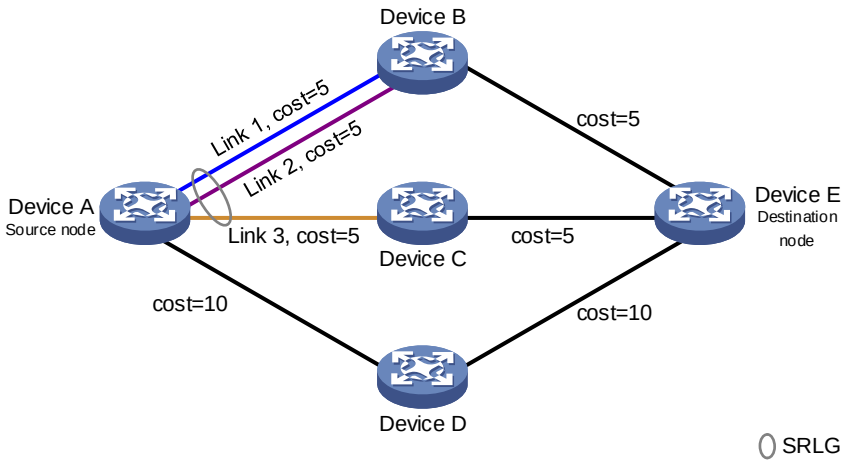
TI-LFA FRR is not supported on static SR networks.

In SR-MPLS and LDP interworking scenarios, you must configure the device to prefer the IS-IS SRLSP to the LDP LSP. This rule ensures the effectiveness of the TI-LFA FRR backup next hop.

By default, no backup path can be calculated by TI-LFA FRR if the next hops of equal-cost primary paths on the source node are different. To address this issue, you can add all equal-cost primary paths on the source node to the same SRLG.

As shown in Figure 16, three equivalent paths Link 1, Link 2, and Link 3 are available from source node Device A to destination node Device E, with their next hops different. For TI-LFA FRR to calculate backup paths, you must add Link 1, Link 2, and Link 3 to the same SRLG.

Figure 16 Using TI-LFA FRR to calculate backup paths in the IS-IS ECMP scenario



TI-LFA FRR tasks at a glance

1. Enabling TI-LFA FRR
2. (Optional.) Disabling an interface from participating in TI-LFA calculation

   On the source node, disable TI-LFA on the route's outgoing interface to the next hop on the primary path.
3. (Optional.) Configuring microloop avoidance

Enabling TI-LFA FRR

Enabling TI-LFA FRR for IS-IS

1. Enter system view.

   system-view
2. Enter IS-IS view.

   isis *process-id*
3. Enter IS-IS IPv4 unicast address family view.

   address-family ipv4
4. Enable LFA FRR for IS-IS.

   fast-reroute lfa [**level-1** | **level-2**]

   By default, LFA FRR is disabled for IS-IS.
5. Enable TI-LFA FRR for IS-IS.

   fast-reroute ti-lfa [**per-prefix**] [**route-policy** *route-policy-name* | **host**] [**level-1** | **level-2**]

   By default, TI-LFA FRR is disabled for IS-IS.
6. (Optional.) Set the priority for the FRR backup path selection policy.

   fast-reroute tiebreaker { **lowest-cost** | **node-protecting** } **preference** *preference* [**level-1** | **level-2**]

   By default, the priority values of the lowest-cost and node-protection backup path selection policies are 20 and 40, respectively.

Disabling an interface from participating in TI-LFA calculation

1. Enter system view.

   system-view
2. Enter the view of IS-IS interface.

   interface *interface-type interface-number*
3. Disable the interface from participating in TI-LFA calculation.

   isis fast-reroute ti-lfa disable [**per-prefix**] [**route-policy** *route-policy-name* | **host**] [**level-1** | **level-2**]

   By default, an IS-IS interface participates in TI-LFA calculation.

Configuring microloop avoidance

About this task

The following microloop avoidance methods are available:

- **FRR microloop avoidance**—Provides microloop avoidance after a network failure.

On a network deployed with TI-LFA FRR, when a node or link fails, traffic will be switched to the backup path calculated by TI-LFA. If devices on the backup path have not finished route convergence, a loop is formed between the source node (failed node or the previous node along the link) and a device on the backup path. The loop exists until the devices on the backup path finish route convergence.

To resolve this issue, FRR microloop avoidance first switches traffic to the backup path calculated by TI-LFA after the node or link failure. Then, the source node waits for the nodes on the backup path to finish route convergence before starting route convergence.

- **SR microloop avoidance**—Provides microloop avoidance both after a network failure and after a failure recovery.

  After a network failure occurs or recovers, route convergence occurs on relevant network devices. Because of non-simultaneous convergence on network devices, microloops might be formed. After you configure SR microloop avoidance, the devices will forward traffic along the specified path before route convergence is finished on all the relevant network devices. Because the forwarding path is independent of route convergence, microloops are avoided.

  To leave sufficient time for IGP to finish route convergence, set the SR microloop avoidance RIB-update-delay timer. Before the timer expires, failure relevant devices will forward traffic along the specified path. After the timer expires, IGP route convergence should have finished and traffic will traverse along the IGP-calculated path.

Restrictions and guidelines

If you configure both FRR microloop avoidance and SR microloop avoidance, SR microloop avoidance takes effect.

Configuring FRR microloop avoidance for IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis *process-id*
3. Enter IS-IS IPv4 unicast address family view.
   address-family ipv4
4. Enable FRR microloop avoidance for IS-IS.
   fast-reroute microloop-avoidance enable [**level-1** | **level-2**]
   By default, FRR microloop avoidance is disabled for IS-IS.
5. (Optional.) Set the FRR microloop avoidance RIB-update-delay time.
   fast-reroute microloop-avoidance rib-update-delay *delay-time* [**level-1** | **level-2**]
   By default, the FRR microloop avoidance RIB-update-delay time is 5000 ms.

Configuring SR microloop avoidance for IS-IS

1. Enter system view.
   system-view
2. Enter IS-IS view.
   isis *process-id*
3. Enter IS-IS IPv4 unicast address family view.
   address-family ipv4
4. Enable SR microloop avoidance for IS-IS.
   segment-routing microloop-avoidance enable [**level-1** | **level-2**]
   By default, SR microloop avoidance is disabled for IS-IS.
5. (Optional.) Set the SR microloop avoidance RIB-update-delay time.

segment-routing microloop-avoidance rib-update-delay *delay-time* [**level-1** | **level-2**]

By default, the SR microloop avoidance RIB-update-delay time is 5000 ms.

Display and maintenance commands for SR-MPLS

Execute **display** commands in any view.

Task	Command
Display BGP SR label ranges.	**display bgp** [*instance instance-name*] **segment-routing label-range**
Display IS-IS SR adjacency segment information.	**display isis segment-routing adjacency** [*process-id*]
Display IS-IS SRGB information.	**display isis segment-routing global-block** [**level-1** | **level-2**] [*process-id*]
Display IS-IS SR prefix-SID mappings.	**display isis segment-routing prefix-sid-map** [**active-policy** | **backup-policy**] [*process-id*] [**verbose**]
Display static SRLSP and adjacency segment information.	**display mpls static-sr-mpls** { **lsp** [*lsp-name*] | **adjacency** [*adjacency-path-name*] }
Display static prefix segment information.	**display mpls static-sr-mpls prefix** [**path** *lsp-name* | **destination** *ip-address* [*mask* | *mask-length*]]
Display prefix-SID mappings.	**display segment-routing mapping-server prefix-sid-map** [*ip-address mask-length* | **verbose**]

SR-MPLS configuration examples

Network configuration

As shown in Figure 17, complete the following tasks:

- Start IS-IS and enable IS-IS-based SR-MPLS on the PEs in the same AS.
- Configure PE 1 and ASBR-PE 1 to exchange labeled IPv4 routes through IBGP. Enable BGP-based SR-MPLS.
- Configure PE 2 and ASBR-PE 2 to exchange labeled IPv4 routes through IBGP. Enable BGP-based SR-MPLS.
- Establish an MP-EBGP peer relationship between PE 1 and PE 2 to exchange VPNv4 routes.
- Configure ASBR-PE 1 and ASBR-PE 2 to exchange labeled IPv4 routes through EBGP. Enable BGP-based SR-MPLS.
- Configure dynamic SID allocation on loopback interfaces of the devices. Then, establish an SRLSP based on the allocated SIDs and configure an MPLS TE tunnel over the SRLSP to transmit data.

Figure 17 Network diagram

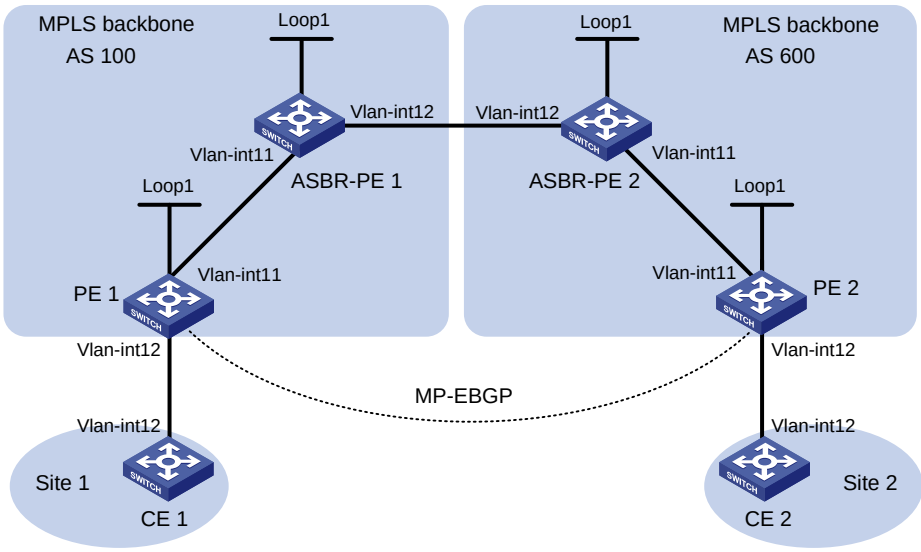


Table 1 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
PE 1	Loop1	2.2.2.9/32	PE 2	Loop1	5.5.5.9/32
	Vlan-int11	1.1.1.2/8		Vlan-int11	9.1.1.2/8
	Vlan-int12	30.0.0.1/24		Vlan-int12	20.0.0.1/24
ASBR-PE 1	Loop1	3.3.3.9/32	ASBR-PE 2	Loop1	4.4.4.9/32
	Vlan-int11	1.1.1.1/8		Vlan-int11	9.1.1.1/8
	Vlan-int12	11.0.0.2/8		Vlan-int12	11.0.0.1/8
CE 1	Vlan-int12	30.0.0.2/24	CE 2	Vlan-int12	20.0.0.2/24

Procedure

1. Configure IP addresses and masks for interfaces. (Details not shown.)
2. Configure CE 1:

   # Assign an IP address to interface VLAN-interface 12.

   ```
   <CE1> system-view
   [CE1] interface vlan-interface 12
   [CE1-Vlan-interface12] ip address 30.0.0.2 24
   [CE1-Vlan-interface12] quit
   ```

   # Establish an EBGP peer relationship with PE 1, and redistribute VPN routes.

   ```
   [CE1] bgp 65001
   [CE1-bgp-default] peer 30.0.0.1 as-number 100
   [CE1-bgp-default] address-family ipv4 unicast
   [CE1-bgp-default-ipv4] peer 30.0.0.1 enable
   [CE1-bgp-default-ipv4] import-route direct
   ```

3. Configure PE 1:

   # Start IS-IS on PE 1, set the LSR ID, and enable MPLS and MPLS TE.

   ```
   <PE1> system-view
   ```

```
[PE1] isis 1
[PE1-isis-1] network-entity 00.0000.0000.0001.00
[PE1-isis-1] cost-style wide
[PE1-isis-1] quit
[PE1] mpls lsr-id 2.2.2.9
[PE1] mpls te
[PE1-te] quit
[PE1] interface vlan-interface 11
[PE1-Vlan-interface11] isis enable 1
[PE1-Vlan-interface11] mpls enable
[PE1-Vlan-interface11] quit
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] mpls enable
[PE1-Vlan-interface12] quit
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 2.2.2.9 32
[PE1-LoopBack1] isis enable 1
[PE1-LoopBack1] quit
```

Enable SR-MPLS in IS-IS IPv4 unicast address family view and configure an IS-IS prefix SID.

```
[PE1] isis 1
[PE1-isis-1] address-family ipv4
[PE1-isis-1-ipv4] segment-routing mpls
[PE1-isis-1-ipv4] quit
[PE1-isis-1] quit
[PE1] interface loopback 1
[PE1-LoopBack1] isis prefix-sid index 20
[PE1-LoopBack1] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] route-distinguisher 11:11
[PE1-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE1-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE1-vpn-instance-vpn1] quit
```

Associate VLAN-interface 12 with VPN instance **vpn1**, and assign an IP address to the interface.

```
[PE1] interface vlan-interface 12
[PE1-Vlan-interface12] ip binding vpn-instance vpn1
[PE1-Vlan-interface12] ip address 30.0.0.1 24
[PE1-Vlan-interface12] quit
```

Create a routing policy and specify a label index value.

```
[PE1] route-policy policy1 permit node 1
[PE1-route-policy-policy1-1] apply label-index 10
[PE1-route-policy-policy1-1] quit
```

Start BGP on PE 1. Enable the capability to exchange labeled routes with IBGP peer 3.3.3.9.

```
[PE1] bgp 100
[PE1-bgp-default] peer 3.3.3.9 as-number 100
[PE1-bgp-default] peer 3.3.3.9 connect-interface loopback 1
[PE1-bgp-default] address-family ipv4 unicast
```

```
[PE1-bgp-default-ipv4] peer 3.3.3.9 enable
[PE1-bgp-default-ipv4] peer 3.3.3.9 label-route-capability
```

Enable SR-MPLS.

```
[PE1-bgp-default-ipv4] segment-routing mpls
```

Redistribute the route of loopback interface 1 to BGP and apply the routing policy.

```
[PE1-bgp-default-ipv4] network 2.2.2.9 32 route-policy policy1
[PE1-bgp-default-ipv4] quit
```

Configure the maximum number of hops from PE 1 to EBGP peer 5.5.5.9 as 10.

```
[PE1-bgp-default] peer 5.5.5.9 as-number 600
[PE1-bgp-default] peer 5.5.5.9 connect-interface loopback 1
[PE1-bgp-default] peer 5.5.5.9 ebgp-max-hop 10
```

Configure peer 5.5.5.9 as a VPNv4 peer.

```
[PE1-bgp-default] address-family vpnv4
[PE1-bgp-default-vpnv4] peer 5.5.5.9 enable
[PE1-bgp-default-vpnv4] quit
```

Configure CE 1 as an EBGP peer of PE 1 and install the learned BGP routes in the routing table of the VPN instance.

```
[PE1-bgp-default] ip vpn-instance vpn1
[PE1-bgp-default-vpn1] peer 30.0.0.2 as-number 65001
[PE1-bgp-default-vpn1] address-family ipv4 unicast
[PE1-bgp-default-ipv4-vpn1] peer 30.0.0.2 enable
```

4. Configure ASBR-PE 1:

Start IS-IS on ASBR-PE 1, set the LSR ID, and enable MPLS and MPLS TE.

```
<ASBR-PE1> system-view
[ASBR-PE1] isis 1
[ASBR-PE1-isis-1] network-entity 00.0000.0000.0002.00
[ASBR-PE1-isis-1] cost-style wide
[ASBR-PE1-isis-1] quit
[ASBR-PE1] mpls lsr-id 3.3.3.9
[ASBR-PE1] mpls te
[ASBR-PE1-te] quit
[ASBR-PE1] interface vlan-interface 11
[ASBR-PE1-Vlan-interface11] isis enable 1
[ASBR-PE1-Vlan-interface11] mpls enable
[ASBR-PE1-Vlan-interface11] quit
[ASBR-PE1] interface vlan-interface 12
[ASBR-PE1-Vlan-interface12] mpls enable
[ASBR-PE1-Vlan-interface12] quit
[ASBR-PE1] interface loopback 1
[ASBR-PE1-LoopBack1] ip address 3.3.3.9 32
[ASBR-PE1-LoopBack1] isis enable 1
[ASBR-PE1-LoopBack1] quit
```

Enable SR-MPLS in IS-IS IPv4 unicast address family view and configure an IS-IS prefix SID.

```
[ASBR-PE1] isis 1
[ASBR-PE1-isis-1] address-family ipv4
[ASBR-PE1-isis-1-ipv4] segment-routing mpls
[ASBR-PE1-isis-1-ipv4] quit
[ASBR-PE1-isis-1] quit
```

```
[ASBR-PE1] interface loopback 1
[ASBR-PE1-LoopBack1] isis prefix-sid index 30
[ASBR-PE1-LoopBack1] quit
```

Start BGP on ASBR-PE 1. Enable the capability to exchange labeled routes with IBGP peer 2.2.2.9.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp-default] peer 2.2.2.9 as-number 100
[ASBR-PE1-bgp-default] peer 2.2.2.9 connect-interface loopback 1
[ASBR-PE1-bgp-default] address-family ipv4 unicast
[ASBR-PE1-bgp-default-ipv4] peer 2.2.2.9 enable
[ASBR-PE1-bgp-default-ipv4] peer 2.2.2.9 label-route-capability
```

Enable SR-MPLS.

```
[ASBR-PE1-bgp-default-ipv4] segment-routing mpls
[ASBR-PE1-bgp-default-ipv4] quit
```

Enable the capability to exchange labeled routes with EBGP peer 11.0.0.1.

```
[ASBR-PE1-bgp-default] peer 11.0.0.1 as-number 600
[ASBR-PE1-bgp-default] address-family ipv4 unicast
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 enable
[ASBR-PE1-bgp-default-ipv4] peer 11.0.0.1 label-route-capability
[ASBR-PE1-bgp-default-ipv4] quit
[ASBR-PE1-bgp-default] quit
```

5. Configure ASBR-PE 2:

Start IS-IS on ASBR-PE 2, set the LSR ID, and enable MPLS and MPLS TE.

```
<ASBR-PE2> system-view
[ASBR-PE2] isis 1
[ASBR-PE2-isis-1] network-entity 00.0000.0000.0003.00
[ASBR-PE2-isis-1] cost-style wide
[ASBR-PE2-isis-1] quit
[ASBR-PE2] mpls lsr-id 4.4.4.9
[ASBR-PE2] mpls te
[ASBR-PE2-te] quit
[ASBR-PE2] interface vlan-interface 11
[ASBR-PE2-Vlan-interface11] isis enable 1
[ASBR-PE2-Vlan-interface11] mpls enable
[ASBR-PE2-Vlan-interface11] quit
[ASBR-PE2] interface vlan-interface 12
[ASBR-PE2-Vlan-interface12] mpls enable
[ASBR-PE2-Vlan-interface12] quit
[ASBR-PE2] interface loopback 1
[ASBR-PE2-LoopBack1] ip address 4.4.4.9 32
[ASBR-PE2-LoopBack1] isis enable 1
[ASBR-PE2-LoopBack1] quit
```

Enable SR-MPLS in IS-IS IPv4 unicast address family view and configure an IS-IS prefix SID.

```
[ASBR-PE2] isis 1
[ASBR-PE2-isis-1] address-family ipv4
[ASBR-PE2-isis-1-ipv4] segment-routing mpls
[ASBR-PE2-isis-1-ipv4] quit
[ASBR-PE2-isis-1] quit
```

```
[ASBR-PE2] interface loopback 1
[ASBR-PE2-LoopBack1] isis prefix-sid index 40
[ASBR-PE2-LoopBack1] quit
```

Start BGP on ASBR-PE 2. Enable the capability to exchange labeled routes with IBGP peer 5.5.5.9.

```
[ASBR-PE2] bgp 600
[ASBR-PE2-bgp-default] peer 5.5.5.9 as-number 600
[ASBR-PE2-bgp-default] peer 5.5.5.9 connect-interface loopback 1
[ASBR-PE2-bgp-default] address-family ipv4 unicast
[ASBR-PE2-bgp-default-ipv4] peer 5.5.5.9 enable
[ASBR-PE2-bgp-default-ipv4] peer 5.5.5.9 label-route-capability
```

Enable SR-MPLS.

```
[ASBR-PE2-bgp-default-ipv4] segment-routing mpls
[ASBR-PE2-bgp-default-ipv4] quit
```

Enable the capability to exchange labeled routes with EBGP peer 11.0.0.2.

```
[ASBR-PE2-bgp-default] peer 11.0.0.2 as-number 100
[ASBR-PE2-bgp-default] address-family ipv4 unicast
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 enable
[ASBR-PE2-bgp-default-ipv4] peer 11.0.0.2 label-route-capability
[ASBR-PE2-bgp-default-ipv4] quit
[ASBR-PE2-bgp-default] quit
```

6. Configure PE 2:

Start IS-IS on PE 2, set the LSR ID, and enable MPLS and MPLS TE.

```
<PE2> system-view
[PE2] isis 1
[PE2-isis-1] network-entity 00.0000.0000.0004.00
[PE2-isis-1] cost-style wide
[PE2-isis-1] quit
[PE2] mpls lsr-id 5.5.5.9
[PE2] mpls te
[PE2-te] quit
[PE2] interface vlan-interface 11
[PE2-Vlan-interface11] isis enable 1
[PE2-Vlan-interface11] mpls enable
[PE2-Vlan-interface11] quit
[PE2] interface vlan-interface 12
[PE2-Vlan-interface12] mpls enable
[PE2-Vlan-interface12] quit
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 5.5.5.9 32
[PE2-LoopBack1] isis enable 1
[PE2-LoopBack1] quit
```

Enable SR-MPLS in IS-IS IPv4 unicast address family view and configure an IS-IS prefix SID.

```
[PE2] isis 1
[PE2-isis-1] address-family ipv4
[PE2-isis-1-ipv4] segment-routing mpls
[PE2-isis-1-ipv4] quit
[PE2-isis-1] quit
```

```
[PE2] interface loopback 1
[PE2-LoopBack1] isis prefix-sid index 50
[PE2-LoopBack1] quit
```

Create VPN instance **vpn1**, and configure the RD and route target attributes.

```
[PE2] ip vpn-instance vpn1
[PE2-vpn-instance-vpn1] route-distinguisher 11:11
[PE2-vpn-instance-vpn1] vpn-target 1:1 2:2 3:3 import-extcommunity
[PE2-vpn-instance-vpn1] vpn-target 3:3 export-extcommunity
[PE2-vpn-instance-vpn1] quit
```

Associate VLAN-interface 12 with VPN instance **vpn1**, and assign an IP address to the interface.

```
[PE2] interface vlan-interface 12
[PE2-Vlan-interface12] ip binding vpn-instance vpn1
[PE2-Vlan-interface12] ip address 20.0.0.1 24
[PE2-Vlan-interface12] quit
```

Create a routing policy and specify a label index value.

```
[PE2] route-policy policy1 permit node 1
[PE2-route-policy-policy1-1] apply label-index 50
[PE2-route-policy-policy1-1] quit
```

Start BGP on PE 2. Enable the capability to exchange labeled routes with IBGP peer 4.4.4.9.

```
[PE2] bgp 600
[PE2-bgp-default] peer 4.4.4.9 as-number 600
[PE2-bgp-default] peer 4.4.4.9 connect-interface loopback 1
[PE2-bgp-default] address-family ipv4 unicast
[PE2-bgp-default-ipv4] peer 4.4.4.9 enable
[PE2-bgp-default-ipv4] peer 4.4.4.9 label-route-capability
```

Enable SR-MPLS.

```
[PE2-bgp-default-ipv4] segment-routing mpls
```

Redistribute the route of loopback interface 1 to BGP and apply the routing policy.

```
[PE2-bgp-default-ipv4] network 5.5.5.9 32 route-policy policy1
[PE2-bgp-default-ipv4] quit
```

Configure the maximum number of hops from PE 2 to EBGP peer 2.2.2.9 as 10.

```
[PE2-bgp-default] peer 2.2.2.9 as-number 100
[PE2-bgp-default] peer 2.2.2.9 connect-interface loopback 1
[PE2-bgp-default] peer 2.2.2.9 ebgp-max-hop 10
```

Configure peer 2.2.2.9 as a VPNv4 peer.

```
[PE2-bgp-default] address-family vpnv4
[PE2-bgp-default-vpnv4] peer 2.2.2.9 enable
[PE2-bgp-default-vpnv4] quit
```

Configure CE 2 as an EBGP peer of PE 2 and install the learned BGP routes in the routing table of the VPN instance.

```
[PE2-bgp-default] ip vpn-instance vpn1
[PE2-bgp-default-vpn1] peer 20.0.0.2 as-number 65001
[PE2-bgp-default-vpn1] address-family ipv4 unicast
[PE2-bgp-default-ipv4-vpn1] peer 20.0.0.2 enable
```

7. Configure CE 2:

Assign an IP address to interface VLAN-interface 12.

```
<CE2> system-view
```

```
[CE2] interface vlan-interface 12
[CE2-Vlan-interface12] ip address 20.0.0.2 24
[CE2-Vlan-interface12] quit
```

Establish an EBGP peer relationship with PE 2, and redistribute VPN routes.

```
[CE2] bgp 65002
[CE2-bgp-default] peer 20.0.0.1 as-number 600
[CE2-bgp-default] address-family ipv4 unicast
[CE2-bgp-default-ipv4] peer 20.0.0.1 enable
[CE2-bgp-default-ipv4] import-route direct
[CE2-bgp-default-ipv4] quit
[CE2-bgp-default] quit
```

Verifying the configuration

Execute the **`display ip routing-table`** command on CE 1 and CE 2 to verify that they have a route to each other. (Details not shown.)

Verify that CE 1 and CE 2 can ping each other. (Details not shown.)

Display MPLS LSP information on PE1.

```
[PE1]display mpls lsp
```

FEC	Proto	In/Out Label	Out Inter/NHLFE/LSINDEX
2.2.2.9/32	BGP	3/-	-
5.5.5.9/32	BGP	-/16040	NHLFE17
5.5.5.9	BGP	-/-	Vlan11
1.1.1.1	Local	-/-	Vlan11
2.2.2.9/32	ISIS	16020/-	-
3.3.3.9/32	ISIS	16030/3	Vlan11
3.3.3.9/32	ISIS	-/3	Vlan11

Example: Configuring SR-MPLS to LDP (IS-IS)

Network configuration

As shown in Figure 18, complete the following tasks so Switch A and Switch C can communicate with each other:

- Configure Switch A, Switch B, and Switch C to run IS-IS.
- Configure Switch A and Switch B to run SR-MPLS and establish an SRLSP between Switch A and Switch B.
- Configure Switch B and Switch C to run LDP to establish a dynamic LSP between Switch B and Switch C.
- Configure Switch B as the SRMS, map the prefix address of the LDP network to an SID, and advertise the mapping to Switch A.

Figure 18 Network diagram

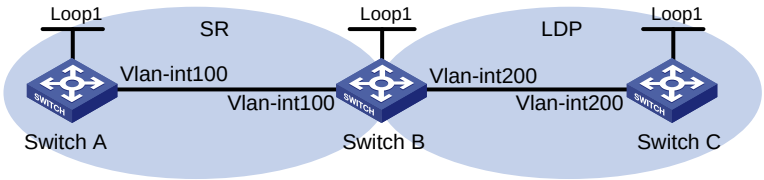


Table 2 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Loop1	1.1.1.1/32	Switch B	Loop1	2.2.2.2/32
	Vlan-Int100	10.0.0.1/24		Vlan-Int100	10.0.0.2/24
Switch C	Loop1	3.3.3.3/32		Vlan-Int200	11.0.0.1/24
	Vlan-Int200	11.0.0.2/24			

Procedure

1. Configure IP addresses and masks for interfaces. (Details not shown.)
2. Configure IS-IS on the switches to achieve network level connectivity and set the IS-IS cost style to wide:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] network-entity 00.0000.0000.0001.00
[SwitchA-isis-1] cost-style wide
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
[SwitchA] interface loopback 1
[SwitchA-LoopBack1] isis enable 1
[SwitchA-LoopBack1] quit
```

   # Configure an IS-IS prefix SID.

```
[SwitchA] interface loopback 1
[SwitchA-LoopBack1] isis prefix-sid index 10
[SwitchA-LoopBack1] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 00.0000.0000.0002.00
[SwitchB-isis-1] cost-style wide
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] isis enable 1
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
[SwitchB] interface loopback 1
[SwitchB-LoopBack1] isis enable 1
[SwitchB-LoopBack1] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
```

```
[SwitchC-isis-1] network-entity 00.0000.0000.0003.00
[SwitchC-isis-1] cost-style wide
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface loopback 1
[SwitchC-LoopBack1] isis enable 1
[SwitchC-LoopBack1] quit
```

3. Configure LSR IDs on the switches:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.1
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.2
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.3
```

4. Configure LDP on Switch B and Switch C:

Configure Switch B.

```
[SwitchB] mpls ldp
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] mpls enable
[SwitchB-Vlan-interface200] mpls ldp enable
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
[SwitchC] mpls ldp
[SwitchC-ldp] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] mpls enable
[SwitchC-Vlan-interface200] mpls ldp enable
[SwitchC-Vlan-interface200] quit
```

5. Enable SR-MPLS on Switch A and Switch B:

Configure Switch A.

```
[SwitchA] isis 1
[SwitchA-isis-1] address-family ipv4
[SwitchA-isis-1-ipv4] segment-routing mpls
[SwitchA-isis-1-ipv4] quit
[SwitchA-isis-1] quit
```

Configure Switch B.

```
[SwitchB] isis 1
[SwitchB-isis-1] address-family ipv4
[SwitchB-isis-1-ipv4] segment-routing mpls
[SwitchB-isis-1-ipv4] quit
[SwitchB-isis-1] quit
```

6. Configure the SRMS:

On Switch B, configure the IGP to advertise locally configured prefix-SID mappings.

```
[SwitchB] isis 1
```

```
[SwitchB-isis-1] address-family ipv4
[SwitchB-isis-1-ipv4] segment-routing mapping-server advertise-local
[SwitchB-isis-1-ipv4] quit
[SwitchB-isis-1] quit
```

On Switch B, configure a prefix-SID mapping.

```
[SwitchB] segment-routing
[SwitchB-segment-routing] mapping-server prefix-sid-map 3.3.3.3 32 100
[SwitchB-segment-routing] quit
```

Verifying the configuration

Display IS-IS SRLSP information on Switch A.

```
[SwitchA] display mpls lsp protocol isis
FEC                         Proto       In/Out Label    Out Inter/NHLFE/LSINDEX
3.3.3.3/32                  ISIS        16100/16100     Vlan100
3.3.3.3/32                  ISIS        -/16100         Vlan100
```

The output shows that the outgoing label of the IS-IS SRLSP entry for Switch C is using the SID assigned to Switch C.

Example: Configuring SR-MPLS over LDP

Network configuration

As shown in Figure 19, complete the following tasks so the two SR networks can communicate across the LDP network:

- Configure Switch A, Switch B, Switch C, Switch D, and Switch E to run IS-IS.
- Configure Switch B, Switch C, and Switch D to run LDP.
- Configure Switch A, Switch B, Switch D, and Switch E to run SR-MPLS.

Figure 19 Network diagram

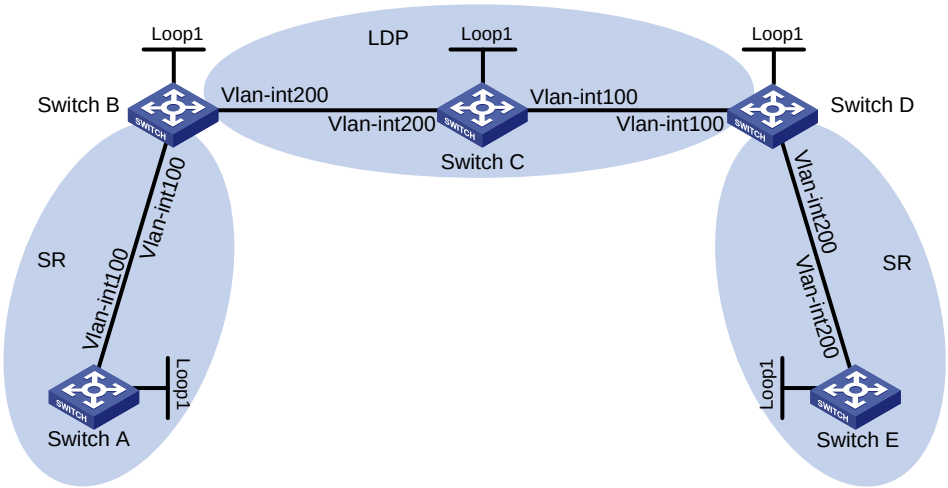


Table 3 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Loop1	1.1.1.1/32	Switch B	Loop1	2.2.2.2/32
	Vlan-Int100	10.0.0.1/24		Vlan-Int100	10.0.0.2/24

Switch C	Loop1	3.3.3.3/32		Vlan-Int200	11.0.0.1/24
	Vlan-Int100	12.0.0.1/24	Switch D	Loop1	4.4.4.4/32
	Vlan-Int200	11.0.0.2/24		Vlan-Int100	12.0.0.2/24
Switch E	Loop1	5.5.5.5/32		Vlan-Int200	13.0.0.1/24
	Vlan-Int200	13.0.0.2/24			

Procedure

1. Configure IP addresses and masks for interfaces. (Details not shown.)
2. Configure IS-IS on the switches to achieve network level connectivity and set the IS-IS cost style to wide:

   # Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] network-entity 00.0000.0000.0001.00
[SwitchA-isis-1] cost-style wide
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
[SwitchA] interface loopback 1
[SwitchA-LoopBack1] isis enable 1
[SwitchA-LoopBack1] quit
```

   # Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 00.0000.0000.0002.00
[SwitchB-isis-1] cost-style wide
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] isis enable 1
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
[SwitchB] interface loopback 1
[SwitchB-LoopBack1] isis enable 1
[SwitchB-LoopBack1] quit
```

   # Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 00.0000.0000.0003.00
[SwitchC-isis-1] cost-style wide
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
```

```
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface loopback 1
[SwitchC-LoopBack1] isis enable 1
[SwitchC-LoopBack1] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] network-entity 00.0000.0000.0004.00
[SwitchD-isis-1] cost-style wide
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] isis enable 1
[SwitchD-Vlan-interface100] quit
[SwitchD] interface vlan-interface 200
[SwitchD-Vlan-interface200] isis enable 1
[SwitchD-Vlan-interface200] quit
[SwitchD] interface loopback 1
[SwitchD-LoopBack1] isis enable 1
[SwitchD-LoopBack1] quit
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] isis 1
[SwitchE-isis-1] network-entity 00.0000.0000.0005.00
[SwitchE-isis-1] cost-style wide
[SwitchE-isis-1] quit
[SwitchE] interface vlan-interface 200
[SwitchE-Vlan-interface200] isis enable 1
[SwitchE-Vlan-interface200] quit
[SwitchE] interface loopback 1
[SwitchE-LoopBack1] isis enable 1
[SwitchE-LoopBack1] quit
```

3. Configure LSR IDs on the switches:

Configure Switch A.

```
[SwitchA] mpls lsr-id 1.1.1.1
```

Configure Switch B.

```
[SwitchB] mpls lsr-id 2.2.2.2
```

Configure Switch C.

```
[SwitchC] mpls lsr-id 3.3.3.3
```

Configure Switch D.

```
[SwitchD] mpls lsr-id 4.4.4.4
```

Configure Switch E.

```
[SwitchE] mpls lsr-id 5.5.5.5
```

4. Configure LDP on Switch B, Switch C, and Switch D:

Configure Switch B.

```
[SwitchB] mpls ldp
```

```
[SwitchB-ldp] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] mpls enable
[SwitchB-Vlan-interface200] mpls ldp enable
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
[SwitchC] mpls ldp
[SwitchC-ldp] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] mpls enable
[SwitchC-Vlan-interface100] mpls ldp enable
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] mpls enable
[SwitchC-Vlan-interface200] mpls ldp enable
[SwitchC-Vlan-interface200] quit
```

Configure Switch D.

```
[SwitchD] mpls ldp
[SwitchD-ldp] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface1001] mpls enable
[SwitchD-Vlan-interface100] mpls ldp enable
[SwitchD-Vlan-interface100] quit
```

5. Enable SR-MPLS on Switch A, Switch B, Switch D, and Switch E:

Configure Switch A.

```
[SwitchA] isis 1
[SwitchA-isis-1] address-family ipv4
[SwitchA-isis-1-ipv4] segment-routing mpls
[SwitchA-isis-1-ipv4] quit
[SwitchA-isis-1] quit
```

Configure Switch B.

```
[SwitchB] isis 1
[SwitchB-isis-1] address-family ipv4
[SwitchB-isis-1-ipv4] segment-routing mpls
[SwitchB-isis-1-ipv4] quit
[SwitchB-isis-1] quit
```

Configure Switch D.

```
[SwitchD] isis 1
[SwitchD-isis-1] address-family ipv4
[SwitchD-isis-1-ipv4] segment-routing mpls
[SwitchD-isis-1-ipv4] quit
[SwitchD-isis-1] quit
```

Configure Switch E.

```
[SwitchE] isis 1
[SwitchE-isis-1] address-family ipv4
[SwitchE-isis-1-ipv4] segment-routing mpls
[SwitchE-isis-1-ipv4] quit
[SwitchE-isis-1] quit
```

6. Configure IS-IS prefix SIDs for Switch A, Switch B, Switch D, and Switch E:

   # Configure Switch A.

```
[SwitchA] interface loopback 1
[SwitchA-LoopBack1] isis prefix-sid index 10
[SwitchA-LoopBack1] quit
```

   # Configure Switch B.

```
[SwitchB] interface loopback 1
[SwitchB-LoopBack1] isis prefix-sid index 20
[SwitchB-LoopBack1] quit
```

   # Configure Switch D.

```
[SwitchD] interface loopback 1
[SwitchD-LoopBack1] isis prefix-sid index 40
[SwitchD-LoopBack1] quit
```

   # Configure Switch E.

```
[SwitchE] interface loopback 1
[SwitchE-LoopBack1] isis prefix-sid index 50
[SwitchE-LoopBack1] quit
```

Verifying the configuration

Display LDP LSP information on Switch B.

```
[SwitchB] display mpls ldp lsp
Status Flags: * - stale, L - liberal, B - backup, N/A - unavailable
FECs: 5           Ingress: 3           Transit: 3           Egress: 2

FEC                In/Out Label        Nexthop         OutInterface/LSINDEX
1.1.1.1/32         2173/-
                   -/2173(L)
2.2.2.2/32         3/-
                   -/2175(L)
3.3.3.3/32         -/3                 11.0.0.2        Vlan200
                   2175/3              11.0.0.2        Vlan200
4.4.4.4/32         -/2174              11.0.0.2        Vlan200
                   2174/2174           11.0.0.2        Vlan200
5.5.5.5/32         -/2172              11.0.0.2        Vlan200
                   2172/2172           11.0.0.2        Vlan200
```

Display IS-IS SRLSP information on Switch B.

```
[SwitchB] display mpls lsp protocol isis
FEC                          Proto       In/Out Label    Out Inter/NHLFE/LSINDEX
1.1.1.1/32                   ISIS        16010/3         Vlan100
1.1.1.1/32                   ISIS        -/3             Vlan100
2.2.2.2/32                   ISIS        16020/-         -
4.4.4.4/32                   ISIS        16040/2174      Vlan200
4.4.4.4/32                   ISIS        -/2174          Vlan200
5.5.5.5/32                   ISIS        16050/2172      Vlan200
5.5.5.5/32                   ISIS        -/2172          Vlan200
```

The output shows that the IS-IS SRLSP entries for Switch D and Switch E are using LDP labels.

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

ACL and QoS Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes fundamentals and configuration procedures for ACL, QoS, time range, and data buffer.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring ACLs 1

About ACLs 1
Numbering and naming ACLs 1
ACL types 1
Match order 1
Comments and remarks for rules 2
Rule numbering 3
Fragment filtering with ACLs 3
Restrictions and guidelines: ACL configuration 3
ACL tasks at a glance 4
Configuring a basic ACL 4
About basic ACLs 4
Configuring an IPv4 basic ACL 4
Configuring an IPv6 basic ACL 5
Configuring an advanced ACL 6
About advanced ACLs 6
Configuring an IPv4 advanced ACL 7
Configuring an IPv6 advanced ACL 21
Configuring a Layer 2 ACL 28
Configuring a user-defined ACL 29
Copying an ACL 30
Configuring the QoS and ACL resource hardware mode 30
Configuring packet filtering with ACLs 31
About packet filtering with ACLs 31
Applying an ACL to filter packets globally 31
Applying an ACL to an interface for packet filtering 31
Configuring the applicable scope of packet filtering on a VLAN interface 32
Configuring logging and SNMP notifications for packet filtering 32
Setting the packet filtering default action 33
Configuring an EM template and its match criteria 33
Display and maintenance commands for ACL 34
ACL configuration examples 34
Example: Configuring interface-based packet filter 34

Configuring ACLs

Only the S5200-HI-G switch series supports user-defined ACLs.

About ACLs

An access control list (ACL) is a set of rules for identifying traffic based on criteria such as source IP address, destination IP address, and port number. The rules are also called permit or deny statements.

ACLs are primarily used for packet filtering. You can also use ACLs in QoS, security, routing, and other modules for identifying traffic. The packet drop or forwarding decisions depend on the modules that use ACLs.

Numbering and naming ACLs

When creating an ACL, you must assign it a number or name for identification. You can specify an existing ACL by its number or name. Each ACL type has a unique range of ACL numbers.

For basic or advanced ACLs with the same number, you must use the **`ipv6`** keyword to distinguish them. For ACLs with the same name, you must use the **`ipv6`**, **`mac`**, and **`user-defined`** keywords to distinguish them.

ACL types

Type	ACL number	IP version	Match criteria
Basic ACLs	2000 to 2999	IPv4	Source IPv4 address.
		IPv6	Source IPv6 address.
Advanced ACLs	3000 to 3999	IPv4	Source IPv4 address, destination IPv4 address, packet priority, protocol number, and other Layer 3 and Layer 4 header fields.
		IPv6	Source IPv6 address, destination IPv6 address, packet priority, protocol number, and other Layer 3 and Layer 4 header fields.
Layer 2 ACLs	4000 to 4999	IPv4 and IPv6	Layer 2 header fields, such as source and destination MAC addresses, 802.1p priority, and link layer protocol type.
User-defined ACLs	5000 to 5999	IPv4 and IPv6	User specified matching patterns in protocol headers.

Match order

The rules in an ACL are sorted in a specific order. When a packet matches a rule, the device stops the match process and performs the action defined in the rule. If an ACL contains overlapping or conflicting rules, the matching result and action to take depend on the rule order.

The following ACL match orders are available:

- **config**—Sorts ACL rules in ascending order of rule ID. A rule with a lower ID is matched before a rule with a higher ID. If you use this method, check the rules and their order carefully.

NOTE:

The match order of user-defined ACLs can only be **config**.

- **auto**—Sorts ACL rules in depth-first order. Depth-first ordering makes sure any subset of a rule is always matched before the rule. The following table lists the sequence of tie breakers that depth-first ordering uses to sort rules for each type of ACL.

Table 1 Sort ACL rules in depth-first order

ACL type	Sequence of tie breakers
IPv4 basic ACL	**1.** VPN instance. **1.** More 0s in the source IPv4 address wildcard (more 0s means a narrower IPv4 address range). **2.** Rule configured earlier.
IPv4 advanced ACL	**1.** VPN instance. **2.** Specific protocol number. **3.** More 0s in the source IPv4 address wildcard mask. **4.** More 0s in the destination IPv4 address wildcard. **5.** Narrower TCP/UDP service port number range. **6.** Rule configured earlier.
IPv6 basic ACL	**1.** VPN instance. **2.** Longer prefix for the source IPv6 address (a longer prefix means a narrower IPv6 address range). **3.** Rule configured earlier.
IPv6 advanced ACL	**1.** VPN instance. **2.** Specific protocol number. **3.** Longer prefix for the source IPv6 address. **4.** Longer prefix for the destination IPv6 address. **5.** Narrower TCP/UDP service port number range. **6.** Rule configured earlier.
Layer 2 ACL	**1.** More 1s in the source MAC address mask (more 1s means a smaller MAC address). **2.** More 1s in the destination MAC address mask. **3.** Rule configured earlier.

A wildcard mask, also called an inverse mask, is a 32-bit binary number represented in dotted decimal notation. In contrast to a network mask, the 0 bits in a wildcard mask represent "do care" bits, and the 1 bits represent "don't care" bits. If the "do care" bits in an IP address are identical to the "do care" bits in an IP address criterion, the IP address matches the criterion. All "don't care" bits are ignored. The 0s and 1s in a wildcard mask can be noncontiguous. For example, 0.255.0.255 is a valid wildcard mask.

Comments and remarks for rules

An ACL can have multiple rules. To identify rules, you can add a comment for each rule one by one, or add the same remark for multiple rules.

- For rules requiring different identification, you can add a comment for a particular rule or add a different comment for each rule.
- For a range of rules requiring the same identification, you can insert a remark before the start rule and a remark after the end rule. This configuration mode eliminates the need for configuring a comment for each rule one by one and improves configuration efficiency.

Rule numbering

ACL rules can be manually numbered or automatically numbered. This section describes how automatic ACL rule numbering works.

Rule numbering step

If you do not assign an ID to the rule you are creating, the system automatically assigns it a rule ID. The rule numbering step sets the increment by which the system automatically numbers rules. For example, the default ACL rule numbering step is 5. If you do not assign IDs to rules you are creating, they are automatically numbered 0, 5, 10, 15, and so on. The wider the numbering step, the more rules you can insert between two rules.

By introducing a gap between rules rather than contiguously numbering rules, you have the flexibility of inserting rules in an ACL. This feature is important for a config-order ACL, where ACL rules are matched in ascending order of rule ID.

The rule numbering step sets the increment by which the system numbers rules automatically. If you do not specify a rule ID when creating an ACL rule, the system automatically assigns it a rule ID. This rule ID is the nearest higher multiple of the numbering step to the current highest rule ID, starting from the start rule ID. For example, if the rule numbering step is 5 and the current highest rule ID is 12, the rule is numbered 15.

The wider the numbering step, the more rules you can insert between two rules. Whenever the step or start rule ID changes, the rules are renumbered, starting from the start rule ID. For example, if there are five rules numbered 0, 5, 9, 10, and 15, changing the step from 5 to 2 causes the rules to be renumbered 0, 2, 4, 6, and 8.

Automatic rule numbering and renumbering

The ID automatically assigned to an ACL rule takes the nearest higher multiple of the numbering step to the current highest rule ID, starting with 0.

For example, if the step is 5, and there are five rules numbered 0, 5, 9, 10, and 12, the newly defined rule is numbered 15. If the ACL does not contain a rule, the first rule is numbered 0.

Whenever the step changes, the rules are renumbered, starting from 0. For example, changing the step from 5 to 2 renumbers rules 5, 10, 13, and 15 as rules 0, 2, 4, and 6.

For an ACL of the match order **auto**, rules are sorted in depth-first order, and are renumbered based on the match order. For example, rules are in the match order of 0, 10, and 5. Changing the numbering step to 2 renumbers rules 0, 10, and 5 (not 0, 5, and 10) as rules 0, 2, 4.

Fragment filtering with ACLs

Traditional packet filtering matches only first fragments of packets, and allows all subsequent non-first fragments to pass through. Attackers can fabricate non-first fragments to attack networks.

To avoid risks, the ACL feature is designed as follows:

- Filters all fragments by default, including non-first fragments.
- Allows for matching criteria modification for efficiency. For example, you can configure the ACL to filter only non-first fragments.

Restrictions and guidelines: ACL configuration

- If you create a numbered ACL, you can enter the view of the ACL by using either of the following commands:
  - **acl** [**ipv6**] **number** *acl-number*.
  - **acl** { [**ipv6**] { **advanced** | **basic** } | **mac** | **user-defined** } *acl-number*.

- If you create a ACL by using the **acl** [**ipv6**] **number** *acl-number* **name** *acl-name* command, you can enter the view of the ACL by using either of the following commands:
  - **acl** [**ipv6**] **name** *acl-name* (for only basic ACLs and advanced ACLs).
  - **acl** [**ipv6**] **number** *acl-number* [**name** *acl-name*].
  - **acl** { [**ipv6**] { **advanced** | **basic** } | **mac** | **user-defined** } **name** *acl-name*.
- If you create a named ACL by using the **acl** { [**ipv6**] { **advanced** | **basic** } | **mac** | **user-defined** } **name** *acl-name* command, you can enter the view of the ACL by using either of the following commands:
  - **acl** [**ipv6**] **name** *acl-name* (for only basic ACLs and advanced ACLs).
  - **acl** { [**ipv6**] { **advanced** | **basic** } | **mac** | **user-defined** } **name** *acl-name*.

ACL tasks at a glance

To configure an ACL, perform the following tasks:

- Configure ACLs according to the characteristics of the packets to be matched
  - Configuring a basic ACL
  - Configuring an advanced ACL
  - Configuring a Layer 2 ACL
  - Configuring a user-defined ACL
- (Optional.) Copying an ACL
- (Optional.) Configuring the QoS and ACL resource hardware mode
- (Optional.) Configuring packet filtering with ACLs
- (Optional.) Configuring an EM template and its match criteria

Configuring a basic ACL

About basic ACLs

Basic ACLs match packets based only on source IP addresses.

Basic ACLs can be used in the following scenarios:

- To improve security when the device acts as an FTP or TFTP server, you can use basic ACLs to allow only matching clients to access the server. For more information, see FTP and TFTP configuration in *Fundamentals Configuration Guide*.
- In a multicast scenario, you can use basic ACLs in a multicast source policy to receive or forward only matching multicast packets. For more information about multicast source policy, see PIM configuration and IPv6 PIM configuration in *IP Multicast Configuration Guide*.
- To filter routes in a routing policy, you can use basic ACLs to receive or send matching routes. For more information about routing policy, see *Layer 3—IP Routing Configuration Guide*.
- To take different QoS actions on different traffic types, you can use basic ACLs to classify traffic. For more information about traffic classes, see "Configuring QoS."

Configuring an IPv4 basic ACL

1. Enter system view.
   system-view

2. Create an IPv4 basic ACL and enter its view. Choose one option as needed:
   - Create an IPv4 basic ACL by specifying an ACL number.

     acl number *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]
   - Create an IPv4 basic ACL by specifying the **basic** keyword.

     acl basic { *acl-number* | **name** *acl-name* } [**match-order** { **auto** | **config** }]
3. (Optional.) Configure a description for the IPv4 basic ACL.

   description *text*

   By default, an IPv4 basic ACL does not have a description.
4. (Optional.) Set the rule numbering step.

   step *step-value* [**start** *start-value*]

   By default, the rule numbering step is 5 and the start rule ID is 0.
5. Create or edit a rule.

   rule [*rule-id*] { **deny** | **permit** } [**counting** | **fragment** | [**flow-logging** | **logging**] | **source** { **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

   The **flow-logging** or **logging** keyword takes effect only when the module (for example, packet filtering) that uses the ACL supports logging.
6. (Optional.) Add or edit a rule comment.

   rule *rule-id* **comment** *text*

   By default, no rule comment is configured.
7. (Optional.) Add a rule remark.

   rule [*rule-id*] **remark** *text*

   By default, no rule remark is configured.

Configuring an IPv6 basic ACL

1. Enter system view.

   system-view
2. Create an IPv6 basic ACL view and enter its view. Choose one option as needed:
   - Create an IPv6 basic ACL by specifying an ACL number.

     acl ipv6 number *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]
   - Create an IPv6 basic ACL by specifying the **basic** keyword.

     acl ipv6 basic { *acl-number* | **name** *acl-name* } [**match-order** { **auto** | **config** }]
3. (Optional.) Configure a description for the IPv6 basic ACL.

   description *text*

   By default, an IPv6 basic ACL does not have a description.
4. (Optional.) Set the rule numbering step.

   step *step-value* [**start** *start-value*]

   By default, the rule numbering step is 5 and the start rule ID is 0.
5. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } [**counting** | **fragment** | [**flow-logging** | **logging**] | **routing** [**type** *routing-type*] | **source** { **object-group** *address-group-name* | *source-address source-prefix* | *source-address*/*source-prefix* | **any** } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

The **flow-logging** or **logging** keyword takes effect only when the module (for example, packet filtering) that uses the ACL supports logging.

6. (Optional.) Add or edit a rule comment.

   rule *rule-id* **comment** *text*

   By default, no rule comment is configured.

7. (Optional.) Add a rule remark.

   rule [*rule-id*] **remark** *text*

   By default, no rule remark is configured.

Configuring an advanced ACL

About advanced ACLs

Advanced ACLs can be used in the following scenarios:

- When the device acts as an FTP or TFTP server, you can use advanced ACLs to allow only matching clients to access the server. For more information, see FTP and TFTP configuration in *Fundamentals Configuration Guide*.
- In a multicast scenario, you can use advanced ACLs in a multicast source policy to receive or forward only matching multicast packets. For more information about multicast source policy, see PIM configuration and IPv6 PIM configuration in *IP Multicast Configuration Guide*.
- To filter routes in a routing policy, you can use advanced ACLs to receive or send matching routes. For more information about routing policy, see *Layer 3—IP Routing Configuration Guide*.
- To take different QoS actions on different traffic types, you can use advanced ACLs to classify traffic. For more information about traffic classification see "Configuring QoS."

Advanced ACLs match packets based on the following criteria:

- Source IP addresses.
- Destination IP addresses.
- Packet priorities.
- Local QoS IDs.
- Protocol types.
- Other protocol header information, such as TCP/UDP source and destination port numbers, TCP flags, ICMP message types, and ICMP message codes.
- Encapsulation types.
- Inner source IPv4 addresses.
- Inner destination IPv4 addresses.
- Inner protocol types.
- Other inner protocol header information, such as inner TCP/UDP source and destination port numbers.

Compared to basic ACLs, advanced ACLs allow more flexible and accurate filtering.

Configuring an IPv4 advanced ACL

1. Enter system view.

   system-view

2. Create an IPv4 advanced ACL and enter its view. Choose one option as needed:

   - Create a numbered IPv4 advanced ACL by specifying an ACL number.

     acl number *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]

   - Create an IPv4 advanced ACL by specifying the **advanced** keyword.

     acl advanced { *acl-number* | **name** *acl-name* } [**match-order** { **auto** | **config** }]

3. (Optional.) Configure a description for the IPv4 advanced ACL.

   description *text*

   By default, an IPv4 advanced ACL does not have a description.

4. (Optional.) Set the rule numbering step.

   step *step-value* [**start** *start-value*]

   By default, the rule numbering step is 5 and the start rule ID is 0.

5. Create or edit a rule. Choose the options to configure as needed:

   - Create or edit a rule for matching non-encapsulated packets.

     rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | [**any-fragment** | **first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **packet-length** *operator length-value1* [*length-value2*] | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **user-group** *group-name* | **vpn-instance** *vpn-instance-name*] *

   - Create or edit a rule for matching encapsulated packets.

     rule [*rule-id*] { **deny** | **permit** } { **gre-encapsulation** | **ipinip-encapsulation** | **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | **inner-source-port** *operator port1* [*port2*] | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence**

inner-precedence } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | [**flow-logging** | **logging**] | **packet-length** *operator length-value1* [*length-value2*] | **time-range** *time-range-name*] *

Parameter	Description
gre-encapsulation	Matches GRE packets. The **source-port** *operator port1* [*port2*] and **vxlan-id** *vxlan-id* options cannot be configured if you specify the **gre-encapsulation** keyword.
ipinip-encapsulation	Matches IP-in-IP encapsulation packets. The **source-port** *operator port1* [*port2*] and **vxlan-id** *vxlan-id* options cannot be configured if you specify the **ipinip-encapsulation** keyword.
vxlan	Matches VXLAN packets by both outer and inner packet information.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

6. Create or edit a rule.

   rule [*rule-id*] { **deny** | **permit** } *protocol* [**counting** | **destination** { *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { **dscp** *dscp* | **precedence** *precedence* } | [**first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | **logging** | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name*] *

7. Create or edit a rule.
   - Create or edit a rule for matching non-encapsulated packets.

     rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

     In a VXLAN network, only the inner header information of VXLAN packets can be matched on edge devices, and only the outer header information of VXLAN packets can be matched on core devices.
   - Create or edit a rule for matching encapsulated packets.

     rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination**

{ *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | **inner-source-port** *operator port1* [*port2*] | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner header information.

8. Create or edit a rule. Choose the options to configure as needed:

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.

geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

9. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { { **precedence** *precedence* | **tos** *tos* } * | **precedence** *precedence* | **dscp** *dscp* } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | **logging** | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name*] *

10. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | [**any-fragment** | **first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name* | **vpn-instance** *vpn-instance-name*] *

The S5200-EI-G switch series does not support the following keywords:

- **any-fragment**
- **first-fragment**
- **non-fragment**
- **non-or-first-fragment**

11. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | **logging** | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

12. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

13. Create or edit a rule.
    - Create or edit a rule for matching non-encapsulated packets.

      rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *
    - Create or edit a rule for matching encapsulated packets.

      rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | **inner-source-port** *operator port1* [*port2*] | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

14. Create or edit a rule.
    - Create or edit a rule for matching non-encapsulated packets.

      rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *
    - Create or edit a rule for matching encapsulated packets.

rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner header information.

15. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

Only the S5560X-HI and S6520X-HI switch series support the **microsegment** parameter.

16. Create or edit a rule.

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port**

operator port1 [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.

17. Create or edit a rule.

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | **logging** | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | **inner-source-port** *operator port1* [*port2*] | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **logging** | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.

18. Create or edit a rule. Choose the options to configure as needed:

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment**

microsegment-id [**mask-length** *mask-length*] | *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | **logging** | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | **logging** | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by outer or inner packet information.

19. Create or edit a rule. Choose the options to configure as needed:

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

  An IPv4 advanced ACL rule containing the **qos-local-id** keyword cannot be used for outbound QoS traffic classification or outbound packet filtering.

  The VPN instance cannot be matched in the outbound direction.

  In the inbound direction on edge devices in a VXLAN network, the inner header information of VXLAN packets is matched first, and the outer header information is matched if no matched is found in the inner header information. On core devices, only the outer header information of VXLAN packets can be matched.

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] *

inner-protocol *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.

20. Create or edit a rule.

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | [**any-fragment** | **first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | **inner-source-port** *operator port1* [*port2*] | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet

	information.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

21. Create or edit a rule.

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | **logging** | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | **logging** | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.

22. Create or edit a rule.

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | **logging** | **qos-local-id** *local-id-value* | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

rule [*rule-id*] { **deny** | **permit** } **vxlan** [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | **logging** | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.

23. Create or edit a rule. Choose the options to configure as needed:

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

24. Create or edit a rule. Choose the options to configure as needed:

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

25. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [**counting** | **destination** { **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | **precedence** *precedence* } | [**first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

26. Create or edit a rule.

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | { { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | **inner-source-port** *operator port1* [*port2*] | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

27. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [**counting** | **destination** { **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** }

| **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { **dscp** *dscp* | { **precedence** *precedence* | **tos** *tos* } * } | [**any-fragment** | **first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

28. Create or edit a rule.

rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { { **precedence** *precedence* | **tos** *tos* } * | **fragment** | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

Only the S6813 switch series supports the **microsegment** parameter.

29. Create or edit a rule. Choose the options to configure as needed:

- Create or edit a rule for matching non-encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** } | **destination-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | { { **precedence** *precedence* | **tos** *tos* } * | { **precedence** *precedence* | **ecn** *ecn* } * | { **dscp** *dscp* | **ecn** *ecn* } * } | [**any-fragment** | **first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **icmp-type** { *icmp-type* [*icmp-code*] | *icmp-message* } | [**flow-logging** | **logging**] | **qos-local-id** *local-id-value* | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-wildcard* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching encapsulated packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**destination** { *dest-address dest-wildcard* | **any** } | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-wildcard* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-wildcard* | **any** } | **inner-source-port** *operator port1* [*port2*] | { { **inner-dscp**

inner-dscp | **inner-ecn** *inner-ecn* } * | { **inner-ecn** *inner-ecn* | **inner-precedence** *inner-precedence* } * | { **inner-precedence** *inner-precedence* | **inner-tos** *inner-tos* } * } | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches VXLAN packets by both outer and inner packet information.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

The **flow-logging** or **logging** keyword takes effect only when the module (for example, packet filtering) that uses the ACL supports logging.

30. (Optional.) Add or edit a rule comment.

    rule *rule-id* **comment** *text*

    By default, no rule comment is configured.

31. (Optional.) Add a rule remark.

    rule [*rule-id*] **remark** *text*

    By default, no rule remark is configured.

Configuring an IPv6 advanced ACL

1. Enter system view.

   system-view

2. Create an IPv6 advanced ACL and enter its view. Choose one option as needed:
   - Create a numbered IPv6 advanced ACL by specifying an ACL number.

     acl ipv6 number *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]
   - Create an IPv6 advanced ACL by specifying the **advanced** keyword.

     acl ipv6 advanced { *acl-number* | **name** *acl-name* } [**match-order** { **auto** | **config** }]

3. (Optional.) Configure a description for the IPv6 advanced ACL.

   description *text*

   By default, an IPv6 advanced ACL does not have a description.

4. (Optional.) Set the rule numbering step.

   step *step-value* [**start** *start-value*]

   By default, the rule numbering step is 5 and the start rule ID is 0.

5. Create or edit a rule. Choose the options to configure as needed:
   - Create or edit a rule for matching non-encapsulated packets.

     rule [*rule-id*] { **deny** | **permit** } *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **counting** | **destination** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **destination-port** { **object-group**

port-group-name | *operator port1* [*port2*] } | **destination-port** *operator port1* [*port2*] | | { **dscp** *dscp* | **ecn** *ecn-value* } * | **dscp** *dscp* | **dscp** *dscp* | **dscp** *dscp* | **dscp** *dscp* | **dscp** *dscp* | **flow-label** *flow-label-value* | [**any-fragment** | **first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **fragment** | [**first-fragment** | **fragment** | **non-fragment** | **non-or-first-fragment**] | **fragment** | **fragment** | **fragment** | **fragment** | **fragment** | **fragment** | **fragment** | **fragment** | **icmp6-type** { *icmp6-type icmp6-code* | *icmp6-message* } | [**flow-logging** | **logging**] | **logging**| | **packet-length** *operator length-value1* [*length-value2*] | **qos-local-id** *local-id-value* | **routing** [**type** *routing-type*] | **hop-by-hop** [**type** *hop-type*] | **source** { **microsegment** *microsegment-id* [**mask-length** *mask-length*] | **object-group** *address-group-name* | *source-address source-prefix* | *source-address/source-prefix* | **any** } | **source-port** { **object-group** *port-group-name* | *operator port1* [*port2*] } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name*| **user-group** *group-name* | **vpn-instance** *vpn-instance-name*] *

- Create or edit a rule for matching IPv4 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **packet-length** *operator length-value1* [*length-value2*] | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**vxlan-id** *vxlan-id*] **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator*

port1 [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | **logging** | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**vxlan-id** *vxlan-id*] **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | **logging** | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | **logging** | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**vxlan-id** *vxlan-id*] **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | **logging** | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN packets.

  rule [*rule-id*] { **deny** | **permit** } **vxlan** [**vxlan-id** *vxlan-id*] **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | **logging** | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] *

inner-protocol *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv4 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

- Create or edit a rule for matching IPv6 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan-ipv6** | **geneve-ipv6** } [**destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **packet-length** *operator length-value1* [*length-value2*] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches IPv4 VXLAN packets by both outer and inner packet information. The following outer header information is not supported for matching IPv4 VXLAN packets: • Outer source IP address. • Outer destination IP address. • Outer source port number. • Outer destination port number.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

vxlan-ipv6	Matches IPv6 VXLAN packets by both outer and inner packet information.
geneve-ipv6	Matches IPv6 Geneve packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

- Create or edit a rule for matching IPv6 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan-ipv6** | **geneve-ipv6** } [**destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches IPv4 VXLAN packets by both outer and inner packet information. The following outer header information is not supported for matching IPv4 VXLAN packets: • Outer source IP address. • Outer destination IP address. • Outer source port number. • Outer destination port number.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.
vxlan-ipv6	Matches IPv6 VXLAN packets by both outer and inner packet information.
geneve-ipv6	Matches IPv6 Geneve packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

- Create or edit a rule for matching IPv6 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan-ipv6** | **geneve-ipv6** } [**destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol**

inner-protocol [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches IPv4 VXLAN packets by both outer and inner packet information. The following outer header information is not supported for matching IPv4 VXLAN packets: • Outer source IP address. • Outer destination IP address. • Outer source port number. • Outer destination port number.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.
vxlan-ipv6	Matches IPv6 VXLAN packets by both outer and inner packet information.
geneve-ipv6	Matches IPv6 Geneve packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

- Create or edit a rule for matching IPv6 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan-ipv6** | **geneve-ipv6** } [**destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | { { **inner-ack** *inner-ack-value* | **inner-fin** *inner-fin-value* | **inner-psh** *inner-psh-value* | **inner-rst** *inner-rst-value* | **inner-syn** *inner-syn-value* | **inner-urg** *inner-urg-value* } * | **inner-established** } | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches IPv4 VXLAN packets by both outer and inner packet information. The following outer header information is not supported for

	matching IPv4 VXLAN packets: • Outer source IP address. • Outer destination IP address. • Outer source port number. • Outer destination port number.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.
vxlan-ipv6	Matches IPv6 VXLAN packets by both outer and inner packet information.
geneve-ipv6	Matches IPv6 Geneve packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

- Create or edit a rule for matching IPv6 VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan-ipv6** | **geneve-ipv6** } [**destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **source-port** *operator port1* [*port2*] | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] * **inner-protocol** *inner-protocol* [**counting** | **inner-destination** { *dest-address dest-prefix* | *dest-address/dest-prefix* | **any** } | **inner-destination-port** *operator port1* [*port2*] | **inner-source** { *source-address source-prefix* | *source-address/source-prefix* | **any** } | **inner-source-port** *operator port1* [*port2*] | { **inner-dscp** *inner-dscp* | **inner-ecn** *inner-ecn* } * | [**flow-logging** | **logging**] | **time-range** *time-range-name*] *

Parameter	Description
vxlan	Matches IPv4 VXLAN packets by both outer and inner packet information. The following outer header information is not supported for matching IPv4 VXLAN packets: • Outer source IP address. • Outer destination IP address. • Outer source port number. • Outer destination port number.
geneve	Matches Generic Network Virtualization Encapsulation (Geneve) packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.
vxlan-ipv6	Matches IPv6 VXLAN packets by both outer and inner packet information.
geneve-ipv6	Matches IPv6 Geneve packets. You can specify the **geneve-id** *geneve-id* and **option-header-length** *option-length* options only after you specify this keyword.

The **flow-logging** or **logging** keyword takes effect only when the module (for example, packet filtering) that uses the ACL supports logging.

6. (Optional.) Add or edit a rule comment.

   rule *rule-id* **comment** *text*

   By default, no rule comment is configured.

7. (Optional.) Add a rule remark.

   rule [*rule-id*] **remark** *text*

   By default, no rule remark is configured.

Configuring a Layer 2 ACL

About this task

Layer 2 ACLs can be used in the following scenarios:

- To improve security when the device acts as a Telnet server, you can use Layer 2 ACLs to allow only matching clients to access the server. For more information, see login management in *Fundamentals Configuration Guide*.
- To take different QoS actions on different traffic types, you can use Layer 2 ACLs to classify traffic. For more information about traffic classes, see "Configuring QoS."

Layer 2 ACLs, also called Ethernet frame header ACLs, match packets based on Layer 2 Ethernet header fields, such as:

- Source MAC address.
- Destination MAC address.
- 802.1p priority (VLAN priority).
- Link layer protocol type.
- Encapsulation type.

Procedure

1. Enter system view.

   system-view

2. Create a Layer 2 ACL and enter its view. Choose one option as needed:
   - Create a Layer 2 ACL by specifying an ACL number.

     acl number *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]
   - Create a Layer 2 ACL by specifying the **mac** keyword.

     acl mac { *acl-number* | **name** *acl-name* } [**match-order** { **auto** | **config** }]

3. (Optional.) Configure a description for the Layer 2 ACL.

   description *text*

   By default, a Layer 2 ACL does not have a description.

4. (Optional.) Set the rule numbering step.

   step *step-value* [**start** *start-value*]

   By default, the rule numbering step is 5 and the start rule ID is 0.

5. Create or edit a rule.
   - Create or edit a rule for matching common packets.

     rule [*rule-id*] { **deny** | **permit** } [**cos** *dot1p* | **counting** | **dest-mac** *dest-address dest-mask* | **type** *protocol-type protocol-type-mask* |

source-mac *source-address source-mask* | **time-range** *time-range-name*] *

- Create or edit a rule for matching VXLAN or Geneve packets.

  rule [*rule-id*] { **deny** | **permit** } { **vxlan** | **geneve** } [**counting** | **dest-mac** *dest-address dest-mask* | **inner-dest-mac** *inner-dest-address inner-dest-mask* | **inner-source-mac** *inner-source-address inner-source-mask* | **inner-type** *inner-protocol-type inner-protocol-type-mask* | **source-mac** *source-address source-mask* | **time-range** *time-range-name* | **type** *protocol-type protocol-type-mask* | **vxlan-id** *vxlan-id* | **geneve-id** *geneve-id* | **option-header-length** *option-length*] *

6. (Optional.) Add or edit a rule comment.

   rule *rule-id* **comment** *text*

   By default, no rule comment is configured.

7. (Optional.) Add a rule remark.

   rule [*rule-id*] **remark** *text*

   By default, no rule remark is configured.

Configuring a user-defined ACL

About this task

User-defined ACLs allow you to customize rules based on information in protocol headers. You can define a user-defined ACL to match packets. A specific number of bytes after an offset (relative to the specified header) are compared against a match pattern after being ANDed with a match pattern mask.

Procedure

1. Enter system view.

   system-view

2. Create a user-defined ACL and enter its view. Choose one option as needed:

   - Create a user-defined ACL by specifying an ACL number.

     acl number *acl-number* [**name** *acl-name*]

   - Create a user-defined ACL by specifying the **user-defined** keyword.

     acl user-defined { *acl-number* | **name** *acl-name* }

3. (Optional.) Configure a description for the user-defined ACL.

   description *text*

   By default, a user-defined ACL does not have a description.

4. Create or edit a rule.

   Command set 1:

   rule [*rule-id*] { **deny** | **permit** } [{ { **ipv4** | **ipv6** | **l2** | **l4** } *rule-string rule-mask offset* }&<1-8>] [**counting** | **time-range** *time-range-name*] *

   Command set 2:

   rule [*rule-id*] { **deny** | **permit** } [**ipv6-protocol**] *protocol* [{ { **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **established** } | **destination** { **object-group** *address-group-name* | *dest-address dest-wildcard* | **any** | **microsegment** *microsegment-id* [**mask-length** *mask-length*] } | **destination-port** { *operator port1* [*port2*] } | { { **precedence** *precedence* | **tos** *tos* } * | **dscp**

dscp } | **source** { **object-group** *address-group-name* | *source-address source-wildcard* | **any** | **microsegment** *microsegment-id* [**mask-length** *mask-length*] } } | **source-port** { *operator port1* [*port2*] } | **vpn-instance** *vpn-instance-name*] * [{ { **ipv4** | **ipv6** | **l2** | **l4** } *rule-string rule-mask offset* }&<1-8>] [**counting** | **time-range** *time-range-name*] *

5. (Optional.) Add or edit a rule comment.

   rule *rule-id* **comment** *text*

   By default, no rule comment is configured.

6. (Optional.) Add a rule remark.

   rule [*rule-id*] **remark** *text*

   By default, no rule remark is configured.

Copying an ACL

About this task

You can create an ACL by copying an existing ACL (source ACL). The new ACL (destination ACL) has the same properties and content as the source ACL, but uses a different number or name than the source ACL.

Restrictions and guidelines

To successfully copy an ACL, make sure:

- The destination ACL is the same type as the source ACL.
- The source ACL already exists, but the destination ACL does not.

Procedure

1. Enter system view.

   system-view

2. Copy an existing ACL to create a new ACL.

   acl [**ipv6** | **mac** | **user-defined**] **copy** { *source-acl-number* | **name** *source-acl-name* } **to** { *dest-acl-number* | **name** *dest-acl-name* }

Configuring the QoS and ACL resource hardware mode

About this task

Different chips produce different packet matching results for QoS and ACL resources. Perform this task to enhance matching capabilities of QoS and ACL resources.

Restrictions and guidelines

This feature occupies more QoS and ACL resources when an ACL is applied.

Procedure

1. Enter system view.

   system-view

2. Configure the QoS and ACL resource hardware mode.

   qos-acl resource hardware-mode *hardware-mode-value*

   By default, no QoS and ACL resource hardware mode is configured.

Configuring packet filtering with ACLs

About packet filtering with ACLs

This section describes procedures for using an ACL to filtering packets. For example, you can apply an ACL to an interface to filter incoming or outgoing packets.

Applying an ACL to filter packets globally

Restrictions and guidelines

If you apply an exact-match ACL to filter packets globally, do not set the packet filtering default action to deny.

Procedure

1. Enter system view.

   system-view

2. Apply an ACL globally.

   packet-filter [**ipv6** | **mac** | **user-defined**] { *acl-number* | **name** *acl-name* } **global** { **inbound** [**exact-match**] | **outbound** } [**hardware-count**]

   By default, global packet filtering is not configured.

Applying an ACL to an interface for packet filtering

Restrictions and guidelines

A VLAN interface does not support using IPv4 ACL rules configured with the **vpn-instance** parameter to filter packets.

When an ACL is applied to the inbound direction of an interface for packet filtering, the ACL cannot match the destination MAC address or customer VLAN field of IPv4 VXLAN packets.

The following interfaces do not support using IPv4 ACL rules configured with the **vpn-instance** parameter to filter packets:

- Layer 3 Ethernet interface/subinterfaces.
- Layer 3 aggregate interface/subinterfaces.

An ACL applied to the outbound direction of a VLAN interface cannot match untagged packets.

The following restrictions apply to the **extend** keyword:

- If equal-cost routes exist from the device to the network accessed through a VLAN interface, the VLAN interface does not support the **extend** keyword.
- support the **extend** keyword only on VLAN interfaces.
- The packet filtering feature that uses expanded ACL resources and the microsegmentation feature are mutually exclusive in the outbound direction of the same VLAN interface. If you have configured one feature in the outbound direction of a VLAN interface, the other feature cannot take effect in the same direction of the VLAN interface.
- The packet filtering feature on a VLAN interface that uses expanded ACL resources and the **packet-filter filter all** command are mutually exclusive.
- If you apply an ACL to the outbound direction of a VLAN interface with the keyword and execute the **packet-filter default deny** command, the deny action takes effect on only Layer 3 packets on that VLAN interface.

If you apply an exact-match ACL to filter packets globally, do not set the packet filtering default action to deny.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Apply an ACL to the interface to filter packets.

   packet-filter [**ipv6** | **mac** | **user-defined**] { *acl-number* | **name** *acl-name* } { **inbound** [**exact-match**] | **outbound** [**extend**] } [**hardware-count**]

   By default, an interface does not filter packets.

Configuring the applicable scope of packet filtering on a VLAN interface

About this task

You can configure the packet filtering on a VLAN interface to filter the following packets:

- Packets forwarded at Layer 3 by the VLAN interface.
- All packets, including packets forwarded at Layer 3 by the VLAN interface and packets forwarded at Layer 2 by the physical ports associated with the VLAN interface.

Procedure

1. Enter system view.

   system-view

2. Create a VLAN interface and enter its view.

   interface vlan-interface *vlan-interface-id*

   If the VLAN interface already exists, you directly enter its view.

   By default, no VLAN interface exists.

3. Specify the applicable scope of packet filtering on the VLAN interface.

   packet-filter filter [**route** | **all**]

   By default, the packet filtering filters packets forwarded at Layer 3.

Configuring logging and SNMP notifications for packet filtering

About this task

You can configure the ACL module to generate log entries or SNMP notifications for packet filtering and output them to the information center or SNMP module at the output interval. When the first packet of a flow matches an ACL rule, the output interval starts. At the end of the interval, the device outputs a log entry or notification to record the number of matching packets and the matched ACL rule during the interval.

For more information about the information center and SNMP, see *Network Management and Monitoring Configuration Guide.*

Procedure

1. Enter system view.

system-view

2. Set the interval for outputting packet filtering logs or notifications.

   acl { **logging** | **trap** } **interval** *interval*

   The default setting is 0 minutes. By default, the device does not generate log entries or SNMP notifications for packet filtering.

Setting the packet filtering default action

1. Enter system view.

   system-view

2. Set the packet filtering default action to deny.

   packet-filter default deny

   By default, the packet filter permits packets that do not match any ACL rule to pass.

Configuring an EM template and its match criteria

About this task

An exact match (EM) template is a template of match criteria. It can be used to match the match criteria in a QoS policy or packet filter, match criteria configured by using the **rule** command in an ACL, or match criteria configured by using the **if-match** command in QoS traffic class view. When the preceding match criteria separately correspond to the match criteria configured by using the **exact-match** command in EM template view, the service module will use EM resources to match and process packets. EM resources are forwarding entry resources of a device, and are stored in the same hardware storage space as MAC address entries and route entries. By using EM resources to match and process packets, you can increase the number of match criteria supported by the device.

To use this feature in a service module, reference an EM template in the service module. A service module referencing an EM template can use EM resources to process the specified match criteria.

Restrictions and guidelines

In the current software version, only the QoS policy and packet filter service modules support this feature. For information about referencing EM templates in a service module, see *ACL and QoS Command Reference* or "Configuring packet filtering with ACLs".

If an EM template has been referenced by a service module, you cannot modify or delete match criteria in the EM template. To do that, first remove the EM template from the service module.

Procedure

1. Enter system view.

   system-view

2. Create an EM template and enter its view.

   acl exact-match template *template-id*

   By default, no EM templates exists.

3. Configure match criteria in the EM template.

   exact-match { **ipv4** | **ipv6** } { **destination-ip** | **destination-port** | **protocol** | **source-ip** | **source-port** } *

   By default, no match criteria are configured in an EM template.

Display and maintenance commands for ACL

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ACL configuration and match statistics.	**display acl** [**ipv6** | **mac** | **user-defined**] { *acl-number* | **all** | **name** *acl-name* }
Display ACL application information for packet filtering.	**display packet-filter** { **global** | **interface** [*interface-type interface-number*] } [**inbound** | **outbound**] [**slot** *slot-number*]
Display match statistics for packet filtering ACLs.	**display packet-filter statistics** { **global** | **interface** *interface-type interface-number* } { **inbound** | **outbound** } [[**ipv6** | **mac** | **user-defined**] { *acl-number* | **name** *acl-name* }] [**brief**]
Display the accumulated statistics for packet filtering ACLs.	**display packet-filter statistics sum** { **inbound** | **outbound** } [**ipv6** | **mac** | **user-defined**] { *acl-number* | **name** *acl-name* } [**brief**]
Display detailed ACL packet filtering information.	**display packet-filter verbose** { **global** | **interface** *interface-type interface-number* } { **inbound** | **outbound** } [[**ipv6** | **mac** | **user-defined**] { *acl-number* | **name** *acl-name* }] [**slot** *slot-number*]
Display QoS and ACL resource usage.	**display qos-acl resource** [**slot** *slot-number*]
Clear ACL statistics.	**reset acl** [**ipv6** | **mac** | **user-defined**] **counter** { *acl-number* | **all** | **name** *acl-name* }
Display the configuration and state of an applied EM template.	**display acl exact-match template** [*template-id*] [**slot** *slot-number*]
Clear match statistics for packet filtering ACLs.	**reset packet-filter statistics** { **global** | **interface** [*interface-type interface-number*] } { **inbound** | **outbound** } [[**ipv6** | **mac** | **user-defined**] { *acl-number* | **name** *acl-name* }]

ACL configuration examples

Example: Configuring interface-based packet filter

Network configuration

A company interconnects its departments through the device. Configure a packet filter to:

- Permit access from the President's office at any time to the financial database server.
- Permit access from the Finance department to the database server only during working hours (from 8:00 to 18:00) on working days.
- Deny access from any other department to the database server.

Figure 1 Network diagram

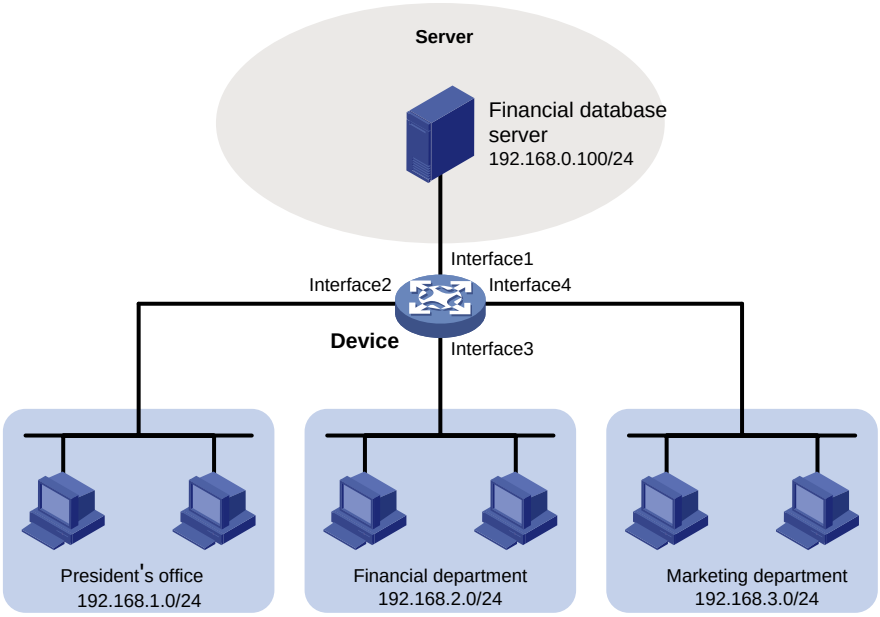


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

Create a periodic time range from 8:00 to 18:00 on working days.

```
<Device> system-view
[Sysname] sysname Device
[Device] time-range work 08:0 to 18:00 working-day
```

Assign IP addresses to interfaces.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] ip address 192.168.0.1 24
[Device-GigabitEthernet1/0/1] quit
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] ip address 192.168.1.1 24
[Device-GigabitEthernet1/0/2] quit
[Device] interface gigabitethernet 1/0/3
[Device-GigabitEthernet1/0/3] ip address 192.168.2.1 24
[Device-GigabitEthernet1/0/3] quit
[Device] interface gigabitethernet 1/0/4
[Device-GigabitEthernet1/0/4] ip address 192.168.3.1 24
[Device-GigabitEthernet1/0/4] quit
```

Configure IS-IS.

```
[Device] isis 1
```

```
[Device-isis-1] network-entity 00.0000.0000.0001.00
[Device-isis-1] is-level level-2
[Device-isis-1] cost-style wide
[Device-isis-1] address-family ipv4
[Device-isis-1-ipv4] quit
[Device-isis-1] quit
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] isis enable 1
[Device-GigabitEthernet1/0/1] quit
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] isis enable 1
[Device-GigabitEthernet1/0/2] quit
[Device] interface gigabitethernet 1/0/3
[Device-GigabitEthernet1/0/3] isis enable 1
[Device-GigabitEthernet1/0/3] quit
[Device] interface gigabitethernet 1/0/4
[Device-GigabitEthernet1/0/4] isis enable 1
[Device-GigabitEthernet1/0/4] quit
```

Create an IPv4 advanced ACL numbered 3000.

```
[Device] acl advanced 3000
```

Configure a rule to permit access from the President's office to the financial database server.

```
[Device-acl-ipv4-adv-3000] rule permit ip source 192.168.1.0 0.0.0.255 destination
192.168.0.100 0
```

Configure a rule to permit access from the Finance department to the database server during working hours.

```
[Device-acl-ipv4-adv-3000] rule permit ip source 192.168.2.0 0.0.0.255 destination
192.168.0.100 0 time-range work
```

Configure a rule to deny access to the financial database server.

```
[Device-acl-ipv4-adv-3000] rule deny ip source any destination 192.168.0.100 0
[Device-acl-ipv4-adv-3000] quit
```

Apply IPv4 advanced ACL 3000 to filter outgoing packets on interface GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] packet-filter 3000 outbound
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Verify that a PC in the Finance department can ping the database server during working hours. (All PCs in this example use Windows XP).

```
C:\> ping 192.168.0.100

Pinging 192.168.0.100 with 32 bytes of data:

Reply from 192.168.0.100: bytes=32 time=1ms TTL=255
Reply from 192.168.0.100: bytes=32 time<1ms TTL=255
Reply from 192.168.0.100: bytes=32 time<1ms TTL=255
Reply from 192.168.0.100: bytes=32 time<1ms TTL=255

Ping statistics for 192.168.0.100:
```

```
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 1ms, Average = 0ms
```

Verify that a PC in the Marketing department cannot ping the database server during working hours.

```
C:\> ping 192.168.0.100

Pinging 192.168.0.100 with 32 bytes of data:

Request timed out.
Request timed out.
Request timed out.
Request timed out.

Ping statistics for 192.168.0.100:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),
```

Display configuration and match statistics for IPv4 advanced ACL 3000 on the device during working hours.

```
[Device] display acl 3000
Advanced IPv4 ACL 3000, 3 rules,
ACL's step is 5
 rule 0 permit ip source 192.168.1.0 0.0.0.255 destination 192.168.0.100 0
 rule 5 permit ip source 192.168.2.0 0.0.0.255 destination 192.168.0.100 0 time-range work (Active)
 rule 10 deny ip destination 192.168.0.100 0
```

The output shows that rule 5 is active. Rule 5 and rule 10 have been matched four times as the result of the ping operations.

Configuration files

```
#
 sysname Device
#
isis 1
 is-level level-2
 cost-style wide
 network-entity 00.0000.0000.0001.00
 #
 address-family ipv4 unicast
#
interface GigabitEthernet1/0/1
 ip address 192.168.0.1 255.255.255.0
 isis enable 1
 packet-filter 3000 outbound
#
interface GigabitEthernet1/0/2
 port link-mode route
 combo enable copper
 ip address 192.168.1.1 255.255.255.0
 isis enable 1
```

```
#
interface GigabitEthernet1/0/3
 ip address 192.168.2.1 255.255.255.0
 isis enable 1
#
interface GigabitEthernet1/0/4
 ip address 192.168.3.1 255.255.255.0
 isis enable 1
#
 time-range work 08:00 to 18:00 working-day
#
acl advanced 3000
 rule 0 permit ip source 192.168.1.0 0.0.0.255 destination 192.168.0.100 0
 rule 5 permit ip source 192.168.2.0 0.0.0.255 destination 192.168.0.100 0 time-range work
 rule 10 deny ip destination 192.168.0.100 0#
```

Contents

QoS overview 1
QoS service models 1
Best-effort service model 1
IntServ model 1
DiffServ model 1
QoS techniques in a network 1
QoS processing flow in a device 2
QoS configuration approaches 3
Configuring a QoS policy 4
About QoS policies 4
QoS policy tasks at a glance 4
Defining a traffic class 4
Defining a traffic behavior 5
Defining a QoS policy 5
Applying the QoS policy 5
Application destinations 5
Restrictions and guidelines for applying a QoS policy 6
Applying the QoS policy to an Ethernet service instance 6
Applying the QoS policy to an interface 7
Applying the QoS policy to VLANs 7
Applying the QoS policy globally 8
Applying the QoS policy to a control plane 9
Applying the QoS policy in control-plane management view 9
Enabling SNMP notifications for QoS 10
Display and maintenance commands for QoS policies 10
Configuring priority mapping 13
About priority mapping 13
About priorities 13
Priority maps 13
Priority mapping configuration methods 14
Priority mapping process 14
Priority mapping tasks at a glance 15
Configuring a priority map 16
Configuring an uncolored priority map 16
Configuring a colored priority map 16
Configuring a port to trust packet priority for priority mapping 16
Changing the port priority of an interface 17
Display and maintenance commands for priority mapping 17
Priority mapping configuration examples 18
Example: Configuring a priority trust mode 18
Example: Configuring priority mapping tables and priority marking 20
Configuring traffic policing, GTS, and rate limit 25
About traffic policing, GTS, and rate limit 25
Traffic evaluation and token buckets 25
Traffic policing 26
GTS 27
Rate limit 28
Restrictions and guidelines: Traffic policing, GTS, and rate limit configuration 29
Configuring traffic policing 29
Configuring traffic policing by using the MQC approach 29
Configuring GTS 30
Configuring queue-based GTS 30
Configuring the rate limit 31
Configuring the rate limit for an interface 31

Display and maintenance commands for traffic policing, GTS, and rate limit ····31
Configuring congestion management ····33
About congestion management ····33
Cause, negative results, and countermeasure of congestion ····33
Congestion management methods ····33
congestion management tasks at a glance ····37
Configuring queuing on an interface ····37
Restrictions and guidelines for queuing configuration ····37
Configuring SP queuing ····37
Configuring WRR queuing ····37
Configuring WFQ queuing ····38
Configuring SP+WRR queuing ····38
Configuring SP+WFQ queuing ····39
Configuring CBQ ····39
Restrictions and guidelines for CBQ configuration ····39
Configure AF and the minimum guaranteed bandwidth ····40
Configuring EF and the maximum bandwidth ····41
Configuring SP ····42
Configuring WFQ ····43
Configuring a queue scheduling profile ····45
About queue scheduling profiles ····45
Restrictions and guidelines for queue scheduling profile configuration ····46
Configuring a queue scheduling profile ····46
Applying a queue scheduling profile ····46
Example: Configuring a queue scheduling profile ····46
Display and maintenance commands for congestion management ····47
Configuring congestion avoidance ····49
About congestion avoidance ····49
Tail drop ····49
RED and WRED ····49
Relationship between WRED and queuing mechanisms ····50
ECN ····50
WRED parameters ····52
Configuring and applying a queue-based WRED table to an interface ····53
Restrictions and guidelines ····53
Procedure ····53
Example: Configuring and applying a queue-based WRED table ····53
Display and maintenance commands for WRED ····54
Configuring traffic filtering ····56
About traffic filtering ····56
Procedure ····56
Traffic filtering configuration examples ····57
Example: Configuring traffic filtering ····57
Configuring priority marking ····59
About priority marking ····59
Configuring priority marking ····59
Configuring nesting ····61
About nesting ····61
Restrictions and guidelines: Nesting configuration ····61
Procedure ····61
Nesting configuration examples ····62
Example: Configuring nesting ····62
Configuring traffic redirecting ····66
About traffic redirecting ····66
Restrictions and guidelines: Traffic redirecting configuration ····66
Procedure ····66

Traffic redirecting configuration examples ····67
Example: Configuring traffic redirecting ····67
Configuring global CAR ····70
About global CAR····70
Aggregate CAR ····70
Hierarchical CAR····70
Restrictions and guidelines: Global CAR configuration····71
Configuring aggregate CAR ····71
Display and maintenance commands for global CAR····72
Global CAR configuration examples ····72
Example: Configuring MQC aggregate CAR····72
Configuring class-based accounting ····75
About class-based accounting ····75
Procedure····75
Class-based accounting configuration examples····76
Example: Configuring class-based accounting ····76
Setting an alarm threshold for protocol packets ····78
About an alarm threshold for protocol packets ····78
Restrictions and guidelines: Alarm threshold for protocol packets configuration ····78
Procedure····78
Display and maintenance commands for alarm threshold for protocol packets····78
Appendixes····79
Appendix A Acronyms····79
Appendix B Default priority maps····80
priority maps····80
Appendix C Introduction to packet precedence ····81
IP precedence and DSCP values····81
802.1p priority ····82

QoS overview

In data communications, Quality of Service (QoS) provides differentiated service guarantees for diversified traffic in terms of bandwidth, delay, jitter, and drop rate, all of which can affect QoS.

QoS manages network resources and prioritizes traffic to balance system resources.

The following section describes typical QoS service models and widely used QoS techniques.

QoS service models

This section describes several typical QoS service models.

Best-effort service model

The best-effort model is a single-service model. The best-effort model is not as reliable as other models and does not guarantee delay-free delivery.

The best-effort service model is the default model for the Internet and applies to most network applications. It uses the First In First Out (FIFO) queuing mechanism.

IntServ model

The integrated service (IntServ) model is a multiple-service model that can accommodate diverse QoS requirements. This service model provides the most granularly differentiated QoS by identifying and guaranteeing definite QoS for each data flow.

In the IntServ model, an application must request service from the network before it sends data. IntServ signals the service request with the RSVP. All nodes receiving the request reserve resources as requested and maintain state information for the application flow.

The IntServ model demands high storage and processing capabilities because it requires all nodes along the transmission path to maintain resource state information for each flow. This model is suitable for small-sized or edge networks. However, it is not suitable for large-sized networks, for example, the core layer of the Internet, where billions of flows are present.

DiffServ model

The differentiated service (DiffServ) model is a multiple-service model that can meet diverse QoS requirements. It is easy to implement and extend. DiffServ does not signal the network to reserve resources before sending data, as IntServ does.

QoS techniques in a network

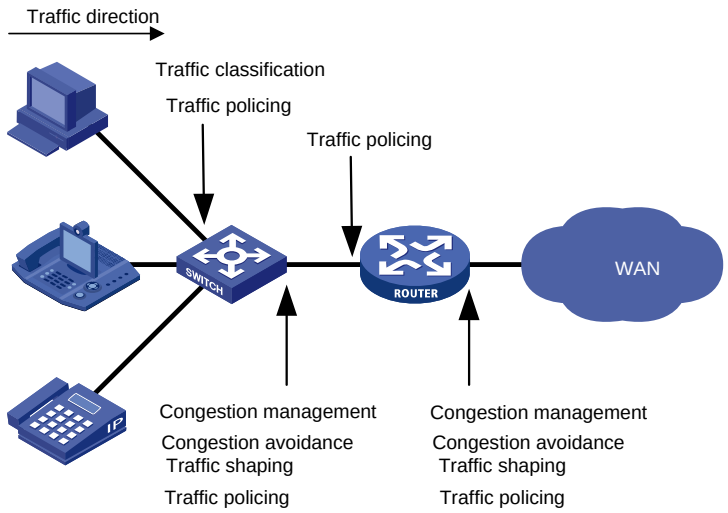
The QoS techniques include the following features:

- Traffic classification.
- Traffic policing.
- Traffic shaping.
- Rate limit.
- Congestion management.
- Congestion avoidance.

The following section briefly introduces these QoS techniques.

All QoS techniques in this document are based on the DiffServ model.

Figure 1 Position of the QoS techniques in a network



As shown in Figure 1, traffic classification, traffic shaping, traffic policing, congestion management, and congestion avoidance mainly implement the following functions:

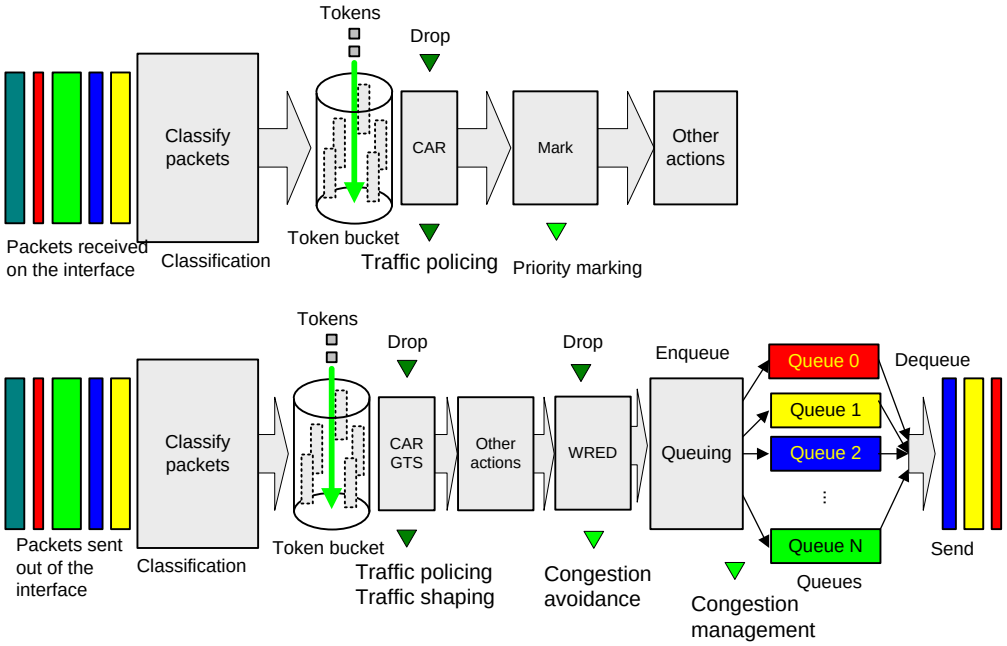
- **Traffic classification**—Uses match criteria to assign packets with the same characteristics to a traffic class. Based on traffic classes, you can provide differentiated services.
- **Traffic policing**—Polices flows and imposes penalties to prevent aggressive use of network resources. You can apply traffic policing to both incoming and outgoing traffic of a port.
- **Traffic shaping**—Adapts the output rate of traffic to the network resources available on the downstream device to eliminate packet drops. Traffic shaping usually applies to the outgoing traffic of a port.
- **Congestion management**—Provides a resource scheduling policy to determine the packet forwarding sequence when congestion occurs. Congestion management usually applies to the outgoing traffic of a port.
- **Congestion avoidance**—Monitors the network resource usage. It is usually applied to the outgoing traffic of a port. When congestion worsens, congestion avoidance reduces the queue length by dropping packets.

QoS processing flow in a device

Figure 2 briefly describes how the QoS module processes traffic.

1. Traffic classifier identifies and classifies traffic for subsequent QoS actions.
2. The QoS module takes various QoS actions on classified traffic as configured, depending on the traffic processing phase and network status. For example, you can configure the QoS module to perform the following operations:
   - Traffic policing for incoming traffic.
   - Traffic shaping for outgoing traffic.
   - Congestion avoidance before congestion occurs.
   - Congestion management when congestion occurs.

Figure 2 QoS processing flow



QoS configuration approaches

You can configure QoS by using the MQC approach or non-MQC approach.

In the modular QoS configuration (MQC) approach, you configure QoS service parameters by using QoS policies. A QoS policy defines QoS actions to take on different classes of traffic and can be applied to an object (such as an interface) to control traffic.

In the non-MQC approach, you configure QoS service parameters without using a QoS policy. For example, you can use the rate limit feature to set a rate limit on an interface without using a QoS policy.

Configuring a QoS policy

About QoS policies

A QoS policy has the following components:

- **Traffic class**—Defines criteria to match packets.
- **Traffic behavior**—Defines QoS actions to take on matching packets.

By associating a traffic class with a traffic behavior, a QoS policy can perform the QoS actions on matching packets.

A QoS policy can have multiple class-behavior associations.

QoS policy tasks at a glance

To configure a QoS policy, perform the following tasks:

1. Defining a traffic class
2. Defining a traffic behavior
3. Defining a QoS policy
4. Applying the QoS policy
   - Applying the QoS policy to an Ethernet service instance
   - Applying the QoS policy to an interface
   - Applying the QoS policy to VLANs
   - Applying the QoS policy globally
   - Applying the QoS policy to a control plane
   - Applying the QoS policy in control-plane management view

Defining a traffic class

Restrictions and guidelines

A match criterion without IPv6 attributes cannot match IPv6 packets.

Procedure

1. Enter system view.
   system-view
2. Create a traffic class and enter traffic class view.
   traffic classifier *classifier-name* [**operator** { **and** | **or** }]
3. (Optional.) Configure a description for the traffic class.
   description *text*

   By default, no description is configured for a traffic class.
4. Configure a match criterion.
   if-match *match-criteria*

   By default, no match criterion is configured.

   For more information, see the **if-match** command in *ACL and QoS Command Reference*.

Defining a traffic behavior

1. Enter system view.

   system-view
2. Create a traffic behavior and enter traffic behavior view.

   traffic behavior *behavior-name*
3. Configure an action in the traffic behavior.

   By default, no action is configured for a traffic behavior.

   For more information about configuring an action, see the subsequent chapters for traffic policing, traffic filtering, priority marking, class-based accounting, and so on.

Defining a QoS policy

1. Enter system view.

   system-view
2. Create a QoS policy and enter QoS policy view.

   qos [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy** *policy-name*

Parameter	Description
ipv6-matching	Represents an IPv6-matching QoS policy.
accounting	Represents an accounting-type QoS policy.
mirroring	Represents a mirroring-type QoS policy.
remarking	Represents a marking-type QoS policy.
tcp-erspan	Represents a TCP ERSPAN-type QoS policy.
exact-matching	Represents an exact-matching QoS policy.

3. Associate a traffic class with a traffic behavior to create a class-behavior association in the QoS policy.

   classifier *classifier-name* **behavior** *behavior-name* [**insert-before** *before-classifier-name*]

   By default, a traffic class is not associated with a traffic behavior.

Applying the QoS policy

Application destinations

You can apply a QoS policy to the following destinations:

- **Ethernet service instance**—The QoS policy can be applied to the traffic received on the Ethernet service instance.
- **Interface**—The QoS policy can be applied to the traffic sent or received on the interface.
- **VLAN**—The QoS policy can be applied to the traffic sent or received on all ports in the VLAN.

- **Globally**—The QoS policy can be applied to the traffic sent or received on all ports.
- **Control plane**—The QoS policy can be applied to the traffic received on the control plane.
- **Management interface control plane**—The QoS policy can be applied to the traffic sent from the management interface to the control plane.

Restrictions and guidelines for applying a QoS policy

You can modify traffic classes, traffic behaviors, and class-behavior associations in a QoS policy even after it is applied (except that it is applied to a user profile). If a traffic class uses an ACL for traffic classification, you can delete or modify the ACL.

If a traffic behavior is configured with multiple actions and one of these actions does not take effect, the whole class-behavior association does not take effect.

A QoS policy applied in control-plane management view does not support matching the following protocols in a traffic class: ARP-snooping, bfd, dhcp, dhcp-snooping, dhcp6, dot1x, lacp, mvrp, oam, portal, pvst, rrpp, smart-link, stp, udp-helper, lldp.

When a QoS policy containing a CAR action is applied on an IRF fabric, the traffic matching the QoS policy might enter the IRF fabric through interfaces on different IRF member devices. In this case, the actual rate limit that takes effect is the sum of the CIR and PIR in the CAR action multiplied by the number of involved IRF member devices by default.

Applying the QoS policy to an Ethernet service instance

Restrictions and guidelines

For configuration commands for Ethernet service instances, see MPLS L2VPN or VPLS in *MPLS Command Reference* or see *VXLAN Command Reference*.

A maximum of five QoS policies (one for each type) can be applied to an Ethernet service instance. Different actions can be taken on the same traffic class if QoS policies of different types are applied to an Ethernet service instance. This rule applies to the devices that support the five QoS policy types: IPv6-matching, accounting, marking, mirroring, and generic.

If you specify the **longest-match** keyword, the system first matches the traffic class with a source or destination address with the longest mask and takes the actions in the associated traffic behavior. Suppose three class-behavior associations are configured in a QoS policy. ACL 3001 in the traffic class of the first class-behavior association is not configured with a source or destination address. ACL 3002 in the traffic class of the second class-behavior association is configured with destination address 10.10.10.0/24. ACL 3003 in the traffic class of the third class-behavior association is configured with destination address 10.10.10.10/32. The match order of the three class-behavior associations is the third class-behavior association, the second class-behavior association, the first class-behavior association. If both source and destination addresses exist in ACLs of different traffic classes, the system matches destination addresses before matching source addresses.

Procedure

1. Enter system view.

   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   - Enter Layer 2 Ethernet interface view:
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view:
     interface bridge-aggregation *interface-number*
3. Create an Ethernet service instance and enter Ethernet service instance view.

   service-instance *instance-id*

4. Apply the QoS policy to the Ethernet service instance.

   qos apply [**ipv6-matching** | **accounting** | **mirroring** | **remarking**] **policy** *policy-name* **inbound** [**longest-match**]

   By default, no QoS policy is applied to an Ethernet service instance.

Applying the QoS policy to an interface

Restrictions and guidelines

A QoS policy can be applied to multiple interfaces. However, only one QoS policy can be applied to one direction (inbound or outbound) of an interface.

A maximum of seven QoS policies (one for each type) can be applied to one direction of an interface. Different actions can be taken on the same traffic class if QoS policies of different types are applied to an interface.

The QoS policy applied to the outgoing traffic on an interface does not regulate local packets. Local packets refer to critical protocol packets sent by the local system for operation maintenance. The most common local packets include link maintenance, RIP, and SSH packets.

If you specify the **longest-match** keyword, the system first matches the traffic class with a source or destination address with the longest mask and takes the actions in the associated traffic behavior. Suppose three class-behavior associations are configured in a QoS policy. ACL 3001 in the traffic class of the first class-behavior association is not configured with a source or destination address. ACL 3002 in the traffic class of the second class-behavior association is configured with destination address 10.10.10.0/24. ACL 3003 in the traffic class of the third class-behavior association is configured with destination address 10.10.10.10/32. The match order of the three class-behavior associations is the third class-behavior association, the second class-behavior association, the first class-behavior association. If both source and destination addresses exist in ACLs of different traffic classes, the system matches destination addresses before matching source addresses.

When a QoS policy is applied to the inbound direction of an interface, it cannot match the destination MAC address or customer VLAN field of IPv4 VXLAN packets.

An ACL rule containing the **vpn-instance** parameter is not supported in a QoS policy applied to an interface.

A QoS policy cannot be applied to the outbound direction of a VSI interface.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Apply the QoS policy to the interface.

   qos apply [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy** *policy-name* { **inbound** | **outbound** [**longest-match**]

   By default, no QoS policy is applied to an interface.

Applying the QoS policy to VLANs

About this task

You can apply a QoS policy to VLANs to regulate the traffic on all ports of the VLANs.

Restrictions and guidelines

QoS policies cannot be applied to dynamic VLANs, including VLANs created by GVRP.

If you specify the **longest-match** keyword, the system first matches the traffic class with a source or destination address with the longest mask and takes the actions in the associated traffic behavior. Suppose three class-behavior associations are configured in a QoS policy. ACL 3001 in the traffic class of the first class-behavior association is not configured with a source or destination address. ACL 3002 in the traffic class of the second class-behavior association is configured with destination address 10.10.10.0/24. ACL 3003 in the traffic class of the third class-behavior association is configured with destination address 10.10.10.10/32. The match order of the three class-behavior associations is the third class-behavior association, the second class-behavior association, the first class-behavior association. If both source and destination addresses exist in ACLs of different traffic classes, the system matches destination addresses before matching source addresses.

When you apply a QoS policy to VLANs, the QoS policy is applied to the specified VLANs on all IRF member devices. If the hardware resources of an IRF member device are insufficient, applying a QoS policy to VLANs might fail on the IRF member device. The system does not automatically roll back the QoS policy configuration already applied to other IRF member devices. To ensure consistency, use the **undo qos vlan-policy vlan** command to manually remove the QoS policy configuration applied to them.

Procedure

1. Enter system view.

   system-view

2. Apply the QoS policy to VLANs.

   qos vlan-policy *policy-name* **vlan** *vlan-id-list* { **inbound** | **outbound** } [**longest-match**]

   By default, no QoS policy is applied to a VLAN.

Applying the QoS policy globally

About this task

You can apply a QoS policy globally to the inbound or outbound direction of all ports.

Restrictions and guidelines

If you do not specify the **longest-match** keyword, the system matches traffic against traffic classes in the order class-behavior associations are configured. When the traffic matches a traffic class, the matching process stops, and the actions in the associated traffic behavior are taken on the matching traffic.

If you specify the **longest-match** keyword, the system first matches the traffic class with a source or destination address with the longest mask and takes the actions in the associated traffic behavior. Suppose three class-behavior associations are configured in a QoS policy. ACL 3001 in the traffic class of the first class-behavior association is not configured with a source or destination address. ACL 3002 in the traffic class of the second class-behavior association is configured with destination address 10.10.10.0/24. ACL 3003 in the traffic class of the third class-behavior association is configured with destination address 10.10.10.10/32. The match order of the three class-behavior associations is the third class-behavior association, the second class-behavior association, the first class-behavior association. If both source and destination addresses exist in ACLs of different traffic classes, the system matches destination addresses before matching source addresses.

If the hardware resources of an IRF member device are insufficient, applying a QoS policy globally might fail on the IRF member device. The system does not automatically roll back the QoS policy configuration already applied to other IRF member devices. To ensure consistency, you must use the **undo qos apply policy global** command to manually remove the QoS policy configuration applied to them.

Procedure

1. Enter system view.

   system-view

2. Apply the QoS policy globally.

   qos apply [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy** *policy-name* **global** { **inbound** | **outbound** } [**longest-match**]

   By default, no QoS policy is applied globally.

Applying the QoS policy to a control plane

About this task

A device provides the user plane and the control plane.

- **User plane**—The units at the user plane are responsible for receiving, transmitting, and switching (forwarding) packets, such as various dedicated forwarding chips. They deliver super processing speeds and throughput.
- **Control plane**—The units at the control plane are processing units running most routing and switching protocols. They are responsible for protocol packet resolution and calculation, such as CPUs. Compared with user plane units, the control plane units allow for great packet processing flexibility but have lower throughput.

When the user plane receives packets that it cannot recognize or process, it transmits them to the control plane. If the transmission rate exceeds the processing capability of the control plane, the control plane will be busy handling undesired packets. As a result, the control plane will fail to handle legitimate packets correctly or timely. As a result, protocol performance is affected.

To address this problem, apply a QoS policy to the control plane to take QoS actions, such as traffic filtering or traffic policing, on inbound traffic. This ensures that the control plane can correctly receive, transmit, and process packets.

A predefined control plane QoS policy uses the protocol type or protocol group type to identify the type of packets sent to the control plane. You can use protocol types or protocol group types in **if-match control-plane** commands in traffic class view for traffic classification. Then you can reconfigure traffic behaviors for these traffic classes as required. You can use the **display qos policy control-plane pre-defined** command to display predefined control plane QoS policies.

Procedure

1. Enter system view.

   system-view

2. Enter control plane view.

   control-plane slot *slot-number*

3. Apply the QoS policy to the control plane.

   qos apply policy *policy-name* **inbound**

   By default, no QoS policy is applied to a control plane.

Applying the QoS policy in control-plane management view

About this task

If the rate of packets from the management interface to the control plane exceeds the processing capability, the control plane will fail to handle the packets correctly or timely. As a result, protocol performance is affected.

This feature allows you to rate limit the packets sent from the management interface to the control plane. This ensures that the control plane can correctly receive, transmit, and process packets from the management interface.

By default, a predefined QoS policy is applied in control-plane management view. To display the predefined QoS policy, use the **display qos policy control-plane management pre-defined** command. The predefined QoS policy uses the protocol type or protocol group type to identify the type of packets sent from the management interface to the control plane. You can use protocol types or protocol group types in **if-match** commands in traffic class view for traffic classification. Then, you can reconfigure traffic behaviors for these traffic classes as required.

Procedure

1. Enter system view.
   system-view
2. Enter control-plane management view.
   control-plane management
3. Apply the QoS policy.
   qos apply policy *policy-name* **inbound**
   By default, no QoS policy is applied in control-plane management view.

Enabling SNMP notifications for QoS

About this task

To report critical QoS events to an SNMP NMS, enable SNMP notifications for QoS.

For QoS notifications to be sent correctly, you must also configure SNMP on the device. For more information about configuring SNMP, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for QoS.
   snmp-agent trap enable qos
   By default, SNMP notifications are disabled for QoS.

Display and maintenance commands for QoS policies

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display QoS policy configuration.	**display qos policy user-defined** [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] [*policy-name* [**classifier** *classifier-name*]] [**slot** *slot-number*]
Display information about QoS policies applied to the control plane.	**display qos policy control-plane slot** *slot-number*

Task	Command
Display information about QoS policies applied in control-plane management view.	**display qos policy control-plane management**
Display information about the predefined QoS policy applied in control-plane management view.	**display qos policy control-plane management pre-defined**
Display information about the predefined QoS policy applied to the control plane.	**display qos policy control-plane pre-defined** [**slot** *slot-number*]
Display information about QoS policies applied globally.	**display qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy global** [**slot** *slot-number*] [**inbound** | **outbound**]
Display information about QoS policies applied to interfaces.	**display qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy interface** [*interface-type interface-number*] [**inbound** | **outbound**]
Display information about QoS policies applied to Ethernet service instances.	**display qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking**] **policy l2vpn-ac** [**interface** *interface-type interface-number* [**service-instance** *instance-id*] [**slot** *slot-number*]] [**inbound**]
Display information about QoS policies applied to VLANs.	**display qos vlan-policy** { **name** *policy-name* | **vlan** [*vlan-id*] } [**slot** *slot-number*] [**inbound** | **outbound**]
Display QoS and ACL resource usage.	**display qos-acl resource** [**slot** *slot-number*]
Display traffic behavior configuration.	**display traffic behavior user-defined** [*behavior-name*] [**slot** *slot-number*]
Display traffic class configuration.	**display traffic classifier user-defined** [*classifier-name*] [**slot** *slot-number*]
Display statistics for the QoS policy applied to a control plane.	**display qos policy control-plane slot** *slot-number* **rule-statistics**
Display statistics for QoS policies applied globally.	**display qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy global** [**slot** *slot-number*] [**inbound** | **outbound**] **rule-statistics**
Display statistics for QoS policies applied to interfaces.	**display qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy interface** [*interface-type interface-number*] [**slot** *slot-number*] [**inbound** | **outbound**] **rule-statistics**
Display statistics for QoS policies applied to Ethernet service instances.	**display qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking**] **policy l2vpn-ac** [**interface** *interface-type interface-number*

Task	Command
	[**service-instance** *instance-id*] [**slot** *slot-number*]] [**inbound**] **rule-statistics**
Display statistics for QoS policies applied to VLANs.	**display qos vlan-policy vlan** [*vlan-id*] [**slot** *slot-number*] [**inbound** | **outbound**] **rule-statistics**
Clear the statistics for the QoS policy applied to the control plane.	**reset qos policy control-plane slot** *slot-number*
Clear the statistics for the QoS policy applied in control-plane management view.	**reset qos policy control-plane management**
Clear the statistics for a QoS policy applied globally.	**reset qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking** | **tcp-erspan** | **exact-matching**] **policy global** [**inbound** | **outbound**]
Clear the statistics about QoS policies applied to Ethernet service instances.	**reset qos** [**ipv6-matching** | **accounting** | **mirroring** | **remarking**] **policy l2vpn-ac** [**interface** *interface-type interface-number* [**service-instance** *instance-id*]] [**inbound**]
Clear the statistics of the QoS policy applied in a certain direction of a VLAN.	**reset qos vlan-policy** [**vlan** *vlan-id*] [**inbound** | **outbound**]

Configuring priority mapping

About priority mapping

When a packet arrives, a device assigns a set of QoS priority parameters to the packet based on either of the following:

- A priority field carried in the packet.
- The port priority of the incoming port.

This process is called priority mapping. During this process, the device can modify the priority of the packet according to the priority mapping rules. The set of QoS priority parameters decides the scheduling priority and forwarding priority of the packet.

Priority mapping is implemented with priority maps and involves the following priorities:

- 802.11e priority.
- 802.1p priority.
- DSCP.
- EXP.
- IP precedence.
- Local precedence.
- Drop priority.

About priorities

Priorities include the following types: priorities carried in packets, and priorities locally assigned for scheduling only.

Packet-carried priorities include 802.1p priority and DSCP precedence. These priorities have global significance and affect the forwarding priority of packets across the network. For more information about these priorities, see "Appendixes."

Locally assigned priorities only have local significance. They are assigned by the device only for scheduling. These priorities include the local precedence, drop priority, and user priority, as follows:

- **Local precedence**—Used for queuing. A local precedence value corresponds to an output queue. A packet with higher local precedence is assigned to a higher priority output queue to be preferentially scheduled.
- **Drop priority**—Used for making packet drop decisions. Packets with the highest drop priority are dropped preferentially.

The device supports only local precedence and drop priority for scheduling.

Priority maps

The device provides various types of priority maps. By looking through a priority map, the device decides which priority value to assign to a packet for subsequent packet processing.

The default priority maps (as shown in Appendix B Default priority maps) are available for priority mapping. They are adequate in most cases. If a default priority map cannot meet your requirements, you can modify the priority map as required.

Priority mapping configuration methods

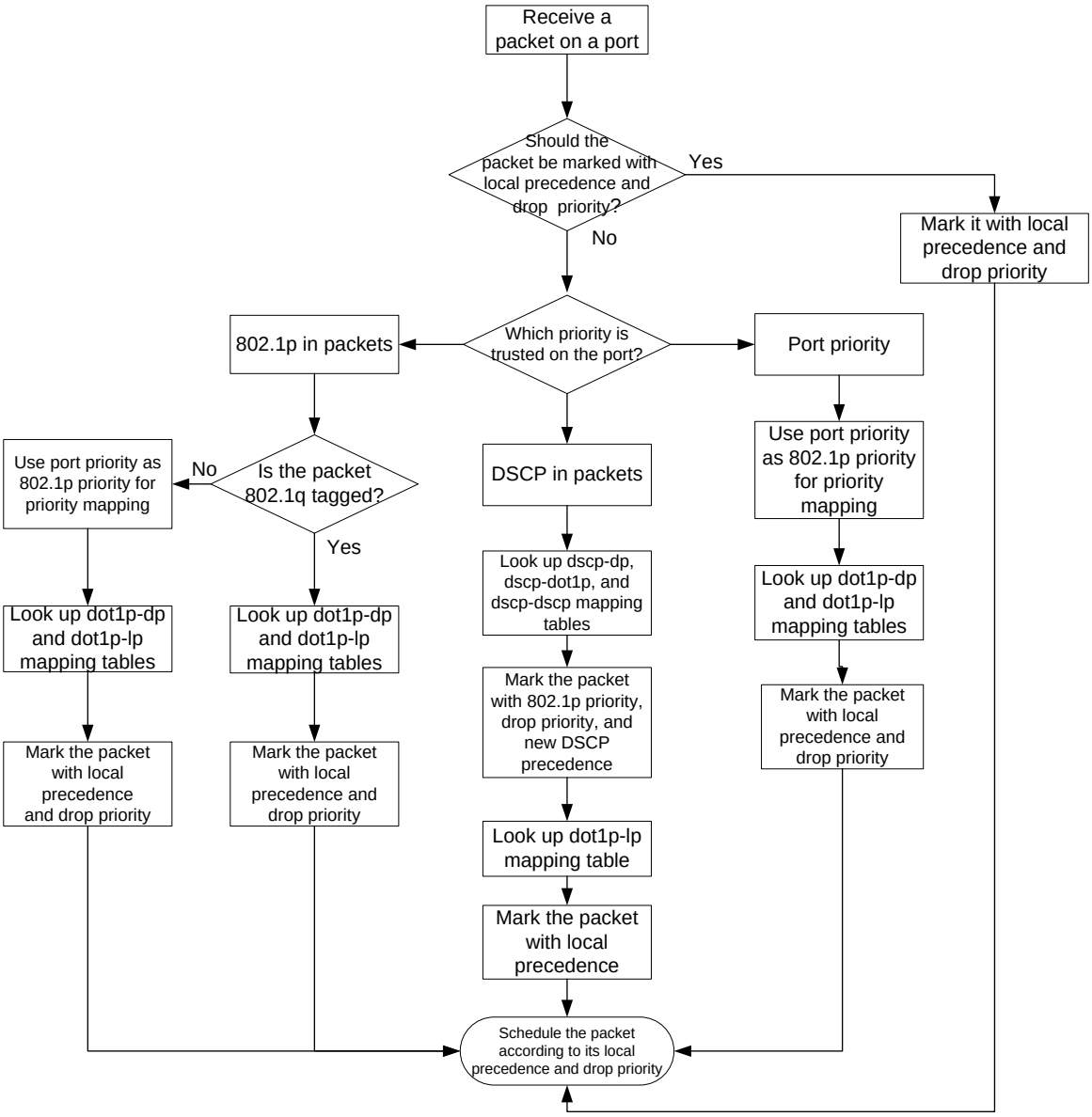
You can configure priority mapping by using any of the following methods:

- **Configuring priority trust mode**—In this method, you can configure a port to look up a trusted priority type (802.1p, for example) in incoming packets in the priority maps. Then, the system maps the trusted priority to the target priority types and values.
- **Changing port priority**—If no packet priority is trusted, the port priority of the incoming port is used. By changing the port priority of a port, you change the priority of the incoming packets on the port.

Priority mapping process

On receiving an Ethernet packet on a port, the switch marks the scheduling priorities (local precedence and drop precedence) for the Ethernet packet. This procedure is done according to the priority trust mode of the receiving port and the 802.1Q tagging status of the packet, as shown in Figure 3.

Figure 3 Priority mapping process for an Ethernet packet



For information about priority marking, see "Configuring priority marking."

Priority mapping tasks at a glance

To configure priority mapping, perform the following tasks:

1. (Optional.) Configuring a priority map
   - Configuring an uncolored priority map
   - Configuring a colored priority map
2. Configure a priority mapping method:
   - Configuring a port to trust packet priority for priority mapping
   - Changing the port priority of an interface

Configuring a priority map

Configuring an uncolored priority map

1. Enter system view.

   system-view

2. Enter priority map view.

   qos map-table { **dot1p-dp** | **dot1p-lp** | **dscp-dp** | **dscp-dscp** | **dscp-lp** | **exp-lp** | **lp-exp** }

3. Configure mappings for the priority map.

   import *import-value-list* **export** *export-value*

   By default, the default priority maps are used. For more information, see "Appendix B Default priority maps."

   If you execute this command multiple times, the most recent configuration takes effect.

Configuring a colored priority map

1. Enter system view.

   system-view

2. Enter colored priority map view.

   qos map-table color { **green** | **yellow** | **red** } **lp-tunnel-dscp**

3. Configure mappings for the colored priority map.

   import *import-value-list* **export** *export-value*

   By default, the default colored priority maps are used. You can use the **display qos map-table color** command to view the default values.

   If you execute this command multiple times, the most recent configuration takes effect.

Configuring a port to trust packet priority for priority mapping

About this task

You can configure the device to trust a particular priority field carried in packets for priority mapping on ports or globally. When you configure the trusted packet priority type on an interface, use the following available keywords:

- **dot1p**—Uses the 802.1p priority of received packets for mapping.
- **dscp**—Uses the DSCP precedence of received IP packets for mapping.

Restrictions and guidelines

For a VXLAN tunnel interface to trust the DSCP priority in the outer IP header of VXLAN packets, you must execute both of the following commands:

- **qos trust tunnel-dscp** (in system view).
- **qos trust dscp** (in interface view for the physical interface of the VXLAN tunnel interface).

The **qos trust tunnel-dscp** command is not supported when the system operating mode is standard. For more information about system operating modes, see device management in *Fundamentals Configuration Guide*.

The term "interface" in this section collectively refers to Layer 2 and Layer 3 Ethernet interfaces. You can use the **port link-mode** command to configure an Ethernet port as a Layer 2 or Layer 3 interface (see *Interface Configuration Guide*).

If you configure both either the following settings and the **qos trust dscp** command on the same interface, only the **qos trust dscp** command takes effect:

- DSCP port priority (**qos priority dscp** *priority-value*).
- DSCP marking action in a QoS policy.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the trusted packet priority type.
   qos trust { **dot1p** | **dscp** }
   By default, an interface does not trust any packet priority.
4. Return to system view.
   quit
5. (Optional.) Configure the global priority trust mode for VXLAN packets.
   qos trust tunnel-dscp
   By default, the global priority trust mode for VXLAN packets is not configured.

Changing the port priority of an interface

About this task

If an interface does not trust any packet priority, the device uses its port priority to look for priority parameters for the incoming packets. By changing port priority, you can prioritize traffic received on different interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the port priority of the interface.
   qos priority *priority-value*
   The default setting is 0.

Display and maintenance commands for priority mapping

Execute **display** commands in any view.

Task	Command
Display priority map configuration.	**display qos map-table** [**dot1p-dp** | **dot1p-lp** | **dscp-dp** | **dscp-dscp** | **dscp-lp** | **exp-lp** | **lp-exp**]
Display colored priority map configuration.	**display qos map-table color** [**green** | **yellow** | **red**] [**lp-tunnel-dscp**]
Display the trusted packet priority type on a port.	**display qos trust interface** [*interface-type interface-number*]

Priority mapping configuration examples

Example: Configuring a priority trust mode

Network configuration

As shown in Figure 4:

- The 802.1p priority of traffic from Device A to Device C is 3.
- The 802.1p priority of traffic from Device B to Device C is 1.

Configure Device C to preferentially process packets from Device A to the server when GigabitEthernet 1/0/3 of Device C is congested.

Figure 4 Network diagram

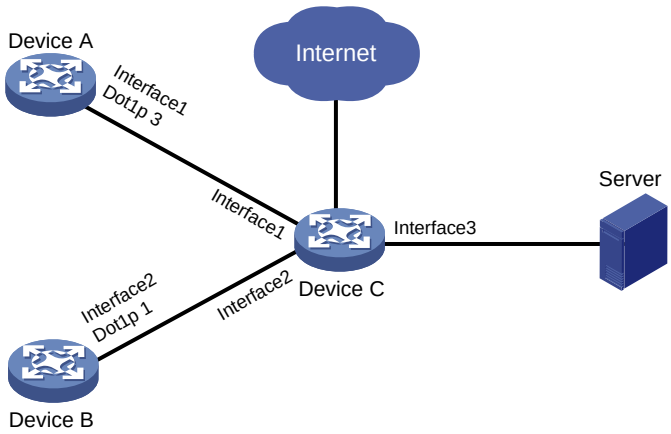


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

(Method 1) Configure Device C to trust packet priority

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to trust the 802.1p priority for priority mapping.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] qos trust dot1p
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] qos trust dot1p
[DeviceC-GigabitEthernet1/0/2] quit
```

(Method 2) Configure Device C to trust port priority

Assign port priority to GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2. Make sure the following requirements are met:

- The priority of GigabitEthernet 1/0/1 is higher than that of GigabitEthernet 1/0/2.
- No trusted packet priority type is configured on GigabitEthernet 1/0/1 or GigabitEthernet 1/0/2.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] qos priority 3
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] qos priority 1
[DeviceC-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Use the **`display qos trust interface`** command to verify that the configuration has been performed successfully.

Configuration files

- Method 1:

```
#
 sysname DeviceC
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 qos trust dot1p
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 qos trust dot1p
#
return
```

- Method 2:

```
#
 sysname DeviceC
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 qos priority 3
#
interface GigabitEthernet1/0/2
```

```
 port link-mode bridge
 qos priority 1
#
return
```

Example: Configuring priority mapping tables and priority marking

Network configuration

As shown in Figure 5:

- The Marketing department connects to GigabitEthernet 1/0/1 of Device, which sets the 802.1p priority of traffic from the Marketing department to 3.
- The R&D department connects to GigabitEthernet 1/0/2 of Device, which sets the 802.1p priority of traffic from the R&D department to 4.
- The Management department connects to GigabitEthernet 1/0/3 of Device, which sets the 802.1p priority of traffic from the Management department to 5.

Configure port priority, 802.1p-to-local mapping table, and priority marking to implement the plan as described in Table 2.

Table 2 Configuration plan

Traffic destination	Traffic priority order	Queuing plan		
		Traffic source	Output queue	Queue priority
Public servers	R&D department > Management department > Marketing department	R&D department	6	High
		Management department	4	Medium
		Marketing department	2	Low
Internet	Management department > Marketing department > R&D department	R&D department	2	Low
		Management department	6	High
		Marketing department	4	Medium

Figure 5 Network diagram

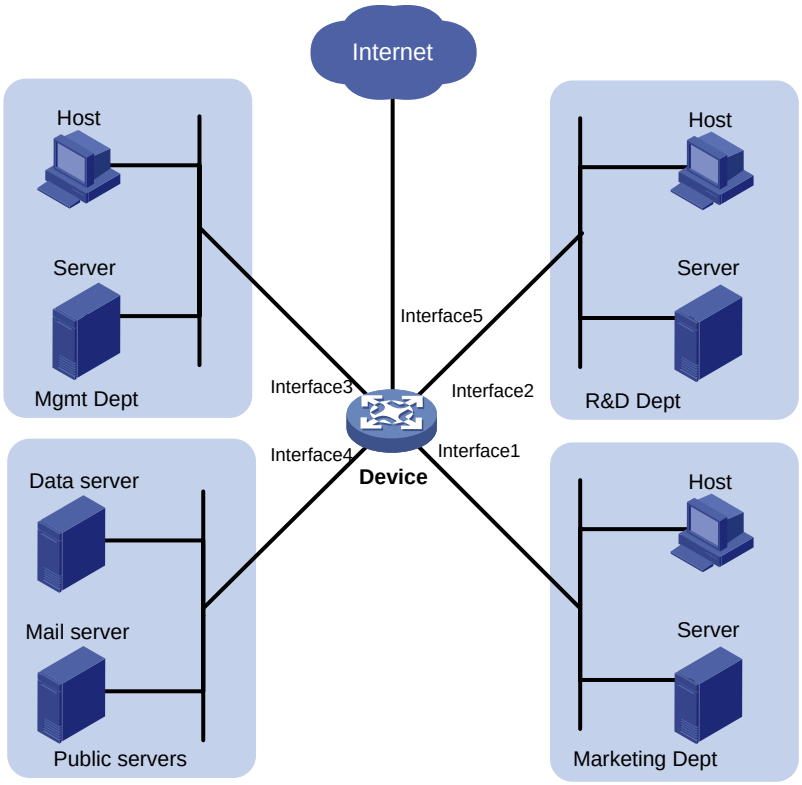


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure trusting port priority:

   # Set the port priority of GigabitEthernet 1/0/1 to 3.

   ```
   <Sysname> system-view
   [Sysname] sysname Device
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] qos priority 3
   [Device-GigabitEthernet1/0/1] quit
   ```

   # Set the port priority of GigabitEthernet 1/0/2 to 4.

   ```
   [Device] interface gigabitethernet 1/0/2
   [Device-GigabitEthernet1/0/2] qos priority 4
   [Device-GigabitEthernet1/0/2] quit
   ```

   # Set the port priority of GigabitEthernet 1/0/3 to 5.

   ```
   [Device] interface gigabitethernet 1/0/3
   [Device-GigabitEthernet1/0/3] qos priority 5
   [Device-GigabitEthernet1/0/3] quit
   ```

2. Configure the 802.1p-to-local mapping table to map 802.1p priority values 3, 4, and 5 to local precedence values 2, 6, and 4.

   This guarantees the R&D department, Management department, and Marketing department decreased priorities to access the public servers.

```
[Device] qos map-table dot1p-lp
[Device-maptbl-dot1p-lp] import 3 export 2
[Device-maptbl-dot1p-lp] import 4 export 6
[Device-maptbl-dot1p-lp] import 5 export 4
[Device-maptbl-dot1p-lp] quit
```

3. Configure priority marking to mark the packets from Management department, Marketing department, and R&D department to the Internet with 802.1p priority values 4, 5, and 3.

   This guarantees the Management department, Marketing department, and R&D department decreased priorities to access the Internet.

   # Create ACL 3000, and configure a rule to match HTTP packets.

```
[Device] acl advanced 3000
[Device-acl-adv-3000] rule permit tcp destination-port eq 80
[Device-acl-adv-3000] quit
```

   # Create a traffic class named **http**, and use ACL 3000 as a match criterion.

```
[Device] traffic classifier http
[Device-classifier-http] if-match acl 3000
[Device-classifier-http] quit
```

   # Create a traffic behavior named **admin**, and configure a marking action for the Management department.

```
[Device] traffic behavior admin
[Device-behavior-admin] remark dot1p 4
[Device-behavior-admin] quit
```

   # Create a QoS policy named **admin**, and associate traffic class **http** with traffic behavior **admin** in QoS policy **admin**.

```
[Device] qos policy admin
[Device-qospolicy-admin] classifier http behavior admin
[Device-qospolicy-admin] quit
```

   # Apply QoS policy **admin** to the inbound direction of GigabitEthernet 1/0/3.

```
[Device] interface gigabitethernet 1/0/3
[Device-GigabitEthernet1/0/3] qos apply policy admin inbound
```

   # Create a traffic behavior named **market**, and configure a marking action for the Marketing department.

```
[Device] traffic behavior market
[Device-behavior-market] remark dot1p 5
[Device-behavior-market] quit
```

   # Create a QoS policy named **market**, and associate traffic class **http** with traffic behavior **market** in QoS policy **market**.

```
[Device] qos policy market
[Device-qospolicy-market] classifier http behavior market
[Device-qospolicy-market] quit
```

   # Apply QoS policy **market** to the inbound direction of GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] qos apply policy market inbound
```

   # Create a traffic behavior named **rd**, and configure a marking action for the R&D department.

```
[Device] traffic behavior rd
```

```
[Device-behavior-rd] remark dot1p 3
[Device-behavior-rd] quit
```

Create a QoS policy named **rd**, and associate traffic class **http** with traffic behavior **rd** in QoS policy **rd**.

```
[Device] qos policy rd
[Device-qospolicy-rd] classifier http behavior rd
[Device-qospolicy-rd] quit
```

Apply QoS policy **rd** to the inbound direction of GigabitEthernet 1/0/2.

```
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] qos apply policy rd inbound
[Device-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Use the **display qos map-table** and **display qos policy interface** commands to verify that the configuration has been performed successfully.

Configuration files

```
#
 sysname Device
#
qos map-table dot1p-lp
 import 3 export 2
 import 4 export 6
 import 5 export 4
#
traffic classifier http operator and
 if-match acl 3000
#
traffic behavior admin
 remark dot1p 4
#
traffic behavior market
 remark dot1p 5
#
traffic behavior rd
 remark dot1p 3
#
qos policy admin
 classifier http behavior admin
#
qos policy market
 classifier http behavior market
#
qos policy rd
 classifier http behavior rd
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 qos priority 3
```

```
 qos apply policy market inbound
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 qos priority 4
 qos apply policy rd inbound
#
interface GigabitEthernet1/0/3
 port link-mode bridge
 qos priority 5
 qos apply policy admin inbound
#
acl advanced 3000
 rule 0 permit tcp destination-port eq www
#
return
```

Configuring traffic policing, GTS, and rate limit

About traffic policing, GTS, and rate limit

Traffic limit helps assign network resources (including bandwidth) and increase network performance. For example, you can configure a flow to use only the resources committed to it in a certain time range. This avoids network congestion caused by burst traffic.

Traffic policing, Generic Traffic Shaping (GTS), and rate limit control the traffic rate and resource usage according to traffic specifications. You can use token buckets for evaluating traffic specifications.

Traffic evaluation and token buckets

Token bucket features

A token bucket is analogous to a container that holds a certain number of tokens. Each token represents a certain forwarding capacity. The system puts tokens into the bucket at a constant rate. When the token bucket is full, the extra tokens cause the token bucket to overflow.

Evaluating traffic with the token bucket

A token bucket mechanism evaluates traffic by looking at the number of tokens in the bucket. If the number of tokens in the bucket is enough for forwarding the packets:

- The traffic conforms to the specification (called conforming traffic).
- The corresponding tokens are taken away from the bucket.

Otherwise, the traffic does not conform to the specification (called excess traffic).

A token bucket has the following configurable parameters:

- Mean rate at which tokens are put into the bucket, which is the permitted average rate of traffic. It is usually set to the committed information rate (CIR).
- Burst size or the capacity of the token bucket. It is the maximum traffic size permitted in each burst. It is usually set to the committed burst size (CBS). The set burst size must be greater than the maximum packet size.

Each arriving packet is evaluated.

Complicated evaluation

You can set two token buckets, bucket C and bucket E, to evaluate traffic in a more complicated environment and achieve more policing flexibility. The following are main mechanisms used for complicated evaluation:

- **Single rate two color**—Uses one token bucket and the following parameters:
  - **CIR**—Rate at which tokens are put into bucket C. It sets the average packet transmission or forwarding rate allowed by bucket C.
  - **CBS**—Size of bucket C, which specifies the transient burst of traffic that bucket C can forward.

  When a packet arrives, the following rules apply:
  - If bucket C has enough tokens to forward the packet, the packet is colored green.
  - Otherwise, the packet is colored red.
- **Single rate three color**—Uses two token buckets and the following parameters:

- **CIR**—Rate at which tokens are put into bucket C. It sets the average packet transmission or forwarding rate allowed by bucket C.
- **CBS**—Size of bucket C, which specifies the transient burst of traffic that bucket C can forward.
- **EBS**—Size of bucket E minus size of bucket C, which specifies the transient burst of traffic that bucket E can forward. The EBS cannot be 0. The size of E bucket is the sum of the CBS and EBS.

When a packet arrives, the following rules apply:

- If bucket C has enough tokens, the packet is colored green.
- If bucket C does not have enough tokens but bucket E has enough tokens, the packet is colored yellow.
- If neither bucket C nor bucket E has sufficient tokens, the packet is colored red.

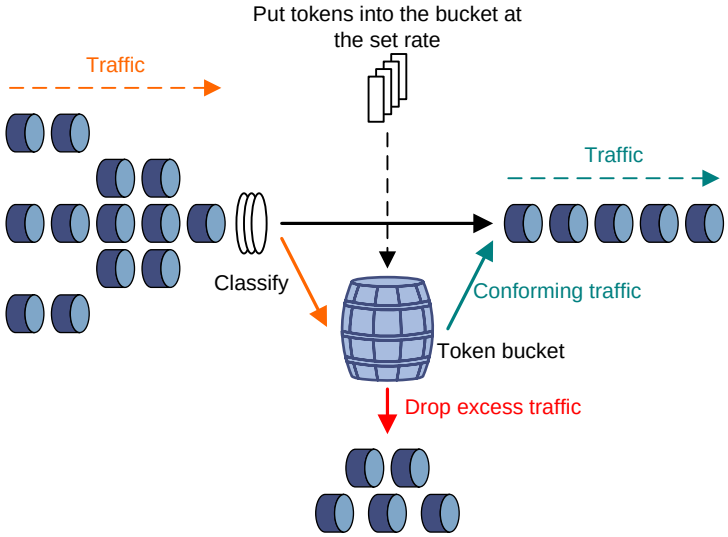
- **Two rate three color**—Uses two token buckets and the following parameters:
  - **CIR**—Rate at which tokens are put into bucket C. It sets the average packet transmission or forwarding rate allowed by bucket C.
  - **CBS**—Size of bucket C, which specifies the transient burst of traffic that bucket C can forward.
  - **PIR**—Rate at which tokens are put into bucket E, which specifies the average packet transmission or forwarding rate allowed by bucket E.
  - **EBS**—Size of bucket E, which specifies the transient burst of traffic that bucket E can forward.

  When a packet arrives, the following rules apply:
  - If bucket C has enough tokens, the packet is colored green.
  - If bucket C does not have enough tokens but bucket E has enough tokens, the packet is colored yellow.
  - If neither bucket C nor bucket E has sufficient tokens, the packet is colored red.

Traffic policing

A typical application of traffic policing is to supervise the specification of traffic entering a network and limit it within a reasonable range. Another application is to "discipline" the extra traffic to prevent aggressive use of network resources by an application. For example, you can limit bandwidth for HTTP packets to less than 50% of the total. If the traffic of a session exceeds the limit, traffic policing can drop the packets or reset the IP precedence of the packets. Figure 6 shows an example of policing outbound traffic on an interface.

Figure 6 Traffic policing



Traffic policing can classify the policed traffic and take predefined policing actions on each packet depending on the evaluation result:

- Forwarding the packet.
- Dropping the packet.
- Forwarding the packet with its precedence re-marked.
- Delivering the packet to next-level traffic policing with its precedence re-marked.

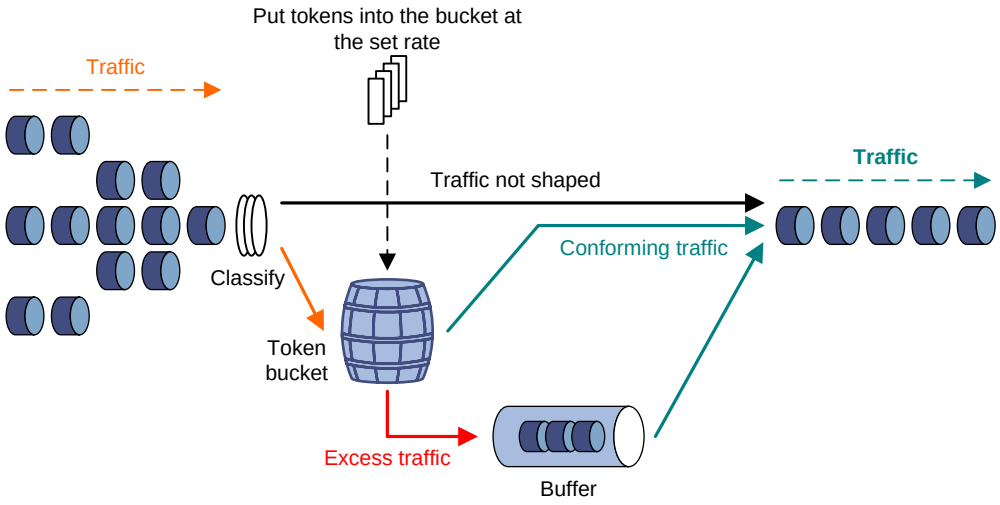
GTS

GTS limits the traffic rate by buffering exceeding traffic. You can use GTS to adapt the traffic output rate on a device to the input traffic rate of its connected device to avoid packet loss.

The differences between traffic policing and GTS are as follows:

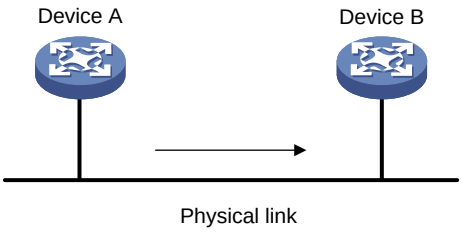
- Packets to be dropped with traffic policing are retained in a buffer or queue with GTS, as shown in Figure 7. When enough tokens are in the token bucket, the buffered packets are sent at an even rate.
- GTS can result in additional delay and traffic policing does not.

Figure 7 GTS



For example, in Figure 8, Device B performs traffic policing on packets from Device A and drops packets exceeding the limit. To avoid packet loss, you can perform GTS on the outgoing interface of Device A so that packets exceeding the limit are cached in Device A. Once resources are released, GTS takes out the cached packets and sends them out.

Figure 8 GTS application

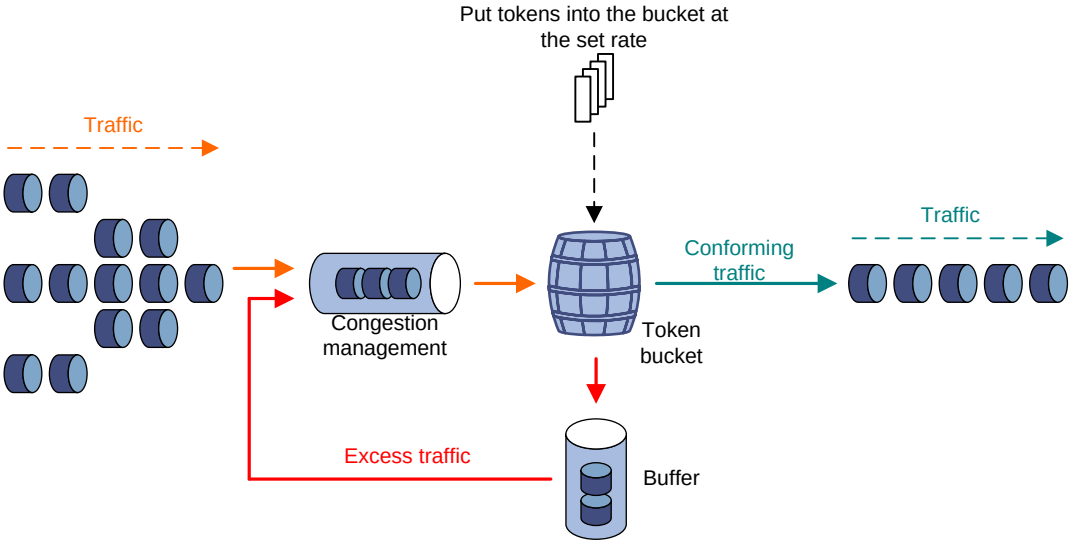


Rate limit

The rate limit of an interface or PW specifies the maximum rate for forwarding packets (excluding critical packets).

Rate limit also uses token buckets for traffic control. When rate limit is configured on an interface, a token bucket handles all packets to be sent through the interface for rate limiting. If enough tokens are in the token bucket, packets can be forwarded. Otherwise, packets are put into QoS queues for congestion management. In this way, the traffic passing the interface is controlled.

Figure 9 Rate limit implementation



The token bucket mechanism limits traffic rate when accommodating bursts. It allows bursty traffic to be transmitted if enough tokens are available. If tokens are scarce, packets cannot be transmitted until efficient tokens are generated in the token bucket. It restricts the traffic rate to the rate for generating tokens.

Rate limit controls the total rate of all packets on an interface. It is easier to use than traffic policing in controlling the total traffic rate.

Restrictions and guidelines: Traffic policing, GTS, and rate limit configuration

The specified CIR does not take traffic transmitted in interframe gaps into account, and the actually allowed rate on an interface is greater than the specified CIR.

An interframe gap is a time interval for transmitting 12 bits between frames. This gap serves the following roles:

- Allows the device to differentiate one frame from another.
- Allows for time for the device to process the current frame and to prepare for receiving the next frame.

Configuring traffic policing

Configuring traffic policing by using the MQC approach

Restrictions and guidelines

A traffic policing action takes effect only when a QoS policy is applied to the following application destinations:

- Interface.
- VLANs.
- Globally.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match *match-criteria*
      By default, no match criterion is configured.
      For more information about the **if-match** command, see *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure a traffic policing action.
      car cir *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]][**green** *action* | **red** *action* | **yellow** *action*] *
      car cir *committed-information-rate* [**cbs** *committed-burst-size*] **pir** *peak-information-rate* [**ebs** *excess-burst-size*] [**green** *action* | **red** *action* | **yellow** *action*] *
      By default, no traffic policing action is configured.
   c. Return to system view.
      quit
4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.
      qos policy *policy-name*
   b. Associate the traffic class with the traffic behavior in the QoS policy.
      classifier *classifier-name* **behavior** *behavior-name*
      By default, a traffic class is not associated with a traffic behavior.
   c. Return to system view.
      quit
5. Apply the QoS policy.
   For more information, see "Applying the QoS policy."
   By default, no QoS policy is applied.

Configuring GTS

Configuring queue-based GTS

Restrictions and guidelines

A queue-based GTS action takes effect only on outgoing traffic in the specified queue.

The term "interface" in this section collectively refers to Layer 2 and Layer 3 Ethernet interfaces. You can use the **port link-mode** command to configure an Ethernet port as a Layer 2 or Layer 3 interface (see *Interface Configuration Guide*).

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure GTS for a queue.
   qos gts queue *queue-id* **cir** *committed-information-rate* [**cbs** *committed-burst-size*]
   By default, GTS is not configured on an interface queue.

Configuring the rate limit

Configuring the rate limit for an interface

Restrictions and guidelines

The term "interface" in this section collectively refers to Layer 2 and Layer 3 Ethernet interfaces. You can use the **port link-mode** command to configure an Ethernet port as a Layer 2 or Layer 3 interface (see *Layer 2—LAN Switching Configuration Guide*).

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the rate limit for the interface.
   qos lr { **inbound** | **outbound** } **cir** *committed-information-rate* [**cbs** *committed-burst-size*]
   By default, no rate limit is configured on an interface.

Display and maintenance commands for traffic policing, GTS, and rate limit

Execute **display** commands in any view.

Task	Command
Display ACL rules in the ACL whitelist.	**display acl whitelist** [**ipv6**] **slot** *slot-number*
Display GTS configuration and statistics for interfaces.	**display qos gts interface** [*interface-type interface-number*]
Display rate limit configuration and statistics.	**display qos lr interface** [*interface-type*

	interface-number]
Display QoS and ACL resource usage.	**display qos-acl resource** [**slot** *slot-number*]
Display traffic behavior configuration.	**display traffic behavior user-defined** [*behavior-name*]

Configuring congestion management

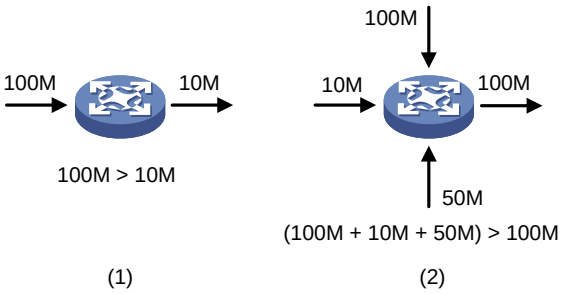
About congestion management

Cause, negative results, and countermeasure of congestion

Congestion occurs on a link or node when traffic size exceeds the processing capability of the link or node. It is typical of a statistical multiplexing network and can be caused by link failures, insufficient resources, and various other causes.

Figure 10 shows two typical congestion scenarios.

Figure 10 Traffic congestion scenarios



Congestion produces the following negative results:

- Increased delay and jitter during packet transmission.
- Decreased network throughput and resource use efficiency.
- Network resource (memory, in particular) exhaustion and even system breakdown.

Congestion is unavoidable in switched networks and multiuser application environments. To improve the service performance of your network, take measures to manage and control it.

The key to congestion management is defining a resource dispatching policy to prioritize packets for forwarding when congestion occurs.

Congestion management methods

Congestion management uses queuing and scheduling algorithms to classify and sort traffic leaving a port.

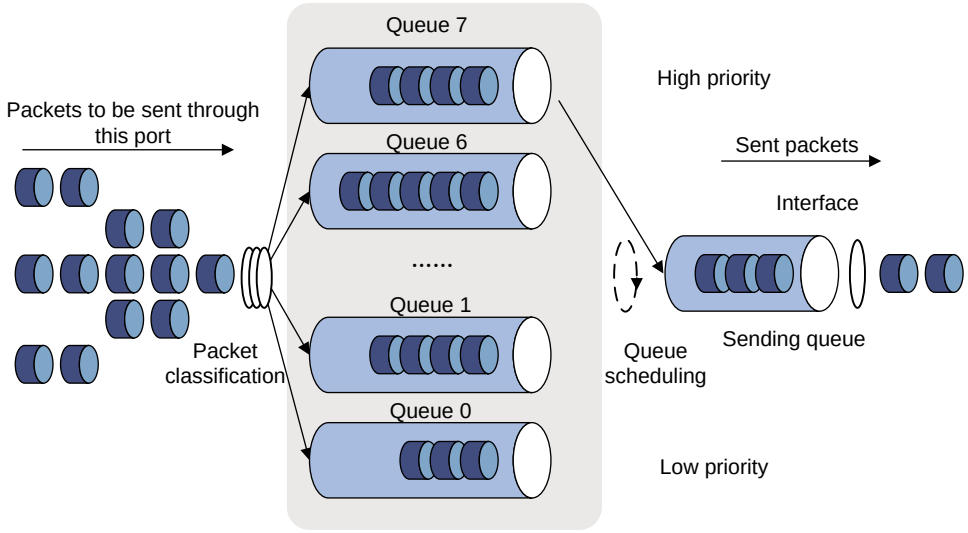
The device supports the following queuing mechanisms:

- SP.
- WRR.
- WFQ.
- CBQ.

SP queuing

SP queuing is designed for mission-critical applications that require preferential service to reduce the response delay when congestion occurs.

Figure 11 SP queuing



In Figure 11, SP queuing classifies eight queues on a port into eight classes, numbered 7 to 0 in descending priority order.

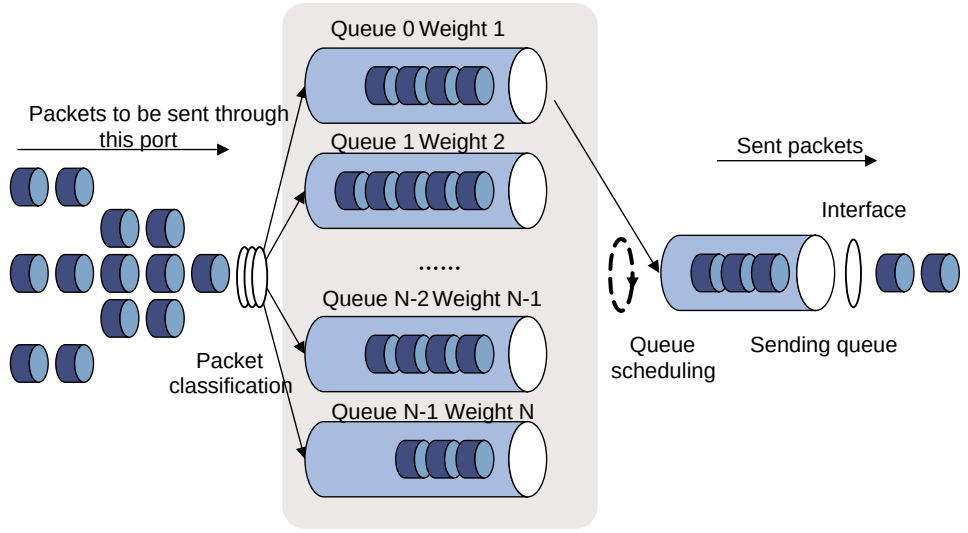
SP queuing schedules the eight queues in the descending order of priority. SP queuing sends packets in the queue with the highest priority first. When the queue with the highest priority is empty, it sends packets in the queue with the second highest priority, and so on. You can assign mission-critical packets to a high priority queue to make sure they are always served first. Common service packets can be assigned to low priority queues to be transmitted when high priority queues are empty.

The disadvantage of SP queuing is that packets in the lower priority queues cannot be transmitted if packets exist in the higher priority queues. In the worst case, lower priority traffic might never get serviced.

WRR queuing

WRR queuing schedules all the queues in turn to ensure that every queue is served for a certain time, as shown in Figure 12.

Figure 12 WRR queuing



Assume a port provides eight output queues. WRR assigns each queue a weight value (represented by w7, w6, w5, w4, w3, w2, w1, or w0). The weight value of a queue decides the proportion of resources assigned to the queue. On a 100 Mbps port, you can set the weight values to 50, 30, 10, 10, 50, 30, 10, and 10 for w7 through w0. In this way, the queue with the lowest priority can get a minimum of 5 Mbps of bandwidth. WRR solves the problem that SP queuing might fail to serve packets in low-priority queues for a long time.

Another advantage of WRR queuing is that when the queues are scheduled in turn, the service time for each queue is not fixed. If a queue is empty, the next queue will be scheduled immediately. This improves bandwidth resource use efficiency.

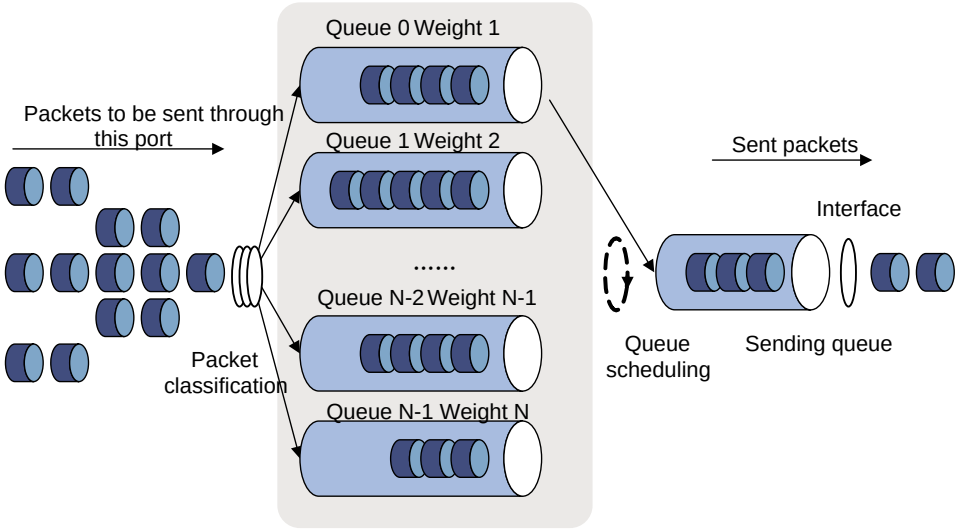
WRR queuing includes the following types:

- **Basic WRR queuing**—Contains multiple queues. You can set the weight for each queue, and WRR schedules these queues based on the user-defined parameters in a round robin manner.
- **Group-based WRR queuing**—All the queues are scheduled by WRR. You can divide output queues to WRR priority queue group 1 and WRR priority queue group 2. Round robin queue scheduling is performed for group 1 first. If group 1 is empty, round robin queue scheduling is performed for group 2.

On an interface enabled with group-based WRR queuing, you can assign queues to the SP group. Queues in the SP group are scheduled with SP. The SP group has higher scheduling priority than the WRR groups.

WFQ queuing

Figure 13 WFQ queuing



WFQ is similar to WRR. WFQ queuing includes basic WFQ queuing and group-based WFQ queuing.

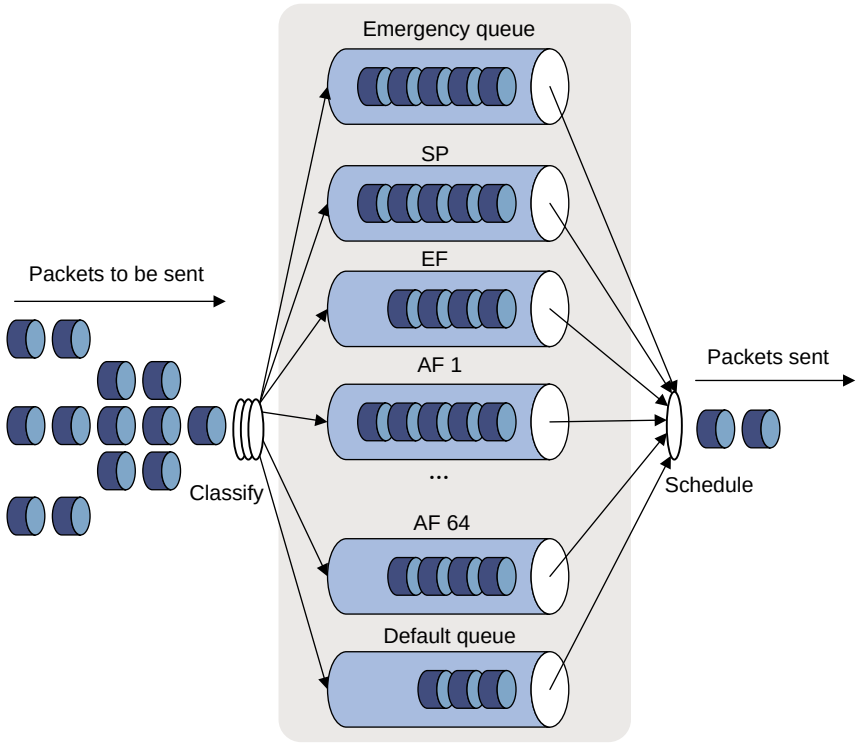
On an interface with group-based WFQ queuing enabled, you can also assign queues to the SP group. The difference is that WFQ enables you to set guaranteed bandwidth that a WFQ queue can get during congestion.

SP+WFQ queuing schedules traffic in the following order:

1. Schedules the traffic conforming to the minimum guaranteed bandwidth of each queue in the WFQ group.
2. Schedules the queues in the SP group based on their priorities.
3. Schedules the traffic in each queue in the WFQ group according to the configured weights when all queues in the SP group are empty.

CBQ

Figure 14 CBQ



Class-based queuing (CBQ) extends WFQ by supporting user-defined classes. When network congestion occurs, CBQ uses user-defined traffic match criteria to enqueue packets. Before packets are enqueued, congestion avoidance actions, such as tail drop or WRED and bandwidth restriction check, are performed. When being dequeued, packets are scheduled by WFQ.

CBQ provides the following queues:

- **Emergency queue**—Enqueues emergent packets. The emergency queue is a FIFO queue without bandwidth restriction.
- **Strict priority (SP) queuing**—SP queuing is designed for mission-critical applications. Mission-critical applications require preferential treatment to reduce the response delay when congestion occurs. With SP queuing, CBQ can provide strict priority service. A maximum of 64 SP queues are supported.
- **Low Latency Queuing (LLQ)**—An EF queue. Because packets are fairly treated in CBQ, delay-sensitive flows like video and voice packets might not be transmitted timely. To solve this problem, LLQ combines PQ and CBQ to preferentially transmit delay-sensitive flows like voice packets. When defining traffic classes for LLQ, you can configure a class of packets to be preferentially transmitted. The packets of all priority classes are assigned to the same priority queue. Bandwidth restriction on each class of packets is checked before the packets are enqueued. During the dequeuing operation, packets in the priority queue are transmitted first. To reduce the delay of the other queues except the priority queue, LLQ assigns the maximum available bandwidth to each priority class. The bandwidth value polices traffic during congestion. When no congestion is present, a priority class can use more than the bandwidth assigned to it. During congestion, the packets of each priority class exceeding the assigned bandwidth are discarded.
- **Bandwidth queuing (BQ)**—An AF queue. The BQ provides guaranteed bandwidth for AF traffic, and schedules the AF classes proportionally. The system supports up to 64 AF queues.
- **Default queue**—A WFQ queue. The default queue transmits the BE traffic by using the remaining interface bandwidth.

The system matches packets with classification rules in the following order:

- Match packets with priority classes and then the other classes.
- Match packets with priority classes in the configuration order.
- Match packets with other classes in the configuration order.
- Match packets with classification rules in a class in the configuration order.

congestion management tasks at a glance

To configure congestion management, perform the following tasks:

1.
2. Configuring queuing on an interface
   - Configuring SP queuing
   - Configuring WRR queuing
   - Configuring WFQ queuing
   - Configuring SP+WRR queuing
   - Configuring SP+WFQ queuing
3. Configuring a queue scheduling profile

Configuring queuing on an interface

Restrictions and guidelines for queuing configuration

The term "interface" in this section collectively refers to Layer 2 and Layer 3 Ethernet interfaces. You can use the **port link-mode** command to configure an Ethernet port as a Layer 2 or Layer 3 interface (see *Layer 2—LAN Switching Configuration Guide*).

Configuring SP queuing

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure SP queuing.
   qos sp
   By default, an interface uses SP queuing.

Configuring WRR queuing

Configuring group-based WRR queuing

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable WRR queuing.

qos wrr { **byte-count** | **weight** }

By default, an interface uses SP queuing.

4. Assign a queue to a WRR group, and configure scheduling parameters for the queue.

qos wrr *queue-id* **group** { **1** | **2** } { **byte-count** | **weight** } *schedule-value*

By default, all queues on a WRR-enabled interface are in WRR group 1, and queues 0 through 7 have a weight of 1, 2, 3, 4, 5, 9, 13, and 15, respectively.

Configuring WFQ queuing

Configuring grouped WFQ queues

1. Enter system view.

system-view

2. Enter interface view.

interface *interface-type interface-number*

3. Enable WFQ queuing.

qos wfq { **byte-count** | **weight** }

By default, an interface uses SP queuing.

4. Assign a queue to a WFQ group, and configure scheduling parameters for the queue.

qos wfq *queue-id* **group** { **1** | **2** } { **byte-count** | **weight** } *schedule-value*

By default, all queues on a WFQ-enabled interface are in WFQ group 1 and have a weight of 1.

Configuring SP+WRR queuing

Restrictions and guidelines

To configure the scheduling weight, you must specify the same scheduling unit as specified when enabling WRR queuing.

Procedure

1. Enter system view.

system-view

2. Enter interface view.

interface *interface-type interface-number*

3. Enable byte-count or packet-count WRR queuing.

qos wrr { **byte-count** | **weight** }

By default, an interface uses SP queuing.

4. Assign a queue to the SP group.

qos wrr *queue-id* **group sp**

By default, all queues on a WRR-enabled interface are in WRR group 1.

5. Assign a queue to a WRR group, and configure a scheduling weight for the queue.

qos wrr *queue-id* **group** { **1** | **2** } { **byte-count** | **weight** } *schedule-value*

By default, all queues on a WRR-enabled interface are in WRR group 1, and queues 0 through 7 have a weight of 1, 2, 3, 4, 5, 9, 13, and 15, respectively.

Configuring SP+WFQ queuing

Restrictions and guidelines

To configure the scheduling weight, you must specify the same scheduling unit as specified when enabling WFQ queuing.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable byte-count or packet-count WFQ queuing.
   qos wfq { **byte-count** | **weight** }
   By default, an interface uses SP queuing.
4. Assign a queue to the SP group.
   qos wfq *queue-id* **group sp**
   By default, all queues on a WFQ-enabled interface are in WFQ group 1.
5. Assign a queue to a WFQ queue scheduling group, and configure a scheduling weight for the queue.
   qos wfq *queue-id* **group** { **1** | **2** } { **byte-count** | **weight** } *schedule-value*
   By default, all queues on a WFQ-enabled interface are in WFQ group 1 and have a weight of 1.
6. (Optional.) Set the minimum guaranteed bandwidth for a queue.
   qos bandwidth queue *queue-id* **min** *bandwidth-value*
   By default, the minimum guaranteed bandwidth for a WFQ queue in an WFQ group is not set.

Configuring CBQ

Restrictions and guidelines for CBQ configuration

When you configure CBQ, follow these restrictions and guidelines:

- Only one of the following queuing actions can be configured in the same behavior:
  - AF.
  - EF.
  - SP.
  - WFQ.
- To configure **queue af** for multiple classes of a policy, you must configure them in one of the following units:
  - Bandwidth in kbps.
  - Percentage of the available bandwidth.
  - Percentage of the remaining bandwidth.
- In a QoS policy, if AF is configured in percentage of the remaining bandwidth, EF can be configured in absolute value or in percentage.
- In a QoS policy, if AF is configured in absolute value or in percentage, EF can be configured in absolute value or in percentage.

- You must configure the rate limit on the main interface for the CBQ feature to take effect on a subinterface.
- The default class cannot be associated with a traffic behavior including EF or SP.
- You can associate the traffic behavior that contains a WFQ action only with the default class.
- RTPQ and CBQ are mutually exclusive. CBQ can use LLQ to provide priority service for real-time traffic.
-

Configure AF and the minimum guaranteed bandwidth

Restrictions and guidelines

- The **undo queue af** command also cancels the **queue-length** command configuration and all WRED-related settings.
- The **wred** and **queue-length** commands are mutually exclusive in a traffic behavior.
- The **undo wred** [**dscp** | **ip-precedence**] command also cancels all other WRED-related settings.
- A QoS policy that contains an AF action can be applied to the outbound direction of an interface or the inbound direction for a user profile. To modify a QoS policy applied to a user profile, follow these steps:
  a. Remove the QoS policy from the user profile.
  b. Configure a new QoS policy.
  c. Apply the new QoS policy to the user profile.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
   traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
   if-match [**not**] *match-criteria*
   By default, no match criterion is configured.
   For more information about configuring match criteria, see *ACL and QoS Command Reference*.
   c. Return to system view.
   quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
   traffic behavior *behavior-name*
   b. Configure AF and the minimum guaranteed bandwidth.
   queue af bandwidth { *bandwidth* [**pir** *peak-information-rate*] | **pct** *percentage* | **remaining-pct** *remaining-percentage* }
   By default, AF is not configured.
   c. (Optional.) Set a weight for the AF queue.
   weight *weight-value*
   The default setting is 1.

d. (Optional.) Set the maximum queue size.

queue-length *queue-length*

The default for this command varies by device model.

If the burst traffic is too heavy, increase the queue length to make queue scheduling more accurate.

e. Enable WRED.

wred [**dscp** | **ip-precedence**]

By default, WRED is not enabled.

f. Set the exponent for WRED to calculate the average queue size.

wred weighting-constant *exponent*

The default setting is 9.

g. Set the lower limit, upper limit, and drop probability denominator for a DSCP value in WRED

wred dscp *dscp-value* **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]

By default, the lower limit is 10, the upper limit is 30, and the drop probability denominator is 10.

You must enable DSCP-based WRED by using the **wred dscp** command before you configure this command.

h. Set the lower limit, upper limit, and drop probability denominator for an IP precedence in WRED.

wred ip-precedence *precedence* **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]

By default, the lower limit is 10, the upper limit is 30, and the drop probability denominator is 10.

i. Return to system view.

quit

4. Define a QoS policy.

a. Create a QoS policy and enter QoS policy view.

qos policy *policy-name*

b. Associate a traffic behavior with a class in the policy.

classifier *classifier-name* **behavior** *behavior-name*

By default, a class is not associated with a behavior.

c. Return to system view.

quit

5. Apply the QoS policy.

For more information, see "Applying the QoS policy."

By default, no QoS policy is applied.

Configuring EF and the maximum bandwidth

Restrictions and guidelines

A QoS policy that contains an EF action can be applied to the outbound direction of an interface or the inbound direction for a user profile. To modify a QoS policy applied to a user profile, follow these steps:

1. Remove the QoS policy from the user profile.

2. Configure a new QoS policy.

3. Apply the new QoS policy to the user profile.

Procedure

1. Enter system view.

   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.

      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.

      if-match [**not**] *match-criteria*

      By default, no match criterion is configured.

      For more information about configuring match criteria, see *ACL and QoS Command Reference*.
   c. Return to system view.

      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.

      traffic behavior *behavior-name*
   b. Configure EF and the maximum bandwidth.

      queue ef bandwidth { *bandwidth* [**cbs** *burst*] [**pir** *peak-information-rate*] | **pct** *percentage* [**cbs-ratio** *ratio*] }

      By default, EF is not configured.
   c. (Optional.) Set a weight for the EF queue.

      weight *weight-value*

      The default setting is 1.
   d. Return to system view.

      quit
4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.

      qos policy *policy-name*
   b. Associate a traffic behavior with a class in the policy.

      classifier *classifier-name* **behavior** *behavior-name*

      By default, a class is not associated with a behavior.
   c. Return to system view.

      quit
5. Apply the QoS policy.

   For more information, see "Applying the QoS policy."

   By default, no QoS policy is applied.

Configuring SP

Restrictions and guidelines

A QoS policy that contains an SP action can be applied to the outbound direction of an interface or the inbound direction for a user profile. To modify a QoS policy applied to a user profile, follow these steps:

1. Remove the QoS policy from the user profile.
2. Configure a new QoS policy.
3. Apply the new QoS policy to the user profile.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match [**not**] *match-criteria*
      By default, no match criterion is configured.
      For more information about configuring match criteria, see *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure SP.
      queue sp
      By default, SP is not configured.
   c. Return to system view.
      quit
4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.
      qos policy *policy-name*
   b. Associate a traffic behavior with a class in the policy.
      classifier *classifier-name* **behavior** *behavior-name*
      By default, a class is not associated with a behavior.
   c. Return to system view.
      quit
5. Apply the QoS policy.
   For more information, see "Applying the QoS policy."
   By default, no QoS policy is applied.

Configuring WFQ

Restrictions and guidelines

- The **undo queue wfq** command also cancels the **queue-length** command configuration and all WRED-related settings.
- The **wred** and **queue-length** commands are mutually exclusive in a traffic behavior.
- The **undo wred** [**dscp** | **ip-precedence**] command also cancels all other WRED-related settings.

- A QoS policy that contains a WFQ action can be applied to the outbound direction of an interface or the inbound direction for a user profile. To modify a QoS policy applied to a user profile, follow these steps:
  a. Remove the QoS policy from the user profile.
  b. Configure a new QoS policy.
  c. Apply the new QoS policy to the user profile.

Procedure

1. Enter system view.
system-view

2. Define a traffic class.
a. Create a traffic class and enter traffic class view.
traffic classifier *classifier-name* [**operator** { **and** | **or** }]
b. Configure a match criterion.
if-match [**not**] *match-criteria*
By default, no match criterion is configured.
For more information about configuring match criteria, see *ACL and QoS Command Reference*.
c. Return to system view.
quit

3. Define a traffic behavior.
a. Create a traffic behavior and enter traffic behavior view.
traffic behavior *behavior-name*
b. Configure WFQ.
queue wfq [**queue-number** *total-queue-number*]
By default, WFQ is not configured.
c. (Optional.) Set a weight for the WFQ queue.
weight *weight-value*
The default setting is 1.
d. Set the maximum queue size.
queue-length *queue-length*
The default for this command varies by device model.
If the burst traffic is too heavy, increase the queue length to make queue scheduling more accurate.
e. Enable WRED.
wred [**dscp** | **ip-precedence**]
By default, WRED is not enabled.
f. Set the exponent for WRED to calculate the average queue size
wred weighting-constant *exponent*
The default setting is 9.
g. Set the lower limit, upper limit, and drop probability denominator for a DSCP value in WRED
wred dscp *dscp-value* **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]
By default, the lower limit is 10, the upper limit is 30, and the drop probability denominator is 10.

h. Set the lower limit, upper limit, and drop probability denominator for an IP precedence in WRED.

wred ip-precedence *precedence* **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]

By default, the lower limit is 10, the upper limit is 30, and the drop probability denominator is 10.

i. Return to system view.

quit

4. Define a QoS policy.

a. Create a QoS policy and enter QoS policy view.

qos policy *policy-name*

b. Associate a traffic behavior with a class in the policy.

classifier *classifier-name* **behavior** *behavior-name*

By default, a class is not associated with a behavior.

c. Return to system view.

quit

5. Apply the QoS policy.

For more information, see "Applying the QoS policy."

By default, no QoS policy is applied.

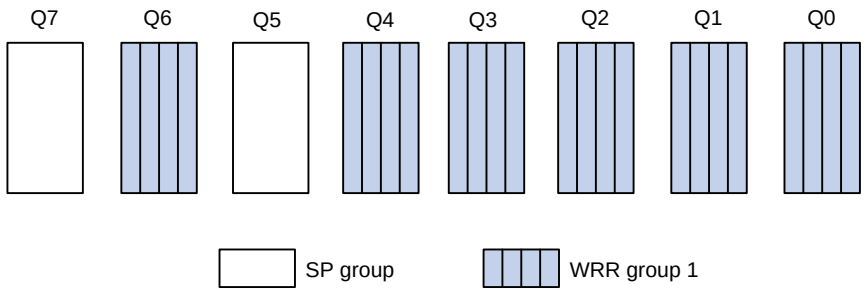
Configuring a queue scheduling profile

About queue scheduling profiles

In a queue scheduling profile, you can configure scheduling parameters for each queue. By applying the queue scheduling profile to an interface or session group profile, you can implement congestion management on the interface or session group profile.

Queue scheduling profiles support three queue scheduling algorithms: SP, WRR, and WFQ. In a queue scheduling profile, you can configure SP + WRR or SP + WFQ. For information about each scheduling algorithm, see "About congestion management." When SP and WRR groups are configured in a queue scheduling profile, Figure 15 shows the scheduling order.

Figure 15 Queue scheduling profile configured with both SP and WRR



- Queue 7 has the highest priority in the SP group. Its packets are sent preferentially.
- Queue 5 has the second highest priority in the SP group. Packets in queue 5 are sent when queue 7 is empty.
- All queues in WRR group 1 are scheduled according to their weights. When queue 7 and queue 5 are empty, WRR group 1 is scheduled.

Restrictions and guidelines for queue scheduling profile configuration

When you configure a queue scheduling profile, follow these restrictions and guidelines:

- The term "interface" in this section collectively refers to Layer 2 and Layer 3 Ethernet interfaces. You can use the **port link-mode** command to configure an Ethernet port as a Layer 2 or Layer 3 interface (see *Layer 2—LAN Switching Configuration Guide*).
- Only one queue scheduling profile can be applied to an interface or session group profile.
- You can modify the scheduling parameters in a queue scheduling profile already applied to an interface, the switching fabric module, or a session group profile.

Configuring a queue scheduling profile

1. Enter system view.

   system-view
2. Create a queue scheduling profile and enter queue scheduling profile view.

   qos qmprofile *profile-name*
3. (Optional.) Configure queue scheduling parameters.
   - Configure a queue to use SP.

     queue *queue-id* **sp**
   - Configure a queue to use WRR.

     queue *queue-id* **wrr group** *group-id* { **weight** | **byte-count** } *schedule-value*
   - Configure a queue to use WFQ.

     queue *queue-id* **wfq group** *group-id* { **weight** | **byte-count** } *schedule-value*

   By default, all queues in a queue scheduling profile are SP queues.
4. (Optional.) Set the minimum guaranteed bandwidth for a WFQ queue.

   bandwidth queue *queue-id* **min** *bandwidth-value*

   By default, the minimum guaranteed bandwidth is not set for a WFQ queue.

Applying a queue scheduling profile

1. Enter system view.

   system-view
2. Execute the following commands in sequence to apply the queue scheduling profile to the outbound direction of an interface:

   interface *interface-type interface-number*

   qos apply qmprofile *profile-name*

   By default, no queue scheduling profile is applied to an interface.

Example: Configuring a queue scheduling profile

Network configuration

Configure a queue scheduling profile to meet the following requirements on GigabitEthernet 1/0/1:

- Queue 7 has the highest priority, and its packets are sent preferentially.
- Queue 0 through queue 6 are in the WRR group and are scheduled according to their packet-count weights, which are 2, 1, 2, 4, 6, 8, and 10, respectively. When queue 7 is empty, the WRR group is scheduled.

Procedure

Enter system view.

```
<Sysname> system-view
```

Create a queue scheduling profile named **qm1**.

```
[Sysname] qos qmprofile qm1
[Sysname-qmprofile-qm1]
```

Configure queue 7 to use SP queuing.

```
[Sysname-qmprofile-qm1] queue 7 sp
```

Assign queue 0 through queue 6 to WRR group 1, with their packet-count weights as 2, 1, 2, 4, 6, 8, and 10, respectively.

```
[Sysname-qmprofile-qm1] queue 0 wrr group 1 weight 2
[Sysname-qmprofile-qm1] queue 1 wrr group 1 weight 1
[Sysname-qmprofile-qm1] queue 2 wrr group 1 weight 2
[Sysname-qmprofile-qm1] queue 3 wrr group 1 weight 4
[Sysname-qmprofile-qm1] queue 4 wrr group 1 weight 6
[Sysname-qmprofile-qm1] queue 5 wrr group 1 weight 8
[Sysname-qmprofile-qm1] queue 6 wrr group 1 weight 10
[Sysname-qmprofile-qm1] quit
```

Apply queue scheduling profile **qm1** to GigabitEthernet 1/0/1.

```
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos apply qmprofile qm1
```

After the configuration is completed, GigabitEthernet 1/0/1 performs queue scheduling as specified in queue scheduling profile **qm1**.

Display and maintenance commands for congestion management

Execute **display** commands in any view.

Task	Command
Display the configuration of queue scheduling profiles.	**display qos qmprofile configuration** [*profile-name*] [**slot** *slot-number*]
Display the queue scheduling profiles applied to interfaces.	**display qos qmprofile interface** [*interface-type interface-number*]
Display outbound queue-based traffic statistics for interfaces.	**display qos queue-statistics interface outbound**
Display SP queuing configuration.	**display qos queue sp interface** [*interface-type interface-number*]
Display WFQ queuing configuration.	**display qos queue wfq interface** [*interface-type*

Task	Command
	interface-number]
Display WRR queuing configuration.	**display qos queue wrr interface** [*interface-type interface-number*]

Configuring congestion avoidance

About congestion avoidance

Avoiding congestion before it occurs is a proactive approach to improving network performance. As a flow control mechanism, congestion avoidance:

- Actively monitors network resources (such as queues and memory buffers).
- Drops packets when congestion is expected to occur or deteriorate.

When dropping packets from a source end, congestion avoidance cooperates with the flow control mechanism at the source end to regulate the network traffic size. The combination of the local packet drop policy and the source-end flow control mechanism implements the following functions:

- Maximizes throughput and network use efficiency.
- Minimizes packet loss and delay.

Tail drop

Congestion management techniques drop all packets that are arriving at a full queue. This tail drop mechanism results in global TCP synchronization. If packets from multiple TCP connections are dropped, these TCP connections go into the state of congestion avoidance and slow start to reduce traffic. However, traffic peak occurs later. Consequently, the network traffic jitters all the time.

RED and WRED

You can use Random Early Detection (RED) or Weighted Random Early Detection (WRED) to avoid global TCP synchronization.

Both RED and WRED avoid global TCP synchronization by randomly dropping packets. When the sending rates of some TCP sessions slow down after their packets are dropped, other TCP sessions remain at high sending rates. Link bandwidth is efficiently used, because TCP sessions at high sending rates always exist.

The RED or WRED algorithm sets an upper threshold and lower threshold for each queue, and processes the packets in a queue as follows:

- When the queue size is shorter than the lower threshold, no packet is dropped.
- When the queue size reaches the upper threshold, all subsequent packets are dropped.
- When the queue size is between the lower threshold and the upper threshold, the received packets are dropped at random. The drop probability in a queue increases along with the queue size under the maximum drop probability.

If the current queue size is compared with the upper threshold and lower threshold to determine the drop policy, burst traffic is not fairly treated. To solve this problem, WRED compares the average queue size with the upper threshold and lower threshold to determine the drop probability.

The average queue size reflects the queue size change trend but is not sensitive to burst queue size changes, and burst traffic can be fairly treated.

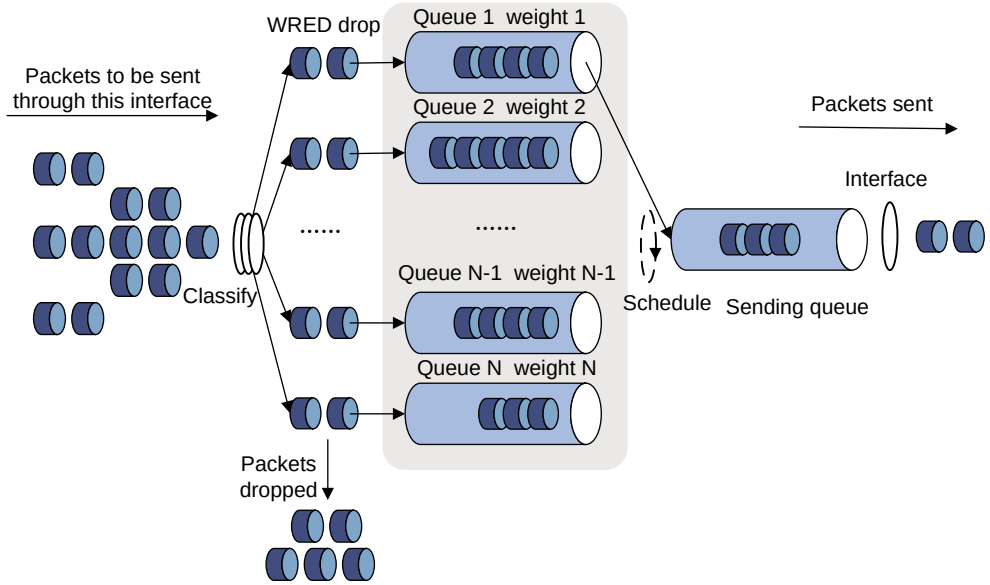
When WFQ queuing is used, you can set the following parameters for packets with different precedence values to provide differentiated drop policies:

- Exponent for average queue size calculation.
- Upper threshold.
- Lower threshold.

- Drop probability.

Relationship between WRED and queuing mechanisms

Figure 16 Relationship between WRED and queuing mechanisms



Through combining WRED with WFQ, the flow-based WRED can be realized. Each flow has its own queue after classification.

- A flow with a smaller queue size has a lower packet drop probability.
- A flow with a larger queue size has a higher packet drop probability.

In this way, the benefits of the flow with a smaller queue size are protected.

ECN

By dropping packets, WRED alleviates the influence of congestion on the network. However, the network resources for transmitting packets from the sender to the device which drops the packets are wasted. When congestion occurs, it is a better idea to perform the following actions:

- Inform the sender of the congestion status.
- Have the sender proactively slow down the packet sending rate or decrease the window size of packets.

This better utilizes the network resources.

RFC 2481 defined an end-to-end congestion notification mechanism named Explicit Congestion Notification (ECN). ECN uses the DS field in the IP header to mark the congestion status along the packet transmission path. An ECN-capable endpoint can determine whether congestion occurs on the transmission path according to the packet contents. Then, it adjusts the packet sending speed to avoid deteriorating congestion. ECN defines the last two bits (ECN field) in the DS field of the IP header as follows:

- Bit 6 indicates whether the sending endpoint device supports ECN, and is called the ECN-Capable Transport (ECT) bit.
- Bit 7 indicates whether the packet has experienced congestion along the transmission path, and is called the Congestion Experienced (CE) bit.

For more information about the DS field, see "Appendixes."

In actual applications, the following packets are considered as packets that an ECN-capable endpoint transmits:

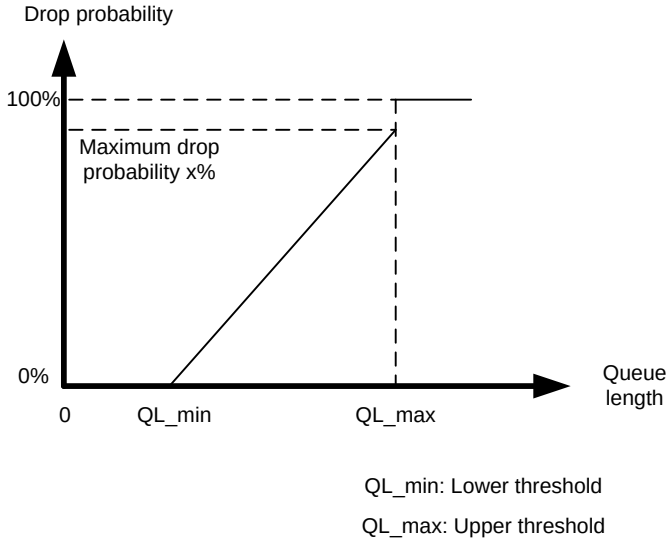
- Packets with ECT set to 1 and CE set to 0.
- Packets with ECT set to 0 and CE set to 1.

To enable ECN, do not specify ToS field bit 6 or 7 as the flag bit in iNQA, and vice versa. For more information about iNQA, see *Network Management and Monitoring Configuration Guide*.

As shown in Figure 17, the WRED algorithm sets an upper threshold and lower threshold for each queue, and processes the packets in a queue as follows if ECN is not enabled:

- When the queue size is shorter than the lower threshold, no packet is dropped.
- When the queue size reaches the upper threshold, all subsequent packets are dropped.
- When the queue size is between the lower threshold and the upper threshold, the received packets are dropped at random. The drop probability in a queue increases along with the queue size under the maximum drop probability.

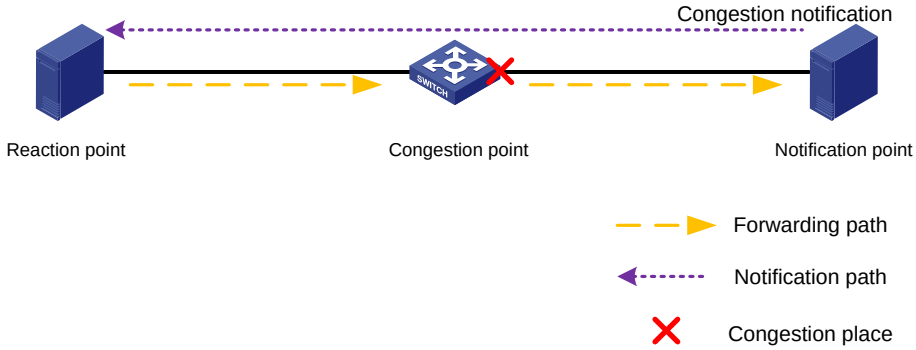
Figure 17 Drop probability and queue length



As shown in Figure 18, the network is deployed with ECN and has the following device roles:

- **Congestion point**—It forwards packets from the packet sender to the packet receiver. It supports ECN and can identify the ECN field. Congestion might occur on the congestion point. ECN must be configured on the congestion point.
- **Notification point**—It is an endpoint that receives packets. Its NIC supports ECN and can identify the ECN values 01, 10, and 11. After receiving a packet with the ECN field as 11, the notification point will notify the reaction point to slow down transmission by sending CE packets at the interval T1.
- **Reaction point**—It is an endpoint that sends packets. Its NIC supports ECN, and the ECN field of packets sent by it can be 01 or 10. After receiving a CE packet, the reaction point slows down transmission and starts a timer (T2, (T2>T1)). If the reaction point does not receive another CE packet when the timer expires, it considers that there is no congestion in the network and restores the previous transmission rate. If the reaction point receives another CE packet before the timer expires, it further reduces the transmission rate.

Figure 18 Network diagram with ECN enabled



After you enable ECN on the congestion point, it processes an incoming packet that enters an output queue as follows:

- When the average queue size is below the lower threshold, the congestion point forwards the packet without modifying the ECN field.
- When the average queue size is between the lower threshold and the upper threshold:
  - If the ECN field is 00, the packet comes from an ECN-incapable endpoint, and the congestion point determines whether to drop the packet according to the drop probability.
  - If the ECN field is 01 or 10, the packet comes from an ECN-capable endpoint, and the congestion point determines whether to modify the ECN field according to the drop probability and then forwards the packet. All packets will not be dropped.
  - If the ECN field is 11, the packet has experienced congestion along the transmission path, and the congestion point forwards the packet without modifying the ECN field.
- When the average queue size is above the upper threshold:
  - If the ECN field is 00, the packet comes from an ECN-incapable endpoint, and the congestion point drops the packet.
  - If the ECN field is 01 or 10, the packet comes from an ECN-capable endpoint, and the congestion point determines whether to modify the ECN field according to the drop probability and then forwards the packet. All packets will not be dropped.
  - If the ECN field is 11, the packet has experienced congestion along the transmission path, and the congestion point forwards the packet without modifying the ECN field.

Proper ECN thresholds can improve the network delay and throughput while alleviating congestion.

WRED parameters

Determine the following parameters before configuring WRED:

- **Upper threshold and lower threshold**—When the average queue size is smaller than the lower threshold, packets are not dropped. When the average queue size is between the lower threshold and the upper threshold, the packets are dropped at random. The longer the queue, the higher the drop probability. When the average queue size exceeds the upper threshold, subsequent packets are dropped.
- **Drop precedence**—A parameter used for packet drop. The value 0 corresponds to green packets, the value 1 corresponds to yellow packets, and the value 2 corresponds to red packets. Red packets are dropped preferentially.
- **Exponent for average queue size calculation**—The greater the exponent, the less sensitive the average queue size is to real-time queue size changes. The formula for calculating the average queue size is:

  Average queue size = (previous average queue size x $(1 – 2^{-n})$) + (current queue size x 2^{-n}), where n is the exponent.

- **Drop probability**—Drop probability in percentage. The greater the value, the higher the drop probability. (Applicable when you configure a WRED table on an interface.)

Configuring and applying a queue-based WRED table to an interface

Restrictions and guidelines

By using a WRED table, WRED randomly drops packets during congestion based on the queues that hold packets.

One WRED table can be applied to multiple interfaces. You can modify the parameters of a WRED table applied to an interface, but you cannot delete the WRED table.

The term "interface" in this section collectively refers to Layer 2 and Layer 3 Ethernet interfaces. You can use the **port link-mode** command to configure an Ethernet port as a Layer 2 or Layer 3 interface (see *Layer 2—LAN Switching Configuration Guide*).

Procedure

1. Enter system view.
   system-view
2. Create a WRED table and enter its view.
   qos wred queue table *table-name*
3. (Optional.) Set the WRED exponent for average queue size calculation.
   queue *queue-id* **weighting-constant** *exponent*
   The default setting is 9.
4. (Optional.) Configure the other WRED parameters.
   queue *queue-id* [**drop-level** *drop-level*] **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]
   By default, the lower limit is 100, the higher limit is 1000, and the drop probability is 10%.
   If you configure the same value for the lower limit and higher limit and set the drop probability to 100%, tail drop is used for the queue.
5. Return to system view.
   quit
6. Enter interface view.
   interface *interface-type interface-number*
7. Apply the WRED table to the interface.
   qos wred apply [*table-name*]
   By default, no WRED table is applied to an interface, and tail drop is used on an interface.

Example: Configuring and applying a queue-based WRED table

Network configuration

Apply a WRED table to GigabitEthernet 1/0/2, so that the packets are dropped as follows when congestion occurs:

- For the interface to preferentially forward higher-priority traffic, set a lower drop probability for a queue with a greater queue number. Set different drop parameters for queue 0, queue 3, and queue 7.
- Drop packets according to their colors.
  - In queue 0, set the drop probability to 25%, 50%, and 75% for green, yellow, and red packets, respectively.
  - In queue 3, set the drop probability to 5%, 10%, and 25% for green, yellow, and red packets, respectively.
  - In queue 7, set the drop probability to 1%, 5%, and 10% for green, yellow, and red packets, respectively.

Procedure

Configure a queue-based WRED table, and set different drop parameters for packets with different drop levels in different queues.

```
<Sysname> system-view
[Sysname] qos wred queue table queue-table1
[Sysname-wred-table-queue-table1] queue 0 drop-level 0 low-limit 128 high-limit 512
discard-probability 25
[Sysname-wred-table-queue-table1] queue 0 drop-level 1 low-limit 128 high-limit 512
discard-probability 50
[Sysname-wred-table-queue-table1] queue 0 drop-level 2 low-limit 128 high-limit 512
discard-probability 75
[Sysname-wred-table-queue-table1] queue 3 drop-level 0 low-limit 256 high-limit 640
discard-probability 5
[Sysname-wred-table-queue-table1] queue 3 drop-level 1 low-limit 256 high-limit 640
discard-probability 10
[Sysname-wred-table-queue-table1] queue 3 drop-level 2 low-limit 256 high-limit 640
discard-probability 25
[Sysname-wred-table-queue-table1] queue 7 drop-level 0 low-limit 512 high-limit 1024
discard-probability 1
[Sysname-wred-table-queue-table1] queue 7 drop-level 1 low-limit 512 high-limit 1024
discard-probability 5
[Sysname-wred-table-queue-table1] queue 7 drop-level 2 low-limit 512 high-limit 1024
discard-probability 10
[Sysname-wred-table-queue-table1] quit
```

Apply the queue-based WRED table to GigabitEthernet 1/0/2.

```
[Sysname] interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2] qos wred apply queue-table1
[Sysname-GigabitEthernet1/0/2] quit
```

Display and maintenance commands for WRED

Execute **display** commands in any view.

Task	Command
Display WRED configuration and statistics for an interface.	**display qos wred interface** [*interface-type interface-number*]
Display the configuration of a WRED table or all WRED tables.	**display qos wred table** [**name** *table-name*] [**slot** *slot-number*]

Configuring traffic filtering

About traffic filtering

You can filter in or filter out traffic of a class by associating the class with a traffic filtering action. For example, you can filter packets sourced from an IP address according to network status.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match *match-criteria*
      By default, no match criterion is configured.
      For more information about configuring match criteria, see *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure the traffic filtering action.
      filter { **deny** | **permit** }
      By default, no traffic filtering action is configured.
      If a traffic behavior has the **filter deny** action, all other actions in the traffic behavior except class-based accounting do not take effect.
      If a traffic behavior has the filter permit action, matching packets will not dropped by a lower-priority configuration even if the action of the lower-priority configuration is drop.
   c. Return to system view.
      quit
4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.
      qos policy *policy-name*
   b. Associate the traffic class with the traffic behavior in the QoS policy.
      classifier *classifier-name* **behavior** *behavior-name*
      By default, a traffic class is not associated with a traffic behavior.
   c. Return to system view.
      quit
5. Apply the QoS policy.
   For more information, see "Applying the QoS policy."

By default, no QoS policy is applied.

6. (Optional.) Display the traffic filtering configuration.

   display traffic behavior user-defined [*behavior-name*]

   This command is available in any view.

Traffic filtering configuration examples

Example: Configuring traffic filtering

Network configuration

As shown in Figure 19, configure traffic filtering on GigabitEthernet 1/0/1 to deny the incoming packets with destination port number 21.

Figure 19 Network diagram

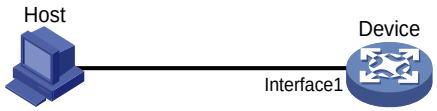


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Create advanced ACL 3000, and configure a rule to match packets whose destination port number is 21.

```
<Sysname> system-view
[Sysname] sysname Device
[Device] acl advanced 3000
[Device-acl-ipv4-adv-3000] rule 0 permit tcp destination-port eq 21
[Device-acl-ipv4-adv-3000] quit
```

Create a traffic class named **classifier_1**, and use ACL 3000 as the match criterion in the traffic class.

```
[Device] traffic classifier classifier_1
[Device-classifier-classifier_1] if-match acl 3000
[Device-classifier-classifier_1] quit
```

Create a traffic behavior named **behavior_1**, and configure the traffic filtering action to drop packets.

```
[Device] traffic behavior behavior_1
[Device-behavior-behavior_1] filter deny
[Device-behavior-behavior_1] quit
```

Create a QoS policy named **policy**, and associate traffic class **classifier_1** with traffic behavior **behavior_1** in the QoS policy.

```
[Device] qos policy policy
[Device-qospolicy-policy] classifier classifier_1 behavior behavior_1
[Device-qospolicy-policy] quit
```

Apply QoS policy **policy** to the incoming traffic of GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] qos apply policy policy inbound
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Use the **display qos policy interface** command to verify that the configuration has been performed successfully.

Configuration files

```
#
 sysname Device
#
traffic classifier classifier_1 operator and
 if-match acl 3000
#
traffic behavior behavior_1
 filter deny
#
qos policy policy
 classifier classifier_1 behavior behavior_1
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 qos apply policy policy inbound
#
acl advanced 3000
 rule 0 permit tcp destination-port eq ftp
#
return
```

Configuring priority marking

About priority marking

Priority marking sets the priority fields or flag bits of packets to modify the priority of packets. For example, you can use priority marking to set IP precedence or DSCP for a class of IP packets to control the forwarding of these packets.

To configure priority marking to set the priority fields or flag bits for a class of packets, perform the following tasks:

1. Configure a traffic behavior with a priority marking action.
2. Associate the traffic class with the traffic behavior.

Priority marking can be used together with priority mapping. For more information, see "Configuring priority mapping."

Configuring priority marking

Restrictions and guidelines

The device supports the following application destinations for priority marking:

- Interface in the inbound direction.
- VLANs in the inbound direction.
- Globally in the inbound direction.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match *match-criteria*
      By default, no match criterion is configured.
      For more information about the **if-match** command, see *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure a priority marking action.
      For configurable priority marking actions, see the **remark** commands in *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
4. Define a QoS policy.

a. Create a QoS policy and enter QoS policy view.
qos policy *policy-name*

b. Associate the traffic class with the traffic behavior in the QoS policy.
classifier *classifier-name* **behavior** *behavior-name*
By default, a traffic class is not associated with a traffic behavior.

c. Return to system view.
quit

5. Apply the QoS policy.
For more information, see "Applying the QoS policy."
By default, no QoS policy is applied.

6. (Optional.) Display the priority marking configuration.
display traffic behavior user-defined [*behavior-name*]
This command is available in any view.

Configuring nesting

About nesting

Nesting adds a VLAN tag to the matching packets to allow the VLAN-tagged packets to pass through the corresponding VLAN. For example, you can add an outer VLAN tag to packets from a customer network to a service provider network. This allows the packets to pass through the service provider network by carrying a VLAN tag assigned by the service provider.

Restrictions and guidelines: Nesting configuration

A VLAN tag adding action takes effect only when a QoS policy is applied to the inbound direction of an interface, VLAN, or globally.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match *match-criteria*
      By default, no match criterion is configured for a traffic class.
      For more information about the match criteria, see the **if-match** command in *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure an outer VLAN tag adding action.
      nest top-most vlan *vlan-id*
      By default, no outer VLAN tag adding action is configured for a traffic behavior.
   c. Return to system view.
      quit
4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.
      qos policy *policy-name*
   b. Associate the traffic class with the traffic behavior in the QoS policy.
      classifier *classifier-name* **behavior** *behavior-name*
      By default, a traffic class is not associated with a traffic behavior.
   c. Return to system view.

```
quit
```

5. Apply the QoS policy.

   For more information, see "Applying the QoS policy."

   By default, no QoS policy is applied.

6. (Optional.) Display the nesting configuration.

   display traffic behavior user-defined [*behavior-name*]

   This command is available in any view.

Nesting configuration examples

Example: Configuring nesting

Network configuration

As shown in Figure 20:

- Site 1 and Site 2 in VPN A are two branches of a company. They use VLAN 5 to transmit traffic.
- Because Site 1 and Site 2 are located in different areas, the two sites use the VPN access service of a service provider. The service provider assigns VLAN 100 to the two sites.

Configure nesting, so that the two branches can communicate through the service provider network.

Figure 20 Network diagram

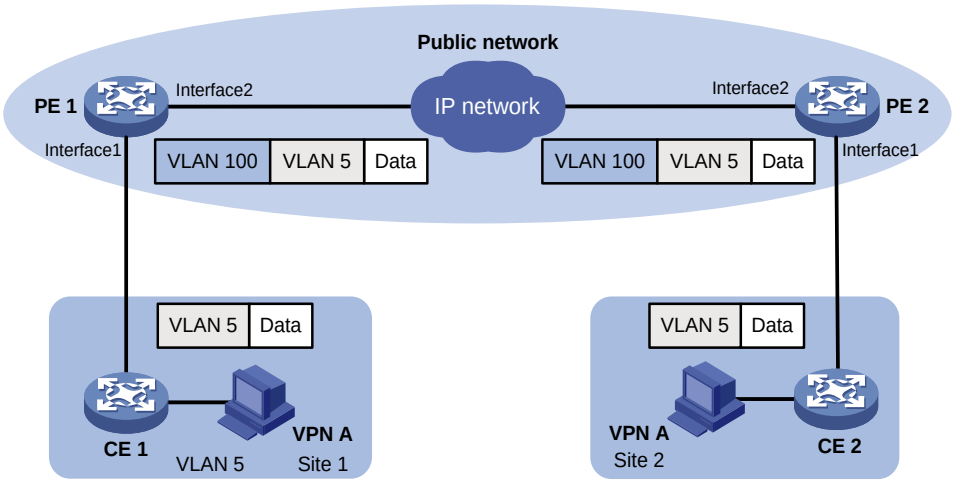


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configuring PE 1:

   # Create a traffic class named **test** to match traffic with VLAN ID 5.

   ```
   <Sysname> system-view
   [Sysname] sysname PE1
   [PE1] traffic classifier test
   ```

```
[PE1-classifier-test] if-match service-vlan-id 5
[PE1-classifier-test] quit
```

Configure an action to add outer VLAN tag 100 in traffic behavior **test**.

```
[PE1] vlan 100
[PE1-vlan100] quit
[PE1] traffic behavior test
[PE1-behavior-test] nest top-most vlan 100
[PE1-behavior-test] quit
```

Create a QoS policy named **test**, and associate class **test** with behavior **test** in the QoS policy.

```
[PE1] qos policy test
[PE1-qospolicy-test] classifier test behavior test
[PE1-qospolicy-test] quit
```

Configure the downlink port (GigabitEthernet 1/0/1) as a hybrid port, and assign the port to VLAN 100 as an untagged member.

```
[PE1] interface gigabitethernet 1/0/1
[PE1-GigabitEthernet1/0/1] port link-type hybrid
[PE1-GigabitEthernet1/0/1] port hybrid vlan 100 untagged
```

Apply QoS policy **test** to the incoming traffic of GigabitEthernet 1/0/1.

```
[PE1-GigabitEthernet1/0/1] qos apply policy test inbound
[PE1-GigabitEthernet1/0/1] quit
```

Configure the uplink port (GigabitEthernet 1/0/2) as a trunk port, and assign it to VLAN 100.

```
[PE1] interface gigabitethernet 1/0/2
[PE1-GigabitEthernet1/0/2] port link-type trunk
[PE1-GigabitEthernet1/0/2] port trunk permit vlan 100
[PE1-GigabitEthernet1/0/2] quit
```

2. Configuring PE 2:

Create a traffic class named **test** to match traffic with VLAN ID 5.

```
<Sysname> system-view
[Sysname] sysname PE2
[PE2] traffic classifier test
[PE2-classifier-test] if-match service-vlan-id 5
[PE2-classifier-test] quit
```

Configure an action to add outer VLAN tag 100 in traffic behavior **test**.

```
[PE2] vlan 100
[PE2-vlan100] quit
[PE2] traffic behavior test
[PE2-behavior-test] nest top-most vlan 100
[PE2-behavior-test] quit
```

Create a QoS policy named **test**, and associate class **test** with behavior **test** in the QoS policy.

```
[PE2] qos policy test
[PE2-qospolicy-test] classifier test behavior test
[PE2-qospolicy-test] quit
```

Configure the downlink port (GigabitEthernet 1/0/1) as a hybrid port, and assign the port to VLAN 100 as an untagged member.

```
[PE2] interface gigabitethernet 1/0/1
[PE2-GigabitEthernet1/0/1] port link-type hybrid
```

```
[PE2-GigabitEthernet1/0/1] port hybrid vlan 100 untagged
```

Apply QoS policy **test** to the incoming traffic of GigabitEthernet 1/0/1.

```
[PE2-GigabitEthernet1/0/1] qos apply policy test inbound
[PE2-GigabitEthernet1/0/1] quit
```

Configure the uplink port (GigabitEthernet 1/0/2) as a trunk port, and assign it to VLAN 100.

```
[PE2] interface gigabitethernet 1/0/2
[PE2-GigabitEthernet1/0/2] port link-type trunk
[PE2-GigabitEthernet1/0/2] port trunk permit vlan 100
[PE2-GigabitEthernet1/0/2] quit
[PE2] commit
```

Verifying the configuration

Site 1 and site 2 can access each other.

Configuration files

- PE 1:

```
#
 sysname PE1
#
vlan 100
#
traffic classifier test operator and
 if-match service-vlan-id 5
#
traffic behavior test
 nest top-most vlan 100
#
qos policy test
 classifier test behavior test
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port link-type hybrid
 port hybrid vlan 1 100 untagged
 qos apply policy test inbound
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 100
#
return
```

- PE 2:

```
#
 sysname PE2
#
vlan 100
#
traffic classifier test operator and
```

```
 if-match service-vlan-id 5
#
traffic behavior test
 nest top-most vlan 100
#
qos policy test
 classifier test behavior test
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 port link-type hybrid
 port hybrid vlan 1 100 untagged
 qos apply policy test inbound
#
interface GigabitEthernet1/0/2
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 100
#
return
```

Configuring traffic redirecting

About traffic redirecting

Traffic redirecting redirects packets matching the specified match criteria to a location for processing.

You can redirect packets to the following destinations:

- CPU.
- Interface.
- Next hop.

Restrictions and guidelines: Traffic redirecting configuration

- A traffic redirecting action takes effect only when a QoS policy is applied to the inbound direction of the following application destinations:
  - Interface.
  - VLANs.
  - Globally.
- If you execute the **redirect** command multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match *match-criteria*
      By default, no match criterion is configured for a traffic class.
      For more information about the match criteria, see the **if-match** command in *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure a traffic redirecting action.
      See the **redirect** commands in the command reference.
      By default, no traffic redirecting action is configured for a traffic behavior.
   c. Return to system view.

quit

4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.
      qos policy *policy-name*
   b. Associate the traffic class with the traffic behavior in the QoS policy.
      classifier *classifier-name* **behavior** *behavior-name*
      By default, a traffic class is not associated with a traffic behavior.
   c. Return to system view.
      quit
5. Apply the QoS policy.
   For more information, see "Applying the QoS policy."
   By default, no QoS policy is applied.
6. (Optional.) Display traffic redirecting configuration.
   display traffic behavior user-defined [*behavior-name*]
   This command is available in any view.

Traffic redirecting configuration examples

Example: Configuring traffic redirecting

Network configuration

As shown in Figure 21:

- Device A is connected to Device B through two links. Device A and Device B are each connected to other devices.
- GigabitEthernet 1/0/1 of Device A is a trunk port and belongs to VLAN 200 and VLAN 201.
- GigabitEthernet 1/0/2 of Device A and GigabitEthernet 1/0/2 of Device B belong to VLAN 200.
- GigabitEthernet 1/0/3 of Device A and GigabitEthernet 1/0/3 of Device B belong to VLAN 201.
- On Device A, the IP address of VLAN-interface 200 is 200.1.1.1/24, and that of VLAN-interface 201 is 201.1.1.1/24.
- On Device B, the IP address of VLAN-interface 200 is 200.1.1.2/24, and that of VLAN-interface 201 is 201.1.1.2/24.

Configure the actions of redirecting traffic to an interface to meet the following requirements:

- Packets with source IP address 2.1.1.1 received on GigabitEthernet 1/0/1 of Device A are forwarded to GigabitEthernet 1/0/2.
- Packets with source IP address 2.1.1.2 received on GigabitEthernet 1/0/1 of Device A are forwarded to GigabitEthernet 1/0/3.
- Other packets received on GigabitEthernet 1/0/1 of Device A are forwarded according to the routing table.

Figure 21 Network diagram

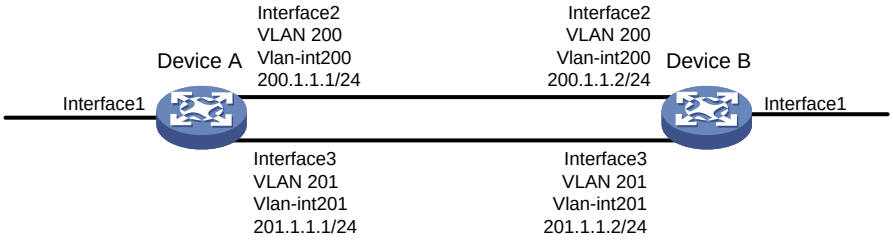


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Create basic ACL 2000, and configure a rule to match packets with source IP address 2.1.1.1.

```
<Sysname> system-view
[Sysname] sysname DeviceA
[DeviceA] acl basic 2000
[DeviceA-acl-ipv4-basic-2000] rule permit source 2.1.1.1 0
[DeviceA-acl-ipv4-basic-2000] quit
```

Create basic ACL 2001, and configure a rule to match packets with source IP address 2.1.1.2.

```
[DeviceA] acl basic 2001
[DeviceA-acl-ipv4-basic-2001] rule permit source 2.1.1.2 0
[DeviceA-acl-ipv4-basic-2001] quit
```

Create a traffic class named **classifier_1**, and use ACL 2000 as the match criterion in the traffic class.

```
[DeviceA] traffic classifier classifier_1
[DeviceA-classifier-classifier_1] if-match acl 2000
[DeviceA-classifier-classifier_1] quit
```

Create a traffic class named **classifier_2**, and use ACL 2001 as the match criterion in the traffic class.

```
[DeviceA] traffic classifier classifier_2
[DeviceA-classifier-classifier_2] if-match acl 2001
[DeviceA-classifier-classifier_2] quit
```

Create a traffic behavior named **behavior_1**, and configure the action of redirecting traffic to GigabitEthernet 1/0/2.

```
[DeviceA] traffic behavior behavior_1
[DeviceA-behavior-behavior_1] redirect interface gigabitethernet 1/0/2
[DeviceA-behavior-behavior_1] quit
```

Create a traffic behavior named **behavior_2**, and configure the action of redirecting traffic to GigabitEthernet 1/0/3.

```
[DeviceA] traffic behavior behavior_2
[DeviceA-behavior-behavior_2] redirect interface gigabitethernet 1/0/3
[DeviceA-behavior-behavior_2] quit
```

Create a QoS policy named **policy**.

```
[DeviceA] qos policy policy
```

Associate traffic class **classifier_1** with traffic behavior **behavior_1** in the QoS policy.

```
[DeviceA-qospolicy-policy] classifier classifier_1 behavior behavior_1
```

Associate traffic class **classifier_2** with traffic behavior **behavior_2** in the QoS policy.

```
[DeviceA-qospolicy-policy] classifier classifier_2 behavior behavior_2
[DeviceA-qospolicy-policy] quit
```

Apply QoS policy **policy** to the incoming traffic of GigabitEthernet 1/0/1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] qos apply policy policy inbound
[DeviceA-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Use the **`display qos policy interface`** command to verify that the configuration has been performed successfully.

Configuration files

```
#
 sysname DeviceA
#
traffic classifier classifier_1 operator and
 if-match acl 2000
#
traffic classifier classifier_2 operator and
 if-match acl 2001
#
traffic behavior behavior_1
 redirect interface GigabitEthernet1/0/2
#
traffic behavior behavior_2
 redirect interface GigabitEthernet1/0/3
#
qos policy policy
 classifier classifier_1 behavior behavior_1
 classifier classifier_2 behavior behavior_2
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 qos apply policy policy inbound
#
acl basic 2000
 rule 5 permit source 2.1.1.1 0
#
acl basic 2001
 rule 5 permit source 2.1.1.2 0
#
return
```

Configuring global CAR

About global CAR

Global committed access rate (CAR) is an approach to policing traffic flows globally. It adds flexibility to common CAR where traffic policing is performed only on a per-traffic class or per-interface basis. In this approach, CAR actions are created in system view and each can be used to police multiple traffic flows as a whole.

Global CAR provides the following CAR actions: aggregate CAR and hierarchical CAR.

Aggregate CAR

An aggregate CAR action is created globally. It can be directly applied to interfaces or used in the traffic behaviors associated with different traffic classes to police multiple traffic flows as a whole. The total rate of the traffic flows must conform to the traffic policing specifications set in the aggregate CAR action.

Hierarchical CAR

A hierarchical CAR action is created globally. It must be used in conjunction with a common CAR or aggregate CAR action. With a hierarchical CAR action, you can limit the total traffic of multiple traffic classes.

A hierarchical CAR action can be used in the common or aggregate CAR action for a traffic class in either AND mode or OR mode.

- In AND mode, the rate of the traffic class is strictly limited under the common or aggregate CAR. This mode applies to flows that must be strictly rate limited.
- In OR mode, the traffic class can use idle bandwidth of other traffic classes associated with the hierarchical CAR. This mode applies to high priority, bursty traffic like video.

By using the two modes appropriately, you can improve bandwidth efficiency.

For example, suppose two flows exist: a low priority data flow and a high priority, bursty video flow. Their total traffic rate cannot exceed 4096 kbps and the video flow must be assured of at least 2048 kbps bandwidth. You can perform the following tasks:

- Configure common CAR actions to set the traffic rate to 2048 kbps for the two flows.
- Configure a hierarchical CAR action to limit their total traffic rate to 4096 kbps.
- Use the action in AND mode in the common CAR action for the data flow.
- Use the action in OR mode in the common CAR action for the video flow.

The video flow is assured of 2048 kbps bandwidth and can use idle bandwidth of the data flow.

In a bandwidth oversubscription scenario, the uplink port bandwidth is lower than the total downlink port traffic rate. You can use hierarchical CAR to meet the following requirements:

- Limit the total rate of downlink port traffic.
- Allow each downlink port to forward traffic at the maximum rate when the other ports are idle.

For example, you can perform the following tasks:

- Use common CAR actions to limit the rates of Internet access flow 1 and flow 2 to both 128 kbps.
- Use a hierarchical CAR action to limit their total traffic rate to 192 kbps.

- Use the hierarchical CAR action for both flow 1 and flow 2 in AND mode.

When flow 1 is not present, flow 2 is transmitted at the maximum rate, 128 kbps. When both flows are present, the total rate of the two flows cannot exceed 192 kbps. As a result, the traffic rate of flow 2 might drop below 128 kbps.

Restrictions and guidelines: Global CAR configuration

The device supports only aggregate CAR in the current software version, and an aggregate CAR action cannot be applied in both the inbound direction and outbound direction.

When a QoS policy containing an aggregate CAR action is applied on an IRF fabric, the traffic matching the QoS policy might enter the IRF fabric through interfaces on different IRF member devices. In this case, the actual rate limit that takes effect is the sum of the CIR and PIR in the aggregate CAR action multiplied by the number of involved IRF member devices by default.

Configuring aggregate CAR

1. Enter system view.

   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.

      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.

      if-match *match-criteria*

      By default, no match criterion is configured.

      For configurable match criteria, see the **if-match** command in *ACL and QoS Command Reference*.
   c. Return to system view.

      quit
3. Configure an aggregate CAR action.

   qos car *car-name* **aggregative cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**green** *action* | **red** *action* | **yellow** *action*] *

   qos car *car-name* **aggregative cir** *committed-information-rate* [**cbs** *committed-burst-size*] **pir** *peak-information-rate* [**ebs** *excess-burst-size*] [**green** *action* | **red** *action* | **yellow** *action*] *

   By default, no aggregate CAR action is configured.
4. Define a traffic behavior.
   a. Enter traffic behavior view.

      traffic behavior *behavior-name*
   b. Use the aggregate CAR in the traffic behavior.

      car name *car-name*

      By default, no aggregate CAR action is used in a traffic behavior.
5. Apply the QoS policy.

   For more information, see "Applying the QoS policy."

   By default, no QoS policy is applied.

Display and maintenance commands for global CAR

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display statistics for global CAR actions.	**display qos car name** [*car-name*]
Clear statistics for global CAR actions.	**reset qos car name** [*car-name*]

Global CAR configuration examples

Example: Configuring MQC aggregate CAR

Network configuration

As shown in Figure 22, configure MQC aggregate CAR to limit the total rate of traffic in VLAN 10 and VLAN 100 to CIR 2560 kbps and CBS 20480 bytes and drop the exceeding packets.

Figure 22 Network diagram

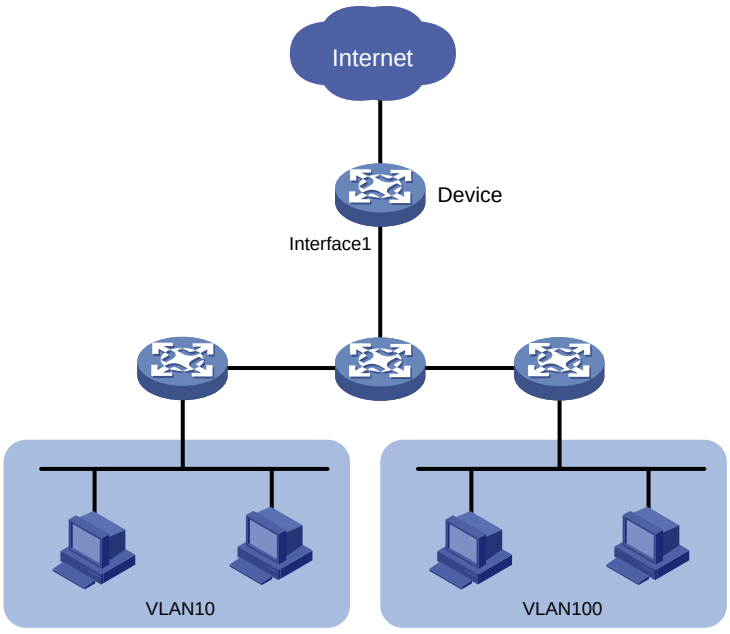


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Configure an aggregate CAR action according to the rate limit requirements.

```
<Sysname> system-view
[Sysname] sysname Device
```

```
[Device] qos car aggcar-1 aggregative cir 2560 cbs 20480 red discard
```

Create traffic class 1, and use VLAN ID 10 as the match criterion in the traffic class.

```
[Device] traffic classifier 1
[Device-classifier-1] if-match service-vlan-id 10
[Device-classifier-1] quit
```

Create traffic behavior 1, and use the aggregate CAR action in the traffic behavior.

```
[Device] traffic behavior 1
[Device-behavior-1] car name aggcar-1
[Device-behavior-1] quit
```

Create traffic class 2, and use VLAN ID 100 as the match criterion in the traffic class.

```
[Device] traffic classifier 2
[Device-classifier-2] if-match service-vlan-id 100
[Device-classifier-2] quit
```

Create traffic behavior 2, and use the aggregate CAR action in the traffic behavior.

```
[Device] traffic behavior 2
[Device-behavior-2] car name aggcar-1
[Device-behavior-2] quit
```

Create a QoS policy named **car**, and associate traffic classes 1 and 2 with traffic behaviors 1 and 2 in the QoS policy, respectively.

```
[Device] qos policy car
[Device-qospolicy-car] classifier 1 behavior 1
[Device-qospolicy-car] classifier 2 behavior 2
[Device-qospolicy-car] quit
```

Apply QoS policy **car** to the incoming traffic of GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] qos apply policy car inbound
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Use the **display qos policy interface** command to verify that the configuration has been performed successfully.

Configuration files

```
#
 sysname Device
#
 qos car aggcar-1 aggregative cir 2560 cbs 20480 ebs 0 green pass red discard yellow pass
#
traffic classifier 1 operator and
 if-match service-vlan-id 10
#
traffic classifier 2 operator and
 if-match service-vlan-id 100
#
traffic behavior 1
 car name aggcar-1
#
traffic behavior 2
```

```
 car name aggcar-1
#
qos policy car
 classifier 1 behavior 1
 classifier 2 behavior 2
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 qos apply policy car inbound
#
return
```

Configuring class-based accounting

About class-based accounting

Class-based accounting collects statistics on a per-traffic class basis. For example, you can define the action to collect statistics for traffic sourced from a certain IP address. By analyzing the statistics, you can determine whether anomalies have occurred and what action to take.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match *match-criteria*
      By default, no match criterion is configured.
      For more information about the **if-match** command, see *ACL and QoS Command Reference*.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure an accounting action.
      accounting [**byte** | **packet**]
      By default, no traffic accounting action is configured.
   c. Return to system view.
      quit
4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.
      qos policy *policy-name*
   b. Associate the traffic class with the traffic behavior in the QoS policy.
      classifier *classifier-name* **behavior** *behavior-name*
      By default, a traffic class is not associated with a traffic behavior.
   c. Return to system view.
      quit
5. Apply the QoS policy.
   For more information, see "Applying the QoS policy."
   By default, no QoS policy is applied.
6. (Optional.) Display the class-based accounting configuration.
   display traffic behavior user-defined [*behavior-name*]

Class-based accounting configuration examples

Example: Configuring class-based accounting

Network configuration

As shown in Figure 23, configure class-based accounting on GigabitEthernet 1/0/1 to collect statistics for incoming traffic from 1.1.1.1/24.

Figure 23 Network diagram



Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Create basic ACL 2000, and configure a rule to match packets with source IP address 1.1.1.1.

```
<Sysname> system-view
[Sysname] sysname Device
[Device] acl basic 2000
[Device-acl-ipv4-basic-2000] rule permit source 1.1.1.1 0
[Device-acl-ipv4-basic-2000] quit
```

Create a traffic class named **classifier_1**, and use ACL 2000 as the match criterion in the traffic class.

```
[Device] traffic classifier classifier_1
[Device-classifier-classifier_1] if-match acl 2000
[Device-classifier-classifier_1] quit
```

Create a traffic behavior named **behavior_1**, and configure the class-based accounting action.

```
[Device] traffic behavior behavior_1
[Device-behavior-behavior_1] accounting packet
[Device-behavior-behavior_1] quit
```

Create a QoS policy named **policy**, and associate traffic class **classifier_1** with traffic behavior **behavior_1** in the QoS policy.

```
[Device] qos policy policy
[Device-qospolicy-policy] classifier classifier_1 behavior behavior_1
[Device-qospolicy-policy] quit
```

Apply QoS policy **policy** to the incoming traffic of GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] qos apply policy policy inbound
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Display traffic statistics to verify the configuration.

```
[Device] display qos policy interface gigabitethernet 1/0/1
Interface: GigabitEthernet1/0/1
  Direction: Inbound
  Policy: policy
   Classifier: classifier_1
     Operator: AND
     Rule(s) :
      If-match acl 2000
     Behavior: behavior_1
      Accounting enable:
        28529 (Packets)
```

Configuration files

```
#
 sysname Device
#
traffic classifier classifier_1 operator and
 if-match acl 2000
#
traffic behavior behavior_1
 accounting packet
#
qos policy policy
 classifier classifier_1 behavior behavior_1
#
interface GigabitEthernet1/0/1
 port link-mode bridge
 qos apply policy policy inbound
#
acl basic 2000
 rule 5 permit source 1.1.1.1 0
#
return
```

Setting an alarm threshold for protocol packets

About an alarm threshold for protocol packets

The device generates an alarm when the rate of protocol packets sent to the CPU exceeds the alarm threshold.

Restrictions and guidelines: Alarm threshold for protocol packets configuration

This feature does not support the **commit** command to deploy target configuration. For more information about target configuration, see target configuration management in *Fundamentals Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Set an alarm threshold for protocol packets sent to the CPU.
   qos soft-car protocol *protocol-type* **alarm-threshold** *alarm-threshold-rate* **slot** *slot-number*
   By default, the device does not generate alarms for protocol packets sent to the CPU.

Display and maintenance commands for alarm threshold for protocol packets

Execute **display** commands in any view.

Task	Command
Display the alarm threshold configuration for protocol packets.	**display qos soft-car protocol** [*protocol-type*] **slot** *slot-number*

Appendixes

Appendix A Acronyms

Table 9 Appendix A Acronyms

Acronym	Full spelling
AF	Assured Forwarding
BE	Best Effort
BQ	Bandwidth Queuing
CAR	Committed Access Rate
CBS	Committed Burst Size
CBQ	Class Based Queuing
CE	Congestion Experienced
CIR	Committed Information Rate
CQ	Custom Queuing
DCBX	Data Center Bridging Exchange Protocol
DiffServ	Differentiated Service
DSCP	Differentiated Services Code Point
EBS	Excess Burst Size
ECN	Explicit Congestion Notification
EF	Expedited Forwarding
FIFO	First in First out
FQ	Fair Queuing
GMB	Guaranteed Minimum Bandwidth
GTS	Generic Traffic Shaping
INT	Inband Network Telemetry
IntServ	Integrated Service
ISP	Internet Service Provider
LLQ	Low Latency Queuing
LSP	Label Switched Path
MPLS	Multiprotocol Label Switching
PE	Provider Edge
PIR	Peak Information Rate
PQ	Priority Queuing
PW	Pseudowire
QoS	Quality of Service
QPPB	QoS Policy Propagation Through the Border Gateway Protocol

Acronym	Full spelling
RED	Random Early Detection
RSVP	Resource Reservation Protocol
RTP	Real-Time Transport Protocol
SP	Strict Priority
ToS	Type of Service
VoIP	Voice over IP
VPN	Virtual Private Network
WFQ	Weighted Fair Queuing
WRED	Weighted Random Early Detection
WRR	Weighted Round Robin

Appendix B Default priority maps

priority maps

For the default **dscp-dscp**, **exp-lp**, and **lp-exp** priority maps, an input value yields a target value equal to it.

Table 10 Default dot1p-lp and dot1p-dp priority map

Input priority value	dot1p-lp map	dot1p-dp map
dot1p	**lp**	**dp**
0	2	0
1	0	0
2	1	0
3	3	0
4	4	0
5	5	0
6	6	0
7	7	0

Table 11 Default dscp-lp priority map

Input priority value	dscp-lp map	dscp-dp map
dscp	**lp**	**dp**
0 to 7	0	0
8 to 15	1	0
16 to 23	2	0
24 to 31	3	0

Input priority value	dscp-lp map	dscp-dp map
32 to 39	4	0
40 to 47	5	0
48 to 55	6	0
56 to 63	7	0

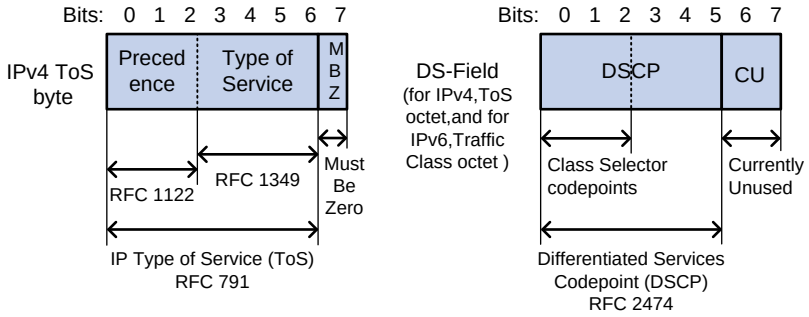
Table 12 Default port priority-local priority map

Port priority	Local precedence
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7

Appendix C Introduction to packet precedence

IP precedence and DSCP values

Figure 24 ToS and DS fields



As shown in Figure 24, the ToS field in the IP header contains 8 bits. The first 3 bits (0 to 2) represent IP precedence from 0 to 7. According to RFC 2474, the ToS field is redefined as the differentiated services (DS) field. A DSCP value is represented by the first 6 bits (0 to 5) of the DS field and is in the range 0 to 63. The remaining 2 bits (6 and 7) are reserved.

Table 13 IP precedence

IP precedence (decimal)	IP precedence (binary)	Description
0	000	Routine
1	001	priority
2	010	immediate

IP precedence (decimal)	IP precedence (binary)	Description
3	011	flash
4	100	flash-override
5	101	critical
6	110	internet
7	111	network

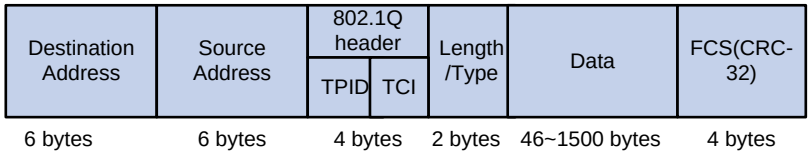
Table 14 DSCP values

DSCP value (decimal)	DSCP value (binary)	Description
46	101110	ef
10	001010	af11
12	001100	af12
14	001110	af13
18	010010	af21
20	010100	af22
22	010110	af23
26	011010	af31
28	011100	af32
30	011110	af33
34	100010	af41
36	100100	af42
38	100110	af43
8	001000	cs1
16	010000	cs2
24	011000	cs3
32	100000	cs4
40	101000	cs5
48	110000	cs6
56	111000	cs7
0	000000	be (default)

802.1p priority

802.1p priority lies in the Layer 2 header. It applies to occasions where Layer 3 header analysis is not needed and QoS must be assured at Layer 2.

Figure 25 An Ethernet frame with an 802.1Q tag header



As shown in Figure 25, the 4-byte 802.1Q tag header contains the 2-byte tag protocol identifier (TPID) and the 2-byte tag control information (TCI). The value of the TPID is 0x8100. Figure 26 shows the format of the 802.1Q tag header. The Priority field in the 802.1Q tag header is called 802.1p priority, because its use is defined in IEEE 802.1p. Figure 26Table 15 shows the values for 802.1p priority.

Figure 26 802.1Q tag header

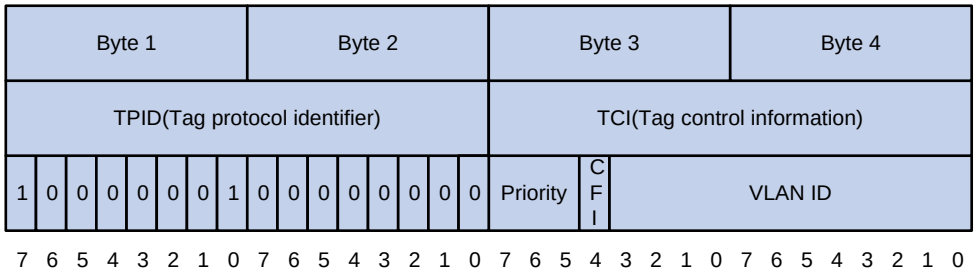


Table 15 Description on 802.1p priority

802.1p priority (decimal)	802.1p priority (binary)	Description
0	000	best-effort
1	001	background
2	010	spare
3	011	excellent-effort
4	100	controlled-load
5	101	video
6	110	voice
7	111	network-management

Contents

Configuring data buffers 1
About data buffers 1
Data buffer types 1
Cell resources 1
Fixed area and shared area 1
Restrictions and guidelines: Data buffer configuration 2
Data buffer tasks at a glance 2
Configuring data buffer alarms 3
About data buffer alarms 3
Restrictions and guidelines for data buffer alarm configuration 3
Configuring per-interface alarm thresholds for the ingress or egress buffer 3
Configuring per-queue alarm thresholds for the ingress or egress buffer 3
Configuring per-interface alarm thresholds for the Headroom buffer 4
Configuring packet-drop alarms 4
Configuring TCB 5
About TCB 5
Overview 5
How TBC works 5
Restrictions and guidelines: TCB configuration 6
Enabling and configuring TCB 6
Enabling and configuring TCB in system view 6
Enabling and configuring TCB in interface view 7
Display and maintenance commands for TCB 7
TCB configuration examples 7
Example: Configuring TCB on an interface 7

Configuring data buffers

About data buffers

Sudden traffic in a short period on the network might cause congestion, while device interfaces have limited processing capabilities, potentially leading to packet loss. Therefore, high-speed forwarding devices use data buffers to cache packets, reducing the probability of packet loss due to congestion. If the network experiences significant sudden traffic, resources in the data buffer might be exhausted, making it impossible to continue caching packets, which might still be discarded. Properly allocating data buffer resources allows high-priority queues or interface traffic to be cached, minimizing congestion impact.

The size and processing capability of the buffer depend entirely on the hardware forwarding chip and interfaces. Generally, all interfaces on the forwarding chip share the chip's data buffer resources, while queues on an interface share the interface's data buffer resources.

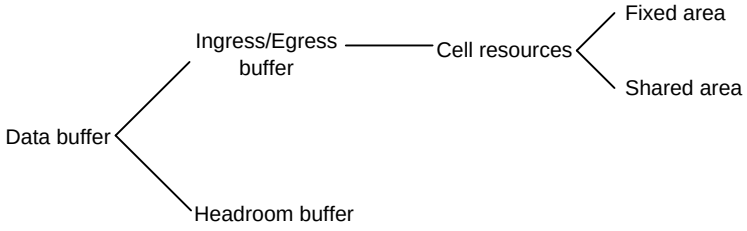
Data buffer types

The following data buffers are available:

- **Ingress buffer**—Stores incoming packets when the CPU is busy.
- **Egress buffer**—Stores outgoing packets when network congestion occurs.
- **Headroom buffer**—Stores packets when the ingress buffer or egress buffer is used up. The headroom buffer is mainly used for the Priority-based Flow Control (PFC) feature. There is a period of time between when the local PFC triggers the sending of the PFC PAUSE frame and when the peer end stops sending packets. During this period of time, the local device uses the headroom data buffer to cache packets. For more information about PFC, see PFC configuration in *Intelligent Lossless Network Configuration Guide*.

Figure 1 shows the structure of ingress and egress buffers.

Figure 1 Data buffer structure



Cell resources

A buffer uses cell resources to store packets based on packet sizes. Suppose a cell resource provides 208 bytes. The buffer allocates one cell resource to a 128-byte packet and two cell resources to a 300-byte packet.

Fixed area and shared area

The cell resources have a fixed area and a shared area.

- **Fixed area**—Partitioned into queues, each of which is equally divided by all the interfaces on the switch, as shown in Figure 2.When congestion occurs or the CPU is busy, the following rules apply:

a. An interface first uses the relevant queues of the fixed area to store packets.

b. When a queue is full, the interface uses the corresponding queue of the shared area.

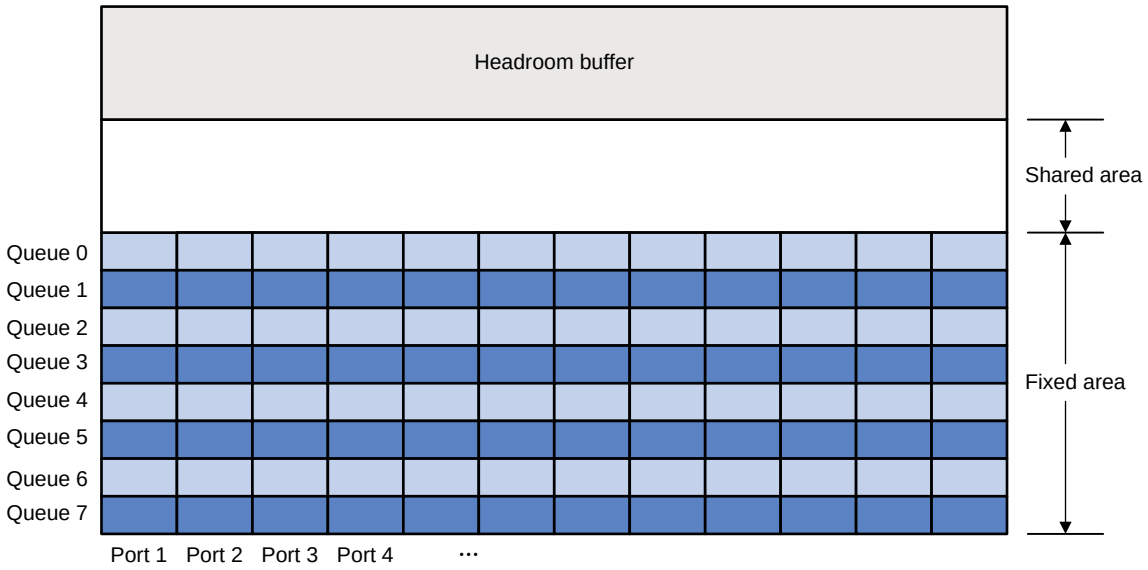
c. When the queue in the shared area is also full, the interface discards subsequent packets.

The system allocates the fixed area among queues as specified by the user. Even if a queue is not full, other queues cannot preempt its space. Similarly, the share of a queue for an interface cannot be preempted by other interfaces even if it is not full.

- **Shared area**—Partitioned into queues, each of which is not equally divided by the interfaces, as shown in Figure 2. The system determines the actual shared-area space for each queue according to user configuration and the number of packets actually received and sent. If a queue is not full, other queues can preempt its space.

  The system puts packets received or sent on all interfaces into a queue in the order they arrive. When the queue is full, subsequent packets are dropped.

Figure 2 Fixed area and shared area



Restrictions and guidelines: Data buffer configuration

IRF physical interfaces do not support this feature.

Data buffer tasks at a glance

To configure the data buffer, perform the following tasks:

- (Optional.) Configuring data buffer alarms
  - Configuring per-interface alarm thresholds for the ingress or egress buffer
  - Configuring per-queue alarm thresholds for the ingress or egress buffer
  - Configuring per-interface alarm thresholds for the Headroom buffer
  - Configuring packet-drop alarms

Configuring data buffer alarms

About data buffer alarms

This feature works with a network management system (such as IMC). Data buffer alarms include threshold-crossing alarms and packet drop alarms. The device reports these alarms to the network management system for displaying the data buffer usage.

You can configure data buffer alarms on a per-interface basis or a per-queue basis.

Restrictions and guidelines for data buffer alarm configuration

Configuring per-queue alarm thresholds is supported for only the egress buffer.

Configuring per-interface alarm thresholds for the ingress or egress buffer

1. Enter system view.

   system-view
2. Configure the alarm thresholds. Choose the options to configure as needed:
   - Configure the global alarm threshold.

     buffer { **egress** | **ingress** } **usage threshold slot** *slot-number* **null-queue ratio** *ratio*

     The default setting is 100%.
   - Execute the following commands in sequence to configure the alarm threshold on an interface:

     interface *interface-type interface-number*

     port buffer { **egress** | **ingress** } **usage threshold null-queue ratio** *ratio*

     By default, an interface uses the global alarm threshold.

   If you configure both the global alarm threshold and the alarm threshold on an interface, the latter takes effect.
3. Enable threshold-crossing alarms.

   buffer threshold alarm { **egress** | **ingress** } **enable**

   By default, threshold-crossing alarms are disabled.
4. (Optional.) Set the interval for sending threshold-crossing alarms.

   buffer threshold alarm { **egress** | **ingress** } **interval** *interval*

   The default setting is 5 seconds.

Configuring per-queue alarm thresholds for the ingress or egress buffer

1. Enter system view.

   system-view
2. Configure the alarm thresholds.

- Execute the following commands in sequence to configure the alarm threshold for a queue on an interface:

  interface *interface-type interface-number*

  buffer egress usage threshold queue *queue-id* **ratio** *ratio*

  The default setting is 100%.

3. Enable threshold-crossing alarms.

   buffer threshold alarm { **egress** | **ingress** } **enable**

   By default, threshold-crossing alarms are disabled.

4. (Optional.) Set the interval for sending threshold-crossing alarms.

   buffer threshold alarm { **egress** | **ingress** } **interval** *interval*

   The default setting is 5 seconds.

Configuring per-interface alarm thresholds for the Headroom buffer

1. Enter system view.

   system-view

2. Configure the alarm thresholds. Choose the options to configure as needed:
   - Configure the global alarm threshold.

     buffer usage threshold headroom slot *slot-number* **ratio** *ratio*

     The default setting is 100%.
   - Execute the following commands in sequence to configure the alarm threshold on an interface:

     interface *interface-type interface-number*

     port buffer usage threshold headroom null-queue ratio *ratio*

     By default, an interface uses the global alarm threshold.

   If you configure both the global alarm threshold and the alarm threshold on an interface, the latter takes effect.

3. Enable threshold-crossing alarms.

   buffer threshold alarm headroom enable

   By default, threshold-crossing alarms are disabled.

4. (Optional.) Set the interval for sending threshold-crossing alarms.

   buffer threshold alarm headroom interval *interval*

   The default setting is 5 seconds.

Configuring packet-drop alarms

About this task

This feature allows the device to periodically send packet-drop information to the NMS.

Restrictions and guidelines

This feature does not take effect on the Headroom buffer.

Procedure

1. Enter system view.

   system-view

2. Enable packet-drop alarms.

   buffer packet-drop alarm enable

   By default, packet-drop alarms are disabled.

3. (Optional.) Set the interval for sending packet-drop alarms.

   buffer packet-drop alarm interval *interval*

   The default setting is 5 seconds.

Configuring TCB

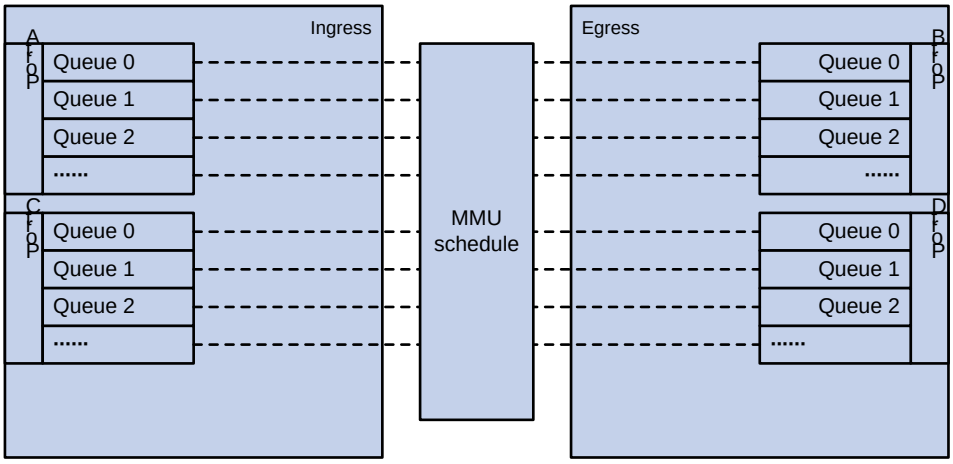
About TCB

Overview

The Transient Capture Buffer (TCB) feature monitors packet drop events on a set of memory management unit (MMU) resources. When a packet is dropped on a queue, the system collects drop time, drop reason, packet metadata, and other information and reports them to an NMS through gRPC.

An MMU manages the allocation of the ingress buffer and egress buffer and flexibly schedules queues to deal with traffic bursts.

Figure 3 MMU queues



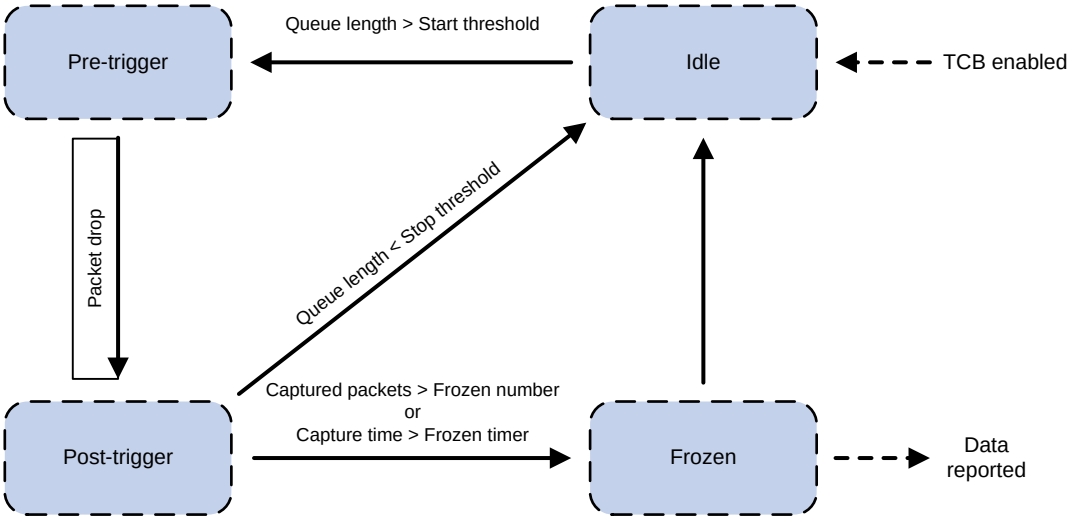
How TBC works

The TCB state machine works as follows:

1. The state machine enters the idle state when TCB is enabled.
2. When the queue length is greater than *start-threshold-value*, the state machine transitions from the idle state to the pre-trigger state. The system captures packets by using the **pre-sample-rate**.
3. When packet drops occur on the queue, the state machine transitions from the pre-trigger to the post-trigger state.
4. When the queue length is smaller than *stop-threshold-value*, the state machine transitions from the post-trigger to the idle state. The system stops capturing packets.
5. When the queue length is greater than *start-threshold-value*, step 2 is repeated.

6. When the number of captured packets reaches the *frozen-number* or the capture timer expires, the state machine transitions from the post-trigger to the frozen state.
7. The state machine analyzes the drop reason and dropped packet characteristics and sends the analysis results to the gRPC module, which reports them to an NMS.
8. The state machine transitions from the frozen to the idle state.

Figure 4 TCB state machine



Restrictions and guidelines: TCB configuration

You can configure TCB in system view or interface view, but not in both views at the same time. For example, if you have configured TCB in system view, to configure TCB in interface view, you must first delete the TCB configuration in system view.

The TCB configuration in system view takes effect on all TCB-capable interfaces. The TCB configuration in interface view takes effect only on the specified interface.

Only one interface on the device can be configured with TCB when you configure TCB in interface view.

TCB cannot identify tunnel-encapsulated TCP or UDP packets if local analysis is enabled or an ACL is referenced.

During congestion, packet drop information is displayed for only TCP or UDP packets.

The gRPC sampling paths for TCB are **tcb/tcbpacketinfoevent** and **tcb/tcbrawpacketinfoevent**. For more information about gRPC, see "Configuring gRPC."

Enabling and configuring TCB

Enabling and configuring TCB in system view

1. Enter system view.

   system-view
2. Enable and configure TCB.

   buffer transient-capture global [**slot** *slot-number*] **egress enable** [**no-local-analysis**] [**acl**{ *ipv4-acl-number*|**name** *ipv4-acl-name*}]

[**start-threshold** *start-threshold-value* **stop-threshold** *stop-threshold-value*] [**frozen-number** *frozen-number* **frozen-timer** *timer-value*] [**pre-sample-rate** *pre-sample-rate* **post-sample-rate** *post-sample-rate*] [**poll-frequency** *frequency-value*]

By default, TCB is disabled.

Enabling and configuring TCB in interface view

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable and configure TCB.

   buffer transient-capture egress [**queue** *queue-id*] **enable** [**no-local-analysis**] [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* }] [**start-threshold** *start-threshold-value* **stop-threshold** *stop-threshold-value*] [**frozen-number** *frozen-number* **frozen-timer** *timer-value*] [**pre-sample-rate** *pre-sample-rate* **post-sample-rate** *post-sample-rate*] [**poll-frequency** *frequency-value*]

   By default, TCB is disabled.

Display and maintenance commands for TCB

Execute **display** commands in any view.

Task	Command
Display packet drop information.	**display buffer transient-capture packet-drop slot** *slot-number* { **ethernet** | **ipv4** | **ipv6** }

TCB configuration examples

Example: Configuring TCB on an interface

Network configuration

As shown in Figure 5, configure TCB and gRPC on the device to monitor packet drops and report packet drop information to the collector through gRPC.

Figure 5 Network diagram

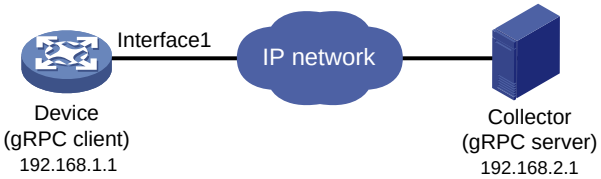


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Assign IP addresses to the device and the collector and configure routes. Make sure the network connections are available. (Details not shown.)
2. Configure TCB:

   # Create an IPv4 advanced ACL numbered 3001, and configure a rule to match packets with source IP address 192.168.5.2.

   ```
   <Device> system-view
   [Device] acl advanced 3001
   [Device-acl-ipv4-adv-3001] rule permit ip source 192.168.5.2 0
   [Device-acl-ipv4-adv-3001] quit
   ```

   # Configure TCB on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] buffer transient-capture egress queue 1 enable no-local-analysis acl 3001 start-threshold 10000 stop-threshold 5000 frozen-number 1000 frozen-timer 400 pre-sample-rate 10 post-sample-rate 10 poll-frequency 600
   [Device-GigabitEthernet1/0/1] quit
   ```

3. Configure gRPC:

   # Enable the gRPC service.

   ```
   [Device] grpc enable
   ```

   # Create a sensor group named **test**, and add sensor paths **tcb/tcbpacketinfoevent** and **tcb/tcbrawpacketinfoevent**.

   ```
   [Device] telemetry
   [Device-telemetry] sensor-group test
   [Device-telemetry-sensor-group-test] sensor path tcb/tcbpacketinfoevent
   [Device-telemetry-sensor-group-test] sensor path tcb/tcbrawpacketinfoevent
   [Device-telemetry-sensor-group-test] quit
   ```

   # Create a destination group named **collector1**. Specify a collector that uses IPv4 address 192.168.2.1 and port number 50050.

   ```
   [Device-telemetry] destination-group collector1
   [Device-telemetry-destination-group-collector1] ipv4-address 192.168.2.1 port 50050
   [Device-telemetry-destination-group-collector1] quit
   ```

   # Configure a subscription named **A** to bind sensor group **test** with destination group **collector1**. Set the sampling interval to 10 seconds.

   ```
   [Device-telemetry] subscription A
   [Device-telemetry-subscription-A] sensor-group test sample-interval 10
   [Device-telemetry-subscription-A] destination-group collector1
   [Device-telemetry-subscription-A] quit
   [Device-telemetry] quit
   ```

Verifying the configuration

Verify that the collector receives packet drop information when packet drops occur on a queue enabled with TCB. (Details not shown.)

Contents

Configuring time ranges 1
About time ranges 1
Restrictions and guidelines: Time range configuration 1
Procedure 1
Display and maintenance commands for time ranges 1
Time range configuration examples 2
Example: Configuring a time range 2

Configuring time ranges

About time ranges

You can implement a service based on the time of the day by applying a time range to it. A time-based service takes effect only in time periods specified by the time range. For example, you can implement time-based ACL rules by applying a time range to them.

The following basic types of time ranges are available:

- **Periodic time range**—Recurs periodically on a day or days of the week.
- **Absolute time range**—Represents only a period of time and does not recur.

The active period of a time range is calculated as follows:

1. Combining all periodic statements.
2. Combining all absolute statements.
3. Taking the intersection of the two statement sets as the active period of the time range.

Restrictions and guidelines: Time range configuration

When you configure the ACL hardware mode, follow these restrictions and guidelines:

- If a time range does not exist, the service based on the time range does not take effect.
- You can create a maximum of 1024 time ranges, each with a maximum of 32 periodic statements and 12 absolute statements.

Procedure

1. Enter system view.

   system-view

2. Create or edit a time range.

   time-range *time-range-name* { *start-time* **to** *end-time days* [**from** *time1 date1*] [**to** *time2 date2*] | **from** *time1 date1* [**to** *time2 date2*] | **to** *time2 date2* }

   If an existing time range name is provided, this command adds a statement to the time range.

Display and maintenance commands for time ranges

Execute the **display** command in any view.

Task	Command
Display time range configuration and status.	**display time-range** { *time-range-name* | **all** }

Time range configuration examples

Example: Configuring a time range

Network configuration

As shown in Figure 1, configure an ACL on the device to allow Host A to access the server only during 8:00 and 18:00 on working days from June 2015 to the end of the year.

Figure 1 Network diagram

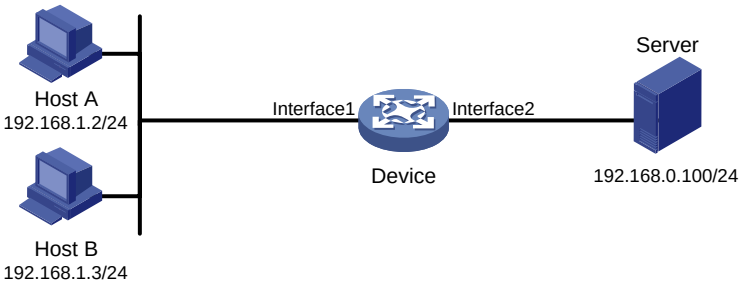


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

Create a periodic time range during 8:00 and 18:00 on working days from June 2015 to the end of the year.

```
<Device> system-view
[Device] time-range work 8:0 to 18:0 working-day from 0:0 6/1/2015 to 24:00 12/31/2015
```

Create an IPv4 basic ACL numbered 2001, and configure a rule in the ACL to permit packets only from 192.168.1.2/32 during the time range **work**.

```
[Device] acl basic 2001
[Device-acl-ipv4-basic-2001] rule permit source 192.168.1.2 0 time-range work
[Device-acl-ipv4-basic-2001] rule deny source any time-range work
[Device-acl-ipv4-basic-2001] quit
```

Apply IPv4 basic ACL 2001 to filter outgoing packets on GigabitEthernet 1/0/2.

```
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] packet-filter 2001 outbound
[Device-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Verify that the time range **work** is active on the device.

```
[Device] display time-range all
Current time is 13:58:35 6/19/2015 Friday

Time-range : work (Active)
 08:00 to 18:00 working-day
 from 00:00 6/1/2015 to 00:00 1/1/2016
```

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Security Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the security fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring AAA 1

About AAA 1
AAA implementation 1
AAA network diagram 1
RADIUS 2
HWTACACS 5
LDAP 7
User management based on ISP domains and user access types 11
Authentication, authorization, and accounting methods 12
AAA extended functions 14
AAA for VPNs 14
RADIUS server feature of the device 15
Protocols and standards 16
FIPS compliance 16
AAA tasks at a glance 16
Configuring local users 17
About local users 17
Local user configuration tasks at a glance 18
Configuring attributes for device management users 18
Configuring attributes for network access users 20
Configuring local guest attributes 21
Configuring user group attributes 23
Managing local guests 23
Configuring the local user auto-delete feature 25
Configuring SNMP notifications for local users 25
Display and maintenance commands for local users and local user groups 26
Configuring RADIUS 26
RADIUS tasks at a glance 26
Restrictions and guidelines for RADIUS configuration 27
Configuring an EAP profile 27
Configuring a test profile for RADIUS server status detection 28
Creating a RADIUS scheme 29
Specifying RADIUS authentication servers 29
Specifying the RADIUS accounting servers 30
Specifying the shared keys for secure RADIUS communication 31
Specifying the MPLS L3VPN instance for a RADIUS scheme 31
Setting the status of RADIUS servers 32
Setting RADIUS timers 33
Specifying the source IP address for outgoing RADIUS packets 34
Setting the username format and traffic statistics units 36
Setting the maximum number of RADIUS request transmission attempts 37
Setting the maximum number of real-time accounting attempts 37
Setting the DSCP priority for RADIUS packets 38
Configuring the format of the RADIUS NAS-Port attribute 38
Configuring the Login-Service attribute check method for SSH, FTP, and terminal users 38
Interpreting the RADIUS class attribute as CAR parameters 39
Configuring the MAC address format for the RADIUS Called-Station-Id attribute 40
Configuring the MAC address format for the RADIUS Calling-Station-Id attribute 40
Configuring the format of RADIUS attribute 87 41
Setting the data measurement unit for the Remanent_Volume attribute 41
Including subattribute 218 of vendor 25506 in outgoing RADIUS packets 42
Configuring the RADIUS attribute translation feature 42
Associating a microsegment with a VSI 44
Configuring RADIUS stop-accounting packet buffering 45
Enabling forcibly sending stop-accounting packets 45
Enabling the RADIUS server load sharing feature 46
Specifying a RADIUS server selection mode for reauthentication 46

Configuring the RADIUS accounting-on feature ····47
Configuring the RADIUS session-control feature ····47
Configuring the RADIUS DAS feature ····48
Configuring the device to preferentially process RADIUS authentication requests ····49
Enabling SNMP notifications for RADIUS ····49
Enabling RADIUS server parallel authentication ····50
Configuring the idle timeout for TLS connections in RADIUS authentication ····51
Disabling the RADIUS service ····51
Display and maintenance commands for RADIUS ····52
Configuring HWTACACS ····53
HWTACACS tasks at a glance ····53
Creating an HWTACACS scheme ····53
Specifying the HWTACACS authentication servers ····53
Specifying the HWTACACS authorization servers ····54
Specifying the HWTACACS accounting servers ····55
Specifying the shared keys for secure HWTACACS communication ····56
Specifying an MPLS L3VPN instance for the scheme ····56
Setting HWTACACS timers ····56
Specifying the source IP address for outgoing HWTACACS packets ····58
Setting the username format and traffic statistics units ····59
Setting the DSCP priority for HWTACACS packets ····59
Configuring HWTACACS stop-accounting packet buffering ····60
Enabling SNMP notifications for HWTACACS ····60
Display and maintenance commands for HWTACACS ····61
Configuring LDAP ····61
LDAP tasks at a glance ····61
Creating an LDAP server ····61
Configuring the IP address of the LDAP server ····62
Specifying the LDAP version ····62
Setting the LDAP server timeout period ····62
Configuring administrator attributes ····63
Configuring LDAP user attributes ····63
Configuring an LDAP attribute map ····64
Creating an LDAP scheme ····65
Specifying the LDAP authentication server ····65
Specifying the LDAP authorization server ····65
Specifying an LDAP attribute map for LDAP authorization ····65
Display and maintenance commands for LDAP ····66
Creating an ISP domain ····66
About ISP domains ····66
Restrictions and guidelines for ISP domain configuration ····66
Creating an ISP domain ····67
Specifying the default ISP domain ····67
Specifying an ISP domain for users that are assigned to nonexistent domains ····67
Configuring ISP domain attributes ····67
Setting ISP domain status ····67
Configuring an authorization profile ····68
Configuring authorization attributes for an ISP domain ····68
Including the idle timeout period in the user online duration to be sent to the server ····70
Enabling temporary redirect ····70
Specifying the user address type in an ISP domain ····71
Configuring the authentication failure policy for users in an ISP domain ····71
Configuring AAA methods for an ISP domain ····71
Configuring authentication methods for an ISP domain ····71
Configuring authorization methods for an ISP domain ····73
Configuring accounting methods for an ISP domain ····75
Display and maintenance commands for ISP domains ····76
Configuring a critical profile ····77
Configuring AAA fail-permit and recovery ····77
Setting the maximum number of concurrent login users ····79
Configuring a NAS-ID ····79
Configuring the device ID ····80

Enabling password change prompt logging ····80
Configuring the RADIUS server feature ····81
RADIUS server feature tasks at a glance ····81
Restrictions and guidelines for the RADIUS server feature ····81
Configuring RADIUS users ····81
Specifying RADIUS clients ····82
Activating the RADIUS server configuration ····82
Display and maintenance commands for RADIUS users and clients ····82
Configuring the connection recording policy ····83
About the connection recording policy ····83
Restrictions and guidelines ····83
Procedure ····83
Display and maintenance commands for the connection recording policy ····83
Displaying user MAC and UUID binding entries ····83
Configuring user online and offline recording ····84
About user online and offline recording ····84
Restrictions and guidelines for user online and offline recording configuration ····84
Enabling user online failure recording ····84
Enabling user offline recording ····84
Enabling real-time recording of AAA user associations and disassociations ····85
Display and maintenance commands for user online and offline recording ····85
Configuring the AAA test feature ····86
Configuring authorization information cache for AAA users ····89
Enabling RADIUS authorization information cache for LAN users ····89
Display and maintenance commands for authorization information cache for AAA users ····89
Enable authorization information cache for AAA users globally ····90
AAA configuration examples ····90
Example: Configuring AAA for SSH users by an HWTACACS server ····90
Example: Configuring local authentication, HWTACACS authorization, and RADIUS accounting for SSH users ····92
Example: Configuring authentication and authorization for SSH users by a RADIUS server ····94
Example: Configuring authentication for SSH users by an LDAP server ····97
Example: Configuring AAA for 802.1X users by a RADIUS server ····100
Example: Configuring and managing a local guest ····109
Example: Configuring authentication and authorization for 802.1X users by the device as a RADIUS server ····111
Troubleshooting AAA ····116
RADIUS authentication failure ····116
RADIUS packet delivery failure ····116
RADIUS accounting error ····117
Troubleshooting HWTACACS ····117
LDAP authentication failure ····117
Appendixes ····118
Appendix A Commonly used RADIUS attributes ····118
Appendix B Descriptions for commonly used standard RADIUS attributes ····119
Appendix C RADIUS subattributes (vendor ID 25506) ····121
Appendix D Format of dynamic authorization ACLs ····124

Configuring AAA

About AAA

AAA implementation

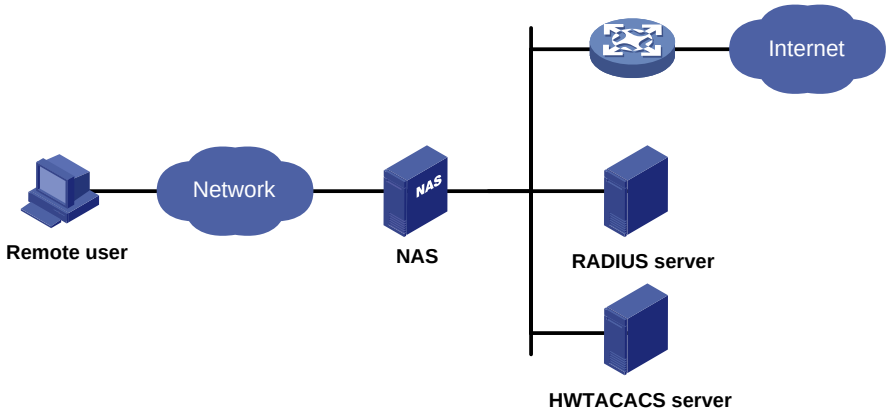
Authentication, Authorization, and Accounting (AAA) provides a uniform framework for implementing network access management. This feature specifies the following security functions:

- **Authentication**—Identifies users and verifies their validity.
- **Authorization**—Grants different users different rights, and controls the users' access to resources and services. For example, you can permit office users to read and print files and prevent guests from accessing files on the device.
- **Accounting**—Records network usage details of users, including the service type, start time, and traffic. This function enables time-based and traffic-based charging and user behavior auditing.

AAA network diagram

AAA uses a client/server model. The client runs on the access device, or the network access server (NAS), which authenticates user identities and controls user access. The server maintains user information centrally. See Figure 1.

Figure 1 AAA network diagram



To access networks or resources beyond the NAS, a user sends its identity information to the NAS. The NAS transparently passes the user information to AAA servers and waits for the authentication, authorization, and accounting result. Based on the result, the NAS determines whether to permit or deny the access request.

AAA has various implementations, including RADIUS, HWTACACS, and LDAP. RADIUS is most often used.

You can use different servers to implement different security functions. For example, you can use an HWTACACS server for authentication and authorization, and use a RADIUS server for accounting.

You can choose the security functions provided by AAA as needed. For example, if your company wants employees to be authenticated before they access specific resources, you would deploy an authentication server. If network usage information is needed, you would also configure an accounting server.

The device performs dynamic password authentication.

RADIUS

Remote Authentication Dial-In User Service (RADIUS) is a distributed information interaction protocol that uses a client/server model. The protocol can protect networks against unauthorized access and is often used in network environments that require both high security and remote user access.

The RADIUS authorization process is combined with the RADIUS authentication process, and user authorization information is piggybacked in authentication responses. RADIUS uses UDP port 1812 for authentication and UDP port 1813 for accounting.

RADIUS was originally designed for dial-in user access, and has been extended to support additional access methods, such as Ethernet and ADSL.

Client/server model

The RADIUS client runs on the NASs located throughout the network. It passes user information to RADIUS servers and acts on the responses to, for example, reject or accept user access requests.

The RADIUS server runs on the computer or workstation at the network center and maintains information related to user authentication and network service access.

The RADIUS server operates using the following process:

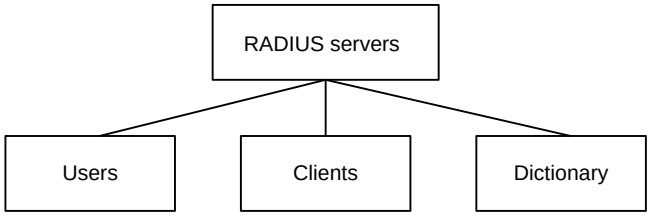
1. Receives authentication, authorization, and accounting requests from RADIUS clients.
2. Performs user authentication, authorization, or accounting.
3. Returns user access control information (for example, rejecting or accepting the user access request) to the clients.

The RADIUS server can also act as the client of another RADIUS server to provide authentication proxy services.

The RADIUS server maintains the following databases:

- **Users**—Stores user information, such as the usernames, passwords, applied protocols, and IP addresses.
- **Clients**—Stores information about RADIUS clients, such as shared keys and IP addresses.
- **Dictionary**—Stores RADIUS protocol attributes and their values.

Figure 2 RADIUS server databases



Information exchange security mechanism

The RADIUS client and server exchange information between them with the help of shared keys, which are preconfigured on the client and server. A RADIUS packet has a 16-byte field called Authenticator. This field includes a signature generated by using the MD5 algorithm, the shared key, and some other information. The receiver of the packet verifies the signature and accepts the packet only when the signature is correct. This mechanism ensures the security of information exchanged between the RADIUS client and server.

The shared keys are also used to encrypt user passwords that are included in RADIUS packets.

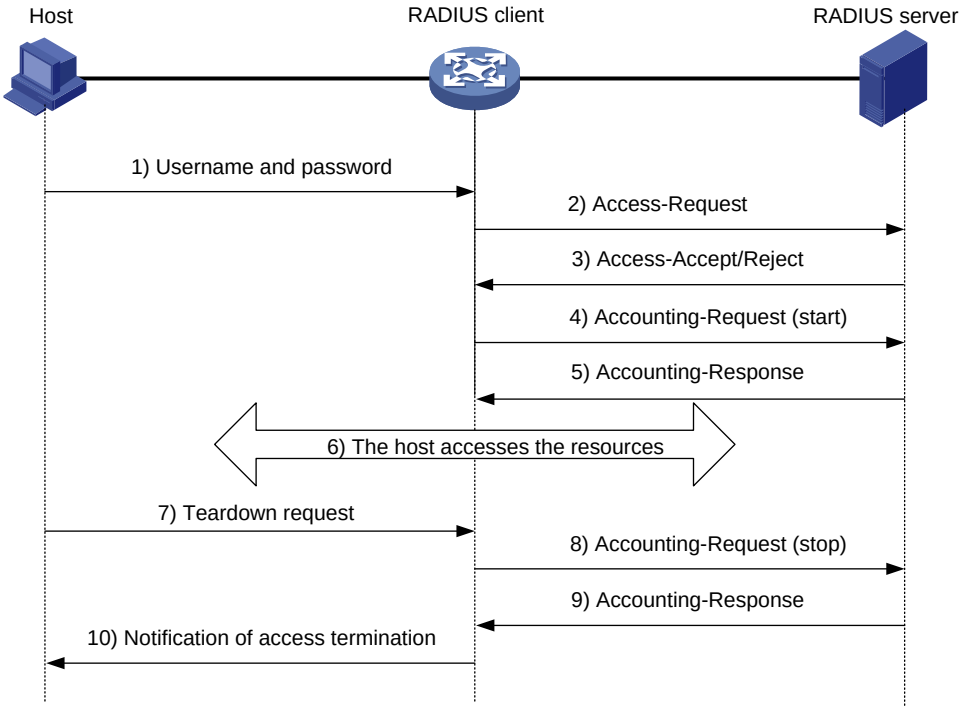
User authentication methods

The RADIUS server supports multiple user authentication methods, such as PAP, CHAP, and EAP.

Basic RADIUS packet exchange process

Figure 3 illustrates the interactions between a user host, the RADIUS client, and the RADIUS server.

Figure 3 Basic RADIUS packet exchange process



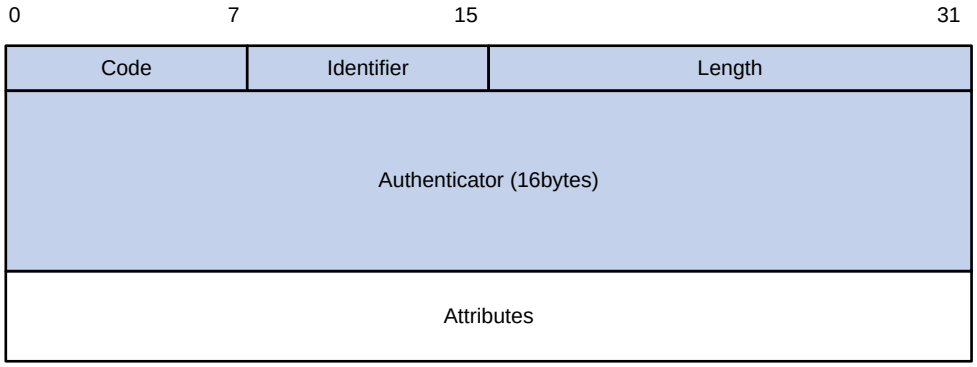
RADIUS uses in the following workflow:

1. The host sends a connection request that includes the user's username and password to the RADIUS client.
2. The RADIUS client sends an authentication request (Access-Request) to the RADIUS server. The request includes the user's password, which has been processed by the MD5 algorithm and shared key.
3. The RADIUS server authenticates the username and password. If the authentication succeeds, the server sends back an Access-Accept packet that contains the user's authorization information. If the authentication fails, the server returns an Access-Reject packet.
4. The RADIUS client permits or denies the user according to the authentication result. If the result permits the user, the RADIUS client sends a start-accounting request (Accounting-Request) packet to the RADIUS server.
5. The RADIUS server returns an acknowledgment (Accounting-Response) packet and starts accounting.
6. The user accesses the network resources.
7. The host requests the RADIUS client to tear down the connection.
8. The RADIUS client sends a stop-accounting request (Accounting-Request) packet to the RADIUS server.
9. The RADIUS server returns an acknowledgment (Accounting-Response) and stops accounting for the user.
10. The RADIUS client notifies the user of the termination.

RADIUS packet format

RADIUS uses UDP to transmit packets. The protocol also uses a series of mechanisms to ensure smooth packet exchange between the RADIUS server and the client. These mechanisms include the timer mechanism, the retransmission mechanism, and the backup server mechanism.

Figure 4 RADIUS packet format



Descriptions of the fields are as follows:

- The Code field (1 byte long) indicates the type of the RADIUS packet. Table 1 gives the main values and their meanings.

  Table 1 Main values of the Code field

Code	Packet type	Description
1	Access-Request	From the client to the server. A packet of this type includes user information for the server to authenticate the user. It must contain the User-Name attribute and can optionally contain the attributes of NAS-IP-Address, User-Password, and NAS-Port.
2	Access-Accept	From the server to the client. If all attribute values included in the Access-Request are acceptable, the authentication succeeds, and the server sends an Access-Accept response.
3	Access-Reject	From the server to the client. If any attribute value included in the Access-Request is unacceptable, the authentication fails, and the server sends an Access-Reject response.
4	Accounting-Request	From the client to the server. A packet of this type includes user information for the server to start or stop accounting for the user. The Acct-Status-Type attribute in the packet indicates whether to start or stop accounting.
5	Accounting-Response	From the server to the client. The server sends a packet of this type to notify the client that it has received the Accounting-Request and has successfully recorded the accounting information.

- The Identifier field (1 byte long) is used to match response packets with request packets and to detect duplicate request packets. The request and response packets of the same exchange process for the same purpose (such as authentication or accounting) have the same identifier.
- The Length field (2 bytes long) indicates the length of the entire packet (in bytes), including the Code, Identifier, Length, Authenticator, and Attributes fields. Bytes beyond this length are considered padding and are ignored by the receiver. If the length of a received packet is less than this length, the packet is dropped.
- The Authenticator field (16 bytes long) is used to authenticate responses from the RADIUS server and to encrypt user passwords. There are two types of authenticators: request authenticator and response authenticator.

- The Attributes field (variable in length) includes authentication, authorization, and accounting information. This field can contain multiple attributes, each with the following subfields:
  - **Type**—Type of the attribute.
  - **Length**—Length of the attribute in bytes, including the Type, Length, and Value subfields.
  - **Value**—Value of the attribute. Its format and content depend on the Type subfield.

Extended RADIUS attributes

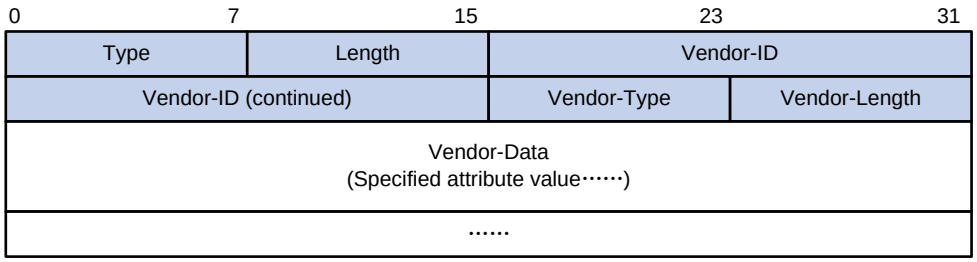
The RADIUS protocol features excellent extensibility. The Vendor-Specific attribute (attribute 26) allows a vendor to define extended attributes. The extended attributes can implement functions that the standard RADIUS protocol does not provide.

A vendor can encapsulate multiple subattributes in the TLV format in attribute 26 to provide extended functions. As shown in Figure 5, a subattribute encapsulated in attribute 26 consists of the following parts:

- **Vendor-ID**—ID of the vendor. The most significant byte is 0. The other three bytes contains a code compliant to RFC 1700.
- **Vendor-Type**—Type of the subattribute.
- **Vendor-Length**—Length of the subattribute.
- **Vendor-Data**—Contents of the subattribute.

The device supports private RADIUS subattributes with a vendor ID of 25506. For more information, see "Appendix C RADIUS subattributes (vendor ID 25506)."

Figure 5 Format of attribute 26



HWTACACS

HW Terminal Access Controller Access Control System (HWTACACS) is an enhanced security protocol based on TACACS (RFC 1492). HWTACACS is similar to RADIUS, and uses a client/server model for information exchange between the NAS and the HWTACACS server.

HWTACACS typically provides AAA services for PPP, VPDN, and terminal users. In a typical HWTACACS scenario, terminal users need to log in to the NAS. Working as the HWTACACS client, the NAS sends users' usernames and passwords to the HWTACACS server for authentication. After passing authentication and obtaining authorized rights, a user logs in to the device and performs operations. The HWTACACS server records the operations that each user performs.

Differences between HWTACACS and RADIUS

HWTACACS and RADIUS have many features in common, such as using a client/server model, using shared keys for data encryption, and providing flexibility and scalability. Table 2 lists the primary differences between HWTACACS and RADIUS.

Table 2 Primary differences between HWTACACS and RADIUS

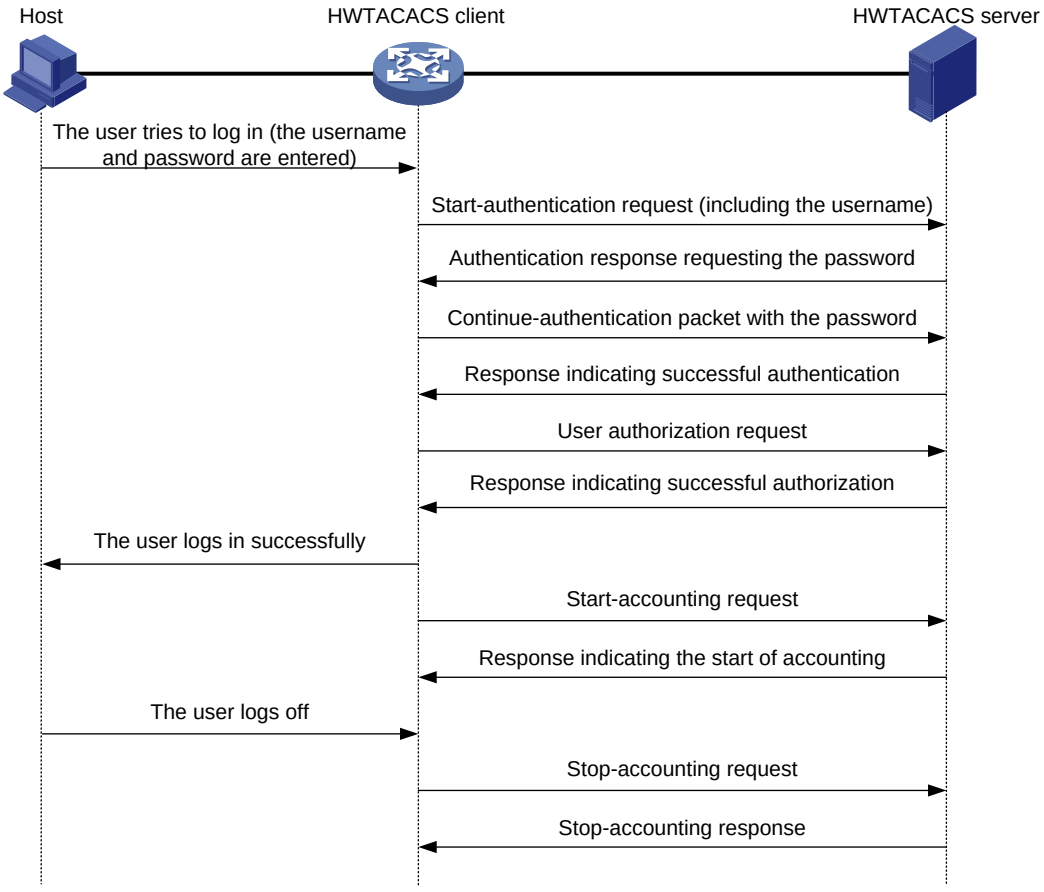
HWTACACS	RADIUS
Uses TCP, which provides reliable network	Uses UDP, which provides high transport efficiency.

HWTACACS	RADIUS
transmission.	
Encrypts the entire packet except for the HWTACACS header.	Encrypts only the user password field in an authentication packet.
Protocol packets are complicated and authorization is independent of authentication. Authentication and authorization can be deployed on different HWTACACS servers.	Protocol packets are simple and the authorization process is combined with the authentication process.
Supports authorization of configuration commands. Access to commands depends on both the user's roles and authorization. A user can use only commands that are permitted by the user roles and authorized by the HWTACACS server.	Does not support authorization of configuration commands. Access to commands solely depends on the user's roles. For more information about user roles, see *Fundamentals Configuration Guide*.

Basic HWTACACS packet exchange process

Figure 6 describes how HWTACACS performs user authentication, authorization, and accounting for a Telnet user.

Figure 6 Basic HWTACACS packet exchange process for a Telnet user



HWTACACS operates using in the following workflow:

1. A Telnet user requests to Telnet to the device and provides the username and password as instructed by the system.
2. The HWTACACS client sends a start-authentication request that includes the username to the HWTACACS server when it receives the Telnet request.

3. The HWTACACS server sends back an authentication response to request the login password.
4. Upon receipt of the response, the HWTACACS client sends the HWTACACS server a continue-authentication packet that includes the login password.
5. If the authentication succeeds, the HWTACACS server sends back an authentication response to indicate that the user has passed authentication.
6. The HWTACACS client sends a user authorization request packet to the HWTACACS server.
7. If the authorization succeeds, the HWTACACS server sends back an authorization response, indicating that the user is now authorized.
8. Knowing that the user is now authorized, the HWTACACS client pushes its CLI to the user and permits the user to log in.
9. The HWTACACS client sends a start-accounting request to the HWTACACS server.
10. The HWTACACS server sends back an accounting response, indicating that it has received the start-accounting request.
11. The user logs off.
12. The HWTACACS client sends a stop-accounting request to the HWTACACS server.
13. The HWTACACS server sends back a stop-accounting response, indicating that the stop-accounting request has been received.

LDAP

The Lightweight Directory Access Protocol (LDAP) provides standard multiplatform directory service. LDAP was developed on the basis of the X.500 protocol. It improves the following functions of X.500:

- Read/write interactive access.
- Browse.
- Search.

LDAP is suitable for storing data that does not often change. The protocol is used to store user information. For example, LDAP server software Active Directory Server is used in Microsoft Windows operating systems. The software stores the user information and user group information for user login authentication and authorization.

LDAP directory service

LDAP uses directories to maintain the organization information, personnel information, and resource information. The directories are organized in a tree structure and include entries. An entry is a set of attributes with distinguished names (DNs). The attributes are used to store information such as usernames, passwords, emails, computer names, and phone numbers.

LDAP uses a client/server model, and all directory information is stored in the LDAP server. Commonly used LDAP server products include Microsoft Active Directory Server, IBM Tivoli Directory Server, and Sun ONE Directory Server.

LDAP authentication and authorization

AAA can use LDAP to provide authentication and authorization services for users. LDAP defines a set of operations to implement its functions. The main operations for authentication and authorization are the bind operation and search operation.

- The bind operation allows an LDAP client to perform the following operations:
  - Establish a connection with the LDAP server.
  - Obtain the access rights to the LDAP server.
  - Check the validity of user information.
- The search operation constructs search conditions and obtains the directory resource information of the LDAP server.

In LDAP authentication, the client completes the following tasks:

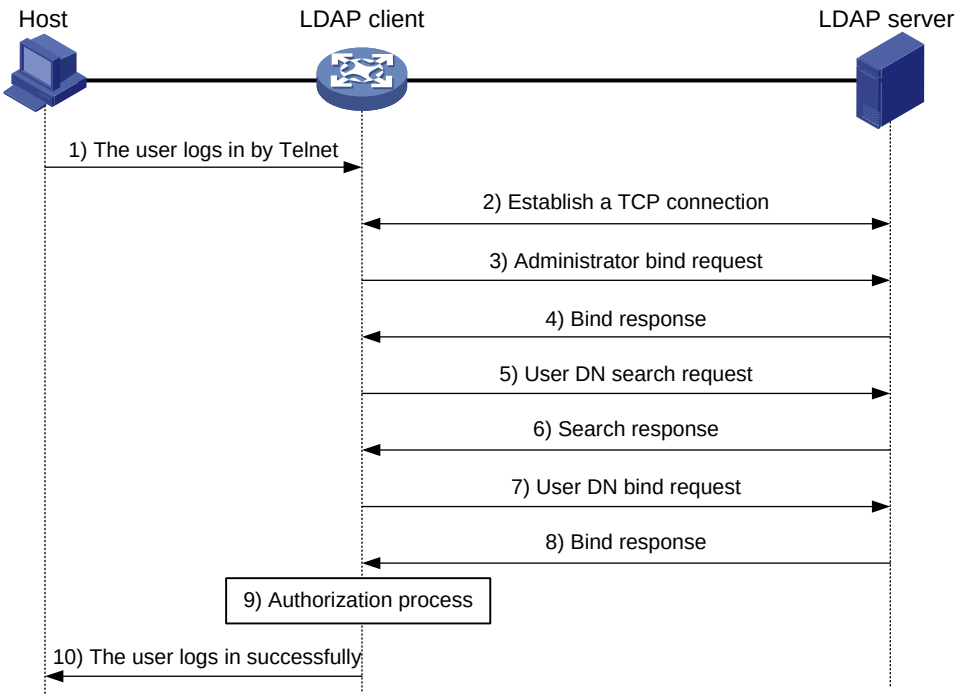
1. Uses the LDAP server administrator DN to bind with the LDAP server. After the binding is created, the client establishes a connection to the server and obtains the right to search.
2. Constructs search conditions by using the username in the authentication information of a user. The specified root directory of the server is searched and a user DN list is generated.
3. Binds with the LDAP server by using each user DN and password. If a binding is created, the user is considered legal.

In LDAP authorization, the client performs the same tasks as in LDAP authentication. When the client constructs search conditions, it obtains both authorization information and the user DN list.

Basic LDAP authentication process

The following example illustrates the basic LDAP authentication process for a Telnet user.

Figure 7 Basic LDAP authentication process for a Telnet user

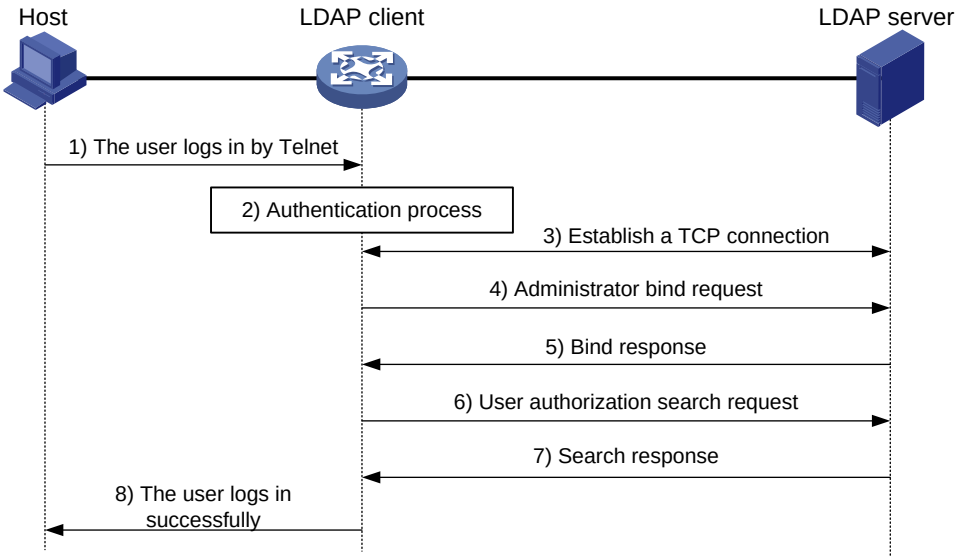


The following shows the basic LDAP authentication process:

1. A Telnet user initiates a connection request and sends the username and password to the LDAP client.
2. After receiving the request, the LDAP client establishes a TCP connection with the LDAP server.
3. To obtain the right to search, the LDAP client uses the administrator DN and password to send an administrator bind request to the LDAP server.
4. The LDAP server processes the request. If the bind operation is successful, the LDAP server sends an acknowledgment to the LDAP client.
5. The LDAP client sends a user DN search request with the username of the Telnet user to the LDAP server.
6. After receiving the request, the LDAP server searches for the user DN by the base DN, search scope, and filtering conditions. If a match is found, the LDAP server sends a response to notify the LDAP client of the successful search. There might be one or more user DNs found.
7. The LDAP client uses the obtained user DN and the entered user password as parameters to send a user DN bind request to the LDAP server. The server will check whether the user password is correct.

8. The LDAP server processes the request, and sends a response to notify the LDAP client of the bind operation result. If the bind operation fails, the LDAP client uses another obtained user DN as the parameter to send a user DN bind request to the LDAP server. This process continues until a DN is bound successfully or all DNs fail to be bound. If all user DNs fail to be bound, the LDAP client notifies the user of the login failure and denies the user's access request.
9. The LDAP client saves the user DN that has been bound and exchanges authorization packets with the authorization server.
   - If LDAP authorization is used, see the authorization process shown in Figure 8.
   - If another method is expected for authorization, the authorization process of that method applies.
10. After successful authorization, the LDAP client notifies the user of the successful login.

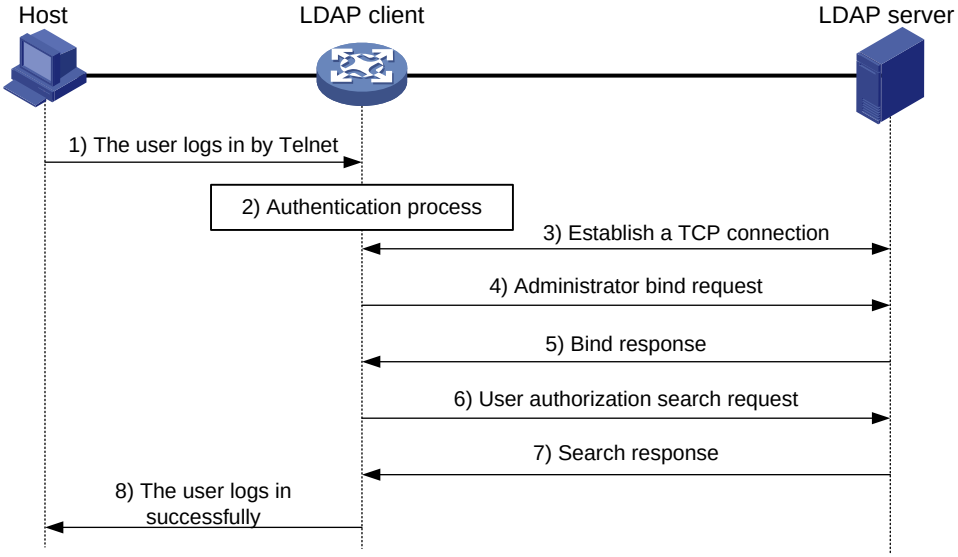
Figure 8 Basic LDAP authorization process for a Telnet user



Basic LDAP authorization process

The following example illustrates the basic LDAP authorization process for a Telnet user.

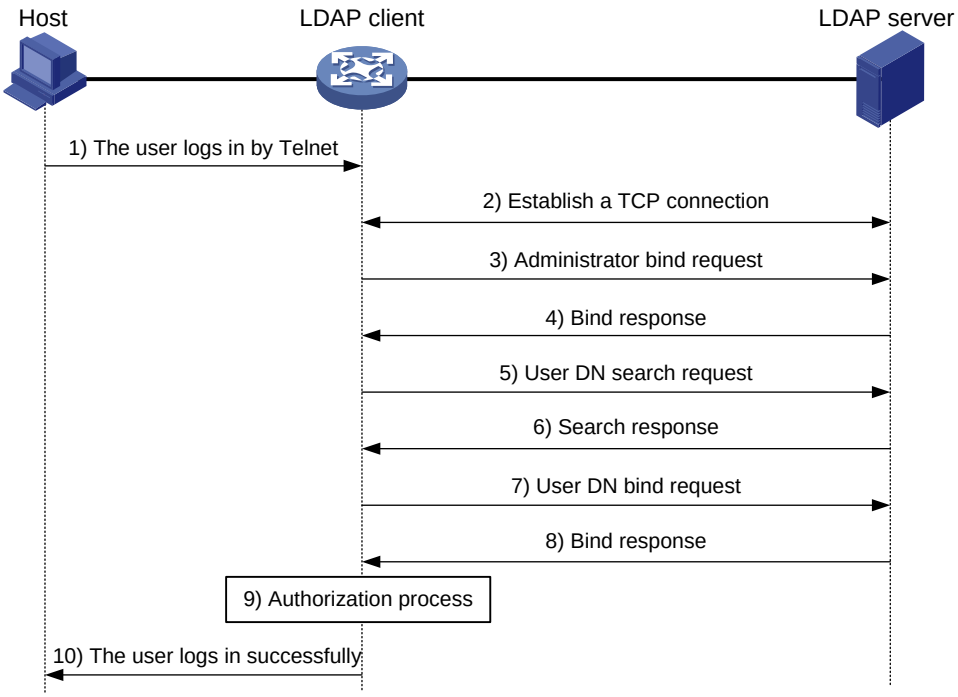
Figure 9 Basic LDAP authorization process for a Telnet user



The following shows the basic LDAP authorization process:

1. A Telnet user initiates a connection request and sends the username and password to the device. The device will act as the LDAP client during authorization.
2. After receiving the request, the device exchanges authentication packets with the authentication server for the user:
   - If LDAP authentication is used, see the authentication process shown in Figure 10.
     - If the device (the LDAP client) uses the same LDAP server for authentication and authorization, skip to step 6.
     - If the device (the LDAP client) uses different LDAP servers for authentication and authorization, skip to step 4.
   - If another authentication method is used, the authentication process of that method applies. The device acts as the LDAP client. Skip to step 3.
3. The LDAP client establishes a TCP connection with the LDAP authorization server.
4. To obtain the right to search, the LDAP client uses the administrator DN and password to send an administrator bind request to the LDAP server.
5. The LDAP server processes the request. If the bind operation is successful, the LDAP server sends an acknowledgment to the LDAP client.
6. The LDAP client sends an authorization search request with the username of the Telnet user to the LDAP server. If the user uses the same LDAP server for authentication and authorization, the client sends the request with the saved user DN of the Telnet user to the LDAP server.
7. After receiving the request, the LDAP server searches for the user information by the base DN, search scope, filtering conditions, and LDAP attributes. If a match is found, the LDAP server sends a response to notify the LDAP client of the successful search.
8. After successful authorization, the LDAP client notifies the user of the successful login.

Figure 10 Basic LDAP authentication process for a Telnet user

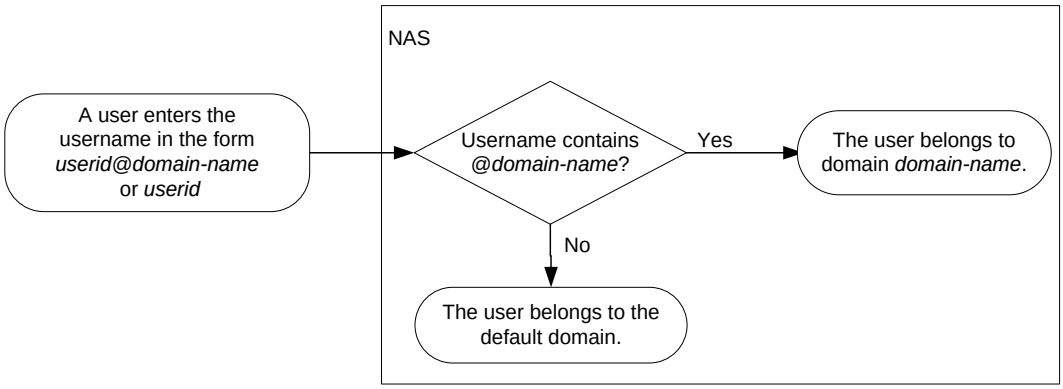


User management based on ISP domains and user access types

AAA manages users based on the users' ISP domains and access types.

On a NAS, each user belongs to one ISP domain. The NAS determines the ISP domain to which a user belongs based on the username entered by the user at login.

Figure 11 Determining the ISP domain for a user by username



AAA manages users in the same ISP domain based on the users' access types. The device supports the following user access types:

- **LAN**—LAN users must pass 802.1X or MAC authentication to come online.
- **Login**—Login users include SSH, Telnet, FTP, and terminal users that log in to the device. Terminal users can access through a console port.
- **Portal**—Portal users must pass portal authentication to access the network.
- **HTTP/HTTPS**—Users log in to the device through HTTP or HTTPS.

The device also provides authentication modules (such as 802.1X) for implementation of user authentication management policies. If you configure these authentication modules, the ISP domains for users of the access types depend on the configuration of the authentication modules.

The device provides the following types of ISP domains for you to control access of wired 802.1, Web authentication, and MAC authentication users to network resources at different phases of authentication:

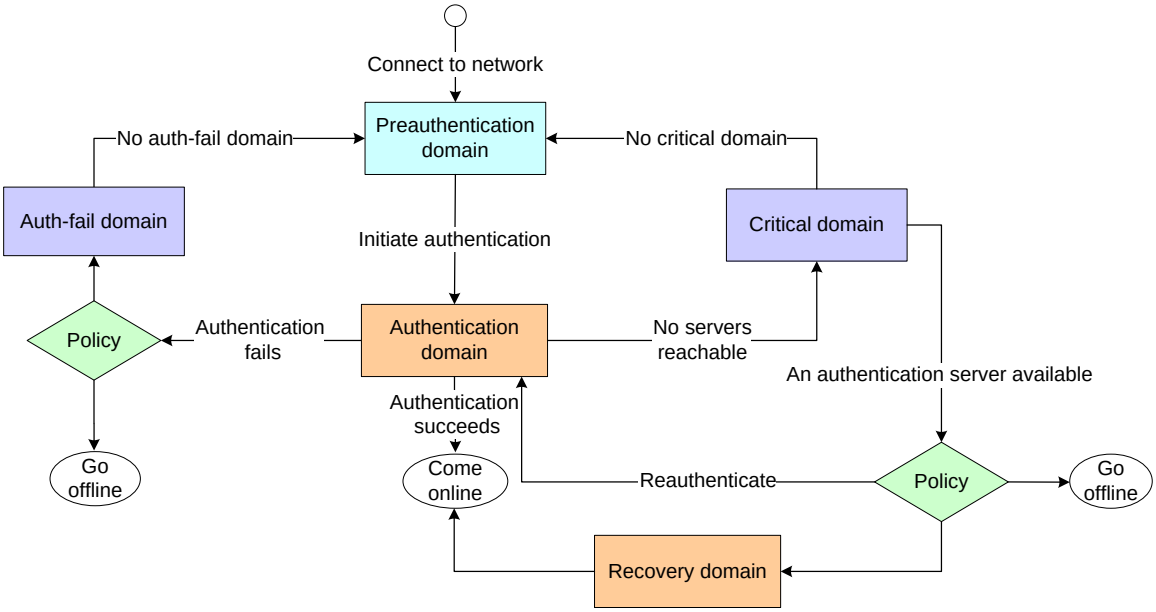
- **Preauthentication domain**—The domain to which users belong before authentication starts for them. For example, users are placed in the preauthentication domain in the following situations:
  - Before they authenticate to the network for the first time.
  - When they fail authentication in an ISP domain that does not have an auth-fail domain.
  - When no authentication servers are available, they are assigned to the specified critical domain. However, the specified critical domain does not exist.

  The preauthentication domain is applicable only to 802.1X and Web authentication users.
- **Authentication domain**—The ISP domain in which users are authenticated and come online.
- **Auth-fail domain**—The ISP domain specified to accommodate users that have failed authentication in an authentication domain.
- **Critical domain**—The ISP domain specified to accommodate users that access an authentication domain when all RADIUS servers in that authentication domain are unavailable.
- **Recovery domain**—The ISP domain to which users in the critical domain are moved after a RADIUS authentication server in their authentication domain becomes available.

IMPORTANT:

In the current version, you must specify the authentication domain as the recovery domain.

Figure 12 Movement of users between different types of ISP domains

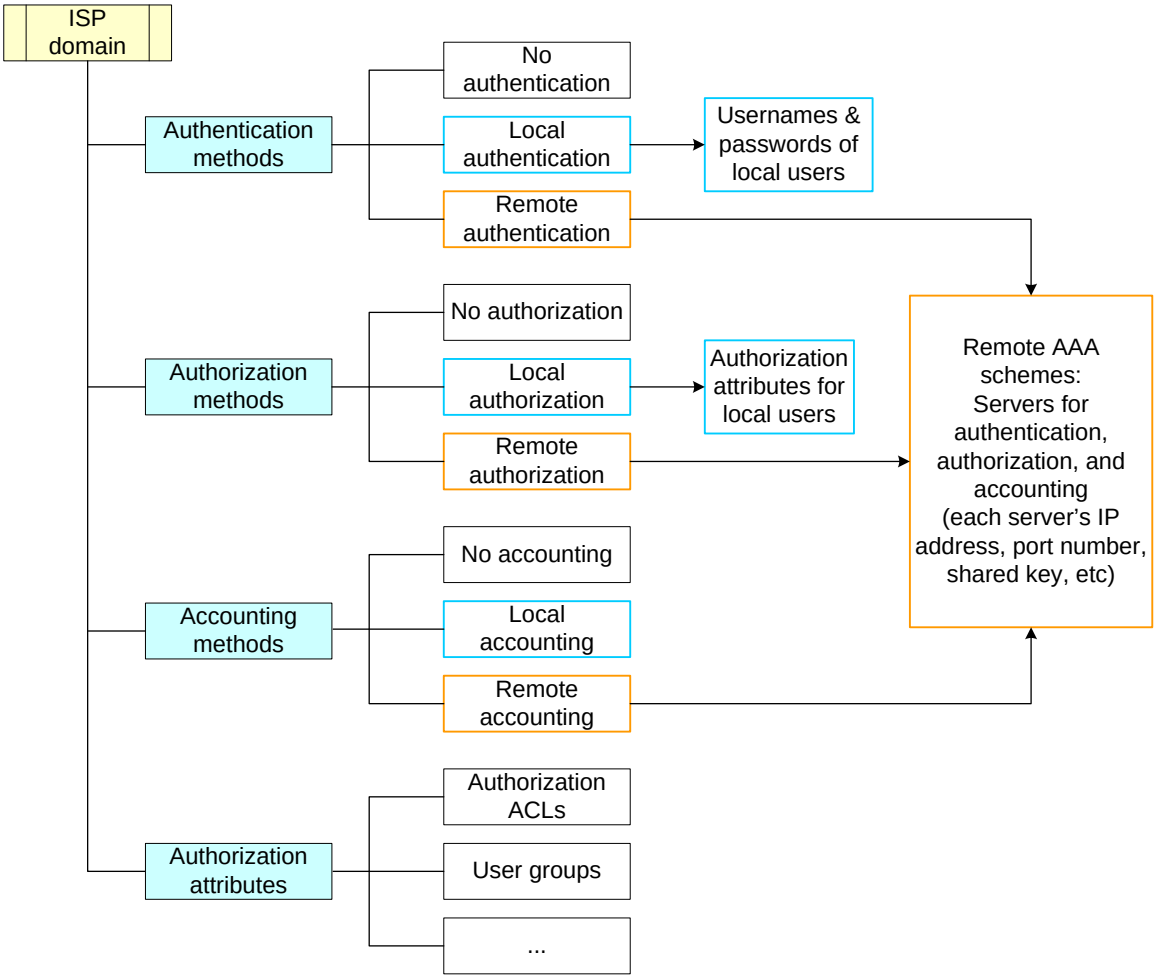


Authentication, authorization, and accounting methods

As shown in Figure 13, an ISP domain represents a set of authentication, authorization, and accounting methods for different types of users in an ISP domain. The NAS determines the ISP domain and access type of a user. The NAS also uses the methods configured for the access type in the domain to control the user's access.

AAA also supports configuring a set of default methods for an ISP domain. These default methods are applied to users for whom no AAA methods are configured.

Figure 13 AAA configuration structure for an ISP domain



Authentication methods

The device supports the following authentication methods:

- **No authentication**—This method trusts all users and does not perform authentication. For security purposes, do not use this method.
- **Local authentication**—The NAS authenticates users by itself, based on the locally configured user information including the usernames, passwords, and attributes. Local authentication allows high speed and low cost, but the amount of information that can be stored is limited by the size of the storage space.
- **Remote authentication**—The NAS works with a remote server to authenticate users. The NAS communicates with the remote server through the RADIUS, HWTACACS, or LDAP protocol. The server manages user information in a centralized manner. Remote authentication provides high capacity, reliable, and centralized authentication services for multiple NASs. You can configure backup methods to be used when the remote server is not available.

Authorization methods

The device supports the following authorization methods:

- **No authorization**—The NAS performs no authorization exchange. The following default authorization information applies after users pass authentication:
  - Login users obtain the level-0 user role. For more information about the level-0 user role, see RBAC configuration in *Fundamentals Configuration Guide*.

    - The working directory for FTP, SFTP, and SCP login users is the root directory of the NAS. However, the users do not have permission to access the root directory.
    - Non-login users can access the network.
- **Local authorization**—The NAS performs authorization according to the user attributes locally configured for users.
- **Remote authorization**—The NAS works with a remote server to authorize users. RADIUS authorization is bound with RADIUS authentication. RADIUS authorization can work only after RADIUS authentication is successful, and the authorization information is included in the Access-Accept packet. HWTACACS or LDAP authorization is separate from authentication, and the authorization information is included in the authorization response after successful authentication. You can configure backup methods to be used when the remote server is not available.

Accounting methods

The device supports the following accounting methods:

- **No accounting**—The NAS does not perform accounting for the users.
- **Local accounting**—Local accounting is implemented on the NAS. It counts and controls the number of concurrent users that use the same local user account, but does not provide statistics for charging.
- **Remote accounting**—The NAS works with a RADIUS server or HWTACACS server for accounting. You can configure backup methods to be used when the remote server is not available.

Authorization attribute configuration methods

You can configure authorization attributes for an ISP domain on the NAS, on a server, or on both the NAS and a server. Server-assigned authorization attributes have higher priority than NAS-assigned settings. NAS-assigned attributes can take effect only if they do not conflict with server-assigned attributes. If a NAS-assigned attribute conflicts with a server-assigned attribute, the server-assigned attribute takes effect.

AAA extended functions

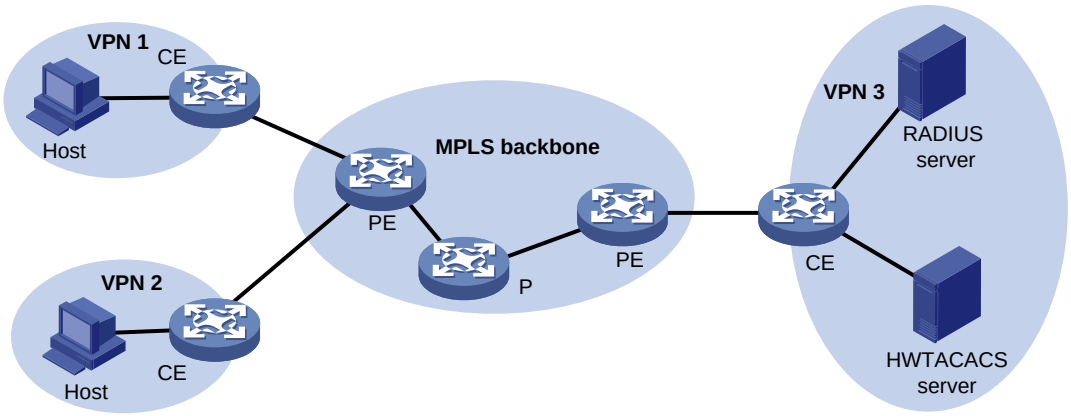
The device provides the following login services to enhance device security:

- **Command authorization**—Enables the NAS to let the authorization server determine whether a command entered by a login user is permitted. Login users can execute only commands permitted by the authorization server. For more information about command authorization, see *Fundamentals Configuration Guide*.
- **Command accounting**—When command authorization is disabled, command accounting enables the accounting server to record all valid commands executed on the device. When command authorization is enabled, command accounting enables the accounting server to record all authorized commands. For more information about command accounting, see *Fundamentals Configuration Guide*.
- **User role authentication**—Authenticates each user that wants to obtain another user role without logging out or getting disconnected. For more information about user role authentication, see *Fundamentals Configuration Guide*.

AAA for VPNs

You can deploy AAA across VPNs to enable forwarding of authentication, authorization, and accounting packets across VPNs. For example, as shown in Figure 14, the PE at the left side of the MPLS backbone acts as a NAS. The NAS transparently delivers the AAA packets of private users in VPN 1 and VPN 2 to the AAA servers in VPN 3 for centralized authentication. Authentication packets of private users in different VPNs do not affect each other.

Figure 14 Network diagram



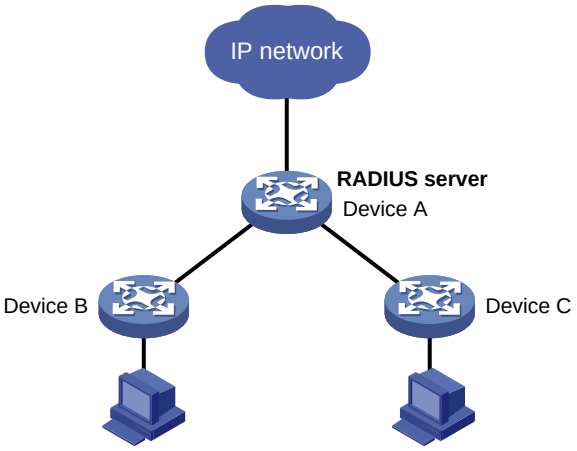
This feature can also help an MCE to implement portal authentication for VPNs. For more information about MCE, see *MCE Configuration Guide*. For more information about portal authentication, see "Configuring portal authentication."

RADIUS server feature of the device

Enable the RADIUS server feature of the device to work with RADIUS clients for user authentication and authorization. The device can act as a dedicated RADIUS server or as both a RADIUS server and a RADIUS client at the same time.

The RADIUS server feature provides for flexible networks with less cost. As shown in Figure 15, Device A provides RADIUS server functions at the distribution layer; Device B and Device C are configured with RADIUS schemes to implement user authentication and authorization at the access layer.

Figure 15 Network diagram



The RADIUS server feature supports the following operations:

- Manages RADIUS user data, which is generated from local user information and includes user name, password, description, authorization ACL, authorization VLAN, and expiration time.
- Manages RADIUS clients. You can add, modify, and delete RADIUS clients. A RADIUS client is identified by the IP address, and it includes attribute information such as the shared key. The RADIUS server feature processes authentication requests only from the managed RADIUS clients and ignores requests from unknown clients.

- Authenticates and authorizes users of the network access type. The server does not provide accounting.

When the RADIUS server receives a RADIUS packet, it performs the following actions:

1. Verifies that the packet is sent from a managed RADIUS client.
2. Verifies the packet with the shared key.
3. Verifies that the user account exists, the password is correct, and other attributes meet the requirements (for example, the account is in the validity period).
4. Determines the authentication result and authorizes specific privileges to the authenticated user.

The RADIUS server feature of the device has the following restrictions:

- The authentication port is fixed at UDP 1812 and cannot be modified.
- The feature is supported on IPv4 networks, but not on IPv6 networks.
- The server provides only PAP and CHAP authentication methods.
- User names sent to the RADIUS server cannot include a domain name.

Protocols and standards

- RFC 2865, *Remote Authentication Dial In User Service (RADIUS)*
- RFC 2866, *RADIUS Accounting*
- RFC 2867, *RADIUS Accounting Modifications for Tunnel Protocol Support*
- RFC 2868, *RADIUS Attributes for Tunnel Protocol Support*
- RFC 2869, *RADIUS Extensions*
- RFC 3576, *Dynamic Authorization Extensions to Remote Authentication Dial In User Service (RADIUS)*
- RFC 4818, *RADIUS Delegated-IPv6-Prefix Attribute*
- RFC 5176, *Dynamic Authorization Extensions to Remote Authentication Dial In User Service (RADIUS)*
- RFC 1492, *An Access Control Protocol, Sometimes Called TACACS*
- RFC 1777, *Lightweight Directory Access Protocol*
- RFC 2251, *Lightweight Directory Access Protocol (v3)*

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see "Configuring FIPS") and non-FIPS mode.

AAA tasks at a glance

To configure AAA, perform the following tasks:

1. Configuring AAA schemes

   If local authentication is used, configure local users and the related attributes. If remote authentication is used, configure the required RADIUS, HWTACACS, or LDAP schemes.

   - Configuring local users
   - Configuring RADIUS

- Configuring HWTACACS
- Configuring LDAP

2. Configuring an ISP domain
   a. Creating an ISP domain
   b. Configuring ISP domain attributes
3. Configuring AAA methods for an ISP domain

   Configure authentication, authorization, and accounting methods for an ISP domain as needed. These methods use existing AAA schemes.
   - Configuring authentication methods for an ISP domain
   - Configuring authorization methods for an ISP domain
   - Configuring accounting methods for an ISP domain
4. (Optional.) Configuring advanced AAA features
   - Configuring AAA fail-permit and recovery
   - Setting the maximum number of concurrent login users
   - Configuring a NAS-ID
   - Configuring the device ID
   - Enabling password change prompt logging
   - Configuring the RADIUS server feature
   - Configuring the connection recording policy
   - Displaying user MAC and UUID binding entries
   - Configuring the AAA test feature
   - Configuring authorization information cache for AAA users

Configuring local users

About local users

To implement local authentication, authorization, and accounting, create local users and configure user attributes on the device. The local users and attributes are stored in the local user database on the device. A local user is uniquely identified by the combination of a username and a user type.

Local users are classified into the following types:

- **Device management user**—User that logs in to the device for device management.
- **Network access user**—User that accesses network resources through the device.

  Network access users also include guests that access the network temporarily. Guests can use only LAN and portal services.

The following shows the configurable local user attributes:

- **Description**—Descriptive information of the user.
- **Service type**—Services that the user can use. Local authentication checks the service types of a local user. If none of the service types is available, the user cannot pass authentication.
- **User state**—Whether or not a local user can request network services. There are two user states: active and blocked. A user in active state can request network services, but a user in blocked state cannot.
- **Upper limit of concurrent logins using the same user name**—Maximum number of users that can concurrently access the device by using the same user name. When the number reaches the upper limit, no more local users can access the device by using the user name.

- **User group**—Each local user belongs to a local user group and has all attributes of the group. The attributes include the password control attributes and authorization attributes. For more information about local user group, see "Configuring user group attributes."
- **Binding attributes**—Binding attributes control the scope of users, and are checked during local authentication of a user. If the attributes of a user do not match the binding attributes configured for the local user account, the user cannot pass authentication.
- **Authorization attributes**—Authorization attributes indicate the user's rights after it passes local authentication.

  Configure the authorization attributes based on the service type of local users.

  You can configure an authorization attribute in user group view or local user view. The setting of an authorization attribute in local user view takes precedence over the attribute setting in user group view.

  The attribute configured in user group view takes effect on all local users in the user group.

  The attribute configured in local user view takes effect only on the local user.
- **Password control attributes**—Password control attributes help control password security for local users. Password control attributes include password aging time, minimum password length, password composition checking, password complexity checking, and login attempt limit.

  You can configure a password control attribute in system view, user group view, or local user view. A password control attribute with a smaller effective range has a higher priority. For more information about password management and global password configuration, see "Configuring password control."
- **Validity period**—Time period in which a network access user is considered valid for authentication.

Local user configuration tasks at a glance

To configure local users, perform the following tasks:

1. Configuring local user attributes
   - Configuring attributes for device management users
   - Configuring attributes for network access users
   - Configuring local guest attributes
2. (Optional.) Configuring user group attributes
3. (Optional.) Managing local guests
4. (Optional.) Configuring the local user auto-delete feature
5. (Optional.) Configuring SNMP notifications for local users

Configuring attributes for device management users

Restrictions and guidelines

When you configure the interface binding attribute for a device management user, follow these restrictions and guidelines to avoid authentication failure:

- Specify the actual access interface of the user as the binding interface for the user.
- Make sure the user's authentication packets include the user's access interface.

If password control is globally enabled for device management users by using the **`password-control enable`** command, the device neither displays local user passwords nor retains them in the running configuration. When you globally disable password control for device management users, local user passwords are automatically restored to the running configuration. To display the running configuration, use the **`display current-configuration`** command.

You can configure authorization attributes and password control attributes in local user view or user group view. The setting in local user view takes precedence over the setting in user group view.

Procedure

1. Enter system view.

   system-view

2. Add a device management user and enter device management user view.

   local-user *user-name* **class manage**

3. Configure a password for the device management user.

   In non-FIPS mode:

   password [{ **hash** | **simple** } *string*]

   A non-password-protected user passes authentication if the user provides the correct username and passes attribute checks. To enhance security, configure a password for each device management user.

   In FIPS mode:

   password

   Only password-protected users can pass authentication. You must set the password in interactive mode for a device management user.

4. Assign services to the device management user.

   In non-FIPS mode:

   service-type { **ftp** | { **http** | **https** | **ssh** | **telnet** | **terminal** } * }

   In FIPS mode:

   service-type { **https** | **ssh** | **terminal** } *

   By default, no services are authorized to a device management user.

5. (Optional.) Set the status of the device management user.

   state { **active** | **block** }

   By default, a device management user is in active state and can request network services.

6. (Optional.) Set the upper limit of concurrent logins using the device management username.

   access-limit *max-user-number*

   By default, the number of concurrent logins is not limited for a device management user.

   This command takes effect only when local accounting is configured for device management users. This command does not apply to FTP, SFTP, or SCP users that do not support accounting.

7. (Optional.) Configure the interface binding attribute for the device management user.

   bind-attribute location interface *interface-type interface-number*

   By default, no interface binding attribute is configured for a device management user.

8. (Optional.) Configure authorization attributes for the device management user.

   authorization-attribute { **idle-cut** *minutes* | **user-role** *role-name* | **work-directory** *directory-name* } *

   The following default settings apply:

   - The working directory for FTP, SFTP, and SCP users is the root directory of the NAS. However, the users do not have permission to access the root directory.
   - The network-operator user role is assigned to the device management local users that are created by a network-admin, system-admin, or level-15 user.

9. (Optional.) Configure password control attributes for the device management user. Choose the following tasks as needed:

   - Set the password aging time.

password-control aging *aging-time*

- Set the minimum password length.
  password-control length *length*
- Configure the password composition policy.
  password-control composition type-number *type-number* [**type-length** *type-length*]
- Configure the password complexity checking policy.
  password-control complexity { **same-character** | **user-name** } **check**
- Configure the maximum login attempts and the action to take if there is a login failure.
  password-control login-attempt *login-times* [**exceed** { **lock** | **lock-time** *time* | **unlock** }]

By default, a device management user uses password control attributes of the user group to which the user belongs.

10. (Optional.) Assign the device management user to a user group.
group *group-name*
By default, a device management user belongs to user group **system**.

Configuring attributes for network access users

Restrictions and guidelines

If password control is globally enabled for network access users by using the **password-control enable network-class** command, the device neither displays local user passwords nor retains them in the running configuration. When you globally disable password control for network access users, local user passwords are automatically restored to the running configuration. To display the running configuration, use the **display current-configuration** command.

You can configure authorization attributes and password control attributes in local user view or user group view. The setting in local user view takes precedence over the setting in user group view.

Configure the **location** binding attribute based on the service types of users.

- For 802.1X users, specify the 802.1X-enabled Layer 2 Ethernet interfaces or Layer 2 aggregate interfaces through which the users access the device.
- For MAC authentication users, specify the MAC authentication-enabled Layer 2 Ethernet interfaces or Layer 2 aggregate interfaces through which the users access the device.
- For Web authentication users, specify the Web authentication-enabled Layer 2 Ethernet interfaces through which the users access the device.
- For portal users, specify the portal-enabled interfaces through which the users access the device. Specify the Layer 2 Ethernet interfaces if portal is enabled on VLAN interfaces and the **portal roaming enable** command is not used.

Procedure

1. Enter system view.
system-view

2. Add a network access user and enter network access user view.
local-user *user-name* **class network**

3. (Optional.) Configure a password for the network access user.
password { **cipher** | **simple** } *string*

4. (Optional.) Configure a description for the network access user.
description *text*

By default, no description is configured for a local user.

5. Assign services to the network access user.

   service-type { **lan-access** | **portal** }

   By default, no services are authorized to a network access user.

6. (Optional.) Set the status of the network access user.

   state { **active** | **block** }

   By default, a network access user is in active state and can request network services.

7. (Optional.) Set the upper limit of concurrent logins using the network access username.

   access-limit *max-user-number*

   By default, the number of concurrent logins is not limited for a network access user.

8. (Optional.) Configure binding attributes for the network access user.

   bind-attribute { **ip** *ip-address* | **location interface** *interface-type interface-number* | **mac** *mac-address* | **vlan** *vlan-id* } *

   By default, no binding attributes are configured for a network access user.

9. (Optional.) Configure authorization attributes for the network access user.

   authorization-attribute { **acl** *acl-number* | **idle-cut** *minutes* | **ip-pool** *ipv4-pool-name* | **ipv6-pool** *ipv6-pool-name* | **session-timeout** *minutes* | **url** *url-string* | **vlan** *vlan-id* } *

   By default, a network access user does not have authorization attributes.

10. (Optional.) Configure password control attributes for the network access user. Choose the following tasks as needed:

    - Set the minimum password length.

      password-control length *length*

    - Configure the password composition policy.

      password-control composition type-number *type-number* [**type-length** *type-length*]

    - Configure the password complexity checking policy.

      password-control complexity { **same-character** | **user-name** } **check**

    By default, a network access user uses password control attributes of the user group to which the user belongs.

11. (Optional.) Assign the network access user to a user group.

    group *group-name*

    By default, a network access user belongs to user group **system**.

12. (Optional.) specify the validity period for the local user.

    validity-datetime { **from** *start-date start-time* **to** *expiration-date expiration-time* | **from** *start-date start-time* | **to** *expiration-date expiration-time* }

    By default, the validity period for a network access user does not expire.

Configuring local guest attributes

About this task

Create local guests and configure guest attributes to control temporary network access behavior. Guests can access the network after passing local authentication. You can configure the recipient addresses and email attribute information to the local guests and guest sponsors.

Procedure

1. Enter system view.
   system-view
2. Create a local guest and enter local guest view.
   local-user *user-name* **class network guest**
3. (Optional.) Configure a password for the local guest.
   password { **cipher** | **simple** } *string*
4. Configure basic information for the local guest. Choose the following tasks as needed:
   - Configure a description for the local guest.
     description *text*
     By default, no description is configured for a local guest.
   - Specify the name of the local guest.
     full-name *name-string*
     By default, no name is specified for a local guest.
   - Specify the company of the local guest.
     company *company-name*
     By default, no company is specified for a local guest.
   - Specify the phone number of the local guest.
     phone *phone-number*
     By default, no phone number is specified for a local guest.
   - Specify the email address of the local guest.
     email *email-string*
     By default, no email address is specified for a local guest.
   - Specify the sponsor name for the local guest.
     sponsor-full-name *name-string*
     By default, no sponsor name is specified for a local guest.
   - Specify the sponsor department for the local guest.
     sponsor-department *department-string*
     By default, no sponsor department is specified for a local guest.
   - Specify the sponsor email address for the local guest.
     sponsor-email *email-string*
     By default, no sponsor email address is specified for a local guest.
5. (Optional.) Configure the validity period for the local guest.
   validity-datetime from *start-date start-time* **to** *expiration-date expiration-time*
   By default, a local guest does not expire.
6. (Optional.) Assign the local guest to a user group.
   group *group-name*
   By default, a local guest belongs to the system-defined user group **system**.
7. (Optional.) Configure the local guest status.
   state { **active** | **block** }
   By default, a local guest is in active state and is allowed to request network services.

Configuring user group attributes

About this task

User groups simplify local user configuration and management. A user group contains a group of local users and has a set of local user attributes. You can configure local user attributes for a user group to implement centralized user attributes management for the local users in the group. Local user attributes that are manageable include authorization attributes.

Procedure

1. Enter system view.

   system-view

2. Create a user group and enter user group view.

   user-group *group-name*

   By default, a system-defined user group exists. The group name is **system**.

3. Configure authorization attributes for the user group.

   authorization-attribute { **acl** *acl-number* | **idle-cut** *minutes* | **ip-pool** *ipv4-pool-name* | **ipv6-pool** *ipv6-pool-name* | **session-timeout** *minutes* | **url** *url-string* | **vlan** *vlan-id* | **work-directory** *directory-name* } *

   By default, no authorization attributes are configured for a user group.

4. (Optional.) Configure password control attributes for the user group. Choose the following tasks as needed:

   - Set the password aging time.

     password-control aging *aging-time*

   - Set the minimum password length.

     password-control length *length*

   - Configure the password composition policy.

     password-control composition type-number *type-number* [**type-length** *type-length*]

   - Configure the password complexity checking policy.

     password-control complexity { **same-character** | **user-name** } **check**

   - Configure the maximum login attempts and the action to take for login failures.

     password-control login-attempt *login-times* [**exceed** { **lock** | **lock-time** *time* | **unlock** }]

   By default, a user group uses the global password control settings. For more information, see "Configuring password control."

Managing local guests

About this task

The local guest management features are for registration, approval, maintenance, and access control of local guests.

The device provides the following local guest management features:

- **Registration and approval**—The device creates local guests after the guest registration information is approved by a guest manager.

  The registration and approval processes are as follows:

  a. The device pushes the portal user registration page to a user that wants to access the network as a local guest.

b. The user submits account information for registration, including the user name, password, and email address.

c. The device forwards the registration request to the guest manager in an email notification.

d. The guest manager adds supplementary information as needed and approves the registration information.

The guest manager must process the registration request before the waiting-approval timeout timer expires. The device automatically deletes expired registration request information.

e. The device creates a local guest account and sends an email notification to the user and guest sponsor. The email contains local guest account, password, validity period, and other account information.

The user can access the network as a local guest.

- **Email notification**—The device notifies the local guests, guest sponsors, or guest managers by email of the guest account information or guest registration requests.
- **Local guest creation in batch**—Create a batch of local guests.
- **Local guest import**—Import guest account information from a .csv file to create local guests on the device based on the imported information.
- **Local guest export**—Export local guest account information to a .csv file. You can import the account information to other devices as needed.

Procedure

1. Enter system view

system-view

2. Configure the email notification feature for local guests.

a. Configure the subject and body of email notifications.

local-guest email format to { **guest** | **manager** | **sponsor** } { **body** *body-string* | **subject** *sub-string* }

By default, no subject or body is configured.

b. Configure the email sender address in the email notifications sent by the device for local guests.

local-guest email sender *email-address*

By default, no email sender address is configured for the email notifications sent by the device.

c. Specify an SMTP server for sending email notifications of local guests.

local-guest email smtp-server *url-string*

By default, no SMTP server is specified.

3. Configure the guest manager's email address.

local-guest manager-email *email-address*

By default, the guest manager's email address is not configured.

4. (Optional.) Set the waiting-approval timeout timer for guest registration requests.

local-guest timer waiting-approval *time-value*

By default, the waiting-approval timeout timer for guest registration requests is 24 hours.

5. (Optional.) Create local guests in batch.

local-guest generate username-prefix *name-prefix* [**password-prefix** *password-prefix*] **suffix** *suffix-number* [**group** *group-name*] **count** *user-count* **validity-datetime** *start-date start-time* **to** *expiration-date expiration-time*

Batch generated local guests share the same name prefix. You can also configure a password prefix to be shared by the guests.

6. (Optional.) Import guest account information from a .csv file in the specified path to create local guests based on the imported information.

   local-user-import class network guest url *url-string* **validity-datetime** *start-date start-time* **to** *expiration-date expiration-time* [**auto-create-group** | **override** | **start-line** *line-number*] *

7. (Optional.) Export local guest account information to a .csv file in the specified path.

   local-user-export class network guest url *url-string*

8. (Optional.) Send email notifications to the local guest or the guest sponsor.

   a. Return to user view.

      quit

   b. Send email notifications to the local guest or the guest sponsor. The email contents include the user name, password, and validity period of the guest account.

      local-guest send-email user-name *user-name* **to** { **guest** | **sponsor** }

Configuring the local user auto-delete feature

About this task

This feature enables the device to examine the validity of local users at fixed time periods of 10 minutes and automatically delete expired local users.

Procedure

1. Enter system view

   system-view

2. Enable the local user auto-delete feature.

   local-user auto-delete enable

   By default, the local user auto-delete feature is disabled.

Configuring SNMP notifications for local users

About this task

With SNMP notifications enabled for password changes of local users, the system generates alarms when one of the following cases occurs:

- The password of a local user is configured or changed from the CLI.
- The password of a local user is changed through NETCONF or MIB.
- The password of a portal user is changed online on the login page.

For RADIUS SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view

   system-view

2. Enable SNMP notifications for password changes of local users.

   snmp-agent trap enable local-server [**manage-password-change** | **network-password-change**] *

By default, all SNMP notifications are disabled for password changes of local users.

Display and maintenance commands for local users and local user groups

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display pending registration requests for local guests.	**display local-guest waiting-approval** [**user-name** *user-name*]
Display the local user configuration and online user statistics.	**display local-user** [**class** { **manage** | **network** [**guest**] } | **idle-cut** { **disable** | **enable** } | **service-type** { **ftp** | **http** | **https** | **lan-access** | **portal** | **ssh** | **telnet** | **terminal** } | **state** { **active** | **block** } | **user-name** *user-name* **class** { **manage** | **network** [**guest**] } | **vlan** *vlan-id*]
Display user group configuration.	**display user-group** { **all** | **name** *group-name* }
Clear pending registration requests for local guests.	**reset local-guest waiting-approval** [**user-name** *user-name*]

Configuring RADIUS

RADIUS tasks at a glance

To configure RADIUS, perform the following tasks:

1. Configuring an EAP profile

   To perform EAP-based RADIUS server status detection, you must configure an EAP profile and specify the EAP profile in a test profile.
2. Configuring a test profile for RADIUS server status detection

   To detect the status of a RADIUS server, you must configure a test profile and configure the RADIUS server to use the test profile in a RADIUS scheme.
3. Creating a RADIUS scheme
4. Specifying RADIUS authentication servers
5. Specifying the RADIUS accounting servers
6. Specifying the shared keys for secure RADIUS communication

   Perform this task if no shared keys are specified when configuring RADIUS authentication or accounting servers.
7. Specifying the MPLS L3VPN instance for a RADIUS scheme

   Perform this task if no MPLS L3VPN instances are specified when configuring RADIUS authentication or accounting servers.
8. (Optional.) Setting the status of RADIUS servers
9. (Optional.) Setting RADIUS timers
10. (Optional.) Configuring parameters for RADIUS packets
    - Specifying the source IP address for outgoing RADIUS packets

- Setting the username format and traffic statistics units
- Setting the maximum number of RADIUS request transmission attempts
- Setting the maximum number of real-time accounting attempts
- Setting the DSCP priority for RADIUS packets

11. (Optional.) Configuring parameters for RADIUS attributes
    - Configuring the format of the RADIUS NAS-Port attribute
    - Configuring the Login-Service attribute check method for SSH, FTP, and terminal users
    - Interpreting the RADIUS class attribute as CAR parameters
    - Configuring the MAC address format for the RADIUS Called-Station-Id attribute
    - Configuring the MAC address format for the RADIUS Calling-Station-Id attribute
    - Configuring the format of RADIUS attribute 87
    - Setting the data measurement unit for the Remanent_Volume attribute
    - Including subattribute 218 of vendor 25506 in outgoing RADIUS packets
    - Configuring the RADIUS attribute translation feature
12. (Optional.) Configuring extended RADIUS features
    - Associating a microsegment with a VSI
    - Configuring RADIUS stop-accounting packet buffering
    - Enabling forcibly sending stop-accounting packets
    - Enabling the RADIUS server load sharing feature
    - Specifying a RADIUS server selection mode for reauthentication
    - Configuring the RADIUS accounting-on feature
    - Configuring the RADIUS session-control feature
    - Configuring the RADIUS DAS feature
    - Configuring the device to preferentially process RADIUS authentication requests
    - Enabling SNMP notifications for RADIUS
    - Disabling the RADIUS service

Restrictions and guidelines for RADIUS configuration

If the authentication server in a RADIUS scheme is provided by the RADIUS server feature on the device, you need to configure only the following items for the RADIUS scheme:

- RADIUS authentication server.
- Shared key for RADIUS communication.
- Username format for interaction with the RADIUS server.

Configuring an EAP profile

About this task

An EAP profile is a collection of EAP authentication settings, including the EAP authentication method and the CA certificate file to be used for some EAP authentication methods.

Restrictions and guidelines

You can specify an EAP profile in multiple test profiles.

You can configure a maximum of 16 EAP profiles.

Prerequisites

Before you specify a CA certificate file, use FTP or TFTP to transfer the CA certificate file to the root directory of the default storage medium on the device.

In an IRF fabric, make sure the CA certificate file exists in the root directory of the default storage medium on the master device before you specify the file.

Procedure

1. Enter system view.

   system-view

2. Create an EAP profile and enter EAP profile view.

   eap-profile *eap-profile-name*

3. Specify the EAP authentication method.

   method { **md5** | **peap-gtc** | **peap-mschapv2** | **ttls-gtc** | **ttls-mschapv2** }

   By default, the EAP authentication method is MD5-challenge.

4. Specify a CA certificate file for EAP authentication.

   ca-file *file-name*

   By default, no CA certificate file is specified for EAP authentication.

   You must specify a CA certificate file to verify the RADIUS server certificate if the EAP authentication method is PEAP-GTC, PEAP-MSCHAPv2, TTLS-GTC, or TTLS-MSCHAPv2.

Configuring a test profile for RADIUS server status detection

About this task

To detect the reachability or availability of a RADIUS authentication server, specify a test profile for the RADIUS server when you specify the server in a RADIUS scheme. With the test profile, the device refreshes the RADIUS server status at each detection interval according to the detection result. If the server is unreachable or unavailable, the device sets the status of the server to blocked. If the server is reachable or available, the device sets the status of the server to active.

The device supports the following RADIUS server status detection methods:

- **Simple detection**—For a RADIUS server, the device simulates an authentication request with the username and password specified in the test profile used by the server. The authentication request is sent to the RADIUS server within each detection interval. The device determines that the RADIUS server is reachable if the device receives a response from the server within the interval.
- **EAP-based detection**—For a RADIUS server, the device simulates an EAP authentication with the username and password specified in the test profile used by the server. The simulated EAP authentication starts at the beginning of each detection interval. If the EAP authentication completes within a detection interval, the device determines that the RADIUS server is available.

Simulating a complete EAP authentication process, EAP-based detection provides more reliable detection results than simple detection. As a best practice, configure EAP-based detection on a network environment where EAP authentication is configured.

Restrictions and guidelines

You can configure multiple test profiles in the system.

The device starts detecting the status of a RADIUS authentication server only if an existing test profile is specified for the server.

If you specify a nonexistent EAP profile in a test profile, the device performs simple detection for the RADIUS servers that use the test profile. After the EAP profile is configured, the device will start EAP-based detection at the next detection interval.

The device stops detecting the status of a RADIUS server when one of the following operations is performed:

- The RADIUS server is removed from the RADIUS scheme.
- The test profile configuration for the RADIUS server is removed in RADIUS scheme view.
- The test profile specified for the RADIUS server is deleted.
- The RADIUS server is manually set to the blocked state.
- The RADIUS scheme that contains the RADIUS server is deleted.

Procedure

1. Enter system view.

   system-view

2. Configure a test profile for detecting the status of RADIUS authentication servers.

   radius-server test-profile *profile-name* **username** *name* [**password** { **cipher** | **simple** } *string*] [**interval** *interval*] [**confirm** | **eap-profile** *eap-profile-name*]

Creating a RADIUS scheme

Restrictions and guidelines

You can configure a maximum of 16 RADIUS schemes. A RADIUS scheme can be used by multiple ISP domains.

Procedure

1. Enter system view.

   system-view

2. Create a RADIUS scheme and enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

Specifying RADIUS authentication servers

About this task

A RADIUS authentication server completes authentication and authorization together, because authorization information is piggybacked in authentication responses sent to RADIUS clients.

You can specify one primary authentication server and a maximum of 16 secondary authentication servers for a RADIUS scheme. Secondary servers provide AAA services when the primary server becomes unreachable. The device searches for an active server in the order the secondary servers are configured.

When RADIUS server load sharing is enabled, the device distributes the workload over all servers without considering the primary and secondary server roles. The device checks the weight value and number of currently served users for each active server, and then determines the most appropriate server in performance to receive an authentication request.

Restrictions and guidelines

If redundancy is not required, specify only the primary server.

A RADIUS authentication server can function as the primary authentication server for one scheme and a secondary authentication server for another scheme at the same time.

Two authentication servers in a scheme, primary or secondary, cannot have the same combination of VPN instance, host name, IP address, and port number.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Specify the primary RADIUS authentication server.

   primary authentication { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **ssl-client-policy** *policy-name* | **test-profile** *profile-name* | **vpn-instance** *vpn-instance-name* | **weight** *weight-value*] *

   By default, no primary RADIUS authentication server is specified.

   The **weight** keyword takes effect only when the RADIUS server load sharing feature is enabled for the RADIUS scheme.

4. (Optional.) Specify a secondary RADIUS authentication server.

   secondary authentication { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **ssl-client-policy** *policy-name* | **test-profile** *profile-name* | **vpn-instance** *vpn-instance-name* | **weight** *weight-value*] *

   By default, no secondary RADIUS authentication servers are specified.

   The **weight** keyword takes effect only when the RADIUS server load sharing feature is enabled for the RADIUS scheme.

Specifying the RADIUS accounting servers

About this task

You can specify one primary accounting server and a maximum of 16 secondary accounting servers for a RADIUS scheme. Secondary servers provide AAA services when the primary server becomes unavailable. The device searches for an active server in the order the secondary servers are configured.

When RADIUS server load sharing is enabled, the device distributes the workload over all servers without considering the primary and secondary server roles. The device checks the weight value and number of currently served users for each active server, and then determines the most appropriate server in performance to receive an accounting request.

Restrictions and guidelines

If redundancy is not required, specify only the primary server.

A RADIUS accounting server can function as the primary accounting server for one scheme and a secondary accounting server for another scheme at the same time.

Two accounting servers in a scheme, primary or secondary, cannot have the same combination of VPN instance, host name, IP address, and port number.

RADIUS does not support accounting for FTP, SFTP, and SCP users.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

radius scheme *radius-scheme-name*

3. Specify the primary RADIUS accounting server.
primary accounting { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **ssl-client-policy** *policy-name* | **vpn-instance** *vpn-instance-name* | **weight** *weight-value*] *

   By default, no primary RADIUS accounting server is specified.

   The **weight** keyword takes effect only when the RADIUS server load sharing feature is enabled for the RADIUS scheme.

4. (Optional.) Specify a secondary RADIUS accounting server.
secondary accounting { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **ssl-client-policy** *policy-name* | **vpn-instance** *vpn-instance-name* | **weight** *weight-value*] *

   By default, no secondary RADIUS accounting servers are specified.

   The **weight** keyword takes effect only when the RADIUS server load sharing feature is enabled for the RADIUS scheme.

Specifying the shared keys for secure RADIUS communication

About this task

The RADIUS client and server use the MD5 algorithm and shared keys to generate the Authenticator value for packet authentication and user password encryption. The client and server must use the same key for each type of communication.

A key configured in this task is for all servers of the same type (accounting or authentication) in the scheme. The key has a lower priority than a key configured individually for a RADIUS server.

Restrictions and guidelines

The shared key configured on the device must be the same as the shared key configured on the RADIUS server.

Procedure

1. Enter system view.
system-view
2. Enter RADIUS scheme view.
radius scheme *radius-scheme-name*
3. Specify a shared key for secure RADIUS communication.
key { **accounting** | **authentication** } { **cipher** | **simple** } *string*

   By default, no shared key is specified for secure RADIUS communication.

Specifying the MPLS L3VPN instance for a RADIUS scheme

About this task

The VPN instance specified for a RADIUS scheme applies to all authentication and accounting servers in that scheme. If a VPN instance is also configured for an individual RADIUS server, the VPN instance specified for the RADIUS scheme does not take effect on that server.

Procedure

1. Enter system view.
system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Specify a VPN instance for the RADIUS scheme.

   vpn-instance *vpn-instance-name*

   By default, a RADIUS scheme belongs to the public network.

Setting the status of RADIUS servers

About this task

To control the RADIUS servers with which the device communicates when the current servers are no longer available, set the status of RADIUS servers to blocked or active. You can specify one primary RADIUS server and multiple secondary RADIUS servers. The secondary servers function as the backup of the primary server. When the RADIUS server load sharing feature is disabled, the device chooses servers based on the following rules:

- When the primary server is in active state, the device first tries to communicate with the primary server. If the primary server is unreachable, the device searches for an active secondary server in the order the servers are configured.
- When one or more servers are in active state, the device tries to communicate with these active servers only, even if the servers are unavailable.
- When all servers are in blocked state, the device only tries to communicate with the primary server.
- If a server is unreachable, the device performs the following operations:
  - Changes the server status to blocked.
  - Starts a quiet timer for the server.
  - Tries to communicate with the next secondary server in active state that has the highest priority.
- When the quiet timer of a server expires or you manually set the server to the active state, the status of the server changes back to active. The device does not check the server again during the authentication or accounting process.
- The search process continues until the device finds an available secondary server or has checked all secondary servers in active state. If no server is reachable, the device considers the authentication or accounting attempt a failure.
- When you remove a server in use, communication with the server times out. The device looks for a server in active state by first checking the primary server, and then checking secondary servers in the order they are configured.
- When a RADIUS server's status changes automatically, the device changes this server's status accordingly in all RADIUS schemes in which this server is specified.
- When a RADIUS server is manually set to blocked, server detection is disabled for the server, regardless of whether a test profile has been specified for the server. When the RADIUS server is set to active state, server detection is enabled for the server on which an existing test profile is specified.

By default, the device sets the status of all RADIUS servers to active. However, in some situations, you must change the status of a server. For example, if a server fails, you can change the status of the server to blocked to avoid communication attempts to the server.

Restrictions and guidelines

The configured server status cannot be saved to any configuration file, and can only be viewed by using the **display radius scheme** command.

After the device restarts, all servers are restored to the active state.

The device selects a reachable server for the authentication or accounting of a new user according to the server selection rules in this section if the RADIUS server load sharing feature is disabled. However, these rules are inapplicable to the reauthentication of online users if the RADIUS server selection mode for reauthentication is set to inherit by using the **reauthentication server-select inherit** command.

Procedure

1. Enter system view.

   system-view
2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*
3. Set the RADIUS server status. Choose the following tasks as needed:
   - Set the status of the primary RADIUS authentication server.

     state primary authentication { **active** | **block** }
   - Set the status of the primary RADIUS accounting server.

     state primary accounting { **active** | **block** }
   - Set the status of a secondary RADIUS authentication server.

     state secondary authentication [{ *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **vpn-instance** *vpn-instance-name*] *] { **active** | **block** }
   - Set the status of a secondary RADIUS accounting server.

     state secondary accounting [{ *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **vpn-instance** *vpn-instance-name*] *] { **active** | **block** }

   By default, a RADIUS server is in active state.

Setting RADIUS timers

About this task

The device uses the following types of timers to control communication with a RADIUS server:

- **Server response timeout timer** (**response-timeout**)—Defines the RADIUS request retransmission interval. The timer starts immediately after a RADIUS request is sent. If the device does not receive a response from the RADIUS server before the timer expires, it resends the request.
- **Server quiet timer** (**quiet**)—Defines the duration to keep an unreachable server in blocked state. If one server is not reachable, the device changes the server status to blocked, starts this timer for the server, and tries to communicate with another server in active state. After the server quiet timer expires, the device changes the status of the server back to active.
- **Real-time accounting timer** (**realtime-accounting**)—Defines the intervals at which the device sends real-time accounting packets to the RADIUS accounting server for online users.

Restrictions and guidelines

Consider the number of secondary servers when you configure the maximum number of RADIUS packet transmission attempts and the RADIUS server response timeout timer. If the RADIUS scheme includes many secondary servers, the retransmission process might be too long and the client connection in the access module, such as Telnet, can time out.

When the client connections have a short timeout period, a large number of secondary servers can cause the initial authentication or accounting attempt to fail. In this case, reconnect the client rather than adjusting the RADIUS packet transmission attempts and server response timeout timer.

Typically, the next attempt will succeed, because the device has blocked the unreachable servers to shorten the time to find a reachable server.

Make sure the server quiet timer is set correctly. A timer that is too short might result in frequent authentication or accounting failures. This is because the device will continue to attempt to communicate with an unreachable server that is in active state. A timer that is too long might temporarily block a reachable server that has recovered from a failure. This is because the server will remain in blocked state until the timer expires.

A short real-time accounting interval helps improve accounting precision but requires many system resources. When there are 1000 or more users, set the interval to 15 minutes or longer.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Set RADIUS timers. Choose the following tasks as needed:
   - Set the RADIUS server response timeout timer.
     timer response-timeout *seconds*
     The default setting is 3 seconds.
   - Set the quiet timer for the servers.
     timer quiet *minutes*
     The default setting is 5 minutes.
   - Set the real-time accounting timer.
     timer realtime-accounting *interval* [**second**]
     The default setting is 12 minutes.

Specifying the source IP address for outgoing RADIUS packets

About this task

The source IP address of RADIUS packets that a NAS sends must match the IP address of the NAS configured on the RADIUS server. A RADIUS server identifies a NAS by its IP address. Upon receiving a RADIUS packet, the RADIUS server checks the source IP address of the packet.

- If it is the IP address of a managed NAS, the server processes the packet.
- If it is not the IP address of a managed NAS, the server drops the packet.

Before sending a RADIUS packet, the NAS selects a source IP address in the following order:

1. The source IP address specified for the RADIUS scheme.
2. The source IP address specified in system view for the VPN or public network, depending on where the RADIUS server resides.
3. The IP address of the outbound interface specified by the route.

In an M-LAG system, you must specify the source IP address for outgoing RADIUS packets on both the M-LAG member devices. Each M-LAG member device selects a source IP address in order as described previously. To make user traffic switchover caused by primary/secondary M-LAG member device switchover transparent to the RADIUS server, the source IP address of outgoing RADIUS packets must remain unchanged after switchover. You must specify the source IP address depending on the load sharing mode for user authentication on M-LAG interfaces:

- In centralized mode, you must specify the same source IP address on both the M-LAG member devices.

  When an M-LAG member device fails, the active M-LAG member device uses the source IP address for outgoing RADIUS packets sent for users authenticated on the failed M-LAG member device.
- In distributed mode, use either of the following methods to specify source IP addresses for outgoing RADIUS packets:
  - Specify different source addresses on the M-LAG member devices. The M-LAG member devices will automatically advertise their source addresses to each other.
  - Specify a local and peer source IP address pair on both of the M-LAG member devices. Make sure the peer source IP address on one M-LAG member device is the local IP address on the other M-LAG member device.

  When one M-LAG member device fails, the other member device uses the automatically obtained or manually configured peer source IP address to send RADIUS packets for users authenticated on the failed member device.

Restrictions and guidelines for source IP address configuration

You can specify a source IP address for outgoing RADIUS packets in RADIUS scheme view or in system view.

- The IP address specified in RADIUS scheme view applies only to one RADIUS scheme.
- The IP address specified in system view applies to all RADIUS schemes.

The source IP address of RADIUS packets that a NAS sends must match the IP address of the NAS that is configured on the RADIUS server.

As a best practice, specify a loopback interface address as the source IP address for outgoing RADIUS packets to avoid RADIUS packet loss caused by physical port errors.

The source address of outgoing RADIUS packets is typically the IP address of an egress interface on the NAS to communicate with the RADIUS server. However, in some situations, you must change the source IP address. For example, when VRRP is configured for stateful failover, configure the virtual IP of the uplink VRRP group as the source address.

You can directly specify a source IP address for outgoing RADIUS packets or specify a source interface to provide the source IP address for outgoing RADIUS packets. The source interface configuration and the source IP address configuration overwrite each other.

In an M-LAG system, you must use the **nas-ip** command to specify a virtual IP address for the M-LAG system as the source IP address of RADIUS packets. For more information about virtual IP addresses for the M-LAG system, see M-LAG configuration in *Layer 2—LAN Switching Configuration Guide*.

Specifying a source interface or source IP address for all RADIUS schemes

1. Enter system view.

   system-view
2. Specify a source interface or source IP address for outgoing RADIUS packets.

   radius nas-ip { **interface** *interface-type interface-number* | { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] }

   By default, the source IP address of an outgoing RADIUS packet is the primary IPv4 address or the IPv6 address of the outbound interface.

Specifying a source interface or source IP address for a RADIUS scheme

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Specify a source interface or source IP address for outgoing RADIUS packets.

   nas-ip [**m-lag** { **local** | **peer** }] { *ipv4-address* | **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }

   By default, the source IP address of an outgoing RADIUS packet is that specified by using the **radius nas-ip** command in system view. If the **radius nas-ip** command is not used, the source IP address is the primary IP address of the outbound interface.

   The **m-lag** { **local** | **peer** } parameters are applicable only to an M-LAG system. Specify a source IP address for outgoing RADIUS packets on each M-LAG member device as follows, depending on the load sharing mode for user authentication on M-LAG interfaces:

   - In centralized mode, execute the **nas-ip** command with or without the **m-lag local parameters to specify the local source IP address. The two methods have the same result except that you must use the same method on both devices.**
   - In distributed mode, execute the **nas-ip** command with the **m-lag local parameters and with the m-lag peer parameters to specify the local and peer source IP addresses, respectively. Alternatively,** execute the **nas-ip** command with or without the **m-lag local parameters to specify the local source IP address. These methods have the same result, except that you must use the same method on both devices**.

   If you do not specify the keywords, the system does not send accounting-on packets when the device restarts.

Setting the username format and traffic statistics units

About this task

A username is in the *userid@isp-name* format, where the *isp-name* part represents the user's ISP domain name. By default, the ISP domain name is included in a username. However, older RADIUS servers might not recognize usernames that contain the ISP domain names. In this case, you can configure the device to remove the domain name of each username to be sent.

The device reports online user traffic statistics in accounting packets. The traffic measurement units are configurable.

Restrictions and guidelines

If two or more ISP domains use the same RADIUS scheme, configure the RADIUS scheme to keep the ISP domain name in usernames for domain identification.

For accounting accuracy, make sure the traffic statistics units configured on the device and on the RADIUS accounting servers are the same.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Set the format for usernames sent to the RADIUS servers.

   user-name-format { **keep-original** | **with-domain** | **without-domain** }

   By default, the ISP domain name is included in a username.

   If the device is specified as the RADIUS server in the scheme, the username format must be set to **without-domain**.

4. Set the data flow and packet measurement units for traffic statistics.

   data-flow-format { **data** { **byte** | **giga-byte** | **kilo-byte** | **mega-byte** } | **packet** { **giga-packet** | **kilo-packet** | **mega-packet** | **one-packet** } }*

   By default, traffic is counted in bytes and packets.

Setting the maximum number of RADIUS request transmission attempts

About this task

RADIUS uses UDP packets to transfer data. Because UDP communication is not reliable, RADIUS uses a retransmission mechanism to improve reliability. A RADIUS request is retransmitted if the NAS does not receive a server response for the request within the response timeout timer. For more information about the RADIUS server response timeout timer, see "Setting the status of RADIUS servers."

You can set the maximum number for the NAS to retransmit a RADIUS request to the same server. When the maximum number is reached, the NAS tries to communicate with other RADIUS servers in active state. If no other servers are in active state at the time, the NAS considers the authentication or accounting attempt a failure.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Set the maximum number of RADIUS request transmission attempts.

   retry *retries*

   By default, the maximum number is 3 for RADIUS request transmission attempts.

Setting the maximum number of real-time accounting attempts

About this task

If you set the maximum number of real-time accounting attempts, the device will disconnect users from whom no accounting responses are received within the permitted attempts.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Set the maximum number of real-time accounting attempts.

   retry realtime-accounting *retries*

   By default, the maximum number is 5 for real-time accounting attempts.

Setting the DSCP priority for RADIUS packets

About this task

DSCP priority determines the transmission priority of RADIUS packets. A larger value represents a higher priority. DSCP priority is contained in the ToS field of the IPv4 header and in the Traffic Class field of the IPv6 header.

Procedure

1. Enter system view.
   system-view
2. Set the DSCP priority for RADIUS packets.
   radius [**ipv6**] **dscp** *dscp-value*
   By default, the DSCP priority is 0 for RADIUS packets.

Configuring the format of the RADIUS NAS-Port attribute

About this task

The device supports the following formats for the RADIUS NAS-Port attribute (attribute 5):

- **Default format**—Contains the following portions:
  - 8-bit IRF member ID.
  - 4-bit slot number.
  - 8-bit interface index.
  - 12-bit VLAN ID.
- **Port format**—Contains the last segment for the interface number of the interface through which a user accesses the device. For example, if a user accesses the device through GigabitEthernet 1/0/2, 2 is used as the value for the RADIUS NAS-Port attribute.

Restrictions and guidelines

RADIUS servers of different types might have different requirements for the format of the NAS-Port attribute. To ensure correct RADIUS packet exchange, make sure the format of the NAS-Port attribute meets the requirements of the RADIUS servers.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Configure the format of the RADIUS NAS-Port attribute as the port format.
   attribute 5 format port
   By default, the RADIUS NAS-Port attribute uses the default format.

Configuring the Login-Service attribute check method for SSH, FTP, and terminal users

About this task

The Login-Service attribute (RADIUS attribute 15) identifies the service that can be used to connect a user to the device. For example, a value of 0 represents the Telnet service. This attribute is

contained in the Access-Accept message sent from the RADIUS server to the device. To log in to the device, the access service type of a RADIUS authentication user must match the Login-Service attribute value in the Access-Accept message.

The RADIUS standards do not assign a Login-Service attribute value for the SSH, FTP, HTTP, HTTPS, or terminal users. To perform service type consistency checks for these types of login users, H3C defines the extended Login-Service attribute values in Table 3.

Table 3 Extended Login-Service attribute values

Attribute value	Service type
50	SSH
51	**FTP**
52	**Terminal**
53	**HTTP**
54	**HTTPS**

The device provides the following Login-Service attribute check methods:

- **Strict**—Checks the service types of login users based on the standard and extended attribute values. SSH, FTP, HTTP, HTTPS, and terminal users can pass the check only if the Access-Accept messages for them contain the extended Login-Service attribute values specified for them.
- **Loose**—Checks the service types of login users based only on the standard attribute values. SSH, FTP, HTTP, HTTPS, and terminal users pass the check if the attribute value is 0.

Restrictions and guidelines

Use the loose check method if the server does not issue the extended Login-Service attribute values for SSH, FTP, HTTP, HTTPS, and terminal users.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Configure the Login-Service attribute check method for SSH, FTP, and terminal users.
   attribute 15 check-mode { **loose** | **strict** }
   The default check method is strict.

Interpreting the RADIUS class attribute as CAR parameters

About this task

A RADIUS server may deliver CAR parameters for user-based traffic monitoring and control by using the RADIUS class attribute (attribute 25) in RADIUS packets. You can configure the device to interpret the class attribute to CAR parameters.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*

3. Interpret the RADIUS class attribute as CAR parameters.

   attribute 25 car

   By default, the RADIUS class attribute is not interpreted as CAR parameters.

Configuring the MAC address format for the RADIUS Called-Station-Id attribute

Restrictions and guidelines

RADIUS servers of different types might have different requirements for the MAC address format in the RADIUS Called-Station-Id attribute (attribute 30). Configure the MAC address format for this attribute to meet the requirements of the RADIUS servers.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Configure the MAC address format for the RADIUS Called-Station-Id attribute.

   attribute 30 mac-format section { **one** | { **six** | **three** } **separator** *separator-character* } { **lowercase** | **uppercase** }

   By default, the MAC address in the RADIUS Called-Station-Id attribute is in the format of HH-HH-HH-HH-HH-HH. The MAC address is separated by hyphen (-) into six sections with letters in upper case.

Configuring the MAC address format for the RADIUS Calling-Station-Id attribute

Restrictions and guidelines

RADIUS servers of different types might have different requirements for the MAC address format in the RADIUS Calling-Station-Id attribute (attribute 31). Configure the MAC address format for this attribute to meet the requirements of the RADIUS servers.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Configure the MAC address format for the RADIUS Calling-Station-Id attribute.

   attribute 31 mac-format section { **one** | { **six** | **three** } **separator** *separator-character* } { **lowercase** | **uppercase** }

   By default, the MAC address in the RADIUS Calling-Station-Id attribute is in the format of HH-HH-HH-HH-HH-HH. The MAC address is separated by hyphen (-) into six sections with letters in upper case.

Configuring the format of RADIUS attribute 87

About this task

RADIUS attribute 87 is the NAS-Port-Id attribute. This attribute has the following format types:

- **Default format**—The default format varies by user access type.
  - For portal users, the NAS-Port-Id attribute in this format contains the following portions:
    - 2-bit slot number.
    - 2-bit 0s.
    - 3-bit interface index.
    - 9-bit VLAN ID.
  - For 802.1X and MAC authentication users, the NAS-Port-Id attribute is in the format of **slot=***xx***;subslot=***xx***;port=***xx***;vlanid=***xx*.
    - **slot**—IRF member ID.
    - **subslot**—Slot number.
    - **port**—Interface index.
    - **vlanid**—VLAN ID.
  - For login users, the NAS-Port-Id attribute is not included in RADIUS packets.
- **Interface name format**—Contains the name of the interface through which a user accesses the device. For example, if a user accesses the device through GigabitEthernet 1/0/1, GigabitEthernet 1/0/1 is used as the value for the RADIUS NAS-Port-Id attribute.

Restrictions and guidelines

RADIUS servers of different types might have different requirements for the format of the NAS-Port-Id attribute. To ensure correct RADIUS packet exchange, configure the format of the NAS-Port-Id attribute to meet the requirements of the RADIUS servers.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Configure the format of RADIUS attribute 87 as the interface name format.
   attribute 87 format interface-name
   By default, the default format is used for RADIUS attribute 87.

Setting the data measurement unit for the Remanent_Volume attribute

About this task

The RADIUS server uses the Remanent_Volume attribute in authentication or real-time accounting responses to notify the device of the current amount of data available for online users.

Restrictions and guidelines

Make sure the configured measurement unit is the same as the user data measurement unit on the RADIUS server.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Set the data measurement unit for the Remanent_Volume attribute.

   attribute remanent-volume unit { **byte** | **giga-byte** | **kilo-byte** | **mega-byte** }

   By default, the data measurement unit is kilobyte.

Including subattribute 218 of vendor 25506 in outgoing RADIUS packets

About this task

The RADIUS Vendor-Specific attribute (attribute 26) allows vendors to define extended attributes to implement functions that the standard RADIUS protocol does not provide. Vendor 25506 defines subattribute 218 to carry user DHCP option information.

To send user DHCP option information to RADIUS servers, perform this task to include subattribute 218 of vendor 25506 in outgoing RADIUS start-accounting and update-accounting requests.

In the current software version, only DHCP Option 55 and DHCP Option 61 can be carried in the subattribute.

Restrictions and guidelines

You can repeat the **include-attribute 218 vendor-id 25506** command to configure the device to encapsulate both DHCP Option 55 and DHCP Option 61 in the subattribute. The length of each option is limited to 246 bytes.

If you repeat this command multiple times with the same DHCP option specified, the most recent configuration takes effect.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Include subattribute 218 of vendor 25506 in outgoing RADIUS packets.

   include-attribute 218 vendor-id 25506 dhcp-option { **55** | **61** } * { **format1** | **format2** }

   By default, the device uses format 1 to encapsulate DHCP Option 61 in subattribute 218 of vendor 25506 in RADIUS packets.

Configuring the RADIUS attribute translation feature

About this task

The RADIUS attribute translation feature enables the device to work correctly with the RADIUS servers of different vendors that support RADIUS attributes incompatible with the device.

RADIUS attribute translation has the following implementations:

- **Attribute conversion**—Converts source RADIUS attributes into destination RADIUS attributes based on RADIUS attribute conversion rules.
- **Attribute rejection**—Rejects RADIUS attributes based on RADIUS attribute rejection rules.

When the RADIUS attribute translation feature is enabled, the device processes RADIUS packets as follows:

- For the sent RADIUS packets:
  - Deletes the rejected attributes from the packets.
  - Uses the destination RADIUS attributes to replace the attributes that match RADIUS attribute conversion rules in the packets.
- For the received RADIUS packets:
  - Ignores the rejected attributes in the packets.
  - Interprets the attributes that match RADIUS attribute conversion rules as the destination RADIUS attributes.

To identify proprietary RADIUS attributes, you can define the attributes as extended RADIUS attributes, and then convert the extended RADIUS attributes to device-supported attributes.

Restrictions and guidelines for RADIUS attribute translation configuration

Configure either conversion rules or rejection rules for a RADIUS attribute.

Configure either direction-based rules or packet type-based rules for a RADIUS attribute.

For direction-based translation of a RADIUS attribute, you can configure a rule for each direction (inbound or outbound). For packet type-based translation of a RADIUS attribute, you can configure a rule for each RADIUS packet type (RADIUS Access-Accept, RADIUS Access-Request, or RADIUS accounting).

Configuring the RADIUS attribute translation feature for a RADIUS scheme

1. Enter system view.

   system-view

2. (Optional.) Define an extended RADIUS attribute.

   radius attribute extended *attribute-name* [**vendor** *vendor-id*] **code** *attribute-code* **type** { **binary** | **date** | **integer** | **interface-id** | **ip** | **ipv6** | **ipv6-prefix** | **octets** | **string** }

3. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

4. Enable the RADIUS attribute translation feature.

   attribute translate

   By default, this feature is disabled.

5. Configure a RADIUS attribute conversion rule or a RADIUS attribute reject rule. Choose the following tasks as needed:
   - Configure a RADIUS attribute conversion rule.

     attribute convert *src-attr-name* **to** *dest-attr-name* { { **access-accept** | **access-request** | **accounting** } * | { **received** | **sent** } * }

     By default, no RADIUS attribute conversion rules are configured.
   - Configure a RADIUS attribute rejection rule.

     attribute reject *attr-name* { { **access-accept** | **access-request** | **accounting** } * | { **received** | **sent** } * }

     By default, no RADIUS attribute rejection rules are configured.

Configuring the RADIUS attribute translation feature for a RADIUS DAS

1. Enter system view.
   system-view
2. (Optional.) Define an extended RADIUS attribute.
   radius attribute extended *attribute-name* [**vendor** *vendor-id*] **code** *attribute-code* **type** { **binary** | **date** | **integer** | **interface-id** | **ip** | **ipv6** | **ipv6-prefix** | **octets** | **string** }
3. Enter RADIUS DAS view.
   radius dynamic-author server
4. Enable the RADIUS attribute translation feature.
   attribute translate
   By default, this feature is disabled.
5. Configure a RADIUS attribute conversion rule or a RADIUS attribute rejection rule. Choose the following tasks as needed:
   - Configure a RADIUS attribute conversion rule.
     attribute convert *src-attr-name* **to** *dest-attr-name* { { **coa-ack** | **coa-request** } * | { **received** | **sent** } * }
     By default, no RADIUS attribute conversion rules are configured.
   - Configure a RADIUS attribute rejection rule.
     attribute reject *attr-name* { { **coa-ack** | **coa-request** } * | { **received** | **sent** } * }
     By default, no RADIUS attribute rejection rules are configured.

Associating a microsegment with a VSI

About this task

Use this feature when microsegment-based access control is deployed on a VXLAN network.

When the RADIUS server assigns the microsegment and VSI attributes to a user, the device directly assigns the microsegment and VSI to the user so the user can access the related VXLAN resources.

When the RADIUS server assigns only the microsegment attribute but no VSI attribute to a user, the device will search for a VSI associated with the microsegment.

- If the device finds an associated VSI, it assigns the microsegment and the VSI to the user.
- If the device does not find an associated VSI, it assigns only the microsegment to the user.

Restrictions and guidelines

In a RADIUS scheme, a microsegment can be associated with only one VSI. Multiple microsegments can be associated with the same VSI. If you repeat the **microsegment associate** command to associate a microsegment with different VSIs, the most recent configuration takes effect.

Make sure the device can interpret microsegment IDs assigned by the RADIUS server. If the device cannot directly interpret microsegment IDs, use the attribute translation feature so the device can translate the server-assigned microsegment ID attribute to the H3C-Microsegment-Id attribute.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*

3. Associate a microsegment with a VSI.

   microsegment *microsegment-id* **associate vsi** *vsi-name*

   By default, no VSI is associated with a microsegment.

Configuring RADIUS stop-accounting packet buffering

About this task

The device sends RADIUS stop-accounting requests when it receives connection teardown requests from hosts or connection teardown commands from an administrator. However, the device might fail to receive a response for a stop-accounting request in a single transmission. Enable the device to buffer RADIUS stop-accounting requests that have not received responses from the accounting server. The device will resend the requests until responses are received.

To limit the transmission times, set a maximum number of transmission attempts that can be made for individual RADIUS stop-accounting requests. When the maximum attempts are made for a request, the device discards the buffered request.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Enable buffering of RADIUS stop-accounting requests to which no responses have been received.

   stop-accounting-buffer enable

   By default, the buffering feature is enabled.

4. (Optional.) Set the maximum number of transmission attempts for individual RADIUS stop-accounting requests.

   retry stop-accounting *retries*

   The default setting is 500.

Enabling forcibly sending stop-accounting packets

About this task

Typically, if the device does not send a start-accounting packet to the RADIUS server for an authenticated user, it does not send a stop-accounting packet when the user goes offline. If the server has generated a user entry for the user without start-accounting packets, it does not release the user entry when the user goes offline. This feature forces the device to send stop-accounting packets to the RADIUS server when the user goes offline for timely releasing the user entry on the server.

Procedure

1. Enter system view.

   system-view

2. Enter RADIUS scheme view.

   radius scheme *radius-scheme-name*

3. Enable the device to send stop-accounting packets when users for which no start-accounting packets are sent go offline.

   stop-accounting-packet send-force

By default, forcibly sending stop-accounting packets is disabled. The device does not send stop-accounting packets when users for which no start-accounting packets are sent go offline.

Enabling the RADIUS server load sharing feature

About this task

By default, the device communicates with RADIUS servers based on the server roles. It first attempts to communicate with the primary server, and, if the primary server is unreachable, it then searches for the secondary servers in the order they are configured. The first secondary server in active state is used for communication. In this process, the workload is always placed on the active server.

Use the RADIUS server load sharing feature to dynamically distribute the workload over multiple servers regardless of their server roles. The device forwards an AAA request to the most appropriate server of all active servers in the scheme after it compares the weight values and numbers of currently served users. Specify a weight value for each RADIUS server based on the AAA capacity of the server. A larger weight value indicates a higher AAA capacity.

In RADIUS server load sharing, once the device sends a start-accounting request to a server for a user, it forwards all subsequent accounting requests of the user to the same server. If the accounting server is unreachable, the device returns an accounting failure message rather than searching for another active accounting server.

Procedure

1. Enter system view.

   `system-view`
2. Enter RADIUS scheme view.

   `radius scheme` *`radius-scheme-name`*
3. Enable the RADIUS server load sharing feature.

   `server-load-sharing enable`

   By default, this feature is disabled.

Specifying a RADIUS server selection mode for reauthentication

About this task

Use this feature to configure the RADIUS server selection mechanism in reauthentication. The following RADIUS server selection modes are available:

- **Inherit**—The device uses the RADIUS server that performed authentication for a user to reauthenticate that user. This mode reduces the amount of time used in reauthentication. However, if the RADIUS server is unreachable, the reauthentication will fail.
- **Reselect**—The device searches for a reachable RADIUS server to reauthenticate a user. This mode requires more time than the inherit mode. However, this mode ensures that the device uses the optimal reachable RADIUS server for reauthentication. The following factors affect the RADIUS server selection:
  - Server configuration in the RADIUS scheme, including the configuration order.
  - Enabling status of the RADIUS server load sharing feature.
  - Status of the RADIUS servers in the RADIUS scheme.

Procedure

1. Enter system view.

   `system-view`

2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Specify a RADIUS server selection mode for reauthentication.
   reauthentication server-select { **inherit** | **reselect** }
   By default, the inherit mode is used.

Configuring the RADIUS accounting-on feature

About this task

When the accounting-on feature is enabled, the device automatically sends an accounting-on packet to the RADIUS server after the entire device reboots. Upon receiving the accounting-on packet, the RADIUS server logs out all online users so they can log in again through the device. Without this feature, users cannot log in again after the reboot, because the RADIUS server considers them to come online.

You can configure the interval for which the device waits to resend the accounting-on packet and the maximum number of retries.

The extended accounting-on feature enhances the accounting-on feature in a distributed architecture.

The extended accounting-on feature is applicable to LAN users. The user data is saved to the IRF member devices through which the users access the system. When the extended accounting-on feature is enabled, the system automatically sends an accounting-on packet to the RADIUS server after a member device reboots. The packet contains the member device identifier. Upon receiving the accounting-on packet, the RADIUS server logs out all online users that access the system through the member device. If no users have come online through the member device, the IRF fabric does not send an accounting-on packet after the member device reboots.

Restrictions and guidelines

For the extended accounting-on feature to take effect, the RADIUS server must run on IMC and the accounting-on feature must be enabled.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Enable accounting-on.
   accounting-on enable [**interval** *interval* | **send** *send-times*] *
   By default, the accounting-on feature is disabled.
4. (Optional.) Enable extended accounting-on.
   accounting-on extended
   By default, extended accounting-on is disabled.

Configuring the RADIUS session-control feature

About this task

Enable this feature for the RADIUS server to dynamically change the user authorization information (for example, authorization ACL, VLAN, user group, VSI, or blackhole MAC) or forcibly disconnect users by using session-control packets. This task enables the device to receive RADIUS session-control packets on UDP port 1812.

To verify the session-control packets sent from a RADIUS server, specify the RADIUS server as a session-control client to the device.

Restrictions and guidelines

The RADIUS session-control feature can only work with RADIUS servers running on IMC. The session-control client configuration takes effect only when the session-control feature is enabled.

Procedure

1. Enter system view.

   system-view

2. Enable the session-control feature.

   radius session-control enable

   By default, the session-control feature is disabled.

3. Specify a session-control client.

   radius session-control client { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**key** { **cipher** | **simple** } *string* | **vpn-instance** *vpn-instance-name*] *

   By default, no session-control clients are specified.

Configuring the RADIUS DAS feature

About this task

Dynamic Authorization Extensions (DAE) to RADIUS, defined in RFC 5176, can log off online users and change online user authorization information.

In a RADIUS network, the RADIUS server typically acts as the DAE client (DAC) and the NAS acts as the DAE server (DAS).

When the RADIUS DAS feature is enabled, the NAS performs the following operations:

1. Listens to the default or specified UDP port to receive DAE requests.
2. Logs off online users that match the criteria in the requests, changes their authorization information, shuts down or reboots their access ports, or reauthenticates the users.
3. Sends DAE responses to the DAC.

DAE defines the following types of packets:

- **Disconnect Messages (DMs)**—The DAC sends DM requests to the DAS to log off specific online users.
- **Change of Authorization Messages (CoA Messages)**—The DAC sends CoA requests to the DAS to change the authorization information of specific online users.

Procedure

1. Enter system view.

   system-view

2. Enable the RADIUS DAS feature and enter RADIUS DAS view.

   radius dynamic-author server

   By default, the RADIUS DAS feature is disabled.

3. Specify a RADIUS DAC.

   client { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**key** { **cipher** | **simple** } *string* | **ssl-client-policy** *policy-name* | **vpn-instance** *vpn-instance-name*] *

   By default, no RADIUS DACs are specified.

4. (Optional.) Specify the RADIUS DAS port.

port *port-number*

By default, the RADIUS DAS port is 3799.

5. (Optional.) Set the idle timeout for TLS connections of the RADIUS DAE service.
   timer tls-idle-timeout *timeout-value*

   By default, the idle timeout is 60 seconds for TLS connections.
6. (Optional.) Specify the TLS connection listening port for the RADIUS DAE service.
   tls-port *port-number*

   By default, the TLS connection listening port for the RADIUS DAE service is 2083.

Configuring the device to preferentially process RADIUS authentication requests

About this task

RADIUS requests include RADIUS authentication requests, RADIUS accounting-start requests, RADIUS accounting-update requests, and RADIUS accounting-stop requests. By default, the device processes the RADIUS requests in the sequence that the requests are initiated.

When a large number of users go offline and then try to come online immediately, authentication might fail for these users because of authentication request timeout. To resolve this issue, configure the device to preferentially process authentication requests.

Restrictions and guidelines

Do not perform this task if the RADIUS server identifies users by the username and does not allow repeated authentication for the same username. A violation might cause authentication failure for users that try to come online immediately after going offline.

As a best practice, do not perform this task when the device has online users.

Procedure

1. Enter system view.
   system-view
2. Configure the device to preferentially process RADIUS authentication requests.
   radius authentication-request first

   By default, the device processes RADIUS requests in the sequence that the requests are initiated.

Enabling SNMP notifications for RADIUS

About this task

When SNMP notifications are enabled for RADIUS, the SNMP agent supports the following notifications generated by RADIUS:

- **RADIUS server unreachable notification**—The RADIUS server cannot be reached. RADIUS generates this notification if it does not receive a response to an accounting or authentication request within the specified number of RADIUS request transmission attempts.
- **RADIUS server reachable notification**—The RADIUS server can be reached. RADIUS generates this notification for a previously blocked RADIUS server after the quiet timer expires.
- **Excessive authentication failures notification**—The number of authentication failures compared to the total number of authentication attempts exceeds the specified threshold.

For RADIUS SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Make sure the RADIUS server status change notifications sent by the device can be recognized by the NMS. Choose a MIB node version depending on the NMS requirements. For more information about the MIB node versions, see AAA in *Security Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for RADIUS.

   snmp-agent trap enable radius [**accounting-server-down** | **accounting-server-up** | **authentication-error-threshold** | **authentication-server-down** | **authentication-server-up**] *

   By default, all SNMP notifications are disabled for RADIUS.

3. Set the version of RADIUS server status change MIB nodes.

   radius trap-version { **v1** | **v2** } [**accounting-server-down** | **accounting-server-up** | **authentication-server-down** | **authentication-server-up**] *

   By default, the device sends notifications about RADIUS server status change MIB nodes over SNMPv1.

Enabling RADIUS server parallel authentication

About this task

The implementation mechanism of this feature varies by authentication method as follows:

- For 802.1X EAP termination authentication or MAC authentication, the device simultaneously sends authentication requests to all servers in Active state, without distinguishing between primary and secondary servers. If any one of the servers responds with successful authentication, the user authentication is considered successful. In the case where user authentication fails on all servers, authentication fail-permit is triggered for the user. User authentication is considered failed if no server responds or if all servers respond with a failure result at least once.
- For 802.1X EAP relay authentication, the device uses the primary and secondary server mode for authentication. If the primary server responds with a failure result, the device checks whether any other servers are in Active state. If an active server is detected, the system notifies the device to send an authentication request to the server for re-authentication. If no active server exists, the device considers that the user authentication fails. In the authentication process, an authentication request is sent to the next active server only when the previous authentication server responds with a fail or does not respond. If any one of the servers responds with successful authentication, the user authentication is considered successful. In the case where user authentication fails on all servers, authentication fail-permit is triggered for the user. User authentication is considered failed if no server responds or if all servers respond with a failure result at least once.

Restrictions and guidelines

If you enable both this feature and RADIUS server load sharing (**server-load-sharing enable**), RADIUS server load sharing does not take effect.

If you enable this feature and specify a RADIUS server selection mode for authentication (**reauthentication server-select**), the specified selection mode does not take effect.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Enable RADIUS server parallel authentication.
   server-parallel-auth enable
   By default, RADIUS server parallel authentication is disabled.

Configuring the idle timeout for TLS connections in RADIUS authentication

About this task

After the device processes the last authentication or accounting request, it uses the idle timeout configured for TLS connections in the RADIUS scheme of the last request to update the idle timeout timer. If the timer expires but no authentication or accounting request is received, the system disconnects the TLS connection. Setting the idle timeout to 0 disables the system from disconnecting TLS connections due to timeout.

If you change the idle timeout under a RADIUS scheme from 0 to a non-zero value, the system traverses the connections of the servers under the scheme and updates the idle timeout to the new value.

Procedure

1. Enter system view.
   system-view
2. Enter RADIUS scheme view.
   radius scheme *radius-scheme-name*
3. Specify the idle timeout for TLS connections in RADIUS authentication.
   timer tls-idle-timeout *timeout-value*
   By default, the idle timeout is 60 seconds for TLS connections in RADIUS authentication.

Disabling the RADIUS service

About this task

By default, the RADIUS service is enabled. The device can send and receive RADIUS packets. Attackers might use RADIUS session-control and DAE ports to attack the device. To protect the device when such an attack occurs, disable the RADIUS service temporarily on the device. After the network is secure, re-enable the RADIUS service.

If settings on the RADIUS servers require modification or the RADIUS servers cannot provide services temporarily, you can temporarily disable the RADIUS service on the device.

When the RADIUS service is disabled, the device stops sending and receiving RADIUS packets. If a new user comes online, the device uses the backup authentication, authorization, or accounting method to process that user. If the device has not finished requesting authentication or accounting for a user before the RADIUS service is disabled, it uses the following rules to process that user:

- If the device has sent RADIUS authentication requests for that user to a RADIUS server, the device processes that user depending on whether it receives a response from the RADIUS server.

    - If the device receives a response from the RADIUS server, it uses the response to determine whether that user has passed authentication. If that user has passed authentication, the device assigns authorization information to that user according to the response.
    - If the device does not receive any response from the RADIUS server, it attempts to use the backup authentication method to authenticate that user.
- If the device has sent RADIUS start-accounting requests for that user to a RADIUS server, the device processes that user depending on whether it receives a response from the RADIUS server.
    - If the device receives a response from the RADIUS server, it allows that user to come online. However, the device cannot send out accounting-update or stop-accounting requests to the RADIUS server. It cannot buffer the accounting requests, either. When that user goes offline, the RADIUS server cannot log off that user in time. The accounting result might be inaccurate.
    - If the device does not receive any response from the RADIUS server, it attempts to use the backup accounting method.

Restrictions and guidelines

Disabling the RADIUS service does not affect the RADIUS server feature of the device.

The authentication, authorization, and accounting processes undertaken by other methods are not switched to RADIUS when you re-enable the RADIUS service.

Procedure

1. Enter system view.

   `system-view`
2. Disable the RADIUS service.

   `undo radius enable`

   By default, the RADIUS service is enabled.

   To re-enable the RADIUS service, use the **`radius enable`** command.

Display and maintenance commands for RADIUS

Execute **`display`** commands in any view and **`reset`** commands in user view.

Task	Command
Display information about globally cached RADIUS packets.	**`display radius buffer`**
Display the RADIUS scheme configuration.	**`display radius scheme`** [*`radius-scheme-name`*]
Display authentication and accounting load statistics for all RADIUS servers.	**`display radius server-load statistics`**
Display RADIUS packet statistics.	**`display radius statistics`**
Display information about buffered RADIUS stop-accounting requests to which no responses have been received.	**`display stop-accounting-buffer`** { **`radius-scheme`** *`radius-scheme-name`* | **`session-id`** *`session-id`* | **`time-range`** *`start-time end-time`* | **`user-name`** *`user-name`* }
Clear history authentication and accounting load statistics for all RADIUS servers.	**`reset radius server-load statistics`**

Task	Command
Clear RADIUS statistics.	**reset radius statistics**
Clear the buffered RADIUS stop-accounting requests to which no responses have been received.	**reset stop-accounting-buffer** { **radius-scheme** *radius-scheme-name* | **session-id** *session-id* | **time-range** *start-time end-time* | **user-name** *user-name* }

Configuring HWTACACS

HWTACACS tasks at a glance

To configure HWTACACS, perform the following tasks:

1. Creating an HWTACACS scheme
2. Specifying the HWTACACS authentication servers
3. Specifying the HWTACACS authorization servers
4. Specifying the HWTACACS accounting servers
5. Specifying the shared keys for secure HWTACACS communication

   Perform this task if no shared keys are specified when configuring HWTACACS servers.
6. Specifying an MPLS L3VPN instance for the scheme

   Perform this task if no MPLS L3VPN instances are specified when configuring HWTACACS servers.
7. (Optional.) Setting HWTACACS timers
8. (Optional.) Configuring parameters for HWTACACS packets
   - Specifying the source IP address for outgoing HWTACACS packets
   - Setting the username format and traffic statistics units
   - Setting the DSCP priority for HWTACACS packets
9. (Optional.) Configuring HWTACACS stop-accounting packet buffering

Creating an HWTACACS scheme

Restrictions and guidelines

You can configure a maximum of 16 HWTACACS schemes. An HWTACACS scheme can be used by multiple ISP domains.

Procedure

1. Enter system view.

   system-view
2. Create an HWTACACS scheme and enter HWTACACS scheme view.

   hwtacacs scheme *hwtacacs-scheme-name*

Specifying the HWTACACS authentication servers

About this task

You can specify one primary authentication server and a maximum of 16 secondary authentication servers for an HWTACACS scheme. When the primary server is unreachable, the device searches

for the secondary servers in the order they are configured. The first secondary server in active state is used for communication.

Restrictions and guidelines

If redundancy is not required, specify only the primary server.

An HWTACACS server can function as the primary authentication server in one scheme and as the secondary authentication server in another scheme at the same time.

Two HWTACACS authentication servers in a scheme, primary or secondary, cannot have the same combination of VPN instance, host name, IP address, and port number.

Procedure

1. Enter system view.

   system-view

2. Enter HWTACACS scheme view.

   hwtacacs scheme *hwtacacs-scheme-name*

3. Specify the primary HWTACACS authentication server.

   primary authentication { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **single-connection** | **vpn-instance** *vpn-instance-name*] *

   By default, no primary HWTACACS authentication server is specified.

4. (Optional.) Specify a secondary HWTACACS authentication server.

   secondary authentication { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **single-connection** | **vpn-instance** *vpn-instance-name*] *

   By default, no secondary HWTACACS authentication servers are specified.

Specifying the HWTACACS authorization servers

About this task

You can specify one primary authorization server and a maximum of 16 secondary authorization servers for an HWTACACS scheme. When the primary server is not available, the device searches for the secondary servers in the order they are configured. The first secondary server in active state is used for communication.

Restrictions and guidelines

If redundancy is not required, specify only the primary server.

An HWTACACS server can function as the primary authorization server of one scheme and as the secondary authorization server of another scheme at the same time.

Two HWTACACS authorization servers in a scheme, primary or secondary, cannot have the same combination of VPN instance, host name, IP address, and port number.

Procedure

1. Enter system view.

   system-view

2. Enter HWTACACS scheme view.

   hwtacacs scheme *hwtacacs-scheme-name*

3. Specify the primary HWTACACS authorization server.

primary authorization { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **single-connection** | **vpn-instance** *vpn-instance-name*] *

By default, no primary HWTACACS authorization server is specified.

4. (Optional.) Specify a secondary HWTACACS authorization server.

secondary authorization { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **single-connection** | **vpn-instance** *vpn-instance-name*] *

By default, no secondary HWTACACS authorization servers are specified.

Specifying the HWTACACS accounting servers

About this task

You can specify one primary accounting server and a maximum of 16 secondary accounting servers for an HWTACACS scheme. When the primary server is not available, the device searches for the secondary servers in the order they are configured. The first secondary server in active state is used for communication.

Restrictions and guidelines

If redundancy is not required, specify only the primary server.

An HWTACACS server can function as the primary accounting server of one scheme and as the secondary accounting server of another scheme at the same time.

Two HWTACACS accounting servers in a scheme, primary or secondary, cannot have the same combination of VPN instance, host name, IP address, and port number.

HWTACACS does not support accounting for FTP, SFTP, and SCP users.

Procedure

1. Enter system view.

system-view

2. Enter HWTACACS scheme view.

hwtacacs scheme *hwtacacs-scheme-name*

3. Specify the primary HWTACACS accounting server.

primary accounting { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **single-connection** | **vpn-instance** *vpn-instance-name*] *

By default, no primary HWTACACS accounting server is specified.

4. (Optional.) Specify a secondary HWTACACS accounting server.

secondary accounting { *host-name* | *ipv4-address* | **ipv6** *ipv6-address* } [*port-number* | **key** { **cipher** | **simple** } *string* | **single-connection** | **vpn-instance** *vpn-instance-name*] *

By default, no secondary HWTACACS accounting servers are specified.

Specifying the shared keys for secure HWTACACS communication

About this task

The HWTACACS client and server use the MD5 algorithm and shared keys to generate the Authenticator value for packet authentication and user password encryption. The client and server must use the same key for each type of communication.

Perform this task to configure shared keys for servers in an HWTACACS scheme. The keys take effect on all servers for which a shared key is not individually configured.

Restrictions and guidelines

Make sure the shared key configured on the device is the same as the shared key configured on the HWTACACS server.

Procedure

1. Enter system view.
   system-view
2. Enter HWTACACS scheme view.
   hwtacacs scheme *hwtacacs-scheme-name*
3. Specify a shared key for secure HWTACACS authentication, authorization, or accounting communication.
   key { **accounting** | **authentication** | **authorization** } { **cipher** | **simple** } *string*
   By default, no shared key is specified for secure HWTACACS communication.

Specifying an MPLS L3VPN instance for the scheme

About this task

The VPN instance specified for an HWTACACS scheme applies to all servers in that scheme. If a VPN instance is also configured for an individual HWTACACS server, the VPN instance specified for the HWTACACS scheme does not take effect on that server.

Procedure

1. Enter system view.
   system-view
2. Enter HWTACACS scheme view.
   hwtacacs scheme *hwtacacs-scheme-name*
3. Specify a VPN instance for the HWTACACS scheme.
   vpn-instance *vpn-instance-name*
   By default, an HWTACACS scheme belongs to the public network.

Setting HWTACACS timers

About this task

The device uses the following timers to control communication with an HWTACACS server:

- **Server response timeout timer (response-timeout)**—Defines the HWTACACS server response timeout timer. The device starts this timer immediately after an HWTACACS authentication, authorization, or accounting request is sent. If the device does not receive a

response from the server within the timer, it sets the server to blocked. Then, the device sends the request to another HWTACACS server.

- **Real-time accounting timer (realtime-accounting)**—Defines the intervals at which the device sends real-time accounting packets to the HWTACACS accounting server for online users.
- **Server quiet timer (quiet)**—Defines the duration to keep an unreachable server in blocked state. If a server is not reachable, the device changes the server status to blocked, starts this timer for the server, and tries to communicate with another server in active state. After the server quiet timer expires, the device changes the status of the server back to active.

The server quiet timer setting affects the status of HWTACACS servers. If the scheme includes one primary HWTACACS server and multiple secondary HWTACACS servers, the device communicates with the HWTACACS servers based on the following rules:

- When the primary server is in active state, the device communicates with the primary server. When the primary server is unreachable, the device researches a secondary server in active status in the order they are configured.
- When one or more servers are in active state, the device tries to communicate with these servers only, even if they are unreachable.
- When all servers are in blocked state, the device only tries to communicate with the primary server.
- If the primary server is unreachable, the device changes the server status to blocked and starts a quiet timer for the server. When the quiet timer of the server expires, the status of the server changes back to active. The device does not check the server again during the authentication, authorization, or accounting process.
- The search process continues until the device finds an available secondary server or has checked all secondary servers in active state. If no server is available, the device considers the authentication, authorization, or accounting attempt a failure.
- When you remove a server in use, communication with the server times out. The device looks for a server in active state by first checking the primary server, and then checking secondary servers in the order they are configured.
- When an HWTACACS server's status changes automatically, the device changes this server's status accordingly in all HWTACACS schemes in which this server is specified.

Restrictions and guidelines

A short real-time accounting interval helps improve accounting precision but requires many system resources. When there are 1000 or more users, set a real-time accounting interval longer than 15 minutes.

Procedure

1. Enter system view.

   system-view

2. Enter HWTACACS scheme view.

   hwtacacs scheme *hwtacacs-scheme-name*

3. Set the HWTACACS timers. Choose the following tasks as needed:
   - Set the HWTACACS server response timeout timer.

     timer response-timeout *seconds*

     By default, the HWTACACS server response timeout timer is 5 seconds.
   - Set the real-time accounting interval.

     timer realtime-accounting *minutes*

     By default, the real-time accounting interval is 12 minutes.
   - Set the server quiet timer.

timer quiet *minutes*

By default, the server quiet timer is 5 minutes.

Specifying the source IP address for outgoing HWTACACS packets

About this task

The source IP address of HWTACACS packets that a NAS sends must match the IP address of the NAS configured on the HWTACACS server. An HWTACACS server identifies a NAS by IP address. When the HWTACACS server receives a packet, it checks the source IP address of the packet.

- If it is the IP address of a managed NAS, the server processes the packet.
- If it is not the IP address of a managed NAS, the server drops the packet.

Before sending an HWTACACS packet, the NAS selects a source IP address in the following order:

1. The source IP address specified for the HWTACACS scheme.
2. The source IP address specified in system view for the VPN or public network, depending on where the HWTACACS server resides.
3. The IP address of the outbound interface specified by the route.

Restrictions and guidelines for source IP address configuration

You can specify the source IP address for outgoing HWTACACS packets in HWTACACS scheme view or in system view.

- The IP address specified in HWTACACS scheme view applies to one HWTACACS scheme.
- The IP address specified in system view applies to all HWTACACS schemes.

The source IP address of HWTACACS packets that a NAS sends must match the IP address of the NAS that is configured on the HWTACACS server.

As a best practice, specify a loopback interface address as the source IP address for outgoing HWTACACS packets to avoid HWTACACS packet loss caused by physical port errors.

To communicate with the HWTACACS server, the source address of outgoing HWTACACS packets is typically the IP address of an egress interface on the NAS. However, in some situations, you must change the source IP address. For example, when VRRP is configured for stateful failover, configure the virtual IP of the uplink VRRP group as the source address.

You can directly specify a source IP address for outgoing HWTACACS packets or specify a source interface to provide the source IP address for outgoing HWTACACS packets. The source interface configuration and the source IP address configuration overwrite each other.

Specifying a source interface or source IP address for all HWTACACS schemes

1. Enter system view.

   system-view

2. Specify a source interface or source IP address for outgoing HWTACACS packets.

   hwtacacs nas-ip { **interface** *interface-type interface-number* | { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] }

   By default, the source IP address of an HWTACACS packet sent to the server is the primary IPv4 address or the IPv6 address of the outbound interface.

Specifying a source interface or source IP address for an HWTACACS scheme

1. Enter system view.

   system-view

2. Enter HWTACACS scheme view.

   hwtacacs scheme *hwtacacs-scheme-name*

3. Specify a source interface or source IP address for outgoing HWTACACS packets.

   nas-ip { *ipv4-address* | **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }

   By default, the source IP address of an outgoing HWTACACS packet is that configured by using the **hwtacacs nas-ip** command in system view. If the **hwtacacs nas-ip** command is not used, the source IP address is the primary IP address of the outbound interface.

Setting the username format and traffic statistics units

About this task

A username is typically in the *userid@isp-name* format, where the *isp-name* part represents the user's ISP domain name. By default, the ISP domain name is included in a username. If HWTACACS servers do not recognize usernames that contain ISP domain names, you can configure the device to send usernames without domain names to the servers.

The device reports online user traffic statistics in accounting packets.

Restrictions and guidelines

If two or more ISP domains use the same HWTACACS scheme, configure the HWTACACS scheme to keep the ISP domain name in usernames for domain identification.

For accounting accuracy, make sure the traffic measurement units configured on the device are the same as the traffic measurement units configured on the HWTACACS accounting servers.

Procedure

1. Enter system view.

   system-view

2. Enter HWTACACS scheme view.

   hwtacacs scheme *hwtacacs-scheme-name*

3. Set the format of usernames sent to the HWTACACS servers.

   user-name-format { **keep-original** | **with-domain** | **without-domain** }

   By default, the ISP domain name is included in a username.

4. Set the data flow and packet measurement units for traffic statistics.

   data-flow-format { **data** { **byte** | **giga-byte** | **kilo-byte** | **mega-byte** } | **packet** { **giga-packet** | **kilo-packet** | **mega-packet** | **one-packet** } }*

   By default, traffic is counted in bytes and packets.

Setting the DSCP priority for HWTACACS packets

About this task

DSCP priority determines the transmission priority of HWTACACS packets. A larger value represents a higher priority.

DSCP priority is contained in the ToS field of the IPv4 header and in the Traffic Class field of the IPv6 header.

Procedure

1. Enter system view.

   system-view

2. Set the DSCP priority for HWTACACS packets.

   hwtacacs [**ipv6**] **dscp** *dscp-value*

   By default, the DSCP priority is 0 for HWTACACS packets.

Configuring HWTACACS stop-accounting packet buffering

About this task

The device sends HWTACACS stop-accounting requests when it receives connection teardown requests from hosts or connection teardown commands from an administrator. However, the device might fail to receive a response for a stop-accounting request in a single transmission. Enable the device to buffer HWTACACS stop-accounting requests that have not received responses from the accounting server. The device will resend the requests until responses are received.

To limit the transmission times, set a maximum number of attempts that can be made for transmitting individual HWTACACS stop-accounting requests. When the maximum attempts are made for a request, the device discards the buffered request.

Procedure

1. Enter system view.

   system-view

2. Enter HWTACACS scheme view.

   hwtacacs scheme *hwtacacs-scheme-name*

3. Enable buffering of HWTACACS stop-accounting requests to which no responses have been received.

   stop-accounting-buffer enable

   By default, the buffering feature is enabled.

4. (Optional.) Set the maximum number of transmission attempts for individual HWTACACS stop-accounting requests.

   retry stop-accounting *retries*

   The default setting is 100.

Enabling SNMP notifications for HWTACACS

About this task

When SNMP notifications are enabled for HWTACACS, the SNMP agent supports the following notifications generated by HWTACACS:

- **HWTACACS server unreachable notification**—The HWTACACS server cannot be reached. The NAS generates this notification if it does not receive a response to an accounting, authentication, or authorization request within the specified number of HWTACACS request transmission attempts.
- **HWTACACS server reachable notification**—The HWTACACS server can be reached. The NAS generates this notification for a previously blocked HWTACACS server after the quiet timer expires and sets the server status to active.

For HWTACACS SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for HWTACACS.

   snmp-agent trap enable hwtacacs [**accounting-server-down** | **accounting-server-up** | **authentication-server-down** | **authentication-server-up** | **authorization-server-down** | **authorization-server-up**] *

   By default, all types of SNMP notifications for HWTACACS are disabled.

Display and maintenance commands for HWTACACS

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the configuration or server statistics of HWTACACS schemes.	**display hwtacacs scheme** [*hwtacacs-scheme-name* [**statistics**]]
Display information about buffered HWTACACS stop-accounting requests to which no responses have been received.	**display stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name*
Display HWTACACS packet statistics.	**display hwtacacs statistics**
Clear HWTACACS statistics.	**reset hwtacacs statistics** { **accounting** | **all** | **authentication** | **authorization** }
Clear the buffered HWTACACS stop-accounting requests to which no responses have been received.	**reset stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name*

Configuring LDAP

LDAP tasks at a glance

To configure LDAP, perform the following tasks:

1. Configuring an LDAP server
   a. Creating an LDAP server
   b. Configuring the IP address of the LDAP server
   c. (Optional.) Specifying the LDAP version
   d. (Optional.) Setting the LDAP server timeout period
   e. Configuring administrator attributes
   f. Configuring LDAP user attributes
2. (Optional.) Configuring an LDAP attribute map
3. Creating an LDAP scheme
4. Specifying the LDAP authentication server
5. (Optional.) Specifying the LDAP authorization server
6. (Optional.) Specifying an LDAP attribute map for LDAP authorization

Creating an LDAP server

1. Enter system view.

system-view

2. Create an LDAP server and enter LDAP server view.
   ldap server *server-name*

Configuring the IP address of the LDAP server

Restrictions and guidelines

You can configure either an IPv4 address or an IPv6 address for an LDAP server. If you configure the IP address for an LDAP server multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Enter LDAP server view.
   ldap server *server-name*
3. Configure the IP address of the LDAP server.
   { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**port** *port-number*] [**vpn-instance** *vpn-instance-name*]
   By default, an LDAP server does not have an IP address.

Specifying the LDAP version

Restrictions and guidelines

The device supports LDAPv2 and LDAPv3.

A Microsoft LDAP server supports only LDAPv3.

The LDAP version specified on the device must be consistent with the version specified on the LDAP server.

Procedure

1. Enter system view.
   system-view
2. Enter LDAP server view.
   ldap server *server-name*
3. Specify the LDAP version.
   protocol-version { **v2** | **v3** }
   By default, LDAPv3 is used.

Setting the LDAP server timeout period

About this task

If the device sends a bind or search request to an LDAP server without receiving the server's response within the server timeout period, the authentication or authorization request times out. Then, the device tries the backup authentication or authorization method. If no backup method is configured in the ISP domain, the device considers the authentication or authorization attempt a failure.

Procedure

1. Enter system view.

system-view

2. Enter LDAP server view.

   ldap server *server-name*

3. Set the LDAP server timeout period.

   server-timeout *time-interval*

   By default, the LDAP server timeout period is 10 seconds.

Configuring administrator attributes

About this task

To configure the administrator DN and password for binding with the LDAP server during LDAP authentication:

Procedure

1. Enter system view.

   system-view

2. Enter LDAP server view.

   ldap server *server-name*

3. Specify the administrator DN.

   login-dn *dn-string*

   By default, no administrator DN is specified.

   The administrator DN specified on the device must be the same as the administrator DN configured on the LDAP server.

4. Configure the administrator password.

   login-password { **cipher** | **simple** } *string*

   By default, no administrator password is specified.

Configuring LDAP user attributes

About this task

To authenticate a user, an LDAP client must complete the following operations:

1. Establish a connection to the LDAP server.
2. Obtain the user DN from the LDAP server.
3. Use the user DN and the user's password to bind with the LDAP server.

LDAP provides a DN search mechanism for obtaining the user DN. According to the mechanism, an LDAP client sends search requests to the server based on the search policy determined by the LDAP user attributes of the LDAP client.

The LDAP user attributes include:

- Search base DN.
- Search scope.
- Username attribute.
- Username format.
- User object class.

Restrictions and guidelines

If the LDAP server contains many directory levels, a user DN search starting from the root directory can take a long time. To improve efficiency, you can change the start point by specifying the search base DN.

Procedure

1. Enter system view.
   system-view
2. Enter LDAP server view.
   ldap server *server-name*
3. Specify the user search base DN.
   search-base-dn *base-dn*
   By default, no user search base DN is specified.
4. (Optional.) Specify the user search scope.
   search-scope { **all-level** | **single-level** }
   By default, the user search scope is **all-level**.
5. (Optional.) Specify the username attribute.
   user-parameters user-name-attribute { *name-attribute* | **cn** | **uid** }
   By default, the username attribute is **cn**.
6. (Optional.) Specify the username format.
   user-parameters user-name-format { **with-domain** | **without-domain** }
   By default, the username format is **without-domain**.
7. (Optional.) Specify the user object class.
   user-parameters user-object-class *object-class-name*
   By default, no user object class is specified, and the default user object class on the LDAP server is used. The default user object class for this command varies by server model.

Configuring an LDAP attribute map

About this task

Configure an LDAP attribute map to define a list of LDAP-AAA attribute mapping entries. To apply the LDAP attribute map, specify the name of the LDAP attribute map in the LDAP scheme used for authorization.

The LDAP attribute map feature enables the device to convert LDAP attributes obtained from an LDAP authorization server to device-recognizable AAA attributes based on the mapping entries. Because the device ignores unrecognized LDAP attributes, configure the mapping entries to include important LDAP attributes that should not be ignored.

An LDAP attribute can be mapped only to one AAA attribute. Different LDAP attributes can be mapped to the same AAA attribute.

Procedure

1. Enter system view.
   system-view
2. Create an LDAP attribute map and enter LDAP attribute map view.
   ldap attribute-map *map-name*
3. Configure a mapping entry.

map ldap-attribute *ldap-attribute-name* [**prefix** *prefix-value* **delimiter** *delimiter-value*] **aaa-attribute user-group**

Creating an LDAP scheme

Restrictions and guidelines

You can configure a maximum of 16 LDAP schemes. An LDAP scheme can be used by multiple ISP domains.

Procedure

1. Enter system view.
   system-view
2. Create an LDAP scheme and enter LDAP scheme view.
   ldap scheme *ldap-scheme-name*

Specifying the LDAP authentication server

1. Enter system view.
   system-view
2. Enter LDAP scheme view.
   ldap scheme *ldap-scheme-name*
3. Specify the LDAP authentication server.
   authentication-server *server-name*
   By default, no LDAP authentication server is specified.

Specifying the LDAP authorization server

1. Enter system view.
   system-view
2. Enter LDAP scheme view.
   ldap scheme *ldap-scheme-name*
3. Specify the LDAP authorization server.
   authorization-server *server-name*
   By default, no LDAP authorization server is specified.

Specifying an LDAP attribute map for LDAP authorization

About this task

Specify an LDAP attribute map for LDAP authorization to convert LDAP attributes obtained from the LDAP authorization server to device-recognizable AAA attributes.

Restrictions and guidelines

You can specify only one LDAP attribute map in an LDAP scheme.

Procedure

1. Enter system view.
   system-view

2. Enter LDAP scheme view.
   ldap scheme *ldap-scheme-name*
3. Specify an LDAP attribute map.
   attribute-map *map-name*
   By default, no LDAP attribute map is specified.

Display and maintenance commands for LDAP

Execute **display** commands in any view.

Task	Command
Display the configuration of LDAP schemes.	**display ldap scheme** [*ldap-scheme-name*]

Creating an ISP domain

About ISP domains

In a networking scenario with multiple ISPs, the device can connect to users of different ISPs. These users can have different user attributes, such as different username and password structures, different service types, and different rights. To manage users of different ISPs, configure authentication, authorization, and accounting methods and domain attributes for each ISP domain as needed.

The device supports a maximum of 16 ISP domains, including the system-defined ISP domain **system**. You can specify one of the ISP domains as the default domain.

On the device, each user belongs to an ISP domain. If a user does not provide an ISP domain name at login, the device considers the user belongs to the default ISP domain.

Each ISP domain has a set of system-defined AAA methods, which are local authentication, local authorization, and local accounting. If you do not configure any AAA methods for an ISP domain, the device uses the system-defined AAA methods for users in the domain.

The device chooses an authentication domain for each user in the following order:

1. The authentication domain specified for the access module.
2. The ISP domain in the username.
3. The default ISP domain of the device.

If the chosen domain does not exist on the device, the device searches for the ISP domain that accommodates users assigned to nonexistent domains. (Support for the authentication domain configuration depends on the access module.) If no such ISP domain is configured, user authentication fails.

Restrictions and guidelines for ISP domain configuration

An ISP domain cannot be deleted when it is the default ISP domain. Before you use the **undo domain** command, change the domain to a non-default ISP domain by using the **undo domain default enable** command.

You can modify the settings of the system-defined ISP domain **system**, but you cannot delete the domain.

To avoid RADIUS authentication, authorization, or accounting failures, use short domain names to ensure that usernames containing a domain name do not exceed 253 characters.

To avoid RADIUS accounting failures, make sure the domain name contained in usernames sent to the RADIUS server does not exceed 247 characters.

Creating an ISP domain

1. Enter system view.
   system-view
2. Create an ISP domain and enter ISP domain view.
   domain *isp-name*
   By default, a system-defined ISP domain exists. The domain name is **system**.

Specifying the default ISP domain

1. Enter system view.
   system-view
2. Specify the default ISP domain.
   domain default enable *isp-name*
   By default, the default ISP domain is the system-defined ISP domain **system**.

Specifying an ISP domain for users that are assigned to nonexistent domains

1. Enter system view.
   system-view
2. Specify the ISP domain to accommodate users that are assigned to nonexistent domains.
   domain if-unknown *isp-name*
   By default, no ISP domain is specified to accommodate users that are assigned to nonexistent domains.

Configuring ISP domain attributes

Setting ISP domain status

About this task

By placing the ISP domain in active or blocked state, you allow or deny network service requests from users in the domain.

Procedure

1. Enter system view.
   system-view
2. Enter ISP domain view.
   domain *isp-name*
3. Set the status of the ISP domain.
   state { **active** | **block** }

By default, an ISP domain is in active state, and users in the domain can request network services.

Configuring an authorization profile

About this task

An authorization profile defines a set of authorization attributes and can be used to authorize network resources on a VPN basis, for example, assigning different microsegments to different VPN users. If you apply an authorization profile to an authentication domain, all users in the domain can obtain network resources defined in the profile.

Restrictions and guidelines

Authorization profiles take effect only on 802.1X, MAC, and Web authentication users.

You can configure a maximum of 16 authorization profiles.

Deleting or changing an authorization profile does not affect users that are using the profile. The configuration change takes effect only on new clients that come online afterward.

Procedure

1. Enter system view.

   system-view

2. Create an authorization profile and enter its view.

   aaa author-profile *profile-name*

3. Specify a VPN instance and microsegment binding.

   if-match vpn-instance *vpn-instance-name* **microsegment** *microsegment-id* [**vsi** *vsi-name*]

   By default, no VPN instance and microsegment binding is specified.

   You can bind a VPN instance only to one microsegment but a microsegment to multiple VPN instances.

4. Specify the default microsegment.

   default microsegment *microsegment-id* [**vsi** *vsi-name*]

   By default, no default microsegment exists.

   The system authorizes the default microsegment to non-VPN users and VPN users that do not match any VPN instance specified by the **if-match vpn-instance microsegment** command.

Configuring authorization attributes for an ISP domain

About this task

The device supports the following authorization attributes:

- **ACL**—The device restricts authenticated users to access only the network resources permitted by the ACL.
- **Idle cut**—The device logs out a user if the user's total traffic in the idle timeout period at the specified direction is less than the specified minimum traffic.
- **Maximum number of multicast groups**—The attribute restricts the maximum number of multicast groups that an authenticated user can join concurrently.
- **IPv4 address pool**—The device assigns IPv4 addresses from the pool to authenticated users in the domain.

- **IPv6 address pool**—The device assigns IPv6 addresses from the pool to authenticated users in the domain.
- **Microsegment**—The device authorizes the authenticated users to access the network resources in the specified microsegment.
- **Redirect URL**—The device redirects users in the domain to the URL after they pass authentication.
- **User group**—Authenticated users in the domain obtain all attributes of the user group.
- **VLAN**—The device authorizes the authenticated users to access the network resources in the specified VLAN.
- **VSI**—The device authorizes the authenticated users to access the network resources in the specified VSI.

Restrictions and guidelines

To avoid user logoff caused by authorization attribute conflicts, do not assign an authorization VSI and an authorization VLAN at the same time. The conflict occurs in the following cases:

- The server authorizes a VSI through session control or CoA messages to an online user that is authorized with a VLAN at association.
- The server authorizes a VLAN through session control or CoA messages to an online user that is authorized with a VSI at association.
- The server authorizes a VSI through session control or CoA messages to an online user that uses the default authorization VLAN because no VLAN or VSI is authorized at association.

When you specify matching rules for user terminals, follow these restrictions and guidelines:

- Vertical bars (|) are used as the delimiter in matching rules and cannot be input as part of a rule.
- Do not input vertical bars (|) consecutively, such as ||, |||, or | |.
- Matching rules can include spaces, but spaces do not participate in the match and will occupy string length.

Procedure

1. Enter system view.

   system-view

2. Enter ISP domain view.

   domain *isp-name*

3. Configure authorization attributes for authenticated users in the ISP domain.

   authorization-attribute { **acl** *acl-number* | **author-profile** *profile-name* | **idle-cut** *minutes* [*flow*] [**traffic** { **both** | **inbound** | **outbound** }] | **igmp max-access-number** *max-access-number* | **ip-pool** *ipv4-pool-name* | **ipv6-pool** *ipv6-pool-name* | **microsegment** *microsegment-id* | **mld max-access-number** *max-access-number* | **session-group-profile** *session-group-profile-name* | **redirect** { **deny** | **permit** } **user-agent match** *match-string* | **url** *url-string* | **user-group** *user-group-name* | **vlan** *vlan-id* | **vsi** *vsi-name* }

   The default settings are as follows:

   - The idle cut feature is disabled.
   - An IPv4 user can concurrently join a maximum of four IGMP multicast groups.
   - An IPv6 user can concurrently join a maximum of four MLD multicast groups.
   - No other authorization attributes exist.

Including the idle timeout period in the user online duration to be sent to the server

About this task

If a user goes offline due to connection failure or malfunction, the user's online duration sent to the server includes the idle timeout period assigned by the authorization server. The online duration generated on the server is longer than the actual online duration of the user.

For portal users, the device includes the idle timeout period set for the online portal user detection feature in the user online duration. For more information about online detection for portal users, see "Configuring portal authentication."

Procedure

1. Enter system view.
   `system-view`
2. Enter ISP domain view.
   `domain` *`isp-name`*
3. Configure the device to include the idle timeout period in the user online duration to be sent to the server.
   `session-time include-idle-time`
   By default, the user online duration sent to the server does not include the idle timeout period.

Enabling temporary redirect

About this task

Typically, the device carries the redirect URL coded in JavaScript in HTTP or HTTPS responses sent to users, and the status code in the responses is 200. The users obtain the redirect URL by parsing the JavaScript codes. If the endpoint of a user (application, for example) does not support JavaScript, the user will fail to be redirected.

To resolve this issue, enable the temporary redirect feature. This feature enables the device to send HTTP or HTTPS responses with status code 302 to users so that the users can obtain the redirect URL.

Restrictions and guidelines

As a best practice, enable this feature only when users that do not support decoding Java Script in HTTP/HTTPS packets exist in the network.

This feature takes effect only on wired 802.1X and MAC authentication users authorized with a redirection URL.

If you change the temporary redirect setting, the change takes effect only on users that come online afterwards or users that are reauthorized.

Procedure

1. Enter system view.
   `system-view`
2. Enter ISP domain view.
   `domain` *`isp-name`*
3. Enable temporary redirect.
   `redirect move-temporarily enable`
   By default, temporary redirect is disabled.

Specifying the user address type in an ISP domain

1. Enter system view.
   system-view
2. Enter ISP domain view.
   domain *isp-name*
3. Specify the user address type in the ISP domain.
   user-address-type { **ds-lite** | **ipv6** | **nat64** | **private-ds** | **private-ipv4** | **public-ds** | **public-ipv4** }
   By default, no user address type is specified.

Configuring the authentication failure policy for users in an ISP domain

About this task

By default, users cannot come online if they fail authentication. To have authentication-failed users in an authentication domain stay online and access a limited set of network resources, specify an auth-fail domain to accommodate them. In the auth-fail domain, you can specify a set of authorization and accounting schemes to assign resources to the authentication-failed users.

Restrictions and guidelines

The authen-fail online feature is applicable only to wired 802.1X, Web authentication, and MAC authentication users.

The authen-fail online feature does not apply to users that fail authentication for one of the following reasons:

- Authentication times out, for example, because no authentication servers respond or no matching local users are found.
- The authentication domain is in blocked state or is a denied domain.

Procedure

1. Enter system view.
   system-view
2. Enter ISP domain view.
   domain *isp-name*
3. Configure the authentication failure policy for users in the ISP domain.
   authen-fail { **offline** | **online domain** *new-isp-name* **no-authen** }
   By default, the device logs out users in an ISP domain if the users fail authentication.

Configuring AAA methods for an ISP domain

Configuring authentication methods for an ISP domain

Restrictions and guidelines

When you configure remote authentication, follow these restrictions and guidelines:

- For high availability, you can specify one primary authentication method and multiple backup authentication methods. When the primary method is invalid, the device attempts to use the

backup methods in sequence. For example, the **radius-scheme** *radius-scheme-name* **local none** parameters specify a primary RADIUS authentication method and two backup methods (local authentication and no authentication). The device performs RADIUS authentication by default and performs local authentication when the RADIUS server is invalid. The device does not perform authentication when both of the previous methods are invalid.

The remote authentication method is invalid in the following situations:

- The specified authentication scheme does not exist.
- The device fails to send authentication request packets.
- The device does not receive any authentication response packets from an authentication server.

The local authentication method is invalid if the device fails to find the matching local user configuration.

- If the authentication method uses a RADIUS scheme and the authorization method does not use a RADIUS scheme, AAA accepts only the authentication result from the RADIUS server. The Access-Accept message from the RADIUS server also includes the authorization information, but the device ignores the information.
- If an HWTACACS scheme is specified, the device uses the entered username for role authentication. If a RADIUS scheme is specified, the device uses username **$enab***n***$** on the RADIUS server for role authentication. The variable *n* represents a user role level. For more information about user role authentication, see *Fundamentals Configuration Guide*.

The **none** keyword is not supported in FIPS mode.

When the primary authentication method is local, the following rules apply to the authentication of a user:

- The device uses the backup authentication methods in sequence only if local authentication is invalid for one of the following reasons:
  - An exception occurs in the local authentication process.
  - The device connection is terminated or the user account is not configured on the device.
- The device does not turn to the backup authentication methods if local authentication is invalid because of any other reason. Authentication fails for the user.

Prerequisites

Before configuring authentication methods, complete the following tasks:

1. Determine the access type or service type to be configured. With AAA, you can configure an authentication method for each access type and service type.
2. Determine whether to configure the default authentication method for all access types or service types. The default authentication method applies to all access users. However, the method has a lower priority than the authentication method that is specified for an access type or service type.

Procedure

1. Enter system view.

   system-view

2. Enter ISP domain view.

   domain *isp-name*

3. (Optional.) Specify default authentication methods for all types of users.

   authentication default { **hwtacacs-scheme** *hwtacacs-scheme-name* [**radius-scheme** *radius-scheme-name*] [**local**] [**none**] | **ldap-scheme** *ldap-scheme-name* [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name* | **hwtacacs-scheme** *hwtacacs-scheme-name*] * [**none**] | **local** [**ldap-scheme** *ldap-scheme-name*] | **none** | **radius-scheme**

radius-scheme-name [**hwtacacs-scheme** *hwtacacs-scheme-name*] [**local**] [**none**] }

By default, the default authentication method is **local**.

4. Specify authentication methods for a user type or a service.
   - Specify authentication methods for LAN users.

     authentication lan-access { **ldap-scheme** *ldap-scheme-name* [**local**] [**none**] | **local** [**ldap-scheme** *ldap-scheme-name* | **radius-scheme** *radius-scheme-name*] [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**local**] [**none**] }

     By default, the default authentication methods are used for LAN users.
   - Specify authentication methods for login users.

     authentication login { **hwtacacs-scheme** *hwtacacs-scheme-name* [**radius-scheme** *radius-scheme-name*] [**local**] [**none**] | **ldap-scheme** *ldap-scheme-name* [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name* | **hwtacacs-scheme** *hwtacacs-scheme-name*] * [**none**] | **local** [**ldap-scheme** *ldap-scheme-name*] [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**hwtacacs-scheme** *hwtacacs-scheme-name*] [**local**] [**none**] }

     By default, the default authentication methods are used for login users.
   - Specify authentication methods for portal users.

     authentication portal { **ldap-scheme** *ldap-scheme-name* [**local**] [**none**] | **local** [**ldap-scheme** *ldap-scheme-name* | **radius-scheme** *radius-scheme-name*] [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**local**] [**none**] }

     By default, the default authentication methods are used for portal users.
   - Specify authentication methods for obtaining a temporary user role.

     authentication super { **hwtacacs-scheme** *hwtacacs-scheme-name* | **radius-scheme** *radius-scheme-name* } *

     By default, the default authentication methods are used for obtaining a temporary user role.

Configuring authorization methods for an ISP domain

Restrictions and guidelines

For high availability, you can specify one primary authorization method and multiple backup authorization methods. When the primary method is invalid, the device attempts to use the backup methods in sequence. For example, the **radius-scheme** *radius-scheme-name* **local none** parameters specify a primary RADIUS authorization method and two backup methods (local authorization and no authorization). The device performs RADIUS authorization by default and performs local authorization when the RADIUS server is invalid. The device does not perform authorization when both of the previous methods are invalid.

The remote authorization method is invalid in the following situations:

- The specified authorization scheme does not exist.
- The device fails to send authorization request packets.
- The device does not receive any authorization response packets from an authorization server.

The local authorization method is invalid if the device fails to find the matching local user configuration.

Only SSL VPN users support LDAP authorization in the current software version.

To use a RADIUS scheme as the authorization method, specify the name of the RADIUS scheme that is configured as the authentication method for the ISP domain. If an invalid RADIUS scheme is specified as the authorization method, RADIUS authentication and authorization fail.

The **none** keyword is not supported in FIPS mode.

When the primary authorization method is local, the following rules apply:

- The device uses the backup authorization methods in sequence only if local authorization is invalid for one of the following reasons:
  - An exception occurs in the local authorization process.
  - The device connection is terminated or the user account is not configured on the device
- The device does not turn to the backup authorization methods if local authorization is invalid because of any other reason. Authorization fails for the user.

Prerequisites

Before configuring authorization methods, complete the following tasks:

1. Determine the access type or service type to be configured. With AAA, you can configure an authorization scheme for each access type and service type.
2. Determine whether to configure the default authorization method for all access types or service types. The default authorization method applies to all access users. However, the method has a lower priority than the authorization method that is specified for an access type or service type.

Procedure

1. Enter system view.

   system-view

2. Enter ISP domain view.

   domain *isp-name*

3. (Optional.) Specify default authorization methods for all types of users.

   authorization default { **hwtacacs-scheme** *hwtacacs-scheme-name* [**radius-scheme** *radius-scheme-name*] [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name* | **hwtacacs-scheme** *hwtacacs-scheme-name*] * [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**hwtacacs-scheme** *hwtacacs-scheme-name*] [**local**] [**none**] }

   By default, the authorization method is **local**.

4. Specify authorization methods for a user type or a service.
   - Specify command authorization methods.

     authorization command { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] [**none**] | **local** [**none**] | **none** }

     By default, the default authorization methods are used for command authorization.
   - Specify authorization methods for LAN users.

     authorization lan-access { **local** [**radius-scheme** *radius-scheme-name*] [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**local**] [**none**] }

     By default, the default authorization methods are used for LAN users.
   - Specify authorization methods for login users.

     authorization login { **hwtacacs-scheme** *hwtacacs-scheme-name* [**radius-scheme** *radius-scheme-name*] [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name* | **hwtacacs-scheme** *hwtacacs-scheme-name*] * [**none**] | **none** | **radius-scheme**

radius-scheme-name [**hwtacacs-scheme** *hwtacacs-scheme-name*] [**local**] [**none**] }

By default, the default authorization methods are used for login users.

- Specify authorization methods for portal users.

  authorization portal { **local** [**radius-scheme** *radius-scheme-name*] [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**local**] [**none**] }

  By default, the default authorization methods are used for portal users.

Configuring accounting methods for an ISP domain

Restrictions and guidelines

For high availability, you can specify one primary accounting method and multiple backup accounting methods. When the primary method is invalid, the device attempts to use the backup methods in sequence. For example, the **radius-scheme** *radius-scheme-name* **local none** parameters specify a primary RADIUS accounting method and two backup methods (local accounting and no accounting). The device performs RADIUS accounting by default and performs local accounting when the RADIUS server is invalid. The device does not perform accounting when both of the previous methods are invalid.

The remote accounting method is invalid in the following situations:

- The specified accounting scheme does not exist.
- The device fails to send accounting request packets.
- The device does not receive any accounting response packets from an accounting server.

The local accounting method is invalid if the device fails to find the matching local user configuration.

FTP, SFTP, and SCP users do not support accounting.

Local accounting does not provide statistics for charging. It only counts and controls the number of concurrent users that use the same local user account. The threshold is configured by using the **access-limit** command.

The **none** keyword is not supported in FIPS mode.

When the primary accounting method is local, the following rules apply:

- The device uses the backup accounting methods in sequence only if local accounting is invalid for one of the following reasons:
  - An exception occurs in the local accounting process.
  - The device connection is terminated or the user account is not configured on the device
- The device does not turn to the backup accounting methods if local accounting is invalid because of any other reason. Accounting fails for the user.

Prerequisites

Before configuring accounting methods, complete the following tasks:

1. Determine the access type or service type to be configured. With AAA, you can configure an accounting method for each access type and service type.
2. Determine whether to configure the default accounting method for all access types or service types. The default accounting method applies to all access users. However, the method has a lower priority than the accounting method that is specified for an access type or service type.

Procedure

1. Enter system view.

   system-view
2. Enter ISP domain view.

domain *isp-name*

3. (Optional.) Specify default accounting methods for all types of users.

   accounting default { **hwtacacs-scheme** *hwtacacs-scheme-name* [**radius-scheme** *radius-scheme-name*] [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name* | **hwtacacs-scheme** *hwtacacs-scheme-name*] * [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**hwtacacs-scheme** *hwtacacs-scheme-name*] [**local**] [**none**] }

   By default, the accounting method is **local**.

4. Specify accounting methods for a user type.
   - Specify the command accounting method.

     accounting command hwtacacs-scheme *hwtacacs-scheme-name*

     By default, the default accounting methods are used for command accounting.
   - Specify accounting methods for LAN users.

     accounting lan-access { **broadcast radius-scheme** *radius-scheme-name1* **radius-scheme** *radius-scheme-name2* [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name*] [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**local**] [**none**] }

     By default, the default accounting methods are used for LAN users.
   - Specify accounting methods for login users.

     accounting login { **hwtacacs-scheme** *hwtacacs-scheme-name* [**radius-scheme** *radius-scheme-name*] [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name* | **hwtacacs-scheme** *hwtacacs-scheme-name*] * [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**hwtacacs-scheme** *hwtacacs-scheme-name*] [**local**] [**none**] }

     By default, the default accounting methods are used for login users.
   - Specify accounting methods for portal users.

     accounting portal { **broadcast radius-scheme** *radius-scheme-name1* **radius-scheme** *radius-scheme-name2* [**local**] [**none**] | **local** [**radius-scheme** *radius-scheme-name*] [**none**] | **none** | **radius-scheme** *radius-scheme-name* [**local**] [**none**] }

     By default, the default accounting methods are used for portal users.

5. (Optional.) Configure extended accounting policies.
   - Configure access control for users that encounter accounting-start failures.

     accounting start-fail { **offline** | **online** }

     By default, the device allows users that encounter accounting-start failures to stay online.
   - Configure access control for users that have failed all their accounting-update attempts.

     accounting update-fail { [**max-times** *max-times*] **offline** | **online** }

     By default, the device allows users that have failed all their accounting-update attempts to stay online.
   - Configure access control for users that have used up their data or time accounting quotas.

     accounting quota-out { **offline** | **online** }

     By default, the device logs off users that have used up their accounting quotas.

Display and maintenance commands for ISP domains

Execute **display** commands in any view.

Task	Command
Display configuration information about an ISP domain or all ISP domains.	**display domain** [*isp-name*]

Configuring a critical profile

About this task

During authentication, if all authentication servers in the authentication scheme are unreachable, the device will not receive authentication responses from the server and users will fail authentication.

To resolve this issue, you can configure a critical profile and apply the profile to a port to accommodate the users when all authentication servers are unreachable. The users in a critical profile can access only network resources in the critical profile, such as the critical microsegment. For users in a VPN instance, the users can access the critical microsegment specified for the VPN instance. If no critical microsegment is specified for the VPN instance, the users can access the default critical microsegment.

Restrictions and guidelines

This feature is applicable only to 802.1X and MAC authentication users.

If you remove a critical profile from a port or edit the configuration in the critical profile, the change does not take effect on users already in the critical profile.

Procedure

1. Enter system view.

   system-view
2. Create a critical profile and enter its view.

   aaa critical-profile *profile-name*
3. Specify the default critical microsegment.

   default critical-microsegment *microsegment-id* [**vsi** *vsi-name*] [**url-user-logoff**]

   The **url-user-logoff** keyword is applicable only to MAC authentication users.
4. Specify the critical microsegment for a VPN instance.

   if-match vpn-instance *vpn-instance-name* **critical-microsegment** *microsegment-id* [**vsi** *vsi-name*] [**url-user-logoff**]

   By default, no critical microsegment is specified for a VPN instance.

   The **url-user-logoff** keyword is applicable only to MAC authentication users.

Configuring AAA fail-permit and recovery

About this task

This feature is used to resolve the issue that users cannot come online when all RADIUS servers are unavailable. The feature contains the following settings in a user authentication domain:

- In the user authentication domain, specify a critical domain (also known as fail-permit domain) to accommodate users that access the authentication domain when all RADIUS servers are unavailable. The users can come online in the critical domain without being authenticated when all RADIUS servers are unavailable.
- In the user authentication domain, specify an action to take on users that have been assigned to the critical domain when a RADIUS server becomes available.

- To perform authentication, authorization, and accounting for the users, log off the users.
- To allow the users to stay online without being authenticated, specify a recovery domain. When a RADIUS server becomes available, the users in the critical domain are assigned to the recovery domain.
- To reauthenticate users in the original authentication domain without users being aware of the process, specify the reauthentication action.

For the device to obtain the status of RADIUS authentication servers in time, it detects the status of the RADIUS authentication servers in each RADIUS scheme at intervals. In addition, the device notifies access modules to remove users that use a RADIUS scheme from the critical domain when that RADIUS scheme has reachable RADIUS servers.

Restrictions and guidelines

This feature takes effect only on the 802.1X authentication, Web authentication, and MAC authentication for wired users.

When you specify a critical domain for an ISP domain, follow these restrictions and guidelines:

- If non-none authentication, authorization, or accounting methods are configured in the critical domain for an ISP domain, the non-none authentication or authorization methods cannot take effect on users. However, the non-none accounting methods in the critical domain can take effect on users.
- If an ISP domain has been specified as a critical domain, do not specify a critical domain for that ISP domain. If you do so, the critical domain specified for that ISP domain cannot take effect.
- If a critical domain has been specified for an ISP domain, do not specify that ISP domain as a critical domain. If you do so, that ISP domain cannot act as a critical domain.

When you specify a recovery domain for an ISP domain, follow these restrictions and guidelines:

- If the **none** method is configured as the backup authentication method in the original authentication domain before the users are assigned to the critical domain, the users are still assigned to the recovery domain when a RADIUS server becomes available.
- As a best practice to accurately identify whether a RADIUS authentication server is available and the recovery configuration can take effect in time, configure RADIUS server status detection.
- In the current software version, you can specify only an ISP domain as its own recovery domain in the view of the ISP domain.

When you set the interval at which the device detects the status of RADIUS authentication servers, follow these restrictions and guidelines:

- A too short interval consumes too many system resources for access services. A too long interval cannot detect server status changes in time.
- As a best practice, consider the processing efficiency for access services and the accuracy for fail-permit and recovery when a large number of users come online in a short time.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Set the interval at which the device detects the status of RADIUS authentication servers.

   radius-server authen-state-check interval *interval*

   By default, the device detects the status of RADIUS authentication servers at intervals of 10 minutes.

3. Enter ISP domain view.

   domain *isp-name*

4. Specify a critical domain to accommodate users that access the ISP domain when all RADIUS servers are unavailable.

   authen-radius-unavailable online domain *new-isp-name*

   By default, no critical domain is specified for an ISP domain to accommodate users that access the ISP domain when all RADIUS servers are unavailable.

5. Specify an action to take on users in the critical domain when a RADIUS server becomes available.

   authen-radius-recover { **offline** | **online domain** *new-isp-name* | **re-authen** }

   By default, no action is specified for users in the critical domain when a RADIUS server becomes available.

Setting the maximum number of concurrent login users

About this task

Perform this task to set the maximum number of concurrent users that can log on to the device through a specific protocol, regardless of their authentication methods. The authentication methods include no authentication, local authentication, and remote authentication.

Procedure

1. Enter system view.

   system-view

2. Set the maximum number of concurrent login users.

   In non-FIPS mode:

   aaa session-limit { **ftp** | **http** | **https** | **ssh** | **telnet** } *max-sessions*

   In FIPS mode:

   aaa session-limit { **https** | **ssh** } *max-sessions*

   By default, the maximum number of concurrent login users is 32 for each user type.

Configuring a NAS-ID

About this task

During RADIUS authentication, the device uses a NAS-ID to set the NAS-Identifier attribute of RADIUS packets so that the RADIUS server can identify the access location of users.

The device selects the NAS-ID for the NAS-Identifier attribute in the following order:

1. NAS-ID in a VSRP instance. For more information, see VSRP configuration in *High Availability Configuration Guide*.
2. NAS-ID bound with VLANs in a NAS-ID profile.

If no NAS-ID is configured, the device uses the device name (set by using the **sysname** command) as the NAS-ID.

Configure a NAS-ID profile to maintain NAS-ID and VLAN bindings on the device so that the device can send different NAS-Identifier attribute strings in RADIUS requests from different VLANs.

Restrictions and guidelines

You can apply a NAS-ID profile to portal- or port security-enabled interfaces. For more information, see "Configuring portal authentication" and "Configuring port security."

You can configure multiple NAS-ID and VLAN bindings in a NAS-ID profile.

A NAS-ID can be bound with more than one VLAN, but a VLAN can be bound with only one NAS-ID. If you configure multiple bindings for the same VLAN, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Create a NAS-ID profile and enter NAS-ID profile view.
   aaa nas-id profile *profile-name*
3. Configure a NAS-ID and VLAN binding in the profile.
   nas-id *nas-identifier* **bind vlan** *vlan-id*

Configuring the device ID

About this task

RADIUS uses the value of the Acct-Session-ID attribute as the accounting ID for a user. The device generates an Acct-Session-ID value for each online user based on the system time, random digits, and device ID.

Restrictions and guidelines

On a VSRP network where multiple devices use the same accounting server, accounting ID conflicts might occur. To avoid duplicate accounting IDs, perform this task to assign a unique device ID to each device.

Procedure

1. Enter system view.
   system-view
2. Configure the device ID.
   aaa device-id *device-id*

   By default, the device ID is 0.

Enabling password change prompt logging

About this task

Use this feature to enhance the protection of passwords for Telnet, SSH, HTTP, HTTPS, NETCONF over SSH, and NETCONF over SOAP users and improve the system security.

This feature enables the device to generate logs to prompt users to change their weak passwords at an interval of 24 hours and at the users' login.

A password is a weak password if it does not meet the following requirements:

- Password composition restriction configured by using the **password-control composition** command.
- Minimum password length restriction set by using the **password-control length** command.
- Password complexity checking policy configured by using the **password-control complexity** command.

For a NETCONF over SSH or NETCONF over SOAP user, the device also generates a password change prompt log if any of the following conditions exists:

- The current password of the user is the default password or has expired.
- The user logs in to the device for the first time or uses a new password to log in after global password control is enabled.

The device will no longer generate password change prompt logs for a user when one of the following conditions exists:

- The password change prompt logging feature is disabled.
- The user has changed the password and the new password meets the password control requirements.
- The enabling status of a related password control feature has changed so the current password of the user meets the password control requirements.
- The password composition policy or the minimum password length has changed.

Restrictions and guidelines

You can use the **display password-control** command to display password control configuration. For more information about password control commands, see password control commands in *Security Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enable password change prompt logging.
   local-server log change-password-prompt
   By default, password change prompt logging is enabled.

Configuring the RADIUS server feature

RADIUS server feature tasks at a glance

To configure the RADIUS server feature, perform the following tasks:

1. Configuring RADIUS users
2. Specifying RADIUS clients
3. Activating the RADIUS server configuration

Restrictions and guidelines for the RADIUS server feature

To ensure correct operation of the RADIUS server feature, disable RADIUS session-control on the device.

Configuring RADIUS users

To configure RADIUS users, you must configure network access users, which are the basis of RADIUS user data.

A RADIUS user has the following attributes: user name, password, description, authorization ACL, authorization VLAN, and expiration time. For more information, see "Configuring attributes for network access users."

Specifying RADIUS clients

About this task

Perform this task to specify RADIUS clients and shared keys for centralized management. The RADIUS server feature does not accept requests from RADIUS clients that are not managed by the system.

Restrictions and guidelines

The IP address of a RADIUS client must be the same as the source IP address for outgoing RADIUS packets specified on the RADIUS client.

The shared key of a RADIUS client specified on the RADIUS server must be the same as the setting on the RADIUS client.

Procedure

1. Enter system view.
   `system-view`
2. Specify a RADIUS client.
   `radius-server client ip` *`ipv4-address`* **`key`** { **`cipher`** | **`simple`** } *`string`*

Activating the RADIUS server configuration

About this task

At the device startup, the RADIUS server configuration is automatically activated, including RADIUS users and RADIUS clients. You can immediately activate the most recent RADIUS server configuration if you have added, modified, or deleted RADIUS clients and network access users from which RADIUS user data is generated.

Procedure

1. Enter system view.
   `system-view`
2. Activate the RADIUS server configuration.
   `radius-server activate`

   Executing this command restarts the RADIUS server process and an authentication service interruption will occur during the restart.

Display and maintenance commands for RADIUS users and clients

Execute **`display`** commands in any view.

Task	Command
Display information about activated RADIUS users.	**`display radius-server active-user`** [*`user-name`*]
Display information about activated RADIUS clients.	**`display radius-server active-client`**

Configuring the connection recording policy

About the connection recording policy

Use this feature on scenarios where the device acts as an FTP, SSH, SFTP, or Telnet login client to establish a connection with a login server. This feature enables the device to provide an accounting server with the connection start and termination information. When the login client establishes a connection with the login server, the system sends a start-accounting request to the accounting server. When the connection is terminated, the system sends a stop-accounting request to the accounting server.

Restrictions and guidelines

The device includes the username entered by a user in the accounting packets to be sent to the AAA server for connection recording. The username format configured by using the **user-name-format** command in the accounting scheme does not take effect.

Procedure

1. Enter system view.
   system-view
2. Create a connection recording policy and enter its view.
   aaa connection-recording policy
3. Specify the accounting method for the connection recording policy.
   accounting hwtacacs-scheme *hwtacacs-scheme-name*

Display and maintenance commands for the connection recording policy

Execute **display** commands in any view.

Task	Command
Display the connection recording policy configuration.	**display aaa connection-recording policy**

Displaying user MAC and UUID binding entries

About this task

To help administrators identify whether a universally unique identifier (UUID) has been assigned to a user and maintain the MAC and UUID binding entries for users, use this feature.

In some scenarios, administrators configure and manage MAC address and UUID binding entries for users on the authentication server, depending on the user access control policies. The binding entries will be assigned to the users after they pass authentication. The device automatically generates the MAC and UUID binding entry of a user when the user comes online and automatically deletes the binding entry when the user goes offline.

When a user's client requests the DHCP server to assign or release an IP address, the DHCP snooping module first searches for the UUID bound to the client MAC address. Then, the device replaces the value of Option 61 (client ID) in the client's DHCP packets with the UUID and forwards the DHCP packets to the DHCP server. The DHCP server will assign or reclaim an IP address based on the UUID in the received DHCP packets. UUID-based IP address assignment flexibly meets the IP address requirements of different scenarios.

Restrictions and guidelines

The device supports user MAC and UUID binding entries only for 802.1X and MAC authentication users that use the iNode client.

Procedure

To display user MAC and UUID binding entries, execute the following command in any view:

display mubm record [**mac** *mac-address*] [**interface** *interface-type interface-number* | **slot** *slot-number*] [**access-type** { **dot1x** | **mac-auth** }]

Configuring user online and offline recording

About user online and offline recording

This feature enables the device to record user information when the users fail to come online, or the users go offline normally or abnormally. These records help the administrator to quickly identify the causes of user online and offline faults. The feature occupies the system memory. To reduce the memory usage, you can disable this feature.

Restrictions and guidelines for user online and offline recording configuration

This feature is applicable only to portal, login, and LAN users.

Enabling user online failure recording

1. Enter system view.
   system-view
2. Enable user online failure recording.
   aaa online-fail-record enable
   By default, user online failure recording is disabled.

Enabling user offline recording

1. Enter system view.
   system-view
2. Enable user offline recording.
   aaa offline-record enable
   By default, user offline recording is disabled.
3. Enable user normal offline recording.
   aaa normal-offline-record enable
   By default, user normal offline recording is disabled.

4. Enable user abnormal offline recording.

   aaa abnormal-offline-record enable

   By default, user abnormal offline recording is disabled.

Enabling real-time recording of AAA user associations and disassociations

1. Enter system view.

   system-view

2. Enabling real-time recording of AAA user associations and disassociations.

   aaa { **abnormal-offline-record** | **normal-offline-record** | **online-fail-record** } * **realtime-backup enable**

   By default, real-time recording of AAA user associations and disassociations is disabled.

Display and maintenance commands for user online and offline recording

Execute **display** commands in any view and **reset** commands in user view and **reset** commands in user view.

Task	Command
Display user abnormal offline records.	**display aaa abnormal-offline-record** { **access-type** { **lan-access** | **login** | **portal** } | **domain** *domain-name* | **interface** *interface-type interface-number* | { **ip** *ipv4-address* | **ipv6** *ipv6-address* } | **mac-address** *mac-address* | **s-vlan** *svlan-id* [**c-vlan** *cvlan-id*] | **slot** *slot-number* | **username** *user-name* [**fuzzy-match**] | **reason-code** *code* } * [**time** *begin-time end-time* [**date** *begin-date end-date*]] [**brief** | **count** *count*] **display aaa abnormal-offline-record offline-reason** { **idle-cut** | **quota-out** | **realtime-acct-fail** | **session-timeout** | **user-detect-fail** } [**brief**] **display aaa abnormal-offline-record time** *begin-time end-time* [**date** *begin-date end-date*] [**brief**] **display aaa abnormal-offline-record**
Display user normal offline records.	**display aaa normal-offline-record** { **access-type** { **lan-access** | **login** | **portal** } | **domain** *domain-name* | **interface** *interface-type interface-number* | { **ip** *ipv4-address* | **ipv6** *ipv6-address* } | **mac-address** *mac-address* | **s-vlan** *svlan-id* [**c-vlan** *cvlan-id*] | **slot** *slot-number* | **username** *user-name* [**fuzzy-match**] | **reason-code** *code* } * [**time** *begin-time end-time* [**date** *begin-date end-date*]] [**brief** | **count** *count*] **display aaa normal-offline-record time**

	begin-time end-time [**date** *begin-date end-date*] [**brief**] **display aaa normal-offline-record**
Display user offline records.	**display aaa offline-record** { **access-type** { **lan-access** | **login** | **portal** } | **domain** *domain-name* | **interface** *interface-type interface-number* | { **ip** *ipv4-address* | **ipv6** *ipv6-address* } | **mac-address** *mac-address* | **s-vlan** *svlan-id* [**c-vlan** *cvlan-id*] | **slot** *slot-number* | **username** *user-name* [**fuzzy-match**] | **reason-code** *code* } * [**time** *begin-time end-time* [**date** *begin-date end-date*]] [**brief** | **count** *count*] **display aaa offline-record time** *begin-time end-time* [**date** *begin-date end-date*] [**brief**] **display aaa offline-record**
Display user online failure records.	**display aaa online-fail-record** { **access-type** { **lan-access** | **login** | **portal** } | **domain** *domain-name* | **interface** *interface-type interface-number* | { **ip** *ipv4-address* | **ipv6** *ipv6-address* } | **mac-address** *mac-address* | **s-vlan** *svlan-id* [**c-vlan** *cvlan-id*] | **slot** *slot-number* | **username** *user-name* [**fuzzy-match**] | **reason-code** *code* } * [**time** *begin-time end-time* [**date** *begin-date end-date*]] [**brief** | **count** *count*] **display aaa online-fail-record time** *begin-time end-time* [**date** *begin-date end-date*] [**brief**] **display aaa online-fail-record**
Clear user abnormal offline records.	**reset aaa abnormal-offline-record**
Clear user normal offline records.	**reset aaa normal-offline-record**
Clear user offline records.	**reset aaa offline-record**
Clear user online failure records.	**reset aaa online-fail-record**

Configuring the AAA test feature

About this task

This feature enables the device to send authentication or accounting requests to the specified AAA servers to simulate an authentication or accounting process of a user. Use this feature to identify the reasons for the failure of the interaction between the device and the AAA servers. This feature is applicable only to RADIUS.

When performing an AAA test, the device ignores the status of the specified AAA servers and the RADIUS server load sharing feature. The process of an AAA test is as follows:

1. The device sends authentication requests that carry the specified username and password to the specified authentication server or to the authentication servers in the specified RADIUS scheme. The device tries to communicate with the authentication servers in the specified scheme in sequence.

The process goes to the next step in the following situations:

- The device receives an authentication response (no matter the authentication succeeds or fails).
- The device does not receive any authentication response after making all authentication request attempts.

This step is skipped if no correct authentication server is specified for the AAA test or no authentication servers are configured in the specified RADIUS scheme.

2. The device sends start-accounting requests to the specified accounting server or to the accounting servers in the specified RADIUS scheme. The device tries to communicate with the accounting servers in the specified scheme in sequence.

   The process goes to the next step in the following situations:

   - The device receives a start-accounting response (no matter the accounting succeeds or fails).
   - The device does not receive any start-accounting response after making all start-accounting request attempts.

   This step and the next step are skipped if no correct accounting server is specified for the AAA test or no accounting servers are configured in the specified RADIUS scheme.

3. The device sends stop-accounting requests to the accounting servers to which it has sent a start-accounting request.

   The process finishes in the following situations:

   - The device receives a stop-accounting response.
   - The device does not receive any stop-accounting response after making all stop-accounting request attempts.

To identify attributes that cause authentication or accounting failures, you can configure the device to carry specific attributes in RADIUS requests or define values for specific attributes in the requests. Table 4 shows the attributes that RADIUS requests carry by default.

Table 4 Attributes that RADIUS requests carry by default

Packet type	Attributes that the type of packets carry by default
RADIUS authentication request	User-Name CHAP-Password (or User-Password) CHAP-Challenge NAS-IP-Address (or NAS-IPv6-Address) Service-Type Framed-Protocol NAS-Identifier NAS-Port-Type Acct-Session-Id
RADIUS accounting request	User-Name Acct-Status-Type NAS-IP-Address (or NAS-IPv6-Address) NAS-Identifier Acct-Session-Id Acct-Delay-Time Acct-Terminate-Cause

Restrictions and guidelines

When you perform an AAA test, follow these restrictions and guidelines:

- The device might communicate with the AAA servers incorrectly during an AAA test. Make sure no users come online or go offline during an AAA text.
- If the configuration of the specified RADIUS scheme changes, the new configuration does not affect the current AAA test. The modification will take effect in the next test.
- The system can have only one AAA test at a time. Another AAA test can be performed only after the current test finishes.

When you configure attributes to be included in or excluded from RADIUS requests, follow these restrictions and guidelines:

- Before you include an attribute that is already configured to be excluded from RADIUS requests, you must cancel the exclusion configuration by using the **undo exclude** command.
- Before you exclude an attribute that is already configured to be included in RADIUS requests, you must cancel the inclusion configuration by using the **undo include** command.

Prerequisites

Before you perform an AAA test, you must configure a RADIUS scheme that contains the RADIUS servers to be tested.

Plan the RADIUS attributes to be included in RADIUS requests. Besides the attributes carried by default, the device adds the specified attributes to RADIUS packets in the order that they are specified by using the **include** command. Additional attributes cannot be added to a RADIUS request if the length of the RADIUS request reaches 4096 bytes.

Procedure

1. (Optional.) Configure a RADIUS attribute test group:
   a. Enter system view.
      system-view
   b. Create a RADIUS attribute test group and enter its view.
      radius attribute-test-group *attr-test-group-name*
      You can create multiple RADIUS attribute test groups.
   c. Include an attribute in RADIUS requests.
      include { **accounting** | **authentication** } { **name** *attribute-name* | [**vendor** *vendor-id*] **code** *attribute-code* } **type** { **binary** | **date** | **integer** | **interface-id** | **ip** | **ipv6** | **ipv6-prefix** | **octets** | **string** } value *attribute-value*
      Use this command to add attributes that RADIUS requests do not carry by default to the RADIUS requests.
      For an attribute that RADIUS requests carry by default, you can use this command to change its attribute value.
   d. Exclude an attribute from RADIUS requests.
      exclude { **accounting** | **authentication** } **name** *attribute-name*
      Use this command to exclude an attribute that RADIUS requests carry by default from the RADIUS requests sent during an AAA test to help troubleshoot authentication or accounting failures.
   e. Return to system view.
      quit
   f. Return to user view.
      quit
2. Perform an AAA test in user view.
   test-aaa user *user-name* **password** *password* **radius-scheme** *radius-scheme-name* [**radius-server** { *ipv4-address* | **ipv6**

ipv6-address } *port-number* [**vpn-instance** *vpn-instance-name*]] [**chap** | **pap**] [**attribute-test-group** attr-*test-group-name*] [**trace**]

Configuring authorization information cache for AAA users

Enabling RADIUS authorization information cache for LAN users

About this task

With this feature enabled, when a LAN user passes RADIUS authentication and comes online in an ISP domain, the system records the authorization information returned by the RADIUS server in the authorization information cache. When the user tries to perform RADIUS authentication and come online in the same ISP domain again, if no server is available, the system queries the cached authorization information for the user. If a match is found, the system uses the cached information to perform authentication and authorization and returns authentication success to the user.

With this feature disabled, the system does not record the authorization information returned by the RADIS server for users.

Restrictions and guidelines

For this feature to take effect, you must also enable authorization information cache for AAA users globally (**aaa authorization-cache enable**).

To use this feature, if the accounting scheme and authentication scheme are the same in the ISP domain, configure RADIUS-based remote authentication as the primary accounting method and configure no authentication as the backup accounting method, or allow users that have failed all their accounting-update attempts to stay online.

Procedure

1. Enter system view.
   system-view
2. Enable authorization information cache for AAA users.
   aaa authorization-cache enable
   By default, authorization information cache is disabled for AAA users.
3. Specify the aging time of authorization information cache entries.
   aaa authorization-cache expiry *expiry-time*
   By default, the aging time of authorization information cache entries is 336 hours.
4. Enter ISP domain view.
   domain *isp-name*
5. Enable RADIUS authorization information cache for LAN users.
   authorization lan-access radius-cache enable
   By default, RADIUS authorization information cache is disabled for LAN users.

Display and maintenance commands for authorization information cache for AAA users

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the recorded authorization information cache entries.	**display aaa authorization-cache** [**dot1x** | **mac-auth** | **static** | **web-auth**] [**mac** *mac-address*]
Clear the recorded authorization information cache entries.	**reset aaa authorization-cache** [**dot1x** | **mac-auth** | **static** | **web-auth**] [mac *mac-address*] [**username** *username*]

Enable authorization information cache for AAA users globally

About this task

This feature enables the system to record and query specific authorization information cache of users during the authentication process in each ISP domain. After you disable this feature, the system stops recording and querying user authorization information cache and deletes existing authorization information cache entries.

Procedure

1. Enter system view.
   system-view
2. Enable authorization information cache for AAA users.
   aaa authorization-cache enable
   By default, authorization information cache is disabled for AAA users.
3. Specify the aging time of authorization information cache entries.
   aaa authorization-cache expiry *expiry-time*
   By default, the aging time of authorization information cache entries is 336 hours.
4. Specify the maximum number of authorization information cache entries for AAA users.
   aaa authorization-cache max-number max-number
   By default, the system can record a maximum of 8000 authorization information cache entries for AAA users.
5. Specify the aging time of authorization information cache entries for AAA users matching the specified ACL.
   aaa authorization-cache acl *acl-number* **expiry** *expiry-time*
   By default, the aging time is determined by the configuration of the **aaa authorization-cache expiry** command.

AAA configuration examples

Example: Configuring AAA for SSH users by an HWTACACS server

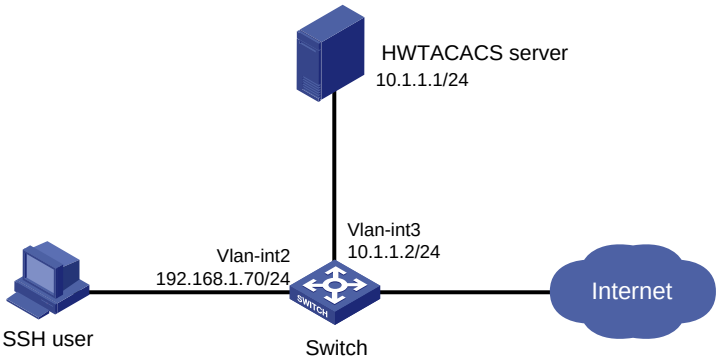
Network configuration

As shown in Figure 16, configure the switch to meet the following requirements:

- Use the HWTACACS server for SSH user authentication, authorization, and accounting.

- Assign the default user role **network-operator** to SSH users after they pass authentication.
- Exclude domain names from the usernames sent to the HWTACACS server.
- Use **expert** as the shared keys for secure HWTACACS communication.

Figure 16 Network diagram



Configuring the HWTACACS server

Set the shared keys to **expert** for secure communication with the switch, add an account for the SSH user, and specify the password. (Details not shown.)

Configuring the switch

Configure IP addresses for the interfaces. (Details not shown.)

Create an HWTACACS scheme.

```
<Switch> system-view
[Switch] hwtacacs scheme hwtac
```

Specify the primary authentication server.

```
[Switch-hwtacacs-hwtac] primary authentication 10.1.1.1 49
```

Specify the primary authorization server.

```
[Switch-hwtacacs-hwtac] primary authorization 10.1.1.1 49
```

Specify the primary accounting server.

```
[Switch-hwtacacs-hwtac] primary accounting 10.1.1.1 49
```

Set the shared keys to **expert** in plaintext form for secure HWTACACS communication.

```
[Switch-hwtacacs-hwtac] key authentication simple expert
[Switch-hwtacacs-hwtac] key authorization simple expert
[Switch-hwtacacs-hwtac] key accounting simple expert
```

Exclude domain names from the usernames sent to the HWTACACS server.

```
[Switch-hwtacacs-hwtac] user-name-format without-domain
[Switch-hwtacacs-hwtac] quit
```

Create an ISP domain named **bbb** and configure the domain to use the HWTACACS scheme for authentication, authorization, and accounting of login users.

```
[Switch-isp-bbb] authentication login hwtacacs-scheme hwtac
[Switch-isp-bbb] authorization login hwtacacs-scheme hwtac
[Switch-isp-bbb] accounting login hwtacacs-scheme hwtac
[Switch-isp-bbb] quit
```

Create local RSA and DSA key pairs.

```
[Switch] public-key local create rsa
[Switch] public-key local create dsa
```

Enable the Stelnet service.

```
[Switch] ssh server enable
```

Enable scheme authentication for user lines VTY 0 through VTY 63.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Enable the default user role feature to assign authenticated SSH users the default user role **network-operator**.

```
[Switch] role default-role enable
```

Verifying the configuration

Initiate an SSH connection to the switch, and enter the correct username and password. The user logs in to the switch. The username must carry the domain name **bbb** in the *username*@**bbb** format. (Details not shown.)

Verify that the user can use the commands permitted by the network-operator user role. (Details not shown.)

Example: Configuring local authentication, HWTACACS authorization, and RADIUS accounting for SSH users

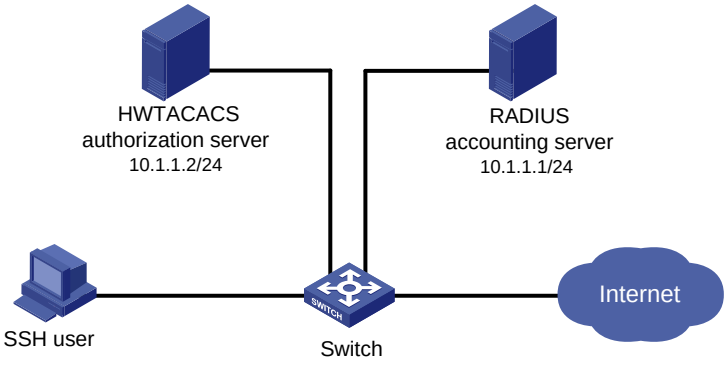
Network configuration

As shown in Figure 17, configure the switch to meet the following requirements:

- Perform local authentication for SSH users.
- Use the HWTACACS server and RADIUS server for SSH user authorization and accounting, respectively.
- Exclude domain names from the usernames sent to the servers.
- Assign the default user role **network-operator** to SSH users after they pass authentication.

Configure an account named **hello** for the SSH user. Configure the shared keys to **expert** for secure communication with the HWTACACS server and RADIUS server.

Figure 17 Network diagram



Configuring the HWTACACS server

Set the shared keys to **expert** for secure communication with the switch, add an account for the SSH user, and specify the password. (Details not shown.)

Configuring the RADIUS server

Set the shared keys to **expert** for secure communication with the switch, add an account for the SSH user, and specify the password. (Details not shown.)

Configuring the switch

Configure IP addresses for interfaces. (Details not shown.)

Create local RSA and DSA key pairs.

```
<Switch> system-view
[Switch] public-key local create rsa
[Switch] public-key local create dsa
```

Enable the Stelnet service.

```
[Switch] ssh server enable
```

Enable scheme authentication for user lines VTY 0 through VTY 63.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Configure an HWTACACS scheme.

```
[Switch] hwtacacs scheme hwtac
[Switch-hwtacacs-hwtac] primary authorization 10.1.1.2 49
[Switch-hwtacacs-hwtac] key authorization simple expert
[Switch-hwtacacs-hwtac] user-name-format without-domain
[Switch-hwtacacs-hwtac] quit
```

Configure a RADIUS scheme.

```
[Switch] radius scheme rd
[Switch-radius-rd] primary accounting 10.1.1.1 1813
[Switch-radius-rd] key accounting simple expert
[Switch-radius-rd] user-name-format without-domain
[Switch-radius-rd] quit
```

Create a device management user.

```
[Switch] local-user hello class manage
```

Assign the SSH service to the local user.

```
[Switch-luser-manage-hello] service-type ssh
```

Set the password to **123456TESTplat&!** in plaintext form for the local user. In FIPS mode, you must set the password in interactive mode.

```
[Switch-luser-manage-hello] password simple 123456TESTplat&!
[Switch-luser-manage-hello] quit
```

Create an ISP domain named **bbb** and configure the login users to use local authentication, HWTACACS authorization, and RADIUS accounting.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login local
[Switch-isp-bbb] authorization login hwtacacs-scheme hwtac
[Switch-isp-bbb] accounting login radius-scheme rd
[Switch-isp-bbb] quit
```

Enable the default user role feature to assign authenticated SSH users the default user role **network-operator**.

```
[Switch] role default-role enable
```

Verifying the configuration

Initiate an SSH connection to the switch, and enter username **hello@bbb** and the correct password. The user logs in to the switch. (Details not shown.)

Verify that the user can use the commands permitted by the network-operator user role. (Details not shown.)

Example: Configuring authentication and authorization for SSH users by a RADIUS server

Network configuration

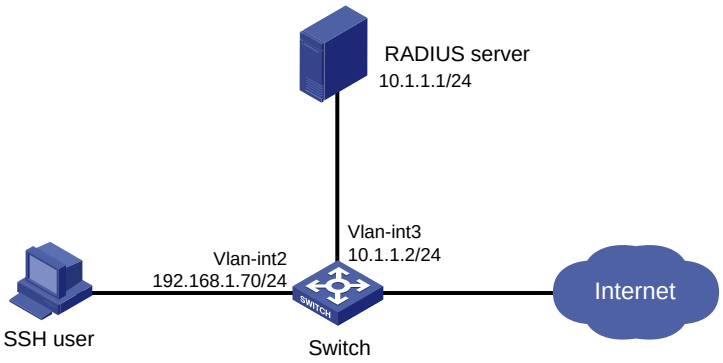
As shown in Figure 18, configure the switch to meet the following requirements:

- Use the RADIUS server for SSH user authentication and authorization.
- Include domain names in the usernames sent to the RADIUS server.
- Assign the **network-admin** user role to SSH users after they pass authentication.

The RADIUS server runs IMC PLAT 7.3 (E0605) and IMC UAM 7.3 (E0512). Add an account with username **hello@bbb** on the RADIUS server.

The RADIUS server and the switch use **expert** as the shared key for secure RADIUS communication. The ports for authentication and accounting are **1812** and **1813**, respectively.

Figure 18 Network diagram



Configuring the RADIUS server

1. Add the switch to the IMC Platform as an access device:

   Log in to IMC, click the **User** tab, and select **User Access Policy** > **Access Device Management** > **Access Device** from the navigation tree. Then, click **Add** to configure an access device as follows:

   a. Set the ports for authentication and accounting to 1812 and 1813, respectively.

   b. Select **Device Management Service** from the **Service Type** list.

   c. Select **H3C (General)** from the **Access Device Type** list.

   d. Set the shared key to **expert** for secure RADIUS communication.

   e. Select an access device from the device list or manually add an access device. In this example, the device IP address is 10.1.1.2.

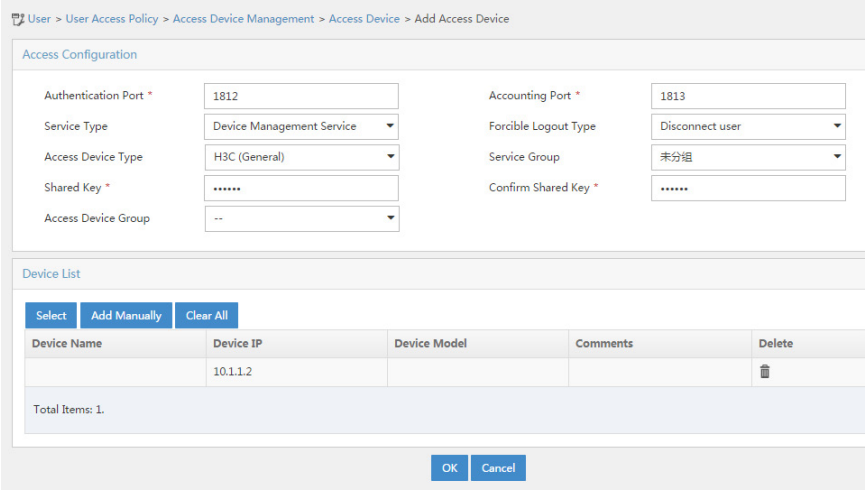
   f. Use the default values for other parameters and click **OK**.

   The IP address of the access device specified here must be the same as the source IP address of the RADIUS packets sent from the switch. The source IP address is chosen in the following order on the switch:

   - IP address specified by using the **`nas-ip`** command.

- IP address specified by using the **`radius nas-ip`** command.
- IP address of the outbound interface (the default).

Figure 19 Adding the switch as an access device



2. Add an account for device management:

Click the **User** tab, and select **Device User** > **Device User** from the navigation tree. Then, click **Add** to configure a device management account as follows:

a. Enter account name **hello@bbb** and specify the password.

b. Select **SSH** from the **Login Type** list.

c. Enter user role **network-admin**.

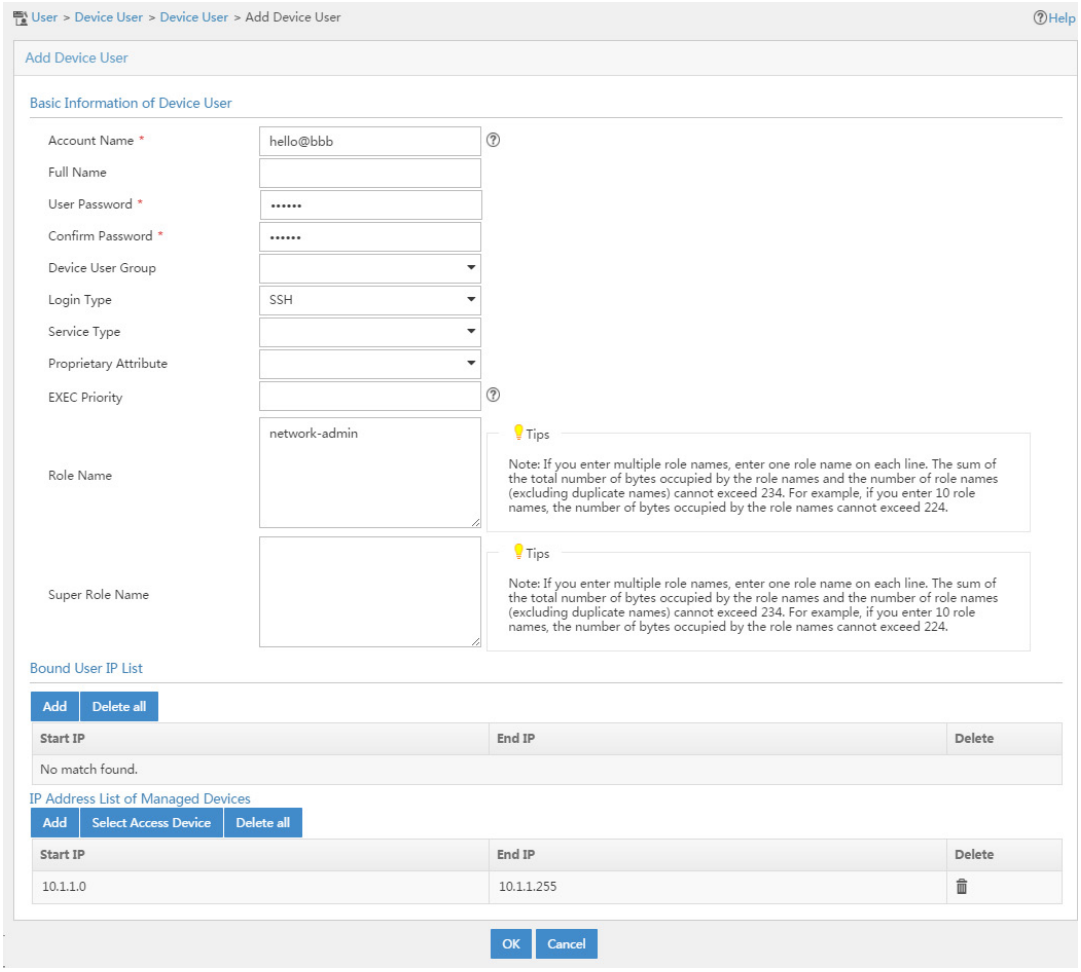
d. Specify 10.1.1.0 to 10.1.1.255 as the IP address range of the hosts to be managed.

e. Click **OK**.

NOTE:

The IP address range must contain the IP address of the switch.

Figure 20 Adding an account for device management



Configuring the switch

Configure IP addresses for interfaces. (Details not shown.)

Create local RSA and DSA key pairs.

```
<Switch> system-view
[Switch] public-key local create rsa
[Switch] public-key local create dsa
```

Enable the Stelnet service.

```
[Switch] ssh server enable
```

Enable scheme authentication for user lines VTY 0 through VTY 63.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Create a RADIUS scheme.

```
[Switch] radius scheme rad
```

Specify the primary authentication server.

```
[Switch-radius-rad] primary authentication 10.1.1.1 1812
```

Set the shared key to **expert** in plaintext form for secure communication with the server.

```
[Switch-radius-rad] key authentication simple expert
```

Include domain names in the usernames sent to the RADIUS server.

```
[Switch-radius-rad] user-name-format with-domain
[Switch-radius-rad] quit
```

Create an ISP domain named **bbb** and configure authentication, authorization, and accounting methods for login users.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login radius-scheme rad
[Switch-isp-bbb] authorization login radius-scheme rad
[Switch-isp-bbb] accounting login none
[Switch-isp-bbb] quit
```

Verifying the configuration

Initiate an SSH connection to the switch, and enter username **hello@bbb** and the correct password. The user logs in to the switch. (Details not shown.)

Verify that the user can use the commands permitted by the network-admin user role. (Details not shown.)

Example: Configuring authentication for SSH users by an LDAP server

Network configuration

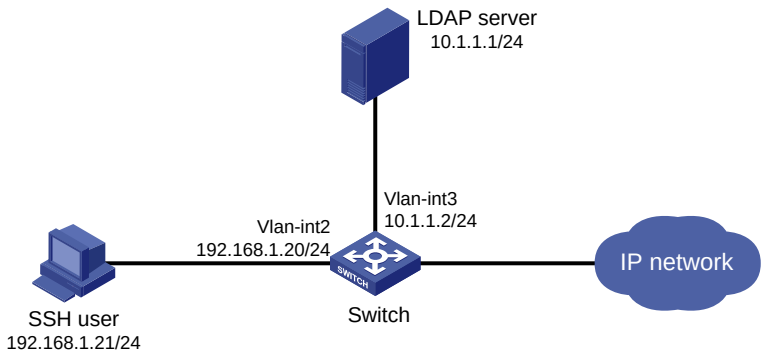
As shown in Figure 21, the LDAP server uses domain **ldap.com** and runs Microsoft Windows 2003 Server Active Directory.

Configure the switch to meet the following requirements:

- Use the LDAP server to authenticate SSH users.
- Assign the level-0 user role to SSH users after they pass authentication.

On the LDAP server, set the administrator password to **admin!123456**, add a user named **aaa**, and set the user's password to **ldap!123456**.

Figure 21 Network diagram



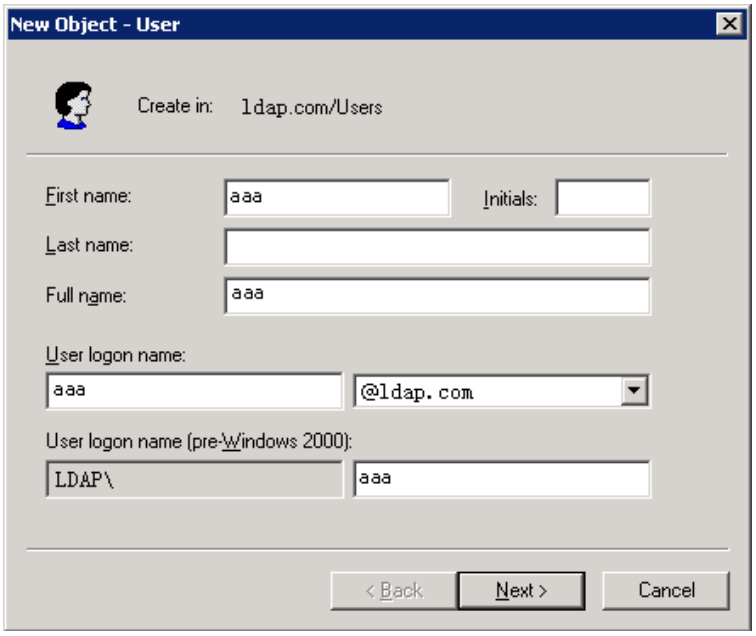
Configuring the LDAP server

1. Add a user named **aaa** and set the password to **ldap!123456**:
   a. On the LDAP server, select **Start** > **Control Panel** > **Administrative Tools**.
   b. Double-click **Active Directory Users and Computers**.
      The **Active Directory Users and Computers** window is displayed.
   c. From the navigation tree, click **Users** under the **ldap.com** node.

d. Select **Action** > **New** > **User** from the menu to display the dialog box for adding a user.

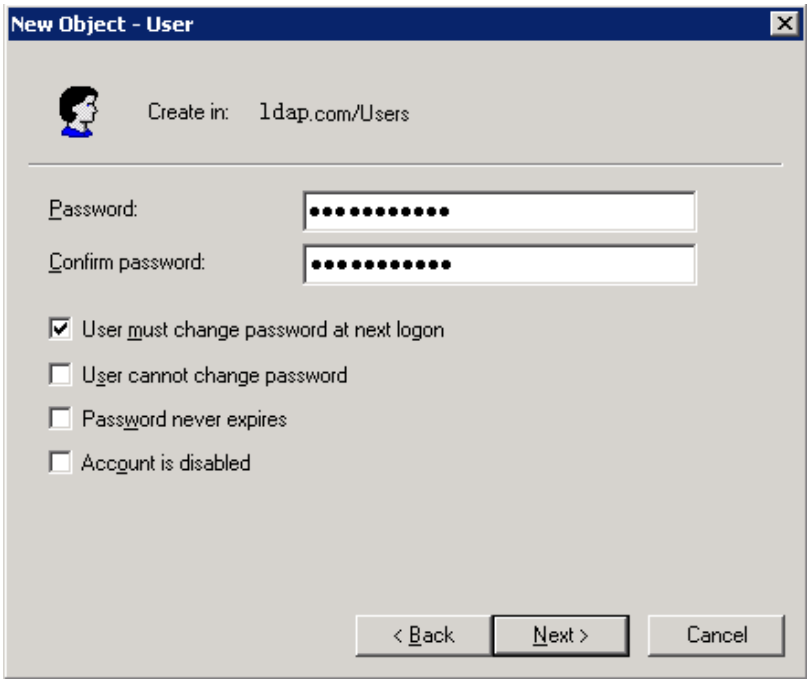
e. Enter logon name **aaa** and click **Next**.

Figure 22 Adding user aaa



f. In the dialog box, enter password **ldap!123456**, select options as needed, and click **Next**.

Figure 23 Setting the user's password



g. Click **OK**.

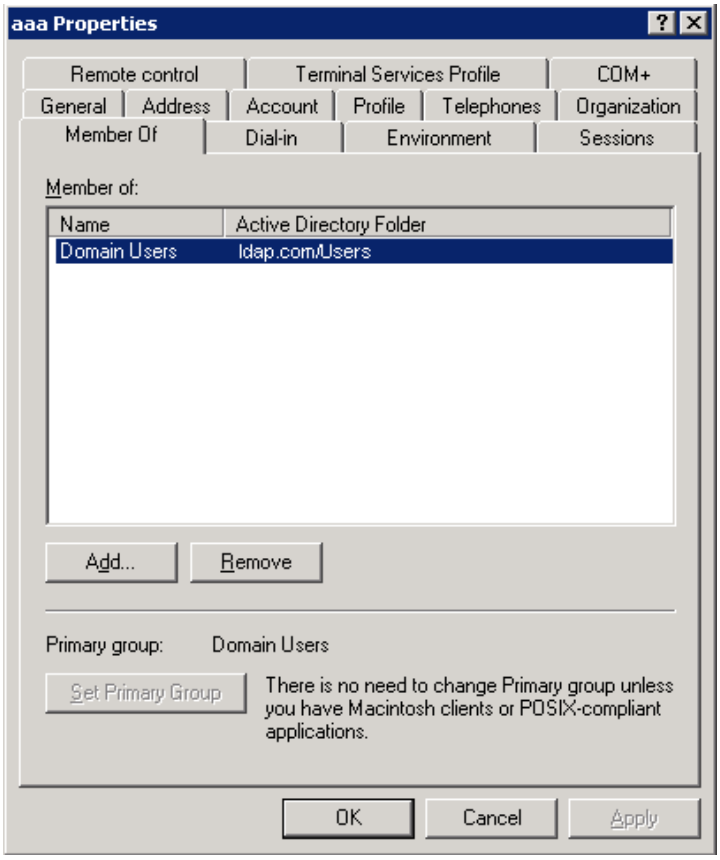
2. Add user **aaa** to group **Users**:

a. From the navigation tree, click **Users** under the **ldap.com** node.

b. In the right pane, right-click user **aaa** and select **Properties**.

c. In the dialog box, click the **Member Of** tab and click **Add**.

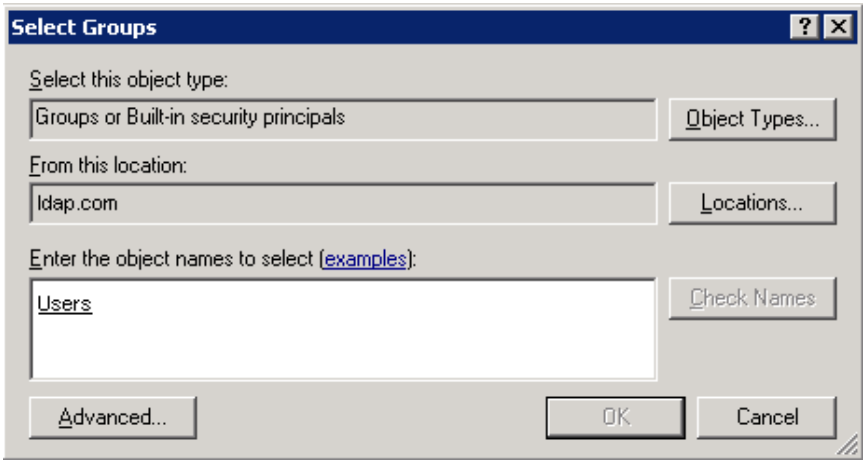
Figure 24 Modifying user properties



d. In the **Select Groups** dialog box, enter **Users** in the **Enter the object names to select** field, and click **OK**.

User **aaa** is added to group **Users**.

Figure 25 Adding user aaa to group Users



3. Set the administrator password to **admin!123456**:

a. In the right pane, right-click user **Administrator** and select **Set Password**.

b. In the dialog box, enter the administrator password. (Details not shown.)

Configuring the switch

Configure IP addresses for interfaces. (Details not shown.)

Create local RSA and DSA key pairs.

```
<Switch> system-view
[Switch] public-key local create rsa
[Switch] public-key local create dsa
```

Enable the Stelnet service.

```
[Switch] ssh server enable
```

Enable scheme authentication for user lines VTY 0 through VTY 63.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Configure an LDAP server.

```
[Switch] ldap server ldap1
```

Specify the IP address of the LDAP authentication server.

```
[Switch-ldap-server-ldap1] ip 10.1.1.1
```

Specify the administrator DN.

```
[Switch-ldap-server-ldap1] login-dn cn=administrator,cn=users,dc=ldap,dc=com
```

Specify the administrator password.

```
[Switch-ldap-server-ldap1] login-password simple admin!123456
```

Configure the base DN for user search.

```
[Switch-ldap-server-ldap1] search-base-dn dc=ldap,dc=com
[Switch-ldap-server-ldap1] quit
```

Create an LDAP scheme.

```
[Switch] ldap scheme ldap-shm1
```

Specify the LDAP authentication server.

```
[Switch-ldap-ldap-shm1] authentication-server ldap1
[Switch-ldap-ldap-shm1] quit
```

Create an ISP domain named **bbb** and configure authentication, authorization, and accounting methods for login users.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login ldap-scheme ldap-shm1
[Switch-isp-bbb] authorization login none
[Switch-isp-bbb] accounting login none
[Switch-isp-bbb] quit
```

Verifying the configuration

Initiate an SSH connection to the switch, and enter username **aaa@bbb** and password **ldap!123456**. The user logs in to the switch. (Details not shown.)

Verify that the user can use the commands permitted by the level-0 user role. (Details not shown.)

Example: Configuring AAA for 802.1X users by a RADIUS server

Network configuration

As shown in Figure 26, configure the switch to meet the following requirements:

- Use the RADIUS server for authentication, authorization, and accounting of 802.1X users.

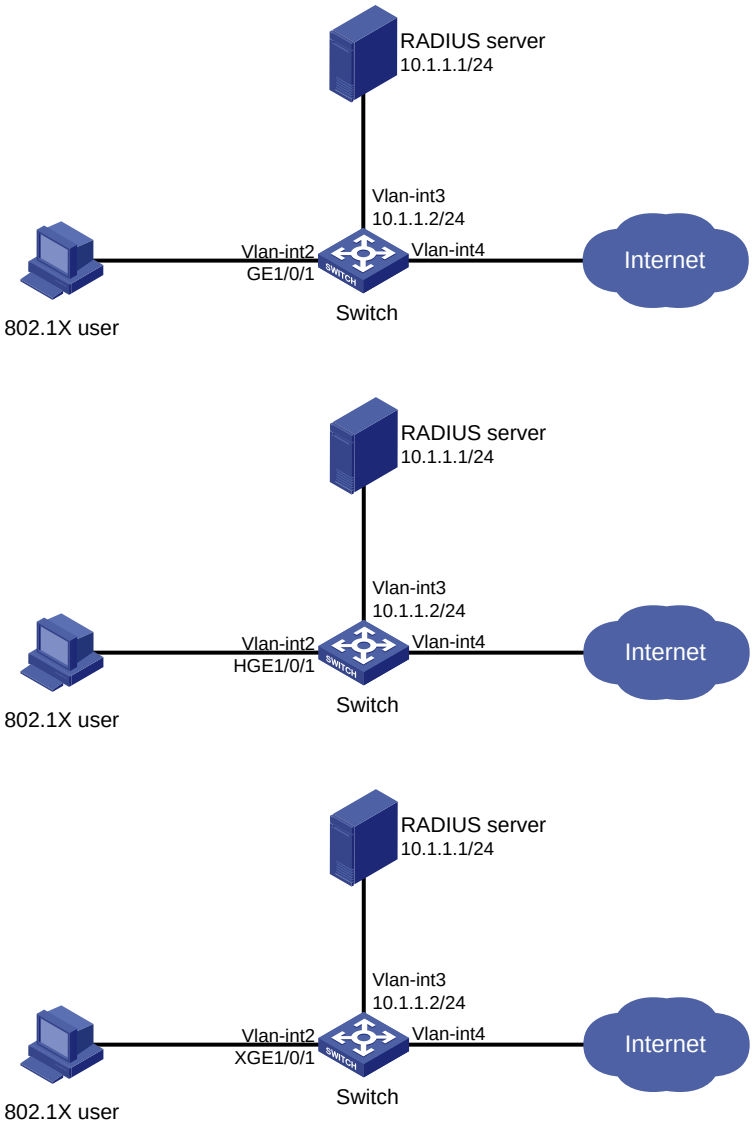
- Use MAC-based access control on GigabitEthernet 1/0/1 to authenticate all 802.1X users on the port separately.
- Include domain names in the usernames sent to the RADIUS server.

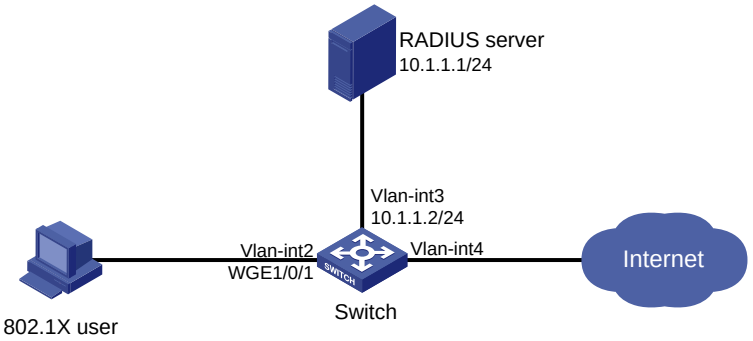
In this example, the RADIUS server runs IMC PLAT 7.1 (E0303), IMC UAM 7.1 (E0304), and IMC CAMS 7.1 (E0301). On the RADIUS server, perform the following tasks:

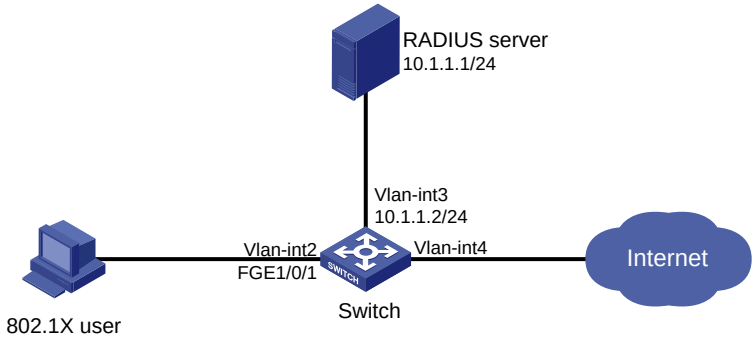
- Add a service that charges 120 dollars for up to 120 hours per month and assigns authenticated users to VLAN 4.
- Configure a user with name **dot1x@bbb** and assign the service to the user.

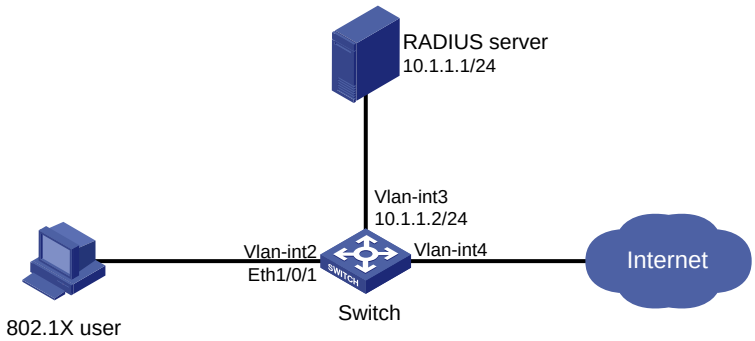
Set the shared keys to **expert** for secure RADIUS communication. Set the ports for authentication and accounting to 1812 and 1813, respectively.

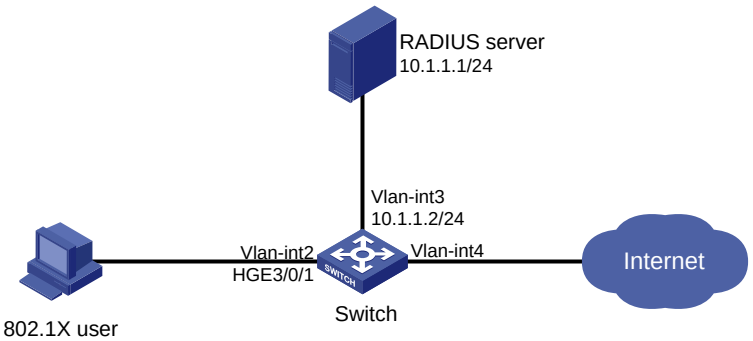
Figure 26 Network diagram

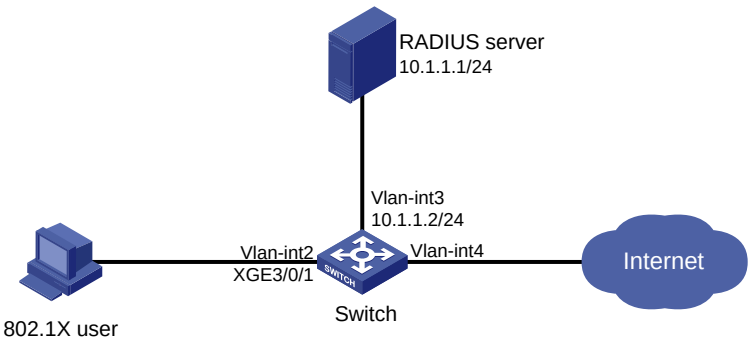


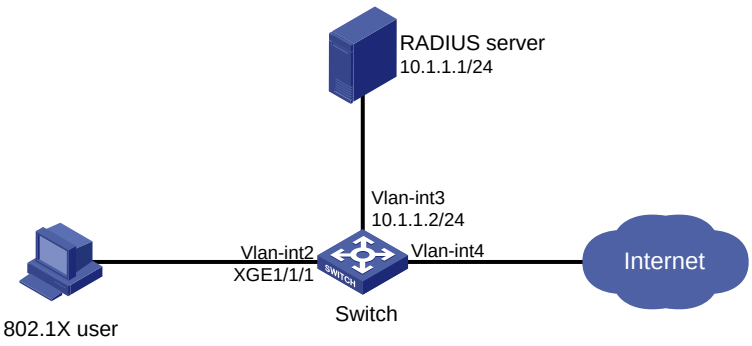

RADIUS server
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
WGE1/0/1
Vlan-int4
Internet
802.1X user
Switch

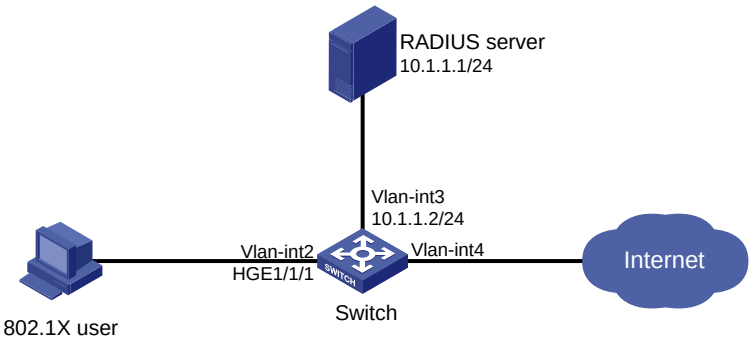

RADIUS server
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
FGE1/0/1
Vlan-int4
Internet
802.1X user
Switch

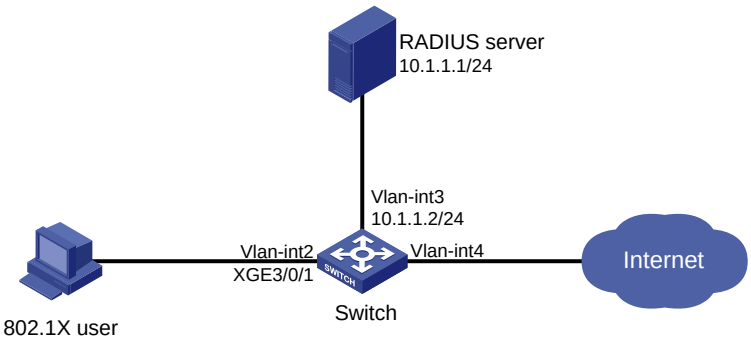

RADIUS server
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
Eth1/0/1
Vlan-int4
Internet
802.1X user
Switch


RADIUS server
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
HGE3/0/1
Vlan-int4
Internet
802.1X user
Switch









Configuring the RADIUS server

1. Add the switch to the IMC Platform as an access device:

   Log in to IMC, click the **User** tab, and select **User Access Policy** > **Access Device Management** > **Access Device** from the navigation tree. Then, click **Add** to configure an access device as follows:

   a. Set the ports for authentication and accounting to 1812 and 1813, respectively.

b. Select **LAN Access Service** from the **Service Type** list.

c. Select **H3C(General)** from the **Access Device Type** list.

d. Set the shared key to **expert** for secure authentication and accounting communication.

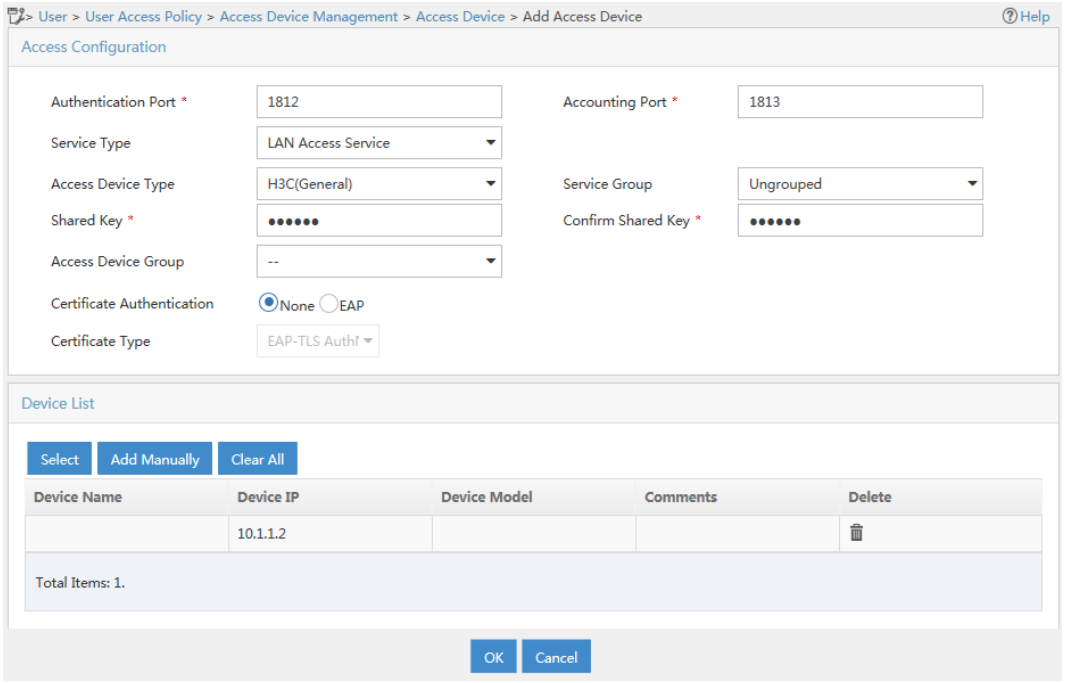
e. Select an access device from the device list or manually add an access device. In this example, the device IP address is 10.1.1.2.

f. Use the default values for other parameters and click **OK**.

The IP address of the access device specified here must be the same as the source IP address of the RADIUS packets sent from the switch. The source IP address is chosen in the following order on the switch:

- IP address specified by using the **`nas-ip`** command.
- IP address specified by using the **`radius nas-ip`** command.
- IP address of the outbound interface (the default).

Figure 27 Adding the switch as an access device



2. Add a charging plan:

Click the **User** tab, and select **Accounting Manager** > **Charging Plans** from the navigation tree to enter the charging plan configuration page. Then, click **Add** to configure a charging plan as follows:

a. Add a plan named **UserAcct**.

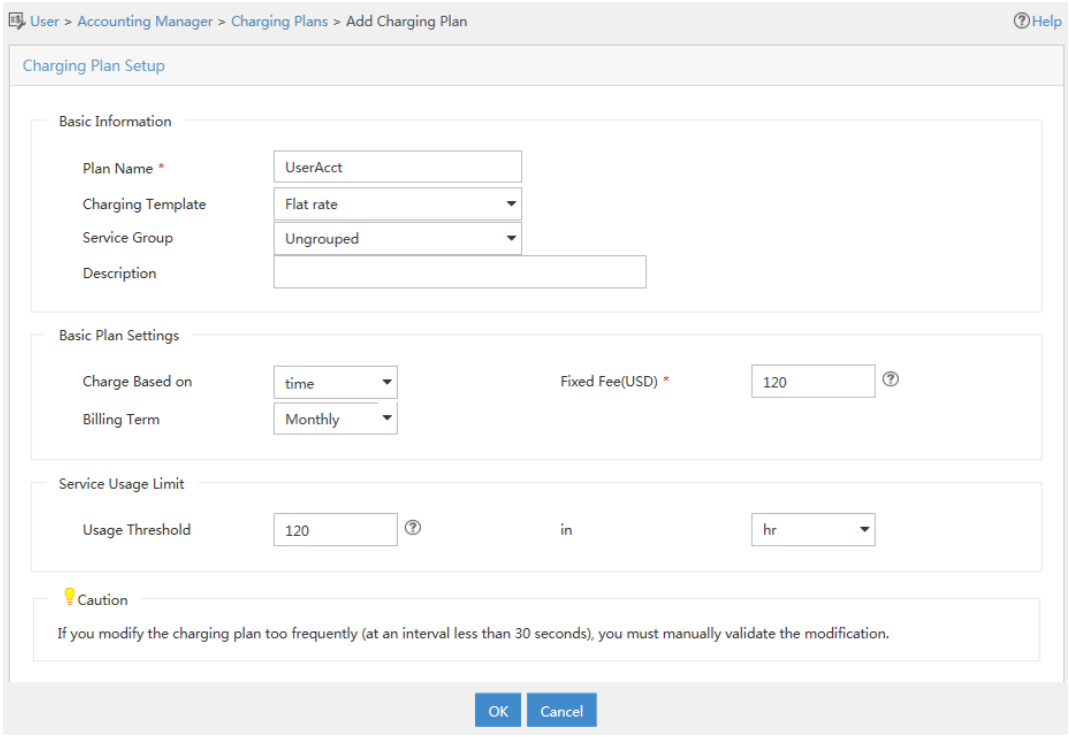
b. Select **Flat rate** from the **Charging Template** list.

c. Select **time** for **Charge Based on**, select **Monthly** for **Billing Term**, and enter **120** in the **Fixed Fee** field.

d. Enter **120** in the **Usage Threshold** field and select **hr** (hours) for the **in** field. The configuration allows the user to access the Internet for up to 120 hours per month.

e. Click **OK**.

Figure 28 Adding a charging plan



3. Add an access policy:

   Click the **User** tab, and select **User Access Policy** > **Access Policy** from the navigation tree. Then, click **Add** to configure an access policy as follows:

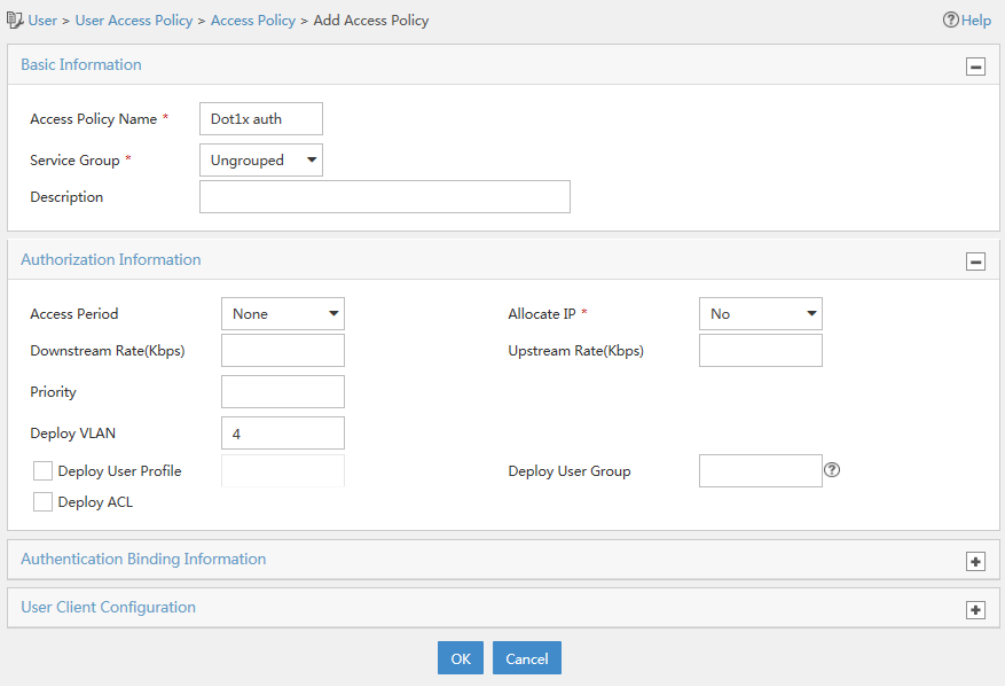
   a. Enter access policy name **Dot1x auth**.

   b. Set the ID of the VLAN to be deployed to **4**.

   c. Configure other parameters as needed.

   d. Click **OK**.

Figure 29 Adding an access policy



4. Add an access service:

   Click the **User** tab, and select **User Access Policy** > **Access Service** from the navigation tree. Then, click **Add** to configure a service as follows:

   a. Add a service named **Dot1x Service,** and set the service suffix to **bbb**, the authentication domain for the 802.1X user. With the service suffix configured, you must configure the access device to send usernames that include domain names to the RADIUS server.

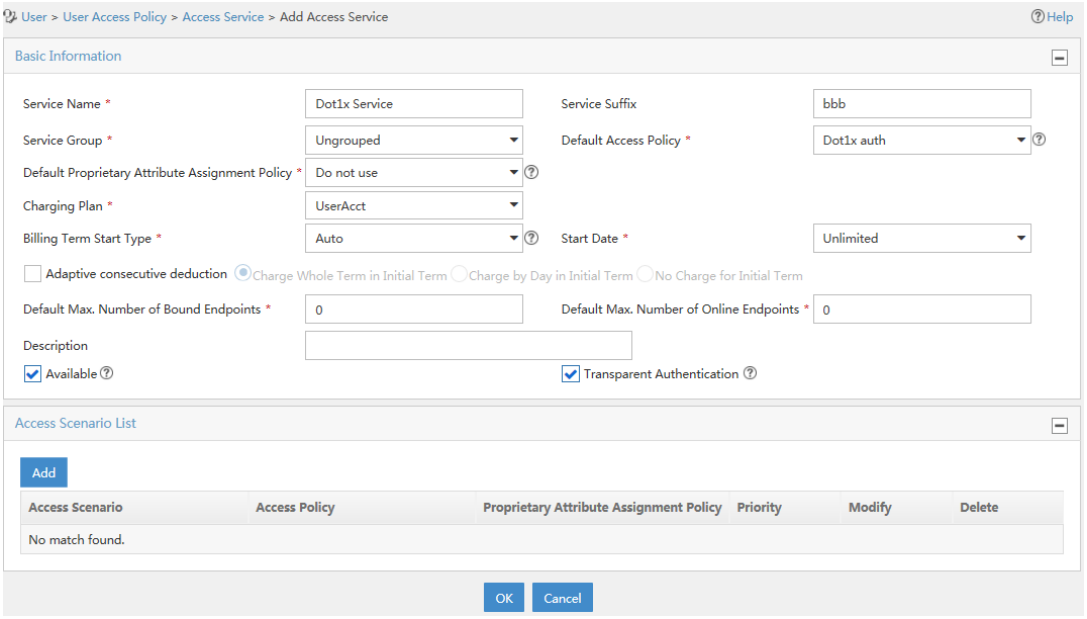
   b. Select **UserAcct** from the **Charging Plan** list.

   c. Select **Dot1x auth** from the **Default Access Policy** list.

   d. Configure other parameters as needed.

   e. Click **OK**.

Figure 30 Adding an access service



5. Add a user:

   Click the **User** tab, and select **All Access Users** from the navigation tree to enter the **All Access Users** page. Then, click **Add** to configure a user as follows:

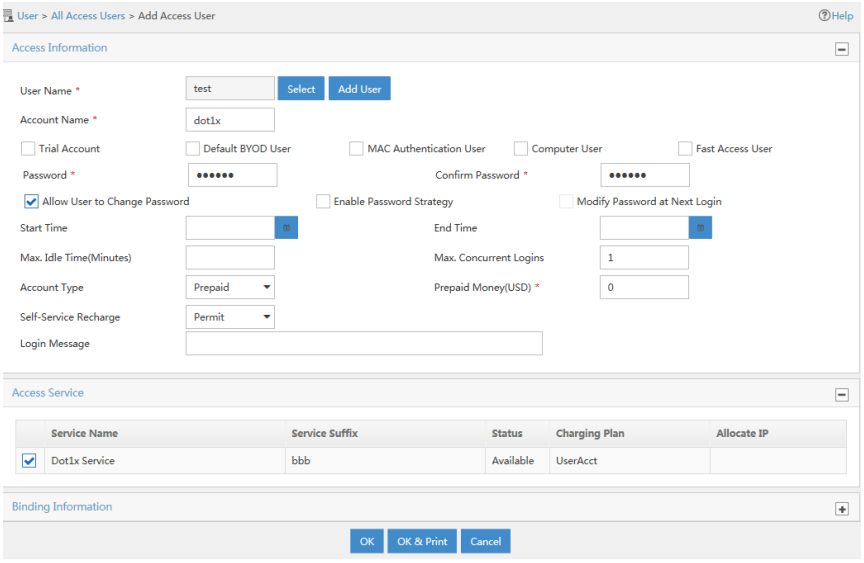
   a. Select the user or add a user named **test**.

   b. Specify the account name as **dot1x** and configure the password.

   c. Select **Dot1x Service** in the **Access Service** area.

   d. Configure other parameters as needed and click **OK**.

Figure 31 Adding an access user account



Configuring the switch

1. Configure a RADIUS scheme:

   # Create a RADIUS scheme named **rad** and enter RADIUS scheme view.

   ```
   <Switch> system-view
   ```

```
[Switch] radius scheme rad
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rad] primary authentication 10.1.1.1
[Switch-radius-rad] primary accounting 10.1.1.1
[Switch-radius-rad] key authentication simple expert
[Switch-radius-rad] key accounting simple expert
```

Include domain names in the usernames sent to the RADIUS server.

```
[Switch-radius-rad] user-name-format with-domain
[Switch-radius-rad] quit
```

2. Configure an ISP domain:

   # Create an ISP domain named **bbb** and enter ISP domain view.

   ```
   [Switch] domain bbb
   ```

   # Configure the ISP domain to use RADIUS scheme **rad** for authentication, authorization, and accounting of LAN users.

   ```
   [Switch-isp-bbb] authentication lan-access radius-scheme rad
   [Switch-isp-bbb] authorization lan-access radius-scheme rad
   [Switch-isp-bbb] accounting lan-access radius-scheme rad
   [Switch-isp-bbb] quit
   ```

3. Configure 802.1X authentication:

   # Enable 802.1X globally.

   ```
   [Switch] dot1x
   ```

   # Enable 802.1X for GigabitEthernet 1/0/1.

   ```
   [Switch] interface gigabitethernet 1/0/1
   [Switch-GigabitEthernet1/0/1] dot1x
   ```

   # Configure the access control method. By default, an 802.1X-enabled port uses the MAC-based access control.

   ```
   [Switch-GigabitEthernet1/0/1] dot1x port-method macbased
   ```

Verifying the configuration

1. On the host, use account **dot1x@bbb** to pass 802.1X authentication:

   # If the host runs the Windows XP 802.1X client, configure the network connection properties as follows:

   a. Click the **Authentication** tab of the properties window.
   b. Select the **Enable IEEE 802.1X authentication for this network** option.
   c. Select MD5 challenge as the EAP type.
   d. Click **OK**.

   The user passes authentication after entering the correct username and password on the authentication page.

   # If the host runs the iNode client, no advanced authentication options are required. The user can pass authentication after entering username **dot1x@bbb** and the correct password on the client property page.

   IMPORTANT:

   Make sure the client can update its IP address to access the resources in the authorized VLAN after passing authentication.

2. On the switch, verify that the server assigns the port connecting the client to VLAN 4 after the user passes authentication. (Details not shown.)

3. Display 802.1X connection information on the switch.

```
[Switch] display dot1x connection
```

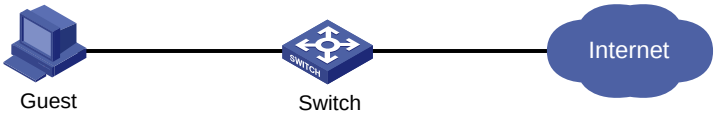
Example: Configuring and managing a local guest

Network configuration

As shown in Figure 32, create an 802.1X local guest named **user1** for Jack. Configure local guest attributes and manage the local guest on the switch as follows:

- Configure attributes for the local guest, including the password, user group, validity period, and sponsor information.
- Enable the local user auto-delete feature.
- Specify an SMTP server and email sender address for the device to send local guest email notifications.
- Configure email addresses for the local guest, guest manager, and guest sponsor.
- Configure the subject and body of the email notifications to be sent to the guest, guest sponsor, and guest manager.
- Send email notifications of the local guest account information to the guest and guest sponsor.

Figure 32 Network diagram



Procedure

1. Configure 802.1X settings. Make sure the guest can pass 802.1X authentication to access the network. (Details not shown.)
2. Manage local guests:

   # Enable the local user auto-delete feature for expired local guests.

   ```
   <Switch> system-view
   [Switch] local-user auto-delete enable
   ```

   # Specify an SMTP server to send local guest email notifications.

   ```
   [Switch] local-guest email smtp-server smtp://192.168.0.112/smtp
   ```

   # Specify the email sender address as **bbb@ccc.com** in the email notifications sent by the device for local guests.

   ```
   [Switch] local-guest email sender bbb@ccc.com
   ```

   # Specify the email address of the guest manager as **guest-manager@ccc.com**.

   ```
   [Switch] local-guest manager-email guest-manager@ccc.com
   ```

   # Configure the subject and body of the email notifications to be sent to the local guest.

   ```
   [Switch] local-guest email format to guest subject Guest account information
   [Switch] local-guest email format to guest body A guest account has been created for
   you. The username, password, and validity period of the account are given below.
   ```

   # Configure the subject and body of the email notifications to be sent to the guest sponsor.

   ```
   [Switch] local-guest email format to sponsor subject Guest account information
   [Switch] local-guest email format to sponsor body A guest account has been created
   for you. The username, password, and validity period of the account are given below.
   ```

   # Configure the subject and body of the email notifications to be sent to the guest manager.

   ```
   [Switch] local-guest email format to manager subject Guest registration information
   ```

```
[Switch] local-guest email format to manager body A guest account has been registered.
The username of the account is given below. Please approve the registration
information.
```

3. Configure the local guest:

Create a user group named **guest1**.

```
[Switch] user-group guest1
[Switch-ugroup-guest1] quit
```

Create a local guest named **user1** and enter local guest view.

```
[Switch] local-user user1 class network guest
```

Set the guest password to **123456** in plain text.

```
[Switch-luser-network(guest)-user1] password simple 123456
```

Assign the guest to user group **guest1**.

```
[Switch-luser-network(guest)-user1] group guest1
```

Specify the name of the local guest.

```
[Switch-luser-network(guest)-user1] full-name Jack
```

Specify the company of the local guest.

```
[Switch-luser-network(guest)-user1] company cc
```

Configure the email address of the local guest.

```
[Switch-luser-network(guest)-user1] email Jack@cc.com
```

Configure the phone number of the local guest.

```
[Switch-luser-network(guest)-user1] phone 131129237
```

Configure a description for the local guest.

```
[Switch-luser-network(guest)-user1] description A guest from company cc
```

Configure the validity period of the local guest.

```
[Switch-luser-network(guest)-user1] validity-datetime from 2015/4/1 08:00:00 to
2015/4/3 18:00:00
```

Specify the guest sponsor name as **Sam**.

```
[Switch-luser-network(guest)-user1] sponsor-full-name Sam
```

Configure the email address of the guest sponsor.

```
[Switch-luser-network(guest)-user1] sponsor-email Sam@aa.com
```

Configure the department of the guest sponsor as **security**.

```
[Switch-luser-network(guest)-user1] sponsor-department security
[Switch-luser-network(guest)-user1] quit
[Switch] quit
```

4. Configure the device to send guest email notifications:

Send an email notification to the guest sponsor.

```
<Switch> local-guest send-email user-name user1 to sponsor
```

Send an email notification to the guest.

```
<Switch> local-guest send-email user-name user1 to guest
```

Verifying the configuration

Display local guest information.

```
<Switch> display local-user user-name user1 class network guest
Total 1 local users matched.

Network access guest user1:
  State:                    Active
```

```
Service type:              LAN-access/Portal
User group:                guest1
Full name:                 Jack
Company:                   cc
Email:                     Jack@cc.com
Phone:                     131129237
Description:               A guest from company cc
Sponsor full name:         Sam
Sponsor department:        security
Sponsor email:             Sam@aa.com
Description:               A guest from company cc
Validity period:
  Start date and time:     2015/04/01-08:00:00
  Expiration date and time:2015/04/03-18:00:00
```

Verify that Jack can use username **user1** and password **123456** to pass local authentication and come online during the validity period. (Details not shown.)

Example: Configuring authentication and authorization for 802.1X users by the device as a RADIUS server

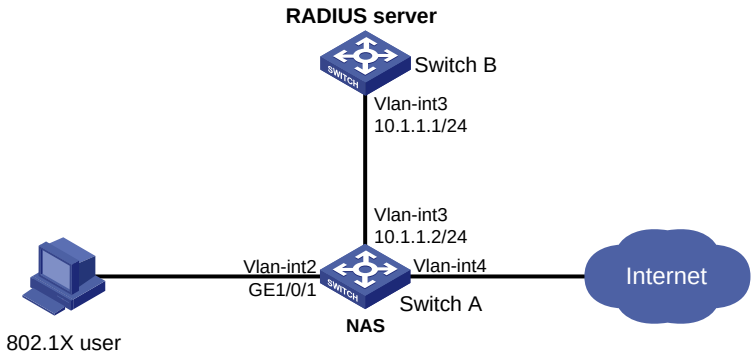
Network configuration

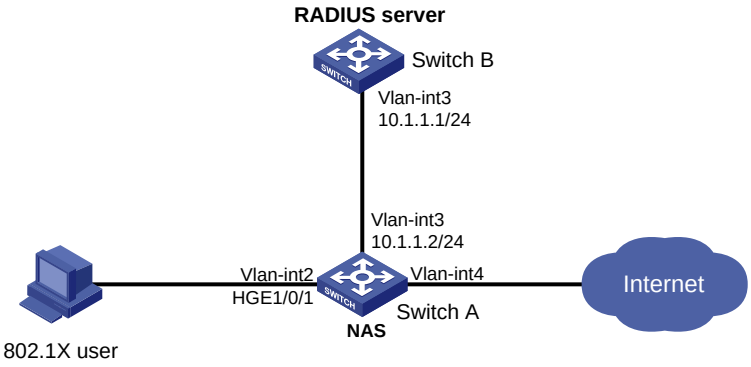
As shown in Figure 33, Switch B acts as the RADIUS server for authentication and authorization of 802.1X users connected to the NAS (Switch A).

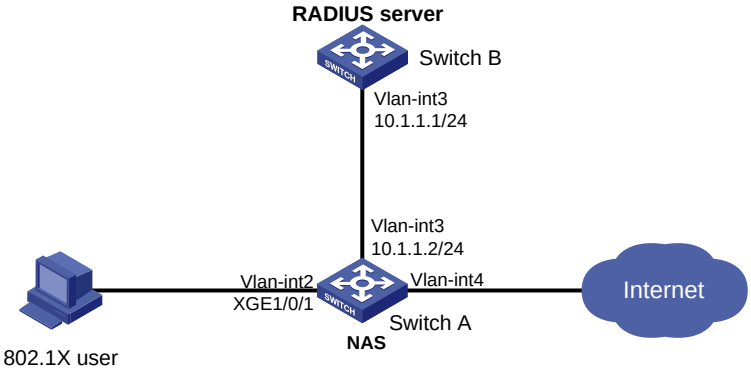
Configure the switches to meet the following requirements:

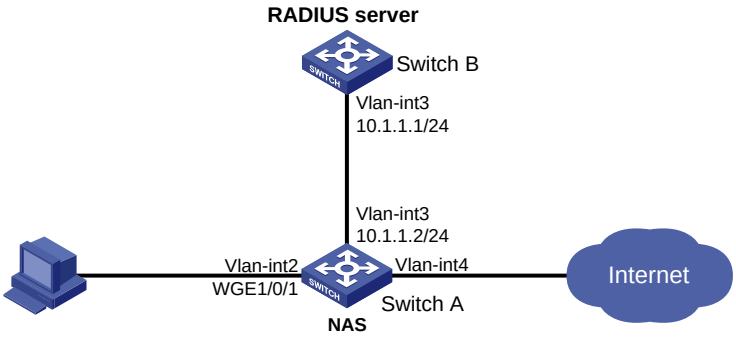
- Perform 802.1X user authentication on GigabitEthernet 1/0/1 of the NAS.
- The shared key is **expert** and the authentication port is 1812.
- Exclude domain names from the usernames sent to the RADIUS server.
- The user name for 802.1X authentication is **dot1x**.
- After the user passes authentication, the RADIUS server authorizes VLAN 4 to the NAS port that the user is connecting to.

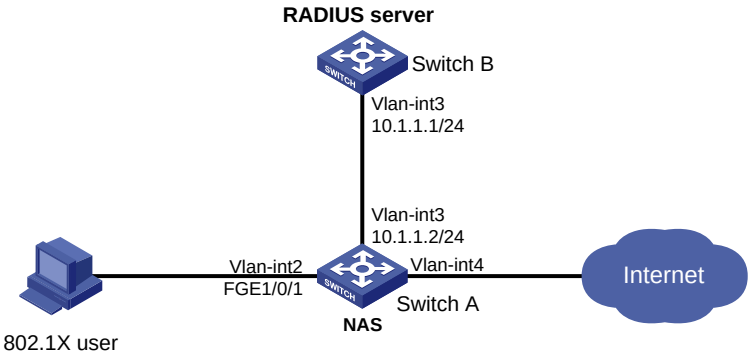
Figure 33 Network diagram

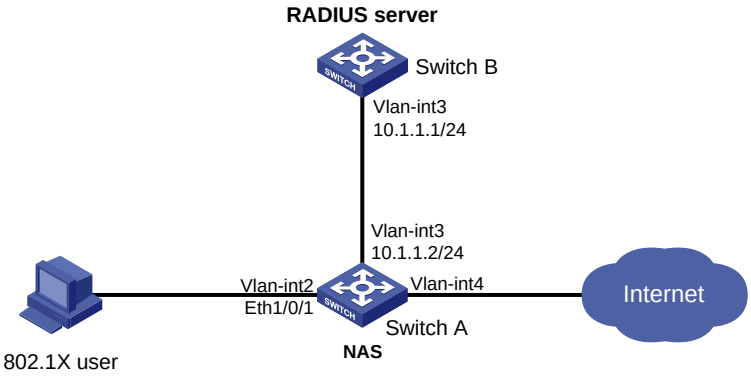


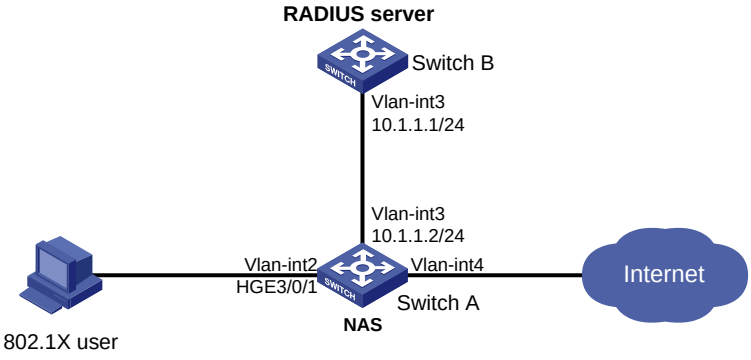

RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
HGE1/0/1
Vlan-int4
Internet
Switch A
NAS
802.1X user

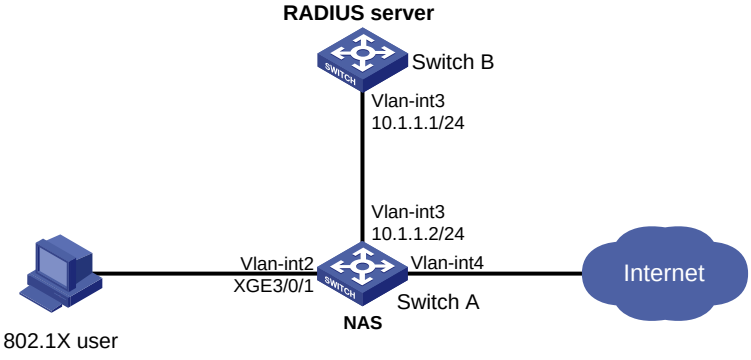

RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
XGE1/0/1
Vlan-int4
Internet
Switch A
NAS
802.1X user

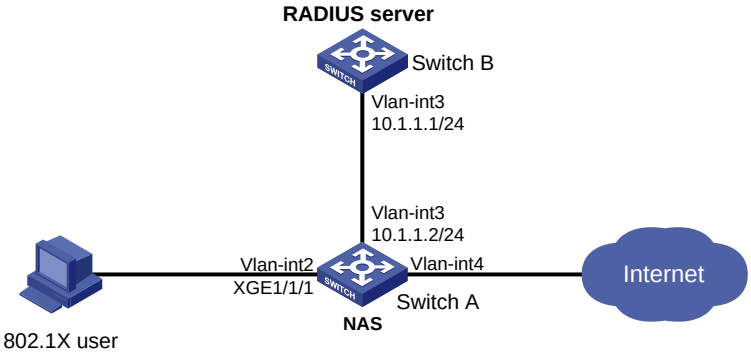

RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
WGE1/0/1
Vlan-int4
Internet
Switch A
NAS

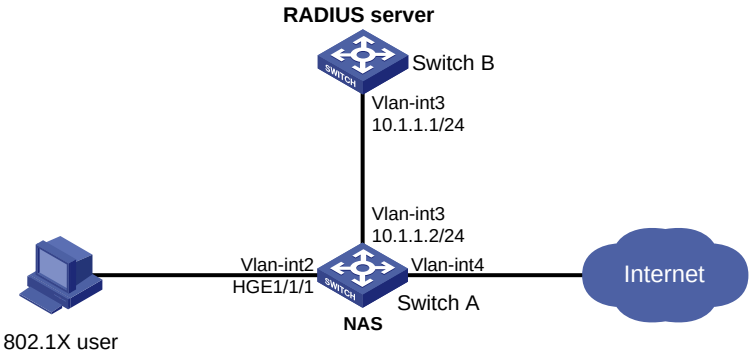

RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
FGE1/0/1
Vlan-int4
Internet
Switch A
NAS
802.1X user

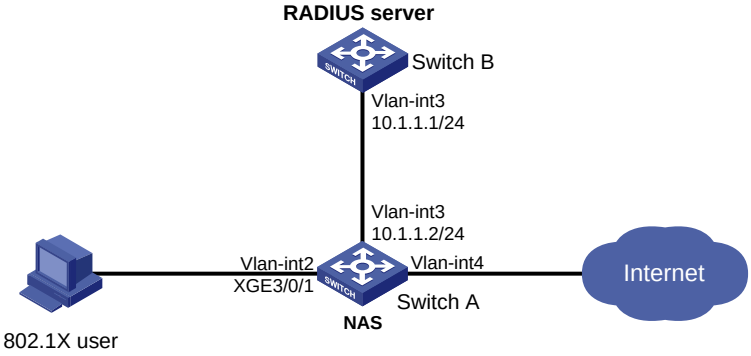

RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
Eth1/0/1
Vlan-int4
Internet
Switch A
NAS
802.1X user


RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
HGE3/0/1
Vlan-int4
Internet
Switch A
NAS
802.1X user


RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
XGE3/0/1
Vlan-int4
Internet
Switch A
NAS
802.1X user


RADIUS server
Switch B
Vlan-int3
10.1.1.1/24
Vlan-int3
10.1.1.2/24
Vlan-int2
XGE1/1/1
Vlan-int4
Internet
Switch A
NAS
802.1X user





Procedure

1. Configure the NAS:

   a. Configure a RADIUS scheme:

   # Configure a RADIUS scheme named **rad** and enter RADIUS scheme view.

   ```
   <SwitchA> system-view
   [SwitchA] radius scheme rad
   ```

   # Specify the primary authentication server with IP address 10.1.1.1 and set the shared key to **expert** in plaintext form.

   ```
   [SwitchA-radius-rad] primary authentication 10.1.1.1 key simple expert
   ```

   # Exclude domain names from the usernames sent to the RADIUS server.

   ```
   [SwitchA-radius-rad] user-name-format without-domain
   [SwitchA-radius-rad] quit
   ```

   b. Configure an ISP domain:

   # Create an ISP domain named **bbb** and enter ISP domain view.

   ```
   [SwitchA] domain bbb
   ```

   # Configure the ISP domain to use RADIUS scheme **rad** for authentication and authorization of LAN users and not to perform accounting for LAN users.

   ```
   [SwitchA-isp-bbb] authentication lan-access radius-scheme rad
   [SwitchA-isp-bbb] authorization lan-access radius-scheme rad
   [SwitchA-isp-bbb] accounting lan-access none
   [SwitchA-isp-bbb] quit
   ```

   c. Configure 802.1X authentication:

   # Enable 802.1X for GigabitEthernet 1/0/1.

   ```
   [SwitchA] interface gigabitethernet 1/0/1
   [SwitchA-GigabitEthernet1/0/1] dot1x
   ```

Specify **bbb** as the mandatory authentication domain for 802.1X users on the interface.

```
[SwitchA-GigabitEthernet1/0/1] dot1x mandatory-domain bbb
[SwitchA-GigabitEthernet1/0/1] quit
```

Enable 802.1X globally.

```
[SwitchA] dot1x
```

2. Configure the RADIUS server:

Create a network access user named **dot1x**.

```
<SwitchB> system-view
[SwitchB] local-user dot1x class network
```

Configure the password as **123456** in plaintext form.

```
[SwitchB-luser-network-dot1x] password simple 123456
```

Configure VLAN 4 as the authorization VLAN.

```
[SwitchB-luser-network-dot1x] authorization-attribute vlan 4
[SwitchB-luser-network-dot1x] quit
```

Configure the IP address of the RADIUS client as 10.1.1.2 and the shared key as **expert** in plaintext form.

```
[SwitchB] radius-server client ip 10.1.1.2 key simple expert
```

Activate the RADIUS server configuration.

```
[SwitchB] radius-server activate
```

Verifying the configuration

1. On the RADIUS server, display the activated RADIUS clients and users.

```
[SwitchB] display radius-server active-client
Total 1 RADIUS clients.
Client IP: 10.1.1.2
[SwitchB] display radius-server active-user dot1x
Total 1 RADIUS users matched.
Username: dot1x
  Description: Not configured
  Authorization attributes:
    VLAN ID: 4
    ACL number: Not configured
  Validity period:
    Expiration time: Not configured
```

2. On the host, use account **dot1x** for 802.1X authentication.

   If the host runs the Windows built-in 802.1X client, configure the network connection properties as follows:

   a. Click the **Authentication** tab of the properties window.

   b. Select the **Enable IEEE 802.1X authentication for this network** option.

   c. Select MD5 challenge as the EAP type.

   d. Click **OK**.

   If the host runs the iNode client, no advanced authentication options are required. The user passes authentication after entering the correct user name and password on the authentication page or the iNode client.

IMPORTANT:

Make sure the client can update its IP address to access the resources in the authorized VLAN after passing authentication.

3. On the NAS, verify that the RADIUS server assigns the port to VLAN 4 after the user passes authentication. (Details not shown.)
4. On the NAS, display online 802.1X user information.

```
[SwitchA] display dot1x connection
```

Troubleshooting AAA

RADIUS authentication failure

Symptom

User authentication always fails.

Analysis

Possible reasons include:

- A communication failure exists between the NAS and the RADIUS server.
- The username is not in the *userid@isp-name* format, or the ISP domain is not correctly configured on the NAS.
- The user is not configured on the RADIUS server.
- The password entered by the user is incorrect.
- The RADIUS server and the NAS are configured with different shared keys.

Solution

To resolve the problem:

1. Verify the following items:
   - The NAS and the RADIUS server can ping each other.
   - The username is in the *userid@isp-name* format and the ISP domain is correctly configured on the NAS.
   - The user is configured on the RADIUS server.
   - The correct password is entered.
   - The same shared key is configured on both the RADIUS server and the NAS.
2. If the problem persists, contact H3C Support.

RADIUS packet delivery failure

Symptom

RADIUS packets cannot reach the RADIUS server.

Analysis

Possible reasons include:

- A communication failure exists between the NAS and the RADIUS server.
- The NAS is not configured with the IP address of the RADIUS server.
- The authentication and accounting UDP ports configured on the NAS are incorrect.
- The RADIUS server's authentication and accounting port numbers are being used by other applications.

Solution

To resolve the problem:

1. Verify the following items:
   - The link between the NAS and the RADIUS server works well at both the physical and data link layers.
   - The IP address of the RADIUS server is correctly configured on the NAS.
   - The authentication and accounting UDP port numbers configured on the NAS are the same as those of the RADIUS server.
   - The RADIUS server's authentication and accounting port numbers are available.
2. If the problem persists, contact H3C Support.

RADIUS accounting error

Symptom

A user is authenticated and authorized, but accounting for the user is not normal.

Analysis

The accounting server configuration on the NAS is not correct. Possible reasons include:

- The accounting port number configured on the NAS is incorrect.
- The accounting server IP address configured on the NAS is incorrect. For example, the NAS is configured to use a single server to provide authentication, authorization, and accounting services, but in fact the services are provided by different servers.

Solution

To resolve the problem:

1. Verify the following items:
   - The accounting port number is correctly configured.
   - The accounting server IP address is correctly configured on the NAS.
2. If the problem persists, contact H3C Support.

Troubleshooting HWTACACS

Similar to RADIUS troubleshooting. See "RADIUS authentication failure," "RADIUS packet delivery failure," and "RADIUS accounting error."

LDAP authentication failure

Symptom

User authentication fails.

Analysis

Possible reasons include:

- A communication failure exists between the NAS and the LDAP server.
- The LDAP server IP address or port number configured on the NAS is not correct.
- The username is not in the *userid@isp-name* format, or the ISP domain is not correctly configured on the NAS.
- The user is not configured on the LDAP server.
- The password entered by the user is incorrect.
- The administrator DN or password is not configured.

- Some user attributes (for example, the username attribute) configured on the NAS are not consistent with those configured on the server.
- No user search base DN is specified for the LDAP scheme.

Solution

To resolve the problem:

1. Verify the following items:
   - The NAS and the LDAP server can ping each other.
   - The IP address and port number of the LDAP server configured on the NAS match those of the server.
   - The username is in the correct format and the ISP domain for the user authentication is correctly configured on the NAS.
   - The user is configured on the LDAP server.
   - The correct password is entered.
   - The administrator DN and the administrator password are correctly configured.
   - The user attributes (for example, the username attribute) configured on the NAS are consistent with those configured on the LDAP server.
   - The user search base DN for authentication is specified.
2. If the problem persists, contact H3C Support.

Appendixes

Appendix A Commonly used RADIUS attributes

Commonly used RADIUS attributes are defined in RFC 2865, RFC 2866, RFC 2867, and RFC 2868.

Table 5 Commonly used RADIUS attributes

No.	Attribute	No.	Attribute
1	User-Name	45	Acct-Authentic
2	User-Password	46	Acct-Session-Time
3	CHAP-Password	47	Acct-Input-Packets
4	NAS-IP-Address	48	Acct-Output-Packets
5	NAS-Port	49	Acct-Terminate-Cause
6	Service-Type	50	Acct-Multi-Session-Id
7	Framed-Protocol	51	Acct-Link-Count
8	Framed-IP-Address	52	Acct-Input-Gigawords
9	Framed-IP-Netmask	53	Acct-Output-Gigawords
10	Framed-Routing	54	(unassigned)
11	Filter-ID	55	Event-Timestamp
12	Framed-MTU	56-59	(unassigned)
13	Framed-Compression	60	CHAP-Challenge
14	Login-IP-Host	61	NAS-Port-Type
15	Login-Service	62	Port-Limit

No.	Attribute	No.	Attribute
16	Login-TCP-Port	63	Login-LAT-Port
17	(unassigned)	64	Tunnel-Type
18	Reply-Message	65	Tunnel-Medium-Type
19	Callback-Number	66	Tunnel-Client-Endpoint
20	Callback-ID	67	Tunnel-Server-Endpoint
21	(unassigned)	68	Acct-Tunnel-Connection
22	Framed-Route	69	Tunnel-Password
23	Framed-IPX-Network	70	ARAP-Password
24	State	71	ARAP-Features
25	Class	72	ARAP-Zone-Access
26	Vendor-Specific	73	ARAP-Security
27	Session-Timeout	74	ARAP-Security-Data
28	Idle-Timeout	75	Password-Retry
29	Termination-Action	76	Prompt
30	Called-Station-Id	77	Connect-Info
31	Calling-Station-Id	78	Configuration-Token
32	NAS-Identifier	79	EAP-Message
33	Proxy-State	80	Message-Authenticator
34	Login-LAT-Service	81	Tunnel-Private-Group-ID
35	Login-LAT-Node	82	Tunnel-Assignment-id
36	Login-LAT-Group	83	Tunnel-Preference
37	Framed-AppleTalk-Link	84	ARAP-Challenge-Response
38	Framed-AppleTalk-Network	85	Acct-Interim-Interval
39	Framed-AppleTalk-Zone	86	Acct-Tunnel-Packets-Lost
40	Acct-Status-Type	87	NAS-Port-Id
41	Acct-Delay-Time	88	Framed-Pool
42	Acct-Input-Octets	89	(unassigned)
43	Acct-Output-Octets	90	Tunnel-Client-Auth-id
44	Acct-Session-Id	91	Tunnel-Server-Auth-id

Appendix B Descriptions for commonly used standard RADIUS attributes

No.	Attribute	Description
1	User-Name	Name of the user to be authenticated.
2	User-Password	User password for PAP authentication, only present in Access-Request packets when PAP authentication is used.

No.	Attribute	Description
3	CHAP-Password	Digest of the user password for CHAP authentication, only present in Access-Request packets when CHAP authentication is used.
4	NAS-IP-Address	IP address for the server to use to identify the client. Typically, a client is identified by the IP address of its access interface. This attribute is only present in Access-Request packets.
5	NAS-Port	Physical port of the NAS that the user accesses.
6	Service-Type	Type of service that the user has requested or type of service to be provided.
7	Framed-Protocol	Encapsulation protocol for framed access.
8	Framed-IP-Address	IP address assigned to the user.
11	Filter-ID	Name of the filter list. If the name starts with a digit, it indicates an ACL number.
12	Framed-MTU	MTU for the data link between the user and NAS. For example, this attribute can be used to define the maximum size of EAP packets allowed to be processed in 802.1X EAP authentication.
14	Login-IP-Host	IP address of the NAS interface that the user accesses.
15	Login-Service	Type of service that the user uses for login.
18	Reply-Message	Text to be displayed to the user, which can be used by the server to communicate information, for example, the cause of the authentication failure.
26	Vendor-Specific	Vendor-specific proprietary attribute. A packet can contain one or more proprietary attributes, each of which can contain one or more subattributes.
27	Session-Timeout	Maximum service duration for the user before termination of the session.
28	Idle-Timeout	Maximum idle time permitted for the user before termination of the session.
31	Calling-Station-Id	User identification that the NAS sends to the server. For the LAN access service provided by an H3C device, this attribute includes the MAC address of the user.
32	NAS-Identifier	Identification that the NAS uses to identify itself to the RADIUS server.
40	Acct-Status-Type	Type of the Accounting-Request packet. Possible values include: • **1**—Start. • **2**—Stop. • **3**—Interim-Update. • **4**—Reset-Charge. • **7**—Accounting-On. (Defined in the 3rd Generation Partnership Project.) • **8**—Accounting-Off. (Defined in the 3rd Generation Partnership Project.) • **9 to 14**—Reserved for tunnel accounting. • **15**—Reserved for failed.
45	Acct-Authentic	Authentication method used by the user. Possible values include: • **1**—RADIUS. • **2**—Local. • **3**—Remote.

No.	Attribute	Description
60	CHAP-Challenge	CHAP challenge generated by the NAS for MD5 calculation during CHAP authentication.
61	NAS-Port-Type	Type of the physical port of the NAS that is authenticating the user. Possible values include: • **15**—Ethernet. • **16**—Any type of ADSL. • **17**—Cable. (With cable for cable TV.) • **19**—WLAN-IEEE 802.11. • **201**—VLAN. • **202**—ATM. If the port is an ATM or Ethernet one and VLANs are implemented on it, the value of this attribute is 201.
64	Tunnel-Type	Tunneling protocols used. The value **13** represents VLAN. If the value is 13, the device interprets the Tunnel-Type, Tunnel-Medium-Type, and Tunnel-Private-Group-ID attributes as attributes to assign VLANs.
65	Tunnel-Medium-Type	Transport medium type to use for creating a tunnel. For VLAN assignment, the value must be **6** to indicate the 802 media plus Ethernet.
79	EAP-Message	Used to encapsulate EAP packets to allow RADIUS to support EAP authentication.
80	Message-Authenticator	Used for authentication and verification of authentication packets to prevent spoofing Access-Requests. This attribute is present when EAP authentication is used.
81	Tunnel-Private-Group-ID	Group ID for a tunnel session. To assign VLANs, the NAS conveys VLAN IDs by using this attribute.
87	NAS-Port-Id	String for describing the port of the NAS that is authenticating the user.
168	Framed-IPv6-Address	Server-assigned IPv6 address for the NAS to assign to the host. The address must be unique.

Appendix C RADIUS subattributes (vendor ID 25506)

Table 6 lists all RADIUS subattributes with a vendor ID of 25506. Support for these subattributes depends on the device model.

Table 6 RADIUS subattributes (vendor ID 25506)

No.	Subattribute	Description
1	Input-Peak-Rate	Peak rate in the direction from the user to the NAS, in bps.
2	Input-Average-Rate	Average rate in the direction from the user to the NAS, in bps.
3	Input-Basic-Rate	Basic rate in the direction from the user to the NAS, in bps.
4	Output-Peak-Rate	Peak rate in the direction from the NAS to the user, in bps.
5	Output-Average-Rate	Average rate in the direction from the NAS to the user, in bps.
6	Output-Basic-Rate	Basic rate in the direction from the NAS to the user, in bps.
15	Remanent_Volume	Total amount of data available for the connection, in different units for different server types.

No.	Subattribute	Description
17	ISP-ID	ISP domain where the user obtains authorization information.
20	Command	Operation for the session, used for session control. Possible values include: • **1**—Trigger-Request. • **2**—Terminate-Request. • **3**—SetPolicy. • **4**—Result. • **5**—PortalClear.
25	Result_Code	Result of the Trigger-Request or SetPolicy operation, zero for success and any other value for failure.
26	Connect_ID	Index of the user connection.
27	PortalURL	PADM redirect URL assigned to PPPoE users.
28	Ftp_Directory	FTP, SFTP, or SCP user working directory. When the RADIUS client acts as the FTP, SFTP, or SCP server, this attribute is used to set the working directory for an FTP, SFTP, or SCP user on the RADIUS client.
29	Exec_Privilege	EXEC user priority.
32	NAT-IP-Address	Public IP address assigned to the user when the source IP address and port are translated.
33	NAT-Start-Port	Start port number of the port range assigned to the user when the source IP address and port are translated.
34	NAT-End-Port	End port number of the port range assigned to the user when the source IP address and port are translated.
59	NAS_Startup_Timestamp	Startup time of the NAS in seconds, which is represented by the time elapsed after 00:00:00 on Jan. 1, 1970 (UTC).
60	Ip_Host_Addr	User IP address and MAC address included in authentication and accounting requests, in the format A.B.C.D hh:hh:hh:hh:hh:hh. A space is required between the IP address and the MAC address.
61	User_Notify	Information that must be sent from the server to the client transparently.
62	User_HeartBeat	Hash value assigned after an 802.1X user passes authentication, which is a 32-byte string. This attribute is stored in the user list on the NAS and verifies the handshake packets from the 802.1X user. This attribute only exists in Access-Accept and Accounting-Request packets.
98	Multicast_Receive_Group	IP address of the multicast group that the user's host joins as a receiver. This subattribute can appear multiple times in a multicast packet to indicate that the user belongs to multiple multicast groups.
100	IP6_Multicast_Receive_Group	IPv6 address of the multicast group that the user's host joins as a receiver. This subattribute can appear multiple times in a multicast packet to indicate that the user belongs to multiple multicast groups.
101	MLD-Access-Limit	Maximum number of MLD multicast groups that the user can join concurrently.
102	local-name	L2TP local tunnel name.
103	IGMP-Access-Limit	Maximum number of IGMP multicast groups that the user can join concurrently.

No.	Subattribute	Description
104	VPN-Instance	MPLS L3VPN instance to which a user belongs.
105	ANCP-Profile	ANCP profile name.
135	Client-Primary-DNS	IP address of the primary DNS server.
136	Client-Secondary-DNS	IP address of the secondary DNS server.
140	User_Group	User groups assigned after the user passes authentication. Typically, a user can belong to only one user group.
144	Acct_IPv6_Input_Octets	Bytes of IPv6 packets in the inbound direction. The measurement unit depends on the configuration on the device.
145	Acct_IPv6_Output_Octets	Bytes of IPv6 packets in the outbound direction. The measurement unit depends on the configuration on the device.
146	Acct_IPv6_Input_Packets	Number of IPv6 packets in the inbound direction. The measurement unit depends on the configuration on the device.
147	Acct_IPv6_Output_Packets	Number of IPv6 packets in the outbound direction. The measurement unit depends on the configuration on the device.
148	Acct_IPv6_Input_Gigawords	Bytes of IPv6 packets in the inbound direction. The measurement unit is 4G bytes.
149	Acct_IPv6_Output_Gigawords	Bytes of IPv6 packets in the outbound direction. The measurement unit is 4G bytes.
155	User-Roles	List of space-separated user roles.
182	Microsegment-Id	Microsegment ID.
210	Av-Pair	User-defined attribute pair. Available attribute pairs include: • Server-assigned voice VLAN in the format of device-traffic-class=voice. • Server-assigned user role in the format of shell:role=xxx. • Server-assigned ACL in the format of url-redirect-acl=xxx. • Server-assigned IPv4 Web redirect URL in the format of url-redirect=xxx. • Server-assigned IPv6 Web redirect URL in the format of urlipv6-redirect=xxx. • Server-deployed command to reboot a port, in the format of subscriber:command=bounce-host-port. • Server-assigned port shutdown duration in the format of bounce:seconds=xxx. • Server-deployed command to shut down a port, in the format of subscriber:command=disable-host-port. • Server-assigned blackhole MAC address attribute in the format of mac:block-mac=xxx. • Server-assigned MAC authentication offline detect timer (in seconds) in the format of mac-authentication: offline-detect-time=xxx. Value 0 of xxx indicates that MAC authentication offline detection is disabled. • Server-assigned MAC authentication offline detection flag in the format of mac-authentication: offline-detect-check=x. x has the following values: ○ **0**—The device does not search for the ARP snooping entry or ND snooping entry of the MAC address. ○ **1**—The device searches for the ARP snooping entry or ND snooping entry of the MAC address. • Server-assigned dynamic ACL. For more information about

No.	Subattribute	Description
		the format of this attribute, see "Appendix D Format of dynamic authorization ACLs." Support for this attribute in 802.1X authentication and MAC authentication depends on the device model. If the server assigns a user both this attribute and the Filter-ID attribute, the device will ignore this attribute. The device does not support using CoA messages to change the content assigned by this attribute or assign another ACL to the user.
230	NAS-Port-Name	Interface through which the user is connected to the NAS.
246	Auth_Detail_Result	Accounting details. The server sends Access-Accept packets with subattributes 246 and 250 in the following situations: • **1**—The subscriber charge is overdue. The subscriber is allowed to access network resources in the whitelist. If the subscriber accesses other network resources, the device redirects it to the URL specified by subattribute 250. • **2**—The broadband lease of the subscriber expires. The device redirects the subscriber to the URL specified by subattribute 250 when the subscriber requests to access webpages for the first time.
247	Input-Committed-Burst-Size	Committed burst size from the user to the NAS, in bits. The total length cannot exceed 4 bytes for this field. This subattribute must be assigned together with the Input-Average-Rate attribute.
248	Output-Committed-Burst-Size	Committed burst size from the NAS to the user, in bits. The total length cannot exceed 4 bytes for this field. This subattribute must be assigned together with the Output-Average-Rate attribute.
249	authentication-type	Authentication type. The value can be: • **1**—Intranet access authentication. • **2**—Internet access authentication. If the packet does not contain this subattribute, common authentication applies.
250	WEB-URL	Redirect URL for users. In single-stack scenarios, use this attribute to assign the authorized IPv4 URL or IPv6 URL. In dual-stack scenarios, use this attribute to assign the authorized IPv4 URL and use user-defined attribute pair **urlipv6-redirect** to assign the authorized IPv6 URL.
251	Subscriber-ID	Family plan ID.
252	Subscriber-Profile	QoS policy name for the family plan of the subscriber.
255	Product_ID	Product name.

Appendix D Format of dynamic authorization ACLs

Vendor ID 25506

The server might assign a dynamic authorization ACL that contains multiple rules to a user in different ways:

- Assign only one Av-Pair subattribute to the user. In this subattribute, the dynamic ACL contains multiple rules separated by question marks (?).

- Assign multiple Av-Pair subattributes to the user. All the subattributes contain the same dynamic ACL and a different rule. Support for this method depends on the server model.

The format of a dynamic ACL rule is as follows:

aclrule?same?*acl-name*?*acl-type*?*ver-type*?*rule-id*?protocol=*protocol-type*?counting?dst-ip=*ip-addr* ?src-ip=*ip-addr*?dst-port=*port-value*?src-port=*port-value*?action=*action-type*

The fields in the rule are described in Table 7. The following is an example of a dynamic ACL rule:

aclrule?same?test?1?1?1?protocol=3?counting?dst-ip=1.1.1.1/1.1.1.1?src-ip=1.1.1.1/0?dst-port=1 .2000?src-port=5.2000-3000?action=1

Table 7 Fields in a dynamic ACL rule

Field	Description	Remarks
aclrule	Indicates that the following part is information about a dynamic ACL rule.	Required
same	Indicates that the current user will inherit the dynamic ACL rules that have been successfully assigned to another authenticated user. If the server assigns one user the same dynamic ACL as another user but the rules are different for the two users, the device applies the dynamic ACL rules of the user that comes online first to the other user.	Required
acl-name	ACL name, a case-insensitive string of 1 to 63 characters. The ACL name must begin with a letter and it cannot be **all** or the same as an existing static ACL on the device.	Required
acl-type	ACL type. **1** indicates an advanced ACL. In the current software version, only advanced ACLs are supported.	Required
ver-type	IP protocol type: • **1**—IPv4. • **2**—IPv6.	Required
rule-id	ACL rule number, in the range of 0 to 65534.	Required
protocol-type	Protocol type: • **1**—IP. • **2**—ICMP. • **3**—TCP. • **4**—UDP. • **5**—ICMPv6. • **6**—IPv6.	Optional
counting	Indicates that rule match statistics is enabled. If a rule does not include this field, rule match statistics is disabled for the rule.	Optional
ip-addr	IP address information. An IPv4 address must be in the format of X.X.X.X/X.X.X.X or X.X.X.X/0. Non-host IPv4 addresses can only be in the X.X.X.X/X.X.X.X format. An IPv6 address must be in the format of X:X::X:X/M. For example, 1.1.1.1/1.1.1.1, 1.1.1.1/0, or 3::3/128. If the value is **any**, the rule matches any IP address.	Optional
port-value	TCP or UDP port information, in the *X.YYY* format. • ***X***—Operator. ○ **1**—Equal to. ○ **2**—Greater than. ○ **3**—Smaller than. ○ **4**—Not equal to.	Optional

Field	Description	Remarks
	○ **5**—In the range of. • ***YYY***—Port number information. For example, 1.3000 and 5.2000-3000.	
action-type	Action type: • **1**—Deny. • **2**—Permit.	Optional

The following restrictions apply to dynamic authorization ACLs:

- For the former six fields (aclrule?same?*acl-name*?*acl-type*?*ver-type*?*rule-id*) of a dynamic ACL rule, their positions are fixed. The device cannot interpret a dynamic ACL rule if the positions of the six fields are changed. For the other fields, position changes are allowed.
- The settings for all the fields in the rules must meet the configuration logic of ACL rules on the device so the device can correctly interpret the rules.
- All dynamic ACL rules in one authorization must belong to the same ACL name.
- A dynamic ACL must have rules and the format of the rules must be valid.

Vendor ID 9

The server might assign a dynamic ACL to a user by using the CISCO-AV-Pair subattribute (vendor ID 9). The formats of the subattribute are as follows:

- **CiscoSecure-Defined-ACL=#ACSACL#-IP-***name***-***number*

  The *name* argument represents the ACL name and the *number* argument represents the version number.

  For example: **CiscoSecure-Defined-ACL=#ACSACL#-IP-UACL-0**.

- **ip:inacl#***number***=***rule*

  The *number* argument represents the ACL rule number and the *rule* argument represents the rule content.

  For example: **ip:inacl#1=permit udp any any eq 21862**.

Currently, only some device models support the CISCO-AV-Pair subattribute for 802.1X and MAC authentication. If the RADIUS server assigns dynamic ACLs to a user by using both the CISCO-AV-Pair (vendor ID 9) and Av-Pair (vendor ID 25506) subattributes, the device ignores the CISCO-AV-Pair subattribute.

The server assigns a dynamic ACL to a user as follows:

1. The server assigns a dynamic ACL name to a user by using the CISCO-AV-Pair subattribute in the format of **CiscoSecure-Defined-ACL=#ACSACL#-IP-***name***-***number* after the user passes authentication.
2. The device uses the dynamic ACL name as the username of the user to initiate authentication again.
3. The server assigns one or multiple ACL rules to the user by using the CISCO-AV-Pair subattribute in the format of **ip:inacl#***number***=***rule* after the user passes authentication.
4. The device parses the ACL rules received from the server, creates the ACL and its rules, and deploys the ACL and its rules to the user.

   In the current software version, the device can parse only the dynamic ACL rules in the following formats:

   { **deny** | **permit** } *protocol* [**destination** { *dest-address dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | **source** { *source-address source-wildcard* | **any** } | **source-port** *operator port1* [*port2*]]

The value for the *operator* argument can be **lt** (lower than), **gt** (greater than), **eq** (equal to), or **neq** (not equal to).

Contents

802.1X overview 3

About the 802.1X protocol 3
802.1X architecture 3
Controlled/uncontrolled port and port authorization status 3
Packet exchange methods 4
Packet formats 5
802.1X authentication procedures 7
802.1X authentication initiation 9
Access control methods 10
802.1X VLAN manipulation 10
Authorization VLAN 10
Guest VLAN 13
Auth-Fail VLAN 14
Critical VLAN 15
802.1X VSI manipulation 17
802.1X support for VXLANs 17
Authorization VSI 17
Guest VSI 18
Auth-Fail VSI 19
Critical VSI 20
802.1X microsegment assignment 22
Authorization microsegment 22
Critical microsegment 22
ACL assignment 23
Redirect URL assignment 24
802.1X support for critical profiles 24
Periodic 802.1X reauthentication 25
EAD assistant 25

Configuring 802.1X 26

Restrictions and guidelines: 802.1X configuration 26
802.1X tasks at a glance 28
Prerequisites for 802.1X 29
Enabling 802.1X 29
Enabling EAP relay or EAP termination 30
Setting the port authorization state 30
Specifying an access control method 31
Specifying a mandatory authentication domain on a port 31
Setting the 802.1X authentication timeout timers 32
Configuring 802.1X reauthentication 32
Setting the quiet timer 33
Configuring an 802.1X guest VLAN 34
Enabling 802.1X guest VLAN assignment delay 35
Configuring an 802.1X Auth-Fail VLAN 35
Configuring an 802.1X critical VLAN 36
Configuring an 802.1X guest VSI 37
Enabling 802.1X guest VSI assignment delay 38
Configuring an 802.1X Auth-Fail VSI 38
Configuring an 802.1X critical VSI 39
Configuring an 802.1X critical microsegment 39
Specifying a critical profile for 802.1X users on a port 40
Configuring an 802.1X access profile 41
Configuring 802.1X unauthenticated user aging 41
Sending EAP-Success packets on assignment of users to the 802.1X Auth-Fail VLAN or VSI 42
Sending EAP-Success packets on assignment of users to the 802.1X critical VLAN or VSI 43
Enabling 802.1X online user synchronization 43
Configuring the authentication trigger feature 44

Setting the maximum number of concurrent 802.1X users on a port ····45
Setting the alarm threshold and alarm clear threshold for 802.1X user access ratio ····45
Setting the maximum number of authentication request attempts ····46
Sending EAP-Success packets to 802.1X clients after successful authorization ····46
Discarding duplicate 802.1X EAPOL-Start requests ····47
Configuring online user handshake ····47
Configuring 802.1X offline detection ····49
Enabling online 802.1X users to escape from offline detection and stay online when no reachable RADIUS authentication servers are available ····50
Specifying supported domain name delimiters ····50
Removing the VLAN tags of 802.1X protocol packets sent out of a port ····51
Setting the maximum number of 802.1X authentication attempts for MAC authenticated users ····52
Enabling 802.1X user IP freezing ····52
Enabling generation of dynamic IPSG binding entries for 802.1X authenticated users ····53
Configuring 802.1X MAC address binding ····53
Configuring the EAD assistant feature ····54
Setting the maximum size of EAP-TLS fragments sent to the server ····56
Setting the unicast-trigger quiet timer ····56
Logging off 802.1X users ····57
Enabling 802.1X user logging ····57
Configuring SNMP notifications for 802.1X ····57
Display and maintenance commands for 802.1X ····58
802.1X authentication configuration examples ····59
Example: Configuring basic 802.1X authentication ····59
Example: Configuring 802.1X guest VLAN and authorization VLAN ····61
Example: Configuring 802.1X with ACL assignment ····64
Example: Configuring 802.1X guest VSI and authorization VSI configuration (port-based) ····66
Example: Configuring 802.1X guest VSI and authorization VSI (MAC-based) ····69
Example: Configuring 802.1X with EAD assistant (with DHCP relay agent) ····71
Example: Configuring 802.1X with EAD assistant (with DHCP server) ····73
Troubleshooting 802.1X ····76
EAD assistant Web redirection failure ····76

802.1X overview

About the 802.1X protocol

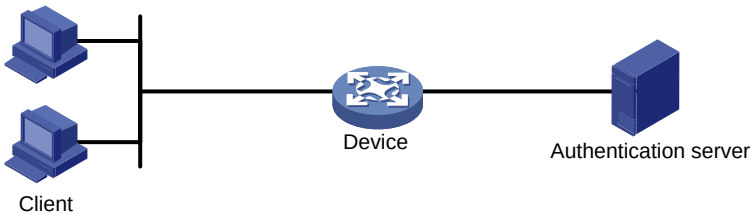
802.1X is a port-based network access control protocol widely used on Ethernet networks. The protocol controls network access by authenticating the devices connected to 802.1X-enabled LAN ports.

802.1X architecture

802.1X operates in the client/server model. As shown in Figure 1, 802.1X authentication includes the following entities:

- **Client (supplicant)**—A user terminal seeking access to the LAN. The terminal must have 802.1X software to authenticate to the access device.
- **Access device (authenticator)**—Authenticates the client to control access to the LAN. In a typical 802.1X environment, the access device uses an authentication server to perform authentication.
- **Authentication server**—Provides authentication services for the access device. The authentication server first authenticates 802.1X clients by using the data sent from the access device. Then, the server returns the authentication results to the access device to make access decisions. The authentication server is typically a RADIUS server. In a small LAN, you can use the access device as the authentication server.

Figure 1 802.1X architecture

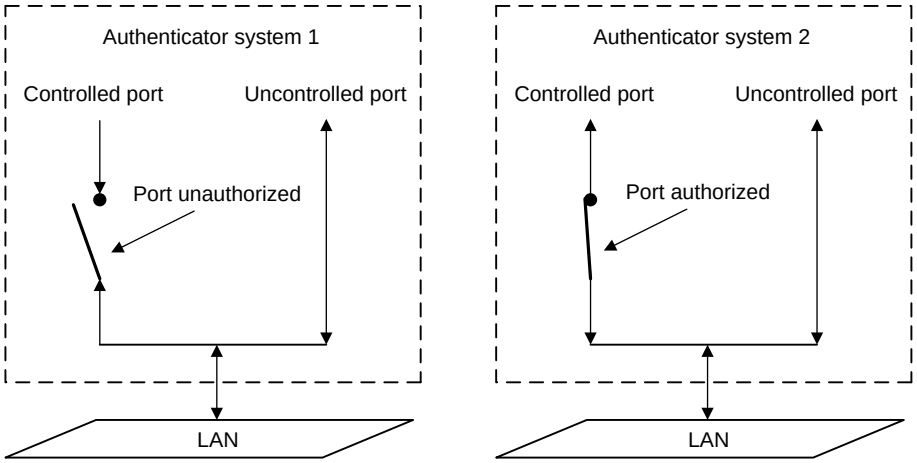


Controlled/uncontrolled port and port authorization status

802.1X defines two logical ports for the network access port: controlled port and uncontrolled port. Any packet arriving at the network access port is visible to both logical ports.

- **Uncontrolled port**—Is always open to receive and transmit authentication packets.
- **Controlled port**—Filters packets depending on the port state.
  - **Authorized state**—The controlled port is in authorized state when the client has passed authentication. The port allows traffic to pass through.
  - **Unauthorized state**—The port is in unauthorized state when the client has failed authentication. The port controls traffic by using one of the following methods:
    - Performs bidirectional traffic control to deny traffic to and from the client.
    - Performs unidirectional traffic control to deny traffic from the client. The device supports only unidirectional traffic control.

Figure 2 Authorization state of a controlled port



Packet exchange methods

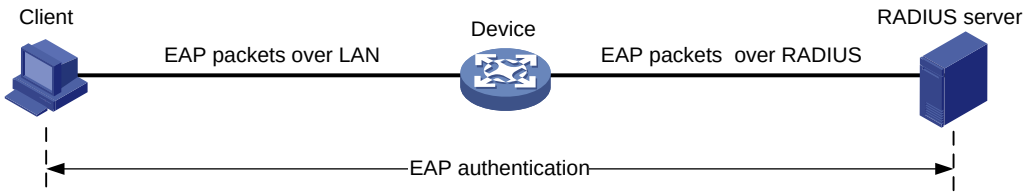
802.1X uses the Extensible Authentication Protocol (EAP) to transport authentication information for the client, the access device, and the authentication server. EAP is an authentication framework that uses the client/server model. The framework supports a variety of authentication methods, including MD5-Challenge, EAP-Transport Layer Security (EAP-TLS), and Protected EAP (PEAP).

802.1X defines EAP over LAN (EAPOL) for passing EAP packets between the client and the access device over a wired or wireless LAN. Between the access device and the authentication server, 802.1X delivers authentication information by either EAP relay or EAP termination.

EAP relay

EAP relay is defined in IEEE 802.1X. In this mode, the network device uses EAP over RADIUS (EAPOR) packets to send authentication information to the RADIUS server, as shown in Figure 3.

Figure 3 EAP relay



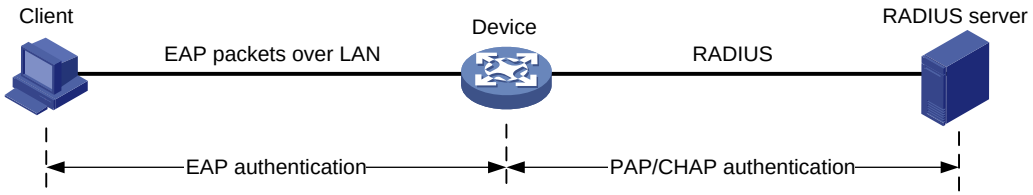
In EAP relay mode, the client must use the same authentication method as the RADIUS server. On the access device, you only need to use the **dot1x authentication-method eap** command to enable EAP relay.

EAP termination

As shown in Figure 4, the access device performs the following operations in EAP termination mode:

1. Terminates the EAP packets received from the client.
2. Encapsulates the client authentication information in standard RADIUS packets.
3. Uses PAP or CHAP to authenticate to the RADIUS server.

Figure 4 EAP termination



Comparing EAP relay and EAP termination

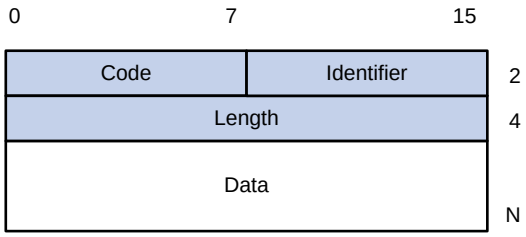
Packet exchange method	Benefits	Limitations
EAP relay	• Supports various EAP authentication methods. • The configuration and processing are simple on the access device.	The RADIUS server must support the EAP-Message and Message-Authenticator attributes, and the EAP authentication method used by the client.
EAP termination	Works with any RADIUS server that supports PAP or CHAP authentication.	• Supports only the following EAP authentication methods: ○ MD5-Challenge EAP authentication. ○ The username and password EAP authentication initiated by an iNode 802.1X client. • The processing is complex on the access device.

Packet formats

EAP packet format

Figure 5 shows the EAP packet format.

Figure 5 EAP packet format

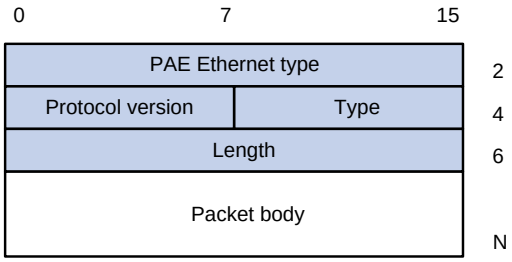


- **Code**—Type of the EAP packet. Options include Request (1), Response (2), Success (3), or Failure (4).
- **Identifier**—Used for matching Responses with Requests.
- **Length**—Length (in bytes) of the EAP packet. The EAP packet length is the sum of the Code, Identifier, Length, and Data fields.
- **Data**—Content of the EAP packet. This field appears only in a Request or Response EAP packet. The **Data** field contains the request type (or the response type) and the type data. Type 1 (Identity) and type 4 (MD5-Challenge) are two examples for the type field.

EAPOL packet format

Figure 6 shows the EAPOL packet format.

Figure 6 EAPOL packet format



- **PAE Ethernet type**—Protocol type. It takes the value 0x888E for EAPOL.
- **Protocol version**—The EAPOL protocol version used by the EAPOL packet sender.
- **Type**—Type of the EAPOL packet. Table 1 lists the types of EAPOL packets supported by the 802.1X implementation of the device.

  Table 1 Types of EAPOL packets

Value	Type	Description
0x00	EAP-Packet	The client and the access device uses EAP-Packets to transport authentication information.
0x01	EAPOL-Start	The client sends an EAPOL-Start message to initiate 802.1X authentication to the access device.
0x02	EAPOL-Logoff	The client sends an EAPOL-Logoff message to tell the access device that the client is logging off.

- **Length**—Data length in bytes, or length of the Packet body. If packet type is EAPOL-Start or EAPOL-Logoff, this field is set to 0, and no Packet body field follows.
- **Packet body**—Content of the packet. When the EAPOL packet type is EAP-Packet, the Packet body field contains an EAP packet.

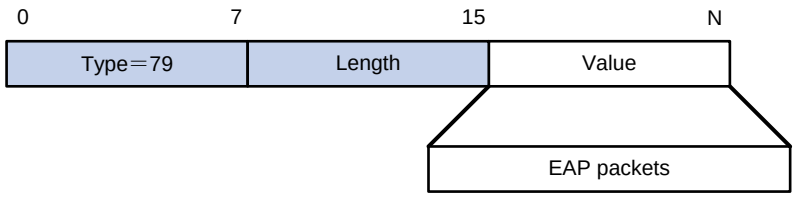
EAP over RADIUS

RADIUS adds two attributes, EAP-Message and Message-Authenticator, for supporting EAP authentication. For the RADIUS packet format, see "Configuring AAA."

- EAP-Message.

  RADIUS encapsulates EAP packets in the EAP-Message attribute, as shown in Figure 7. The Type field takes 79, and the Value field can be up to 253 bytes. If an EAP packet is longer than 253 bytes, RADIUS encapsulates it in multiple EAP-Message attributes.

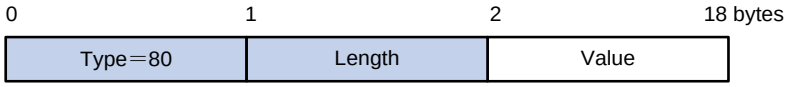
  Figure 7 EAP-Message attribute format



- Message-Authenticator.

  As shown in Figure 8, RADIUS includes the Message-Authenticator attribute in all packets that have an EAP-Message attribute to check their integrity. The packet receiver drops the packet if the calculated packet integrity checksum is different from the Message-Authenticator attribute value. The Message-Authenticator prevents EAP authentication packets from being tampered with during EAP authentication.

Figure 8 Message-Authenticator attribute format



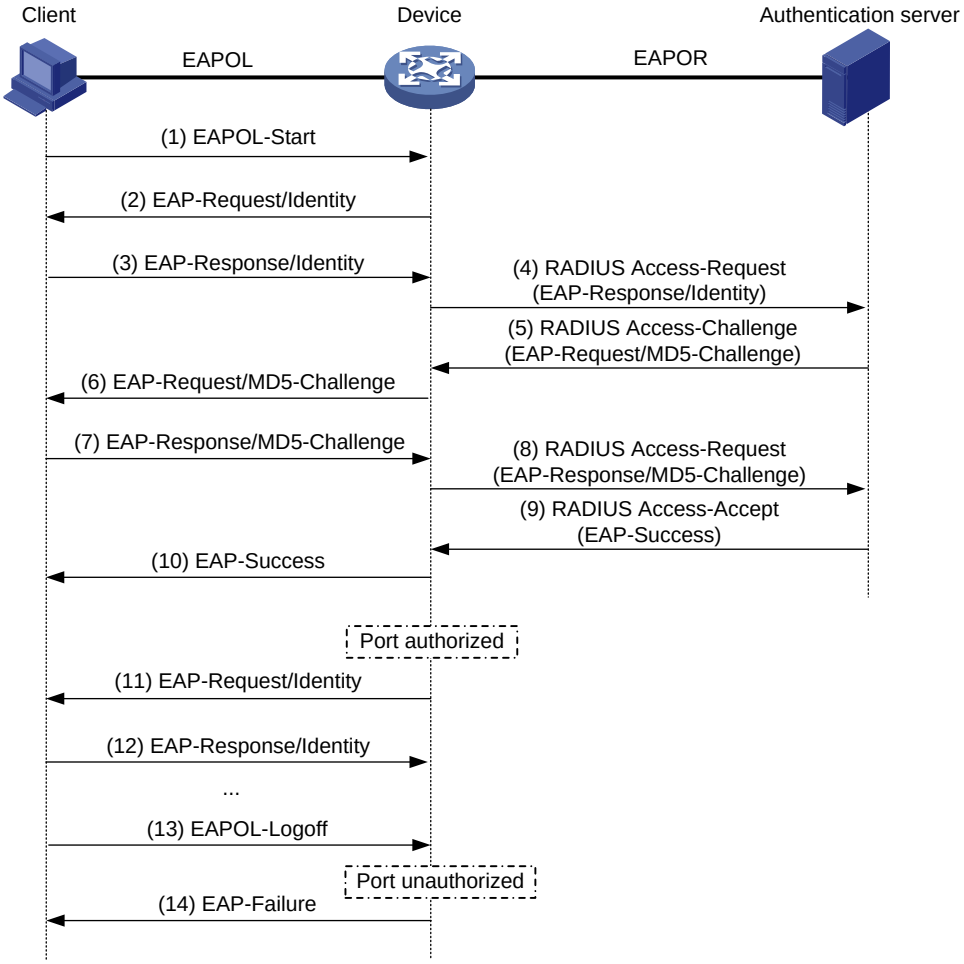
802.1X authentication procedures

802.1X authentication has two methods: EAP relay and EAP termination. You choose either mode depending on support of the RADIUS server for EAP packets and EAP authentication methods.

EAP relay

Figure 9 shows the basic 802.1X authentication procedure in EAP relay mode, assuming that MD5-Challenge EAP authentication is used.

Figure 9 802.1X authentication procedure in EAP relay mode



The following steps describe the 802.1X authentication procedure:

2. When a user launches the 802.1X client and enters a registered username and password, the 802.1X client sends an EAPOL-Start packet to the access device.
3. The access device responds with an EAP-Request/Identity packet to ask for the client username.
4. In response to the EAP-Request/Identity packet, the client sends the username in an EAP-Response/Identity packet to the access device.

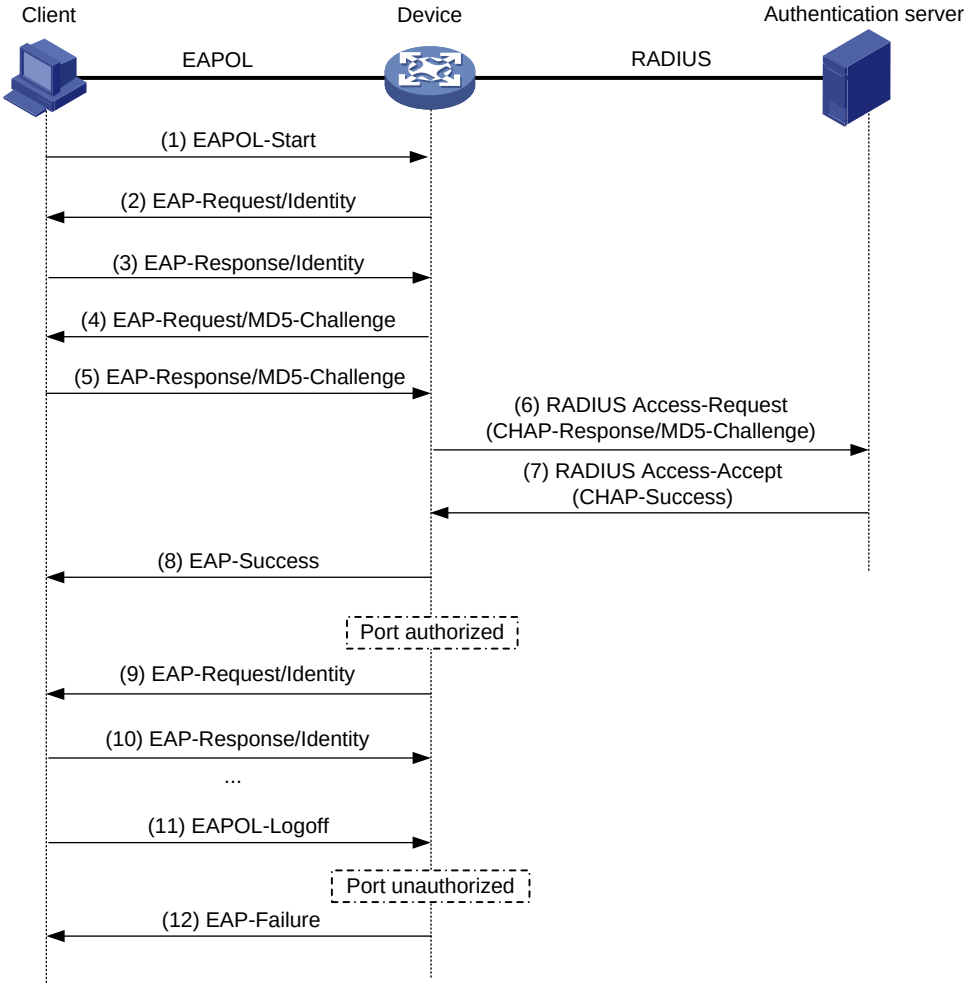
5. The access device relays the EAP-Response/Identity packet in a RADIUS Access-Request packet to the authentication server.
6. The authentication server uses the identity information in the RADIUS Access-Request to search its user database. If a matching entry is found, the server uses a randomly generated challenge (EAP-Request/MD5-Challenge) to encrypt the password in the entry. Then, the server sends the challenge in a RADIUS Access-Challenge packet to the access device.
7. The access device transmits the EAP-Request/MD5-Challenge packet to the client.
8. The client uses the received challenge to encrypt the password, and sends the encrypted password in an EAP-Response/MD5-Challenge packet to the access device.
9. The access device relays the EAP-Response/MD5-Challenge packet in a RADIUS Access-Request packet to the authentication server.
10. The authentication server compares the received encrypted password with the encrypted password it generated at step 5. If the two passwords are identical, the server considers the client valid and sends a RADIUS Access-Accept packet to the access device.
11. Upon receiving the RADIUS Access-Accept packet, the access device performs the following operations:
    a. Sends an EAP-Success packet to the client.
    b. Sets the controlled port in authorized state.

    The client can access the network.
12. After the client comes online, the access device periodically sends handshake requests to check whether the client is still online. By default, if two consecutive handshake attempts fail, the device logs off the client.
13. Upon receiving a handshake request, the client returns a response. If the client fails to return a response after a number of consecutive handshake attempts (two by default), the access device logs off the client. This handshake mechanism enables timely release of the network resources used by 802.1X users that have abnormally gone offline.
14. The client can also send an EAPOL-Logoff packet to ask the access device for a logoff.
15. In response to the EAPOL-Logoff packet, the access device changes the status of the controlled port from authorized to unauthorized. Then, the access device sends an EAP-Failure packet to the client.

EAP termination

Figure 10 shows the basic 802.1X authentication procedure in EAP termination mode, assuming that CHAP authentication is used.

Figure 10 802.1X authentication procedure in EAP termination mode



In EAP termination mode, the access device rather than the authentication server generates an MD5 challenge for password encryption. The access device then sends the MD5 challenge together with the username and encrypted password in a standard RADIUS packet to the RADIUS server.

802.1X authentication initiation

Both the 802.1X client and the access device can initiate 802.1X authentication.

802.1X client as the initiator

The client sends an EAPOL-Start packet to the access device to initiate 802.1X authentication. The destination MAC address of the packet is the IEEE 802.1X specified multicast address 01-80-C2-00-00-03 or the broadcast MAC address. If any intermediate device between the client and the authentication server does not support the multicast address, you must use an 802.1X client that can send broadcast EAPOL-Start packets. For example, you can use the iNode 802.1X client.

Access device as the initiator

If the client cannot send EAPOL-Start packets, configure the access device to initiate authentication.

The access device supports the following modes:

- **Multicast trigger mode**—The access device multicasts EAP-Request/Identity packets to initiate 802.1X authentication at the identity request interval.

- **Unicast trigger mode**—Upon receiving a frame from an unknown MAC address, the access device sends an EAP-Request/Identity packet out of the receiving port to the MAC address. The device retransmits the packet if no response has been received within the identity request timeout interval. This process continues until the maximum number of request attempts set by using the **`dot1x retry`** command is reached.

The username request timeout timer sets both the identity request interval for the multicast trigger and the identity request timeout interval for the unicast trigger.

Access control methods

The device implements port-based access control as defined in the 802.1X protocol, and extends the protocol to support MAC-based access control.

- **Port-based access control**—Once an 802.1X user passes authentication on a port, any subsequent user can access the network through the port without authentication. When the authenticated user logs off, all other users are logged off.
- **MAC-based access control**—Each user is separately authenticated on a port. When a user logs off, no other online users are affected.

802.1X VLAN manipulation

Authorization VLAN

The authorization VLAN controls the access of an 802.1X user to authorized network resources. The device supports authorization VLANs assigned locally or by a remote server.

IMPORTANT:

Only remote servers can assign tagged authorization VLANs.

Remote VLAN authorization

In remote VLAN authorization, you must configure an authorization VLAN for a user on the remote server. After the user authenticates to the server, the server assigns authorization VLAN information to the device. Then, the device assigns the user access port to the authorization VLAN as a tagged or untagged member.

The device supports assignment of the following authorization VLAN information by the remote server:

- VLAN ID.
- VLAN name, which must be the same as the VLAN description on the access device.
- A string of VLAN IDs and VLAN names.

  In the string, some VLANs are represented by their IDs, and some VLANs are represented by their names.
- VLAN group name.

  For more information about VLAN groups, see *Layer 2—LAN Switching Configuration Guide*.
- VLAN ID with a suffix of **t** or **u**.

  The **t** and **u** suffixes require the device to assign the access port to the VLAN as a tagged or untagged member, respectively. For example, **2u** indicates assigning the port to VLAN 2 as an untagged member.

If a VLAN name or VLAN group name is assigned, the device converts the information into a VLAN ID before VLAN assignment.

IMPORTANT:

For a VLAN represented by its VLAN name to be assigned successfully, you must make sure the VLAN has been created on the device.

To assign VLAN IDs with suffixes, make sure the user access port is a hybrid or trunk port that performs port-based access control.

To ensure a successful assignment, the authorization VLANs assigned by the remote server cannot be any of the following types:

- Dynamically learned VLANs.
- Reserved VLANs.

If the server assigns a group of VLANs, the access device selects a VLAN as described in Table 2.

Table 2 Authorization VLAN selection from a group of VLANs

VLAN information	Authorization VLAN selection
VLANs by IDs VLANs by names VLAN group name	If the 802.1X-enabled port performs MAC-based access control, the device selects an authorization VLAN from the VLAN group for a user according to the following rules: • On a hybrid port with MAC-based VLAN enabled: ○ If the port does not have online users, the device selects the VLAN with the lowest ID. ○ If the port has online users, the device selects the VLAN that has the fewest online users. If two VLANs have the same number of online 802.1X users, the device selects the VLAN with the lower ID. • On an access, trunk, or MAC-based VLAN disabled hybrid port: ○ If the port does not have online users, the device selects the VLAN with the lowest ID. ○ If the port has online users, the device examines the VLAN group for the VLAN of the online users. If the VLAN is found, the VLAN is assigned to the user as the authorization VLAN. If the VLAN is not found, VLAN authorization fails. If the 802.1X-enabled port performs port-based access control, the device selects the VLAN with the lowest ID from the VLAN group. All subsequent 802.1X users are assigned to that VLAN.
VLAN IDs with suffixes	1. The device selects the leftmost VLAN ID without a suffix, or the leftmost VLAN ID suffixed by **u** as an untagged VLAN, whichever is more leftmost. 2. The device assigns the untagged VLAN to the port as the PVID, and it assigns the remaining as tagged VLANs. If no untagged VLAN is assigned, the PVID of the port does not change. The port permits traffic from these tagged and untagged VLANs to pass through. For example, the authentication server sends the string **1u 2t 3** to the access device for a user. The device assigns VLAN 1 as an untagged VLAN and all remaining VLANs (including VLAN 3) as tagged VLANs. VLAN 1 becomes the PVID.

Local VLAN authorization

To perform local VLAN authorization for a user, specify the VLAN ID in the authorization attribute list of the local user account for that user. For each local user, you can specify only one authorization VLAN ID. The user access port is assigned to the VLAN as an untagged member.

IMPORTANT:

Local VLAN authorization does not support assignment of tagged VLANs.

For more information about local user configuration, see "Configuring AAA."

Authorization VLAN manipulation on an 802.1X-enabled port

Table 3 describes how the access device handles VLANs (except for the VLANs specified with suffixes) on an 802.1X-enabled port.

Table 3 VLAN manipulation

Port access control method	Access port	Trunk port	Hybrid port
Port-based	Make sure the authorization VLAN assigned by the server has the untagged attribute. VLAN assignment will fail if the server issues a VLAN that has the tagged attribute. The device assigns the port to the first authenticated user's authorization VLAN and sets the VLAN as the PVID if that authorization VLAN has the untagged attribute.	If the authorization VLAN has the untagged attribute, the device assigns the port to the authorization VLAN and sets the VLAN as the PVID. If the authorization VLAN has the tagged attribute, the device assigns the port to the VLAN without changing the PVID.	If the authorization VLAN has the untagged attribute, the device assigns the port to the authorization VLAN as an untagged member and sets the VLAN as the PVID. If the authorization VLAN has the tagged attribute, the device assigns the port to the VLAN as a tagged member without changing the PVID.
MAC-based	Make sure the authorization VLAN assigned by the server has the untagged attribute. VLAN assignment will fail if the server issues a VLAN that has the tagged attribute. The device assigns the port to the first authenticated user's authorization VLAN and sets the VLAN as the PVID if that authorization VLAN has the untagged attribute.	If the authorization VLAN has the untagged attribute, the device assigns the port to the authorization VLAN and sets the VLAN as the PVID. If the authorization VLAN has the tagged attribute, the device assigns the port to the VLAN without changing the PVID.	If the authorization VLAN has the untagged attribute, the device assigns the port to the authorization VLAN as an untagged member and sets the VLAN as the PVID. If the authorization VLAN has the tagged attribute, the device assigns the port to the VLAN as a tagged member without changing the PVID. On a hybrid port with MAC-based VLAN enabled, if an untagged VLAN is assigned to a user, the device dynamically creates a VLAN and MAC address mapping for the user, without changing the PVID of the port. On a hybrid port with MAC-based VLAN disabled, the device assigns the port to the first authenticated user's authorization VLAN and sets that VLAN as the PVID if it has the untagged attribute.
	IMPORTANT: Always assign the same untagged VLAN to users on a port if the port is an access port, trunk port, or hybrid port with MAC MAC-based VLAN disabled. You can assign different untagged VLANs to different users only on a hybrid port with MAC-based VLAN enabled. After a user passes authentication and comes online with an untagged authorization VLAN, any subsequent users will be unable to pass authentication if they are assigned a different authorization VLAN. You can assign different tagged VLANs to different users, whether or not MAC-based		

Port access control method	Access port	Trunk port	Hybrid port
	VLAN is enabled. The device allows one tagged VLAN to be assigned to only one user.		

The VLAN assigned by the server to a user as an authorization VLAN might have been configured on the user access port but with a different tagging mode. For example, the server assigns an authorization VLAN with the tagged attribute, but the same VLAN configured on the port has the untagged attribute. In this situation, the VLAN settings that take effect on the user depend on the link type of the port.

- If the link type of the port is access or trunk, the authorization VLAN settings assigned by the server always take effect on the user as long as the user is online. After the user goes offline, the VLAN settings on the port take effect.
- If the link type of the port is hybrid, the VLAN settings configured on the port take effect. For example, the server assigns VLAN 30 with the untagged attribute to a user on the hybrid port. However, VLAN 30 has been configured on the port with the tagged attribute by using the **`port hybrid vlan tagged`** command. Then, the VLAN has the tagged attribute on the port.

For an 802.1X authenticated user to access the network on a hybrid port when no authorization VLAN is configured for the user, perform one of the following tasks:

- If the port receives tagged authentication packets from the user in a VLAN, use the **`port hybrid vlan`** command to configure the port as a tagged member in the VLAN.
- If the port receives untagged authentication packets from the user in a VLAN, use the **`port hybrid vlan`** command to configure the port as an untagged member in the VLAN.

On a port with periodic online user reauthentication enabled, the MAC-based VLAN feature does not take effect on a user that has been online since before this feature was enabled. The access device creates a MAC-to-VLAN mapping for the user when the following requirements are met:

- The user passes reauthentication.
- The authorization VLAN for the user is changed.

For more information about VLAN configuration and MAC-based VLANs, see *Layer 2—LAN Switching Configuration Guide*.

Guest VLAN

The 802.1X guest VLAN on a port accommodates users that have not performed 802.1X authentication. Users in the guest VLAN can access a limited set of network resources, such as a software server, to download antivirus software and system patches. Once a user in the guest VLAN passes 802.1X authentication, it is removed from the guest VLAN and can access authorized network resources.

The access device handles VLANs on an 802.1X-enabled port based on its 802.1X access control method.

Port-based access control

Authentication status	VLAN manipulation
A user accesses the 802.1X-enabled port when the port is in auto state.	The device assigns the port to the 802.1X guest VLAN. All 802.1X users on this port can access only resources in the guest VLAN. The guest VLAN assignment varies by port link mode. For more information, see "Authorization VLAN manipulation on an 802.1X-enabled port."

Authentication status	VLAN manipulation
A user in the 802.1X guest VLAN fails 802.1X authentication.	If an 802.1X Auth-Fail VLAN is available, the device assigns the port to the Auth-Fail VLAN. All users on this port can access only resources in the Auth-Fail VLAN. If no Auth-Fail VLAN is configured, the port is still in the 802.1X guest VLAN. All users on the port are in the guest VLAN. For information about the 802.1X Auth-Fail VLAN, see "Auth-Fail VLAN."
A user in the 802.1X guest VLAN passes 802.1X authentication.	The device removes the port from the 802.1X guest VLAN and assigns the port to the authorization VLAN of the user. If the authentication server does not assign an authorization VLAN, the initial PVID of the port applies. The user and all subsequent 802.1X users are assigned to the initial port VLAN. After the user logs off, the port is assigned to the guest VLAN again. **NOTE:** The initial PVID of an 802.1X-enabled port refers to the PVID used by the port before the port is assigned to any 802.1X VLANs.

IMPORTANT:

When the port receives a packet with a VLAN tag, the packet will be forwarded within the tagged VLAN if the VLAN is not the guest VLAN.

MAC-based access control

Authentication status	VLAN manipulation
A user accesses the 802.1X-enabled port and has not performed 802.1X authentication.	The device creates a mapping between the MAC address of the user and the 802.1X guest VLAN. The user can access only resources in the guest VLAN.
A user in the 802.1X guest VLAN fails 802.1X authentication.	If an 802.1X Auth-Fail VLAN is available, the device remaps the MAC address of the user to the Auth-Fail VLAN. The user can access only resources in the Auth-Fail VLAN. If no 802.1X Auth-Fail VLAN is configured, the user is removed from the guest VLAN and added to the initial PVID.
A user in the 802.1X guest VLAN passes 802.1X authentication.	The device remaps the MAC address of the user to the authorization VLAN. If the authentication server does not assign an authorization VLAN, the device remaps the MAC address of the user to the initial PVID on the port.

Auth-Fail VLAN

The 802.1X Auth-Fail VLAN on a port accommodates users that have failed 802.1X authentication because of the failure to comply with the organization's security policy. For example, the VLAN accommodates users that have entered a wrong password. Users in the Auth-Fail VLAN can access a limited set of network resources, such as a software server, to download antivirus software and system patches.

The access device handles VLANs on an 802.1X-enabled port based on its 802.1X access control method.

Port-based access control

Authentication status	VLAN manipulation
A user accesses the port and fails 802.1X authentication.	The device assigns the port to the Auth-Fail VLAN. All 802.1X users on this port can access only resources in the Auth-Fail VLAN. The Auth-Fail VLAN assignment varies by port link mode. For more information, see "Authorization VLAN manipulation on an 802.1X-enabled port."
A user in the 802.1X Auth-Fail VLAN fails 802.1X authentication.	The port is still in the Auth-Fail VLAN, and all 802.1X users on this port are in this VLAN.
A user in the 802.1X Auth-Fail VLAN passes 802.1X authentication.	The device assigns the port to the authorization VLAN of the user, and it removes the port from the Auth-Fail VLAN. If the authentication server does not assign an authorization VLAN, the initial PVID of the port applies. The user and all subsequent 802.1X users are assigned to the initial PVID. After the user logs off, the port is assigned to the guest VLAN. If no guest VLAN is configured, the port is assigned to the initial PVID of the port.

MAC-based access control

Authentication status	VLAN manipulation
A user accesses the port and fails 802.1X authentication.	The device maps the MAC address of the user to the 802.1X Auth-Fail VLAN. The user can access only resources in the Auth-Fail VLAN.
A user in the 802.1X Auth-Fail VLAN fails 802.1X authentication.	The user is still in the Auth-Fail VLAN.
A user in the 802.1X Auth-Fail VLAN passes 802.1X authentication.	The device remaps the MAC address of the user to the authorization VLAN. If the authentication server does not assign an authorization VLAN, the device remaps the MAC address of the user to the initial PVID on the port.

Critical VLAN

The 802.1X critical VLAN on a port accommodates 802.1X users that have failed authentication because none of the RADIUS servers in their ISP domain are reachable. Users in the critical VLAN can access a limited set of network resources depending on the configuration.

The critical VLAN feature takes effect when 802.1X authentication is performed only through RADIUS servers. If an 802.1X user fails local authentication after RADIUS authentication, the user is not assigned to the critical VLAN. For more information about the authentication methods, see "Configuring AAA."

The access device handles VLANs on an 802.1X-enabled port based on its 802.1X access control method.

Port-based access control

Authentication status	VLAN manipulation
A user accesses the port and fails 802.1X authentication because all the RADIUS servers are unreachable. A user that has passed authentication fails reauthentication because all the RADIUS servers are unreachable, and the user is logged out of the device.	The device assigns the port to the critical VLAN. The 802.1X user and all subsequent 802.1X users on this port can access only resources in the 802.1X critical VLAN. The critical VLAN assignment varies by port link mode. For more information, see "Authorization VLAN manipulation on an 802.1X-enabled port."
A user in the 802.1X critical VLAN fails authentication because all the RADIUS servers are unreachable.	The port is still in the critical VLAN.
A user in the 802.1X critical VLAN fails authentication for any reason other than unreachable servers.	If an 802.1X Auth-Fail VLAN has been configured, the port is assigned to the Auth-Fail VLAN. If no 802.1X Auth-Fail VLAN is configured, the port is assigned to the initial PVID of the port.
A user in the 802.1X critical VLAN passes 802.1X authentication.	The device assigns the port to the authorization VLAN of the user, and it removes the port from the 802.1X critical VLAN. If the authentication server does not assign an authorization VLAN, the initial PVID of the port applies. The user and all subsequent 802.1X users are assigned to this port VLAN. After the user logs off, the port is assigned to the guest VLAN. If no 802.1X guest VLAN is configured, the initial PVID of the port is restored.

If the port is added to the critical VLAN because no RADIUS servers are reachable, the device performs the following operations after it detects a reachable RADIUS server:

1. Removes the port from the critical VLAN.
2. Sends a multicast EAP-Request/Identity message out of the port to trigger authentication.

MAC-based access control

Authentication status	VLAN manipulation
A user accesses the port and fails 802.1X authentication because all the RADIUS servers are unreachable.	The device maps the MAC address of the user to the 802.1X critical VLAN. The user can access only resources in the 802.1X critical VLAN.
A user in the 802.1X critical VLAN fails authentication because all the RADIUS servers are unreachable.	The user is still in the critical VLAN.
A user in the 802.1X critical VLAN fails 802.1X authentication for any reason other than unreachable servers.	If an 802.1X Auth-Fail VLAN has been configured, the device remaps the MAC address of the user to the Auth-Fail VLAN ID. If no 802.1X Auth-Fail VLAN has been configured, the device remaps the MAC address of the user to the initial PVID.
A user in the 802.1X critical VLAN passes 802.1X authentication.	The device remaps the MAC address of the user to the authorization VLAN. If the authentication server does not assign an authorization VLAN to the user, the device remaps the MAC address of the user to the initial PVID on the port.

If a user is added to the critical VLAN because no RADIUS servers are reachable, the device performs the following operations after it detects a reachable RADIUS server:

1. Removes the user from the critical VLAN.
2. Sends a unicast EAP-Request/Identity message out of the port to the user for reauthentication.

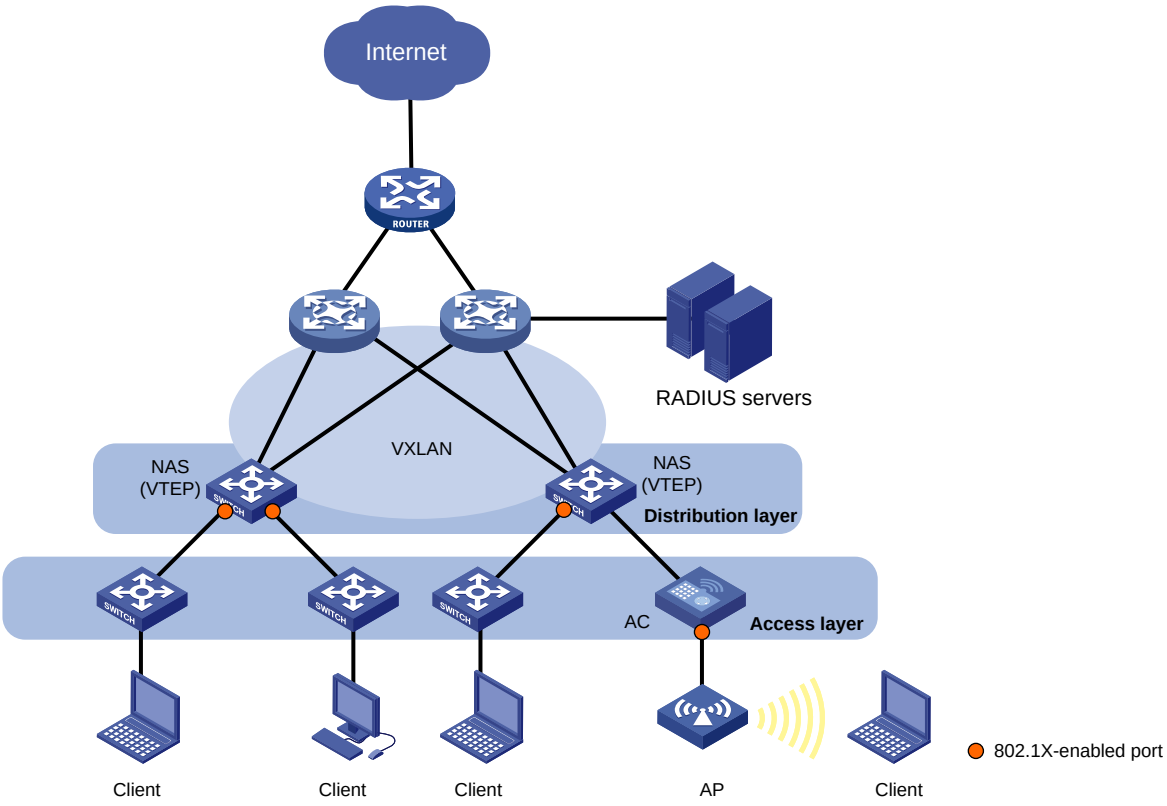
802.1X VSI manipulation

802.1X support for VXLANs

As shown in Figure 11, when the device acts as both a VXLAN VTEP and a NAS, users' service information cannot be identified by VLANs. To resolve this issue, you must configure the RADIUS server to assign VSIs to authenticated 802.1X users. The NAS will map a user's traffic to the VXLAN that is associated with the user's authorization VSI. The mapping criteria include the user's access VLAN, access port, and MAC address. MAC address information is required only when the access port performs MAC-based access control.

For information about VSIs and VXLANs, see *VXLAN Configuration Guide*.

Figure 11 VXLAN network diagram for 802.1X authentication



Authorization VSI

An authorization VSI is associated with a VXLAN that has network resources inaccessible to unauthenticated users.

802.1X supports remote VSI authorization. If the VTEP does not receive authorization VSI information for an 802.1X user from the remote server, the user cannot access resources in any

VXLAN after passing authentication. If the VTEP receives authorization VSI information for a user from the remote server, it handles the authorization VSIs on an 802.1X-enabled port as shown in Table 4.

Table 4 VSI manipulation

Port access control method	VSI manipulation
Port-based	The VTEP dynamically creates an Ethernet service instance according to the user's access port and VLAN, and it maps the AC to the authorization VSI. Subsequent users in the same VLAN can access resources in the VXLAN associated with the VSI without authentication.
MAC-based	The VTEP dynamically creates an Ethernet service instance according to the user's access port, VLAN, and MAC address, and it then maps the AC to the authorization VSI. The user then can access resources in the VXLAN associated with the VSI.

For information about dynamic creation of Ethernet service instances, see *VXLAN configuration Guide*.

Guest VSI

The 802.1X guest VSI on a port accommodates users that have not performed 802.1X authentication. You can deploy a limited set of network resources in the VXLAN that is associated with the guest VSI. For example, deploy a software server for users to download anti-virus software and system patches. Once a user in the guest VSI passes 802.1X authentication, the user is removed from the guest VSI and can access authorized network resources.

The VTEP handles VSIs on an 802.1X-enabled port based on its 802.1X access control method.

Port-based access control

Authentication status	VSI manipulation
A user accesses the port and has not performed 802.1X authentication.	The VTEP assigns the port to the 802.1X guest VSI. All 802.1X users from the same VLAN on this port can access only resources in the VXLAN associated with the guest VSI.
A user in the 802.1X guest VSI fails 802.1X authentication.	If an 802.1X Auth-Fail VSI is available (see "Auth-Fail VSI"), the VTEP assigns the port to the Auth-Fail VSI. All users from the same VLAN on this port can access only resources in the VXLAN associated with the Auth-Fail VSI. If no Auth-Fail VSI is configured, the port is still in the 802.1X guest VSI.
A user in the 802.1X guest VSI passes 802.1X authentication.	The VTEP removes the port from the 802.1X guest VSI and assigns the port to the authorization VSI of the user.

MAC-based access control

Authentication status	VSI manipulation
A user accesses the port and has not performed 802.1X authentication.	The VTEP maps the user's MAC address and access VLAN to the 802.1X guest VSI on the port. The user can access only resources in the VXLAN associated with the guest VSI.
A user in the 802.1X guest VSI fails 802.1X authentication.	If an 802.1X Auth-Fail VSI is available on the port, the VTEP remaps the user's MAC address and access VLAN to the Auth-Fail VSI. The user can access only resources in the VXLAN associated with the Auth-Fail VSI. If no 802.1X Auth-Fail VSI is configured on the port, the user is removed from the 802.1X guest VSI.
A user in the 802.1X guest VSI passes 802.1X authentication.	The VTEP removes the user from the 802.1X guest VSI and remaps the user's MAC address and access VLAN to the authorization VSI.

Auth-Fail VSI

The 802.1X Auth-Fail VSI on a port accommodates users that have failed 802.1X authentication because of the failure to comply with the organization's security policy. For example, the VSI accommodates users with wrong passwords entered. Users in the Auth-Fail VSI can access a limited set of network resources in the VXLAN associated with this VSI. You can deploy a software server in the Auth-Fail VSI for users to download antivirus software and system patches.

The VTEP handles VSIs on an 802.1X-enabled port based on its 802.1X access control method.

Port-based access control

Authentication status	VSI manipulation
A user accesses the port and fails 802.1X authentication.	The VTEP assigns the port to the Auth-Fail VSI. All 802.1X users from the same VLAN on this port can access only resources in the VXLAN associated with the Auth-Fail VSI.
A user in the 802.1X Auth-Fail VSI fails 802.1X authentication because of any reason other than unreachable servers.	The port is still in the Auth-Fail VSI.
A user passes 802.1X authentication.	The VTEP removes the port from the Auth-Fail VSI and assigns the port to the authorization VSI of the user. After the user logs off, the port is removed from the authorization VSI. If the 802.1X guest VSI is configured on the port, the VTEP assigns the port to the guest VSI.

MAC-based access control

Authentication status	VSI manipulation
A user accesses the port and fails 802.1X authentication.	The VTEP maps the user's MAC address and access VLAN to the 802.1X Auth-Fail VSI on the port. The user can access only resources in the VXLAN associated with the Auth-Fail VSI.

Authentication status	VSI manipulation
A user in the 802.1X Auth-Fail VSI fails 802.1X authentication because of any reason other than unreachable servers.	The user is still in the Auth-Fail VSI.
A user in the 802.1X Auth-Fail VSI passes 802.1X authentication.	The VTEP removes the user from the 802.1X Auth-Fail VSI and remaps the user's MAC address and access VLAN to the authorization VSI.

Critical VSI

The 802.1X critical VSI on a port accommodates 802.1X users that have failed authentication because none of the RADIUS servers in their ISP domain are reachable. Users in the critical VSI can access a limited set of network resources in the VXLAN associated with this VSI.

The critical VSI feature takes effect when 802.1X authentication is performed only through RADIUS servers. If an 802.1X user fails local authentication after RADIUS authentication, the user is not assigned to the critical VSI. For more information about the authentication methods, see "Configuring AAA."

The VTEP handles VSIs on an 802.1X-enabled port based on its 802.1X access control method.

Port-based access control

Authentication status	VSI manipulation
A user accesses the port and fails 802.1X authentication because all the RADIUS servers are unreachable.	The VTEP assigns the port to the critical VSI. The 802.1X user and all subsequent 802.1X users from the same VLAN on this port can access only resources in the VXLAN associated with the critical VSI.
A user in the 802.1X critical VSI fails authentication because all the RADIUS servers are unreachable.	The port is still in the critical VSI, and all 802.1X users from the same VLAN on this port are in the critical VSI.
A user in the 802.1X critical VSI fails authentication for any reason other than unreachable servers.	If an 802.1X Auth-Fail VSI has been configured for the port, the VTEP removes the port to the 802.1X Auth-Fail VSI. If no 802.1X Auth-Fail VSI is configured for the port, the VTEP logs off the user.
A user in the 802.1X critical VSI passes 802.1X authentication.	The VTEP removes the port from the 802.1X critical VSI and assigns the port to the authorization VSI. After the user logs off, the port is removed from the authorization VSI. If the port has been configured with the 802.1X guest VSI, the VTEP assigns the port to the guest VSI.

Authentication status	VSI manipulation
A user in the 802.1X guest VSI fails authentication because all the RADIUS servers are unreachable.	The VTEP assigns the port to the critical VSI. All 802.1X users on this port can access only resources in the VXLAN associated with the 802.1X critical VSI.
A user in the 802.1X Auth-Fail VSI fails authentication because all the RADIUS servers are unreachable.	The PVID of the port remains unchanged. All 802.1X users on this port can access only resources in the VXLAN associated with the 802.1X Auth-Fail VSI.
A user that has passed authentication fails reauthentication because all the RADIUS servers are unreachable, and the user is logged out of the device.	The VTEP assigns the port to the critical VSI.

MAC-based access control

Authentication status	VSI manipulation
A user accesses the port and fails 802.1X authentication because all the RADIUS servers are unreachable.	The VTEP maps the user's MAC address and access VLAN to the 802.1X critical VSI on the port. The user can access only resources in the VXLAN associated with the critical VSI.
A user in the 802.1X critical VSI fails authentication because all the RADIUS servers are unreachable.	The user is still in the critical VSI.
A user in the 802.1X critical VSI fails 802.1X authentication for any reason other than unreachable servers.	If an 802.1X Auth-Fail VSI has been configured on the port, the VTEP remaps the user's MAC address and access VLAN to the Auth-Fail VSI. If no 802.1X Auth-Fail VSI has been configured on the port, the VTEP logs off the user.
A user in the 802.1X critical VSI passes 802.1X authentication.	The VTEP remaps the user's MAC address and access VLAN to the authorization VSI.
A user in the 802.1X guest VSI fails authentication because all the RADIUS servers are unreachable.	The VTEP maps the user's MAC address and access VLAN to the 802.1X critical VSI on the port. The user can access only resources in the VXLAN associated with the critical VSI.
A user in the 802.1X Auth-Fail VSI fails authentication because all the RADIUS servers are unreachable.	The user remains in the 802.1X Auth-Fail VSI.

802.1X microsegment assignment

Authorization microsegment

A microsegment is a group of traffic that matches the same criteria. The microsegmentation feature provides group-based traffic control by applying policy to traffic on a per-microsegment basis. The RADIUS server can assign a microsegment to a user after that user passes 802.1X authentication. Then, the device performs microsegment-based traffic filtering for that user on the user's access port to make sure the user can access only the network resources available for the microsegment.

For more information about microsegmentation, see "Configuring microsegmentation."

IMPORTANT:

The following restrictions apply to 802.1X authentication on a microsegmented network:

- The server must assign an authorization ACL, redirect URL, and microsegment to an 802.1X user if that user must pass Web authentication.
- If the server assigns both an authorization ACL and an authorization microsegment to an 802.1X user that does not require Web authentication, only the authorization microsegment takes effect.

Critical microsegment

The 802.1X critical microsegment on a port accommodates users that have failed 802.1X authentication because no RADIUS authentication servers are reachable. Users in an 802.1X critical microsegment can access only network resources in the critical microsegment. The microsegment can belong to a VLAN or VXLAN.

The critical microsegment feature takes effect when 802.1X authentication is performed only through RADIUS servers. If an 802.1X user fails local authentication after RADIUS authentication, the user is not assigned to the critical microsegment. For more information about the authentication methods, see "Configuring AAA."

The access device handles users on an 802.1X-enabled port based on its 802.1X access control method.

Port-based access control

The access device assigns or removes 802.1X users to or from the critical microsegment on a port depending on user authentication status, as shown in Table 5.

Table 5 Microsegment manipulation

Authentication status	VSI manipulation
A user accesses the port and fails 802.1X authentication because all the RADIUS servers are unreachable.	The device assigns the port to the critical microsegment. The 802.1X user and all subsequent 802.1X users on this port can access only resources in the critical microsegment.
A user in the 802.1X critical microsegment fails authentication because all the RADIUS servers are unreachable.	The port is still in the critical microsegment, and all 802.1X users on this port are in the critical microsegment.
A user in the 802.1X critical microsegment fails authentication for any reason other than unreachable servers.	The device removes the port from the 802.1X critical microsegment.

Authentication status	VSI manipulation
A user in the 802.1X critical microsegment passes 802.1X authentication.	The device removes the port from the 802.1X critical microsegment and assigns the port to the authorization microsegment. If no authorization microsegment is assigned, the device does not add the port to any microsegment.

MAC-based access control

The access device assigns or removes an 802.1X user to or from the critical microsegment on a port depending on the user's authentication status, as shown in Table 6.

Table 6 Microsegment manipulation

Authentication status	Microsegment manipulation
A user fails 802.1X authentication because all the RADIUS servers are unreachable.	The access device assigns the user to the 802.1X critical microsegment. The access device retains the user in the 802.1X critical microsegment if 802.1X reauthentication fails because none of the RADIUS servers is reachable.
A user in the 802.1X critical microsegment fails 802.1X for any reason other than server unreachable.	The access device removes the user from the 802.1X critical microsegment.
A user in the 802.1X critical microsegment passes 802.1X authentication.	The access device assigns the user to the authorization microsegment assigned by the authentication server.

ACL assignment

You can specify an authorization ACL for an 802.1X user on a remote server or the access device to control the user's access to network resources. After the user passes 802.1X authentication, the server or access device assigns the authorization ACL to the user access port. Then, the port permits or drops the matching traffic for the user depending on the rules in the ACL.

The device supports assignment of static and dynamic authorization ACLs.

- **Assignment of static authorization ACLs**—Static ACLs can be assigned by a RADIUS server or the access device. When the server or access device assigns a static ACL to a user, it assigns only the ACL number. You must manually create the ACL and configure its rules on the access device.

  To change the access permissions of a user, you can use one of the following methods:

  - Modify ACL rules in the authorization ACL on the access device.
  - Assign another ACL to the user as the authorization ACL from the RADIUS server or the access device.

  Static ACLs and their rules can be manually deleted from the access device.
- **Assignment of dynamic authorization ACLs**—Dynamic ACLs and their rules are automatically deployed by a RADIUS server, which are not configurable on the access device. Dynamic ACLs can only be named ACLs. After the device receives a server-deployed dynamic ACL and its rules, it automatically creates the ACL and configures its rules.

If the dynamic ACL assigned by the server to a user has the same name as a static ACL, the dynamic ACL cannot be issued and the user cannot come online.

A dynamic ACL and its rules are automatically deleted from the access device after all its users go offline.

Dynamic ACLs and their rules cannot be manually modified or deleted on the access device. To display information about dynamic ACLs and their rules, use the **`display dot1x connection`** or **`display acl`** command.

IMPORTANT:

The supported authorization ACLs include the following types:

- Basic ACLs, which are numbered in the range of 2000 to 2999.
- Advanced ACLs, which are numbered in the range of 3000 to 3999.
- Layer 2 ACLs, which are numbered in the range of 4000 to 4999.
- User-defined ACLs, which are numbered in the range of 5000 to 5999.

For an authorization ACL to take effect, make sure the ACL exists with rules and none of the rules contains the **`counting`**, **`established`**, **`fragment`**, **`source-mac`**, or **`logging`** keyword.

For more information about ACLs, see *ACL and QoS Command Reference*.

IMPORTANT:

If the server assigns both an authorization ACL and microsegment to an 802.1X authentication user on a microsegmented network, only the authorization microsegment takes effect.

Redirect URL assignment

The device supports the URL attribute assigned by a RADIUS server when the 802.1X-enabled port performs MAC-based access control and the port authorization state is **auto**. During authentication, the HTTP or HTTPS requests of an 802.1X user are redirected to the Web interface specified by the server-assigned URL attribute. After the user passes the Web authentication, the RADIUS server records the MAC address of the user and uses a DM (Disconnect Message) to log off the user. When the user initiates 802.1X authentication again, it will pass the authentication and come online successfully.

This feature is mutually exclusive with the EAD assistant feature.

By default, the device listens to port 6654 for HTTPS requests to be redirected. To change the redirect listening port number, see configuring HTTP redirect in *Layer 3—IP Services Configuration Guide*.

802.1X support for critical profiles

If all RADIUS servers in a RADIUS scheme are not reachable for user authentication, users will be unable to get authenticated to come online. To make sure users can access certain resources in this situation, you can specify a critical profile for them on their access port. For example, you can define critical microsegment resources in the critical profile.

For users in a VPN instance, the resources defined for that VPN instance in a critical profile take precedence over the default resources in the critical profile.

Periodic 802.1X reauthentication

Periodic 802.1X reauthentication tracks the connection status of online users and updates the authorization attributes (such as ACL and VLAN) assigned by the server.

The device reauthenticates online 802.1X users at the periodic reauthentication interval when the periodic online user reauthentication feature is enabled. The interval is controlled by a timer and the timer is user configurable. A change to the periodic reauthentication timer applies to online users only after the old timer expires and the users pass authentication.

The server-assigned session timeout timer (Session-Timeout attribute) and termination action (Termination-Action attribute) together can affect the periodic online user reauthentication feature. To display the server-assigned Session-Timeout and Termination-Action attributes, use the **`display dot1x connection`** command (see *Security Command Reference*).

- If the termination action is **Default** (logoff), periodic online user reauthentication on the device takes effect only when the periodic reauthentication timer is shorter than the session timeout timer.
- If the termination action is **Radius-request**, the periodic online user reauthentication settings on the device do not take effect. The device reauthenticates the online 802.1X users after the session timeout timer expires.

If no session timeout timer is assigned by the server, whether the device performs periodic 802.1X reauthentication depends on the periodic reauthentication configuration on the device. Support for the assignment of Session-Timeout and Termination-Action attributes depends on the server model.

With the RADIUS DAS feature enabled, the device immediately reauthenticates a user upon receiving a CoA message that carries the reauthentication attribute from a RADIUS authentication server. In this case, reauthentication will be performed regardless of whether 802.1X periodic reauthentication is enabled on the device. For more information about RADIUS DAS configuration, see "Configuring AAA."

By default, the device logs off online 802.1X users if no server is reachable for 802.1X reauthentication. The keep-online feature keeps authenticated 802.1X users online when no server is reachable for 802.1X reauthentication.

The VLANs assigned to an online user before and after reauthentication can be the same or different.

EAD assistant

Endpoint Admission Defense (EAD) is an integrated endpoint access control solution to improve the threat defensive capability of a network. The solution enables the security client, security policy server, access device, and third-party server to operate together. If a terminal device seeks to access an EAD network, it must have an EAD client, which performs 802.1X authentication.

The EAD assistant feature allows unauthenticated users to access freely accessible network segments and enables the access device to redirect the HTTP or HTTPS requests of users to a redirect URL for downloading and installing EAD clients. This feature eliminates the administrative task to deploy EAD clients.

EAD assistant is implemented by the following functionality:

- Free IP.

  A free IP is a freely accessible network segment, which has a limited set of network resources such as software and DHCP servers. To ensure security policy compliance, an unauthenticated user can access only this segment to perform operations. For example, the user can download EAD client from a software server or obtain a dynamic IP address from a DHCP server.
- Free microsegment.

A free microsegment is a freely accessible microsegment, which has a limited set of network resources such as software and DHCP servers. To ensure security policy compliance, an unauthenticated user can access only this microsegment to perform operations. For example, the user can download EAD client from a software server or obtain a dynamic IP address from a DHCP server.

- Redirect URL.

  If an unauthenticated 802.1X user is using a Web browser to access the network, EAD assistant redirects the network access requests of the user to a specific URL. For example, you can use this feature to redirect the user to the EAD client software download page.

The EAD assistant feature creates an ACL-based EAD rule automatically to open access to the free segment for each unauthenticated user.

EAD rules are implemented by using ACL resources. When the EAD rule timer expires or the user passes authentication, the rule is removed. If users fail to download EAD client or fail to pass authentication before the timer expires, they must reconnect to the network to access the free segment.

Configuring 802.1X

Restrictions and guidelines: 802.1X configuration

You can configure the port security feature to perform 802.1X. Port security combines and extends 802.1X and MAC authentication. It applies to a network (a WLAN, for example) that requires different authentication methods for different users on a port. For more information about the port security feature, see "Configuring port security."

When RADIUS DAS is enabled, an M-LAG system does not support shutting down or rebooting the access ports of 802.1X users or reauthenticating the users through CoA messages. For more information about RADIUS DAS, see "Configuring AAA."

When you configure 802.1X settings on an interface, follow these restrictions and guidelines:

- 802.1X is supported only on Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces.
- If you add a Layer 2 Ethernet interface to an aggregation group, the 802.1X settings configured on it will not take effect until it is removed from the aggregation group.
- Do not delete a Layer 2 aggregate interface if the interface has online 802.1X users.
- Do not change the link type of a port when the 802.1X guest VLAN, Auth-Fail VLAN, or critical VLAN on the port has users.
- To ensure that online 802.1X users on a Layer 2 aggregate interface with MAC-based VLAN enabled can correctly go offline, enable MAC-based VLAN also on all member ports of the aggregate interface.

On an interface, the 802.1X critical profile settings are mutually exclusive with the 802.1X critical VSI, critical VLAN, and critical microsegment settings.

On an interface, the 802.1X critical profile settings are mutually exclusive with the critical VSI, critical VLAN, and critical microsegment settings for MAC authentication.

When you configure 802.1X VLANs, VSIs, or miscrosegments, follow these restrictions and guidelines:

- If the server assigns both an authorization VSI and authorization VLAN to a user, the device uses only the authorization VLAN.
- On an interface avoid the incompatible 802.1X settings in Table 7.

Table 7 Settings incompatible with 802.1X VLANs, VSIs, or miscrosegments on a port

Settings	Incompatible settings
802.1X guest VLAN, Auth-Fail VLAN, and critical VLAN settings	802.1X guest VSI, Auth-Fail VSI, and critical VSI settings 802.1X critical microsegment settings
802.1X guest VSI, Auth-Fail VSI, and critical VSI settings	802.1X guest VLAN, Auth-Fail VLAN, and critical VLAN settings 802.1X critical microsegment settings
802.1X critical microsegment settings	802.1X guest VLAN, Auth-Fail VLAN, and critical VLAN settings 802.1X guest VSI, Auth-Fail VSI, and critical VSI settings

- On an interface, avoid the incompatible 802.1X and MAC authentication settings in Table 8.

  Table 8 MAC authentication settings incompatible with 802.1X settings on a port

Settings	Incompatible settings
802.1X guest VLAN, Auth-Fail VLAN, and critical VLAN settings	Guest VSI and critical VSI settings for MAC authentication Critical microsegment settings for MAC authentication
802.1X guest VSI, Auth-Fail VSI, and critical VSI settings	Guest VLAN and critical VLAN settings for MAC authentication Critical microsegment settings for MAC authentication

- To ensure a successful authentication, you must configure the authentication server to assign authorization microsegments, VLANs, or VSIs to the 802.1X users attached to a port in the following situations:
  - If the port is configured with the guest VLAN, Auth-Fail VLAN, or critical VLAN, you must configure VLAN authorization for the 802.1X users.
  - If the port is configured with the guest VSI, Auth-Fail VSI, or critical VSI, you must configure VSI authorization for the 802.1X users.
  - If the port is configured with critical microsegment settings, you must configure microsegment authorization for the 802.1X users.
  - If both 802.1X and MAC authentication are configured on the port, be careful with the VSI settings of the two authentication features.
    - You must configure VSI authorization for the 802.1X users if the port has a guest or critical VSI for MAC authentication.
    - Likewise, you must configure VSI authorization for the MAC authentication users if the port has an Auth-Fail, guest, or critical VSI for 802.1X authentication.
- As a best practice, do not use Web authentication in conjunction with any of the following 802.1X settings on a port:
  - Guest VLAN, Auth-Fail VLAN, and critical VLAN settings.
  - Guest VSI, Auth-Fail VSI, and critical VSI settings.
  - Critical microsegment settings.
- As a best practice, do not use port security preauthentication domain in conjunction with any of the following 802.1X settings on a port:
  - Guest VLAN, Auth-Fail VLAN, and critical VLAN settings.
  - Guest VSI, Auth-Fail VSI, and critical VSI settings.
  - Critical microsegment settings.
  - Critical profile settings.
- If you have configured multiple default authentication methods with RADIUS authentication as the primary method for an ISP domain by using the **`authentication default`** command, the following restrictions and guidelines apply:

- If users fail RADIUS authentication because all RADIUS servers are unreachable, the system tries backup authentication methods to authenticate them, without adding them to the critical VLAN, VSI, or microsegmentthe critical VLAN or VSI.
- The users will be added to the critical VLAN, VSI, or microsegment only if they fail RADIUS authentication upon unreachability of all RADIUS servers and no other authentication methods are available.

To ensure a successful HTTPS redirect for users in the following situations, make sure VLAN interfaces or VSI interfaces exist for the VLANs or VSIs that transport their packets, respectively:

- The server has assigned a redirect URL to the users.
- EAD assistant is enabled.
-

802.1X tasks at a glance

To configure 802.1X authentication, perform the following tasks:

1. Enabling 802.1X
2. Configuring basic 802.1X features
   - Enabling EAP relay or EAP termination
   - Setting the port authorization state
   - Specifying an access control method
   - (Optional.) Specifying a mandatory authentication domain on a port
   - (Optional.) Setting the 802.1X authentication timeout timers
   - (Optional.) Configuring 802.1X reauthentication
   - (Optional.) Setting the quiet timer
3. (Optional.) Configuring 802.1X VLAN assignment
   - Configuring an 802.1X guest VLAN
   - Enabling 802.1X guest VLAN assignment delay
   - Configuring an 802.1X Auth-Fail VLAN
   - Configuring an 802.1X critical VLAN
4. (Optional.) Configuring 802.1X VSI assignment
   - Configuring an 802.1X guest VSI
   - Enabling 802.1X guest VSI assignment delay
   - Configuring an 802.1X Auth-Fail VSI
   - Configuring an 802.1X critical VSI
5. (Optional.) Configuring an 802.1X critical microsegment
6. (Optional.) Specifying a critical profile for 802.1X users on a port
7. (Optional.) Configuring an 802.1X access profile
8. (Optional.) Setting the upper limit for 802.1X parameters
   - Setting the maximum number of concurrent 802.1X users on a port
   - Setting the alarm threshold and alarm clear threshold for 802.1X user access ratio
   - Setting the maximum number of authentication request attempts
   - Setting the maximum number of 802.1X authentication attempts for MAC authenticated users
9. (Optional.) Configuring other 802.1X features
   - Configuring 802.1X unauthenticated user aging

- Sending EAP-Success packets on assignment of users to the 802.1X Auth-Fail VLAN or VSI
- Sending EAP-Success packets on assignment of users to the 802.1X critical VLAN or VSI
- Enabling 802.1X online user synchronization
- Configuring the authentication trigger feature

  Perform this task when 802.1X clients cannot initiate authentication.
- Sending EAP-Success packets to 802.1X clients after successful authorization
- Discarding duplicate 802.1X EAPOL-Start requests
- Configuring online user handshake
- Configuring 802.1X offline detection
- Enabling online 802.1X users to escape from offline detection and stay online when no reachable RADIUS authentication servers are available
- Specifying supported domain name delimiters
- Removing the VLAN tags of 802.1X protocol packets sent out of a port
- Enabling 802.1X user IP freezing
- Enabling generation of dynamic IPSG binding entries for 802.1X authenticated users
- Configuring 802.1X MAC address binding
- Configuring the EAD assistant feature
- Setting the maximum size of EAP-TLS fragments sent to the server

  Use this feature to reduce the size of authentication packets sent to the server when the device uses EAP-TLS authentication method in EAP relay mode.
- Setting the unicast-trigger quiet timer
- Logging off 802.1X users
- Enabling 802.1X user logging
- Configuring SNMP notifications for 802.1X

Prerequisites for 802.1X

Before you configure 802.1X, complete the following tasks:

- Configure an ISP domain and AAA scheme (local or RADIUS authentication) for 802.1X users.
- If RADIUS authentication is used, create user accounts on the RADIUS server.
- If local authentication is used, create local user accounts on the access device and set the service type to **lan-access**.

Enabling 802.1X

Restrictions and guidelines

For 802.1X to take effect on a port, you must enable it both globally and on the port.

Procedure

1. Enter system view.

   `system-view`
2. Enable 802.1X globally.

   `dot1x`

   By default, 802.1X is disabled globally.

3. Enter interface view.
   interface *interface-type interface-number*
4. Enable 802.1X on a port.
   dot1x
   By default, 802.1X is disabled on a port.

Enabling EAP relay or EAP termination

About this task

Consider the following factors to select a proper EAP mode:

- Support of the RADIUS server for EAP packets.
- Authentication methods supported by the 802.1X client and the RADIUS server.

Restrictions and guidelines

- If EAP relay mode is used, the **user-name-format** command configured in RADIUS scheme view does not take effect. The access device sends the authentication data from the client to the server without any modification. For more information about the **user-name-format** command, see *Security Command Reference*.
- You can use both EAP termination and EAP relay in any of the following situations:
  - The client is using only MD5-Challenge EAP authentication. If EAP termination is used, you must enable CHAP authentication on the access device.
  - The client is an iNode 802.1X client and initiates only the username and password EAP authentication. If EAP termination is used, you can enable either PAP or CHAP authentication on the access device. However, for the purpose of security, you must use CHAP authentication on the access device.
- To use EAP-TLS, PEAP, or any other EAP authentication methods, you must use EAP relay. When you make your decision, see "Comparing EAP relay and EAP termination" for help.

Procedure

1. Enter system view.
   system-view
2. Configure EAP relay or EAP termination.
   dot1x authentication-method { **chap** | **eap** | **pap** }
   By default, the access device performs EAP termination and uses CHAP to communicate with the RADIUS server.

Setting the port authorization state

About this task

The port authorization state determines whether the client is granted access to the network. You can control the following authorization states of a port:

- **Authorized**—Places the port in the authorized state, enabling users on the port to access the network without authentication.
- **Unauthorized**—Places the port in the unauthorized state, denying any access requests from users on the port.
- **Auto**—Places the port initially in unauthorized state to allow only EAPOL packets to pass. After a user passes authentication, sets the port in the authorized state to allow access to the network. You can use this option in most scenarios.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the port authorization state.
   dot1x port-control { **authorized-force** | **auto** | **unauthorized-force** }
   By default, the **auto** state applies.

Specifying an access control method

About this task

The device supports port-based and MAC-based access control methods.

Restrictions and guidelines

If online 802.1X users are present on a port, changing its access control method will cause the online users to go offline.

To avoid authentication anomalies, do not execute the **dot1x port-method portbased** command on an interface that is bound to a port security authentication profile.

802.1X port-based access control is mutually exclusive with MAC authentication, Web authentication and static user authentication.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify an access control method.
   dot1x port-method { **macbased** | **portbased** }
   By default, MAC-based access control applies.

Specifying a mandatory authentication domain on a port

About this task

You can place all 802.1X users in a mandatory authentication domain for authentication, authorization, and accounting on a port. No user can use an account in any other domain to access the network through the port. The implementation of a mandatory authentication domain enhances the flexibility of 802.1X access control deployment.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify a mandatory 802.1X authentication domain on the port.

dot1x mandatory-domain *domain-name*

By default, no mandatory 802.1X authentication domain is specified.

Setting the 802.1X authentication timeout timers

About this task

The network device uses the following 802.1X authentication timeout timers:

- **Client timeout timer**—Starts when the access device sends an EAP-Request/MD5-Challenge packet to a client. If no response is received when this timer expires, the access device retransmits the request to the client.
- **Server timeout timer**—Starts when the access device sends a RADIUS Access-Request packet to the authentication server. If no response is received when this timer expires, the 802.1X authentication fails.

Restrictions and guidelines

In most cases, the default settings are sufficient. You can edit the timers, depending on the network conditions.

- In a low-speed network, increase the client timeout timer.
- In a network with authentication servers of different performance, adjust the server timeout timer.

To avoid forced logoff before the server timeout timer expires, set the server timeout timer to a value that is lower than or equal to the product of the following values:

- The maximum number of RADIUS packet transmission attempts set by using the **retry** command in RADIUS scheme view.
- The RADIUS server response timeout timer set by using the **timer response-timeout** command in RADIUS scheme view.

For information about setting the maximum number of RADIUS packet transmission attempts and the RADIUS server response timeout timer, see "Configuring AAA."

Procedure

1. Enter system view.

   system-view
2. Set the client timeout timer.

   dot1x timer supp-timeout *supp-timeout-value*

   The default is 30 seconds.
3. Set the server timeout timer.

   dot1x timer server-timeout *server-timeout-value*

   The default is 100 seconds.

Configuring 802.1X reauthentication

Restrictions and guidelines

The device selects a periodic reauthentication timer for 802.1X reauthentication in the following order:

1. Server-assigned reauthentication timer.
2. Port-specific reauthentication timer.
3. Global reauthentication timer.

4. Default reauthentication timer.

After you perform a manual reauthentication, the device reauthenticates all online 802.1X users on a port regardless of the server-assigned reauthentication attribute and the periodic reauthentication feature on the port.

Modification to the mandatory authentication domain or EAP message handling method setting does not affect the reauthentication of online 802.1X users. The modified setting takes effect only on 802.1X users that come online after the modification.

If periodic reauthentication is triggered for a user while that user is waiting for online synchronization, the system performs online synchronization and does not perform reauthentication for the user.

Procedure

1. Enter system view.

   system-view
2. Set the periodic reauthentication timer.
   - Set a global periodic reauthentication timer.

     dot1x timer reauth-period *reauth-period-value*

     The default setting is 3600 seconds.
   - Execute the following commands in sequence to set a port-specific periodic reauthentication timer:

     interface *interface-type interface-number*

     dot1x timer reauth-period *reauth-period-value*

     quit

     By default, no periodic reauthentication timer is set on a port. The port uses the global 802.1X periodic reauthentication timer.
3. Enter interface view.

   interface *interface-type interface-number*
4. Enable periodic online user reauthentication.

   dot1x re-authenticate

   By default, the feature is disabled.
5. (Optional.) Manually reauthenticate all online 802.1X users on the port.

   dot1x re-authenticate manual
6. (Optional.) Enable the keep-online feature for 802.1X users.

   dot1x re-authenticate server-unreachable keep-online

   By default, this feature is disabled. The device logs off online 802.1X users if no authentication server is reachable for 802.1X reauthentication.

   Use the keep-online feature according to the actual network condition. In a fast-recovery network, you can use the keep-online feature to prevent 802.1X users from coming online and going offline frequently.

Setting the quiet timer

About this task

The quiet timer enables the access device to wait a period of time before it can process any authentication request from a client that has failed an 802.1X authentication.

Restrictions and guidelines

You can edit the quiet timer, depending on the network conditions.

- In a vulnerable network, set the quiet timer to a high value.
- In a high-performance network with quick authentication response, set the quiet timer to a low value.

Procedure

1. Enter system view.
   `system-view`
2. Enable the quiet timer.
   `dot1x quiet-period`
   By default, the timer is disabled.
3. Set the quiet timer.
   `dot1x timer quiet-period` *`quiet-period-value`*
   The default is 60 seconds.

Configuring an 802.1X guest VLAN

Restrictions and guidelines

- You can configure only one 802.1X guest VLAN on a port. The 802.1X guest VLANs on different ports can be different.
- Assign different IDs to the port VLAN and the 802.1X guest VLAN on a port. The assignment makes sure the port can correctly process incoming VLAN-tagged traffic.
- For the 802.1X guest VLAN feature to work correctly, do not configure this feature together with EAD assistant.
- On a hybrid port, the guest VLAN can only be an untagged VLAN. If the guest VLAN is tagged, you must make it untagged, and then reconfigure it as the guest VLAN.
- When you configure multiple security features on a port, follow the guidelines in Table 9.

Table 9 Relationships of the 802.1X guest VLAN and other security features

Feature	Relationship description	Reference
802.1X Auth-Fail VLAN on a port that performs MAC-based access control	The 802.1X Auth-Fail VLAN has a higher priority than the 802.1X guest VLAN.	See "802.1X VLAN manipulation."
Port intrusion protection actions on a port that performs MAC-based access control	The 802.1X guest VLAN feature has higher priority than the block MAC action. The 802.1X guest VLAN feature has lower priority than the shutdown port action of the port intrusion protection feature.	See "Configuring port security."

Prerequisites

Before you configure an 802.1X guest VLAN, complete the following tasks:

- Create the VLAN to be specified as the 802.1X guest VLAN.
- If the 802.1X-enabled port performs MAC-based access control, perform the following operations for the port:
  - Configure the port as a hybrid port.
  - Enable MAC-based VLAN on the port. For more information about MAC-based VLANs, see *Layer 2—LAN Switching Configuration Guide*.

- If the port type is hybrid, verify that the VLAN to be specified as the guest VLAN is not in the tagged VLAN list on the port.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the 802.1X guest VLAN on the port.
   dot1x guest-vlan *guest-vlan-id*
   By default, no 802.1X guest VLAN exists on a port.

Enabling 802.1X guest VLAN assignment delay

About this task

This feature delays assigning an 802.1X-enabled port to the 802.1X guest VLAN when 802.1X authentication is triggered on the port.

This feature applies only to situations where 802.1X authentication is triggered by EAPOL-Start packets from 802.1X clients or packets from unknown MAC addresses.

To use this feature, the 802.1X-enabled port must perform MAC-based access control. To use the new MAC-triggered 802.1X guest VLAN assignment delay, you must also configure 802.1X unicast trigger on the port.

When 802.1X authentication is triggered on a port, the device performs the following operations:

1. Sends a unicast EAP-Request/Identity packet to the MAC address that triggers the authentication.
2. Retransmits the packet if no response is received within the username request timeout interval set by using the **dot1x timer tx-period** command.
3. Assigns the port to the 802.1X guest VLAN after the maximum number of request attempts set by using the **dot1x retry** command is reached.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable 802.1X guest VLAN assignment delay on the port.
   dot1x guest-vlan-delay { **eapol** | **new-mac** }
   By default, 802.1X guest VLAN assignment delay is disabled on a port.

Configuring an 802.1X Auth-Fail VLAN

Restrictions and guidelines

- Assign different IDs to the port VLAN and the 802.1X Auth-Fail VLAN on a port. The assignment makes sure the port can correctly process VLAN-tagged incoming traffic.
- You can configure only one 802.1X Auth-Fail VLAN on a port. The 802.1X Auth-Fail VLANs on different ports can be different.
- On a hybrid port, the Auth-Fail VLAN can only be an untagged VLAN.

- When you configure multiple security features on a port, follow the following guidelines:
  - On a port that performs MAC-based access control, the 802.1X Auth-Fail VLAN has a high priority over the MAC authentication guest VLAN. For more information about MAC authentication, see "Configuring MAC authentication."
  - For port intrusion protection actions on a port that performs MAC-based access control: The 802.1X Auth-Fail VLAN feature has higher priority than the block MAC action. The 802.1X Auth-Fail VLAN feature has lower priority than the shutdown port action of the port intrusion protection feature.

    For more information about port security, see "Configuring port security."

Prerequisites

Before you configure an 802.1X Auth-Fail VLAN, complete the following tasks:

- Create the VLAN to be specified as the 802.1X Auth-Fail VLAN.
- If the 802.1X-enabled port performs MAC-based access control, perform the following operations for the port:
  - Configure the port as a hybrid port.
  - Enable MAC-based VLAN on the port. For more information about MAC-based VLANs, see *Layer 2—LAN Switching Configuration Guide*.
- If the port type is hybrid, verify that the VLAN to be specified as the Auth-Fail VLAN is not in the tagged VLAN list on the port.

Procedure

1. Enter system view.

   `system-view`
2. Enter interface view.

   `interface` *`interface-type interface-number`*
3. Configure the 802.1X Auth-Fail VLAN on the port.

   `dot1x auth-fail vlan` *`authfail-vlan-id`*

   By default, no 802.1X Auth-Fail VLAN exists on a port.

Configuring an 802.1X critical VLAN

Restrictions and guidelines

- You can configure only one 802.1X critical VLAN on a port. The 802.1X critical VLANs on different ports can be different.
- Assign different IDs to the PVID and the 802.1X critical VLAN on a port. The assignment makes sure the port can correctly process VLAN-tagged incoming traffic.
- If a port configured with port-based access control is already in the 802.1X Auth-Fail VLAN, it will not move to the 802.1X critical VLAN when all authentication servers become unreachable. If all authentication servers are detected unreachable upon reauthentication initiated by a user in the Auth-Fail VLAN on a port configured with MAC-based access control, that user will remain in the Auth-Fail VLAN. The device will not move the user to the 802.1X critical VLAN.
- A port will move to the 802.1X Critical VLAN if all authentication servers are detected unreachable after it has been added to the 802.1X Guest VLAN.
- On a hybrid port, the critical VLAN can only be an untagged VLAN.

Prerequisites

Before you configure an 802.1X critical VLAN, complete the following tasks:

- Create the VLAN to be specified as a critical VLAN.

- If the 802.1X-enabled port performs MAC-based access control, perform the following operations for the port:
  - Configure the port as a hybrid port.
  - Enable MAC-based VLAN on the port. For more information about MAC-based VLANs, see *Layer 2—LAN Switching Configuration Guide*.
- If the port type is hybrid, verify that the VLAN to be specified as the critical VLAN is not in the tagged VLAN list on the port.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the 802.1X critical VLAN on the port.
   dot1x critical vlan *critical-vlan-id*
   By default, no 802.1X critical VLAN exists on a port.

Configuring an 802.1X guest VSI

Restrictions and guidelines

You can configure only one 802.1X guest VSI on a port. The 802.1X guest VSIs on different ports can be different.

For the 802.1X guest VSI feature to work correctly, do not configure this feature together with EAD assistant.

Prerequisites

To ensure that the 802.1X guest VSI on an 802.1X-enabled port can take effect, complete the following tasks:

- Enable L2VPN.
- Create the VSI to be specified as the 802.1X guest VSI, and create a VXLAN for the VSI.
- If the port performs MAC-based access control, enable MAC-based traffic match mode for dynamic Ethernet service instances on the port.

For more information, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the 802.1X guest VSI on the port.
   dot1x guest-vsi *guest-vsi-name*
   By default, no 802.1X guest VSI exists on a port.

Enabling 802.1X guest VSI assignment delay

About this task

This feature delays assigning an 802.1X-enabled port to the 802.1X guest VSI when 802.1X authentication is triggered on the port.

This feature applies only to situations where 802.1X authentication is triggered by EAPOL-Start packets from 802.1X clients or packets from unknown MAC addresses.

To use this feature, the 802.1X-enabled port must perform MAC-based access control.

When 802.1X authentication is triggered on a port, the device performs the following operations:

1. Sends a unicast EAP-Request/Identity packet to the MAC address that triggers the authentication.
2. Retransmits the packet if no response is received within the username request timeout interval set by using the **dot1x timer tx-period** command.
3. Assigns the port to the 802.1X guest VSI after the maximum number of request attempts set by using the **dot1x retry** command is reached.

This feature can work with the parallel processing of MAC authentication and 802.1X authentication feature when a port performs a combination of 802.1X and MAC authentication. The collaboration facilitates the port to perform MAC authentication before it is assigned to the 802.1X guest VSI. For information about the parallel processing of MAC authentication and 802.1X authentication feature, see "Configuring MAC authentication."

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable 802.1X guest VSI assignment delay on the port.

   dot1x guest-vsi-delay { **eapol** | **new-mac** }

   By default, 802.1X guest VSI assignment delay is disabled on a port.

Configuring an 802.1X Auth-Fail VSI

Restrictions and guidelines

You can configure only one 802.1X Auth-Fail VSI on a port. The 802.1X Auth-Fail VSIs on different ports can be different.

Prerequisites

To ensure that the 802.1X Auth-Fail VSI on an 802.1X-enabled port can take effect, complete the following tasks:

- Enable L2VPN.
- Create the VSI to be specified as the 802.1X Auth-Fail VSI, and create a VXLAN for the VSI.
- If the port performs MAC-based access control, enable MAC-based traffic match mode for dynamic Ethernet service instances on the port.

For more information, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the 802.1X Auth-Fail VSI on the port.
   dot1x auth-fail vsi *authfail-vsi-name*
   By default, no 802.1X Auth-Fail VSI exists on a port.

Configuring an 802.1X critical VSI

Restrictions and guidelines

You can configure only one 802.1X critical VSI on a port. The 802.1X critical VSIs on different ports can be different.

The 802.1X critical VSI on a port does not take effect on users already in the 802.1X Auth-Fail VSI.

Prerequisites

To ensure that the 802.1X critical VSI on an 802.1X-enabled port can take effect, complete the following tasks:

- Enable L2VPN.
- Create the VSI to be specified as the 802.1X critical VSI, and create a VXLAN for the VSI.
- If the port performs MAC-based access control, enable MAC-based traffic match mode for dynamic Ethernet service instances on the port.

For more information, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the 802.1X critical VSI on the port.
   dot1x critical vsi *critical-vsi-name*
   By default, no 802.1X critical VSI exists on a port.

Configuring an 802.1X critical microsegment

Restrictions and guidelines

You can configure only one 802.1X critical microsegment on a port. The 802.1X critical microsegments on different ports can be the same or different.

You can specify only one VSI for the critical microsegment on a port. The VSI specified for the critical microsegment does not take effect on users already in a VSI.

The critical microsegment does not take effect on a port that performs port-based access control.

For more information, see the command reference and "Restrictions and guidelines: 802.1X configuration."

Prerequisites

For the 802.1X critical microsegment you specify on a port to take effect, you must create that microsegment. For more information about microsegmentation, see "Configuring microsegmentation."

To specify a VSI fror the critical microsegment on a port, perform the following tasks:

- Enable L2VPN.
- If the port performs MAC-based access control, enable MAC-based traffic match mode for dynamic Ethernet service instances on the port.

For more information, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the 802.1X critical microsegment on the port.
   dot1x critical microsegment *microsegment-id* [**vsi** *vsi-name*]
   By default, no 802.1X critical microsegment exists on a port.

Specifying a critical profile for 802.1X users on a port

About this task

A critical profile specifies a limited set of resources accessible to 802.1X users when they cannot get authenticated because none of the applicable RADIUS servers is available.

To make sure the 802.1X users attached to a port can access the resources in a critical profile in the server unavailable situation, perform this task.

Restrictions and guidelines

You can specify only one critical profile for 802.1X authentication on a port. The critical profiles for 802.1X authentication on different ports can be different.

On an interface, the 802.1X critical profile settings are mutually exclusive with the critical VSI, critical VLAN, and critical microsegment settings.

Prerequisites

As a best practice, configure the critical profile before you perform this task. For more information about configuring critical profiles, see "Configuring AAA."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify a critical profile for 802.1X users on the port.
   dot1x critical profile *profile-name*
   By default, no critical profile is specified for 802.1X users on a port.

Configuring an 802.1X access profile

About this task

The device uses 802.1X access profiles to manage 802.1X access settings in a unified manner. After you bind an 802.1X access profile to a port security authentication profile, the interfaces to which the port security authentication profile is applied will use the settings in the 802.1X access profile to authenticate access users.

After you bind a port security authentication profile to an interface:

- For settings that can be configured both in interface view and 802.1X access profile view (the command lines might differ in the views), the system automatically deletes the settings in interface view regardless of whether the port security authentication profile has been bound to an 802.1X access profile. To use these settings, configure them in the 802.1X access profile bound to the port security authentication profile.
- For settings that can be configured in interface view but cannot be configured in 802.1X access profile view, they can still take effect after you apply an 802.1X access profile to that interface.

For more information about port security authentication profiles, see "Configuring port security."

Restrictions and guidelines

Deleting the 802.1X access profile bound to a port security authentication profile will cause online users to go offline abnormally if the access profile has taken effect.

An 802.1X access profile can be bound to multiple port security authentication profiles, but a port security authentication profile can bind only one 802.1X access profile. To change the 802.1X access profile bound to a port security authentication profile, you must first remove the original binding.

Procedure

1. Enter system view.

   system-view
2. Create an 802.1X access profile and enter its view.

   dot1x access-profile name *profile-name*

   By default, no 802.1X access profiles exist.
3. Configure 802.1X features in the 802.1X access profile.

   For information about the features that can be configured in an 802.1X access profile, see 802.1X commands in the command reference.
4. Enter port security authentication profile view.

   port-security authentication-profile name *profile-name*
5. Bind the 802.1X access profile to the port security authentication profile.

   dot1x access-profile *profile-name*

   By default, no 802.1X access profile is bound to a port security authentication profile.

Configuring 802.1X unauthenticated user aging

About this task

802.1X unauthenticated user aging applies to users added to the following access control segments because they have not been authenticated or have failed authentication:

- 802.1X guest, Auth-Fail, or critical VLANs.
- 802.1X guest, Auth-Fail, or critical VSIs.
- 802.1X critical microsegments.

When a user in one of those access control segments ages out, the device removes the user from the segment and deletes the MAC address entry for the user from the access port.

For users in one of those segments on one port to be authenticated successfully and come online on another port, enable this feature. In any other scenarios, disable this feature as a best practice.

The 802.1X user aging mechanism on a port depends on its access control mode.

- If the port uses port-based access control, an 802.1X unauthenticated user aging timer starts when the port is assigned to one of the following access control segments:
  - Critical microsegment.
  - Auth-Fail VLAN or VSI.
  - Critical VLAN or VSI.

  When the aging timer expires, the port is removed from the segment and all MAC address entries for users in the segment are also removed.
- If the port uses MAC-based access control, an 802.1X unauthenticated user aging timer starts for each 802.1X user when they are assigned to one of the following access control segments:
  - Auth-Fail VLAN or VSI.
  - Critical VLAN or VSI.
  - Guest VLAN or VSI.

  When the aging timer for a user expires, the device removes that user from the segment.

The removed users will be unable to access any network resources until after another authentication is triggered.

Restrictions and guidelines

As a best practice, use this feature on a port only if you want to have its unauthenticated users to be authenticated and come online on a different port.

Procedure

1. Enter system view.

   system-view
2. Set the user aging timer for a type of 802.1X microsegment, VLAN, or VSI.

   dot1x timer user-aging { **auth-fail-vlan** | **auth-fail-vsi** | **critical-microsegment** | **critical-vlan** | **critical-vsi** | **guest-vlan** | **guest-vsi** } *aging-time-value*

   By default, the user aging timers for all applicable types of 802.1X microsegments, VLANs, and VSIs are 1000 seconds.
3. Enter interface view.

   interface *interface-type interface-number*
4. Enable 802.1X unauthenticated user aging.

   dot1x unauthenticated-user aging enable

   By default, 802.1X unauthenticated user aging is enabled.

Sending EAP-Success packets on assignment of users to the 802.1X Auth-Fail VLAN or VSI

About this task

By default, the device sends EAP-Failure packets to 802.1X clients when the client users are assigned to the 802.1X Auth-Fail VLAN or VSI on a port. However, some 802.1X clients cannot send DHCP requests for IP addresses after they receive EAP-Failure packets.

To have these clients obtain IP addresses and access resources after they are assigned to the 802.1X Auth-Fail VLAN or VSI, perform this task.

This task enables the device to send EAP-Success packets instead of EAP-Failure packets to 802.1X clients when the client users are assigned to the 802.1X Auth-Fail VLAN or VSI on the port.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the device to send an EAP-Success packet to a client when the client user is assigned to the 802.1X Auth-Fail VLAN or VSI on the port.

   dot1x auth-fail eapol

   By default, the device sends an EAP-Failure packet to a client when the client user is assigned to the 802.1X Auth-Fail VLAN or VSI.

Sending EAP-Success packets on assignment of users to the 802.1X critical VLAN or VSI

About this task

By default, the device sends EAP-Failure packets to 802.1X clients when the client users are assigned to the 802.1X critical VLAN or VSI. Some 802.1X clients cannot respond to the EAP-Request/Identity packet from the device for reauthentication if they have received an EAP-Failure packet. As a result, reauthentication for these clients will fail after the authentication server becomes reachable.

To avoid this situation, enable the device to send EAP-Success packets instead of EAP-Failure packets to 802.1X clients when the client users are assigned to the 802.1X critical VLAN or VSI.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the device to send an EAP-Success packet to an 802.1X client when its client user is assigned to the critical VLAN or VSI on the port.

   dot1x critical eapol

   By default, the device sends an EAP-Failure packet to an 802.1X client when its client user is assigned to the critical VLAN or VSI.

Enabling 802.1X online user synchronization

About this task

IMPORTANT:

This feature takes effect only when the device uses an IMC RADIUS server to authenticate 802.1X users.

To ensure that the RADIUS server maintains the same online 802.1X user information as the device after the server state changes from unreachable to reachable, use this feature.

This feature synchronizes online 802.1X user information between the device and the RADIUS server when the RADIUS server state is detected having changed from unreachable to reachable.

When synchronizing online 802.1X user information on a port with the RADIUS server, the device initiates 802.1X authentication in turn for each authenticated online 802.1X user to the RADIUS server.

If synchronization fails for an online user, the device logs off that user unless the failure occurs because the server has become unreachable again.

Restrictions and guidelines

The amount of time required to complete online user synchronization increases as the number of online users grows. This might result in an increased delay for new 802.1X users and users in the critical VLAN or VSI to authenticate or reauthenticate to the RADIUS server and come online.

To have this feature take effect, you must use it in conjunction with the RADIUS server status detection feature, which is configurable with the **radius-server test-profile** command. When you configure this feature, make sure the detection interval is shorter than the RADIUS server quiet timer configured by using the **timer quiet** command in RADIUS scheme view. The server state changes to active on expiration of the quiet timer regardless of its actual reachability. Setting a shorter detection interval than the quiet timer prevents the RADIUS server status detection feature from falsely reporting the server reachability.

For more information about the RADIUS server status detection feature, see "Configuring AAA."

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable 802.1X online user synchronization.

   dot1x server-recovery online-user-sync

   By default, 802.1X online user synchronization is disabled.

Configuring the authentication trigger feature

About this task

The authentication trigger feature enables the access device to initiate 802.1X authentication when 802.1X clients cannot initiate authentication.

This feature provides the multicast trigger and unicast trigger (see 802.1X authentication initiation in "802.1X overview").

Restrictions and guidelines

- Enable the multicast trigger on a port when the clients attached to the port cannot send EAPOL-Start packets to initiate 802.1X authentication.
- As a best practice to conserve link bandwidth, disable the multicast trigger if a lot of VLANs are configured on the port.
- Enable the unicast trigger on a port if only a few 802.1X clients are attached to the port and these clients cannot initiate authentication.
- To avoid duplicate authentication packets, do not enable both triggers on a port.

- As a best practice, do not use the unicast trigger on a port that performs port-based access control. If you do so, users on the port might fail to come online.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Set the username request timeout timer.
   dot1x timer tx-period *tx-period-value*
   The default is 30 seconds.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable an authentication trigger.
   dot1x { **multicast-trigger** | **unicast-trigger** }
   By default, the multicast trigger is enabled and the unicast trigger is disabled.

Setting the maximum number of concurrent 802.1X users on a port

About this task

Perform this task to prevent the system resources from being overused.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the maximum number of concurrent 802.1X users on a port.
   dot1x max-user [**preauth-domain** | **auth-fail-domain**] *max-number*
   The default is 4294967295.

Setting the alarm threshold and alarm clear threshold for 802.1X user access ratio

About this task

The 802.1X user access ratio on an interface is the percentage of current online 802.1X users to the maximum number of concurrent 802.1X users on the interface (configured by using the **dot1x max-user** command).

After you use the **snmp-agent trap enable dot1x** command to enable SNMP notifications for 802.1X user access ratio and use the **dot1x max-user-alarm** command to set the alarm threshold and alarm clear threshold:

- If the percentage of 802.1X user access ratio on an interface reaches the specified alarm threshold for the first time, or increases to the alarm threshold from a value below or equal to the alarm clear threshold, the device sends an alarm notification.
- If the percentage of 802.1X user access ratio on the interface drops to the alarm clear threshold from a value above or equal to the alarm threshold, the device sends an alarm clear notification.

If the user access ratio on an interface has reached or exceeded the alarm threshold before SNMP notifications is enabled, the alarm is triggered in the following conditions after SNMP notifications is enabled:

- A user comes online and the alarm threshold is exceeded.
- A user goes offline and the alarm threshold is reached or exceeded.

Procedure

1. Enter system view.

   system-view

2. Set the alarm threshold and alarm clear threshold for 802.1X user access ratio in percentage.

   dot1x max-user-alarm trigger-threshold *trigger-threshold* **clear-threshold** *clear-threshold*

   By default, the alarm threshold and alarm clear threshold for 802.1X user access ratio in percentage is 100 and 50, respectively.

Setting the maximum number of authentication request attempts

About this task

The access device retransmits an authentication request if it does not receive any responses to the request from the client within a period of time. To set the time, use the **dot1x timer tx-period** *tx-period-value* command or the **dot1x timer supp-timeout** *supp-timeout-value* command. The access device stops retransmitting the request if it has made the maximum number of request transmission attempts but still receives no response.

Procedure

1. Enter system view.

   system-view

2. Set the maximum number of attempts for sending an authentication request.

   dot1x retry *retries*

   The default setting is 2.

Sending EAP-Success packets to 802.1X clients after successful authorization

About this task

By default, the device sends an EAP-Success packet to a client after it receives a RADIUS Access-Accept packet sent by the RADIUS server for the user of that client. After that, the device grants the user access to the network through the user's access port. Upon receiving the EAP-Success packet, the client immediately initiates a DHCP request to obtain an IP address. However, if the device has not yet received authorization information assigned by the RADIUS server, the client directly obtains an IP address from its access network segment. If its access network segment is different from its authorization network segment after the RADIUS server assigns the authorization information, the client will be unable to come online.

To ensure that each 802.1X client obtains an IP address from the authorization network segment, use this feature to configure the device to send EAP-Success packets to 802.1X clients after authorization for the users of the clients succeeds.

Restrictions and guidelines

When this feature is configured, authentication for some users might fail due to untimely authorization if the RADIUS server or device needs to process a large number of authentication processes, resulting in authentication response timeout. As a best practice, use this feature based on the service processing capabilities of the device and the RADIUS server in the actual application scenario.

Procedure

1. Enter system view.

   system-view

2. Configure the device to send EAP-Success packets to 802.1X clients after authorization for the users of the clients succeeds.

   dot1x eap-success post-authorization

   By default, the device sends EAP-Success packets to 802.1X clients immediately after authentication for the users of the clients succeeds.

Discarding duplicate 802.1X EAPOL-Start requests

About this task

During 802.1X authentication, the device might receive duplicate EAPOL-Start requests from an 802.1X user. By default, the device delivers the duplicate EAPOL-Start requests to the authentication server as long as they are legal. However, this mechanism might result in authentication failure if the authentication server cannot respond to duplicate EAPOL-Start requests. To resolve this issue, perform this task on the user access interface to discard duplicate EAPOL-Start requests.

Restrictions and guidelines

As a best practice, perform this task only if the server cannot respond to duplicate EAPOL-Start requests. Do not perform this task in other situations.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Discard duplicate EAPOL-Start requests on the interface.

   dot1x duplicate-eapol-start discard

   By default, the device does not discard duplicate EAPOL-Start requests on an interface if the requests are legal.

Configuring online user handshake

About this task

The online user handshake feature checks the connectivity status of online 802.1X users. The access device sends handshake requests (EAP-Request/Identity) to online users at the interval specified by the **dot1x timer handshake-period** command. If the device does not receive any EAP-Response/Identity packets from an online user after it has made the maximum handshake

attempts, the device sets the user to offline state. To set the maximum handshake attempts, use the **dot1x retry** command.

Typically, the device does not reply to 802.1X clients' EAP-Response/Identity packets with EAP-Success packets. Some 802.1X clients will go offline if they do not receive the EAP-Success packets for handshake. To avoid this issue, enable the online user handshake reply feature.

If iNode clients are deployed, you can also enable the online user handshake security feature to check authentication information in the handshake packets from clients. This feature can prevent 802.1X users that use illegal client software from bypassing iNode security check, such as dual network interface cards (NICs) detection. If a user fails the handshake security checking, the device sets the user to the offline state.

Restrictions and guidelines

- If the network has 802.1X clients that cannot exchange handshake packets with the access device, disable the online user handshake feature. This operation prevents the 802.1X connections from being incorrectly torn down.
- To use the online user handshake security feature, make sure the online user handshake feature is enabled.
- The online user handshake security feature takes effect only on the network where the iNode client and IMC server are used.
- Enable the online user handshake reply feature only if 802.1X clients will go offline without receiving EAP-Success packets from the device.
- To ensure online user handshake and new user authentication when a large number of users are present, set the following parameters to a large value:
  - Handshake timer (set by using the **dot1x timer handshake-period** command).
  - Maximum number of attempts to send an authentication request to a client (set by using the **dot1x retry** command).

Procedure

1. Enter system view.

   system-view

2. (Optional.) Set the handshake timer.

   dot1x timer handshake-period *handshake-period-value*

   The default is 15 seconds.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable the online user handshake feature.

   dot1x handshake

   By default, the feature is enabled.

5. (Optional.) Enable the online user handshake security feature.

   dot1x handshake secure

   By default, the feature is disabled.

6. (Optional.) Enable the 802.1X online user handshake reply feature.

   dot1x handshake reply enable

   By default, the device does not reply to 802.1X clients' EAP-Response/Identity packets during the online handshake process.

Configuring 802.1X offline detection

About this task

The 802.1X offline detection feature monitors the online status of 802.1X users. This feature uses an offline detection timer to set the interval that the device must wait for traffic from a user before the device determines that the user is idle. If the device has not received traffic from a user before the timer expires, the device logs off that user and requests the accounting server to stop accounting for the user.

The traffic monitored by this feature includes 802.1X protocol packets sent from 802.1X clients.

This feature takes effect on both of the following types of users:

- Users in preauthentication domains, authentication domains, critical domains, and Auth-Fail domains.
- Users assigned to guest VLANs or VSIs, Auth-Fail VLANs or VSIs, and critical VLANs, VSIs, or microsegments.

Restrictions and guidelines

The 802.1X offline detection feature takes effect only on a port that performs MAC-based access control. If you change the port access mode to port-based, the 802.1X offline detection feature cannot take effect.

If you enable 802.1X offline detection on a Layer 2 aggregate interface, delay exists for the device to log off an idle user.

For this feature to operate as expected, do not set the offline detection timer to the same value as either of the following timers:

- Handshake timer (set by using the **dot1x timer handshake-period** command).
- Periodic reauthentication timer (set by using the **dot1x timer reauth-period** command).

Procedure

1. Enter system view.
   system-view
2. Set the 802.1X offline detection timer.
   dot1x timer offline-detect *offline-detect-value*
   By default, the 802.1X offline detection timer is 300 seconds.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable 802.1X offline detection.
   dot1x offline-detect enable
   By default, 802.1X offline detection is disabled.

Enabling online 802.1X users to escape from offline detection and stay online when no reachable RADIUS authentication servers are available

About this task

The 802.1X offline detection feature logs off an online 802.1X user if it has not detected traffic from that user before the offline-detect timer expires. To prevent online 802.1X users on a port from being forced offline by the 802.1X offline detection feature and stay online when none of the RADIUS servers in their ISP domain is reachable, perform this task.

Restrictions and guidelines

Use this feature only when RADIUS scheme authentication is configured for the 802.1X users on a port. If local or no authentication is also configured, offline detection might be undesirably disabled for the 802.1X users when they come online through local authentication or without authentication.

Procedure

1. Enter system view.

   `system-view`

2. Enter interface view.

   `interface` *`interface-type interface-number`*

3. Enable online 802.1X users to escape from 802.1X offline detection and stay online when no reachable RADIUS authentication servers are available.

   `dot1x auth-server-unavailable escape`

   By default, online 802.1X users cannot escape from 802.1X offline detection (if enabled) when no reachable RADIUS authentication servers are available.

Specifying supported domain name delimiters

About this task

By default, the access device supports the at sign (@) as the delimiter. You can also configure the access device to accommodate 802.1X users that use other domain name delimiters. The configurable delimiters include the at sign (@), backslash (\), dot (.), and forward slash (/). Usernames that include domain names can use the format of *username***@***domain-name*, *domain-name******username*, *username***.***domain-name*, or *username***/***domain-name*.

If an 802.1X username string contains multiple configured delimiters, the rightmost delimiter is the domain name delimiter. For example, if you configure the backslash (\), dot (.), and forward slash (/) as delimiters, the domain name delimiter for the username string 121.123/22\@abc is the backslash (\). The username is @**abc** and the domain name is **121.123/22**.

Restrictions and guidelines

If a username string contains none of the delimiters, the access device authenticates the user in the mandatory or default ISP domain.

If you configure the access device to send usernames with domain names to the RADIUS server, make sure the domain delimiter can be recognized by the RADIUS server. For username format configuration, see the **`user-name-format`** command in *Security Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Specify a set of domain name delimiters for 802.1X users.

   dot1x domain-delimiter *string*

   By default, only the at sign (@) delimiter is supported.

Removing the VLAN tags of 802.1X protocol packets sent out of a port

About this task

This feature operates on a hybrid port to have it send 802.1X protocol packets with their VLAN tags removed, regardless of whether the port is a tagged or untagged member of a VLAN.

Use this feature if the 802.1X-enabled hybrid port is a tagged member of its PVID and the attached 802.1X clients cannot recognize VLAN-tagged 802.1X protocol packets.

Restrictions and guidelines

This feature removes the VLAN tags of all 802.1X protocol packets sent out of the port to 802.1X clients. Do not use this feature if VLAN-aware 802.1X clients are attached to the port. As a best practice, perform this task only in the described applicable scenario.

Prerequisites

Set the link type of the 802.1X-enabled port to hybrid. For more information, see VLAN configuration in *Layer 2 LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Remove the VLAN tags of all 802.1X protocol packets sent out of the port to 802.1X clients.

   dot1x eapol untag

   By default, whether the device removes the VLAN tags of all 802.1X protocol packets sent out of a port to 802.1X clients depends on the configuration in the VLAN module.

CAUTION:

This command removes the VLAN tags of all 802.1X protocol packets sent out of the port to 802.1X clients. Do not use this command if VLAN-aware 802.1X clients are attached to the port. As a best practice, use this command only in the described applicable scenario.

Setting the maximum number of 802.1X authentication attempts for MAC authenticated users

About this task

When a port uses both 802.1X authentication and MAC authentication, the device accepts 802.1X authentication requests from MAC authenticated users. If a MAC authenticated user passes 802.1X authentication, the user will come online as an 802.1X user. If the user fails 802.1X authentication, the user continues to make 802.1X authentication attempts depending on client configuration.

Perform this task to limit the number of 802.1X authentication attempts made by a MAC authenticated user.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set the maximum number of 802.1X authentication attempts for MAC authenticated users on the port.

   dot1x after-mac-auth max-attempt *max-attempts*

   By default, the number of 802.1X authentication attempts for MAC authenticated users is not limited on a port.

Enabling 802.1X user IP freezing

About this task

This feature works with the IP source guard feature. 802.1X-based IP source guard requires that 802.1X clients support sending user IP addresses to the access device. The device uses information such as user MAC addresses and IP addresses obtained through 802.1X to generate IPSG bindings to filter out IPv4 packets from unauthenticated 802.1X users. For information about IP source guard, see "Configuring IP source guard."

This feature prevents any authenticated 802.1X users on a port from changing their IP addresses. After you enable this feature, the port does not update the IP addresses in dynamic IPSG bindings for 802.1X users. If an 802.1X user uses an IP address different from the IP address in its IPSG binding entry, the port denies the user access.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable 802.1X user IP freezing.

   dot1x user-ip freeze

   By default, 802.1X user IP freezing is disabled.

Enabling generation of dynamic IPSG binding entries for 802.1X authenticated users

About this task

IMPORTANT:

This feature must operate in conjunction with the IP source guard (IPSG) feature.

By default, the device generates a dynamic IPv4SG or IPv6SG binding entry for an 802.1X authenticated user after the user obtains a static or DHCP assigned IP address.

To allow only 802.1X users with DHCP assigned IP addresses to access the network, perform the following operations:

- Enable IPSG.
- Disable generation of dynamic IPv4SG or IPv6SG binding entries for 802.1X authenticated users.
- Enable DHCP snooping. The device will generate IPv4SG or IPv6SG binding entries for the users based on DHCP snooping.

For more information about IPSG, see IP source guard in *Security Configuration Guide*.

Restrictions and guidelines

This feature takes effect only on 802.1X users that come online after the feature is enabled. If the IP address of an online 802.1X user changes, the device will update the dynamic IPv4SG or IPv6SG binding entry for the user.

Disabling this feature does not delete the existing dynamic IPv4SG or IPv6SG binding entries for online 802.1X users. If the IP address of an online 802.1X user changes after the feature is disabled, the device will delete the dynamic IPv4SG or IPv6SG binding entry for the user.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable generation of dynamic IPv4SG or IPv6SG binding entries for 802.1X authenticated users.

   dot1x { **ip-verify-source** | **ipv6-verify-source** } **enable**

   By default, generation of dynamic IPv4SG or IPv6SG binding entries for 802.1X authenticated users is enabled.

Configuring 802.1X MAC address binding

About this task

This feature can automatically bind MAC addresses of authenticated 802.1X users to the users' access port and generate 802.1X MAC address binding entries. You can also use the **dot1x mac-binding** *mac-address* command to manually add 802.1X MAC address binding entries.

802.1X MAC address binding entries never age out. They can survive a user logoff or a device reboot. If users in the 802.1X MAC address binding entries perform 802.1X authentication on another port, they cannot pass authentication.

Restrictions and guidelines

The 802.1X MAC address binding feature takes effect only when the port performs MAC-based access control.

To delete an 802.1X MAC address binding entry, use the **undo dot1x mac-binding** *mac-address* command. An 802.1X MAC address binding entry cannot be deleted when the user in the entry is online.

After the number of 802.1X MAC address binding entries reaches the upper limit of concurrent 802.1X users (set by using the **dot1x max-user** command), the following restrictions exist:

- Users not in the binding entries will fail authentication even after users in the binding entries go offline.
- New 802.1X MAC address binding entries are not allowed.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the 802.1X MAC address binding feature.
   dot1x mac-binding enable
   By default, the feature is disabled.
4. (Optional.) Manually add an 802.1X MAC address binding entry.
   dot1x mac-binding *mac-address*
   By default, no 802.1X MAC address binding entries exist on a port.

Configuring the EAD assistant feature

Restrictions and guidelines

- If you enable EAD assistant in conjunction with port security, EAD assistant cannot take effect on ports operating in autoLearn port security mode.
- To make the EAD assistant feature take effect on an 802.1X-enabled port, you must set the port authorization mode to **auto**.
- When port security is enabled, the free IPs cannot take effect.
- EAD assistant might cause malfunction of 802.1X Auth-Fail VLANs or VSIs and critical VLANs or VSIs. To ensure that users can correctly access Auth-Fail VLANs or VSIs and critical VLANs or VSIs, use the **dot1x ead-assistant permit authentication-escape** command.
- The **dot1x ead-assistant permit authentication-escape** command cannot take effect on ports in port-based access control.
- If you use free IP and Auth-Fail VLAN features together, make sure the resources in the Auth-Fail VLAN are on the free IP segments.
- The server that provides the redirect URL must be on the free IP accessible to unauthenticated users.
- If you configure both EAD assistant and 802.1X unicast trigger on a port that performs port-based access control, 802.1X unicast trigger cannot take effect.

When you use both EAD assistant and MAC authentication on the device, follow these restrictions and guidelines:

- If both EAD assistant and MAC authentication are configured, the device does not mark the MAC address of a user that has failed MAC authentication as a silent MAC address. If the user

has never passed MAC authentication, packets from the user can trigger MAC authentication again only after the user's EAD entry ages out.

- As a best practice, do not configure MAC authentication guest VSIs, guest VLANs, critical VSIs, or critical VLANs. The VLANs or VSIs might fail to work correctly when both EAD assistant and MAC authentication are configured on the device.
- As a best practice, do not configure the Web authentication or IP source guard feature. These features might fail to work correctly when both EAD assistant and MAC authentication are configured on the device.
- If the MAC address of a user has been marked as a silent MAC address before you enable EAD assistant, packets from the user can trigger 802.1X or MAC authentication only after the quiet timer expires.

Procedure

1. Enter system view.

   system-view

2. Enable the EAD assistant feature.

   dot1x ead-assistant enable

   By default, this feature is disabled.

3. Configure a free IP.

   dot1x ead-assistant free-ip *ip-address* { *mask-length* | *mask-address* }

   Repeat this command to configure multiple free IPs.

4. Configure a free microsegment.

   dot1x ead-assistant free-microsegment *free-microsegment-id*

   Repeat this command to configure a maximum of 20 free microsegments.

5. (Optional.) Configure the redirect URL if users will use Web browsers to access the network.

   dot1x ead-assistant url *url-string* [**secondary**] [**track** *track-entry-number*]

   By default, no redirect URL exists.

   By default, the device listens to port 6654 for HTTPS requests to be redirected. To change the redirect listening port number, see configuring HTTP redirect in *Layer 3—IP Services Configuration Guide*.

6. (Optional.) Set the EAD rule timer.

   dot1x timer ead-timeout *ead-timeout-value*

   The default setting is 30 minutes.

   To avoid using up ACL resources when a large number of EAD users exist, you can shorten the EAD rule timer.

7. (Optional.) Enable support for correctly using Auth-Fail VLANs or VSIs and critical VLANs or VSIs when 802.1X EAD assistant is enabled.

   dot1x ead-assistant permit authentication-escape

   By default, 802.1X Auth-Fail VLANs or VSIs and critical VLANs or VSIs cannot function correctly when 802.1X EAD assistant is enabled.

Setting the maximum size of EAP-TLS fragments sent to the server

About this task

When the device uses EAP-TLS authentication method in EAP relay mode, the RADIUS packets might exceed the maximum packet size supported by the RADIUS server. This situation typically occurs when long EAP-TLS messages are encapsulated in the EAP-Message attribute of the RADIUS packet sent to the RADIUS server.

To avoid authentication failures caused by oversized packets, fragment the EAP-TLS messages depending on the maximum RADIUS packet size supported by the remote RADIUS server.

For example, the maximum packet length allowed by the server is 1200 bytes and the length of a RADIUS packet (excluding the EAP-Message attribute) is 800 bytes. To make sure the maximum length of a RADIUS packet does not exceed 1200 bytes, you must set the maximum length of an EAP-TLS fragment to a value less than 400 bytes.

Restrictions and guidelines

802.1X EAP-TLS fragmentation takes effect only when EAP relay mode is used. For more information about enabling EAP relay, see "Enabling EAP relay or EAP termination."

Procedure

1. Enter system view.

   system-view
2. Enable 802.1X EAP-TLS fragmentation and set the maximum EAP-TLS fragment size.

   dot1x eap-tls-fragment to-server *eap-tls-max-length*

   By default, EAP-TLS messages are not fragmented.

Setting the unicast-trigger quiet timer

About this task

To suppress the unicast trigger from requesting an unknown source MAC address for 802.1X authentication after it has made the maximum number of failed attempts, perform this task.

The 802.1X unicast-trigger quiet timer sets the duration that the unicast trigger is suppressed from sending EAP-Request/Identity requests to an unknown MAC address. This timer starts when the device fails to receive a response from an unknown MAC address after the number of unicast EAP-Request/Identity requests sent to that MAC address has reached the limit. The limit is configurable with the **dot1x retry** command.

Restrictions and guidelines

The network access delay introduced by the unicast-trigger quiet timer increases as the timer value increases. When you set the unicast-trigger quiet timer, make sure its value does not introduce undesired delay for 802.1X authentication users to come online.

You cannot use the unicast-trigger quiet timer in conjunction with the guest VLAN, guest VSI, and EAD assistant feature.

Procedure

1. Enter system view.

   system-view
2. Set the unicast-trigger quiet timer.

   dot1x timer unicast-trigger quiet-period *quiet-period-value*

By default, the unicast-trigger timer is set to 0 seconds, which indicates that the unicast-trigger timer is not enabled.

Logging off 802.1X users

About this task

Perform this task to log off specified 802.1X users and clear information about these users from the device. These users must perform 802.1X authentication to come online again.

Procedure

To log off 802.1X users, execute the following command in user view:

reset dot1x access-user [**interface** *interface-type interface-number* | **mac** *mac-address* | **microsegment** *microsegment-id* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **username** *username* | **vlan** *vlan-id* | **vsi** *vsi-name*]

Enabling 802.1X user logging

About this task

This feature enables the device to generate logs about 802.1X users and send the logs to the information center. For the logs to be output correctly, you must also configure the information center on the device. For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

To prevent excessive 802.1X user log entries, use this feature only if you need to analyze abnormal 802.1X user logins or logouts.

Procedure

1. Enter system view.

   system-view

2. Enable 802.1X user logging.

   dot1x access-user log enable [**abnormal-logoff** | **failed-login** | **ip-change** | **maxnum-clear-threshold** | **maxnum-trigger-threshold** | **normal-logoff** | **successful-login**] *

   By default, 802.1X user logging is disabled.

   If you do not specify any parameters, this command enables all types of 802.1X user logs.

Configuring SNMP notifications for 802.1X

About this task

This feature enables generating SNMP notifications for the following 802.1X events:

- If you specify the **dot1x-not-support** keyword, the system generates a notification when an interface does not support 802.1X.
- If you specify the **resource-not-enough** keyword, the system generates a notification when ACL resources become insufficient.
- If you specify the **unicast-not-support** keyword, the system generates a notification when an interface does not support unicast trigger.

- If you specify the **user-maxnum** keyword, the system generates a notification for a 802.1X user access ratio event.

After you enable SNMP notifications for 802.1X, the 802.1X module notifies the NMS of important module events. The notifications are sent to the device's SNMP module. For the notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP notifications, see *Network Management and Monitoring Configuration Guide*.

After you configure this feature to enable SNMP notifications for 802.1X user access ratio and use the **dot1x max-user-alarm** command to set the alarm threshold and alarm clear threshold:

- If the percentage of 802.1X user access ratio on an interface reaches the specified alarm threshold for the first time, or increases to the alarm threshold from a value below or equal to the alarm clear threshold, the device sends an alarm notification.
- If the percentage of 802.1X user access ratio on the interface drops to the alarm clear threshold from a value above or equal to the alarm threshold, the device sends an alarm clear notification.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for 802.1X.

   snmp-agent trap enable dot1x [**dot1x-not-support** | **resource-not-enough** | **unicast-not-support** | **user-maxnum**] *

   By default, SNMP notifications for 802.1X are disabled.

Display and maintenance commands for 802.1X

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display 802.1X session information, statistics, or configuration information of specified or all ports.	**display dot1x** [**sessions** | **statistics**] [**interface** *interface-type interface-number*]
Display information about 802.1X access profiles.	**display dot1x access-profile** [**name** *profile-name*]
Display online 802.1X user information.	**display dot1x connection** [**open**] [[**m-lag** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **slot** *slot-number* | **user-name** *name-string*] | **user-mac** *mac-address*]
Display the MAC addresses of 802.1X users in a type of 802.1X microsegment, VLAN, or VSI.	**display dot1x mac-address** { **auth-fail-vlan** | **auth-fail-vsi** | **critical-microsegment** | **critical-vlan** | **critical-vsi** | **guest-vlan** | **guest-vsi** } [**interface** *interface-type interface-number*]
Display unknown source MAC addresses in the unicast-trigger quiet period.	**display dot1x unicast-trigger quiet-mac** [**interface** *interface-type interface-number*]
Remove users from the 802.1X guest VLAN on a port.	**reset dot1x guest-vlan interface** *interface-type interface-number*

Task	Command
	[**mac-address** *mac-address*]
Remove users from the 802.1X guest VSI on a port.	**reset dot1x guest-vsi interface** *interface-type interface-number* [**mac-address** *mac-address*]
Clear 802.1X statistics.	**reset dot1x statistics** [**interface** *interface-type interface-number*]
Remove the records of unknown source MAC addresses in the unicast-trigger quiet period.	**reset dot1x unicast-trigger quiet-mac** [**interface** *interface-type interface-number* | **mac-address** *mac-address*]

802.1X authentication configuration examples

Example: Configuring basic 802.1X authentication

Network configuration

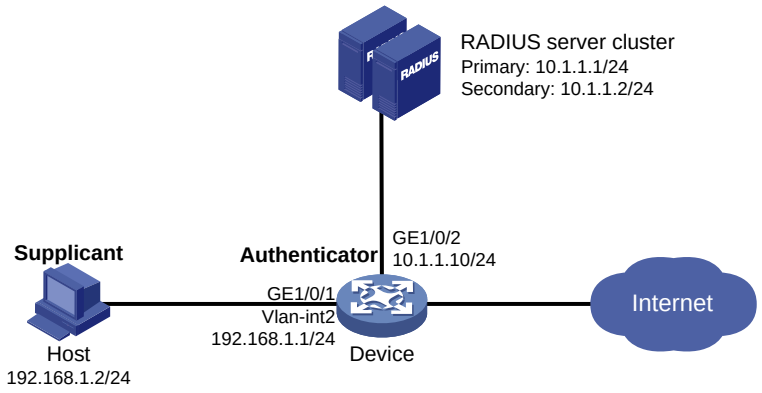
As shown in Figure 12, the access device performs 802.1X authentication for users that connect to GigabitEthernet 1/0/1. Implement MAC-based access control on the port, so the logoff of one user does not affect other online 802.1X users.

Use RADIUS servers to perform authentication, authorization, and accounting for the 802.1X users. If RADIUS authentication fails, perform local authentication on the access device.

Configure the RADIUS server at 10.1.1.1/24 as the primary authentication and accounting server, and the RADIUS server at 10.1.1.2/24 as the secondary authentication and accounting server. Assign all users to ISP domain **bbb**.

Set the shared key to **name** for packets between the access device and the authentication server. Set the shared key to **money** for packets between the access device and the accounting server.

Figure 12 Network diagram



Procedure

For information about the RADIUS commands used on the access device in this example, see *Security Command Reference*.

1. Configure the RADIUS servers and add user accounts for the 802.1X users. Make sure the RADIUS servers can provide authentication, authorization, and accounting services. (Details not shown.)

2. Assign an IP address to each interface. (Details not shown.)
3. Configure user accounts for the 802.1X users on the access device:

   # Add a local network access user with username **localuser** and password **123456TESTplat&!** in plaintext. (Make sure the username and password are the same as those configured on the RADIUS servers.)

   ```
   <Device> system-view
   [Device] local-user localuser class network
   [Device-luser-network-localuser] password simple 123456TESTplat&!
   ```

   # Set the service type to **lan-access**.

   ```
   [Device-luser-network-localuser] service-type lan-access
   [Device-luser-network-localuser] quit
   ```

4. Configure a RADIUS scheme on the access device:

   # Create a RADIUS scheme named **radius1** and enter RADIUS scheme view.

   ```
   [Device] radius scheme radius1
   ```

   # Specify the IP addresses of the primary authentication and accounting RADIUS servers.

   ```
   [Device-radius-radius1] primary authentication 10.1.1.1
   [Device-radius-radius1] primary accounting 10.1.1.1
   ```

   # Configure the IP addresses of the secondary authentication and accounting RADIUS servers.

   ```
   [Device-radius-radius1] secondary authentication 10.1.1.2
   [Device-radius-radius1] secondary accounting 10.1.1.2
   ```

   # Specify the shared key between the access device and the authentication server.

   ```
   [Device-radius-radius1] key authentication simple name
   ```

   # Specify the shared key between the access device and the accounting server.

   ```
   [Device-radius-radius1] key accounting simple money
   ```

   # Exclude the ISP domain names from the usernames sent to the RADIUS servers.

   ```
   [Device-radius-radius1] user-name-format without-domain
   [Device-radius-radius1] quit
   ```

   NOTE:

   The access device must use the same username format as the RADIUS server. If the RADIUS server includes the ISP domain name in the username, so must the access device.

5. Configure an ISP domain on the access device:

   # Create an ISP domain named **bbb** and enter ISP domain view.

   ```
   [Device] domain bbb
   ```

   # Apply RADIUS scheme **radius1** to the ISP domain, and specify local authentication as the secondary authentication method.

   ```
   [Device-isp-bbb] authentication lan-access radius-scheme radius1 local
   [Device-isp-bbb] authorization lan-access radius-scheme radius1 local
   [Device-isp-bbb] accounting lan-access radius-scheme radius1 local
   [Device-isp-bbb] quit
   ```

6. Configure 802.1X on the access device:

   # Enable 802.1X on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] dot1x
   ```

   # Enable MAC-based access control on the port. By default, the port uses MAC-based access control.

   ```
   [Device-GigabitEthernet1/0/1] dot1x port-method macbased
   ```

Specify ISP domain **bbb** as the mandatory domain.

```
[Device-GigabitEthernet1/0/1] dot1x mandatory-domain bbb
[Device-GigabitEthernet1/0/1] quit
```

Enable 802.1X globally.

```
[Device] dot1x
```

7. Configure the 802.1X client. If an iNode client is used, do not select the **Carry version info** option in the client configuration. (Details not shown.)

Verifying the configuration

Verify the 802.1X configuration on GigabitEthernet 1/0/1.

```
[Device] display dot1x interface gigabitethernet 1/0/1
```

Display the user connection information after an 802.1X user passes authentication.

```
[Device] display dot1x connection
```

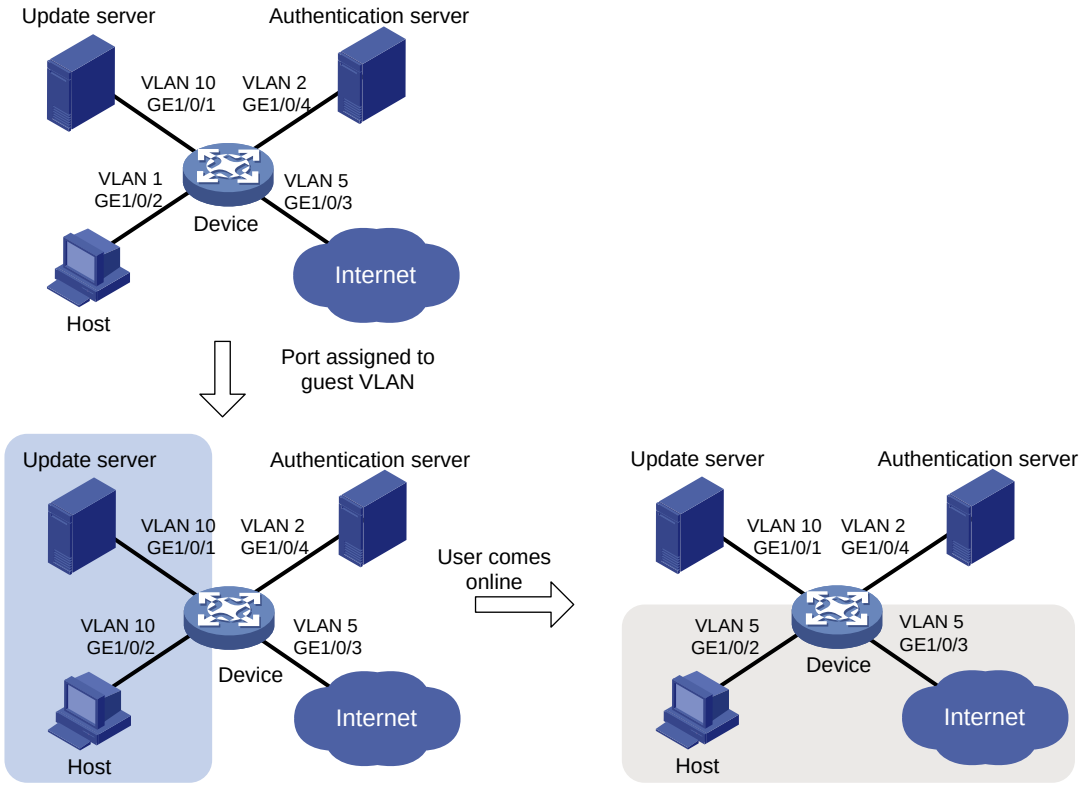
Example: Configuring 802.1X guest VLAN and authorization VLAN

Network configuration

As shown in Figure 13:

- Use RADIUS servers to perform authentication, authorization, and accounting for 802.1X users that connect to GigabitEthernet 1/0/2. Implement port-based access control on the port.
- Configure VLAN 10 as the 802.1X guest VLAN on GigabitEthernet 1/0/2. The host and the update server are both in VLAN 10, and the host can access the update server and download the 802.1X client software.
- Configure a QoS policy to deny packets destined for the Internet (5.1.1.1) and apply the QoS policy to the outbound direction of VLAN 10. The configuration prevents users in the 802.1X guest VLAN from accessing the Internet before they pass 802.1X authentication.
- After the host passes 802.1X authentication, the access device assigns the host to VLAN 5 where GigabitEthernet 1/0/3 is. The host can access the Internet.

Figure 13 Network diagram



Procedure

For information about the RADIUS commands used on the access device in this example, see *Security Command Reference*.

1. Configure the RADIUS server to provide authentication, authorization, and accounting services. Configure user accounts and authorization VLAN (VLAN 5 in this example) for the users. (Details not shown.)
2. Create VLANs, and assign ports to the VLANs on the access device.

   NOTE:

   By default, VLAN 1 exists and all ports belong to the VLAN. You do not need to create the VLAN or assign GigabitEthernet 1/0/2 to the VLAN.

   ```
   <Device> system-view
   [Device] vlan 10
   [Device-vlan10] port gigabitethernet 1/0/1
   [Device-vlan10] quit
   [Device] vlan 2
   [Device-vlan2] port gigabitethernet 1/0/4
   [Device-vlan2] quit
   [Device] vlan 5
   [Device-vlan5] port gigabitethernet 1/0/3
   [Device-vlan5] quit
   ```

3. Configure a QoS policy:

   # Configure advanced ACL 3000 to match packets destined for 5.1.1.1.

   ```
   [Device] acl advanced 3000
   ```

```
[Device-acl-ipv4-adv-3000] rule permit ip destination 5.1.1.1 0.0.0.255
[Device-acl-ipv4-adv-3000] quit
```

Specify advanced ACL 3000 in traffic class **classifier_1** to match traffic.

```
[Device] traffic classifier classifier_1
[Device-classifier-classifier_1] if-match acl 3000
[Device-classifier-classifier_1] quit
```

Configure traffic behavior **behavior_1** to deny matching packets.

```
[Device] traffic behavior behavior_1
[Device-behavior-behavior_1] filter deny
[Device-behavior-behavior_1] quit
```

Configure QoS policy **policy_1** to associate traffic class **classifier_1** with traffic behavior **behavior_1**.

```
[Device] qos policy policy_1
[Device-qospolicy-policy_1] classifier classifier_1 behavior behavior_1
[Device-qospolicy-policy_1] quit
```

Apply QoS policy **policy_1** to the outbound direction of VLAN 10.

```
[Device] qos vlan-policy policy_1 vlan 10 outbound
```

4. Configure a RADIUS scheme on the access device:

Create RADIUS scheme **2000** and enter RADIUS scheme view.

```
[Device] radius scheme 2000
```

Specify the server at 10.11.1.1 as the primary authentication server, and set the authentication port to 1812.

```
[Device-radius-2000] primary authentication 10.11.1.1 1812
```

Specify the server at 10.11.1.1 as the primary accounting server, and set the accounting port to 1813.

```
[Device-radius-2000] primary accounting 10.11.1.1 1813
```

Set the shared key to **abc** in plain text for secure communication between the authentication server and the device.

```
[Device-radius-2000] key authentication simple abc
```

Set the shared key to **abc** in plain text for secure communication between the accounting server and the device.

```
[Device-radius-2000] key accounting simple abc
```

Exclude the ISP domain names from the usernames sent to the RADIUS server.

```
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

5. Configure an ISP domain on the access device:

Create ISP domain **bbb** and enter ISP domain view.

```
[Device] domain bbb
```

Apply RADIUS scheme 2000 to the ISP domain for authentication, authorization, and accounting.

```
[Device-isp-bbb] authentication lan-access radius-scheme 2000
[Device-isp-bbb] authorization lan-access radius-scheme 2000
[Device-isp-bbb] accounting lan-access radius-scheme 2000
[Device-isp-bbb] quit
```

6. Configure 802.1X on the access device:

Enable 802.1X on GigabitEthernet 1/0/2.

```
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] dot1x
```

Implement port-based access control on the port.

```
[Device-GigabitEthernet1/0/2] dot1x port-method portbased
```

Set the port authorization mode to **auto**. By default, the port uses the auto mode.

```
[Device-GigabitEthernet1/0/2] dot1x port-control auto
```

Specify VLAN 10 as the 802.1X guest VLAN on GigabitEthernet 1/0/2.

```
[Device-GigabitEthernet1/0/2] dot1x guest-vlan 10
[Device-GigabitEthernet1/0/2] quit
```

Enable 802.1X globally.

```
[Device] dot1x
```

7. Configure the 802.1X client. Make sure the 802.1X client can update its IP address after the access port is assigned to the guest VLAN or an authorization VLAN. (Details not shown.)

Verifying the configuration

Verify the 802.1X guest VLAN configuration on GigabitEthernet 1/0/2.

```
[Device] display dot1x interface gigabitethernet 1/0/2
```

Verify that GigabitEthernet 1/0/2 is assigned to VLAN 10 before any user passes authentication on the port.

```
[Device] display vlan 10
```

After a user passes authentication, display information on GigabitEthernet 1/0/2. Verify that GigabitEthernet 1/0/2 is assigned to VLAN 5.

```
[Device] display interface gigabitethernet 1/0/2
```

Example: Configuring 802.1X with ACL assignment

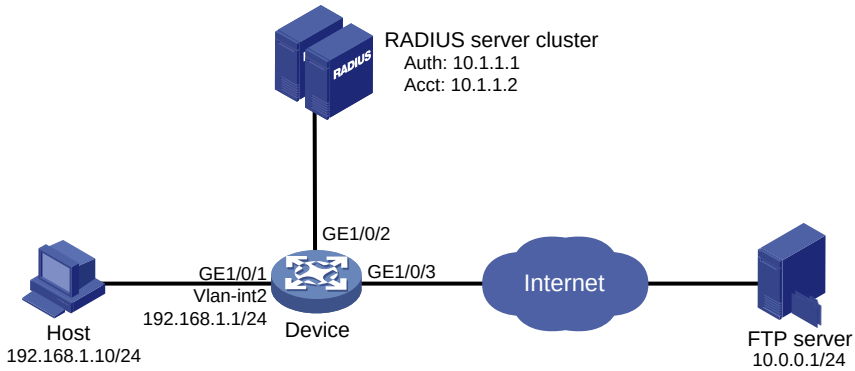
Network configuration

As shown in Figure 14, the host that connects to GigabitEthernet 1/0/1 must pass 802.1X authentication to access the Internet.

Perform 802.1X authentication on GigabitEthernet 1/0/1. Use the RADIUS server at 10.1.1.1 as the authentication and authorization server, and the RADIUS server at 10.1.1.2 as the accounting server.

Configure ACL assignment on GigabitEthernet 1/0/1 to deny access of 802.1X users to the FTP server from 8:00 to 18:00 on weekdays.

Figure 14 Network diagram



Procedure

For information about the RADIUS commands used on the access device in this example, see *Security Command Reference*.

1. Configure the RADIUS servers to provide authentication, authorization, and accounting services. Add user accounts and specify the ACL (ACL 3000 in this example) for the users. (Details not shown.)
2. Assign an IP address to each interface, as shown in Figure 14. (Details not shown.)
3. Configure a RADIUS scheme on the access device:

   # Create RADIUS scheme **2000** and enter RADIUS scheme view.

   ```
   <Device> system-view
   [Device] radius scheme 2000
   ```

   # Specify the server at 10.1.1.1 as the primary authentication server, and set the authentication port to 1812.

   ```
   [Device-radius-2000] primary authentication 10.1.1.1 1812
   ```

   # Specify the server at 10.1.1.2 as the primary accounting server, and set the accounting port to 1813.

   ```
   [Device-radius-2000] primary accounting 10.1.1.2 1813
   ```

   # Set the shared key to **abc** in plain text for secure communication between the authentication server and the device.

   ```
   [Device-radius-2000] key authentication simple abc
   ```

   # Set the shared key to **abc** in plain text for secure communication between the accounting server and the device.

   ```
   [Device-radius-2000] key accounting simple abc
   ```

   # Exclude the ISP domain names from the usernames sent to the RADIUS server.

   ```
   [Device-radius-2000] user-name-format without-domain
   [Device-radius-2000] quit
   ```

4. Configure an ISP domain on the access device:

   # Create ISP domain **bbb** and enter ISP domain view.

   ```
   [Device] domain bbb
   ```

   # Apply RADIUS scheme 2000 to the ISP domain for authentication, authorization, and accounting.

   ```
   [Device-isp-bbb] authentication lan-access radius-scheme 2000
   [Device-isp-bbb] authorization lan-access radius-scheme 2000
   [Device-isp-bbb] accounting lan-access radius-scheme 2000
   [Device-isp-bbb] quit
   ```

5. Configure a time range named **ftp** from 8:00 to 18:00 on weekdays on the access device.

   ```
   [Device] time-range ftp 8:00 to 18:00 working-day
   ```

6. Configure ACL 3000 to deny packets destined for the FTP server at 10.0.0.1 during the specified time range on the access device.

   ```
   [Device] acl advanced 3000
   [Device-acl-ipv4-adv-3000] rule 0 deny ip destination 10.0.0.1 0 time-range ftp
   [Device-acl-ipv4-adv-3000] quit
   ```

7. Configure 802.1X on the access device:

   # Enable 802.1X on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] dot1x
   [Device-GigabitEthernet1/0/1] quit
   ```

   # Enable 802.1X globally.

   ```
   [Device] dot1x
   ```

8. Configure the 802.1X client. Make sure the client is able to update its IP address after the access port is assigned to the 802.1X guest VLAN or an authorization VLAN. (Details not shown.)

Verifying the configuration

Use the user account to pass authentication. (Details not shown.)

Verify that the user cannot ping the FTP server at any time from 8:00 to 18:00 on any weekday.

```
C:\>ping 10.0.0.1

Pinging 10.0.0.1 with 32 bytes of data:

Request timed out.
Request timed out.
Request timed out.
Request timed out.

Ping statistics for 10.0.0.1:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),
```

The output shows that ACL 3000 is active on the user, and the user cannot access the FTP server.

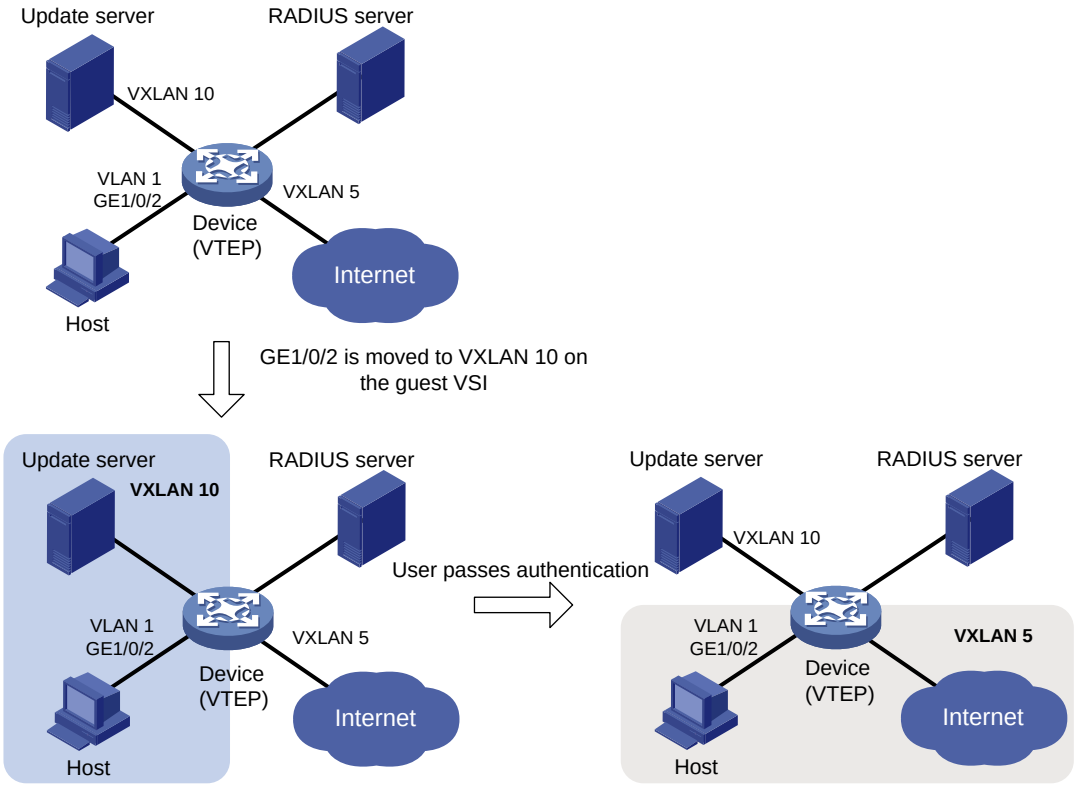
Example: Configuring 802.1X guest VSI and authorization VSI configuration (port-based)

Network configuration

As shown in Figure 15:

- The device acts as both a VXLAN VTEP and a network access device. It uses the RADIUS server to perform authentication, authorization, and accounting for 802.1X users that connect to GigabitEthernet 1/0/2.
- GigabitEthernet 1/0/2 uses port-based access control and is configured with the 802.1X guest VSI. VXLAN 10 is created on the guest VSI. Users in the guest VSI can access the update server in VXLAN 10 and download the 802.1X client software.
- The RADIUS server assigns an authorization VSI to the host. The VSI is associated with VXLAN 5 on the device. After passing authentication, the host can access the Internet.

Figure 15 Network diagram



Procedure

For information about the RADIUS commands used on the access device in this example, see *Security Command Reference*.

1. Configure the RADIUS server to provide authentication, authorization, and accounting services. Configure user accounts and authorization VSI (VSI **vpn5** in this example) for the users. (Details not shown.)

   If an ADCAM server is used for authentication and authorization, configure VSIs on the server. The server will assign these VSIs to the device. You do not need to configure VSIs on the device.

2. Enable L2VPN on the access device.

   ```
   <Device> system-view
   [Device] l2vpn enable
   ```

3. Create VSIs and the corresponding VXLANs on the access device.

   ```
   [Device] vsi vpn10
   [Device-vsi-vpn10] vxlan 10
   [Device-vsi-vpn10-vxlan-10] quit
   [Device-vsi-vpn10] quit
   [Device] vsi vpn5
   [Device-vsi-vpn5] vxlan 5
   [Device-vsi-vpn5-vxlan-5] quit
   [Device-vsi-vpn5] quit
   ```

4. Configure a RADIUS scheme on the access device:

   # Create RADIUS scheme **2000** and enter RADIUS scheme view.

   ```
   [Device] radius scheme 2000
   ```

Specify the server at 10.11.1.1 as the primary authentication server, and set the authentication port to 1812.

```
[Device-radius-2000] primary authentication 10.11.1.1 1812
```

Specify the server at 10.11.1.1 as the primary accounting server, and set the accounting port to 1813.

```
[Device-radius-2000] primary accounting 10.11.1.1 1813
```

Set the shared key to **abc** in plain text for secure communication between the authentication server and the device.

```
[Device-radius-2000] key authentication simple abc
```

Set the shared key to **abc** in plain text for secure communication between the accounting server and the device.

```
[Device-radius-2000] key accounting simple abc
```

Exclude the ISP domain names from the usernames sent to the authentication and accounting servers.

```
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

5. Configure an ISP domain on the access device:

Create ISP domain **bbb** and enter ISP domain view.

```
[Device] domain bbb
```

Apply RADIUS scheme 2000 to the ISP domain for authentication, authorization, and accounting.

```
[Device-isp-bbb] authentication lan-access radius-scheme 2000
[Device-isp-bbb] authorization lan-access radius-scheme 2000
[Device-isp-bbb] accounting lan-access radius-scheme 2000
[Device-isp-bbb] quit
```

6. Configure 802.1X on the access device:

Enable 802.1X on GigabitEthernet 1/0/2.

```
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] dot1x
```

Implement port-based access control on the port.

```
[Device-GigabitEthernet1/0/2] dot1x port-method portbased
```

Set the port authorization mode to **auto**. By default, the port uses the auto mode.

```
[Device-GigabitEthernet1/0/2] dot1x port-control auto
```

Specify VSI **vpn10** as the 802.1X guest VSI on GigabitEthernet 1/0/2.

```
[Device-GigabitEthernet1/0/2] dot1x guest-vsi vpn10
[Device-GigabitEthernet1/0/2] quit
```

Enable 802.1X globally.

```
[Device] dot1x
```

7. Configure the 802.1X client. Make sure the 802.1X client can update its IP address after the access port is assigned to the guest VSI or an authorization VSI. (Details not shown.)

Verifying the configuration

Verify that GigabitEthernet 1/0/2 is assigned to VSI **vpn10** before any user passes authentication on the port.

```
[Device] display l2vpn forwarding ac verbose
```

Verify that GigabitEthernet 1/0/2 is assigned to VSI **vpn5** after a user passes authentication on the port.

```
[Device] display l2vpn forwarding ac verbose
```

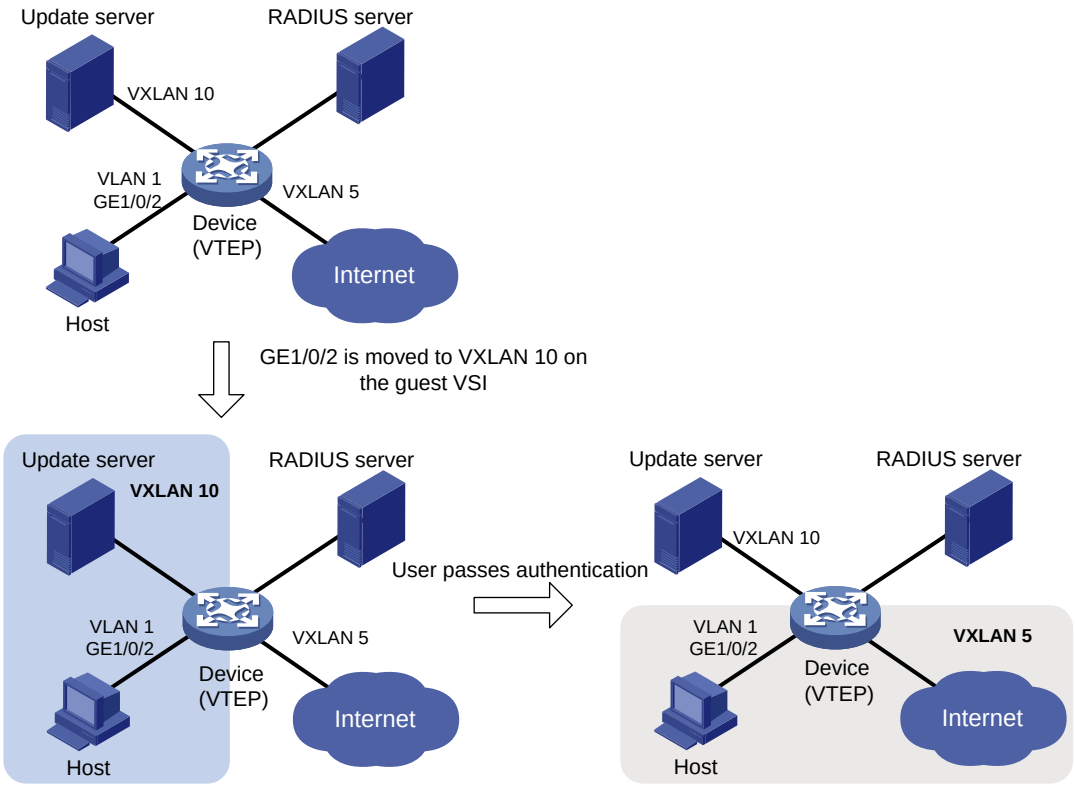
Example: Configuring 802.1X guest VSI and authorization VSI (MAC-based)

Network configuration

As shown in Figure 16:

- The device acts as both a VXLAN VTEP and a network access device. It uses the RADIUS server to perform authentication, authorization, and accounting for 802.1X users that connect to GigabitEthernet 1/0/2.
- GigabitEthernet 1/0/2 uses MAC-based access control and is configured with the 802.1X guest VSI. VXLAN 10 is created on the guest VSI. Users in the guest VSI can access the update server in VXLAN 10 and download the 802.1X client software.
- The RADIUS server assigns an authorization VSI to the host. The VSI is associated with VXLAN 5 on the device. After passing authentication, the host can access the Internet.

Figure 16 Network diagram



Procedure

For information about the RADIUS commands used on the access device in this example, see *Security Command Reference*.

1. Configure the RADIUS server to provide authentication, authorization, and accounting services. Configure user accounts and authorization VSI (VSI **vpn5** in this example) for the users. (Details not shown.)

   If an ADCAM server is used for authentication and authorization, configure VSIs on the server. The server will assign these VSIs to the device. You do not need to configure VSIs on the device.

2. Enable L2VPN on the access device.

```
<Device> system-view
[Device] l2vpn enable
```

3. Create VSIs and the corresponding VXLANs on the access device.

```
[Device] vsi vpn10
[Device-vsi-vpn10] vxlan 10
[Device-vsi-vpn10-vxlan-10] quit
[Device-vsi-vpn10] quit
[Device] vsi vpn5
[Device-vsi-vpn5] vxlan 5
[Device-vsi-vpn5-vxlan-5] quit
[Device-vsi-vpn5] quit
```

4. Configure a RADIUS scheme on the access device:

   # Create RADIUS scheme **2000** and enter RADIUS scheme view.

```
[Device] radius scheme 2000
```

   # Specify the server at 10.11.1.1 as the primary authentication server, and set the authentication port to 1812.

```
[Device-radius-2000] primary authentication 10.11.1.1 1812
```

   # Specify the server at 10.11.1.1 as the primary accounting server, and set the accounting port to 1813.

```
[Device-radius-2000] primary accounting 10.11.1.1 1813
```

   # Set the shared key to **abc** in plain text for secure communication between the authentication server and the device.

```
[Device-radius-2000] key authentication simple abc
```

   # Set the shared key to **abc** in plain text for secure communication between the accounting server and the device.

```
[Device-radius-2000] key accounting simple abc
```

   # Exclude the ISP domain names from the usernames sent to the authentication and accounting servers.

```
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

5. Configure an ISP domain on the access device:

   # Create ISP domain **bbb** and enter ISP domain view.

```
[Device] domain bbb
```

   # Apply RADIUS scheme **2000** to the ISP domain for authentication, authorization, and accounting.

```
[Device-isp-bbb] authentication lan-access radius-scheme 2000
[Device-isp-bbb] authorization lan-access radius-scheme 2000
[Device-isp-bbb] accounting lan-access radius-scheme 2000
[Device-isp-bbb] quit
```

6. Configure 802.1X on the access device:

   # Enable 802.1X on GigabitEthernet 1/0/2.

```
[Device] interface gigabitethernet 1/0/2
[Device-GigabitEthernet1/0/2] dot1x
```

   # Set the port authorization mode to **auto**. By default, the port uses the auto mode.

```
[Device-GigabitEthernet1/0/2] dot1x port-control auto
```

   # Enable MAC-based traffic matching for dynamic Ethernet service instances on GigabitEthernet 1/0/2.

```
[Device-GigabitEthernet1/0/2] mac-based ac
```

Enable 802.1X unicast trigger on GigabitEthernet 1/0/2.

```
[Device-GigabitEthernet1/0/2] dot1x unicast-trigger
```

Specify VSI **vpn10** as the 802.1X guest VSI on GigabitEthernet 1/0/2.

```
[Device-GigabitEthernet1/0/2] dot1x guest-vsi vpn10
[Device-GigabitEthernet1/0/2] quit
```

Enable 802.1X globally.

```
[Device] dot1x
```

7. Configure the 802.1X client. Make sure the 802.1X client can update its IP address after the access port is assigned to the guest VSI or an authorization VSI. (Details not shown.)

Verifying the configuration

Verify that GigabitEthernet 1/0/2 is assigned to VSI **vpn10** if no responses are received from the client after 802.1X authentication is triggered.

```
[Device] display l2vpn forwarding ac verbose
```

Verify that GigabitEthernet 1/0/2 is assigned to VSI **vpn5** after a user passes authentication on the port.

```
[Device] display l2vpn forwarding ac verbose
```

Example: Configuring 802.1X with EAD assistant (with DHCP relay agent)

Network configuration

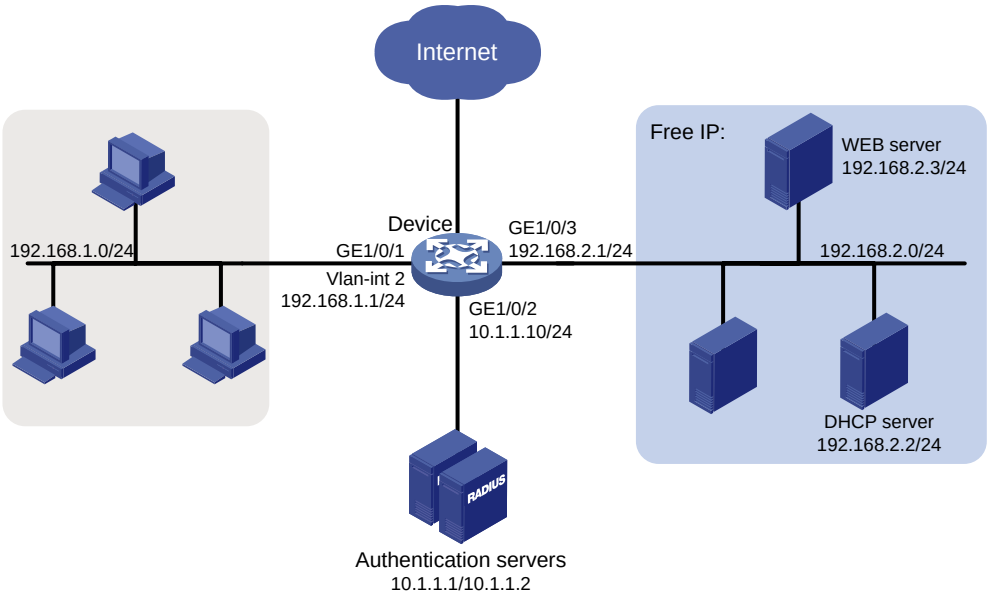
As shown in Figure 17:

- The intranet 192.168.1.0/24 is attached to GigabitEthernet 1/0/1 of the access device.
- The hosts use DHCP to obtain IP addresses.
- A DHCP server and a Web server are deployed on the 192.168.2.0/24 subnet for users to obtain IP addresses and download client software.

Deploy an EAD solution for the intranet to meet the following requirements:

- Allow unauthenticated users and users that have failed 802.1X authentication to access 192.168.2.0/24. The users can obtain IP addresses and download software.
- If these users use a Web browser to access a network other than 192.168.2.0/24, redirect them to the Web server for 802.1X client downloading.
- Allow authenticated 802.1X users to access the network.

Figure 17 Network diagram



Procedure

1. Make sure the DHCP server, the Web server, and the authentication servers have been configured correctly. (Details not shown.)
2. Configure an IP address for each interface. (Details not shown.)
3. Configure DHCP relay:

   # Enable DHCP.

   ```
   <Device> system-view
   [Device] dhcp enable
   ```

   # Enable the DHCP relay agent on VLAN-interface 2.

   ```
   [Device] interface vlan-interface 2
   [Device-Vlan-interface2] dhcp select relay
   ```

   # Specify the DHCP server 192.168.2.2 on the relay agent interface VLAN-interface 2.

   ```
   [Device-Vlan-interface2] dhcp relay server-address 192.168.2.2
   [Device-Vlan-interface2] quit
   ```

4. Configure a RADIUS scheme:

   # Create RADIUS scheme **2000** and enter RADIUS scheme view.

   ```
   [Device] radius scheme 2000
   ```

   # Specify the server at 10.1.1.1 as the primary authentication server, and set the authentication port to 1812.

   ```
   [Device-radius-2000] primary authentication 10.1.1.1 1812
   ```

   # Specify the server at 10.1.1.2 as the primary accounting server, and set the accounting port to 1813.

   ```
   [Device-radius-2000] primary accounting 10.1.1.2 1813
   ```

   # Set the shared key to **abc** in plain text for secure communication between the authentication server and the device.

   ```
   [Device-radius-2000] key authentication simple abc
   ```

   # Set the shared key to **abc** in plain text for secure communication between the accounting server and the device.

   ```
   [Device-radius-2000] key accounting simple abc
   ```

Exclude the ISP domain names from the usernames sent to the RADIUS server.

```
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

5. Configure an ISP domain:

   # Create ISP domain **bbb** and enter ISP domain view.

   ```
   [Device] domain bbb
   ```

   # Apply RADIUS scheme 2000 to the ISP domain for authentication, authorization, and accounting.

   ```
   [Device-isp-bbb] authentication lan-access radius-scheme 2000
   [Device-isp-bbb] authorization lan-access radius-scheme 2000
   [Device-isp-bbb] accounting lan-access radius-scheme 2000
   [Device-isp-bbb] quit
   ```

6. Configure 802.1X:

   # Configure the free IP.

   ```
   [Device] dot1x ead-assistant free-ip 192.168.2.0 24
   ```

   # Configure the redirect URL for client software download.

   ```
   [Device] dot1x ead-assistant url https://192.168.2.3
   ```

   # Enable the EAD assistant feature.

   ```
   [Device] dot1x ead-assistant enable
   ```

   # Enable 802.1X on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] dot1x
   [Device-GigabitEthernet1/0/1] quit
   ```

   # Enable 802.1X globally.

   ```
   [Device] dot1x
   ```

Verifying the configuration

Verify the 802.1X configuration.

```
[Device] display dot1x
```

Verify that you can ping an IP address on the free IP subnet from a host before the user on the host passes 802.1X authentication. (Details not shown.)

Verify that you are redirected to the Web server when you enter in your Web browser an IP address not on the free IP. (Details not shown.)

Example: Configuring 802.1X with EAD assistant (with DHCP server)

Network configuration

As shown in Figure 18:

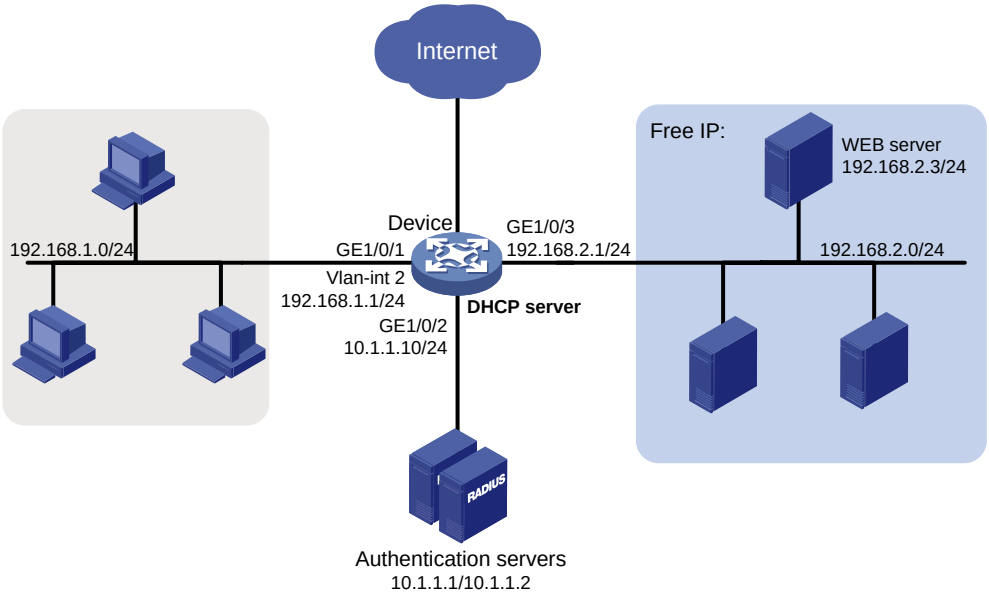
- The intranet 192.168.1.0/24 is attached to GigabitEthernet 1/0/1 of the access device.
- The hosts use DHCP to obtain IP addresses.
- A Web server is deployed on the 192.168.2.0/24 subnet for users to download client software.

Deploy an EAD solution for the intranet to meet the following requirements:

- Allow unauthenticated users and users that have failed 802.1X authentication to access 192.168.2.0/24. The users can download software.

- If these users use a Web browser to access a network other than 192.168.2.0/24, redirect them to the Web server for 802.1X client downloading.
- Allow authenticated 802.1X users to access the network.

Figure 18 Network diagram



Procedure

1. Make sure the Web server and the authentication servers have been configured correctly. (Details not shown.)
2. Configure an IP address for each interface. (Details not shown.)
3. Configure the DHCP server:

   # Enable DHCP.

   ```
   <Device> system-view
   [Device] dhcp enable
   ```

   # Enable the DHCP server on VLAN-interface 2.

   ```
   [Device] interface vlan-interface 2
   [Device-Vlan-interface2] dhcp select server
   [Device-Vlan-interface2] quit
   ```

   # Create DHCP address pool **0**.

   ```
   [Device] dhcp server ip-pool 0
   ```

   # Specify subnet 192.168.1.0/24 in DHCP address pool 0.

   ```
   [Device-dhcp-pool-0] network 192.168.1.0 mask 255.255.255.0
   ```

   # Specify the gateway address 192.168.1.1 in DHCP address pool 0.

   ```
   [Device-dhcp-pool-0] gateway-list 192.168.1.1
   [Device-dhcp-pool-0] quit
   ```

4. Configure a RADIUS scheme:

   # Create RADIUS scheme **2000** and enter RADIUS scheme view.

   ```
   [Device] radius scheme 2000
   ```

   # Specify the server at 10.1.1.1 as the primary authentication server, and set the authentication port to 1812.

   ```
   [Device-radius-2000] primary authentication 10.1.1.1 1812
   ```

Specify the server at 10.1.1.2 as the primary accounting server, and set the accounting port to 1813.

```
[Device-radius-2000] primary accounting 10.1.1.2 1813
```

Set the shared key to **abc** in plain text for secure communication between the authentication server and the device.

```
[Device-radius-2000] key authentication simple abc
```

Set the shared key to **abc** in plain text for secure communication between the accounting server and the device.

```
[Device-radius-2000] key accounting simple abc
```

Exclude the ISP domain names from the usernames sent to the RADIUS server.

```
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

5. Configure an ISP domain:

Create ISP domain **bbb** and enter ISP domain view.

```
[Device] domain bbb
```

Apply RADIUS scheme 2000 to the ISP domain for authentication, authorization, and accounting.

```
[Device-isp-bbb] authentication lan-access radius-scheme 2000
[Device-isp-bbb] authorization lan-access radius-scheme 2000
[Device-isp-bbb] accounting lan-access radius-scheme 2000
[Device-isp-bbb] quit
```

6. Configure 802.1X:

Configure the free IP.

```
[Device] dot1x ead-assistant free-ip 192.168.2.0 24
```

Configure the redirect URL for client software download.

```
[Device] dot1x ead-assistant url https://192.168.2.3
```

Enable the EAD assistant feature.

```
[Device] dot1x ead-assistant enable
```

Enable 802.1X on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] dot1x
[Device-GigabitEthernet1/0/1] quit
```

Enable 802.1X globally.

```
[Device] dot1x
```

Verifying the configuration

Verify the 802.1X configuration.

```
[Device] display dot1x
```

Verify that you can ping an IP address on the free IP subnet from a host before the user on the host passes 802.1X authentication. (Details not shown.)

Verify that you are redirected to the Web server when you enter in your Web browser an IP address not on the free IP. (Details not shown.)

Troubleshooting 802.1X

EAD assistant Web redirection failure

Symptom

Unauthenticated users are not redirected to the specified URL server after they enter external website addresses in their Web browsers.

Analysis

Redirection will not happen for one of the following reasons:

- The address is in the string format. The operating system of the host regards the string as a website name and tries to resolve the string. If the resolution fails, the operating system sends an ARP request, but the target address is not in the dotted decimal notation. The redirection feature does redirect this kind of ARP request.
- The address is within a free IP segment or free microsegment. No redirection will take place, even if no host is present with the address.
- The server is not in a free IP segment or free microsegment.
- No server is using the redirect URL, or the server with the URL does not provide Web services.

Solution

To resolve the issue:

1. Enter a dotted decimal IP address that is not in any free IP segments or free microsegments.
2. Verify that the access device and the server are configured correctly.
3. If the issue persists, contact H3C Support.

Contents

Configuring MAC authentication 1

About MAC authentication 1
User account policies 1
Authentication methods 2
VLAN assignment 3
VSI manipulation 7
Microsegment assignment 9
ACL assignment 10
Redirect URL assignment 11
Blackhole MAC attribute assignment 11
Periodic MAC reauthentication 11
RESTful server-assisted MAC authentication user recovery 12
Restrictions and guidelines: MAC authentication configuration 12
MAC authentication tasks at a glance 14
Prerequisites for MAC authentication 14
Enabling MAC authentication 15
Specifying a MAC authentication method 15
Specifying a MAC authentication domain 16
Configuring user account policy 16
Configuring MAC authentication timers 17
Configuring periodic MAC reauthentication 18
Configuring a MAC authentication guest VLAN 19
Specifying a MAC authentication guest VLAN on a port 19
Enabling guest VLAN reauthentication in MAC authentication 20
Configuring a MAC authentication critical VLAN 20
Configuring a MAC authentication guest VSI 21
Specifying a MAC authentication guest VSI on a port 21
Enabling guest VSI reauthentication in MAC authentication 22
Configuring a MAC authentication critical VSI 22
Specifying a critical profile for MAC authentication users on a port 23
Configuring a MAC authentication critical microsegment 24
Configuring a MAC authentication access profile 24
Configuring unauthenticated MAC authentication user aging 25
Configuring MAC authentication offline detection 26
Enabling online MAC authentication users to escape from offline detection and stay online when no reachable RADIUS authentication servers are available 27
Enabling online user synchronization for MAC authentication 28
Setting the maximum number of concurrent MAC authentication users on a port 29
Enabling MAC authentication multi-VLAN mode on a port 29
Configuring MAC authentication delay 30
Including user IP addresses in MAC authentication requests 30
Enabling parallel MAC authentication and 802.1X authentication 32
Configuring RESTful server-assisted MAC authentication user recovery 33
Tracking the availability of the Web authentication server at a redirect URL 34
Configuring Web proxy ports for URL redirection in MAC authentication 35
Logging off MAC authentication users 35
Enabling MAC authentication user logging 35
Display and maintenance commands for MAC authentication 36
MAC authentication configuration examples 37
Example: Configuring local MAC authentication 37
Example: Configuring RADIUS-based MAC authentication 40
Example: Configuring ACL assignment for MAC authentication 42
Example: Configuring MAC authentication authorization VSI assignment 45

Configuring MAC authentication

About MAC authentication

MAC authentication controls network access by authenticating source MAC addresses on a port. The feature does not require client software, and users do not have to enter a username and password for network access. The device initiates a MAC authentication process when it detects an unknown source MAC address on a MAC authentication-enabled port. If the MAC address passes authentication, the user can access authorized network resources. If the authentication fails, the device marks the MAC address as a silent MAC address, drops the packet, and starts a quiet timer. The device drops all subsequent packets from the MAC address within the quiet time. The quiet mechanism avoids repeated authentication during a short time.

User account policies

MAC authentication supports the following user account policies:

- **Global user account policy**, which can be either of the following options:
  - **MAC-based user account policy**—One MAC-based user account for each user. As shown in Figure 1, the access device uses the source MAC addresses in packets as the usernames and passwords of users for MAC authentication. This policy is suitable for an insecure environment.

    NOTE:

    MAC-based user account policy also supports configuring a password shared by all MAC-based user accounts.

  - **Shared user account policy**—One shared user account for all users. You specify one username and password (which are not necessarily a MAC address) for all MAC authentication users on the access device, as shown in Figure 2. This policy is suitable for a secure environment.
- **MAC range-specific user account policy**—One shared user account for users in a specific MAC address range. You specify one username and password (which are not necessarily a MAC address) for users in a specific MAC address range on the access device, as shown in Figure 3. For example, you can specify a username and password for users with a specific OUI for MAC authentication.

IMPORTANT:

- You can use global and MAC range-specific user account policies together. For users in a MAC address range, the MAC range-specific user account settings have higher priority than the global user account settings.
- If a RADIUS server is used for MAC authentication, you must create the user accounts on the RADIUS server based on the user account policy on the access device.

Figure 1 MAC-based user account policy

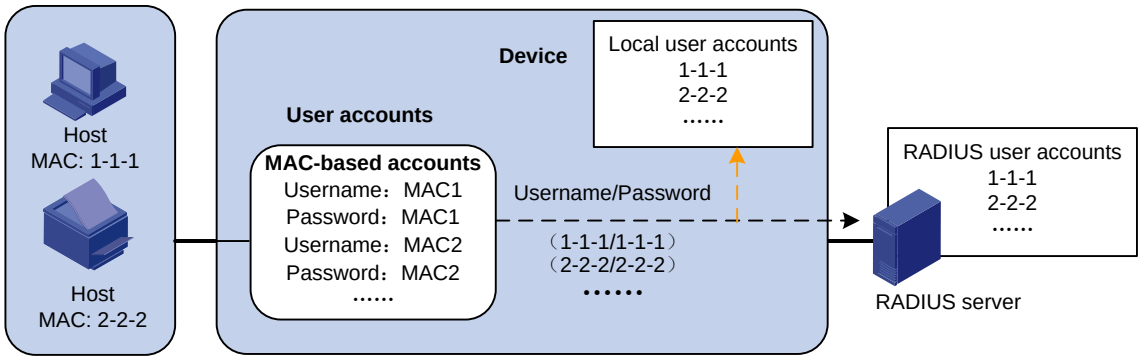


Figure 2 Shared user account policy (global)

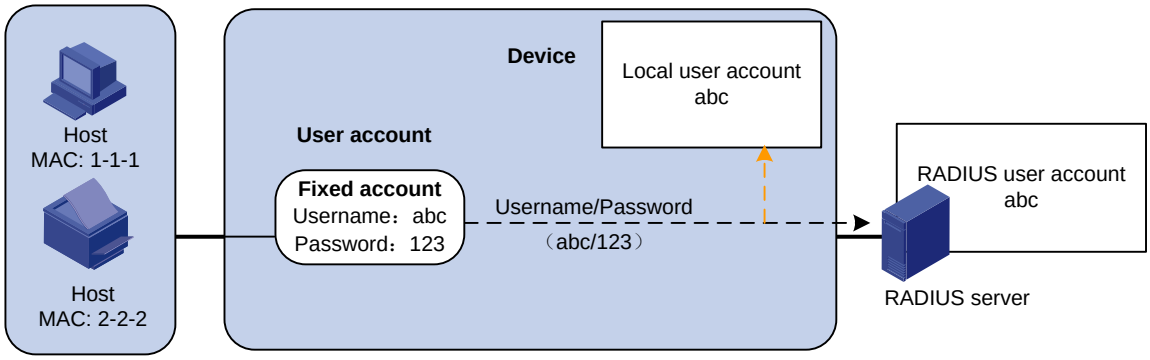
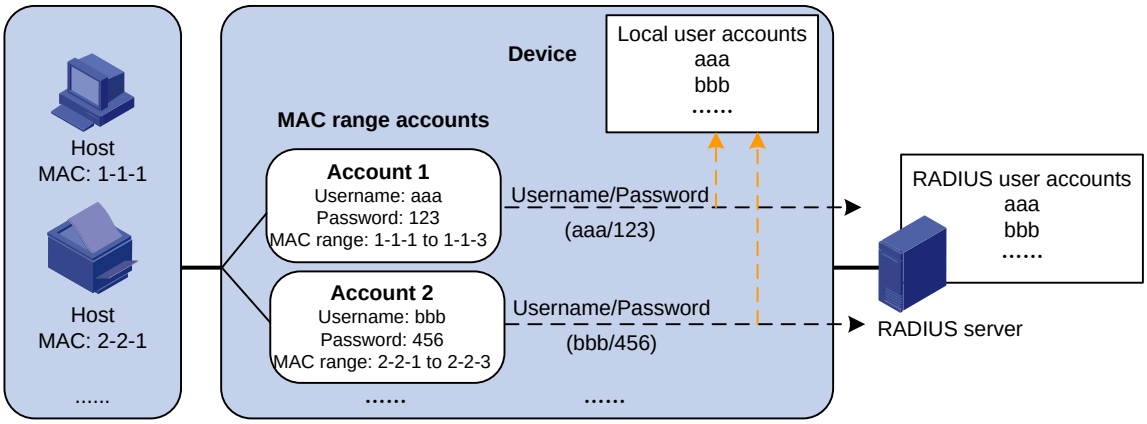


Figure 3 Shared user account policy (specific to MAC address ranges)



Authentication methods

You can perform MAC authentication on the access device (local authentication) or through a RADIUS server.

For more information about configuring local authentication and RADIUS authentication, see "Configuring AAA."

RADIUS authentication

If MAC-based accounts are used, the access device by default sends the source MAC address of a packet as the username and password to the RADIUS server for authentication. If a password is

configured for MAC-based accounts, the access device sends the configured password as the password to the RADIUS server.

If a shared account is used, the access device sends the shared account username and password to the RADIUS server for authentication.

The access device and the RADIUS server use Password Authentication Protocol (PAP) or Challenge Handshake Authentication Protocol (CHAP) for communication.

Local authentication

If MAC-based accounts are used, the access device by default uses the source MAC address of a packet as the username and password to search the local account database for a match. If a password is configured for MAC-based accounts, the device uses the configured password to search the local account database for a match.

If a shared account is used, the access device uses the shared account username and password to search the local account database for a match.

VLAN assignment

Authorization VLAN

The authorization VLAN controls the access of a MAC authentication user to authorized network resources. The device supports authorization VLANs assigned locally or by a remote server.

IMPORTANT:

Only remote servers can assign tagged authorization VLANs.

Remote VLAN authorization

In remote VLAN authorization, you must configure an authorization VLAN for a user on the remote server. After the user authenticates to the server, the server assigns authorization VLAN information to the device. Then, the device assigns the user access port to the authorization VLAN as a tagged or untagged member.

The device supports assignment of the following authorization VLAN information by the remote server:

- VLAN ID.
- VLAN name, which must be the same as the VLAN description on the access device.
- A string of VLAN IDs and VLAN names.

  In the string, some VLANs are represented by their IDs, and some VLANs are represented by their names.
- VLAN group name.

  For more information about VLAN groups, see *Layer 2—LAN Switching Configuration Guide*.
- VLAN ID with a suffix of **t** or **u**.

  The **t** and **u** suffixes require the device to assign the access port to the VLAN as a tagged or untagged member, respectively. For example, **2u** indicates assigning the port to VLAN 2 as an untagged member.

If a VLAN name or VLAN group name is assigned, the device converts the information into a VLAN ID before VLAN assignment.

IMPORTANT:

For a VLAN represented by its VLAN name to be assigned successfully, you must make sure the VLAN has been created on the device.

To assign VLAN IDs with suffixes, make sure the user access port is a hybrid or trunk port.

IMPORTANT:

To ensure a successful assignment, the authorization VLANs assigned by the remote server cannot be any of the following types:

- Dynamically learned VLANs.
- Reserved VLANs.

If the server assigns a group of VLANs, the access device selects a VLAN as described in Table 1.

Table 1 Authorization VLAN selection from a group of VLANs

VLAN information	Authorization VLAN selection
VLANs by IDs VLANs by names VLAN group name	The device selects an authorization VLAN from the VLAN group for a user according to the following rules: • On a hybrid port with MAC-based VLAN enabled: ○ If the port does not have online users, the device selects the VLAN with the lowest ID. ○ If the port has online users, the device selects the VLAN that has the fewest online users. If two VLANs have the same number of online MAC authentication users, the device selects the VLAN with the lower ID. • On an access, trunk, or MAC-based VLAN disabled hybrid port: ○ If the port does not have online users, the device selects the VLAN with the lowest ID. ○ If the port has online users, the device examines the VLAN group for the VLAN of the online users. If the VLAN is found, the VLAN is assigned to the user as the authorization VLAN. If the VLAN is not found, VLAN authorization fails.
VLAN IDs with suffixes	1. The device selects the leftmost VLAN ID without a suffix, or the leftmost VLAN ID suffixed by **u** as an untagged VLAN, whichever is more leftmost. 2. The device assigns the untagged VLAN to the port as the PVID, and it assigns the remaining as tagged VLANs. If no untagged VLAN is assigned, the PVID of the port does not change. The port permits traffic from these tagged and untagged VLANs to pass through. For example, the authentication server sends the string **1u 2t 3** to the access device for a user. The device assigns VLAN 1 as an untagged VLAN and all remaining VLANs (including VLAN 3) as tagged VLANs. VLAN 1 becomes the PVID.

Local VLAN authorization

To perform local VLAN authorization for a user, specify the VLAN ID in the authorization attribute list of the local user account for that user. For each local user, you can specify only one authorization VLAN ID. The user access port is assigned to the VLAN as an untagged member.

IMPORTANT:

Local VLAN authorization does not support assignment of tagged VLANs.

For more information about local user configuration, see "Configuring AAA."

Authorization VLAN manipulation on a MAC authentication-enabled port

Table 2 describes the way the network access device handles authorization VLANs (except for the VLANs specified with suffixes) for MAC authenticated users on a port.

Table 2 VLAN manipulation

Port type	VLAN manipulation
• Access port • Trunk port • Hybrid port with MAC-based VLAN disabled	• The device assigns the port to the first authenticated user's authorization VLAN and sets the VLAN as the PVID if that authorization VLAN has the untagged attribute. • If the authorization VLAN has the tagged attribute, the device assigns the port to the authorization VLAN without changing its PVID. **NOTE:** The tagged attribute is supported only on trunk and hybrid ports.
Hybrid port with MAC-based VLAN enabled	The device maps the MAC address of each user to its own authorization VLAN regardless of whether the port is a tagged member. The PVID of the port does not change.

IMPORTANT:

- If the users are attached to a port whose link type is access, make sure the authorization VLAN assigned by the server has the untagged attribute. VLAN assignment will fail if the server issues a VLAN that has the tagged attribute.
- When you assign VLANs to users attached to a trunk port or a MAC-based VLAN disabled hybrid port, make sure there is only one untagged VLAN. If a different untagged VLAN is assigned to a subsequent user, the user cannot pass authentication.
- As a best practice to enhance network security, do not use the **`port hybrid vlan`** command to assign a hybrid port to an authorization VLAN as a tagged member.

The VLAN assigned by the server to a user as an authorization VLAN might have been configured on the user access port but with a different tagging mode. For example, the server assigns an authorization VLAN with the tagged attribute, but the same VLAN configured on the port has the untagged attribute. In this situation, the VLAN settings that take effect on the user depend on the link type of the port.

- If the link type of the port is access or trunk, the authorization VLAN settings assigned by the server always take effect on the user as long as the user is online. After the user goes offline, the VLAN settings on the port take effect.
- If the link type of the port is hybrid, the VLAN settings configured on the port take effect. For example, the server assigns VLAN 30 with the untagged attribute to a user on the hybrid port. However, VLAN 30 has been configured on the port with the tagged attribute by using the **`port hybrid vlan tagged`** command. Then, the VLAN has the tagged attribute on the port.

For a MAC authenticated user to access the network on a hybrid port when no authorization VLAN is configured for the user, perform one of the following tasks:

- If the port receives tagged authentication packets from the user in a VLAN, use the **`port hybrid vlan`** command to configure the port as a tagged member in the VLAN.
- If the port receives untagged authentication packets from the user in a VLAN, use the **`port hybrid vlan`** command to configure the port as an untagged member in the VLAN.

Guest VLAN

The MAC authentication guest VLAN on a port accommodates users that have failed MAC authentication for any reason other than server unreachable. For example, the VLAN accommodates users with invalid passwords entered.

You can deploy a limited set of network resources in the MAC authentication guest VLAN. For example, a software server for downloading software and system patches.

A hybrid port is always assigned to a MAC authentication guest VLAN as an untagged member. After the assignment, do not reconfigure the port as a tagged member in the VLAN.

The device reauthenticates users in the MAC authentication guest VLAN at a specific interval. Table 3 shows the way that the network access device handles guest VLANs for MAC authentication users.

Table 3 VLAN manipulation

Authentication status	VLAN manipulation
A user in the MAC authentication guest VLAN fails MAC authentication.	The user is still in the MAC authentication guest VLAN.
A user in the MAC authentication guest VLAN passes MAC authentication.	The device remaps the MAC address of the user to the authorization VLAN assigned by the authentication server. If no authorization VLAN is configured for the user on the authentication server, the device remaps the MAC address of the user to the PVID of the port.

Critical VLAN

The MAC authentication critical VLAN on a port accommodates users that have failed MAC authentication because no RADIUS authentication servers are reachable. Users in a MAC authentication critical VLAN can access only network resources in the critical VLAN.

The critical VLAN feature takes effect when MAC authentication is performed only through RADIUS servers. If a MAC authentication user fails local authentication after RADIUS authentication, the user is not assigned to the critical VLAN. For more information about the authentication methods, see "Configuring AAA."

Table 4 shows the way that the network access device handles critical VLANs for MAC authentication users.

Table 4 VLAN manipulation

Authentication status	VLAN manipulation
A user fails MAC authentication because all the RADIUS servers are unreachable.	The device maps the MAC address of the user to the MAC authentication critical VLAN. The user is still in the MAC authentication critical VLAN if the user fails MAC reauthentication because all the RADIUS servers are unreachable. If no MAC authentication critical VLAN is configured, the device maps the MAC address of the user to the PVID of the port.
A user in the MAC authentication guest VLAN fails authentication because all the RADIUS servers are unreachable.	The user remains in the MAC authentication guest VLAN.
A user in the MAC authentication critical VLAN fails MAC authentication for any reason other than server unreachable.	If a guest VLAN has been configured, the device maps the MAC address of the user to the guest VLAN. If no guest VLAN is configured, the device maps the MAC address of the user to the PVID of the port.
A user in the MAC authentication critical VLAN passes MAC authentication.	The device remaps the MAC address of the user to the authorization VLAN assigned by the authentication server. If no authorization VLAN is configured for the user on the authentication server, the device remaps the MAC address of the user to the PVID of the access port.

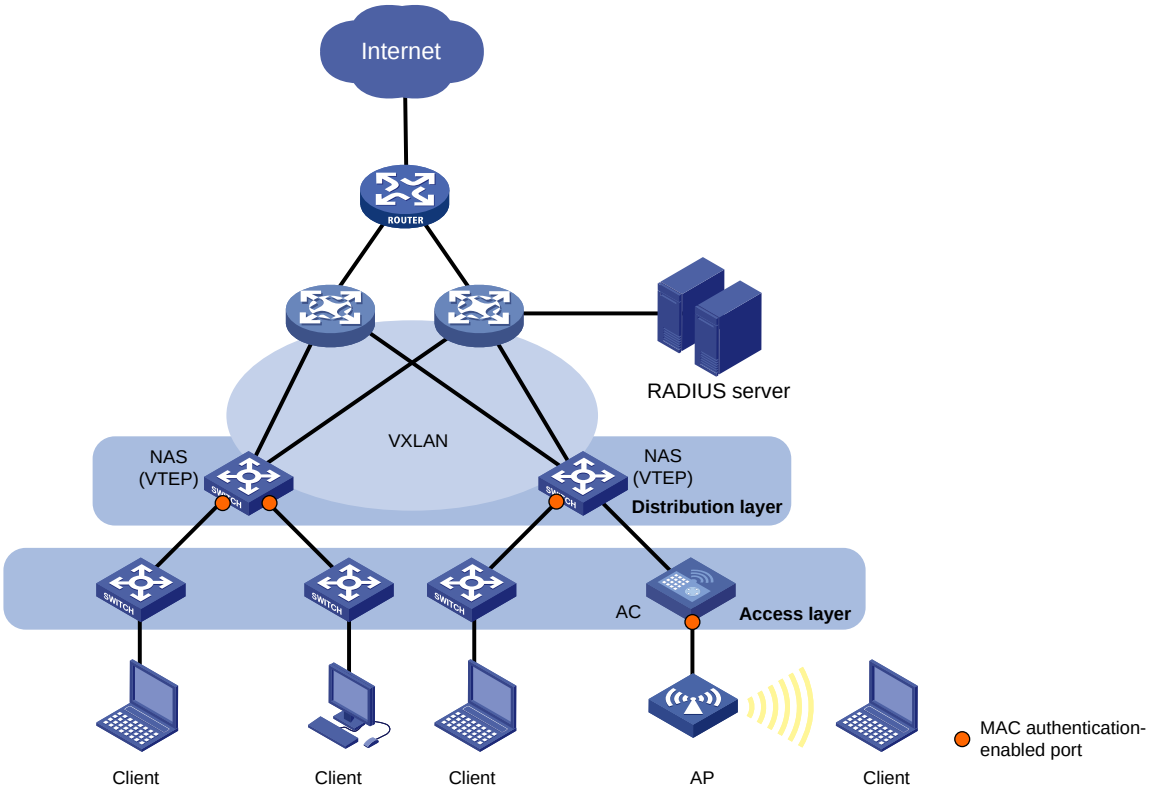
VSI manipulation

MAC authentication support for VXLANs

As shown in Figure 4, when the device acts as both a VXLAN VTEP and a NAS, users' service information cannot be identified by VLANs. To resolve this issue, you must configure the RADIUS server to assign VSIs to MAC authenticated users. The NAS will map a user's traffic to the VXLAN that is associated with the user's authorization VSI. The mapping criteria include the user's access VLAN, access port, and MAC address.

For information about VSIs and VXLANs, see *VXLAN Configuration Guide*.

Figure 4 VXLAN network diagram for MAC authentication



Authorization VSI

An authorization VSI is associated with a VXLAN that has network resources inaccessible to unauthenticated users.

MAC authentication supports remote VSI authorization. If the VTEP does not receive authorization VSI information for a MAC authentication user from the remote server, the user cannot access resources in any VXLAN after passing authentication. If the VTEP receives authorization VSI information for the user from the remote server, it performs the following operations:

- Dynamically creates an Ethernet service instance according to the user's access port, VLAN, and MAC address.
- Maps the Ethernet service instance to the authorization VSI.

  The user then can access resources in the VXLAN associated with the authorization VSI.

For information about dynamic creation of Ethernet service instances, see *VXLAN configuration Guide*.

Guest VSI

The MAC authentication guest VSI on a port accommodates users that have failed MAC authentication for any reason other than server unreachable. For example, the VSI accommodates users with invalid passwords entered.

You can deploy a limited set of network resources in the VXLAN that is associated with the MAC authentication guest VSI. For example, a software server for downloading software and system patches.

Table 5 shows the way that the VTEP handles guest VSIs for MAC authentication users.

Table 5 VSI manipulation

Authentication status	VSI manipulation
A user fails MAC authentication for any reason other than server unreachable.	The VTEP maps the user's MAC address and access VLAN to the MAC authentication guest VSI.
A user in the MAC authentication guest VSI fails MAC authentication for any reason other than server unreachable.	The user is still in the MAC authentication guest VSI.
A user in the MAC authentication guest VSI passes MAC authentication.	The VTEP remaps the user's MAC address and access VLAN to the authorization VSI assigned by the authentication server.

Critical VSI

The MAC authentication critical VSI on a port accommodates users that have failed MAC authentication because no RADIUS authentication servers are reachable. Users in a MAC authentication critical VSI can access only network resources in the VXLAN associated with this VSI.

The critical VSI feature takes effect when MAC authentication is performed only through RADIUS servers. If a MAC authentication user fails local authentication after RADIUS authentication, the user is not assigned to the critical VSI. For more information about the authentication methods, see "Configuring AAA."

Table 6 shows the way that the VTEP handles critical VSIs for MAC authentication users.

Table 6 VSI manipulation

Authentication status	VSI manipulation
A user fails MAC authentication because all the RADIUS servers are unreachable.	The VTEP maps the user's MAC address and access VLAN to the MAC authentication critical VSI. The user is still in the MAC authentication critical VSI if the user fails MAC reauthentication because all the RADIUS servers are unreachable. If no MAC authentication critical VSI is configured, the device logs off the user.
A user in the MAC authentication critical VSI fails MAC authentication for any reason other than server unreachable.	If a guest VSI has been configured, the VTEP maps the user's MAC address and access VLAN to the guest VSI. If no guest VSI is configured, the VTEP logs off the user.
A user in the MAC authentication guest VSI fails authentication because all the RADIUS servers are unreachable.	The user remains in the MAC authentication guest VSI.

Authentication status	VSI manipulation
A user in the MAC authentication critical VSI passes MAC authentication.	The VTEP remaps the user's MAC address and access VLAN to the authorization VSI assigned by the authentication server.

Microsegment assignment

Authorization microsegment

The microsegmentation feature provides group-based traffic control by applying policy to groups of traffic that has same attributes. A microsegment is a group of traffic that matches the same criteria. The RADIUS server can assign a microsegment to a user after that user passes MAC authentication, Then, the device performs microsegment-based traffic filtering for that user on the user's access port to make sure the user can access only the network resources available for the microsegment.

For more information about microsegmentation, see "Configuring microsegmentation."

IMPORTANT:

The following restrictions apply to MAC authentication on a microsegmented network:

- The server must assign an authorization ACL, redirect URL, and microsegment to a MAC user if that user must pass Web authentication.
- If the server assigns both an authorization ACL and an authorization microsegment to a MAC authentication user that does not require Web authentication, only the authorization microsegment takes effect.

Critical microsegment

The MAC authentication critical microsegment on a port accommodates users that have failed MAC authentication because no RADIUS authentication servers are reachable. Users in a MAC authentication critical microsegment can access only network resources in the critical microsegment. The microsegment can belong to a VLAN or VXLAN.

The critical microsegment feature takes effect when MAC authentication is performed only through RADIUS servers. If a MAC authentication user fails local authentication after RADIUS authentication, the user is not assigned to the critical microsegment. For more information about the authentication methods, see "Configuring AAA."

The access device assigns or removes a MAC authentication user to or from the critical microsegment depending on their authentication status, as shown in Table 7.

Table 7 VSI manipulation

Authentication status	Microsegment manipulation
A user fails MAC authentication because all the RADIUS servers are unreachable.	The access device maps the MAC address of the user to the MAC authentication critical microsegment. The MAC address of the user is still in the MAC authentication critical microsegment if the user fails MAC reauthentication because all the RADIUS servers are unreachable.
A user in the MAC authentication critical microsegment fails MAC authentication for any reason other than server unreachable.	The access device removes the MAC address of the user from the MAC authentication critical microsegment.

Authentication status	Microsegment manipulation
A user in the MAC authentication critical microsegment passes MAC authentication.	The access device maps the MAC address of the user to the authorization microsegment assigned by the authentication server.

ACL assignment

You can specify an authorization ACL for a MAC authentication user on a remote server or the access device to control the user's access to network resources. After the user passes MAC authentication, the server or access device assigns the authorization ACL to the user access port. Then, the port permits or drops the matching traffic for the user depending on the rules in the ACL.

The device supports assignment of static and dynamic authorization ACLs.

- **Assignment of static authorization ACLs**—Static ACLs can be assigned by a RADIUS server or the access device. When the server or access device assigns a static ACL to a user, it assigns only the ACL number. You must manually create the ACL and configure its rules on the access device.

  To change the access permissions of a user, use one of the following methods:

  - Modify ACL rules in the authorization ACL on the access device.
  - Assign another ACL to the user as the authorization ACL from the RADIUS server or the access device.

  Static ACLs and their rules can be manually deleted from the access device.
- **Assignment of dynamic authorization ACLs**—Dynamic ACLs and their rules are automatically deployed by a RADIUS server, which are not configurable on the access device. Dynamic ACLs can only be named ACLs. After the device receives a server-deployed dynamic ACL and its rules, it automatically creates the ACL and configures its rules.

  If a dynamic ACL assigned by the server to a user has the same name as a static ACL, the dynamic ACL cannot be issued and the user cannot come online.

  A dynamic ACL and its rules are automatically deleted from the access device after all its users go offline.

  Dynamic ACLs and their rules cannot be manually modified or deleted on the access device. To display information about dynamic ACLs and their rules, use the **`display mac-authentication connection`** or **`display acl`** command.

IMPORTANT:

The supported authorization ACLs include the following types:

- Basic ACLs, which are numbered in the range of 2000 to 2999.
- Advanced ACLs, which are numbered in the range of 3000 to 3999.
- Layer 2 ACLs, which are numbered in the range of 4000 to 4999.
- User-defined ACLs, which are numbered in the range of 5000 to 5999.

For an authorization ACL to take effect, make sure the ACL exists with rules and none of the rules contains the **`counting`**, **`established`**, **`fragment`**, **`source-mac`**, **`cos`**, **`dest-mac`**, **`lsap`**, or **`logging`** keyword.

For more information about ACLs, see *ACL and QoS Command Reference*.

IMPORTANT:

If the server assigns both an authorization ACL and microsegment to a MAC authentication user on a microsegmented network, only the authorization microsegment takes effect.

Redirect URL assignment

The device supports the URL attribute assigned by a RADIUS server. During MAC authentication, the HTTP or HTTPS requests of a user are redirected to the Web interface specified by the server-assigned URL attribute. After the user passes the Web authentication, the RADIUS server records the MAC address of the MAC authentication user and uses a DM (Disconnect Message) to log off the user. When the user initiates MAC authentication again, it will pass the authentication and come online successfully.

By default, the device listens to port 6654 for HTTPS requests to be redirected. To change the redirect listening port number, see configuring HTTP redirect in *Layer 3—IP Services Configuration Guide*.

Blackhole MAC attribute assignment

The device supports the blackhole MAC attribute assigned by the RADIUS authentication server through CoA messages for users that have passed MAC authentication. Upon receiving a CoA message that contains the blackhole MAC attribute for a user that has passed MAC authentication, the device performs the following operations:

1. Logs off the user.
2. Marks the MAC address of the user as a silent MAC address and starts a quiet timer for the MAC address.

The quiet timer is 10 minutes and is not user configurable. The device drops all packets from the MAC address after the quiet timer starts, and it will not authenticate the MAC address until the quiet timer expires.

To display silent MAC addresses, use the **`display mac-authentication`** command.

Periodic MAC reauthentication

Periodic MAC reauthentication tracks the connection status of online users, and updates the authorization attributes assigned by the RADIUS server. The attributes include the ACL and VLAN.

The device reauthenticates online MAC authentication users at the periodic reauthentication interval when the periodic MAC reauthentication feature is enabled. The interval is controlled by a timer and the timer is user configurable. A change to the periodic reauthentication timer applies to online MAC authentication users only after the old timer expires and the MAC authentication users pass authentication.

The server-assigned RADIUS Session-Timeout (attribute 27) and Termination-Action (attribute 29) attributes together can affect the periodic MAC reauthentication feature. To display the server-assigned Session-Timeout and Termination-Action attributes, use the **`display mac-authentication connection`** command.

- If the termination action is to log off users, periodic MAC reauthentication takes effect only when the periodic reauthentication timer is shorter than the session timeout timer. If the session timeout timer is shorter, the device logs off online authenticated users when the session timeout timer expires.
- If the termination action is to reauthenticate users, the periodic MAC reauthentication configuration on the device cannot take effect. The device reauthenticates online MAC authentication users after the server-assigned session timeout timer expires.

If no session timeout timer is assigned by the server, whether the device performs periodic MAC reauthentication depends on the periodic MAC reauthentication configuration on the device. Support for the assignment of Session-Timeout and Termination-Action attributes depends on the server model.

With the RADIUS DAS feature enabled, the device immediately reauthenticates a user upon receiving a CoA message that carries the reauthentication attribute from a RADIUS authentication server. In this case, reauthentication will be performed regardless of whether periodic MAC reauthentication is enabled on the device. For more information about RADIUS DAS configuration, see "Configuring AAA."

By default, the device logs off online MAC authentication users if no server is reachable for MAC reauthentication. The keep-online feature keeps authenticated MAC authentication users online when no server is reachable for MAC reauthentication.

The VLANs assigned to an online user before and after reauthentication can be the same or different.

RESTful server-assisted MAC authentication user recovery

Applicable scenario

After online dumb terminals such as printers go offline, they cannot reauthenticate to come online until after they send a service packet. This might cause the dumb terminals to stay offline for a long time after the device or their attached interface module reboots or after their attached interfaces fail.

To avoid this situation, you can configure the device to work with the authentication server and a RESTful server to trigger MAC reauthentication for dumb terminals or retain their online state.

NOTE:

This feature is available only when you use IMC server as the authentication server and the RESTful server.

Operating mechanism

The following information describes how RESTful server-assisted MAC authentication user recovery works:

1. The authentication server includes the **shutdown-keep-online** proprietary attribute in the authorization attributes assigned to a dumb terminal after it passes authentication.
2. After receiving the authentication result from the authentication server, the device allows the dumb terminal user to come online and marks it as a dumb terminal user.
3. When a device or interface issue occurs, the device handles the dumb terminal user as follows:
   - If the device or the interface module reboots or if the interface recovers from a failure, the device obtains MAC authentication user information from the RESTful server to perform MAC reauthentication.
   - If the interface is physical down, the device retains the online state of the dumb terminal user.

Restrictions and guidelines: MAC authentication configuration

When you configure MAC authentication on an interface, follow these restrictions and guidelines:

- MAC authentication is supported only on Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces.

- If you add a Layer 2 Ethernet interface to an aggregation group, the MAC authentication settings configured on it will not take effect until it is removed from the aggregation group.
- Do not delete a Layer 2 aggregate interface if the interface has online MAC authentication users.
- Do not change the link type of a port when the MAC authentication guest VLAN or critical VLAN on the port has users.

When you configure MAC authentication VLANs or VSIs, follow these restrictions and guidelines:

- If the server assigns both an authorization VSI and authorization VLAN to a user, the device uses only the authorization VLAN.
- On a port, the following MAC authentication settings are mutually exclusive:
  - Guest and critical VLAN settings.
  - Guest and critical VSI settings.
  - Critical microsegment settings.
- On a port, the MAC authentication guest and critical VLAN settings are mutually exclusive with the 802.1X guest, Auth-Fail, and critical VSI settings.
- On a port, the MAC authentication guest and critical VSI settings are mutually exclusive with the 802.1X guest, Auth-Fail, and critical VLAN settings.
- To ensure a successful authentication, you must configure the authentication server to assign authorization VLANs or VSIs to the MAC authentication users attached to a port in the following situations:
  - If the MAC authentication-enabled port is configured with the guest VLAN or critical VLAN, configure the authentication server to assign authorization VLANs to MAC authentication users.
  - If the MAC authentication-enabled port is configured with the guest VSI or critical VSI, configure the authentication server to assign authorization VSIs to MAC authentication users.
  - If both 802.1X and MAC authentication are configured on the port, be careful with the VSI settings of the two authentication features.
    - You must configure VSI authorization for the MAC authentication users if the port has an Auth-Fail, guest, or critical VSI for 802.1X authentication.
    - Likewise, you must configure VSI authorization for the 802.1X users if the port has a guest or critical VSI for MAC authentication.
- As a best practice, do not use Web authentication in conjunction with any of the following MAC authentication settings on a port:
  - Guest VLAN and critical VLAN settings.
  - Guest VSI and critical VSI settings.
  - Critical microsegment settings.
- As a best practice, do not use port security preauthentication domain in conjunction with any of the following MAC authentication settings on a port:
  - Guest VLAN and critical VLAN settings.
  - Guest VSI and critical VSI settings.
  - Critical microsegment settings.
  - Critical profile settings.

If the MAC address that has failed authentication is a static MAC address or a MAC address that has passed any security authentication, the device does not mark the MAC address as a silent address.

To ensure a successful HTTPS redirect for users who have received a server-assigned redirect URL, make sure VLAN interfaces or VSI interfaces exist for the VLANs or VSIs that transport their packets, respectively.

When RADIUS DAS is enabled, an M-LAG system does not support shutting down or rebooting the access ports of MAC authentication users or reauthenticating the users through CoA messages. For more information about RADIUS DAS, see "Configuring AAA."

MAC authentication tasks at a glance

To configure MAC authentication, perform the following tasks:

1. Enabling MAC authentication
2. Configure basic MAC authentication features
   - Specifying a MAC authentication method
   - Specifying a MAC authentication domain
   - Configuring user account policy
   - (Optional.) Configuring MAC authentication timers
   - (Optional.) Configuring periodic MAC reauthentication
3. (Optional.) Configuring MAC authentication VLAN assignment
   - Configuring a MAC authentication guest VLAN
   - Configuring a MAC authentication critical VLAN
4. (Optional.) Configuring MAC authentication VSI assignment
   - Configuring a MAC authentication guest VSI
   - Configuring a MAC authentication critical VSI
5. (Optional.) Configuring a MAC authentication critical microsegment
6. (Optional.) Configuring a MAC authentication access profile
7. (Optional.) Configuring other MAC authentication features
   - Configuring a MAC authentication critical microsegment
   - Configuring unauthenticated MAC authentication user aging
   - Configuring MAC authentication offline detection
   - Enabling online MAC authentication users to escape from offline detection and stay online when no reachable RADIUS authentication servers are available
   - Enabling online user synchronization for MAC authentication
   - Setting the maximum number of concurrent MAC authentication users on a port
   - Enabling MAC authentication multi-VLAN mode on a port

     Perform this task to not reauthenticate online users when VLAN changes occur on a port.
   - Configuring MAC authentication delay
   - Including user IP addresses in MAC authentication requests
   - Enabling parallel MAC authentication and 802.1X authentication
   - Configuring RESTful server-assisted MAC authentication user recovery
   - Tracking the availability of the Web authentication server at a redirect URL
   - Configuring Web proxy ports for URL redirection in MAC authentication
   - Logging off MAC authentication users
   - Enabling MAC authentication user logging

Prerequisites for MAC authentication

Before you configure MAC authentication, complete the following tasks:

1. Make sure the port security feature is disabled. For more information about port security, see "Configuring port security."
2. Configure an ISP domain and specify an AAA method. For more information, see "Configuring AAA."
   - For local authentication, you must also create local user accounts (including usernames and passwords) and specify the **lan-access** service for local users.
   - For RADIUS authentication, make sure the device and the RADIUS server can reach each other and create user accounts on the RADIUS server. If you are using MAC-based accounts, make sure the username and password for each account are the same as the MAC address of each MAC authentication user.

Enabling MAC authentication

Restrictions and guidelines

For MAC authentication to take effect on a port, you must enable this feature globally and on the port.

MAC authentication cannnot take effect on a port if the device has run out of ACL resources when you perform either of the following operations:

- Enable MAC authentication on the port while MAC authentication has been enabled globally.
- Enable MAC authentication globally in system while MAC authentication has been enabled on the port.

Procedure

1. Enter system view.
   `system-view`
2. Enable MAC authentication globally.
   `mac-authentication`
   By default, MAC authentication is disabled globally.
3. Enter interface view.
   `interface` *`interface-type interface-number`*
4. Enable MAC authentication on the port.
   `mac-authentication`
   By default, MAC authentication is disabled on a port.

Specifying a MAC authentication method

About this task

RADIUS-based MAC authentication supports the following authentication methods:

- **PAP**—Transports usernames and passwords in plain text. The authentication method applies to scenarios that do not require high security.
- **CHAP**—Transports usernames in plain text and passwords in encrypted form over the network. CHAP is more secure than PAP.

Restrictions and guidelines

The device must use the same authentication method as the RADIUS server.

Procedure

1. Enter system view.
   `system-view`

2. Specify an authentication method for MAC authentication.

   mac-authentication authentication-method { **chap** | **pap** }

   By default, the device uses PAP for MAC authentication.

Specifying a MAC authentication domain

About this task

By default, MAC authentication users are in the system default authentication domain. To implement different access policies for users, you can use one of the following methods to specify authentication domains for MAC authentication users:

- Specify a global authentication domain in system view. This domain setting applies to all ports enabled with MAC authentication.
- Specify an authentication domain for an individual port in interface view.

MAC authentication chooses an authentication domain for users on a port in this order: the port-specific domain, the global domain, and the default domain. For more information about authentication domains, see "Configuring AAA."

Procedure

1. Enter system view.

   system-view

2. Specify an authentication domain for MAC authentication users.
   - In system view:

     mac-authentication domain *domain-name*
   - In interface view:

     interface *interface-type interface-number*

     mac-authentication domain *domain-name*

   By default, the system default authentication domain is used for MAC authentication users.

Configuring user account policy

Restrictions and guidelines

For users in a MAC address range, the MAC address range-specific user account has higher priority than the global user account settings.

You can configure a maximum of 16 MAC address ranges and must make sure the MAC address ranges do not overlap.

If you configure user account settings multiple times for the same MAC address range, the most recent configuration overwrites the previous configuration.

The MAC range-specific accounts apply only to unicast MAC addresses.

- If you specify a MAC address range that contains only multicast MAC addresses, execution of this command will fail.
- If you specify a MAC address range that contains both unicast and multicast MAC addresses, the command takes effect only on unicast MAC addresses.

The all-zero MAC address is invalid for MAC authentication. Users with the all-zero MAC address cannot pass MAC authentication.

Procedure

1. Enter system view.

system-view

2. Configure the global MAC authentication user account policy.
   - Use one MAC-based user account for each user.

     mac-authentication user-name-format mac-address [{ **with-hyphen** [**separator colon**] | **without-hyphen** } [**lowercase** | **uppercase**]] [**password** { **cipher** | **simple** } *string*]
   - Use one shared user account for all users.

     mac-authentication user-name-format fixed [**account** *name*] [**password** { **cipher** | **simple** } *string*]

   By default, the device uses the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation without hyphens, and letters are in lower case.
3. Specify one shared user account specific to a MAC address range.

   mac-authentication mac-range-account mac-address *mac-address* **mask** { *mask* | *mask-length* } **account** *name* **password** { **cipher** | **simple** } *string*

   By default, no username or password is configured specific to a MAC address range. The global user account policy applies to the users.

Configuring MAC authentication timers

About this task

MAC authentication uses the following timers:

- **Offline detection timer**—Sets the interval that the device must wait for traffic from a user before the device determines that the user is idle. If the device has not received traffic from a user before the timer expires, the device logs off that user and requests the accounting server to stop accounting for the user. This timer takes effect only when the MAC authentication offline detection feature is enabled.

  As a best practice, set the MAC address aging timer to the same value as the offline detection timer. This operation prevents a MAC authenticated user from being logged off within the offline detection interval because of MAC address entry expiration.
- **Quiet timer**—Sets the interval that the device must wait before the device can perform MAC authentication for a user that has failed MAC authentication. All packets from the MAC address are dropped during the quiet time. This quiet mechanism prevents repeated authentication from affecting system performance.
- **Server timeout timer**—Sets the interval that the device waits for a response from a RADIUS server before the device determines that the RADIUS server is unavailable. If the timer expires during MAC authentication, the user fails MAC authentication.
- **Aging timer for temporary MAC authentication users**—Sets the aging timer for temporary MAC authentication users. This timer starts when a user passes MAC authentication on a port operating in **macAddressAndUserLoginSecureExt** mode. If the device has not received 802.1X protocol packets from the MAC authenticated user on that port before the timer expires, the device determines that the user has failed authentication. Then, the device logs off the user. For more information about the port security mode, see "Configuring port security."

Restrictions and guidelines

To avoid forced logoff before the server timeout timer expires, set the server timeout timer to a value that is lower than or equal to the product of the following values:

- The maximum number of RADIUS packet transmission attempts set by using the **retry** command in RADIUS scheme view.

- The RADIUS server response timeout timer set by using the **timer response-timeout** command in RADIUS scheme view.

For information about setting the maximum number of RADIUS packet transmission attempts and the RADIUS server response timeout timer, see "Configuring AAA."

Procedure

1. Enter system view.

   system-view

2. Configure MAC authentication timers.

   mac-authentication timer { **offline-detect** *offline-detect-value* | **quiet** *quiet-value* | **server-timeout** *server-timeout-value* | **temporary-user-aging** *aging-time-value* }

   The default settings are as follows:

   - The offline detection timer is 300 seconds.
   - The quiet timer is 60 seconds.
   - The server timeout timer is 100 seconds.
   - The temporary user aging timer is 60 seconds.

Configuring periodic MAC reauthentication

Restrictions and guidelines

The device selects a periodic reauthentication timer for MAC reauthentication in the following order:

1. Server-assigned reauthentication timer.
2. Port-specific reauthentication timer.
3. Global reauthentication timer.
4. Default reauthentication timer.

Modification to the MAC authentication domain, MAC authentication method, or user account format setting does not affect the reauthentication of online MAC authentication users. The modified setting takes effect only on MAC authentication users that come online after the modification.

If periodic reauthentication is triggered for a user while that user is waiting for online synchronization, the system performs online synchronization and does not perform reauthentication for the user.

Procedure

1. Enter system view.

   system-view

2. Set the periodic MAC reauthentication timer.

   - Set a global periodic reauthentication timer.

     mac-authentication timer reauth-period *reauth-period-value*

     The default setting is 3600 seconds.

   - Execute the following commands in sequence to set a port-specific periodic reauthentication timer:

     interface *interface-type interface-number*

     mac-authentication timer reauth-period *reauth-period-value*

     quit

     By default, no periodic MAC reauthentication timer is set on a port. The port uses the global periodic MAC reauthentication timer.

3. Enter interface view.

interface *interface-type interface-number*

4. Enable periodic MAC reauthentication.

   mac-authentication re-authenticate

   By default, periodic MAC reauthentication is disabled on a port.

5. (Optional.) Enable the keep-online feature for MAC authenticated users on the port.

   mac-authentication re-authenticate server-unreachable keep-online

   By default, the keep-online feature is disabled. The device logs off online MAC authentication users if no server is reachable for MAC reauthentication.

   In a fast-recovery network, you can use the keep-online feature to prevent MAC authentication users from coming online and going offline frequently.

Configuring a MAC authentication guest VLAN

Specifying a MAC authentication guest VLAN on a port

Restrictions and guidelines

When you configure the MAC authentication guest VLAN on a port, follow the guidelines in Table 8.

Table 8 Relationships of the MAC authentication guest VLAN with other security features

Feature	Relationship description	Reference
Quiet feature of MAC authentication	The MAC authentication guest VLAN feature has higher priority. When a user fails MAC authentication, the user can access the resources in the guest VLAN. The user's MAC address is not marked as a silent MAC address.	See "Configuring MAC authentication timers."
Port intrusion protection	The guest VLAN feature has higher priority than the block MAC action but lower priority than the shutdown port action of the port intrusion protection feature.	See "Configuring port security."

Prerequisites

Before you configure the MAC authentication guest VLAN on a port, complete the following tasks:

- Create the VLAN to be specified as the MAC authentication guest VLAN.
- Configure the port as a hybrid port, and configure the VLAN as an untagged member on the port.
- Enable MAC-based VLAN on the port.

For information about VLAN configuration, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify the MAC authentication guest VLAN on the port.

   mac-authentication guest-vlan *guest-vlan-id*

   By default, no MAC authentication guest VLAN is specified on a port.

You can configure only one MAC authentication guest VLAN on a port. The MAC authentication guest VLANs on different ports can be different.

Enabling guest VLAN reauthentication in MAC authentication

About this task

The guest VLAN reauthentication feature of MAC authentication enables the device to reauthenticate users in the MAC authentication guest VLAN on a port at reauthentication intervals.

By default, this feature is enabled. You can disable it as needed to suppress excessive authentication failure log messages, which might occur when a network issue results in a large number of reauthentication failures.

If guest VLAN reauthentication is disabled on a port, the device does not reauthenticate users in the MAC authentication guest VLAN on the port. The guest VLAN users will stay in the guest VLAN until they age out. To control guest VLAN user aging functionality, use the **mac-authentication unauthenticated-user aging enable** command. To configure the aging timer, use the **mac-authentication timer user-aging guest-vlan** *aging-time-value* command.

Restrictions and guidelines

As a best practice, set the reauthentication interval to a value greater than 30 seconds if the number of concurrent MAC authentication users on a port is likely to exceed 300.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the guest VLAN reauthentication feature of MAC authentication on the port.
   mac-authentication guest-vlan re-authenticate
   By default, the guest VLAN reauthentication feature of MAC authentication is enabled on a port.
4. Set the authentication interval for users in the MAC authentication guest VLAN.
   mac-authentication guest-vlan auth-period *period-value*
   The default setting is 30 seconds.

Configuring a MAC authentication critical VLAN

Restrictions and guidelines

When you configure the MAC authentication critical VLAN on a port, follow the guidelines in Table 9.

Table 9 Relationships of the MAC authentication critical VLAN with other security features

Feature	Relationship description	Reference
Quiet feature of MAC authentication	The MAC authentication critical VLAN feature has higher priority. When a user fails MAC authentication because no RADIUS authentication server is reachable, the user can access the resources in the critical VLAN. The user's MAC address is not marked as a silent MAC address.	See "Configuring MAC authentication timers."
Port intrusion protection	The critical VLAN feature has higher priority than the block MAC action but lower priority than the	See "Configuring port security."

Feature	Relationship description	Reference
	shutdown port action of the port intrusion protection feature.	

Prerequisites

Before you configure the MAC authentication critical VLAN on a port, complete the following tasks:

- Create the VLAN to be specified as the MAC authentication critical VLAN.
- Configure the port as a hybrid port, and configure the VLAN as an untagged member on the port.
- Enable MAC-based VLAN on the port.

For information about VLAN configuration, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify the MAC authentication critical VLAN on the port.

   mac-authentication critical vlan *critical-vlan-id* [**url-user-logoff**]

   By default, no MAC authentication critical VLAN is specified on a port.

   You can configure only one MAC authentication critical VLAN on a port. The MAC authentication critical VLANs on different ports can be different.

Configuring a MAC authentication guest VSI

Specifying a MAC authentication guest VSI on a port

Restrictions and guidelines

The MAC authentication guest VSI feature has higher priority than the quiet feature of MAC authentication. When a user fails MAC authentication, the user can access the resources in the guest VSI. The user's MAC address is not marked as a silent MAC address.

You can configure only one MAC authentication guest VSI on a port. The MAC authentication guest VSIs on different ports can be different.

Prerequisites

To ensure that the MAC authentication guest VSI on a port can take effect, complete the following tasks:

- Enable L2VPN.
- Create the VSI to be specified as the MAC authentication guest VSI, and create a VXLAN for the VSI.
- Enable MAC-based traffic match mode for dynamic Ethernet service instances on the port.

For more information, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the MAC authentication guest VSI on the port.
   mac-authentication guest-vsi *guest-vsi-name*
   By default, no MAC authentication guest VSI exists on a port.

Enabling guest VSI reauthentication in MAC authentication

About this task

The guest VSI reauthentication feature of MAC authentication enables the device to reauthenticate users in the MAC authentication guest VSI on a port at reauthentication intervals.

You can enable guest VSI reauthentication by using the **mac-authentication guest-vsi re-authenticate** command or disable the feature by using the **undo** form of the command.

Typically, you disable this feature to suppress excessive authentication failure log messages, which might occur when a network issue results in a large number of reauthentication failures.

If guest VSI reauthentication is disabled on a port, the device does not reauthenticate users in the MAC authentication guest VSI on the port. The guest VSI users will stay in the guest VSI until they age out. To configure the aging timer, use the **mac-authentication timer user-aging guest-vsi** *aging-time-value* command.

Restrictions and guidelines

As a best practice, set the reauthentication interval to a value greater than 30 seconds if the number of concurrent MAC authentication users on a port is likely to exceed 300.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable the guest VSI reauthentication feature of MAC authentication on the port.
   mac-authentication guest-vsi re-authenticate
   By default, the guest VSI reauthentication feature of MAC authentication is enabled on a port.
4. Set the authentication interval for users in the MAC authentication guest VSI.
   mac-authentication guest-vsi auth-period *period-value*
   The default setting is 30 seconds.

Configuring a MAC authentication critical VSI

Restrictions and guidelines

The MAC authentication critical VSI feature has higher priority than the quiet feature of MAC authentication. When a user fails MAC authentication, the user can access the resources in the critical VSI. The user's MAC address is not marked as a silent MAC address.

You can configure only one MAC authentication critical VSI on a port. The MAC authentication critical VSIs on different ports can be different.

Prerequisites

To ensure that the MAC authentication critical VSI on a port can take effect, complete the following tasks:

- Enable L2VPN.
- Create the VSI to be specified as the MAC authentication critical VSI, and create a VXLAN for the VSI.
- Enable MAC-based traffic match mode for dynamic Ethernet service instances on the port.

For more information, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the MAC authentication critical VSI on the port.
   mac-authentication critical vsi *critical-vsi-name* [**url-user-logoff**]
   By default, no MAC authentication critical VSI exists on a port.

Specifying a critical profile for MAC authentication users on a port

About this task

A critical profile specifies a limited set of resources accessible to MAC authentication users when they cannot get authenticated because none of the applicable RADIUS servers is available.

To make sure the MAC authentication users attached to a port can access the resources in a critical profile in the server unavailable situation, perform this task.

Restrictions and guidelines

You can specify only one critical profile for MAC authentication on a port. The critical profiles for MAC authentication on different ports can be different.

On an interface, the critical profile settings are mutually exclusive with the critical VSI, critical VLAN, and critical microsegment settings for MAC authentication.

For more information, see the command reference and "Restrictions and guidelines: MAC authentication configuration."

Prerequisites

As a best practice, configure the critical profile before you perform this task. For more information about configuring critical profiles, see "Configuring AAA."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify a critical profile for MAC authentication users on the port.
   mac-authentication critical profile *profile-name*
   By default, no critical profile is specified for MAC authentication users on a port.

Configuring a MAC authentication critical microsegment

Restrictions and guidelines

The MAC authentication critical microsegment feature has higher priority than the MAC authentication quiet feature. If the device adds a MAC address to a critical microsegment, it will not add that MAC address to the quiet MAC address list.

You can configure only one MAC authentication critical microsegment on a port. The MAC authentication critical microsegments on different ports can be the same or different.

Prerequisites

For the critical microsegment you specify for MAC authentication on a port to take effect, you must create that microsegment. For more information about microsegments, see "Configuring microsegmentation."

To specify a VSI to match user packets in VSIs, perform the following tasks:

- Enable L2VPN.
- Enable MAC-based traffic match mode for dynamic Ethernet service instances on the port.

For more information, see *VXLAN Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify the MAC authentication critical microsegment on the port.

   mac-authentication critical microsegment *microsegment-id* [**vsi** *vsi-name*] [**url-user-logoff**]

   By default, no MAC authentication critical microsegment is configured on a port.

Configuring a MAC authentication access profile

About this task

The device uses MAC authentication access profiles to uniformly manage configurations related to MAC authentication access. After you bind a MAC authentication access profile to a port security authentication profile, the interface bound to the port security authentication profile will use the access profile's configuration to authenticate users.

After you bind a port security authentication profile to an interface:

- For settings that can be configured both in interface view and MAC authentication access profile view (the command lines might differ in the views), the system automatically deletes the settings in interface view regardless of whether the port security authentication profile has been bound to a MAC authentication access profile. To use these settings, configure them in the MAC authentication access profile bound to the port security authentication profile.
- For settings supported in interface view but not in MAC authentication access profile view, the settings in interface view still take effect.

For more information about port security authentication profiles, see "Configuring port security."

Restrictions and guidelines

If a MAC authentication access profile bound to a port security authentication profile has taken effect, deleting the access profile will cause users who have come online using this access profile to disconnect abnormally.

A MAC authentication access profile can be bound to multiple port security authentication profiles, but a port security authentication profile can be bound to only one MAC authentication access profile. To bind another MAC authentication access profile to a port security authentication profile, you must first unbind the existing one.

Procedure

1. Enter system view.

   system-view

2. Create a MAC authentication access profile and enter its view.

   mac-authentication access-profile name *profile-name*

   By default, no MAC authentication access profiles exist.

3. Configure supported MAC authentication features as needed.

   Only some MAC authentication features are supported in MAC authentication access profile view. For more information, see the command reference.

4. Enter a port security authentication profile view.

   port-security authentication-profile name *profile-name*

5. Bind the MAC authentication access profile to the port security authentication profile.

   mac-authentication access-profile *profile-name*

   By default, no MAC authentication access profile is bound to the port security authentication profile.

Configuring unauthenticated MAC authentication user aging

About this task

Unauthenticated MAC authentication user aging applies to users added to the following locations because they have not been authenticated or have failed authentication:

- MAC authentication guest or critical VLANs.
- MAC authentication guest or critical VSIs.
- MAC authentication critical microsegments.

When a user in one of those locations ages out, the device removes the user from the location and deletes the MAC address entry for the user from the access port.

For users in one of those locations on one port to be authenticated successfully and come online on another port, enable this feature. In any other scenarios, disable this feature as a best practice.

Restrictions and guidelines

Users removed from an access control segment upon aging timer expiration will be unable to access the network resources for that segment until after another authentication is triggered. As a best practice, use this feature on a port only if you want to have its unauthenticated users to be authenticated and come online on a different port.

Procedure

1. Enter system view.

   system-view

2. Set the user aging timer for a type of MAC authentication VLAN, VSI, or microsegment.

   mac-authentication timer user-aging { **critical-microsegment** | **critical-vlan** | **critical-vsi** | **guest-vlan** | **guest-vsi** } *aging-time-value*

   By default, the user aging timer is 1000 seconds for all applicable types of MAC authentication VLANs, VSIs, and microsegments.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable unauthenticated MAC authentication user aging.

   mac-authentication unauthenticated-user aging enable

   By default, unauthenticated MAC authentication user aging is enabled.

Configuring MAC authentication offline detection

About this task

Enable MAC authentication offline detection to detect idle users on a port. If the port has not received traffic from a user when the offline detection timer expires, the device logs off that user and requests the accounting server to stop accounting for the users. For information about setting the offline detection timer in system view, see "Configuring MAC authentication timers."

Disabling this feature disables the device from inspecting the online user status.

In addition to port-based MAC authentication offline detection, you can configure offline detection parameters on a per-user basis, as follows:

- Set an offline detection timer specific to a user and control whether to use the ARP snooping or ND snooping table to determine the offline state of the user.
  - If the ARP snooping or ND snooping table is used, the device searches the ARP snooping or ND snooping table before it checks for traffic from the user within the detection interval. If a matching ARP snooping or ND snooping entry is found, the device resets the offline detection timer and the user stays online. If the offline detection timer expires because the device has not found a matching snooping entry for the user or received traffic from the user, the device disconnects the user.
  - If the ARP or ND snooping table is not used, the device disconnects the user if it has not received traffic from that user before the offline detection timer expires.

  When disconnecting the user, the device also notifies the RADIUS server (if any) to stop user accounting.
- Skip offline detection for the user. You can choose this option if the user is a dumb terminal. A dumb terminal might fail to come online again after it is logged off by the offline detection feature.

The device uses the offline detection settings for a user in the following sequence:

1. User-specific offline detection settings.
2. Offline detection settings assigned to the user by the RADIUS server. The settings include the offline detection timer, use of the ARP or ND snooping table in offline detection, and whether to ignore offline detection.
3. Port-based offline detection settings.

MAC authentication offline detection takes effect on both of the following types of users:

- Users in preauthentication domains, authentication domains, critical domains, and Auth-Fail domains.

- Users assigned to guest VLANs or VSIs, Auth-Fail VLANs or VSIs, and critical VLANs, VSIs, or microsegments.

Restrictions and guidelines

For the user-specific offline detection feature to take effect on a user, make sure the MAC authentication offline detection feature is enabled on the user's access port.

The user-specific offline detection settings take effect on the online users immediately after they are configured.

If you enable MAC authentication offline detection on a Layer 2 aggregate interface, delay exists for the device to log off an idle user.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Configure MAC authentication offline detection for a user.

   mac-authentication offline-detect mac-address *mac-address* { **ignore** | **timer** *offline-detect-value* [**check-arp-or-nd-snooping**] }

   By default, offline detection settings configured on access ports take effect and the offline detection timer set in system view is used.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable MAC authentication offline detection.

   mac-authentication offline-detect enable

   By default, MAC authentication offline detection is enabled on a port.

Enabling online MAC authentication users to escape from offline detection and stay online when no reachable RADIUS authentication servers are available

About this task

The offline detection feature for MAC authentication logs off an online MAC authentication user if it has not detected traffic from that user before the offline-detect timer expires. To prevent online MAC authentication users on a port from being forced offline by the offline detection feature and stay online when none of the RADIUS servers in their ISP domain is reachable, perform this task.

Restrictions and guidelines

Use this feature only when RADIUS scheme authentication is configured for the MAC authentication users on a port. If local or no authentication is also configured, offline detection might be undesirably disabled for MAC authentication users that come online through local authentication or without authentication.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable online MAC authentication users to escape from offline detection and stay online when no reachable RADIUS authentication servers are available.

   mac-authentication auth-server-unavailable escape

   By default, online MAC authentication users cannot escape from offline detection (if enabled) when no reachable RADIUS authentication servers are available.

Enabling online user synchronization for MAC authentication

About this task

IMPORTANT:

This feature takes effect only when the device uses an IMC RADIUS server to authenticate MAC authentication users.

To ensure that the RADIUS server maintains the same online MAC authentication user information as the device after the server state changes from unreachable to reachable, use this feature.

This feature synchronizes online MAC authentication user information between the device and the RADIUS server when the RADIUS server state is detected having changed from unreachable to reachable.

When synchronizing online MAC authentication user information on a port with the RADIUS server, the device initiates MAC authentication in turn for each authenticated online MAC authentication user to the RADIUS server.

If synchronization fails for an online user, the device logs off that user unless the failure occurs because the server has become unreachable again.

Restrictions and guidelines

The amount of time required to complete online user synchronization increases as the number of online users grows. This might result in an increased delay for new MAC authentication users and users in the critical VLAN or VSI to authenticate or reauthenticate to the RADIUS server and come online.

To have this feature take effect, you must use it in conjunction with the RADIUS server status detection feature, which is configurable with the **radius-server test-profile** command. For more information about the RADIUS server status detection feature, see "Configuring AAA."

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable online user synchronization for MAC authentication.

   mac-authentication server-recovery online-user-sync

   By default, online user synchronization for MAC authentication is disabled.

Setting the maximum number of concurrent MAC authentication users on a port

About this task

Perform this task to prevent the system resources from being overused.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set the maximum number of concurrent MAC authentication users on the port.

   mac-authentication max-user [**preauth-domain** | **auth-fail-domain**] *max-number*

   The default setting is 4294967295.

Enabling MAC authentication multi-VLAN mode on a port

About this task

By default, MAC authentication single-VLAN mode applies on a port. In this mode, traffic from an online user cannot be sent in different VLANs on a port without service interruption. To accommodate applications that are sensitive to delay or service interruption in a multi-VLAN environment, for example, IP phones, enable MAC authentication multi-VLAN mode.

In multi-VLAN mode, the port forwards traffic from a user in different VLANs without reauthentication if the user has been authenticated and come online in any VLAN on the port. Free of reauthentication, traffic from an online user can be sent in different VLANs without delay or service interruption.

In single-VLAN mode, the port reauthenticates an online user when traffic received from that user contains a VLAN tag different from the VLAN in which the user was authenticated. The authentication process differs depending on the MAC move setting in port security and the authorization VLAN assignment status, as follows:

- If no authorization VLAN has been assigned to the online user, the device first logs off the user and then reauthenticates the user in the new VLAN.
- If the online user has been assigned an authorization VLAN, the device handles the user depending on the MAC move setting in port security.
  - If MAC move is disabled in port security, the user cannot pass authentication and come online from the new VLAN until after it goes offline from the port.
  - If MAC move is enabled in port security, the user can pass authentication on the new VLAN and come online without having to first go offline from the port. After the user passes authentication on the new VLAN, the original authentication session of the user is deleted from the port.

  To enable the port security MAC move feature, use the **port-security mac-move permit** command.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Enable MAC authentication multi-VLAN mode.
   mac-authentication host-mode multi-vlan
   By default, MAC authentication operates in single-VLAN mode on a port.

Configuring MAC authentication delay

About this task

When both 802.1X authentication and MAC authentication are enabled on a port, you can delay MAC authentication so that 802.1X authentication is preferentially triggered.

If no 802.1X authentication is triggered or 802.1X authentication fails within the delay period, the port continues to process MAC authentication.

Restrictions and guidelines

Do not set the port security mode to **mac-else-userlogin-secure** or **mac-else-userlogin-secure-ext** when you use MAC authentication delay. The delay does not take effect on a port in either of the two modes. For more information about port security modes, see "Configuring port security."

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable MAC authentication delay and set the delay timer.
   mac-authentication timer auth-delay *time*
   By default, MAC authentication delay is disabled.

Including user IP addresses in MAC authentication requests

About this task

IMPORTANT:
This feature can only operate in conjunction with an IMC server.

To avoid IP conflicts that result from changes to static IP addresses, use this feature on a port that has MAC authentication users with static IP addresses.

This feature adds user IP addresses to the MAC authentication requests sent to the authentication server. When MAC authentication is triggered for a user, the device checks the user's IP address for invalidity.

- If the IP address is valid, the device sends a MAC authentication request with the IP address included.
- If the IP address is not a valid host IP address or the triggering packet does not contain an IP address, the device does not initiate MAC authentication.

- If the packet is a DHCP packet with a source IP address of all zeros, the device sends a MAC authentication request without including the IP address. In this case, the IMC server does not examine the user IP address when it performs authentication.

Upon receipt of the authentication request that includes a user's IP address, the IMC server compares the user's IP and MAC addresses with its local IP-MAC mappings.

- If an exact match is found or if no match is found, the user passes MAC authentication. In the latter case, the server creates an IP-MAC mapping for the user.
- If a mapping is found for the MAC address but the IP addresses do not match, the user fails the MAC authentication.

After you specify the **exclude-ip acl** keyword, the applied ACL cannot filter out DHCP packets. If a DHCP packet that carries an invalid IP address triggers MAC authentication, user authentication will fail. To resolve this issue, you can specify the **exclude-dhcpv4** or **exclude-dhcpv6** keyword to prevent users from triggering MAC authentication through DHCP packets.

Restrictions and guidelines

Configure DHCP snooping or ARP snooping, so when users' IP addresses change, the device can quickly update the users' IP addresses and send authentication requests with correct user IP addresses to the authentication server. To configure DHCP snooping, execute the **dhcp snooping enable** and **dhcp snooping binding record** commands. To enable ARP snooping, execute the **arp snooping enable** command. For more information about DHCP snooping and ARP snooping, see ARP configuration and DHCP configuration respectively in *Layer 3—IP Services Configuration Guide*.

Do not use this feature in conjunction with the MAC authentication guest VLAN or guest VSI on a port. The device cannot perform MAC authentication for a user once that user is added to the MAC authentication guest VLAN or guest VSI.

You can specify an ACL to identify source IP addresses that can or cannot trigger MAC authentication. When you configure the ACL, follow these guidelines:

- The specified ACL number represents an IPv4 ACL and an IPv6 ACL with the same number. For example, if the ACL number is 2000, you specify both IPv4 ACL 2000 and IPv6 ACL 2000. The IPv4 ACL and the IPv6 ACL will be used to process IPv4 packets and IPv6 packets, respectively.
- Use permit rules to identify source IP addresses that are valid for MAC authentication. Use deny rules to identify source IP addresses that cannot trigger MAC authentication.
- In the rules, only the action keyword (permit or deny) and the source IP match criteria can take effect. If the ACL is a user-defined ACL (numbered in the range of 5000 to 5999), only the **ipv4**, **ipv6**, and **any** source IP match criteria can take effect.
- As a best practice, configure a deny rule to exclude the IPv6 IP addresses that start with fe80 from triggering MAC authentication.
- If you configure permit rules, add a **deny all** rule at the bottom of the ACL.

IMPORTANT:

If the user host is configured with IPv6, the device might receive packets that contain an IPv6 link-local address, which starts with fe80. MAC authentication failure or incorrect MAC-IP binding will occur if this address is used in MAC authentication. To avoid these issues, configure an ACL to exclude the IPv6 IP addresses that start with fe80.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Include user IP addresses in MAC authentication requests.

   mac-authentication carry user-ip [**exclude-ip acl** *acl-number*] [**exclude-dhcpv4**] [**exclude-dhcpv6**]

   By default, a MAC authentication request does not include the user IP address.

Enabling parallel MAC authentication and 802.1X authentication

About this task

This feature enables a port to perform MAC authentication in parallel with 802.1X authentication when it receives a packet from an unknown MAC address.

This feature applies to the following situation in which MAC authentication can be performed only after 802.1X authentication is completed:

- The port is configured with both 802.1X authentication and MAC authentication and performs MAC-based access control for 802.1X authentication.
- The port is enabled with the 802.1X unicast trigger.

With parallel authentication enabled, the port performs MAC authentication while sending a unicast EAP-Request/Identity packet to trigger 802.1 authentication when it receives a packet from an unknown MAC address.

If MAC authentication succeeds before 802.1X authentication is completed, the port is assigned to the MAC authentication authorization VLAN or VSI. When 802.1X authentication completes, the device manipulates VLAN assignment depending on the authentication result, as follows:

- If 802.1X authentication fails, the MAC authentication result takes effect.
- If 802.1X authentication succeeds, the device manipulates VLAN or VSI assignment based on the 802.1X authentication result.

For the port to perform MAC authentication before it is assigned to the 802.1X guest VLAN or VSI, you must also enable new MAC-triggered 802.1X guest VLAN or VSI assignment delay. For information about new MAC-triggered 802.1X guest VLAN or VSI assignment delay, see "Configuring 802.1X."

Restrictions and guidelines

To configure both 802.1X authentication and MAC authentication on a port, use one of the following methods:

- Enable the 802.1X and MAC authentication features separately on the port.
- Enable port security on the port. The port security mode must be **userlogin-secure-or-mac** or **userlogin-secure-or-mac-ext**.

  For information about port security mode configuration, see "Configuring port security."

For the parallel authentication feature to work correctly, do not enable MAC authentication delay on the port. This feature will delay MAC authentication after 802.1X authentication is triggered.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable parallel MAC authentication and 802.1X authentication on the port.

   mac-authentication parallel-with-dot1x

By default, this feature is disabled.

Configuring RESTful server-assisted MAC authentication user recovery

About this task

RESTful server-assisted MAC authentication user recovery supports the following methods:

- **Automatic recovery**—The device automatically obtains MAC authentication user information and performs a reauthentication after the device or an interface module reboots or after an interface recovers from a failure. This feature recovers the online state of MAC authenticated users quickly without waiting for packets from the MAC authentication users to trigger a reauthentication. It is helpful when the network has a large number of dumb terminals or the operation of dumb terminals is important for services.
- **Manual recovery**—You trigger a recovery manually. Then, the device obtains MAC authentication user information from the RESTful server and performs reauthentication. This method is helpful when automatic recovery fails to recover the online state of all MAC authenticated users because of link flapping.

With this feature, you can configure a maximum of 16 profiles for MAC authentication user recovery. Each profile contains a set of parameters for accessing a RESTful server and the URI from which you obtain MAC authentication user information.

Restrictions and guidelines

To use this feature, you must use an IMC server to act as both authentication and RESTful servers.

This feature is mutually exclusive with the RADIUS accounting-on feature. Do not use the two features together. For more information about the RADIUS accounting-on feature, see "Configuring AAA."

To ensure that the MAC authentication user entries obtained from the RESTful server contain user IP addresses, configure the device to include user IP addresses in the MAC authentication requests sent to the authentication server. For more information about this feature, see "Including user IP addresses in MAC authentication requests."

Procedure

1. Enter system view.

   system-view

2. Configure a MAC authentication user-recovery profile:
   a. Create a profile for MAC authentication user recovery and enter its view.

      mac-authentication user-recovery-profile *profile-name*

   b. Configure the IP address and port number of the RESTful server.

      server-address { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**port** *port-number*] [**vpn-instance** *vpn-instance-name*]

      By default, no RESTful servers are configured.

   c. Configure the username and password for accessing the RESTful server.

      login-name *username* [**password** { **cipher** | **simple** } *string*]

      By default, no username or password is configured for accessing the RESTful server.

   d. Configure the NAS IP address used by the device to communicate with the RESTful server.

      nas-ip { *ipv4-address* | **ipv6** *ipv6-address* }

      By default, no NAS IP address is configured for the device to communicate with the RESTful server.

The NAS IP address must be the same as that configured in the authentication RADIUS scheme for the MAC authentication users.

e. Specify the URI for obtaining MAC authentication user information from the RESTful server.

uri *uri-string*

By default, no URIs are specified.

The URI can only be **imcrs/uam/online/notAgingMuteTerminal**. Other URIs cannot take effect.

3. Return to system view.

quit

4. Enable automatic MAC authentication user recovery.

mac-authentication auto-recover-user profile *profile-name*

By default, automatic MAC authentication user recovery is disabled.

5. Return to user view.

quit

6. (Optional.) Manually recover MAC authentication users.

mac-authentication recover-user profile *profile-name* [**interface** *interface-type interface-number*

Tracking the availability of the Web authentication server at a redirect URL

About this task

By default, the device does not track the availability of the Web authentication server at the redirect URL assigned to MAC authentication users. On the device, the state of the Web authentication server will always be active, regardless of its availability. To have the MAC authentication module log off MAC authentication users that have been assigned a redirect URL promptly when the Web authentication server at that URL becomes unavailable, associate that redirect URL with a track entry to detect its connectivity.

As a best practice, configure the track entry associated with the redirect URL to collaborate with an HTTP NQA operation for detection. For more information about Track, see *High Availability Configuration Guide*. For more information about NQA, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Associate a redirect URL for MAC authentication users with a track entry.

mac-authentication redirect-url *url-string* **track** *track-entry-number*

By default, no redirect URLs for MAC authentication users are associated with a track entry.

Configuring Web proxy ports for URL redirection in MAC authentication

About this task

By default, the device redirects the HTTP or HTTPS requests from a MAC authenticated user to the authorized URL only if the requests are sent from a browser with Web proxy disabled. The device discards these requests if they are sent from a browser configured with a Web proxy.

To prevent the HTTP or HTTPS requests from being discarded, add the TCP port numbers of the Web proxy for HTTP and HTTPS to the device. These TCP port numbers are called Web proxy ports in this document.

Restrictions and guidelines

You can configure a maximum of 64 Web proxy ports for MAC authenticated users.

Specify different Web proxy port numbers for HTTP and HTTPS requests.

Adding or removing a Web proxy port will cause the device to log off all online MAC authenticated users that have been assigned an authorization redirect URL.

Procedure

1. Enter system view.
   system-view
2. Specify a Web proxy port used by MAC authentication users for HTTP or HTTPS.
   mac-authentication web-proxy { **http** | **https** } **port** *port-number*
   By default, no Web proxy ports are specified.

Logging off MAC authentication users

About this task

Perform this task to log off specified MAC authentication users and clear information about these users from the device. These users must perform MAC authentication to come online again.

Procedure

To log off MAC authentication users, execute the following command in user view:

reset mac-authentication access-user [**interface** *interface-type interface-number* | **mac** *mac-address* | **microsegment** *microsegment-id* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **username** *username* | **vlan** *vlan-id* | **vsi** *vsi-name*]

Enabling MAC authentication user logging

About this task

This feature enables the device to generate logs about MAC authentication users and send the logs to the information center. For the logs to be output correctly, you must also configure the information center on the device. For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

To prevent excessive MAC authentication user log entries, use this feature only if you need to analyze abnormal MAC authentication user logins or logouts.

Procedure

1. Enter system view.

   system-view

2. Enable MAC authentication user logging.

   mac-authentication access-user log enable [**failed-login** | **ip-change** | **logoff** | **successful-login**] *

   By default, MAC authentication user logging is disabled.

   If you do not specify any parameters, this command enables all types of MAC authentication user logs.

Display and maintenance commands for MAC authentication

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display MAC authentication information.	**display mac-authentication** [**interface** *interface-type interface-number*]
Display configuration information for MAC authentication access profiles.	**display mac-authentication access-profile** [**name** *profile-name*]
Display MAC authentication connections.	**display mac-authentication connection** [**open**] [[**m-lag** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **slot** *slot-number* | **user-name** *user-name*] | **user-mac** *mac-address*]
Display the MAC addresses of MAC authentication users in a type of MAC authentication microsegment, VLAN, or VSI.	**display mac-authentication mac-address** { **critical-microsegment** | **critical-vlan** | **critical-vsi** | **guest-vlan** | **guest-vsi** } [**interface** *interface-type interface-number*]
Clear MAC authentication statistics.	**reset mac-authentication statistics** [**interface** *interface-type*

Task	Command
	interface-number]
Remove users from the MAC authentication critical VLAN on a port.	**reset mac-authentication critical vlan interface** *interface-type interface-number* [**mac-address** *mac-address*]
Remove users from the MAC authentication critical microsegment on a port.	**reset mac-authentication critical microsegment interface** *interface-type interface-number* [**mac-address** *mac-address*]
Remove users from the MAC authentication guest VLAN on a port.	**reset mac-authentication guest-vlan interface** *interface-type interface-number* [**mac-address** *mac-address*]
Remove users from the MAC authentication critical VSI on a port.	**reset mac-authentication critical vsi interface** *interface-type interface-number* [**mac-address** *mac-address*]
Remove users from the MAC authentication guest VSI on a port.	**reset mac-authentication guest-vsi interface** *interface-type interface-number* [**mac-address** *mac-address*]

MAC authentication configuration examples

Example: Configuring local MAC authentication

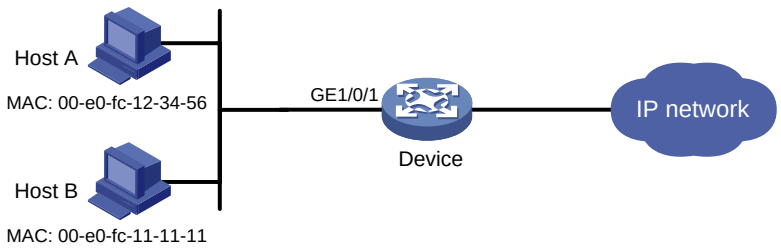
Network configuration

As shown in Figure 5, the device performs local MAC authentication on GigabitEthernet 1/0/1 to control Internet access of users.

Configure the device to meet the following requirements:

- Detect whether a user has gone offline every 180 seconds.
- Deny a user for 180 seconds if the user fails MAC authentication.
- Authenticate all users in ISP domain **bbb**.
- Use the MAC address of each user as both the username and password for authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in lower case.

Figure 5 Network diagram



Procedure

Add a network access local user. In this example, configure both the username and password as Host A's MAC address 08-00-27-12-34-56.

```
<Device> system-view
[Device] local-user 08-00-27-12-34-56 class network
[Device-luser-network-08-00-27-12-34-56] password simple 08-00-27-12-34-56
```

Specify the LAN access service for the user.

```
[Device-luser-network-08-00-27-12-34-56] service-type lan-access
[Device-luser-network-08-00-27-12-34-56] quit
```

Configure ISP domain **bbb** to perform local authentication for LAN users.

```
[Device] domain bbb
[Device-isp-bbb] authentication lan-access local
[Device-isp-bbb] quit
```

Enable MAC authentication on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] mac-authentication
[Device-GigabitEthernet1/0/1] quit
```

Specify ISP domain **bbb** as the MAC authentication domain.

```
[Device] mac-authentication domain bbb
```

Configure MAC authentication timers.

```
[Device] mac-authentication timer offline-detect 180
[Device] mac-authentication timer quiet 180
```

Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in lower case.

```
[Device] mac-authentication user-name-format mac-address with-hyphen lowercase
```

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Verifying the configuration

Display MAC authentication settings and statistics to verify your configuration.

```
[Device] display mac-authentication
 Global MAC authentication parameters:
   MAC authentication                          : Enabled
   Authentication method                       : PAP
   M-LAG member configuration conflict         : Unknown
   User name format                         : MAC address in lowercase(xx-xx-xx-xx-xx-xx)
           Username                            : mac
           Password                            : Not configured
   Offline detect period                       : 180 s
   Quiet period                                : 180 s
   Server timeout                              : 100 s
   Reauth period                               : 3600 s
   User aging period for critical VLAN         : 1000 s
   User aging period for critical VSI          : 1000 s
   User aging period for guest VLAN            : 1000 s
   User aging period for guest VSI             : 1000 s
   User aging period for critical microsegment: 1000 s
```

```
  Authentication domain                    : bbb
  HTTP proxy port list                     : Not configured
  HTTPS proxy port list                    : Not configured
 Online MAC-auth wired users               : 1

 Silent MAC users:
         MAC address       VLAN ID  From port               Port index
         0800-2711-1111    8        GE1/0/1                1

 GigabitEthernet1/0/1 is link-up
  MAC authentication                       : Enabled
  Carry User-IP                            : Disabled
  Authentication domain                    : Not configured
  Auth-delay timer                         : Disabled
  Periodic reauth                          : Disabled
  Re-auth server-unreachable               : Logoff
  Guest VLAN                               : Not configured
  Guest VLAN reauthentication              : Enabled
    Guest VLAN auth-period                 : 30 s
  Critical VLAN                            : Not configured
  Critical voice VLAN                      : Disabled
  Host mode                                : Single VLAN
  Offline detection                        : Enabled
  Authentication order                     : Default
  Unauthenticated-user aging               : Enabled
  Server-recovery online-user-sync         : Enabled

  Guest VSI                                : Not configured
  Guest VSI reauthentication               : Enabled
    Guest VSI auth-period                  : 30 s
  Critical VSI                             : Not configured
  Critical microsegment ID                 : Not configured
  URL user logoff                          : No
  Auto-tag feature                         : Disabled
  VLAN tag configuration ignoring          : Disabled
  Max online users                         : 4294967295
  Max online preauth-domain users          : 4294967295
  Max online Auth-Fail-domain users        : 4294967295
  Auth-server-unavailable escape           : Enabled
  Authentication attempts                  : successful 1, failed 0
  Current online users                     : 1
         MAC address       Auth state
         0800-2712-3456    Authenticated
```

The output shows that Host A has passed MAC authentication and has come online. Host B failed MAC authentication and its MAC address is marked as a silent MAC address.

Example: Configuring RADIUS-based MAC authentication

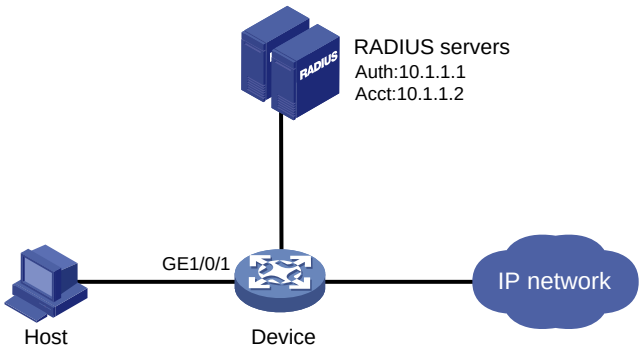
Network configuration

As shown in Figure 6, the device uses RADIUS servers to perform authentication, authorization, and accounting for users. The RADIUS servers use the CHAP authentication method.

To control user access to the Internet by MAC authentication, perform the following tasks:

- Enable MAC authentication globally and on GigabitEthernet 1/0/1.
- Configure the device to use CHAP for MAC authentication.
- Configure the device to detect whether a user has gone offline every 180 seconds.
- Configure the device to deny a user for 180 seconds if the user fails MAC authentication.
- Configure all users to belong to ISP domain **bbb**.
- Use a shared user account for all users, with username **aaa** and password **123456**.

Figure 6 Network diagram



Procedure

Make sure the RADIUS servers and the access device can reach each other.

1. Configure the RADIUS servers to provide authentication, authorization, and accounting services. Create a shared account with username **aaa** and password **123456** for MAC authentication users. (Details not shown.)
2. Configure RADIUS-based MAC authentication on the device:

   # Configure a RADIUS scheme.

   ```
   <Device> system-view
   [Device] radius scheme 2000
   [Device-radius-2000] primary authentication 10.1.1.1 1812
   [Device-radius-2000] primary accounting 10.1.1.2 1813
   [Device-radius-2000] key authentication simple abc
   [Device-radius-2000] key accounting simple abc
   [Device-radius-2000] user-name-format without-domain
   [Device-radius-2000] quit
   ```

   # Specify CHAP as the authentication method for MAC authentication.

   ```
   [Device] mac-authentication authentication-method chap
   ```

   # Apply the RADIUS scheme to ISP domain **bbb** for authentication, authorization, and accounting.

   ```
   [Device] domain bbb
   [Device-isp-bbb] authentication default radius-scheme 2000
   [Device-isp-bbb] authorization default radius-scheme 2000
   ```

```
[Device-isp-bbb] accounting default radius-scheme 2000
[Device-isp-bbb] quit
```

Enable MAC authentication on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] mac-authentication
[Device-GigabitEthernet1/0/1] quit
```

Specify the MAC authentication domain as ISP domain **bbb**.

```
[Device] mac-authentication domain bbb
```

Set MAC authentication timers.

```
[Device] mac-authentication timer offline-detect 180
[Device] mac-authentication timer quiet 180
```

Specify username **aaa** and password **123456** in plain text for the account shared by MAC authentication users.

```
[Device] mac-authentication user-name-format fixed account aaa password simple 123456
```

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Verifying the configuration

Verify the MAC authentication configuration.

```
[Device] display mac-authentication
 Global MAC authentication parameters:
   MAC authentication                         : Enabled
   Authentication method                      : CHAP
   M-LAG member configuration conflict        : Unknown
   Username format                            : Fixed account
           Username                           : aaa
           Password                           : ******
   Offline detect period                      : 180 s
   Quiet period                               : 180 s
   Server timeout                             : 100 s
   Reauth period                              : 3600 s
   User aging period for critical VLAN        : 1000 s
   User aging period for critical VSI         : 1000 s
   User aging period for guest VLAN           : 1000 s
   User aging period for guest VSI            : 1000 s
   User aging period for critical microsegment: 1000 s
   Authentication domain                      : bbb
   HTTP proxy port list                       : Not configured
   HTTPS proxy port list                      : Not configured
 Online MAC-auth wired users                  : 1

 Silent MAC users:
          MAC address       VLAN ID  From port               Port index

 GigabitEthernet1/0/1  is link-up
   MAC authentication                         : Enabled
   Carry User-IP                              : Disabled
   Authentication domain                      : Not configured
   Auth-delay timer                           : Disabled
```

```
Periodic reauth                          : Disabled
Re-auth server-unreachable               : Logoff
Guest VLAN                               : Not configured
Guest VLAN reauthentication              : Enabled
  Guest VLAN auth-period                 : 30 s
Critical VLAN                            : Not configured
Critical voice VLAN                      : Disabled
Host mode                                : Single VLAN
Offline detection                        : Enabled
Authentication order                     : Default
Unauthenticated-user aging               : Enabled
Server-recovery online-user-sync         : Enabled

Guest VSI                                : Not configured
Guest VSI reauthentication               : Enabled
  Guest VSI auth-period                  : 30 s
Critical VSI                             : Not configured
Critical microsegment ID                 : Not configured
URL user logoff                          : No
Auto-tag feature                         : Disabled
VLAN tag configuration ignoring          : Disabled
Max online users                         : 4294967295
Max online preauth-domain users          : 4294967295
Max online Auth-Fail-domain users        : 4294967295
Auth-server-unavailable escape           : Enabled
Authentication attempts                  : successful 1, failed 0
Current online users                     : 1
       MAC address       Auth state
       0800-2712-3456    Authenticated
```

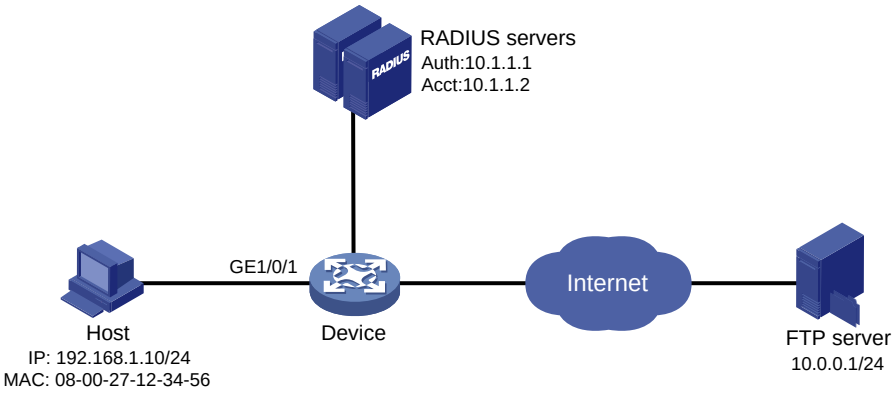
Example: Configuring ACL assignment for MAC authentication

Network configuration

As shown in Figure 7, configure the device to meet the following requirements:

- Use RADIUS servers to perform authentication, authorization, and accounting for users.
- Perform MAC authentication on GigabitEthernet 1/0/1 to control Internet access.
- Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in lower case.
- Use an ACL to deny authenticated users to access the FTP server at 10.0.0.1.

Figure 7 Network diagram



Procedure

Make sure the RADIUS servers and the access device can reach each other.

1. Configure the RADIUS servers:

   # Configure the RADIUS servers to provide authentication, authorization, and accounting services. (Details not shown.)

   # Add a user account with **08-00-27-12-34-56** as both the username and password on each RADIUS server. (Details not shown.)

   # Specify ACL 3000 as the authorization ACL for the user account. (Details not shown.)

2. Configure ACL 3000 to deny packets destined for 10.0.0.1 on the device.

```
<Device> system-view
[Device] acl advanced 3000
[Device-acl-ipv4-adv-3000] rule 0 deny ip destination 10.0.0.1 0
[Device-acl-ipv4-adv-3000] quit
```

3. Configure RADIUS-based MAC authentication on the device:

   # Configure a RADIUS scheme.

```
[Device] radius scheme 2000
[Device-radius-2000] primary authentication 10.1.1.1 1812
[Device-radius-2000] primary accounting 10.1.1.2 1813
[Device-radius-2000] key authentication simple abc
[Device-radius-2000] key accounting simple abc
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

   # Apply the RADIUS scheme to an ISP domain for authentication, authorization, and accounting.

```
[Device] domain bbb
[Device-isp-bbb] authentication default radius-scheme 2000
[Device-isp-bbb] authorization default radius-scheme 2000
[Device-isp-bbb] accounting default radius-scheme 2000
[Device-isp-bbb] quit
```

   # Specify the ISP domain for MAC authentication.

```
[Device] mac-authentication domain bbb
```

   # Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in lower case.

```
[Device] mac-authentication user-name-format mac-address with-hyphen lowercase
```

Enable MAC authentication on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] mac-authentication
[Device-GigabitEthernet1/0/1] quit
```

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Verifying the configuration

Verify the MAC authentication configuration.

```
[Device] display mac-authentication
 Global MAC authentication parameters:
   MAC authentication                          : Enable
   Authentication method                       : PAP
   M-LAG member configuration conflict         : Unknown
   Username format                           : MAC address in lowercase(xx-xx-xx-xx-xx-xx)
           Username                            : mac
           Password                            : Not configured
   Offline detect period                       : 300 s
   Quiet period                                : 60 s
   Server timeout                              : 100 s
   Reauth period                               : 3600 s
   User aging period for critical VLAN         : 1000 s
   User aging period for critical VSI          : 1000 s
   User aging period for guest VLAN            : 1000 s
   User aging period for guest VSI             : 1000 s
   User aging period for critical microsegment: 1000 s
   Authentication domain                       : bbb
   HTTP proxy port list                        : Not configured
   HTTPS proxy port list                       : Not configured
 Online MAC-auth wired users                   : 1

 Silent MAC users:
          MAC address       VLAN ID  From port               Port index

 GigabitEthernet1/0/1  is link-up
   MAC authentication                          : Enabled
   Carry User-IP                               : Disabled
   Authentication domain                       : Not configured
   Auth-delay timer                            : Disabled
   Periodic reauth                             : Disabled
   Re-auth server-unreachable                  : Logoff
   Guest VLAN                                  : Not configured
   Guest VLAN reauthentication                 : Enabled
     Guest VLAN auth-period                    : 30 s
   Critical VLAN                               : Not configured
   Critical voice VLAN                         : Disabled
   Host mode                                   : Single VLAN
   Offline detection                           : Enabled
   Authentication order                        : Default
```

```
   Unauthenticated-user aging                 : Enabled
   Server-recovery online-user-sync           : Enabled

   Guest VSI                                  : Not configured
   Guest VSI reauthentication                 : Enabled
     Guest VSI auth-period                    : 30 s
   Critical VSI                               : Not configured
   Critical microsegment ID                   : Not configured
   URL user logoff                            : No
   Auto-tag feature                           : Disabled
   VLAN tag configuration ignoring            : Disabled
   Max online users                           : 4294967295
   Max online preauth-domain users            : 4294967295
   Max online Auth-Fail-domain users          : 4294967295
   Auth-server-unavailable escape             : Enabled
   Authentication attempts                    : successful 1, failed 0
   Current online users                       : 1
          MAC address       Auth state
          08-00-2712-3456    Authenticated
```

Verify that you cannot ping the FTP server from the host.

```
C:\>ping 10.0.0.1

Pinging 10.0.0.1 with 32 bytes of data:

Request timed out.
Request timed out.
Request timed out.
Request timed out.

Ping statistics for 10.0.0.1:
   Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),
```

The output shows that ACL 3000 has been assigned to GigabitEthernet 1/0/1 to deny access to the FTP server.

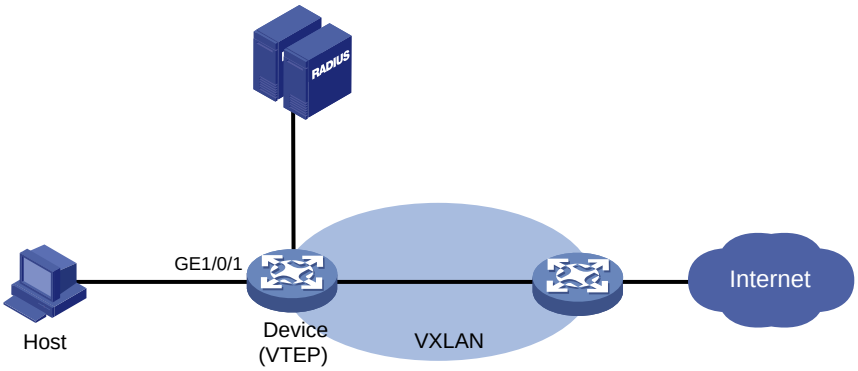
Example: Configuring MAC authentication authorization VSI assignment

Network configuration

As shown in Figure 8, configure the device to meet the following requirements:

- Use RADIUS servers to perform authentication, authorization, and accounting for users.
- Perform MAC authentication on GigabitEthernet 1/0/1 to control Internet access.
- Configure the RADIUS server to assign VSI **bbb** to the host when the host passes MAC authentication.
- Authenticate all users in ISP domain **2000**.
- Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in lower case.

Figure 8 Network diagram



Procedure

Make sure the RADIUS servers and the access device can reach each other.

1. Configure the RADIUS servers:

   # Configure the RADIUS servers to provide authentication, authorization, and accounting services. (Details not shown.)

   # Add a user account with **d4-85-64-be-c6-3e** as both the username and password on each RADIUS server. (Details not shown.)

   # Specify VSI **bbb** as the authorization VSI for the user account. (Details not shown.)

   NOTE:

   If an ADCAM server is used for authentication and authorization, configure VSIs on the server. The server will assign these VSIs to the device. You do not need to configure VSIs on the device.

2. Configure RADIUS-based MAC authentication on the device:

   # Configure a RADIUS scheme.

   ```
   <Device> system-view
   [Device] radius scheme bbb
   [Device-radius-bbb] primary authentication 10.1.1.1
   [Device-radius-bbb] primary accounting 10.1.1.2
   [Device-radius-bbb] key authentication simple bbb
   [Device-radius-bbb] key accounting simple bbb
   [Device-radius-bbb] user-name-format without-domain
   [Device-radius-bbb] quit
   ```

   # Apply the RADIUS scheme to ISP domain **2000** for authentication, authorization, and accounting.

   ```
   [Device] domain 2000
   [Device-isp-2000] authentication lan-access radius-scheme bbb
   [Device-isp-2000] authorization lan-access radius-scheme bbb
   [Device-isp-2000] accounting lan-access radius-scheme bbb
   [Device-isp-2000] quit
   ```

   # Enable MAC authentication on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] mac-authentication
   ```

   # Enable the MAC match mode for dynamic Ethernet service instances on GigabitEthernet 1/0/1.

```
[Device-GigabitEthernet1/0/1] mac-based ac
[Device-GigabitEthernet1/0/1] quit
```

Enable L2VPN.

```
[Device] l2vpn enable
```

Create a VSI named **bbb** and the associated VXLAN.

```
[Device] vsi bbb
[Device-vsi-bbb] vxlan 5
[Device-vsi-bbb-vxlan-5] quit
```

Specify the ISP domain for MAC authentication.

```
[Device] mac-authentication domain 2000
```

Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in lower case.

```
[Device] mac-authentication user-name-format mac-address with-hyphen lowercase
```

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Verifying the configuration

Verify that VSI **bbb** is assigned to the MAC authentication user after the user passes authentication.

```
[Device] display mac-authentication connection
Total connections: 1
Slot ID: 1
User MAC address: d485-64be-c63e
Access interface: GigabitEthernet1/0/1
Username: d4-85-64-be-c6-3e
User access state: Successful
Authentication domain: 2000
IPv4 address: 192.168.1.1
IPv6 address: 2000:0:0:0:1:2345:6789:abcd
IPv4 address source: User packet
IPv6 address source: User packet
Initial VLAN: 1
Authorization untagged VLAN: N/A
Authorization tagged VLAN: N/A
Authorization VSI: bbb
Authorization microsegment ID: N/A
Authorization ACL number/name: 3001
Authorization dynamic ACL name: N/A
Authorization user profile: N/A
Authorization CAR:  N/A
Authorization URL: N/A
Authorization IPv6 URL: N/A
Authorization temporary redirect: Disabled
Authorization redirect match terminal rule: N/A
Start accounting: Successful
Real-time accounting-update failures: 0
Termination action: N/A
Session timeout period: N/A
```

```
Offline detection: 100 sec (server-assigned)
Online from: 2016/06/13 09:06:37
Online duration: 0h 0m 35s
Port-down keep online: Disabled (offline)
```

Verify that a dynamic AC is created for MAC address **d485-64be-c63e**.

```
[Device] display l2vpn forwarding ac verbose
VSI Name: bbb
  Interface: GE1/0/1  Service Instance: 1
    Link ID      : 0
    Access Mode  : VLAN
    Encapsulation: untagged
    Type         : Dynamic (MAC-based)
    MAC address  : d485-64be-c63e
```

Contents

Configuring portal authentication 1

About portal authentication 1
Advantages of portal authentication 1
Extended portal functions 1
Portal system 1
Portal authentication using a remote portal server 2
Local portal service 3
Portal authentication modes 3
Portal authentication process 4
Portal support for EAP 6
Portal filtering rules 7
MAC-based quick portal authentication 8
IP-based quick portal authentication 8
Portal support for M-LAG 9
Portal support for microsegment authorization 10
Restrictions and guidelines: Portal configuration 10
Portal authentication tasks at a glance 11
Prerequisites for portal authentication 12
Configuring a remote portal authentication server 12
Configuring a portal Web server 13
Portal Web server tasks at a glance 13
Configure basic parameters for a portal Web server 13
Enabling the captive-bypass feature 14
Configuring a match rule for URL redirection 15
Configuring unescaped special characters for portal Web server URLs redirected to users 15
Configuring local portal service features 16
About the local portal service 16
Restrictions and guidelines for configuring local portal service features 16
Customizing authentication pages 16
Configuring a local portal Web service 18
Enabling portal authentication on an interface 19
Specifying a portal Web server on an interface 20
Enabling portal to support IPv4/IPv6 dual stack 20
Configuring a portal preauthentication domain 21
Specifying a preauthentication IP address pool 22
Specifying a portal authentication domain 23
About portal authentication domains 23
Restrictions and guidelines for specifying a portal authentication domain 23
Specifying a portal authentication domain on an interface 23
Controlling portal user access 24
Configuring a portal-free rule 24
Configuring an authentication source subnet 25
Configuring an authentication destination subnet 26
Configuring support of Web proxy for portal authentication 26
Checking the issuing of category-2 portal filtering rules 27
Setting the maximum number of portal users 27
Enabling strict-checking on portal authorization information 28
Allowing only users with DHCP-assigned IP addresses to pass portal authentication 28
Enabling portal roaming 29
Configuring the portal fail-permit feature 30
Specifying a critical profile on an interface and setting the aging time for portal users in critical microsegments 30
Configuring portal detection features 31
Configuring online detection of portal users 31
Configuring portal authentication server detection 32
Configuring portal Web server detection 33
Configuring portal user synchronization 33

Configuring portal packet attributes ····34
Configuring the BAS-IP or BAS-IPv6 attribute ····34
Specifying the device ID····35
Configuring attributes for RADIUS packets····35
Specifying a format for the NAS-Port-Id attribute····35
Configuring the NAS-Port-Type attribute ····36
Applying a NAS-ID profile to an interface····36
Configuring MAC-based quick portal authentication····37
Restrictions and guidelines for configuring MAC-based quick portal authentication····37
Configuring a MAC binding server ····37
Specifying a MAC binding server on an interface ····38
Configuring cloud MAC-trigger authentication ····39
Configuring IP-based quick portal authentication ····39
Configuring portal authentication to support M-LAG ····40
Prerequisites ····40
Setting the authentication load sharing mode for portal users on M-LAG interfaces ····40
Configure traffic backup for portal users on M-LAG interfaces ····41
Logging out online portal users····41
Enabling portal user login/logout logging ····42
Disabling the Rule ARP or ND entry feature for portal clients ····42
Configuring Web redirect ····42
Setting the user synchronization interval for portal authentication using OAuth ····43
Obtaining user access information from ARP or ND entries ····43
Display and maintenance commands for portal ····44
Portal configuration examples····45
Example: Configuring direct portal authentication····45
Example: Configuring re-DHCP portal authentication····48
Example: Configuring cross-subnet portal authentication····52
Example: Configuring extended direct portal authentication····55
Example: Configuring extended re-DHCP portal authentication····58
Example: Configuring extended cross-subnet portal authentication····63
Example: Configuring portal server detection and portal user synchronization ····66
Example: Configuring cross-subnet portal authentication for MPLS L3VPNs····69
Example: Configuring direct portal authentication with a preauthentication domain····71
Example: Configuring re-DHCP portal authentication with a preauthentication domain····73
Example: Configuring direct portal authentication using a local portal Web service····75
Example: Configuring MAC-based quick portal authentication····78
Example: Configuring cloud MAC-trigger authentication ····81
Example: Configuring portal authentication to support M-LAG ····83
Troubleshooting portal ····91
No portal authentication page is pushed for users····91
Cannot log out portal users on the access device ····92
Cannot log out portal users on the RADIUS server ····92
Users logged out by the access device still exist on the portal authentication server····92
Re-DHCP portal authenticated users cannot log in successfully ····93

Configuring portal authentication

About portal authentication

Portal authentication controls user access to networks. Portal authenticates a user by the username and password the user enters on a portal authentication page. Typically, portal authentication is deployed on the access layer and vital data entries.

In a portal-enabled network, users can actively initiate portal authentication by visiting the authentication website provided by the portal Web server. Or, they are redirected to the portal authentication page for authentication when they visit other websites.

The device supports Portal 1.0, Portal 2.0, and Portal 3.0.

Advantages of portal authentication

Portal authentication has the following advantages:

- Allows users to perform authentication through a Web browser without installing client software.
- Provides ISPs with diversified management choices and extended functions. For example, the ISPs can place advertisements, provide community services, and publish information on the authentication page.
- Supports multiple authentication modes. For example, re-DHCP authentication implements a flexible address assignment scheme and saves public IP addresses. Cross-subnet authentication can authenticate users who reside in a different subnet than the access device.

Extended portal functions

By forcing patching and anti-virus policies, extended portal functions help hosts to defend against viruses. Portal supports the following extended functions:

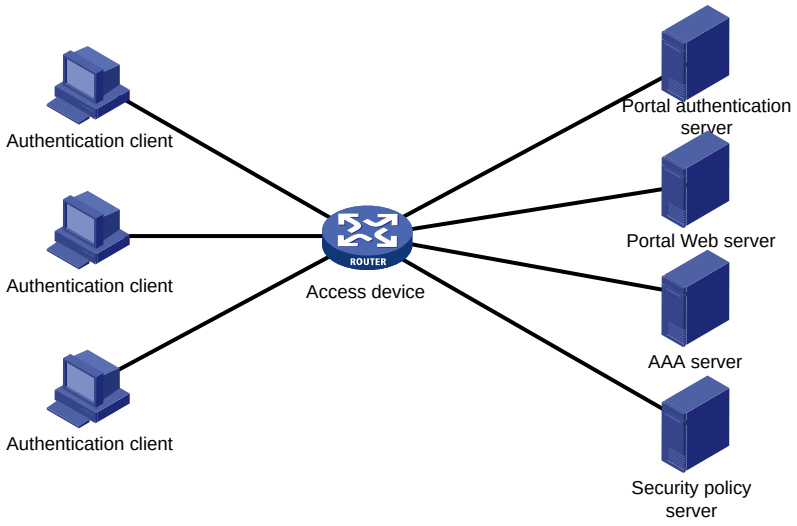
- **Security check**—Detects after authentication whether or not a user host installs anti-virus software, virus definition file, unauthorized software, and operating system patches.
- **Resource access restriction**—Allows an authenticated user to access certain network resources such as the virus server and the patch server. Users can access more network resources after passing security check.

Security check must cooperate with the IMC security policy server and the iNode client.

Portal system

A typical portal system consists of these basic components: authentication client, access device, portal authentication server, portal Web server, AAA server, and security policy server.

Figure 1 Portal system



Authentication client

An authentication client is a Web browser that runs HTTP/HTTPS or a user host that runs a portal client. Security check for the user host is implemented through the interaction between the portal client and the security policy server. Only the iNode client is supported.

Access device

An access device provides access services. It has the following functions:

- Redirects all HTTP or HTTPS requests of unauthenticated users to the portal Web server.
- Interacts with the portal authentication server and the AAA server to complete authentication, authorization, and accounting.
- Allows users that pass portal authentication to access authorized network resources.

Portal server

A portal server collectively refers to a portal authentication server and portal Web server.

The portal Web server pushes the Web authentication page to authentication clients and forwards user authentication information (username and password) to the portal authentication server. The portal authentication server receives authentication requests from authentication clients and interacts with the access device to authenticate users. The portal Web server is typically integrated with the portal authentication server and it can also be an independent server.

AAA server

The AAA server interacts with the access device to implement authentication, authorization, accounting for portal users. In a portal system, a RADIUS server can perform authentication, authorization, accounting for portal users, and an LDAP server can perform authentication for portal users.

Security policy server

The security policy server interacts with the portal client and the access device for security check and authorization for users. Only hosts that run portal clients can interact with the security policy server.

Portal authentication using a remote portal server

The components of a portal system interact as follows:

1. An unauthenticated user initiates authentication by accessing an Internet website through a Web browser. When receiving the HTTP or HTTPS request, the access device redirects it to the

Web authentication page provided by the portal Web server. The user can also visit the authentication website to log in. The user must log in through the iNode client for extended portal functions.

2. The user enters the authentication information on the authentication page/dialog box and submits the information. The portal Web server forwards the information to the portal authentication server. The portal authentication server processes the information and forwards it to the access device.
3. The access device interacts with the AAA server to implement authentication, authorization, accounting for the user.
4. If security policies are not imposed on the user, the access device allows the authenticated user to access networks.

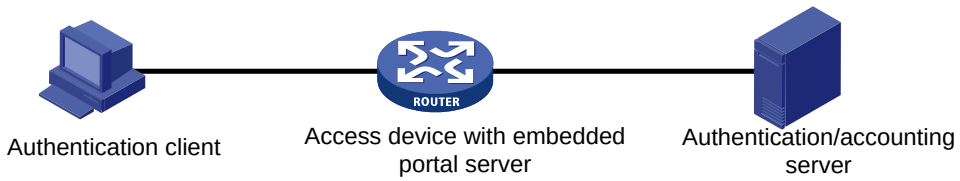
   If security policies are imposed on the user, the portal client, the access device, and the security policy server interact to check the user host. If the user passes the security check, the security policy server authorizes the user to access resources based on the check result.

Local portal service

System components

As shown in Figure 2, a local portal system consists of an authentication client, access device, and AAA server. The access device acts as both the portal Web server and the portal authentication server to provide the local portal Web service for the authentication client. The authentication client can only be a Web browser, and it cannot be a user host that runs a portal client. Therefore, extended portal functions are not supported and no security policy server is required.

Figure 2 System components



Portal page customization

The portal system pushes different authentication pages to portal users at different stages. The local portal service supports portal page customization. You can customize one or more sets of authentication pages, compress each set of authentication pages to a .zip file, and upload the compressed files to the storage medium of the device.

For more information about authentication page customization, see "Customizing authentication pages."

Portal authentication modes

Portal authentication has three modes: direct authentication, re-DHCP authentication, and cross-subnet authentication. In direct authentication and re-DHCP authentication, no Layer 3 forwarding devices exist between the authentication client and the access device. In cross-subnet authentication, Layer 3 forwarding devices can exist between the authentication client and the access device.

Direct authentication

A user manually configures a public IP address or obtains a public IP address through DHCP. Before authentication, the user can access only the portal Web server and predefined authentication-free websites. After passing authentication, the user can access other network resources. The process of direct authentication is simpler than that of re-DHCP authentication.

Re-DHCP authentication

Before a user passes authentication, DHCP allocates an IP address (a private IP address) to the user. The user can access only the portal Web server and predefined authentication-free websites. After the user passes authentication, DHCP reallocates an IP address (a public IP address) to the user. The user then can access other network resources. No public IP address is allocated to users who fail authentication. Re-DHCP authentication saves public IP addresses. For example, an ISP can allocate public IP addresses to broadband users only when they access networks beyond the residential community network.

Only the iNode client supports re-DHCP authentication. IPv6 portal authentication does not support the re-DHCP authentication mode.

Cross-subnet authentication

Cross-subnet authentication is similar to direct authentication, except it allows Layer 3 forwarding devices to exist between the authentication client and the access device.

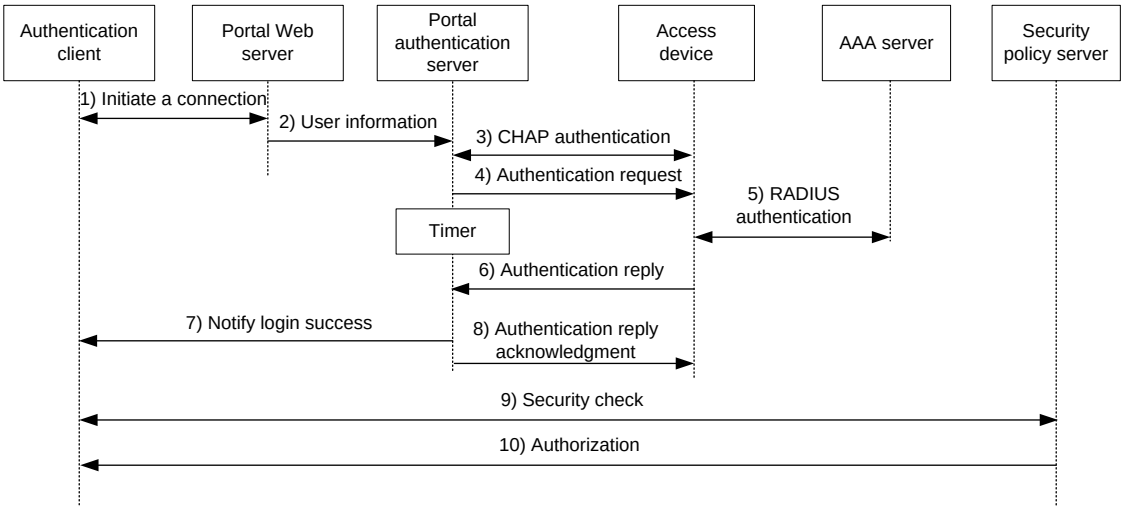
In direct authentication, re-DHCP authentication, and cross-subnet authentication, a user's IP address uniquely identifies the user. After a user passes authentication, the access device generates an ACL for the user based on the user's IP address to control forwarding of the packets from the user. Because no Layer 3 forwarding device exists between authentication clients and the access device in direct authentication and re-DHCP authentication, the access device can learn the user MAC addresses. The access device can enhance its capability of controlling packet forwarding by using the learned MAC addresses.

Portal authentication process

Direct authentication and cross-subnet authentication share the same authentication process. Re-DHCP authentication has a different process as it has two address allocation procedures.

Direct authentication/cross-subnet authentication process (with CHAP/PAP authentication)

Figure 3 Direct authentication/cross-subnet authentication process



The direct/cross-subnet authentication process is as follows:

2. A portal user access the Internet through HTTP or HTTPS, and the HTTP or HTTPS packet arrives at the access device.

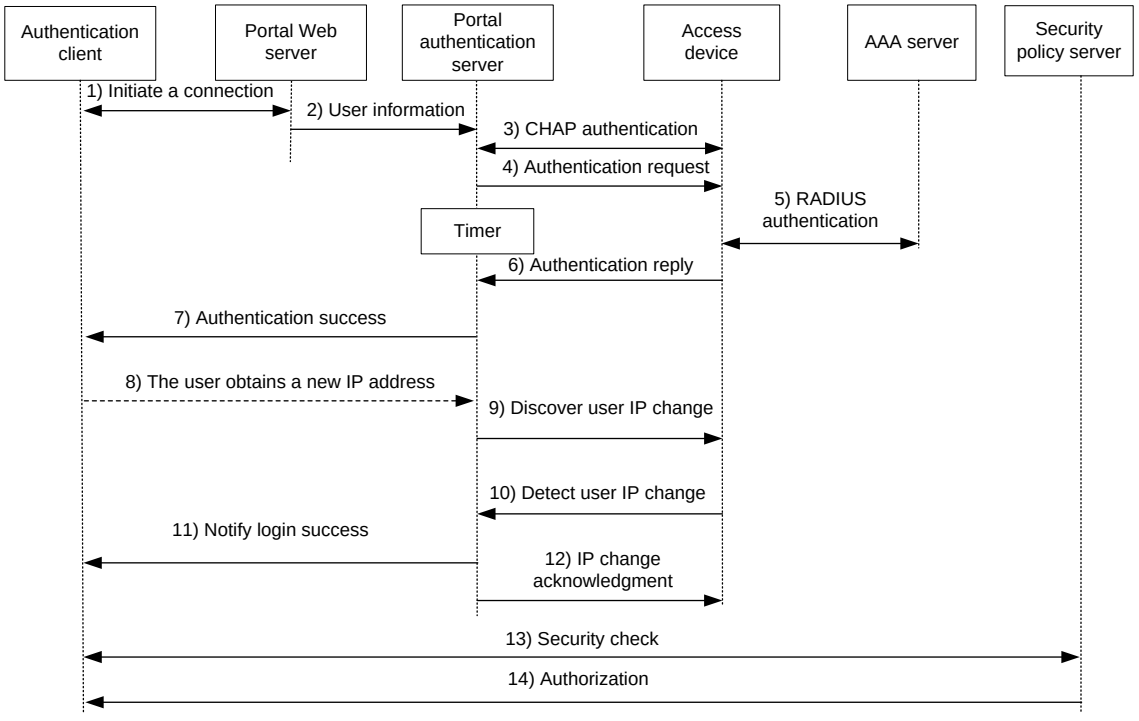
- If the packet matches a portal free rule, the access device allows the packet to pass.
- If the packet does not match any portal-free rule, the access device redirects the packet to the portal Web server. The portal Web server pushes the Web authentication page to the user for him to enter his username and password.

3. The portal Web server submits the user authentication information to the portal authentication server.
4. The portal authentication server and the access device exchange CHAP messages. This step is skipped for PAP authentication. The portal authentication server decides the method (CHAP or PAP) to use.
5. The portal authentication server adds the username and password into an authentication request packet and sends it to the access device. Meanwhile, the portal authentication server starts a timer to wait for an authentication reply packet.
6. The access device and the RADIUS server exchange RADIUS packets.
7. The access device sends an authentication reply packet to the portal authentication server to notify authentication success or failure.
8. The portal authentication server sends an authentication success or failure packet to the client.
9. If the authentication is successful, the portal authentication server sends an authentication reply acknowledgment packet to the access device.

   If the client is an iNode client, the authentication process includes step 9 and step 10 for extended portal functions. Otherwise the authentication process is complete.
10. The client and the security policy server exchange security check information. The security policy server detects whether or not the user host installs anti-virus software, virus definition files, unauthorized software, and operating system patches.
11. The security policy server authorizes the user to access certain network resources based on the check result. The access device saves the authorization information and uses it to control access of the user.

Re-DHCP authentication process (with CHAP/PAP authentication)

Figure 4 Re-DHCP authentication process



The re-DHCP authentication process is as follows:

Step 1 through step 7 are the same as those in the direct authentication/cross-subnet authentication process.

2. A portal user access the Internet through HTTP or HTTPS, and the HTTP or HTTPS packet arrives at the access device.
    - If the packet matches a portal free rule, the access device allows the packet to pass.
    - If the packet does not match any portal-free rule, the access device redirects the packet to the portal Web server. The portal Web server pushes the Web authentication page to the user for him to enter his username and password.
3. The portal Web server submits the user authentication information to the portal authentication server.
4. The portal authentication server and the access device exchange CHAP messages. This step is skipped for PAP authentication. The portal authentication server decides the method (CHAP or PAP) to use.
5. The portal authentication server adds the username and password into an authentication request packet and sends it to the access device. Meanwhile, the portal authentication server starts a timer to wait for an authentication reply packet.
6. The access device and the RADIUS server exchange RADIUS packets.
7. The access device sends an authentication reply packet to the portal authentication server to notify authentication success or failure.
8. The portal authentication server sends an authentication success or failure packet to the client.
9. After receiving the authentication success packet, the client obtains a public IP address through DHCP. The client then notifies the portal authentication server that it has a public IP address.
10. The portal authentication server notifies the access device that the client has obtained a public IP address.
11. The access device detects the IP change of the client through DHCP and then notifies the portal authentication server that it has detected an IP change of the client IP.
12. After receiving the IP change notification packets sent by the client and the access device, the portal authentication server notifies the client of login success.
13. The portal authentication server sends an IP change acknowledgment packet to the access device.
14. The client and the security policy server exchanges security check information. The security policy server detects whether or not the user host installs anti-virus software, virus definition files, unauthorized software, and operating system patches.
15. The security policy server authorizes the user to access certain network resources based on the check result. The access device saves the authorization information and uses it to control access of the user.

Step 13 and step 14 are for extended portal functions.

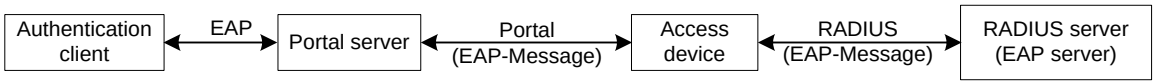
Portal support for EAP

To use portal authentication that supports EAP, the portal authentication server and client must be the IMC portal server and the iNode portal client. Local portal authentication does not support EAP authentication.

Compared with username and password based authentication, digital certificate-based authentication ensures higher security.

The Extensible Authentication Protocol (EAP) supports several digital certificate-based authentication methods, for example, EAP-TLS. Working together with EAP, portal authentication can implement digital certificate-based user authentication.

Figure 5 Portal support for EAP working flow diagram



As shown in Figure 5, the authentication client and the portal authentication server exchange EAP authentication packets. The portal authentication server and the access device exchange portal authentication packets that carry the EAP-Message attributes. The access device and the RADIUS server exchange RADIUS packets that carry the EAP-Message attributes. The RADIUS server that supports the EAP server function processes the EAP packets encapsulated in the EAP-Message attributes, and provides the EAP authentication result.

The access device does not process but only transports EAP-Message attributes between the portal authentication server and the RADIUS server. Therefore, the access device requires no additional configuration to support EAP authentication.

Portal filtering rules

The access device uses portal filtering rules to control user traffic forwarding.

Basic rules

Based on the configuration and authentication status of portal users, the device generates the following categories of portal filtering rules:

- **Category 1**—The rule permits user packets that are destined for the portal Web server and packets that match the portal-free rules to pass through.
- **Category 2**—For an authenticated user with neither ACL nor microsegment authorized, the rule allows the user to access any destination network resources. For an authenticated user with an ACL or microsegment authorized, the rule allows users to access resources permitted by the ACL or microsegment. The device adds the rule when a user comes online and deletes the rule when the user goes offline.

  The device supports the following types of authorization ACLs:

  - Basic ACLs (ACL 2000 to ACL 2999).
  - Advanced ACLs (ACL 3000 to ACL 3999).
  - Layer 2 ACLs (ACL 4000 to ACL 4999).
  - User-defined ACLs (ACL 5000 to ACL 5999).

  For an authorization ACL to take effect, make sure the ACL exists and has ACL rules excluding rules configured with the **`counting`**, **`established`**, **`fragment`**, **`source-mac`**, **`source-ip`**, or **`logging`** keyword. For more information about ACL rules, see ACL commands in *ACL and QoS Command Reference*.

  If the authorization microsegment ID does not exist or microsegmentation is disabled, the authorization microsegment does not take effect. The microsegment authorization feature is supported in only direct portal authentication mode. The restrictions for configurations of authorization ACLs and microsegments are as follows:

  - An authorization ACL and authorization microsegment can both take effect, but source criteria cannot be configured if the ACL is a basic ACL and microsegment cannot be configured if the ACL is an advanced ACL.
  - When ACL resource sharing in shared portal authentication mode is enabled, the microsegment authorization feature does not take effect.

  For more information about microsegmentation, see "Configuring microsegmentation."
- **Category 3**—The rule redirects all HTTP or HTTPS requests from unauthenticated users to the portal Web server.

- **Category 4**—For direct authentication and cross-subnet authentication, the rule forbids any user packets to pass through. For re-DHCP authentication, the device forbids user packets with private source addresses to pass.

After receiving a user packet, the device compares the packet against the filtering rules from category 1 to category 4. Once the packet matches a rule, the matching process completes.

MAC-based quick portal authentication

MAC-based quick portal authentication is applicable to scenarios where users access the network frequently. It allows users to pass authentication without entering a username and password. MAC-based quick portal authentication is also called MAC-trigger authentication or MAC-based transparent authentication.

A MAC binding server is required for MAC-trigger authentication. The MAC binding server records the MAC-to-account bindings of portal users for authentication. The account contains the portal authentication information of the user, including username and password.

Only direct authentication supports MAC-based quick portal authentication.

The authentication is implemented as follows:

1. When a user accesses the network for the first time, the access device generates a MAC-trigger entry that records the user's MAC address and access interface. The user can access the network without performing portal authentication if the user's network traffic is below the free-traffic threshold.
2. When the user's network traffic reaches the threshold, the access device sends a MAC binding query to the MAC binding server.
3. The MAC binding server checks whether the MAC address of the user is bound with a portal user account.
   - If a matching MAC-account binding exists, the MAC binding server sends the user authentication information to the access device to initiate portal authentication. The user is authenticated without entering the username and password.
     - If the user fails portal authentication, an authentication failure message is returned to the user. The MAC-trigger entry of the user on the access device is deleted when the entry ages out.
     - If the user passes portal authentication, the access device deletes the MAC-trigger entry of the user.
   - If no matching MAC-account binding exists, the MAC binding server notifies the access device to perform normal portal authentication for the user.
     - If the user fails portal authentication, an authentication failure message is returned to the user. The whole process is finished.
     - If the user passes portal authentication, the access device sends the user's MAC address and authentication information to the MAC binding server for MAC-account binding. Additionally, the access device deletes the MAC-trigger entry of the user.

NOTE:

For information about MAC binding server configuration, see the user manual of the server.

IP-based quick portal authentication

IP-based quick portal authentication is applicable to scenarios where users access the network frequently. It allows users to pass authentication without entering a username and password. IP-based quick portal authentication is also called IP-trigger authentication or IP-based transparent authentication.

After IP-based quick portal authentication is enabled, the device adds an IP-trigger entry for a user on receiving user traffic that matches the IP-trigger ACL. The IP-trigger entry records the user's IP address, access interface, VLAN ID, and the entry aging time. The device sends the user's authentication credentials (using the user IP address as both the username and password) to the RADIUS server for authentication. In this way, the device provides portal authentication transparently to the user.

Only direct authentication and cross-subnet authentication support IP-based quick portal authentication.

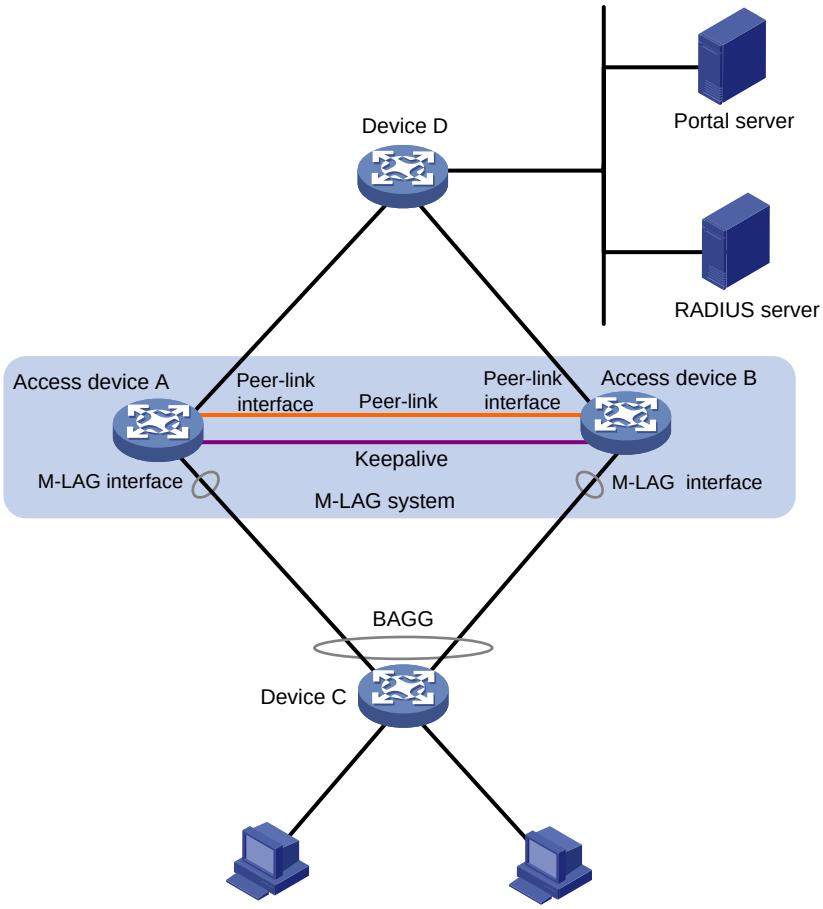
Portal support for M-LAG

Portal authentication supports M-LAG to improve high availability of the access device.

As shown in Figure 6, Device A and Device B form an M-LAG system. The M-LAG system acts as an access device to provide redundancy and load sharing. For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*.

In an M-LAG system, after portal users pass authentication on an M-LAG member device, the M-LAG member device will synchronize data of the users to the peer member device for backup. When the M-LAG member device fails, the peer member device can use the backup user data to process services for the portal users. This ensures user service continuity.

Figure 6 Support of portal for M-LAG



Portal support for microsegment authorization

Authorization microsegment

The microsegmentation feature can perform access control on network access users based on the microsegment groups to which the users belong. When a portal user comes online after passing direct portal authentication (not the shared mode) and an authorization microsegment is assigned by the RADIUS server to the user, the device filters the user traffic on the access interface according to the authorization microsegment. For more information about microsegmentation, see "Configuring microsegmentation."

Portal critical microsegment

The portal critical microsegment feature allows users to access critical microsegments (specific microsegment resources) when all authentication servers are unreachable. This feature takes effect only in RADIUS authentication. If other authentication methods are adopted, users will not be accommodated to the critical microsegments.

In direct portal authentication mode, after a critical profile is specified on the access interface of a user, the user is accommodated to the critical microsegment when no authentication server is reachable during authentication. The user can access the network resources in the authorized critical microsegment. When the user performs authentication again and no authentication server is reachable, the user will stay in the critical microsegment. When the authentication server is reachable and the user performs authentication again, the user leaves the critical microsegment no matter whether authentication succeeded or failed.

Restrictions and guidelines: Portal configuration

In portal authentication, the device can redirect users' HTTPS requests to the portal Web server. You can use one of the following local certificates for the HTTPS redirect service according to the security requirements and the configuration complexity:

- **Self-signed certificate**—Using this type of certificate is simple in configuration but has low security. You do not need to associate an SSL server policy with the HTTPS redirect service and the default SSL parameters are used. However, a self-signed certificate is not trusted by the browser. When the device redirects HTTPS requests to the portal Web server, a certificate security warning prompt might appear on the browser. If you accept the security risks stated in the prompt, you can ignore the prompt to browse the page.
- **CA-signed certificate**—Using this type of certificate is complex in configuration but has high security. You must obtain a CA certificate, request a local certificate from the CA, create an SSL server policy (with policy name of **https_redirect**), and associate the SSL server policy with the HTTPS redirect service.

For more information about digital certificates, see "Configuring PKI". For more information about the SSL server policy configuration, see "Configuring SSL".

Portal authentication through Web does not support security check for users. To implement security check, the client must be the iNode client.

Portal authentication supports NAT traversal whether it is initiated by a Web client or an iNode client. NAT traversal must be configured when the portal client is on a private network and the portal server is on a public network.

If the portal clients and the portal server belong to different VPNs, the Rule ARP entry feature will cause the portal clients to go offline abnormally. To avoid this issue, execute the **`undo portal refresh arp enable`** command to disable the Rule ARP entry feature for portal clients.

Portal authentication tasks at a glance

To configure portal authentication, perform the following tasks:

1. Configuring a remote portal service

   Perform this task if the access device does not act as a portal authentication server or portal Web server.

   - Configuring a remote portal authentication server
   - Configuring a portal Web server

2. Configuring a local portal service

   Perform this task if the access device acts as a portal authentication server and portal Web server.

   - Configuring local portal service features
   - Configuring a portal Web server

3. Enabling portal authentication and specifying a portal Web server
   - Enabling portal authentication on an interface
   - Specifying a portal Web server on an interface
   - Enabling portal to support IPv4/IPv6 dual stack

     Perform this task if dual-stack portal users exist on the network.

4. (Optional.) Configure parameters for preauthentication portal users
   - Configuring a portal preauthentication domain
   - Specifying a preauthentication IP address pool
5. (Optional.) Specifying a portal authentication domain
6. (Optional.) Controlling portal user access
   - Configuring a portal-free rule
   - Configuring an authentication source subnet
   - Configuring an authentication destination subnet
   - Configuring support of Web proxy for portal authentication
   - Checking the issuing of category-2 portal filtering rules
   - Setting the maximum number of portal users
   - Enabling strict-checking on portal authorization information
   - Allowing only users with DHCP-assigned IP addresses to pass portal authentication
   - Enabling portal roaming
   - Configuring the portal fail-permit feature
7. (Optional.) Configuring portal detection features
   - Configuring online detection of portal users
   - Configuring portal authentication server detection
   - Configuring portal Web server detection
   - Configuring portal user synchronization
8. (Optional.) Configuring attributes for portal packets and RADIUS packets
   - Configuring portal packet attributes

     You can configure the BAS-IP or BAS-IPv6 attribute for portal packets and specify the device ID.

   - Configuring attributes for RADIUS packets

You can configure the NAS-Port-Id and NAS-Port-Type attributes and apply a NAS-ID profile to an interface.

9. (Optional.) Configuring MAC-based quick portal authentication
   a. Configuring a MAC binding server
   b. Specifying a MAC binding server on an interface
   c. Configuring cloud MAC-trigger authentication
10. (Optional.) Configuring IP-based quick portal authentication
11. (Optional.) Configuring portal authentication to support M-LAG
12. (Optional.) Configuring online and offline related features for portal users
    - Logging out online portal users
    - Enabling portal user login/logout logging
13. (Optional.) Configuring extended portal authentication features
    - Disabling the Rule ARP or ND entry feature for portal clients
    - Configuring Web redirect
14. Obtaining user access information from ARP or ND entries

    This task is required in a remote Web authentication network.

Prerequisites for portal authentication

The portal feature provides a solution for user identity authentication and security check. To complete user identity authentication, portal must cooperate with RADIUS.

Before you configure portal, you must complete the following tasks:

- The portal authentication server, portal Web server, and RADIUS server have been installed and configured correctly.
- To use the re-DHCP portal authentication mode, make sure the DHCP relay agent is enabled on the access device, and the DHCP server is installed and configured correctly.
- The portal client, access device, and servers can reach each other.
- To use the remote RADIUS server, configure usernames and passwords on the RADIUS server, and configure the RADIUS client on the access device. For information about RADIUS client configuration, see "Configuring AAA."
- To implement extended portal functions, install and configure IMC EAD. Make sure the ACLs configured on the access device correspond to the isolation ACL and the security ACL on the security policy server. For installation and configuration about the security policy server, see *IMC EAD Security Policy Help*.

Configuring a remote portal authentication server

About this task

With portal authentication enabled, the device searches for a portal authentication server for a received portal request packet according to the source IP address and VPN information of the packet.

- If a matching portal authentication server is found, the device regards the packet valid and sends an authentication response packet to the portal authentication server. After a user logs in to the device, the user interacts with the portal authentication server as needed.
- If no matching portal authentication server is found, the device drops the packet.

Restrictions and guidelines

Do not delete a portal authentication server in use. Otherwise, users authenticated by that server cannot log out correctly.

Procedure

1. Enter system view.

   system-view

2. Create a portal authentication server and enter its view.

   portal server *server-name*

   You can create multiple portal authentication servers.

3. Specify the IP address of the portal authentication server.

   IPv4:

   ip *ipv4-address* [**vpn-instance** *vpn-instance-name*] [**key** { **cipher** | **simple** } *string*]

   IPv6:

   ipv6 *ipv6-address* [**vpn-instance** *vpn-instance-name*] [**key** { **cipher** | **simple** } *string*]

4. (Optional.) Set the destination UDP port number used by the device to send unsolicited portal packets to the portal authentication server.

   port *port-number*

   By default, the UDP port number is 50100.

   This port number must be the same as the listening port number specified on the portal authentication server.

5. (Optional.) Specify the portal authentication server type.

   server-type { **cmcc** | **imc** }

   By default, the portal authentication server type is IMC.

   The specified server type must be the same as the type of the portal authentication server actually used.

6. (Optional.) Configure the device to periodically register with the portal authentication server.

   server-register [**interval** *interval-value*]

   By default, the device does not register with a portal authentication server.

Configuring a portal Web server

Portal Web server tasks at a glance

To configure a portal Web server, perform the following tasks:

1. Configure basic parameters for a portal Web server
2. (Optional.) Enabling the captive-bypass feature
3. (Optional.) Configuring a match rule for URL redirection
4. (Optional.) Configuring unescaped special characters for portal Web server URLs redirected to users

Configure basic parameters for a portal Web server

1. Enter system view.

system-view

2. Create a portal Web server and enter its view.

   portal web-server *server-name*

   You can create multiple portal Web servers.

3. Specify the VPN instance to which the portal Web server belongs.

   vpn-instance *vpn-instance-name*

   By default, the portal Web server belongs to the public network.

4. Specify the URL of the portal Web server.

   url *url-string*

   By default, no URL is specified for a portal Web server.

5. Configure the parameters to be carried in the URL when the device redirects it to users.

   url-parameter *param-name* { **original-url** | **source-address** | **source-mac** [**format section** { **1** | **3** | **6** } { **lowercase** | **uppercase** }] [**encryption** { **aes** | **des** } **key** { **cipher** | **simple** } *string*] | **value** *expression* }

   By default, no redirection URL parameters are configured.

6. (Optional.) Specify the portal Web server type.

   server-type { **cmcc** | **imc** | **ise** | **oauth** }

   By default, the portal Web server type is IMC.

   This configuration is applicable to only to the remote portal service.

   The specified server type must be the same as the type of the portal Web server actually used.

Enabling the captive-bypass feature

About this task

Typically, when iOS mobile devices or some Android mobile devices are connected a portal-enabled network, the device pushes the authentication page to the mobile devices.

The captive-bypass feature enables the device to push the portal authentication page to the iOS and Android devices only when the users access the Internet by using a browser. If the users do not perform authentication but press the home button to return to the desktop, the Wi-Fi connection is terminated. To maintain the Wi-Fi connection in such cases, you can enable the optimized captive-bypass feature.

When optimized captive-bypass is enabled, the portal authentication page is automatically pushed to iOS mobile devices after they connects to the network. Users can perform authentication on the page or press the home button to return to the desktop without performing authentication, and the Wi-Fi connection is not terminated.

When an iOS client is connected to a network, it automatically sends a server reachability detection packet to determine whether the Apple server is reachable. If the server is reachable, the Wi-Fi connection displays connected. If the server is not reachable, the Wi-Fi connection is terminated.

When the network condition is poor, the device cannot detect a server reachability detection packet from an iOS mobile client within the captive-bypass detection timeout time. The client cannot receive a response for the server reachability detection packet, and therefore it determines the server to be unreachable and terminates the Wi-Fi connection. To avoid Wi-Fi disconnections caused by server reachability detection failure, set a longer captive-bypass detection timeout time when the network condition is poor.

Procedure

1. Enter system view.

   system-view

2. Enter portal Web server view.

   portal web-server *server-name*

3. Enable the captive-pass feature.

   captive-bypass [**android** | **ios** [**optimize**]] **enable**

   By default, the captive-bypass feature is disabled. The device automatically pushes the portal authentication page to iOS mobile devices and some Android mobile devices when they are connected to a portal-enabled network.

4. (Optional.) Set the captive-bypass detection timeout time.

   a. Return to system view.

      quit

   b. Set the captive-bypass detection timeout time.

      portal captive-bypass optimize delay *seconds*

      By default, the captive-bypass detection timeout is 6 seconds.

Configuring a match rule for URL redirection

About this task

A URL redirection match rule matches HTTP or HTTPS requests by user-requested URL or User-Agent information, and redirects the matching requests to the specified redirection URL. Therefore, URL redirection match rules allow for more flexible URL redirection than the **url** command. The **url** command is only used to redirect HTTP or HTTPS requests from unauthenticated users to the portal Web server for authentication.

Restrictions and guidelines

For a user to successfully access a redirection URL, configure a portal-free rule to allow HTTP or HTTPS requests destined for the redirection URL to pass. For information about configuring portal-free rules, see the **portal free-rule** command.

If both the **url** and **if-match** commands are executed, the **if-match** command takes priority to perform URL redirection.

Procedure

1. Enter system view.

   system-view

2. Enter portal Web server view.

   portal web-server *server-name*

3. Configure a match rule for URL redirection.

   if-match { **original-url** *url-string* **redirect-url** *url-string* [**url-param-encryption** { **aes** | **des** } **key** { **cipher** | **simple** } *string*] | **user-agent** *string* **redirect-url** *url-string* }

Configuring unescaped special characters for portal Web server URLs redirected to users

About this task

By default, the device escapes all special characters in the portal Web server URL redirected to users. This practice converts special characters to a secure form to prevent the clients from misinterpreting them as syntax symbols or instructions.

However, a portal Web server might be unable to recognize the escape characters for certain special characters in URLs. In this case, you must configure the device to not escape these special characters so the portal Web server can correctly deliver the Web authentication page to users.

Prerequisites

Before you specify unescaped special characters, identify the escape characters that the portal Web server cannot recognize.

Procedure

1. Enter system view.

   ```
   system-view
   ```

2. Specify unescaped special characters for portal Web server URLs redirected to users.

   ```
   portal url-unescape-chars character-string
   ```

   By default, the device escapes all special characters in portal Web server URLs redirected to users.

Configuring local portal service features

About the local portal service

After a local portal service is configured, the device acts as the portal Web server and portal authentication server to perform portal authentication on users. The portal authentication page file is saved in the root directory of the device.

Restrictions and guidelines for configuring local portal service features

For an interface to use the local portal service, the URL of the portal Web server specified for the interface must meet the following requirements:

- The IP address in the URL must be the IP address of a Layer 3 interface (except 127.0.0.1) on the device, and the IP address must be reachable to portal clients.
- The URL must be ended with **/portal/**. For example: **https://1.1.1.1/portal/**.

The device provides a default authentication page file. To use customized authentication pages, you must edit the authentication pages, compress them to a file, and then upload the file to the device. On the device, specify the file as the default authentication page file.

Customizing authentication pages

About this task

Authentication pages are HTML files. Local portal authentication requires the following authentication pages:

- Logon page
- Logon success page
- Logon failure page
- Online page
- System busy page
- Logoff success page

You must customize the authentication pages, including the page elements that the authentication pages will use, for example, **back.jpg** for authentication page **Logon.htm**.

Follow the authentication page customization rules when you edit the authentication page files.

File name rules

The names of the main authentication page files are fixed (see Table 1). You can define the names of the files other than the main authentication page files. File names and directory names are case insensitive.

Table 1 Main authentication page file names

Main authentication page	File name
Logon page	logon.htm
Logon success page	logonSuccess.htm
Logon failure page	logonFail.htm
Online page Pushed after the user gets online for online notification	online.htm
System busy page Pushed when the system is busy or the user is in the logon process	busy.htm
Logoff success page	logoffSuccess.htm

Page request rules

The local portal Web server supports only Get and Post requests.

- **Get requests**—Used to get the static files in the authentication pages and allow no recursion. For example, if file **Logon.htm** includes contents that perform Get action on file **ca.htm**, file **ca.htm** cannot include any reference to file **Logon.htm**.
- **Post requests**—Used when users submit username and password pairs, log in, and log out.

Post request attribute rules

1. Observe the following requirements when editing a form of an authentication page:
   - An authentication page can have multiple forms, but there must be one and only one form whose action is **logon.cgi**. Otherwise, user information cannot be sent to the access device.
   - The username attribute is fixed as **PtUser**. The password attribute is fixed as **PtPwd**.
   - The value of the **PtButton** attribute is either **Logon** or **Logoff**, which indicates the action that the user requests.
   - A logon Post request must contain **PtUser**, **PtPwd**, and **PtButton** attributes.
   - A logoff Post request must contain the **PtButton** attribute.
2. Authentication pages **logon.htm** and **logonFail.htm** must contain the logon Post request.

   The following example shows part of the script in page **logon.htm**.

```
<form action=logon.cgi method = post >
<p>User name:<input type="text" name = "PtUser" style="width:160px;height:22px"
maxlength=64>
<p>Password :<input type="password" name = "PtPwd" style="width:160px;height:22px"
maxlength=32>
<p><input type=SUBMIT value="Logon" name = "PtButton" style="width:60px;"
onclick="form.action=form.action+location.search;">
</form>
```

3. Authentication pages **logonSuccess.htm** and **online.htm** must contain the logoff Post request.

   The following example shows part of the script in page **online.htm**.

```
<form action=logon.cgi method = post >
<p><input type=SUBMIT value="Logoff" name="PtButton" style="width:60px;">
</form>
```

Page file compression and saving rules

You must compress the authentication pages and their page elements into a standard zip file.

- The name of a zip file can contain only letters, numbers, and underscores.
- The authentication pages must be placed in the root directory of the zip file.
- Zip files can be transferred to the device through FTP or TFTP and must be saved in the root directory of the device.

  Examples of zip files on the device:

```
<Sysname> dir
Directory of flash:
   1     -rw-      1405  Feb 28 2008 15:53:20   ssid1.zip
   0     -rw-      1405  Feb 28 2008 15:53:31   ssid2.zip
   2     -rw-      1405  Feb 28 2008 15:53:39   ssid3.zip
   3     -rw-      1405  Feb 28 2008 15:53:44   ssid4.zip
2540 KB total (1319 KB free)
```

Redirecting authenticated users to a specific webpage

To make the device automatically redirect authenticated users to a specific webpage, do the following in logon.htm and logonSuccess.htm:

1. In logon.htm, set the target attribute of Form to _**blank**.

   See the contents in gray:

```
<form method=post action=logon.cgi target="_blank">
```

2. Add the function for page loading pt_init() to LogonSuccess.htm.

   See the contents in gray:

```
<html>
<head>
<title>LogonSuccess</title>
<script type="text/javascript" language="javascript"
src="pt_private.js"></script>
</head>
<body onload="pt_init();" onbeforeunload="return pt_unload();">
... ...
</body>
</html>
```

Configuring a local portal Web service

Prerequisites

Before you configure an HTTPS-based local portal Web service, you must complete the following tasks:

- Configure a PKI policy, obtain the CA certificate, and request a local certificate. For more information, see "Configuring PKI."
- Configure an SSL server policy, and specify the PKI domain configured in the PKI policy.

During SSL connection establishment, the user browser might display a message that it cannot verify server identity by certificate. For users to perform portal authentication without checking such a message, configure an SSL server policy to request a client-trusted certificate on the device. The name of the policy must be **https_redirect**. For more information about SSL server policy configuration, see "Configuring SSL."

Procedure

1. Enter system view.

   system-view

2. Enable HTTP- or HTTPS-based local portal Web service and enter its view.

   portal local-web-server { **http** | **https ssl-server-policy** *policy-name* [**tcp-port** *port-number*] }

3. Specify the default authentication page file for the local portal Web service.

   default-logon-page *filename*

   By default, the default authentication page file for a local portal Web service is file **defaultfile.zip**.

4. (Optional.) Configure the HTTP or HTTPS listening TCP port for the local portal Web service.

   tcp-port *port-number*

   By default, the HTTP service listening port number is 80 and the HTTPS service listening port number is the TCP port number set by the **portal local-web-server** command.

5. (Optional.) Bind the endpoint name to an authentication page file.

   logon-page bind device-name *device-name* **file** *file-name*

   By default, no endpoint name is bound to an authentication page file.

Enabling portal authentication on an interface

Restrictions and guidelines

When you enable portal authentication on an interface, follow these restrictions and guidelines:

- Cross-subnet authentication mode (**layer3**) does not require Layer 3 forwarding devices between the access device and the portal authentication clients. However, if a Layer 3 forwarding device exists between the authentication client and the access device, you must use the cross-subnet portal authentication mode.
- You can enable both IPv4 portal authentication and IPv6 portal authentication on an interface.

When you configure re-DHCP portal authentication on an interface, follow these restrictions and guidelines:

- Make sure the interface has a valid IP address before you enable re-DHCP portal authentication on the interface.
- With re-DHCP portal authentication, configure authorized ARP on the interface as a best practice to make sure only valid users can access the network. With authorized ARP configured on the interface, the interface learns ARP entries only from the users who have obtained a public address from DHCP.
- For successful re-DHCP portal authentication, make sure the BAS-IP or BAS-IPv6 attribute value is the same as the device IP address specified on the portal authentication server. To configure the attribute, use the **portal** { **bas-ip** | **bas-ipv6** } command.
- An IPv6 portal server does not support re-DHCP portal authentication.

Portal authentication supports both HTTP redirect and HTTPS redirect. The default HTTPS redirect listening port number on the device is 6654. To change the HTTPS redirect listening port number, see HTTP redirect configuration in *Layer 3—IP Services Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Enable portal authentication.
   IPv4:
   portal enable method { **direct** [**shared**] | **layer3** | **redhcp** [**shared**] }
   IPv6:
   portal ipv6 enable method { **direct** [**shared**] | **layer3** }
   By default, portal authentication on an interface is disabled.

Specifying a portal Web server on an interface

About this task

With a portal Web server specified on an interface, the device redirects the HTTP requests of portal users on the interface to the portal Web server.

You can specify both an IPv4 portal Web server and an IPv6 portal Web server on an interface.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Specify a portal Web server on the interface.
   portal [**ipv6**] **apply web-server** *server-name* [**fail-permit** | **secondary**]
   By default, no portal Web servers are specified on an interface.

Enabling portal to support IPv4/IPv6 dual stack

About this task

When portal is enabled to support IPv4/IPv6 dual stack, dual-stack portal users can access both IPv4 and IPv6 networks after passing one type (IPv4 or IPv6) of portal authentication.

The following describes the mechanism of how portal supports IPv4/IPv6 dual stack:

1. A portal user passes portal authentication of one IP stack (the first stack, IPv4 or IPv6). The user can access network resources of this IP stack.
2. The device records the user's MAC address and IP address in the portal user entry for the user.
3. Upon receiving a packet of the other IP stack (the second stack, IPv6 or IPv4) from the user, the device compares the source MAC address with that in the portal user entry for the user.
   - If the MAC addresses are the same, the device allows the user to access network resources of the second stack without reauthentication.
   - If the MAC addresses are different, the user needs to perform portal authentication of the second stack.

Restrictions and guidelines

This configuration takes effect on an interface only when both direct IPv4 portal authentication and direct IPv6 portal authentication are enabled on the interface.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Enable portal to support IPv4/IPv6 dual stack.
   portal dual-stack enable

   By default, portal support for IPv4/IPv6 dual stack is disabled. Dual-stack portal users can access only one type (IPv4 or IPv6) of network after passing that type of portal authentication.

Configuring a portal preauthentication domain

About this task

A portal preauthentication domain defines user attributes assigned to preauthentication portal users on a portal-enabled interface after the users obtain IP addresses. Before the preauthentication users pass portal authentication, they have limited access to the network based on the assigned user attributes (such as ACL). After the users pass portal authentication, they are assigned new attributes by the AAA server. After the users go offline, they are re-assigned user attributes in the preauthentication domain.

Restrictions and guidelines

The portal preauthentication domain takes effect only on portal users with IP addresses obtained through DHCP or DHCPv6.

The portal preauthentication domain does not take effect on interfaces enabled with cross-subnet portal authentication.

Make sure you specify an existing ISP domain as a portal preauthentication domain. If the specified ISP domain does not exist, the device might operate incorrectly.

You must delete a preauthentication domain (by using the **undo portal** [**ipv6**] **pre-auth domain** command) and reconfigure it in the following situations:

- You create the ISP domain after specifying it as the preauthentication domain.
- You delete the specified ISP domain and then re-create it.

For the authorization ACL in the preauthentication domain, the following rules apply:

- If the traffic of preauthentication users matches a rule in the ACL, the device processes the traffic based on the permit or deny statement of the rule.
- If the ACL does not exist or the permitted destination address of a rule in the ACL is set to **any**, the device does not control user access. Users can access any network resources without passing portal authentication.
- If the ACL does not have any rules, all user accesses are restricted. The device allows users to access network resources only after they pass authentication.
- If the traffic of preauthentication users does not match any rule in the ACL, the device pushes the authentication page to the users. The users can access the network resources after passing authentication.
- If the ACL contains rules that specify a source address, users might not be able to get online. Do not specify a source IPv4, IPv6 or MAC address when you configure a rule in the ACL.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Specify a portal preauthentication domain.
   portal [**ipv6**] **pre-auth domain** *domain-name*

Specifying a preauthentication IP address pool

About this task

You must specify a preauthentication IP address pool on a portal-enabled interface in the following situation:

- Portal users access the network through a subinterface of the portal-enabled interface.
- The subinterface does not have an IP address.
- Portal users need to obtain IP addresses through DHCP.

After a user connects to a portal-enabled interface, the user uses an IP address for portal authentication according to the following rules:

- If the interface is configured with a preauthentication IP address pool, the user uses the following IP address:
  - If the client is configured to obtain an IP address automatically through DHCP, the user obtains an address from the specified IP address pool.
  - If the client is configured with a static IP address, the user uses the static IP address.
- If the interface has an IP address but no preauthentication IP pool specified, the user uses the static IP address or the IP address obtained from a DHCP server.
- If the interface has no IP address or preauthentication IP pool specified, the user cannot perform portal authentication.

After the user passes portal authentication, the AAA server authorizes an IP address pool for re-assigning an IP address to the user. If no authorized IP address pool is deployed, the user continues using the previous IP address.

Restrictions and guidelines

This configuration takes effect only when the direct IPv4 portal authentication is enabled on the interface.

Make sure the specified IP address pool exists and is complete. Otherwise, the user cannot obtain the IP address and cannot perform portal authentication.

If the portal user does not perform authentication or fails to pass authentication, the assigned IP address is still retained.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Specify a preauthentication IP address pool on the interface.
   portal [**ipv6**] **pre-auth ip-pool** *pool-name*

   By default, no preauthentication IP address pool is specified on an interface.

Specifying a portal authentication domain

About portal authentication domains

An authentication domain defines a set of authentication, authorization, and accounting policies. Each portal user belongs to an authentication domain and is authenticated, authorized, and accounted in the domain.

With an authentication domain specified on an interface, the device uses the authentication domain for AAA of portal users. This allows for flexible portal access control.

Restrictions and guidelines for specifying a portal authentication domain

The device selects the authentication domain for a portal user in this order:

1. ISP domain specified for the interface.
2. ISP domain carried in the username.
3. System default ISP domain.

If the chosen domain does not exist on the device, the device searches for the ISP domain configured to accommodate users assigned to nonexistent domains. If no such ISP domain is configured, user authentication fails. For information about ISP domains, see "Configuring AAA."

For the authorization ACL in the authentication domain, the following rules apply:

- If the user traffic matches a rule in the ACL, the device processes the traffic based on the permit or deny statement of the rule.
- If the user traffic does not match any rule in the ACL, the device permits the traffic. To deny such traffic, configure the last rule in the ACL to deny all packets by using the **rule deny ip** command.
- If the ACL contains rules that specify a source address, users might not be able to get online. Do not specify a source IPv4, IPv6, or MAC address when you configure a rule in the ACL.

Specifying a portal authentication domain on an interface

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Specify an portal authentication domain on the interface.
   portal [**ipv6**] **domain** *domain-name*

   By default, no portal authentication domain is specified on an interface.

   You can specify both an IPv4 portal authentication domain and an IPv6 portal authentication domain on an interface.

Controlling portal user access

Configuring a portal-free rule

About this task

A portal-free rule allows specified users to access specified external websites without portal authentication.

The matching items for a portal-free rule include the host name, source/destination IP address, TCP/UDP port number, source MAC address, microsegment attributes (VPN instance and microsegment ID settings) to which the destination IP address belongs, access interface, and VLAN. Packets matching a portal-free rule will not trigger portal authentication, so users sending the packets can directly access the specified external websites.

Restrictions and guidelines for configuring a portal-free rule

If you specify both a VLAN and an interface, the interface must belong to the VLAN. If the interface does not belong to the VLAN, the portal-free rule does not take effect.

You cannot configure two or more portal-free rules with the same filtering criteria. Otherwise, the system prompts that the rule already exists.

Regardless of whether portal authentication is enabled or not, you can only add or remove a portal-free rule. You cannot modify it.

Before configuring the destination microsegment for a portal-free rule, enable microsegmentation. If a VPN instance is configured for the destination microsegment, make sure the VPN instance has been created. Otherwise, the portal-free rule containing the microsegment does not take effect. To immediately invalidate a portal-free rule containing a microsegment when microsegmentation is disabled, manually delete the portal-free rule. For more information about microsegmentation, see "Configuring microsegmentation."

Configuring an IP-based portal-free rule

1. Enter system view.

   system-view

2. Configure an IP-based portal-free rule.

   IPv4:

   portal free-rule *rule-number* { **destination ip** { *ipv4-address* { *mask-length* | *mask* } | **any** | **microsegment** *microsegment-id* [**vpn-instance** *vpn-instance-name*] } [**tcp** *tcp-port-number* | **udp** *udp-port-number*] | **source ip** { *ipv4-address* { *mask-length* | *mask* } | **any** } [**tcp** *tcp-port-number* | **udp** *udp-port-number*] } * [**interface** *interface-type interface-number*]

   IPv6:

   portal free-rule *rule-number* { **destination ipv6** { *ipv6-address prefix-length* | **any** | **microsegment** *microsegment-id* [**vpn-instance** *vpn-instance-name*] } [**tcp** *tcp-port-number* | **udp** *udp-port-number*] | **source ipv6** { *ipv6-address prefix-length* | **any** } [**tcp** *tcp-port-number* | **udp** *udp-port-number*] } * [**interface** *interface-type interface-number*]

Configuring a source-based portal-free rule

1. Enter system view.

   system-view

2. Configure a source-based portal-free rule.

portal free-rule *rule-number* **source** { **interface** *interface-type interface-number* | **mac** *mac-address* | **vlan** *vlan-id* } *

The **vlan** *vlan-id* option takes effect only on portal users that access the network through VLAN interfaces.

Configuring a destination-based portal-free rule

1. Enter system view.

   system-view

2. Configure a destination-based portal-free rule.

   portal free-rule *rule-number* **destination** *host-name*

   Before you configure destination-based portal-free rules, make sure a DNS server is deployed on the network.

Configuring an authentication source subnet

About this task

By configuring authentication source subnets, you specify that only HTTP or HTTPS packets from users on the authentication source subnets can trigger portal authentication. HTTP or HTTPS packets from users not in authentication source subnets will be discarded if the packets do not match any portal-free rules.

Restrictions and guidelines

Authentication source subnets are used only for cross-subnet authentication.

In direct or re-DHCP portal authentication mode, a portal user and its access interface (portal-enabled) are on the same subnet. It is not necessary to specify the subnet as the authentication source subnet.

- In direct mode, the access device regards the authentication source subnet as any source IP address.
- In re-DHCP mode, the access device regards the authentication source subnet on an interface as the subnet to which the private IP address of the interface belongs.

If both authentication source subnets and destination subnets are configured on an interface, only the authentication destination subnets take effect.

You can configure multiple authentication source subnets. If the source subnets overlap, the subnet with the largest address scope (with the smallest mask or prefix) takes effect.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 3 interface view.

   interface *interface-type interface-number*

3. Configure a portal authentication source subnet.

   IPv4:

   portal layer3 source *ipv4-network-address* { *mask-length* | *mask* }

   IPv6:

   portal ipv6 layer3 source *ipv6-network-address prefix-length*.

   By default, no portal authentication source subnets are configured, and users from any subnets can trigger portal authentication.

Configuring an authentication destination subnet

About this task

After authentication destination subnets are configured, the device performs portal authentication on portal users only when they access the specified subnets (excluding the destination IP addresses and subnets specified in portal-free rules). Users can access other subnets without portal authentication.

Restrictions and guidelines

If both authentication source subnets and destination subnets are configured on an interface, only the authentication destination subnets take effect.

You can configure multiple authentication destination subnets. If the destination subnets overlap, the subnet with the largest address scope (with the smallest mask or prefix) takes effect.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 3 interface view.

   interface *interface-type interface-number*

3. Configure a portal authentication destination subnet.

   IPv4:

   portal free-all except destination *ipv4-network-address* { *mask-length* | *mask* }

   IPv6:

   portal ipv6 free-all except destination *ipv6-network-address prefix-length*

   By default, no authentication destination subnets are configured, and users accessing any subnets must pass portal authentication.

Configuring support of Web proxy for portal authentication

About this task

To allow HTTP requests proxied by a Web proxy server to trigger portal authentication, specify the port number of the Web proxy server on the device. If a Web proxy server port is not specified on the device, HTTP requests proxied by the Web proxy server are dropped, and portal authentication cannot be triggered.

Restrictions and guidelines

If a user's browser uses the Web Proxy Auto-Discovery (WPAD) protocol to discover Web proxy servers, you must perform the following tasks on the device:

- Specify port numbers of the Web proxy servers.
- Configure portal-free rules to allow user packets destined for the WPAD server to pass without authentication.

If portal users enable Web proxy in their browsers, the users must add the IP address of the portal authentication server as a proxy exception in their browsers. Thus, HTTP packets that the users send to the portal authentication server will not be sent to Web proxy servers.

You cannot specify Web proxy server port 443 on the device.

You can execute this command multiple times to specify multiple port numbers of Web proxy servers.

Procedure

1. Enter system view.
   system-view
2. Specify the port number of a Web proxy server.
   portal web-proxy port *port-number*
   By default, no port numbers of Web proxy servers are specified. Proxied HTTP requests are dropped.

Checking the issuing of category-2 portal filtering rules

About this task

Category-2 portal filtering rules permit authenticated users to access authorized network resources. By default, the device allows an authenticated user to come online as long as a member device has issued a category-2 portal filtering rule for the user. Users coming online from global interfaces might fail to access network resources because some member ports might not have category-2 rules for the users. To resolve this issue, enable the device to check the issuing of category-2 portal filtering rules. Then, the device allows users to come online only when all member devices have issued category-2 portal filtering rules for the users.

As a best practice, perform this task when portal authentication is enabled on a global interface.

Procedure

1. Enter system view.
   system-view
2. Enable the device to check the issuing of category-2 portal filtering rules.
   portal user-rule assign-check enable
   By default, the device does not check the issuing of category-2 portal filtering rules.

Setting the maximum number of portal users

About this task

Perform this task to control the total number of portal users in the system, and the maximum number of IPv4 or IPv6 portal users on an interface.

Restrictions and guidelines for setting the maximum number of portal users

Make sure the maximum combined number of IPv4 and IPv6 portal users specified on all interfaces does not exceed the global maximum number. Otherwise, the exceeding number of portal users will not be able to log in to the device.

Make sure the sum of the maximum number of portal users for IPv4 and IPv6 on an interface does not exceed the maximum number of all portal users set on the interface. Otherwise, the exceeding number of portal users cannot come online from the interface.

You can set the maximum number of portal users in interface view or system view. The maximum number of portal users allowed on an interface is the smallest value among the following items:

- Maximum number of IPv4 or IPv6 portal users allowed on the interface.
- Maximum number of all portal users allowed on the interface.
- Global maximum number of portal users.

Setting the global maximum number of portal users

1. Enter system view.
   system-view

2. Set the global maximum number of portal users.

   portal max-user *max-number*

   By default, no limit is set on the global number of portal users.

   If you set the global maximum number smaller than the number of current online portal users on the device, this configuration still takes effect. The online users are not affected but the system forbids new portal users to log in.

Setting the maximum number of portal users on an interface

1. Enter system view.

   system-view

2. Enter Layer 3 interface view.

   interface *interface-type interface-number*

3. Set the maximum number of portal users.

   portal { **ipv4-max-user** | **ipv6-max-user** | **max-user** } *max-number*

   By default, no limit is set on the number of portal users on an interface.

   If you set the maximum number smaller than the current number of portal users on an interface, this configuration still takes effect. The online users are not affected but the system forbids new portal users to log in from the interface.

Enabling strict-checking on portal authorization information

About this task

The strict checking feature allows a portal user to stay online only when the authorization information for the user is successfully deployed.

Enabling strict checking on portal authentication information on an interface

1. Enter system view.

   system-view

2. Enter Layer 3 interface view.

   interface *interface-type interface-number*

3. Enable strict checking on portal authorization information.

   portal authorization acl strict-checking

CAUTION:

- The device logs out a portal user if the authorized ACL does not exist on the device or fails to be deployed.

By default, strict checking on portal authorization information is disabled on an interface. Portal users stay online even when the authorized ACL does not exist or fails to be deployed.

Allowing only users with DHCP-assigned IP addresses to pass portal authentication

About this task

This feature allows only users with DHCP-assigned IP addresses to pass portal authentication. Use this feature to ensure that only users with valid IP addresses can access the network.

Restrictions and guidelines

This feature takes effect only when the device acts as both the access device and the DHCP server.

Configuration of this feature does not affect the online portal users.

Allowing only users with DHCP-assigned IP addresses to pass portal authentication on an interface

1. Enter system view.

   system-view
2. Enter Layer 3 interface view.

   interface *interface-type interface-number*
3. Allow only users with DHCP-assigned IP addresses to pass portal authentication.

   portal [**ipv6**] **user-dhcp-only**

CAUTION:

- Users with static IP addresses cannot pass portal authentication to come online.
- If this feature is configured, to ensure that IPv6 users can pass portal authentication, disable the temporary IPv6 address feature on terminal devices. Otherwise, IPv6 users will use temporary IPv6 addresses to access the IPv6 network and will fail portal authentication.

By default, both users with IP addresses obtained through DHCP and users with static IP addresses can pass authentication to come online.

Enabling portal roaming

About this task

If portal roaming is enabled on a VLAN interface, an online portal user can access resources from any Layer 2 port in the VLAN without re-authentication.

If portal roaming is disabled, to access external network resources from a Layer 2 port different from the current access port in the VLAN, the user must do the following:

1. Logs out from the current port.
2. Re-authenticates on the new Layer 2 port.

Restrictions and guidelines

Portal roaming takes effect only on portal users logging in from VLAN interfaces. It does not take effect on portal users logging in from common Layer 3 interface.

You cannot enable portal roaming when online portal users or preauthentication portal users exist on the device.

For portal roaming to take effect, you must disable the Rule ARP or ND entry feature by using the **undo portal refresh** { **arp** | **nd** } **enable** command.

Procedure

1. Enter system view.

   system-view
2. Enable portal roaming.

   portal roaming enable

   By default, portal roaming is disabled.

Configuring the portal fail-permit feature

About this task

Perform this task to configure the portal fail-permit feature on an interface. When the access device detects that the portal authentication server or portal Web server is unreachable, it allows users on the interface to have network access without portal authentication.

If you enable fail-permit for both a portal authentication server and a portal Web server on an interface, the interface does the following:

- Disables portal authentication when either server is unreachable.
- Resumes portal authentication when both servers are reachable.

After portal authentication resumes, unauthenticated users must pass portal authentication to access the network. Users who have passed portal authentication before the fail-permit event can continue accessing the network.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Enable portal fail-permit for a portal authentication server.
   portal [**ipv6**] **fail-permit server** *server-name*
   By default, portal fail-permit is disabled for a portal authentication server.
4. Enable portal fail-permit for portal Web servers.
   portal [**ipv6**] **fail-permit web-server**
   By default, portal fail-permit is disabled for portal Web servers.

Specifying a critical profile on an interface and setting the aging time for portal users in critical microsegments

About this task

In direct portal authentication mode, when no authentication server is reachable, the portal critical microsegment feature allows users to access network resources in microsegments of the critical profiles specified on the access interfaces without performing authentication.

After you specify a critical profile on an interface for portal authentication, the critical profile is assigned to the users on the interface when all authentication servers are unreachable. The users then can access network resources in the critical profile, such as the critical microsegment. For users in a VPN instance, the users can access the critical microsegment specified for the VPN instance. If no critical microsegment is specified for the VPN instance, the users can access the default critical microsegment.

Portal users in critical microsegments have an aging time. Aged portal users in critical microsegments are logged out to avoid occupying too much system resources. Set a proper aging time for portal users in critical microsegments according to your network condition.

Restrictions and guidelines

Before specifying a critical profile for an access interface of portal users, make sure microsegmentation is enabled. The critical profile specified for the interface must be configured correctly. For more information about configuring critical profiles, see "Configuring AAA."

In a portal dual-stack network, the microsegment attribute used by the second stack is restricted by the microsegment attribute obtained from the critical profile of the first stack.

Some features such as online detection do not take effect on portal users in critical microsegments. Modification of critical profile configurations causes online portal users in critical microsegments to go offline.

Procedure

1. Enter system view.
   system-view
2. Configure the aging time for portal users in critical microsegments.
   portal critical-microsegment aging-time *minutes*
   By default, the aging time for portal users in critical microsegments is 10 minutes.
3. Enter Layer 3 interface view.
   interface *interface-type interface-number*
4. Specify a critical profile for portal users on the interface.
   portal [**ipv6**] **critical profile** *profile-name*
   By default, no critical profile for portal users is specified on an interface.

Configuring portal detection features

Configuring online detection of portal users

About this task

Use the online detection feature to quickly detect abnormal logouts of portal users. Configure ARP or ICMP detection for IPv4 portal users. Configure ND or ICMPv6 detection for IPv6 portal users.

If the device receives no packets from a portal user within the idle time, the device detects the user's online status as follows:

- **ICMP or ICMPv6 detection**—Sends ICMP or ICMPv6 requests to the user at configurable intervals to detect the user status.
  - If the device receives a reply within the maximum number of detection attempts, it considers that the user is online and stops sending detection packets. Then the device resets the idle timer and repeats the detection process when the timer expires.
  - If the device receives no reply after the maximum number of detection attempts, the device logs out the user.
- **ARP or ND detection**—Sends ARP or ND requests to the user and detects the ARP or ND entry status of the user at configurable intervals.
  - If the ARP or ND entry of the user is refreshed within the maximum number of detection attempts, the device considers that the user is online and stops detection. Then the device resets the idle timer and repeats the detection process when the timer expires.
  - If the ARP or ND entry of the user is not refreshed after the maximum number of detection attempts, the device logs out the user.

Restrictions and guidelines

ARP detection and ND detection apply only to direct and re-DHCP portal authentication. ICMP detection applies to all portal authentication modes.

Procedure

1. Enter system view.
   system-view

2. Enter Layer 3 interface view.

   interface *interface-type interface-number*

3. Configure online detection of portal users.

   IPv4:

   portal user-detect type { **arp** | **icmp** } [**retry** *retries*] [**interval** *interval*] [**idle** *time*]

   IPv6:

   portal ipv6 user-detect type { **icmpv6** | **nd** } [**retry** *retries*] [**interval** *interval*] [**idle** *time*]

   By default, online detection is disabled for portal users on an interface.

Configuring portal authentication server detection

About this task

During portal authentication, if the communication between the access device and portal authentication server is broken, new portal users are not able to log in. Online portal users are not able to log out normally.

To address this problem, the access device needs to be able to detect the reachability changes of the portal server quickly and take corresponding actions to deal with the changes.

The portal authentication server detection feature enables the device to periodically detect portal packets sent by a portal authentication server to determine the reachability of the server. If the device receives a portal packet within a detection timeout (**timeout** *timeout*) and the portal packet is valid, the device considers the portal authentication server to be reachable. Otherwise, the device considers the portal authentication server to be unreachable.

Portal packets include user login packets, user logout packets, and heartbeat packets. Heartbeat packets are periodically sent by a server. By detecting heartbeat packets, the device can detect the server's actual status more quickly than by detecting other portal packets.

Restrictions and guidelines

The portal authentication server detection feature takes effect only when the device has a portal-enabled interface.

Only the IMC portal authentication server supports sending heartbeat packets. To test server reachability by detecting heartbeat packets, you must enable the server heartbeat feature on the IMC portal authentication server.

You can configure the device to take one or more of the following actions when the server reachability status changes:

- Sending a log message, which contains the name, the current state, and the original state of the portal authentication server.
- Enabling portal fail-permit. When the portal authentication server is unreachable, the portal fail-permit feature on an interface allows users on the interface to have network access. When the server recovers, it resumes portal authentication on the interface. For more information, see "Configuring the portal fail-permit feature."
- Make sure the detection timeout configured on the device is greater than the server heartbeat interval configured on the portal authentication server.

Procedure

1. Enter system view.

   system-view

2. Enter portal authentication server view.

portal server *server-name*

3. Configure portal authentication server detection.

   server-detect [**timeout** *timeout*] **log**

   By default, portal authentication server detection is disabled.

Configuring portal Web server detection

About this task

A portal authentication process cannot complete if the communication between the access device and the portal Web server is broken. To address this problem, you can enable portal Web server detection on the access device.

With the portal Web server detection feature, the access device simulates a Web access process to initiate a TCP connection to the portal Web server. If the TCP connection can be established successfully, the access device considers the detection successful, and the portal Web server is reachable. Otherwise, it considers the detection to have failed.

You can configure the following detection parameters:

- **Detection interval**—Interval at which the device detects the server reachability.
- **Maximum number of consecutive failures**—If the number of consecutive detection failures reaches this value, the access device considers that the portal Web server is unreachable.

You can configure the device to take one or more of the following actions when the server reachability status changes:

- Sending a log message, which contains the name, the current state, and the original state of the portal Web server.
- Enabling portal fail-permit. When the portal Web server is unreachable, the portal fail-permit feature on an interface allows users on the interface to have network access. When the server recovers, it resumes portal authentication on the interface. For more information, see "Configuring the portal fail-permit feature."

Restrictions and guidelines

The portal Web server detection feature takes effect only when the URL of the portal Web server is specified and the device has a portal-enabled interface.

Procedure

1. Enter system view.

   system-view

2. Enter portal Web server view.

   portal web-server *server-name*

3. Configure portal Web server detection.

   server-detect [**interval** *interval*] [**retry** *retries*] **log**

   By default, portal Web server detection is disabled.

Configuring portal user synchronization

About this task

Once the access device loses communication with a portal authentication server, the portal user information on the access device and that on the portal authentication server might be inconsistent after the communication resumes. To address this problem, the device provides the portal user synchronization feature. This feature is implemented by sending and detecting portal synchronization packets, as follows:

1. The portal authentication server sends the online user information to the access device in a synchronization packet at the user heartbeat interval.

   The user heartbeat interval is set on the portal authentication server.

2. Upon receiving the synchronization packet, the access device compares the users carried in the packet with its own user list and performs the following operations:
   - If a user contained in the packet does not exist on the access device, the access device informs the portal authentication server to delete the user. The access device starts the synchronization detection timer (**timeout** *timeout*) immediately when a user logs in.
   - If the user does not appear in any synchronization packet within a synchronization detection interval, the access device considers the user does not exist on the portal authentication server and logs the user out.

Restrictions and guidelines

Portal user synchronization requires a portal authentication server to support the portal user heartbeat function. Only the IMC portal authentication server supports the portal user heartbeat function. To implement the portal user synchronization feature, you also need to configure the user heartbeat function on the portal authentication server. Make sure the user heartbeat interval configured on the portal authentication server is not greater than the synchronization detection timeout configured on the access device.

Deleting a portal authentication server on the access device also deletes the user synchronization configuration for the portal authentication server.

Procedure

1. Enter system view.

   system-view

2. Enter portal authentication server view.

   portal server *server-name*

3. Configure portal user synchronization.

   user-sync timeout *timeout*

   By default, portal user synchronization is disabled.

Configuring portal packet attributes

Configuring the BAS-IP or BAS-IPv6 attribute

About this task

If the device runs Portal 2.0, the unsolicited packets sent to the portal authentication server must carry the BAS-IP attribute. If the device runs Portal 3.0, the unsolicited packets sent to the portal authentication server must carry the BAS-IP or BAS-IPv6 attribute.

After this attribute is configured, the source IP address for unsolicited notification portal packets the device sends to the portal authentication server is the configured BAS-IP or BAS-IPv6 address. If the attribute is not configured, the source IP address of the portal packets is the IP address of the packet output interface.

Restrictions and guidelines

During a re-DHCP portal authentication or mandatory user logout process, the device sends portal notification packets to the portal authentication server. For the authentication or logout process to complete, make sure the BAS-IP/BAS-IPv6 attribute is the same as the device IP address specified on the portal authentication server.

You must configure the BAS-IP or BAS-IPv6 attribute on a portal authentication-enabled interface if the following conditions are met:

- The portal authentication server is an IMC server.
- The portal device IP address specified on the portal authentication server is not the IP address of the portal packet output interface.

Configuring the BAS-IP or BAS-IPv6 attribute on an interface

1. Enter system view.

   system-view
2. Enter Layer 3 interface view.

   interface *interface-type interface-number*
3. Configure the BAS-IP or BAS-IPv6 attribute.

   IPv4:

   portal bas-ip *ipv4-address*

   By default, the BAS-IP attribute of an IPv4 portal reply packet is the source IPv4 address of the packet. The BAS-IP attribute of an IPv4 portal notification packet is the IPv4 address of the packet's output interface.

   IPv6:

   portal bas-ipv6 *ipv6-address*

   By default, the BAS-IPv6 attribute of an IPv6 portal reply packet is the source IPv6 address of the packet. The BAS-IPv6 attribute of an IPv6 portal notification packet is the IPv6 address of the packet's output interface.

Specifying the device ID

About this task

The portal authentication server uses device IDs to identify the devices that send protocol packets to the portal server.

Restrictions and guidelines

Make sure the configured device ID is different than any other access devices communicating with the same portal authentication server.

Procedure

1. Enter system view.

   system-view
2. Specify the device ID.

   portal device-id *device-id*

   By default, a device is not configured with a device ID.

Configuring attributes for RADIUS packets

Specifying a format for the NAS-Port-Id attribute

About this task

RADIUS servers from different vendors might require different formats of the NAS-Port-Id attribute in the RADIUS packets. You can specify the NAS-Port-Id attribute format as required.

The device supports predefined formats (format 1, 2, 3, and 4). For more information about the formats, see portal commands in *Security Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Specify the format for the NAS-Port-Id attribute.

   portal nas-port-id format { **1** | **2** | **3** | **4** }

   By default, the format for the NAS-Port-Id attribute is format 2.

Configuring the NAS-Port-Type attribute

About this task

The NAS-Port-Type attribute in a RADIUS request represents the type of a user's access interface.

The access device might not be able to correctly obtain the type of users' access interfaces when multiple network devices exist between the access device and the portal client. For the access device to send the correct access interface type to the RADIUS server, perform this task to configure the NAS-Port-Type attribute.

Restrictions and guidelines

This configuration takes effect only on portal users that newly come online.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 3 interface view.

   interface *interface-type interface-number*

3. Configure the NAS-Port-Type attribute carried in outgoing RADIUS requests on the interface.

   portal nas-port-type { **802.11** | **adsl-cap** | **adsl-dmt** | **async** | **cable** | **ethernet** | **g.3-fax** | **hdlc** | **idsl** | **isdn-async-v110** | **isdn-async-v120** | **isdn-sync** | **piafs** | **sdsl** | **sync** | **virtual** | **wireless-other** | **x.25** | **x.75** | **xdsl** }

   By default, the NAS-Port-Type carried in outgoing RADIUS requests is Ethernet (attribute value 15).

Applying a NAS-ID profile to an interface

About this task

By default, the device sends its device name in the NAS-Identifier attribute of all RADIUS requests.

A NAS-ID profile enables you to send different NAS-Identifier attribute strings in RADIUS requests from different VLANs. The strings can be organization names, service names, or any user categorization criteria, depending on the administrative requirements.

For example, map the NAS-ID **companyA** to all VLANs of company A. The device will send **companyA** in the NAS-Identifier attribute for the RADIUS server to identify requests from any Company A users.

Restrictions and guidelines

You can apply a NAS-ID profile to a portal-enabled interface. If no NAS-ID profile is specified on the interface or no matching NAS-ID is found in the specified profile, the device uses the device name as the interface NAS-ID.

Procedure

1. Enter system view.
   system-view
2. Create a NAS-ID profile and enter NAS-ID profile view.
   aaa nas-id profile *profile-name*
   For more information about this command, see *Security Command Reference*.
3. Configure a NAS ID and VLAN binding in the profile.
   nas-id *nas-identifier* **bind vlan** *vlan-id*
   For more information about this command, see *Security Command Reference*. Portal access matches only the inner VLAN ID of QinQ packets. For more information about QinQ, see *Layer 2—LAN Switching Configuration Guide*.
4. Specify the NAS-ID profile on the interface.
   a. Return to system view.
      quit
   b. Enter Layer 3 interface view.
      interface *interface-type interface-number*
   c. Specify the NAS-ID profile on the interface.
      portal nas-id-profile *profile-name*
      By default, no NAS-ID profile is specified for an interface.

Configuring MAC-based quick portal authentication

Restrictions and guidelines for configuring MAC-based quick portal authentication

Only direct portal authentication supports MAC-based quick portal authentication.

For MAC-based quick portal authentication to take effect on an interface configured with a portal preauthentication policy, set the free-traffic threshold to 0 bytes.

Configuring a MAC binding server

About this task

You can configure multiple MAC binding servers on the device.

Perform this task to configure MAC binding server parameters, such as the server's IP address, VPN instance, port number, and the pre-shared key for communication between the access device and the server.

Procedure

1. Enter system view.
   system-view
2. Create a MAC binding server and enter its view.
   portal mac-trigger-server *server-name*
   By default, no MAC binder servers exist.

3. Configure a MAC binding server.
   - Specify the IP address of the MAC binding server.

     IPv4:

     ip *ipv4-address* [**vpn-instance** *vpn-instance-name*] [**key** { **cipher** | **simple** } *string*]

     IPv6:

     ipv6 *ipv6-address* [**vpn-instance** *vpn-instance-name*] [**key** { **cipher** | **simple** } *string*]

     By default, no MAC binding server address is specified.
   - (Optional.) Set the UDP port number on which the MAC binding server listens for MAC binding query packets.

     port *port-number*

     By default, the MAC binding server listens for MAC binding query packets on UDP port 50100.
   - (Optional.) Set the maximum number of attempts and the interval for sending MAC binding queries to the MAC binding server.

     binding-retry { *retries* | **interval** *interval* } *

     By default, the maximum number of query attempts is 3 and the query interval is 1 second.
   - (Optional.) Specify the type of the MAC binding server.

     server-type { **cmcc** | **imc** }

     By default, the type of a MAC binding server is IMC.
4. (Optional.) Set the free-traffic threshold.

   free-traffic threshold *value*

   By default, the free-traffic threshold is 0 bytes.
5. (Optional.) Set the NAS-Port-Type value carried in RADIUS requests sent to the RADIUS server.

   nas-port-type *value*

   By default, the NAS-Port-Type value carried in RADIUS requests is 0.
6. (Optional.) Specify the version of the portal protocol.

   version *version-number*

   By default, the version of the portal protocol is 1.
7. (Optional.) Set the timeout the device waits for portal authentication to complete after receiving the MAC binding query response.

   authentication-timeout *minutes*

   By default, the portal authentication timeout time is 3 minutes.
8. (Optional.) Set the aging time for MAC-trigger entries.

   aging-time *seconds*

   By default, the aging time for MAC-trigger entries is 300 seconds.

Specifying a MAC binding server on an interface

About this task

After a MAC binding server is specified on an interface, the device can implement MAC-based quick portal authentication for portal users on the interface.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Specify a MAC binding server on the interface.
   portal [**ipv6**] **apply mac-trigger-server** *server-name*
   By default, no MAC binding server is specified on an interface.

Configuring cloud MAC-trigger authentication

About this task

This feature enables the cloud portal server to act as a MAC binding server to perform cloud MAC-trigger authentication on portal users.

Procedure

1. Enter system view.
   system-view
2. Create a MAC binding server and enter its view.
   portal mac-trigger-server *server-name*
3. Enable cloud MAC-trigger authentication.
   cloud-binding enable
   By default, cloud MAC-trigger authentication is disabled.
4. Specify the URL of the cloud portal authentication server.
   cloud-server url *url-string*
   By default, the cloud portal authentication server URL is not specified. The device uses the URL of the portal Web server as the URL of the cloud portal authentication server.

Configuring IP-based quick portal authentication

Prerequisites

Before configuring IP-based quick portal authentication on the device, configure user accounts on the RADIUS server for IP-trigger authentication. Each of the accounts uses the user's IP address as both the username and password.

Restrictions and guidelines

Only direct authentication and cross-subnet authentication support IP-based quick portal authentication.

If both IP-based and MAC-based quick portal authentication modes are enabled on an interface, the MAC-based mode has a higher priority. The device will perform MAC-based quick portal authentication for a user if it can obtain both the MAC address and IP address of the user.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*

3. Configure IP-based quick portal authentication.

   IPv4:

   portal enable method { **direct** | **layer3** } [**ip-trigger** [**acl** *acl-number*]]

   IPv6:

   portal ipv6 enable method { **direct** | **layer3** } [**ip-trigger** [**acl** *acl-number*]]

   By default, IP-based quick portal authentication is disabled.

4. (Optional.) Configure the IP-trigger entry aging time.

   portal ip-trigger aging-time *seconds*

   By default, the IP-trigger entry aging time is 300 seconds.

Configuring portal authentication to support M-LAG

Prerequisites

Before you configure portal authentication to support M-LAG, you must complete the following tasks on both the primary and secondary M-LAG member devices:

- Enable portal authentication.
- Configure IPv4 portal authentication or IPv6 portal authentication.
- Set the portal authentication mode.
- Enable or disable MAC-based quick portal authentication.
- Set the free-traffic threshold.
- Enable portal fail-permit for a portal authentication server.
- Configure the URL of the portal Web server.
- Configure portal user roaming.
- Enable or disable portal authentication to support IPv4/IPv6 dual stack.

Make sure the above configuration on both the primary and secondary M-LAG member devices are the same.

Setting the authentication load sharing mode for portal users on M-LAG interfaces

About this task

By default, the primary M-LAG member device authenticates portal users on all M-LAG interfaces in an M-LAG system. If a large number of portal users exist, perform this task to set the authentication load sharing mode to improve authentication efficiency.

Restrictions and guidelines

Do not set the same distribution criterion for the distributed authentication load sharing mode on both the M-LAG member devices. A violation might cause that portal users fail to come online.

If online portal users already exist on the M-LAG network, do not change the authentication load sharing mode.

Procedure

1. Enter system view.

   system-view

2. Set the authentication load sharing mode for portal users on M-LAG interfaces.

   portal m-lag load-sharing-mode { **centralized** | **distributed** { **even-ip** | **odd-ip** } }

   By default, the authentication load sharing mode is centralized for portal users on M-LAG interfaces.

Configure traffic backup for portal users on M-LAG interfaces

About this task

In an M-LAG system, the M-LAG member devices synchronize traffic of portal users with each other. A M-LAG member device can periodically back up traffic of portal users to the peer device at the specified interval. It can also back up portal user traffic to the peer device if the amount of incremental traffic it has received since the last backup reaches the backup threshold.

Perform this task to decrease the frequency of traffic backup among M-LAG member devices.

Procedure

1. Enter system view.

   system-view

2. Set the traffic backup interval and threshold for portal users on M-LAG interfaces.

   portal m-lag traffic backup { **interval** *interval-value* | **threshold** *threshold-value* } *

   By default, the traffic backup interval is 10 minutes and the traffic backup threshold is 50 MB for portal users on M-LAG interfaces.

Logging out online portal users

About this task

This feature deletes users that have passed portal authentication and terminates ongoing portal authentications.

Restrictions and guidelines

When the number of online users exceeds 2000, executing the **portal delete-user** command takes a few minutes.

To ensure successful logout of online users, do not disable portal authentication or perform master/subordinate device switchover on the portal-enabled interface during the command execution.

Procedure

1. Enter system view.

   system-view

2. Log out online portal users.

   portal delete-user { *ipv4-address* | **all** | **auth-type** { **cloud** | **local** | **normal** } | **interface** *interface-type interface-number* | **ipv6** *ipv6-address* | **mac** *mac-address* }

Enabling portal user login/logout logging

About this task

This feature logs information about user login and logout events. The information includes the username, user IP address and MAC address, user access interface, VLAN, and login result. The logs are sent to the information center of the device. For the logs to be output correctly, you must also configure the information center on the device. For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Enable portal user login/logout logging.
   `portal log enable`
   By default, portal user login/logout logging is disabled.

Disabling the Rule ARP or ND entry feature for portal clients

About this task

When the Rule ARP or ND entry feature is enabled for portal clients, ARP or ND entries for portal clients are Rule entries after the clients come online. After the portal clients go offline, the Rule ARP or ND entries are converted into dynamic entries and deleted when they age out.

Restrictions and guidelines

Enabling or disabling of this feature does not affect existing Rule/dynamic ARP or ND entries.

If the portal clients and the portal server belong to different VPNs, the Rule ARP entry feature will cause the portal clients to go offline abnormally. To avoid this issue, execute the **`undo portal refresh arp enable`** command to disable the Rule ARP entry feature for portal clients.

Procedure

1. Enter system view.
   `system-view`
2. Disable the Rule ARP or ND entry feature for portal clients.
   `undo portal refresh` { **`arp`** | **`nd`** } **`enable`**
   By default, the Rule ARP or ND entry feature is enabled for portal clients.

Configuring Web redirect

About this task

Web redirect is a simplified portal feature. With Web redirect, a user does not perform portal authentication but is directly redirected to the specified URL on the first Web access attempt in a browser. After the specified redirect interval, the user is redirected from the visiting website to the specified URL again.

Web redirect can provide ISPs with extended services. For example, the ISPs can place advertisements and publish information on the redirected webpage.

Restrictions and guidelines

The Web redirect feature takes effect only on HTTP packets that use the default port number 80.

If the device supports Etherchannel interfaces, you can enable both Web redirect and portal authentication on interfaces that support these features. If the device does not support Etherchannel interfaces, Web redirect does not work when both Web redirect and portal authentication are enabled on an interface.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Configure Web redirect.
   web-redirect [**ipv6**] **url** *url-string* [**interval** *interval*]
   By default, Web redirect is disabled.

Setting the user synchronization interval for portal authentication using OAuth

About this task

If portal authentication uses OAuth, the device periodically reports user information to the portal authentication server for user synchronization on the server. To disable user synchronization from the device to the portal authentication server, set the user synchronization interval to 0 seconds on the device.

Procedure

1. Enter system view.
   system-view
2. Set the user synchronization interval for portal authentication using OAuth.
   portal oauth user-sync interval *interval*
   By default, the user synchronization interval is 60 seconds.

Obtaining user access information from ARP or ND entries

About this task

In a remote Web authentication network, the device cannot obtain user access information from FIB entries upon receiving portal packets from the portal authentication server. This will cause users to fail Web authentication.

To resolve this issue, you can perform this task on the device. When this feature is enabled, the device first attempts to obtain user access information from ARP or ND entries during Web authentication.

Procedure

1. Enter system view.
   system-view

2. Configure the device to obtain user information from ARP or ND entries.
   portal access-info trust { **arp** | **nd** }
   By default, the device obtains user information from FIB entries.

Display and maintenance commands for portal

Execute **display** commands in any view and the **reset** command in user view.

Task	Command
Display portal configuration and portal running state.	**display portal interface** *interface-type interface-number*
Display packet statistics for portal captive-bypass.	**display portal captive-bypass statistics** [**slot** *slot-number*]
Display information about MAC binding servers.	**display portal mac-trigger-server** { **all** | **name** *server-name* }
Display packet statistics for portal authentication servers.	**display portal packet statistics** [**extend-auth-server cloud** | **server** *server-name*]
Display portal rules.	**display portal rule** { **all** | **dynamic** | **static** } **interface** *interface-type interface-number* [**slot** *slot-number*]
Display portal authentication server information.	**display portal server** [*server-name*]
Display session information for portal users or portal-based Web authentication users.	**display portal session user-type** { **portal** | **web-auth** }
Display portal user information.	**display portal user** { **all** | **auth-type** { **cloud** | **local** | **normal** } | **interface** *interface-type interface-number* | **ip** *ipv4-address* | **ipv6** *ipv6-address* | **mac** *mac-address* | **pre-auth** [**interface** *interface-type interface-number* | **ip** *ipv4-address* | **ipv6** *ipv6-address*] | **username** *username* } [**brief** | **verbose**]
Display portal Web server information.	**display portal web-server** [*server-name*]
Display Web redirect rule information.	**display web-redirect rule interface** *interface-type interface-number* [**slot** *slot-number*]
Clear packet statistics for portal captive-bypass.	**reset portal captive-bypass statistics** [**slot** *slot-number*]
Clear packet statistics for portal authentication servers.	**reset portal packet statistics** [**extend-auth-server cloud** | **server** *server-name*]

Portal configuration examples

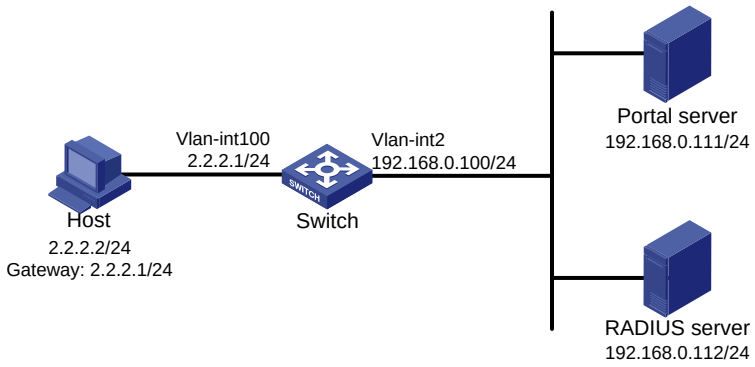
Example: Configuring direct portal authentication

Network configuration

As shown in Figure 7, the host is directly connected to the switch (the access device). The host is assigned a public IP address either manually or through DHCP. An IMC server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure direct portal authentication, so the host can access only the portal server before passing the authentication and access other network resources after passing the authentication.

Figure 7 Network diagram



Prerequisites

- Configure IP addresses for the host, switch, and servers and make sure they can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.
- Configure the portal service on the IMC server, including setting portal server parameters, adding access devices, creating user IP address groups, and configuring device port groups to associate with IP address groups. This ensures coordinated implementation of portal authentication with the device.

Configuring the switch

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.112
[Switch-radius-rs1] primary accounting 192.168.0.112
[Switch-radius-rs1] key authentication simple radius
[Switch-radius-rs1] key accounting simple radius
```

Exclude the ISP domain name from the username sent to the RADIUS server.

```
[Switch-radius-rs1] user-name-format without-domain
[Switch-radius-rs1] quit
```

Enable RADIUS session control.

```
[Switch] radius session-control enable
```

2. Configure an authentication domain:

Create an ISP domain named **dm1** and enter its view.

```
[Switch] domain dm1
```

Configure AAA methods for the ISP domain.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[Switch] domain default enable dm1
```

3. Configure portal authentication:

Configure a portal authentication server.

```
[Switch] portal server newpt
[Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
[Switch-portal-server-newpt] port 50100
[Switch-portal-server-newpt] quit
```

Configure a portal Web server.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[Switch-portal-websvr-newpt] quit
```

Enable direct portal authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal enable method direct
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[Switch–Vlan-interface100] portal apply web-server newpt
```

Configure the BAS-IP as 2.2.2.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[Switch–Vlan-interface100] portal bas-ip 2.2.2.1
[Switch–Vlan-interface100] quit
```

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[Switch] display portal interface vlan-interface 100
 Portal information of Vlan-interface100
     NAS-ID profile: Not configured
     M-LAG group ID: N/A
     M-LAG role    : N/A
     Authorization : Strict checking
     ACL           : Disabled
     User profile  : Disabled
     Max users     : Not configured
 IPv4:
     Portal status: Enabled
     Portal authentication method: Direct    authentication shared mode: Disabled
     Portal web server: newpt(active)
```

```
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: 2.2.2.1
     User detection:  Not configured
     Action for server detection:
         Server type    Server name                         Action
         --             --                                  --
     Layer3 source network:
         IP address              Mask

     Destination authenticate subnet:
         IP address              Mask
IPv6:
     Portal status: Disabled
     Portal authentication method: Disabled    authentication shared mode: Disabled
     Portal web server: Not configured
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ipv6: Not configured
     User detection: Not configured
     Action for server detection:
         Server type    Server name                         Action
         --             --                                  --
     Layer3 source network:
         IP address                                         Prefix length

     Destination authenticate subnet:
         IP address                                         Prefix length
```

A user can perform portal authentication by using the iNode client or through a Web browser. Before passing the authentication, the user can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page. After passing the authentication, the user can access other network resources.

After the user passes authentication, use the following command to display information about the portal user.

```
[Switch] display portal user interface vlan-interface 100
 Total portal users: 1
  Username: abc
  Portal server: newpt
  State: Online
```

```
VPN instance: N/A
MAC               IP              VLAN   Interface
0015-e9a6-7cfe    2.2.2.2         100    Vlan-interface100
Authorization information:
  DHCP IP pool: N/A
  User profile: N/A
  Session group profile: N/A
  ACL number: N/A
  Inbound CAR: N/A
  Outbound CAR: N/A
  Microsegment ID: N/A
```

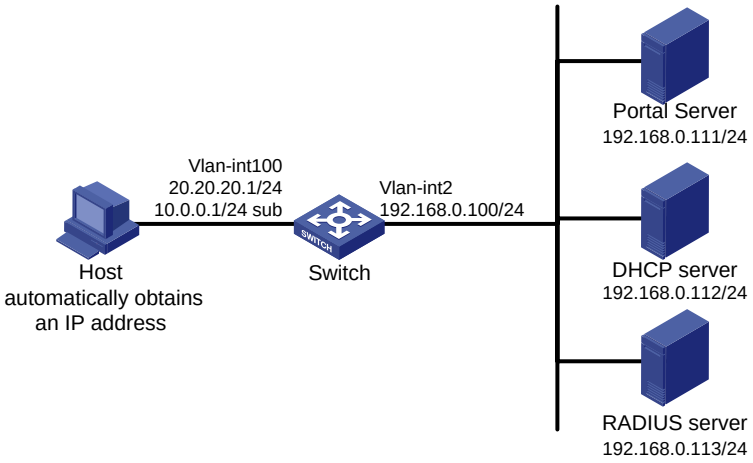
Example: Configuring re-DHCP portal authentication

Network configuration

As shown in Figure 8, the host is directly connected to the switch (the access device). The host obtains an IP address through the DHCP server. A portal server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure re-DHCP portal authentication. Before passing the authentication, the host is assigned a private IP address. After passing the authentication, the host gets a public IP address and can access network resources.

Figure 8 Network diagram



Restrictions and guidelines

- For re-DHCP portal authentication, configure a public address pool (20.20.20.0/24) and a private address pool (10.0.0.0/24) on the DHCP server. (Details not shown.)
- For re-DHCP portal authentication:
  - The switch must be configured as a DHCP relay agent.
  - The portal-enabled interface must be configured with a primary IP address (a public IP address) and a secondary IP address (a private IP address).

  For information about DHCP relay agent configuration, see *Layer 3—IP Services Configuration Guide*.
- Make sure the IP address of the portal device added on the portal server is the public IP address (20.20.20.1) of the switch's interface connecting the host. The private IP address range for the IP address group associated with the portal device is the private subnet 10.0.0.0/24

where the host resides. The public IP address range for the IP address group is the public subnet 20.20.20.0/24.

Prerequisites

- Configure IP addresses for the switch and servers as shown in Figure 8 and make sure the host, switch, and servers can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a RADIUS scheme:

   # Create a RADIUS scheme named **rs1** and enter its view.

   ```
   <Switch> system-view
   [Switch] radius scheme rs1
   ```

   # Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

   ```
   [Switch-radius-rs1] primary authentication 192.168.0.113
   [Switch-radius-rs1] primary accounting 192.168.0.113
   [Switch-radius-rs1] key authentication simple radius
   [Switch-radius-rs1] key accounting simple radius
   ```

   # Exclude the ISP domain name from the username sent to the RADIUS server.

   ```
   [Switch-radius-rs1] user-name-format without-domain
   [Switch-radius-rs1] quit
   ```

   # Enable RADIUS session control.

   ```
   [Switch] radius session-control enable
   ```

2. Configure an authentication domain:

   # Create an ISP domain named **dm1** and enter its view.

   ```
   [Switch] domain dm1
   ```

   # Configure AAA methods for the ISP domain.

   ```
   [Switch-isp-dm1] authentication portal radius-scheme rs1
   [Switch-isp-dm1] authorization portal radius-scheme rs1
   [Switch-isp-dm1] accounting portal radius-scheme rs1
   [Switch-isp-dm1] quit
   ```

   # Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

   ```
   [Switch] domain default enable dm1
   ```

3. Configure DHCP relay and authorized ARP:

   # Configure DHCP relay.

   ```
   [Switch] dhcp enable
   [Switch] dhcp relay client-information record
   [Switch] interface vlan-interface 100
   [Switch–Vlan-interface100] ip address 20.20.20.1 255.255.255.0
   [Switch–Vlan-interface100] ip address 10.0.0.1 255.255.255.0 sub
   [Switch-Vlan-interface100] dhcp select relay
   [Switch-Vlan-interface100] dhcp relay server-address 192.168.0.112
   ```

   # Enable authorized ARP.

   ```
   [Switch-Vlan-interface100] arp authorized enable
   [Switch-Vlan-interface100] quit
   ```

4. Configure portal authentication:

Configure a portal authentication server.

```
[Switch] portal server newpt
[Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
[Switch-portal-server-newpt] port 50100
[Switch-portal-server-newpt] quit
```

Configure a portal Web server.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[Switch-portal-websvr-newpt] quit
```

Enable re-DHCP portal authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal enable method redhcp
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[Switch–Vlan-interface100] portal apply web-server newpt
```

Configure the BAS-IP as 20.20.20.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[Switch–Vlan-interface100] portal bas-ip 20.20.20.1
[Switch–Vlan-interface100] quit
```

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[Switch] display portal interface vlan-interface 100
 Portal information of Vlan-interface100
     NAS-ID profile: Not configured
     M-LAG group ID: N/A
     M-LAG role    : N/A
     Authorization : Strict checking
     ACL           : Disabled
     User profile  : Disabled
     Max users     : Not configured
 IPv4:
     Portal status: Enabled
     Portal authentication method: Redhcp    authentication shared mode: Disabled
     Portal web server: newpt(active)
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: 20.20.20.1
     User detection:  Not configured
     Action for server detection:
         Server type    Server name                          Action
         --             --                                   --
     Layer3 source network:
         IP address              Mask
```

```
     Destination authenticate subnet:
         IP address              Mask
IPv6:
     Portal status: Disabled
     Portal authentication method: Disabled    authentication shared mode: Disabled
     Portal web server: Not configured
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ipv6: Not configured
     User detection: Not configured
     Action for server detection:
         Server type    Server name                        Action
         --             --                                 --
     Layer3 source network:
         IP address                                        Prefix length

     Destination authenticate subnet:
         IP address                                        Prefix length
```

Before passing the authentication, a user that uses the iNode client can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page. After passing the authentication, the user can access other network resources.

After the user passes authentication, use the following command to display information about the portal user.

```
[Switch] display portal user interface vlan-interface 100
Total portal users: 1
Username: abc
  Portal server: newpt
  State: Online
  VPN instance: N/A
  MAC               IP                VLAN   Interface
  0015-e9a6-7cfe    20.20.20.2        100    Vlan-interface100
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL number: N/A
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

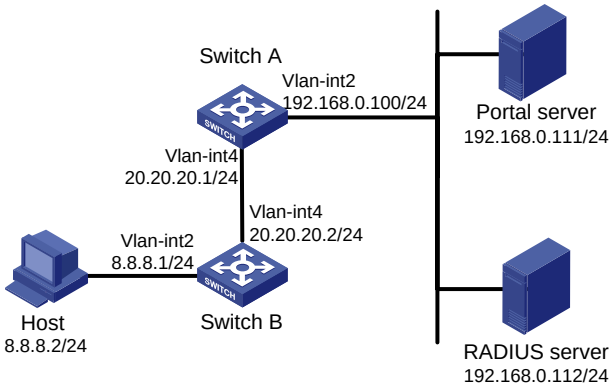
Example: Configuring cross-subnet portal authentication

Network configuration

As shown in Figure 9, Switch A supports portal authentication. The host accesses Switch A through Switch B. A portal server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure Switch A for cross-subnet portal authentication. Before passing the authentication, the host can access only the portal Web server. After passing the authentication, the user can access other network resources.

Figure 9 Network diagram



Restrictions and guidelines

Make sure the IP address of the portal device added on the portal authentication server is the IP address (20.20.20.1) of the switch's interface connecting the host. The IP address group associated with the portal device is the subnet of the host (8.8.8.0/24).

Prerequisites

- Configure IP addresses for the switch and servers as shown in Figure 9 and make sure the host, switch, and servers can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1** and enter its view.

```
<SwitchA> system-view
[SwitchA] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[SwitchA-radius-rs1] primary authentication 192.168.0.112
[SwitchA-radius-rs1] primary accounting 192.168.0.112
[SwitchA-radius-rs1] key authentication simple radius
[SwitchA-radius-rs1] key accounting simple radius
```

Exclude the ISP domain name from the username sent to the RADIUS server.

```
[SwitchA-radius-rs1] user-name-format without-domain
[SwitchA-radius-rs1] quit
```

Enable RADIUS session control.

```
[SwitchA] radius session-control enable
```

2. Configure an authentication domain:

   # Create an ISP domain named **dm1** and enter its view.

   ```
   [SwitchA] domain dm1
   ```

   # Configure AAA methods for the ISP domain.

   ```
   [SwitchA-isp-dm1] authentication portal radius-scheme rs1
   [SwitchA-isp-dm1] authorization portal radius-scheme rs1
   [SwitchA-isp-dm1] accounting portal radius-scheme rs1
   [SwitchA-isp-dm1] quit
   ```

   # Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

   ```
   [SwitchA] domain default enable dm1
   ```

3. Configure portal authentication:

   # Configure a portal authentication server.

   ```
   [SwitchA] portal server newpt
   [SwitchA-portal-server-newpt] ip 192.168.0.111 key simple portal
   [SwitchA-portal-server-newpt] port 50100
   [SwitchA-portal-server-newpt] quit
   ```

   # Configure a portal Web server.

   ```
   [SwitchA] portal web-server newpt
   [SwitchA-portal-websvr-newpt] url http://192.168.0.111:8080/portal
   [SwitchA-portal-websvr-newpt] quit
   ```

   # Enable cross-subnet portal authentication on VLAN-interface 4.

   ```
   [SwitchA] interface vlan-interface 4
   [SwitchA–Vlan-interface4] portal enable method layer3
   ```

   # Specify portal Web server **newpt** on VLAN-interface 4.

   ```
   [SwitchA–Vlan-interface4] portal apply web-server newpt
   ```

   # Configure the BAS-IP as 20.20.20.1 for portal packets sent from VLAN-interface 4 to the portal authentication server.

   ```
   [SwitchA–Vlan-interface4] portal bas-ip 20.20.20.1
   [SwitchA–Vlan-interface4] quit
   ```

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[SwitchA] display portal interface vlan-interface 4
 Portal information of Vlan-interface4
     NAS-ID profile: Not configured
     M-LAG group ID: N/A
     M-LAG role    : N/A
     Authorization : Strict checking
     ACL           : Disabled
     User profile  : Disabled
     Max users     : Not configured
 IPv4:
     Portal status: Enabled
     Portal authentication method: Layer3
     Portal web server: newpt(active)
     Secondary portal Web server: Not configured
```

```
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: 20.20.20.1
     User detection:  Not configured
     Action for server detection:
         Server type    Server name                      Action
         --             --                               --
     Layer3 source network:
         IP address              Mask

     Destination authenticate subnet:
         IP address              Mask
IPv6:
     Portal status: Disabled
     Portal authentication method: Disabled    authentication shared mode: Disabled
     Portal web server: Not configured(active)
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ipv6: Not configured
     User detection: Not configured
     Action for server detection:
         Server type    Server name                      Action
         --             --                               --
     Layer3 source network:
         IP address                                      Prefix length

     Destination authenticate subnet:
         IP address                                      Prefix length
```

A user can perform portal authentication by using the iNode client or through a Web browser. Before passing the authentication, the user can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page. After passing the authentication, the user can access other network resources.

After the user passes authentication, use the following command to display information about the portal user.

```
[SwitchA] display portal user interface vlan-interface 4
Total portal users: 1
Username: abc
  Portal server: newpt
  State: Online
  VPN instance: N/A
```

```
MAC               IP              VLAN   Interface
0000-0000-0000    8.8.8.2         4      Vlan-interface4
Authorization information:
  DHCP IP pool: N/A
  User profile: N/A
  Session group profile: N/A
  ACL number: N/A
  Inbound CAR: N/A
  Outbound CAR: N/A
  Microsegment ID: N/A
```

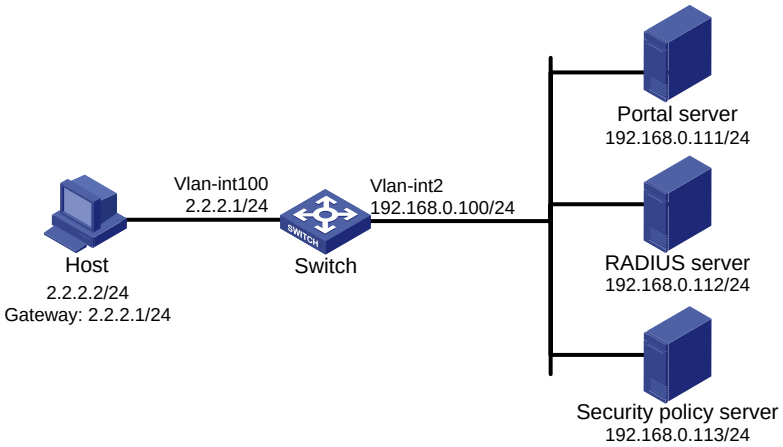
Example: Configuring extended direct portal authentication

Network configuration

As shown in Figure 10, the host is directly connected to the switch (the access device). The host is assigned a public IP address either manually or through DHCP. A portal server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure extended direct portal authentication. If the host fails security check after passing identity authentication, it can access only subnet 192.168.0.0/24. After passing security check, the host can access other network resources.

Figure 10 Network diagram



Prerequisites

- Configure IP addresses for the host, switch, and servers as shown in Figure 10 and make sure they can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a RADIUS scheme:

   # Create a RADIUS scheme named **rs1** and enter its view.

   ```
   <Switch> system-view
   [Switch] radius scheme rs1
   ```

   # Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

   ```
   [Switch-radius-rs1] primary authentication 192.168.0.112
   ```

```
[Switch-radius-rs1] primary accounting 192.168.0.112
[Switch-radius-rs1] key accounting simple radius
[Switch-radius-rs1] key authentication simple radius
[Switch-radius-rs1] user-name-format without-domain
```

Enable RADIUS session control.

```
[Switch] radius session-control enable
```

Specify a session-control client with IP address 192.168.0.112 and shared key 12345 in plaintext form.

```
[Switch] radius session-control client ip 192.168.0.112 key simple 12345
```

2. Configure an authentication domain:

Create an ISP domain named dm1 and enter its view.

```
[Switch] domain dm1
```

Configure AAA methods for the ISP domain.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure domain dm1 as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[Switch] domain default enable dm1
```

3. Configure ACL 3000 as the isolation ACL and ACL 3001 as the security ACL.

```
[Switch] acl advanced 3000
[Switch-acl-ipv4-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[Switch-acl-ipv4-adv-3000] rule deny ip
[Switch-acl-ipv4-adv-3000] quit
[Switch] acl advanced 3001
[Switch-acl-ipv4-adv-3001] rule permit ip
[Switch-acl-ipv4-adv-3001] quit
```

NOTE:

Make sure you specify ACL 3000 as the isolation ACL and ACL 3001 as the security ACL on the security policy server.

4. Configure portal authentication:

Configure a portal authentication server.

```
[Switch] portal server newpt
[Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
[Switch-portal-server-newpt] port 50100
[Switch-portal-server-newpt] quit
```

Configure a portal Web server.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[Switch-portal-websvr-newpt] quit
```

Enable direct portal authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal enable method direct
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[Switch–Vlan-interface100] portal apply web-server newpt
```

Configure the BAS-IP as 2.2.2.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[Switch–Vlan-interface100] portal bas-ip 2.2.2.1
[Switch–Vlan-interface100] quit
```

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[Switch] display portal interface vlan-interface 100
 Portal information of Vlan-interface100
     NAS-ID profile: Not configured
     M-LAG group ID: N/A
     M-LAG role    : N/A
     Authorization : Strict checking
     ACL           : Disabled
     User profile  : Disabled
     Max users     : Not configured
 IPv4:
     Portal status: Enabled
     Portal authentication method: Direct    authentication shared mode: Disabled
     Portal web server: newpt(active)
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: 2.2.2.1
     User detection:  Not configured
     Action for server detection:
         Server type    Server name                      Action
         --             --                               --
     Layer3 source network:
         IP address              Mask

     Destination authenticate subnet:
         IP address              Mask
IPv6:
     Portal status: Disabled
     Portal authentication method: Disabled    authentication shared mode: Disabled
     Portal web server: Not configured
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
```

```
Bas-ipv6: Not configured
User detection: Not configured
Action for server detection:
    Server type    Server name                          Action
    --             --                                   --
Layer3 source network:
    IP address                                          Prefix length

Destination authenticate subnet:
    IP address                                          Prefix length
```

Before passing portal authentication, a user that uses the iNode client can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page.

- The user can access the resources permitted by ACL 3000 after passing only identity authentication.
- The user can access network resources permitted by ACL 3001 after passing both identity authentication and security check.

After the user passes identity authentication and security check, use the following command to display information about the portal user.

```
[Switch] display portal user interface vlan-interface 100
Total portal users: 1
Username: abc
  Portal server: newpt
  State: Online
  VPN instance: N/A
  MAC               IP                VLAN   Interface
  0015-e9a6-7cfe    2.2.2.2           100    Vlan-interface100
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL: 3001
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

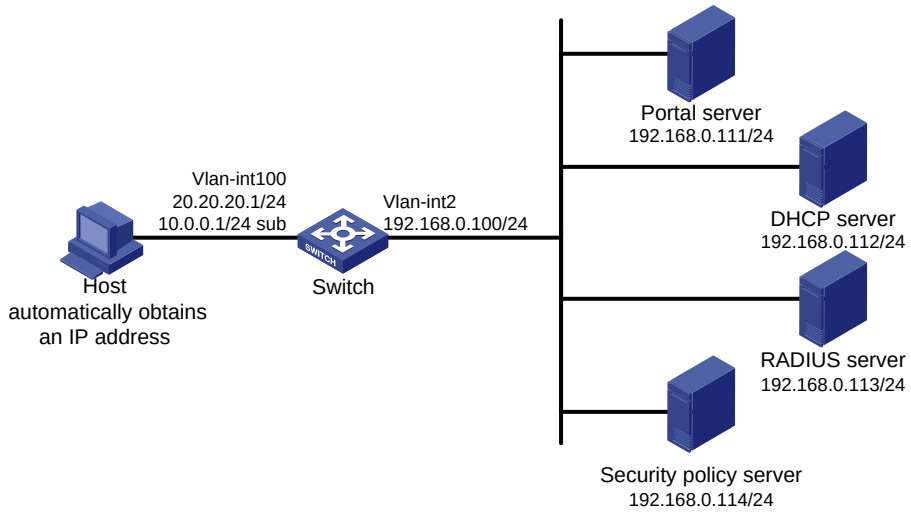
Example: Configuring extended re-DHCP portal authentication

Network configuration

As shown in Figure 11, the host is directly connected to the switch (the access device). The host obtains an IP address through the DHCP server. A portal server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure extended re-DHCP portal authentication. Before passing portal authentication, the host is assigned a private IP address. After passing portal identity authentication, the host obtains a public IP address and accepts security check. If the host fails the security check, it can access only subnet 192.168.0.0/24. After passing the security check, the host can access other network resources.

Figure 11 Network diagram



Restrictions and guidelines

- For re-DHCP portal authentication, configure a public address pool (20.20.20.0/24) and a private address pool (10.0.0.0/24) on the DHCP server. (Details not shown.)
- For re-DHCP portal authentication:
  - The switch must be configured as a DHCP relay agent.
  - The portal-enabled interface must be configured with a primary IP address (a public IP address) and a secondary IP address (a private IP address).

  For information about DHCP relay agent configuration, see *Layer 3—IP Services Configuration Guide*.
- Make sure the IP address of the portal device added on the portal server is the public IP address (20.20.20.1) of the switch's interface connecting the host. The private IP address range for the IP address group associated with the portal device is the private subnet 10.0.0.0/24 where the host resides. The public IP address range for the IP address group is the public subnet 20.20.20.0/24.

Prerequisites

- Configure IP addresses for the switch and servers as shown in Figure 11 and make sure the host, switch, and servers can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.113
[Switch-radius-rs1] primary accounting 192.168.0.113
[Switch-radius-rs1] key accounting simple radius
[Switch-radius-rs1] key authentication simple radius
[Switch-radius-rs1] user-name-format without-domain
```

Enable RADIUS session control.

```
[Switch] radius session-control enable
```

Specify a session-control client with IP address 192.168.0.113 and shared key 12345 in plaintext form.

```
[Switch] radius session-control client ip 192.168.0.113 key simple 12345
```

2. Configure an authentication domain:

Create an ISP domain named **dm1** and enter its view.

```
[Switch] domain dm1
```

Configure AAA methods for the ISP domain.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[Switch] domain default enable dm1
```

3. Configure ACL 3000 as the isolation ACL and ACL 3001 as the security ACL.

```
[Switch] acl advanced 3000
[Switch-acl-ipv4-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[Switch-acl-ipv4-adv-3000] rule deny ip
[Switch-acl-ipv4-adv-3000] quit
[Switch] acl advanced 3001
[Switch-acl-ipv4-adv-3001] rule permit ip
[Switch-acl-ipv4-adv-3001] quit
```

NOTE:

Make sure you specify ACL 3000 as the isolation ACL and ACL 3001 as the security ACL on the security policy server.

4. Configure DHCP relay and authorized ARP:

Configure DHCP relay.

```
[Switch] dhcp enable
[Switch] dhcp relay client-information record
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] ip address 20.20.20.1 255.255.255.0
[Switch–Vlan-interface100] ip address 10.0.0.1 255.255.255.0 sub
[Switch-Vlan-interface100] dhcp select relay
[Switch-Vlan-interface100] dhcp relay server-address 192.168.0.112
```

Enable authorized ARP.

```
[Switch-Vlan-interface100] arp authorized enable
[Switch-Vlan-interface100] quit
```

5. Configure portal authentication:

Configure a portal authentication server.

```
[Switch] portal server newpt
[Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
[Switch-portal-server-newpt] port 50100
[Switch-portal-server-newpt] quit
```

Configure a portal Web server.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[Switch-portal-websvr-newpt] quit
```

Enable re-DHCP portal authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal enable method redhcp
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[Switch–Vlan-interface100] portal apply web-server newpt
```

Configure the BAS-IP as 20.20.20.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[Switch–Vlan-interface100] portal bas-ip 20.20.20.1
[Switch–Vlan-interface100] quit
```

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[Switch] display portal interface vlan-interface 100
 Portal information of Vlan-interface100
     NAS-ID profile: Not configured
     M-LAG group ID: N/A
     M-LAG role    : N/A
     Authorization : Strict checking
     ACL           : Disabled
     User profile  : Disabled
     Max users     : Not configured
 IPv4:
     Portal status: Enabled
     Portal authentication method: Redhcp    authentication shared mode: Disabled
     Portal web server: newpt(active)
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: 20.20.20.1
     User detection:  Not configured
     Action for server detection:
         Server type    Server name                          Action
         --             --                                   --
     Layer3 source network:
         IP address              Mask

     Destination authenticate subnet:
         IP address              Mask
IPv6:
     Portal status: Disabled
     Portal authentication method: Disabled    authentication shared mode: Disabled
     Portal web server: Not configured
```

```
    Secondary portal Web server: Not configured
    Portal mac-trigger-server: Not configured
    Authentication domain: Not configured
    Pre-auth domain: Not configured
    User-dhcp-only: Disabled
    Pre-auth IP pool: Not configured
    Max users: Not configured
    Bas-ipv6: Not configured
    User detection: Not configured
    Action for server detection:
        Server type    Server name                     Action
        --             --                              --
    Layer3 source network:
        IP address                                     Prefix length


    Destination authenticate subnet:
        IP address                                     Prefix length
```

Before passing portal authentication, a user that uses the iNode client can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page.

- The user can access the resources permitted by ACL 3000 after passing only identity authentication.
- The user can access network resources permitted by ACL 3001 after passing both identity authentication and security check.

After the user passes identity authentication and security check, use the following command to display information about the portal user.

```
[Switch] display portal user interface vlan-interface 100
Total portal users: 1
Username: abc
  Portal server: newpt
  State: Online
  VPN instance: N/A
  MAC               IP                VLAN   Interface
  0015-e9a6-7cfe    20.20.20.2        100    Vlan-interface100
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL number: 3001 (active)
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

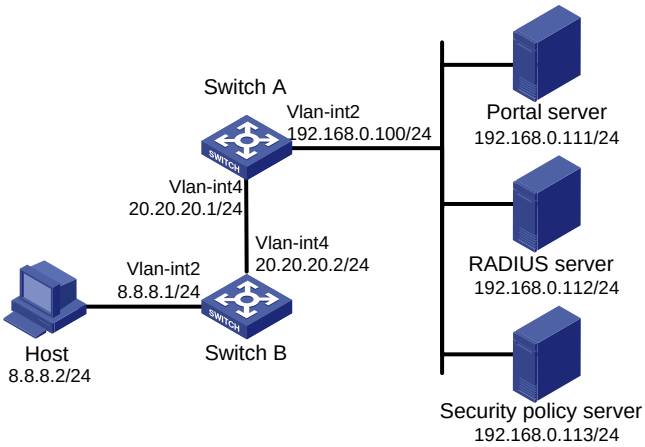
Example: Configuring extended cross-subnet portal authentication

Network configuration

As shown in Figure 12, Switch A supports portal authentication. The host accesses Switch A through Switch B. A portal server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure Switch A for extended cross-subnet portal authentication. Before passing portal authentication, the host can access only the portal server. After passing portal identity authentication, the host accepts security check. If the host fails the security check it can access only the subnet 192.168.0.0/24. After passing the security check, the host can access other network resources.

Figure 12 Network diagram



Restrictions and guidelines

Make sure the IP address of the portal device added on the portal server is the IP address (20.20.20.1) of the switch's interface connecting the host. The IP address group associated with the portal device is the subnet of the host (8.8.8.0/24).

Prerequisites

- Configure IP addresses for the switch and servers as shown in Figure 12 and make sure the host, switch, and servers can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1** and enter its view.

```
<SwitchA> system-view
[SwitchA] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[SwitchA-radius-rs1] primary authentication 192.168.0.112
[SwitchA-radius-rs1] primary accounting 192.168.0.112
[SwitchA-radius-rs1] key accounting simple radius
[SwitchA-radius-rs1] key authentication simple radius
[SwitchA-radius-rs1] user-name-format without-domain
```

Enable RADIUS session control.

```
[SwitchA] radius session-control enable
```

Specify a session-control client with IP address 192.168.0.112 and shared key 12345 in plaintext form.

```
[SwitchA] radius session-control client ip 192.168.0.112 key simple 12345
```

2. Configure an authentication domain:

Create an ISP domain named **dm1** and enter its view.

```
[SwitchA] domain dm1
```

Configure AAA methods for the ISP domain.

```
[SwitchA-isp-dm1] authentication portal radius-scheme rs1
[SwitchA-isp-dm1] authorization portal radius-scheme rs1
[SwitchA-isp-dm1] accounting portal radius-scheme rs1
[SwitchA-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[SwitchA] domain default enable dm1
```

3. Configure ACL 3000 as the isolation ACL and ACL 3001 as the security ACL.

```
[SwitchA] acl advanced 3000
[SwitchA-acl-ipv4-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[SwitchA-acl-ipv4-adv-3000] rule deny ip
[SwitchA-acl-ipv4-adv-3000] quit
[SwitchA] acl advanced 3001
[SwitchA-acl-ipv4-adv-3001] rule permit ip
[SwitchA-acl-ipv4-adv-3001] quit
```

NOTE:

Make sure you specify ACL 3000 as the isolation ACL and ACL 3001 as the security ACL on the security policy server.

4. Configure portal authentication:

Configure a portal authentication server.

```
[SwitchA] portal server newpt
[SwitchA-portal-server-newpt] ip 192.168.0.111 key simple portal
[SwitchA-portal-server-newpt] port 50100
[SwitchA-portal-server-newpt] quit
```

Configure a portal Web server.

```
[SwitchA] portal web-server newpt
[SwitchA-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[SwitchA-portal-websvr-newpt] quit
```

Enable cross-subnet portal authentication on VLAN-interface 4.

```
[SwitchA] interface vlan-interface 4
[SwitchA–Vlan-interface4] portal enable method layer3
```

Specify portal Web server **newpt** on VLAN-interface 4.

```
[SwitchA–Vlan-interface4] portal apply web-server newpt
```

Configure the BAS-IP as 20.20.20.1 for portal packets sent from VLAN-interface 4 to the portal authentication server.

```
[SwitchA–Vlan-interface4] portal bas-ip 20.20.20.1
[SwitchA–Vlan-interface4] quit
```

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[SwitchA] display portal interface vlan-interface 4
 Portal information of Vlan-interface4
     NAS-ID profile: Not configured
     M-LAG group ID: N/A
     M-LAG role    : N/A
     Authorization : Strict checking
     ACL           : Disabled
     User profile  : Disabled
     Max users     : Not configured
 IPv4:
     Portal status: Enabled
     Portal authentication method: Layer3
     Portal web server: newpt(active)
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: 20.20.20.1
     User detection:  Not configured
     Action for server detection:
         Server type    Server name                        Action
         --             --                                 --
     Layer3 source network:
         IP address              Mask

     Destination authenticate subnet:
         IP address              Mask
IPv6:
     Portal status: Disabled
     Portal authentication method: Disabled    authentication shared mode: Disabled
     Portal web server: Not configured(active)
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: Not configured
     User detection: Not configured
     Action for server detection:
         Server type    Server name                        Action
         --             --                                 --
```

```
Layer3 source network:
    IP address                                      Prefix length

Destination authenticate subnet:
    IP address                                      Prefix length
```

Before passing portal authentication, a user that uses the iNode client can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page.

- The user can access the resources permitted by ACL 3000 after passing only identity authentication.
- The user can access network resources permitted by ACL 3001 after passing both identity authentication and security check.

After the user passes identity authentication and security check, use the following command to display information about the portal user.

```
[SwitchA] display portal user interface vlan-interface 4
Total portal users: 1
Username: abc
  Portal server: newpt
  State: Online
  VPN instance: N/A
  MAC               IP                VLAN   Interface
  0015-e9a6-7cfe    8.8.8.2           4    Vlan-interface4
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL number: 3001 (active)
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

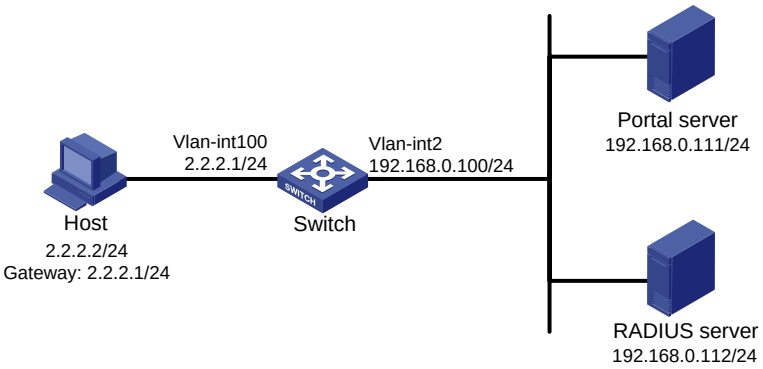
Example: Configuring portal server detection and portal user synchronization

Network configuration

As shown in Figure 13, the host is directly connected to the switch (the access device). The host is assigned a public IP address either manually or through DHCP. An IMC server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

- Configure direct portal authentication on the switch, so the host can access only the portal server before passing the authentication and access other network resources after passing the authentication.
- Configure the switch to detect the reachability state of the portal authentication server, send log messages upon state changes, and disable portal authentication when the authentication server is unreachable.
- Configure the switch to synchronize portal user information with the portal server periodically.

Figure 13 Network diagram



Prerequisites

- Configure IP addresses for the switch and servers and make sure the host, switch, and servers can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.
- Configure the portal service on the IMC server, including setting portal server parameters, adding access devices, creating user IP address groups, and configuring device port groups to associate with IP address groups. This ensures coordinated implementation of portal authentication with the device.

Configuring the switch

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.112
[Switch-radius-rs1] primary accounting 192.168.0.112
[Switch-radius-rs1] key authentication simple radius
[Switch-radius-rs1] key accounting simple radius
```

Exclude the ISP domain name from the username sent to the RADIUS server.

```
[Switch-radius-rs1] user-name-format without-domain
[Switch-radius-rs1] quit
```

Enable RADIUS session control.

```
[Switch] radius session-control enable
```

2. Configure an authentication domain:

Create an ISP domain named **dm1** and enter its view.

```
[Switch] domain dm1
```

Configure AAA methods for the ISP domain.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[Switch] domain default enable dm1
```

3. Configure portal authentication:

   # Configure a portal authentication server.

   ```
   [Switch] portal server newpt
   [Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
   [Switch-portal-server-newpt] port 50100
   ```

   # Configure reachability detection of the portal authentication server: set the server detection interval to 40 seconds, and send log messages upon reachability status changes.

   ```
   [Switch-portal-server-newpt] server-detect timeout 40 log
   ```

   NOTE:

   The value of **timeout** must be greater than or equal to the portal server heartbeat interval.

   # Configure portal user synchronization with the portal authentication server, and set the synchronization detection interval to 600 seconds.

   ```
   [Switch-portal-server-newpt] user-sync timeout 600
   [Switch-portal-server-newpt] quit
   ```

   NOTE:

   The value of **timeout** must be greater than or equal to the portal user heartbeat interval.

   # Configure a portal Web server.

   ```
   [Switch] portal web-server newpt
   [Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
   [Switch-portal-websvr-newpt] quit
   ```

   # Enable direct portal authentication on VLAN-interface 100.

   ```
   [Switch] interface vlan-interface 100
   [Switch–Vlan-interface100] portal enable method direct
   ```

   # Enable portal fail-permit for the portal authentication server **newpt**.

   ```
   [Switch–Vlan-interface100] portal fail-permit server newpt
   ```

   # Specify portal Web server **newpt** on VLAN-interface 100.

   ```
   [Switch–Vlan-interface100] portal apply web-server newpt
   ```

   # Configure the BAS-IP as 2.2.2.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

   ```
   [Switch–Vlan-interface100] portal bas-ip 2.2.2.1
   [Switch–Vlan-interface100] quit
   ```

Verifying the configuration

Use the following command to display information about the portal authentication server.

```
[Switch] display portal server newpt
Portal server: newpt
  Type                  : IMC
  IP                    : 192.168.0.111
  VPN instance          : Not configured
  Port                  : 50100
  Server Detection      : Timeout 40s  Action: log
  User synchronization  : Timeout 600s
  Status                : Up
```

The **Up** status of the portal authentication server indicates that the portal authentication server is reachable. If the access device detects that the portal authentication server is unreachable, the **Status** field in the command output displays **Down**. The access device generates a server unreachable log "Portal server newpt turns down from up." and disables portal authentication on the access interface, so the host can access the external network without authentication.

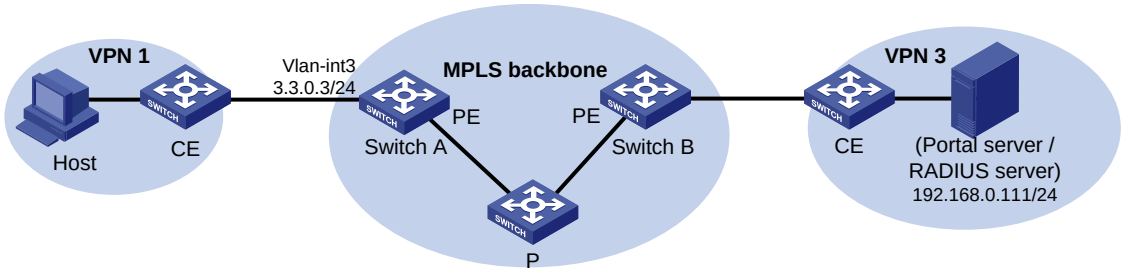
Example: Configuring cross-subnet portal authentication for MPLS L3VPNs

Network configuration

As shown in Figure 14, the PE device Switch A provides portal authentication for the host in VPN 1. A portal server in VPN 3 acts as the portal authentication server, portal Web server, and RADIUS server.

Configure cross-subnet portal authentication on Switch A, so the host can access network resources after passing identity authentication.

Figure 14 Network diagram



Prerequisites

- Before enabling portal authentication, configure MPLS L3VPN and specify VPN targets for VPN 1 and VPN 3 so that VPN 1 and VPN 3 can communicate with each other. This example describes only the access authentication configuration on the user-side PE. For information about MPLS L3VPN configurations, see MPLS L3VPN configuration in *MCE Configuration Guide*.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a RADIUS scheme:

   # Create a RADIUS scheme named **rs1** and enter its view.

   ```
   <SwitchA> system-view
   [SwitchA] radius scheme rs1
   ```

   # For the RADIUS scheme, specify the VPN instance that is bound to the interface connected to the portal/RADIUS server. This example uses VPN instance **vpn3**. (For information about the VPN instance, see the MPLS L3VPN configuration on Switch A.)

   ```
   [SwitchA-radius-rs1] vpn-instance vpn3
   ```

   # Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

   ```
   [SwitchA-radius-rs1] primary authentication 192.168.0.111
   [SwitchA-radius-rs1] primary accounting 192.168.0.111
   [SwitchA-radius-rs1] key accounting simple radius
   [SwitchA-radius-rs1] key authentication simple radius
   ```

   # Exclude the ISP domain name from the username sent to the RADIUS server.

```
[SwitchA-radius-rs1] user-name-format without-domain
```

Specify the source IP address for RADIUS packets to be sent as 3.3.0.3. This address must be the same as that of the portal device specified on the portal authentication server to avoid authentication failures.

```
[SwitchA-radius-rs1] nas-ip 3.3.0.3
[SwitchA-radius-rs1] quit
```

Enable RADIUS session control.

```
[SwitchA] radius session-control enable
```

2. Configure an authentication domain:

Create an ISP domain named **dm1** and enter its view.

```
[SwitchA] domain dm1
```

Configure AAA methods for the ISP domain.

```
[SwitchA-isp-dm1] authentication portal radius-scheme rs1
[SwitchA-isp-dm1] authorization portal radius-scheme rs1
[SwitchA-isp-dm1] accounting portal radius-scheme rs1
[SwitchA-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[SwitchA] domain default enable dm1
```

3. Configure portal authentication:

Configure a portal authentication server.

```
[SwitchA] portal server newpt
[SwitchA-portal-server-newpt] ip 192.168.0.111 vpn-instance vpn3 key simple portal
[SwitchA-portal-server-newpt] port 50100
[SwitchA-portal-server-newpt] quit
```

Configure a portal Web server.

```
[SwitchA] portal web-server newpt
[SwitchA-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[SwitchA-portal-websvr-newpt] vpn-instance vpn3
[SwitchA-portal-websvr-newpt] quit
```

Enable cross-subnet portal authentication on VLAN-interface 3.

```
[SwitchA] interface vlan-interface 3
[SwitchA–Vlan-interface3] portal enable method layer3
```

Specify portal Web server **newpt** on VLAN-interface 3.

```
[SwitchA–Vlan-interface3] portal apply web-server newpt
```

Configure the BAS-IP as 3.3.0.3 for portal packets sent from VLAN-interface 3 to the portal authentication server.

```
[SwitchA–Vlan-interface3] portal bas-ip 3.3.0.3
[SwitchA–Vlan-interface3] quit
```

Verifying the configuration

Verify the portal configuration by executing the **`display portal interface`** command. (Details not shown.)

After the user passes authentication, execute the **`display portal user`** command to display the portal user information.

```
[SwitchA] display portal user all
Total portal users: 1
```

```
Username: abc
  Portal server: newpt
  State: Online
  VPN instance: vpn3
  MAC               IP                VLAN   Interface
  000d-88f7-c268    3.3.0.1           3      Vlan-interface3
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL number: N/A
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

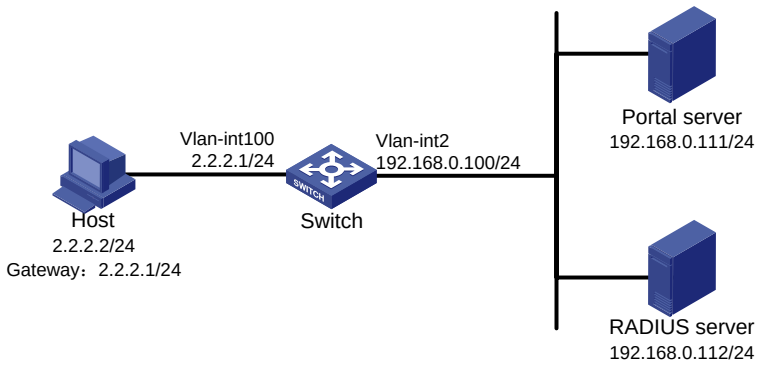
Example: Configuring direct portal authentication with a preauthentication domain

Network configuration

As shown in Figure 15, the host is directly connected to the switch (the access device). The host is assigned a public IP address through DHCP. A portal server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure direct portal authentication, so the host can access only subnet 192.168.0.0/24 before passing the authentication and access other network resources after passing the authentication.

Figure 15 Network diagram



Prerequisites

- Configure IP addresses for the host, switch, and servers as shown in Figure 15 and make sure they can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a preauthentication IP address pool:

   # Configure DHCP address pool **pre** to assign IP addresses and other configuration parameters to clients on subnet 2.2.2.0/24.

   ```
   <Switch> system-view
   [Switch] dhcp server ip-pool pre
   ```

```
[Switch-dhcp-pool-pre] gateway-list 2.2.2.1
[Switch-dhcp-pool-pre] network 2.2.2.0 24
[Switch-dhcp-pool-pre] quit
```

Enable the DHCP server on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] dhcp select server
[Switch–Vlan-interface100] quit
```

2. Configure a portal preauthentication domain:

Create an ISP domain named **abc**.

```
[Switch] domain abc
```

Specify user attribute ACL 3010 in the ISP domain.

```
[Switch-isp-abc] authorization-attribute acl 3010
[Switch-isp-abc] quit
```

In ACL 3010, configure a rule to permit access to the subnet 192.168.0.0/24.

```
[Switch] acl advanced 3010
[Switch-acl-ipv4-adv-3010] rule 1 permit ip destination 192.168.0.0 0.0.0.255
[Switch-acl-ipv4-adv-3010] quit
```

Specify ISP domain **abc** as the portal authentication domain on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal pre-auth domain abc
[Switch–Vlan-interface100] quit
```

3. Configure portal authentication:

Configure a portal authentication server.

```
[Switch] portal server newpt
[Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
[Switch-portal-server-newpt] port 50100
[Switch-portal-server-newpt] quit
```

Configure a portal Web server.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[Switch-portal-websvr-newpt] quit
```

Enable direct portal authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal enable method direct
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[Switch–Vlan-interface100] portal apply web-server newpt
```

Configure the BAS-IP as 2.2.2.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[Switch–Vlan-interface100] portal bas-ip 2.2.2.1
[Switch–Vlan-interface100] quit
```

Verifying the configuration

Verify the portal configuration by executing the **`display portal interface`** command. (Details not shown.)

Display information about preauthentication portal users.

```
[Switch] display portal user pre-auth interface vlan-interface 100
MAC               IP                VLAN   Interface
0015-e9a6-7cfe    10.10.10.4        100     Vlan-interface100
```

```
State: Online
VPN instance: N/A
Authorization information:
  User profile: N/A
  Session group profile: N/A
  ACL number: 3010 (active)
  Inbound CAR: N/A
  Outbound CAR: N/A
  Microsegment ID: N/A
```

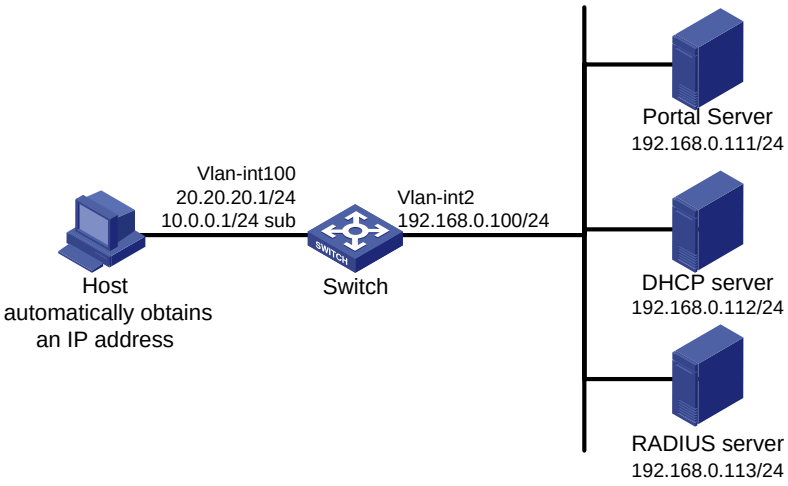
Example: Configuring re-DHCP portal authentication with a preauthentication domain

Network configuration

As shown in Figure 16, the host is directly connected to the switch (the access device). The host obtains an IP address through the DHCP server. A portal server acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure re-DHCP portal authentication. Before passing the authentication, the host is assigned a private IP address and can access only the subnet 192.168.0.0/24. After passing the authentication, the host gets a public IP address and can access other network resources.

Figure 16 Network diagram



Restrictions and guidelines

- For re-DHCP portal authentication, configure a public address pool (20.20.20.0/24) and a private address pool (10.0.0.0/24) on the DHCP server. (Details not shown.)
- For re-DHCP portal authentication:
  - The switch must be configured as a DHCP relay agent.
  - The portal-enabled interface must be configured with a primary IP address (a public IP address) and a secondary IP address (a private IP address).

  For information about DHCP relay agent configuration, see *Layer 3—IP Services Configuration Guide*.
- Make sure the IP address of the portal device added on the portal server is the public IP address (20.20.20.1) of the switch's interface connecting the host. The private IP address range for the IP address group associated with the portal device is the private subnet 10.0.0.0/24

where the host resides. The public IP address range for the IP address group is the public subnet 20.20.20.0/24.

- If you have configured a preauthentication IP address pool on portal-enabled interfaces, configure a DHCP relay address pool with the same name on the device. For the DHCP relay address pool, specify the subnet address where the unauthenticated users reside (with the **`export-router`** keyword specified) and the DHCP server address.

Prerequisites

- Configure IP addresses for the switch and servers as shown in Figure 16 and make sure the host, switch, and servers can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a portal preauthentication domain:

Create an ISP domain named **abc**.

```
<Switch> system-view
[Switch] domain abc
```

Specify user attribute ACL 3010 in the ISP domain.

```
[Switch-isp-abc] authorization-attribute acl 3010
[Switch-isp-abc] quit
```

In ACL 3010, configure a rule to permit access to the subnet 192.168.0.0/24.

```
[Switch] acl advanced 3010
[Switch-acl-ipv4-adv-3010] rule 1 permit ip destination 192.168.0.0 0.0.0.255
[Switch-acl-ipv4-adv-3010] quit
```

Specify ISP domain **abc** as the portal authentication domain on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal pre-auth domain abc
[Switch–Vlan-interface100] quit
```

2. Configure DHCP relay and authorized ARP.

Configure DHCP relay.

```
[Switch] dhcp enable
[Switch] dhcp relay client-information record
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] ip address 20.20.20.1 255.255.255.0
[Switch–Vlan-interface100] ip address 10.0.0.1 255.255.255.0 sub
[Switch-Vlan-interface100] dhcp select relay
[Switch-Vlan-interface100] dhcp relay server-address 192.168.0.112
```

Enable authorized ARP.

```
[Switch-Vlan-interface100] arp authorized enable
[Switch-Vlan-interface100] quit
```

3. Configure portal authentication:

Configure a portal authentication server.

```
[Switch] portal server newpt
[Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
[Switch-portal-server-newpt] port 50100
[Switch-portal-server-newpt] quit
```

Configure a portal Web server.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
```

```
[Switch-portal-websvr-newpt] quit
```

Enable re-DHCP portal authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch-Vlan-interface100] portal enable method redhcp
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[Switch–Vlan-interface100] portal apply web-server newpt
```

Configure the BAS-IP as 20.20.20.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[Switch–Vlan-interface100] portal bas-ip 20.20.20.1
[Switch–Vlan-interface100] quit
```

Verifying the configuration

Verify the portal configuration by executing the **`display portal interface`** command. (Details not shown.)

Display information about preauthentication portal users.

```
[Switch] display portal user pre-auth interface vlan-interface 100
MAC                IP                 VLAN   Interface
0015-e9a6-7cfe     10.10.10.4         100     Vlan-interface100
  State: Online
  VPN instance: N/A
  Authorization information:
    User profile: N/A
    Session group profile: N/A
    ACL number: 3010 (active)
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

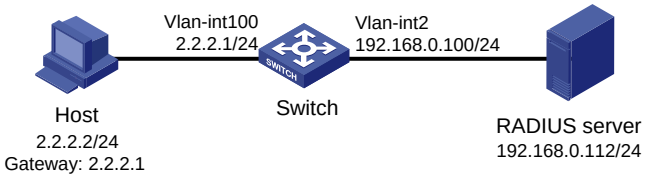
Example: Configuring direct portal authentication using a local portal Web service

Network configuration

As shown in Figure 17, the host is directly connected to the switch (the access device). The host is assigned a public IP address either manually or through DHCP. The switch acts as both a portal authentication server and a portal Web server. A RADIUS server acts as the authentication and accounting server.

Configure direct portal authentication on the switch. Before a user passes portal authentication, the user can access only the portal Web server. After passing portal authentication, the user can access other network resources.

Figure 17 Network diagram



Prerequisites

- Configure IP addresses for the host, switch, and server as shown in Figure 17 and make sure they can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.

Procedure

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.112
[Switch-radius-rs1] primary accounting 192.168.0.112
[Switch-radius-rs1] key authentication simple radius
[Switch-radius-rs1] key accounting simple radius
```

Exclude the ISP domain name from the username sent to the RADIUS server.

```
[Switch-radius-rs1] user-name-format without-domain
[Switch-radius-rs1] quit
```

Enable RADIUS session control.

```
[Switch] radius session-control enable
```

2. Configure an authentication domain:

Create an ISP domain named **dm1** and enter its view.

```
[Switch] domain dm1
```

Configure AAA methods for the ISP domain.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[Switch] domain default enable dm1
```

3. Configure portal authentication:

Configure a portal Web server named **newpt** and specify **http://2.2.2.1:2331/portal** as the URL of the portal Web server. The IP address in the URL must be the IP address of a Layer 3 interface reachable to portal clients or a loopback interface (except 127.0.0.1) on the device.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://2.2.2.1:2331/portal
[Switch-portal-websvr-newpt] quit
```

Enable direct portal authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] portal enable method direct
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[Switch–Vlan-interface100] portal apply web-server newpt
[Switch–Vlan-interface100] quit
```

Create an HTTP-based local portal Web service and enter its view.

```
[Switch] portal local-web-server http
```

Specify file **defaultfile.zip** as the default authentication page file for the local portal Web service. (Make sure the file exists under the root directory of the switch.)

```
[Switch–portal-local-websvr-http] default-logon-page defaultfile.zip
```

Set the HTTP listening port number to 2331 for the local portal Web service.

```
[Switch–portal-local-webserver-http] tcp-port 2331
[Switch–portal-local-websvr-http] quit
```

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[Switch] display portal interface vlan-interface 100
 Portal information of Vlan-interface 100
     M-LAG group ID: N/A
     M-LAG role    : N/A
     Authorization                   Strict checking
     ACL                             Disabled
     User profile                    Disabled
 IPv4:
     Portal status: Enabled
     Portal authentication method: Direct    authentication shared mode: Disabled
     Portal web server: newpt(active)
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ip: Not configured
     User detection:  Not configured
     Action for server detection:
         Server type    Server name                          Action
         --             --                                   --
     Layer3 source network:
         IP address              Mask

     Destination authenticate subnet:
         IP address              Mask
IPv6:
     Portal status: Disabled
     Portal authentication method: Disabled    authentication shared mode: Disabled
     Portal web server: Not configured
     Secondary portal Web server: Not configured
     Portal mac-trigger-server: Not configured
     Authentication domain: Not configured
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
     Max users: Not configured
     Bas-ipv6: Not configured
```

```
    User detection: Not configured
    Action for server detection:
        Server type    Server name                          Action
        --             --                                   --
    Layer3 source network:
        IP address                                          Prefix length

    Destination authenticate subnet:
        IP address                                          Prefix length
```

A user can perform portal authentication through a Web page. Before passing the authentication, the user can access only the authentication page **http://2.2.2.1:2331/portal** and all Web requests will be redirected to the authentication page. After passing the authentication, the user can access other network resources.

After the user passes authentication, use the following command to display information about the portal user.

```
[Switch] display portal user interface vlan-interface 100
Total portal users: 1
Username: abc
  Portal server: newpt
  State: Online
  VPN instance: N/A
  MAC               IP                VLAN   Interface
  0015-e9a6-7cfe    2.2.2.2           100     Vlan-interface100
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL number: N/A
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

Example: Configuring MAC-based quick portal authentication

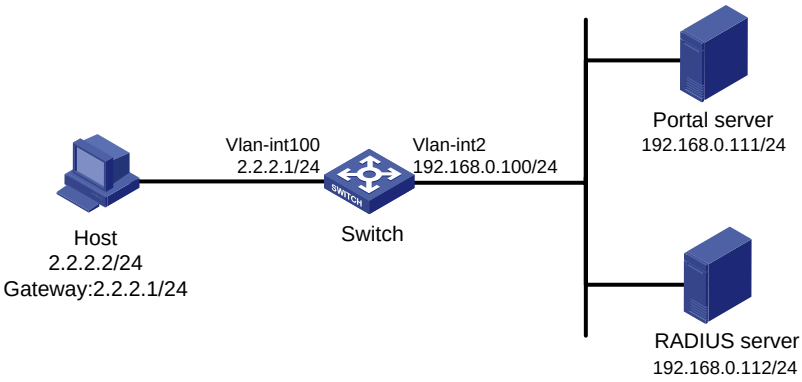
Network configuration

As shown in Figure 18, the host accesses the network through the switch. The host is assigned a public IP address either manually or through DHCP.

An IMC server acts as a portal server (portal authentication server and portal Web server) and a MAC binding server. A RADIUS server acts as the authentication and accounting server.

Configure direct portal authentication, so the host can access only the portal Web server before passing the authentication and can access other network resources after passing the authentication.

Figure 18 Network diagram



Prerequisites

- Configure IP addresses for the host, switch, and servers and make sure they can reach each other.
- Configure the RADIUS server correctly to provide authentication and accounting functions.
- Configure the portal service on the IMC server, including setting portal server parameters, adding access devices, creating user IP address groups, and configuring device port groups to associate with IP address groups. This ensures coordinated implementation of portal authentication with the device.
- Configure the MAC binding server on the IMC server, including adding access policies, configuring access services and access users, setting system parameters (such as MAC-based quick portal authentication and terminal aging policy), and manually activate these parameters to apply them to the system. This helps achieve MAC address binding and MAC-based quick portal authentication for terminals.

Configuring the switch

1. Configure a RADIUS scheme:

   # Create a RADIUS scheme named **rs1** and enter its view.

   ```
   <Switch> system-view
   [Switch] radius scheme rs1
   ```

   # Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

   ```
   [Switch-radius-rs1] primary authentication 192.168.0.112
   [Switch-radius-rs1] primary accounting 192.168.0.112
   [Switch-radius-rs1] key authentication simple radius
   [Switch-radius-rs1] key accounting simple radius
   ```

   # Exclude the ISP domain name from the username sent to the RADIUS server.

   ```
   [Switch-radius-rs1] user-name-format without-domain
   [Switch-radius-rs1] quit
   ```

   # Enable RADIUS session control.

   ```
   [Switch] radius session-control enable
   ```

2. Configure an authentication domain:

   # Create an ISP domain named **dm1** and enter its view.

   ```
   [Switch] domain dm1
   ```

   # Configure AAA methods for the ISP domain.

   ```
   [Switch-isp-dm1] authentication portal radius-scheme rs1
   [Switch-isp-dm1] authorization portal radius-scheme rs1
   ```

```
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[Switch] domain default enable dm1
```

3. Configure portal authentication:

Configure a portal authentication server.

```
[Switch] portal server newpt
[Switch-portal-server-newpt] ip 192.168.0.111 key simple portal
[Switch-portal-server-newpt] port 50100
[Switch-portal-server-newpt] quit
```

Configure a portal Web server.

```
[Switch] portal web-server newpt
[Switch-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[Switch-portal-websvr-newpt] quit
```

Enable direct authentication on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch-Vlan-interface100] portal enable method direct
```

Specify the portal Web server **newpt** on VLAN-interface 100.

```
[Switch-Vlan-interface100] portal apply web-server newpt
```

Configure the BAS-IP as 2.2.2.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[Switch-Vlan-interface100] portal bas-ip 2.2.2.1
[Switch-Vlan-interface100] quit
```

4. Configure MAC-based quick portal authentication:

Create the MAC binding server **mts**.

```
[Switch] portal mac-trigger-server mts
```

Set the free-traffic threshold for portal users to 1024000 bytes.

```
[Switch-portal-mac-trigger-server-mts] free-traffic threshold 1024000
```

Specify the IP address of the MAC binding server as 192.168.0.111.

```
[Switch-portal-mac-trigger-server-mts] ip 192.168.0.111
[Switch-portal-mac-trigger-server-mts] quit
```

Specify the MAC binding server **mts** on VLAN-interface 100.

```
[Switch] interface vlan-interface 100
[Switch-Vlan-interface100] portal apply mac-trigger-server mts
[Switch-Vlan-interface100] quit
```

Verifying the configuration

Display information about the MAC binding server.

```
[Switch] display portal mac-trigger-server name mts
Portal mac trigger server name: mts
  Version                    : 1.0
  Server type                : IMC
  IP                         : 192.168.0.111
  Port                       : 50100
  VPN instance               : Not configured
  Aging time                 : 300 seconds
```

```
  Free-traffic threshold     : 1024000 bytes
  NAS-Port-Type              : Not configured
  Binding retry times        : 3
  Binding retry interval     : 1 seconds
  Authentication timeout     : 3 minutes
```

A user can perform portal authentication by using the iNode client or through a Web browser. Before passing the authentication, the user can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page. After passing the authentication, the user can access other network resources.

For the first portal authentication, the user is required to enter the username and password. When the user goes offline and then accesses the network again, the user does not need to enter the authentication username and password.

Display portal user information.

```
[Switch] display portal user interface vlan-interface 100
Total portal users: 1
Username: Client1
  Portal server: newpt
  State: Online
  VPN instance: N/A
  MAC               IP                VLAN   Interface
  0015-e9a6-7cfe    2.2.2.2           100    Vlan-interface100
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL number: N/A
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

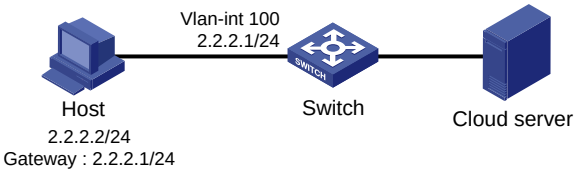
Example: Configuring cloud MAC-trigger authentication

Network configuration

As shown in Figure 19, the host accesses the network through the switch. The host is assigned a public IP address either manually or through DHCP. A cloud server acts as a portal server (portal authentication server and portal Web server) and a MAC binding server.

Configure MAC-trigger authenticaiton (MAC-based quick portal authentication) that uses the cloud server, so the user can access only the portal Web server before passing the authentication and can access other network resources after passing the authentication. The user can pass portal authentication without entering a username or password when the user tries to come online again.

Figure 19 Network diagram



Configuring the cloud server

On the Oasis platform, enable **Auth free for reconnect** on the authentication template for the AC. (Details not shown.)

Configuring the switch

1. Configure IP addresses for the host, switch, and server and make sure they can reach each other. (Details not shown.)
2. Configure basic network functions:

   # Create VLAN 100. The host will use this VLAN to access the network.

   ```
   <Switch> system-view
   [Switch] vlan 100
   [Switch-vlan100] quit
   ```

   # Assign an IP address to the VLAN interface. (Details not shown.)

   # Enable DNS proxy.

   ```
   [Switch] dns proxy enable
   ```

3. Configure an authentication domain:

   # Create an ISP domain named **cloud** and enter its view.

   ```
   [Switch] domain cloud
   ```

   # Configure AAA methods for portal users in the ISP domain as none.

   ```
   [Switch-isp-cloud] authentication portal none
   [Switch-isp-cloud] authorization portal none
   [Switch-isp-cloud] accounting portal none
   [Switch-isp-cloud] quit
   ```

4. Configure cloud MAC-trigger authentication:

   # Configure a portal Web server named **wbs**.

   ```
   [Switch] portal web-server wbs
   [Switch-portal-websvr-wbs] url http://oasisauth.h3c.com
   [Switch-portal-websvr-wbs] server-type oauth
   [Switch-portal-websvr-wbs] url-parameter nas_id value
   cm-0-554212-210235A2G1B205000009
   [Switch-portal-websvr-wbs] url-parameter template_id value 4291
   ```

   # Create an HTTP-based local portal Web service and enter its view.

   ```
   [Switch] portal local-web-server http
   ```

   # Enable cloud MAC-trigger authentication.

   ```
   [Switch] portal mac-trigger-server abc
   [Switch-portal-extend-auth-server-abc] cloud-binding enable
   ```

   # Specify **https://oasisauth.h3c.com** as the URL of the cloud portal authentication server.

   ```
   [Switch-portal-extend-auth-server-abc] cloud-server url http://oasisauth.h3c.com
   [Switch-portal-extend-auth-server-abc] quit
   ```

   # Enable direct portal authentication on interface VLAN-interface 100.

   ```
   [Switch] interface vlan-interface 100
   [Switch-Vlan-interface100] portal enable method direct
   ```

   # Specify the authentication domain, portal Web server, and MAC binding server on interface VLAN-interface 100.

   ```
   [Switch-Vlan-interface100] portal domain cloud
   [Switch-Vlan-interface100] portal apply web-server cloud
   [Switch-Vlan-interface100] portal apply mac-trigger-server abc
   ```

Verifying the configuration

Display packet statistics for the cloud portal authentication server.

```
[Switch] display portal packet statistics extend-auth-server cloud
Extend-auth server : cloud
  Pkt-Type           Success      Error      Timeout      Conn-failure
  REQ_ACCESSTOKEN    1            0          0            0
  REQ_USERINFO       1            0          0            0
  RESP_ACCESSTOKEN   1            0          0            0
  RESP_USERINFO      1            0          0            0
  POST_ONLINEDATA    10           0          0            0
  RESP_ONLINEDATA    10           0          0            0
  POST_OFFLINEUSER   1            0          0            0
  REPORT_ONLINEUSER  2            0          0            0
  REQ_CLOUDBIND      2            0          0            0
  ESP_CLOUDBIND      2            0          0            0
  REQ_BINDUSERINFO   1            0          0            0
  RESP_BINDUSERINFO  1            0          0            0
  AUTHENTICATION     2            0          0            0
```

Before passing portal authentication, a user can access only the portal authentication page **http://oasisauth.h3c.com**. All Web requests from the user will be redirected to the portal authentication page. After passing portal authentication, the user can access other network resources.

The user needs to enter the username and password for the first authentication. If the user goes offline and then tries to come online, the user can directly access the network resources without entering the username and password.

Display information about portal users.

```
[Switch] display portal user interface vlan-interface 100
Total portal users: 1
Username: Client1
  Portal server: newpt
  State: Online
  VPN instance: N/A
  MAC                IP                VLAN   Interface
  0015-e9a6-7cfe     2.2.2.2           --     vlan-interface 100
  Authorization information:
    DHCP IP pool: N/A
    User profile: N/A
    Session group profile: N/A
    ACL number: N/A
    Inbound CAR: N/A
    Outbound CAR: N/A
    Microsegment ID: N/A
```

Example: Configuring portal authentication to support M-LAG

Network configuration

As shown in Figure 20, the hosts access the network through Device C (Layer 2 device). Each host is assigned a public IP address either manually or through DHCP.

An IMC server acts as a portal server (portal authentication server and portal Web server). A RADIUS server acts as the authentication and accounting server.

Configure the devices as follows:

- Configure direct portal authentication, so the hosts can access only the portal server before passing the authentication and access other network resources after passing the authentication.
- Configure Device A and Device B to form an M-LAG system to load share traffic. When a device fails, the other device can take over to ensure service continuity.
- Set the authentication load sharing mode to distributed if a large number of portal users exist.

Figure 20 Network diagram

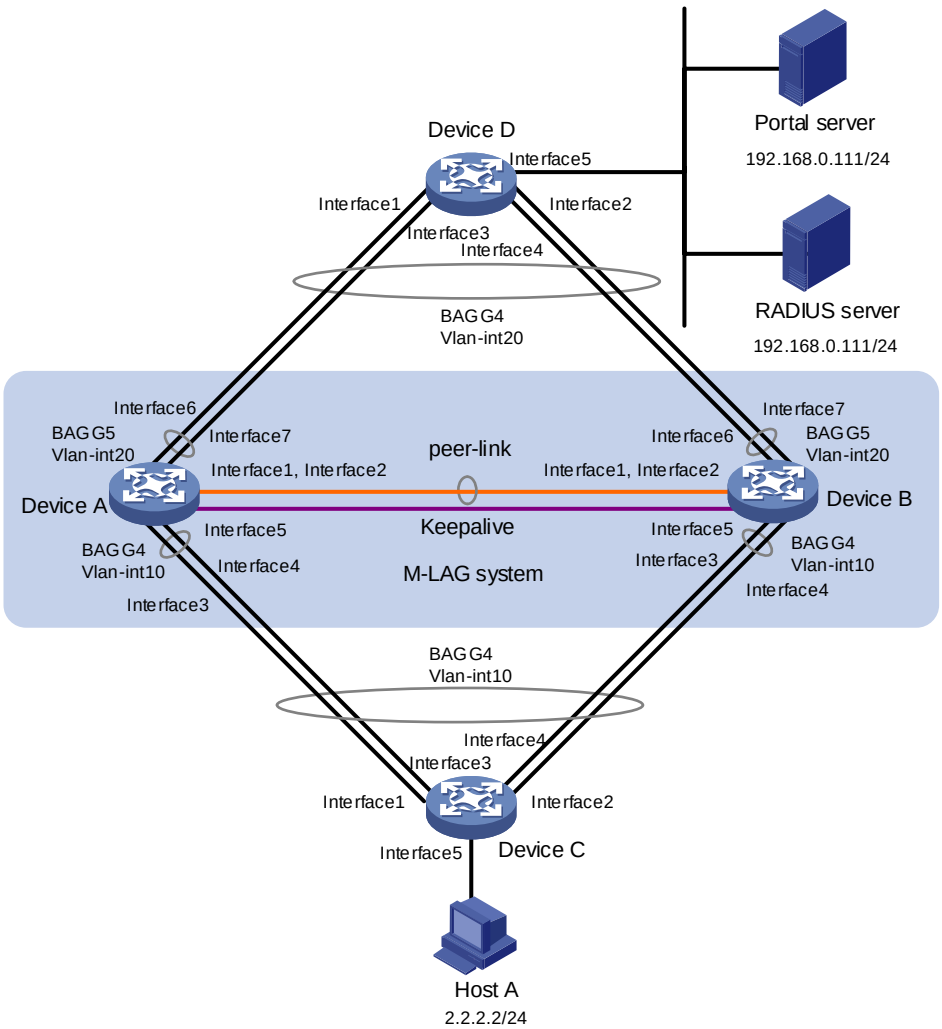


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5
Interface6	GigabitEthernet1/0/6

Interface7	GigabitEthernet1/0/7

Table 3 Interface and IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Device A	Vlan-int10	2.2.2.1/24	Device B	Vlan-int10	2.2.2.1/24
Device A	Vlan-int20	3.3.3.1/24	Device B	Vlan-int20	3.3.3.1/24
Device A	GE1/0/3	1.1.1.1/24	Device B	GE1/0/3	1.1.1.2/24
Device C	Vlan-int10	2.2.2.3/24	Device D	Vlan-int10	3.3.3.3/24

Prerequisites

1. Configure IP addresses for the host, switch, and servers as shown in Figure 20 and make sure they can reach each other.
2. Configure the RADIUS server correctly to provide authentication and accounting functions.
3. Configure the portal service on the IMC server, including setting portal server parameters, adding access devices, creating user IP address groups, and configuring device port groups to associate with IP address groups. This ensures coordinated implementation of portal authentication with the device.

Configuring the M-LAG system

1. Configure Device A.

   # Configure M-LAG system settings.

   ```
   <DeviceA> system-view
   [DeviceA] m-lag system-mac 1-1-1
   [DeviceA] m-lag system-number 1
   [DeviceA] m-lag system-priority 123
   ```

   # Configure the destination and source IP addresses of keepalive packets.

   ```
   [DeviceA] m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
   ```

   # Configure port GigabitEthernet 1/0/3 to operate in Layer 3 mode and configure the source IP address of keepalive packets as the IP address of this port.

   ```
   [DeviceA] interface gigabitethernet 1/0/3
   [DeviceA-GigabitEthernet1/0/3] port link-mode route
   [DeviceA-GigabitEthernet1/0/3] ip address 1.1.1.2 24
   [DeviceA-GigabitEthernet1/0/3] quit
   ```

   # Exclude GigabitEthernet 1/0/5 (keepalive interface) from being shut down upon detection of multi-active collisions.

   ```
   [DeviceA] m-lag mad exclude interface gigabitethernet 1/0/5
   ```

   # Create Layer 2 aggregate interface Bridge-Aggregation 3 and set the aggregation mode to dynamic.

   ```
   [DeviceA] interface bridge-aggregation 3
   [DeviceA-Bridge-Aggregation3] link-aggregation mode dynamic
   [DeviceA-Bridge-Aggregation3] quit
   ```

   # Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 3.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] port link-aggregation group 3
   [DeviceA-GigabitEthernet1/0/1] quit
   [DeviceA] interface gigabitethernet 1/0/2
   [DeviceA-GigabitEthernet1/0/2] port link-aggregation group 3
   [DeviceA-GigabitEthernet1/0/2] quit
   ```

Configure Layer 2 aggregate interface Bridge-Aggregation 3 as a trunk port, add the interface to VLAN 10 and VLAN 20, and configure the interface as a peer-link interface.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] port link-type trunk
[DeviceA-Bridge-Aggregation3] port trunk permit vlan 10 20
[DeviceA-Bridge-Aggregation3] port m-lag peer-link 1
[DeviceA-Bridge-Aggregation3] quit
```

Allow packets from the peer-link interface to prevent packets from the peer end from being discarded due to portal deny rules.

```
[DeviceA] portal free-rule 1 source interface bridge-aggregation 3
```

Create Layer 2 aggregate interface Bridge-Aggregation 4 and set the aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 4
[DeviceA-Bridge-Aggregation4] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 to aggregation group 4.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 4
[DeviceA-GigabitEthernet1/0/3] quit
[DeviceA] interface gigabitethernet 1/0/4
[DeviceA-GigabitEthernet1/0/4] port link-aggregation group 4
[DeviceA-GigabitEthernet1/0/4] quit
```

Add Layer 2 aggregate interface Bridge-Aggregation 4 to VLAN 10, and configure the interface as an M-LAG interface.

```
[DeviceA] interface bridge-aggregation 4
[DeviceA-Bridge-Aggregation4] port access vlan 10
[DeviceA-Bridge-Aggregation4] port m-lag group 4
[DeviceA-Bridge-Aggregation4] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 5 and set the aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 5
[DeviceA-Bridge-Aggregation5] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/6 and GigabitEthernet 1/0/7 to aggregation group 5.

```
[DeviceA] interface gigabitethernet 1/0/6
[DeviceA-GigabitEthernet1/0/6] port link-aggregation group 5
[DeviceA-GigabitEthernet1/0/6] quit
[DeviceA] interface gigabitethernet 1/0/7
[DeviceA-GigabitEthernet1/0/7] port link-aggregation group 5
[DeviceA-GigabitEthernet1/0/7] quit
```

Add Layer 2 aggregate interface Bridge-Aggregation 5 to VLAN 10, and configure the interface as an M-LAG interface.

```
[DeviceA] interface bridge-aggregation 5
[DeviceA-Bridge-Aggregation5] port access vlan 20
[DeviceA-Bridge-Aggregation5] port m-lag group 5
[DeviceA-Bridge-Aggregation5] quit
```

2. Configure Device B in the same way Device A is configured.

3. Configure Device C.

Create Layer 2 aggregate interface Bridge-Aggregation 4 and set the aggregation mode to dynamic.

```
<DeviceC> system-view
[DeviceC] interface bridge-aggregation 4
[DeviceC-Bridge-Aggregation4] link-aggregation mode dynamic
[DeviceC-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to aggregation group 4.

```
[DeviceC] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[DeviceC-if-range] port link-aggregation group 4
[DeviceC-if-range] quit
```

Add Layer 2 aggregate interface Bridge-Aggregation 4 to VLAN 100.

```
[DeviceC] interface bridge-aggregation 4
[DeviceC-Bridge-Aggregation4] port access vlan 10
[DeviceC-Bridge-Aggregation4] quit
```

Add GigabitEthernet 1/0/5 through GigabitEthernet 1/0/7 to VLAN 10.

```
[DeviceC] interface range gigabitethernet 1/0/5 to gigabitethernet 1/0/7
[DeviceC-if-range] port access vlan 10
[DeviceC-if-range] quit
```

4. Configure Device D in the same way Device C is configured.

Configuring portal authentication

1. Configure Device A.

a. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1** and enter its view.

```
<DeviceA> system-view
[DeviceA] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[DeviceA-radius-rs1] primary authentication 192.168.0.112
[DeviceA-radius-rs1] primary accounting 192.168.0.112
[DeviceA-radius-rs1] key authentication simple radius
[DeviceA-radius-rs1] key accounting simple radius
```

Exclude the ISP domain name from the username sent to the RADIUS server.

```
[DeviceA-radius-rs1] user-name-format without-domain
[DeviceA-radius-rs1] quit
```

b. Configure an authentication domain:

Create an ISP domain named **dm1** and enter its view.

```
[DeviceA] domain dm1
```

Configure AAA methods for the ISP domain.

```
[DeviceA-isp-dm1] authentication portal radius-scheme rs1
[DeviceA-isp-dm1] authorization portal radius-scheme rs1
[DeviceA-isp-dm1] accounting portal radius-scheme rs1
[DeviceA-isp-dm1] quit
```

Configure domain **dm1** as the default ISP domain. If a user enters the username without the ISP domain name at login, the authentication and accounting methods of the default domain are used for the user.

```
[DeviceA] domain default enable dm1
```

c. Configure portal authentication:

Configure a portal authentication server.

```
[DeviceA] portal server newpt
[DeviceA-portal-server-newpt] ip 192.168.0.111 key simple portal
[DeviceA-portal-server-newpt] port 50100
[DeviceA-portal-server-newpt] quit
```

Configure a portal Web server.

```
[DeviceA] portal web-server newpt
[DeviceA-portal-websvr-newpt] url http://192.168.0.111:8080/portal
[DeviceA-portal-websvr-newpt] quit
```

Enable direct portal authentication on VLAN-interface 100.

```
[DeviceA] interface vlan 10
[DeviceA–Vlan-interface10] portal enable method direct
```

Specify portal Web server **newpt** on VLAN-interface 100.

```
[DeviceA–Vlan-interface10] portal apply web-server newpt
```

Configure the BAS-IP as 2.2.2.1 for portal packets sent from VLAN-interface 100 to the portal authentication server.

```
[DeviceA–Vlan-interface10] portal bas-ip 2.2.2.1
[DeviceA–Vlan-interface10] quit
```

Set the authentication load sharing mode to distributed and set the distribution criterion to even-IP.

```
[DeviceA] portal m-lag load-sharing-mode distributed even-ip
```

2. Configure Device B.

Set the authentication load sharing mode to distributed and set the distribution criterion to odd-IP.

```
[DeviceB] portal m-lag load-sharing-mode distributed odd-ip
```

Configure other configuration on Device B in the same way Device A is configured.

Verifying the configuration

Verify that the portal configuration has taken effect.

```
[DeviceA] display portal interface vlan-interface 10
 Portal information of Vlan-interface10
     NAS-ID profile: Not configured
     M-LAG Group ID : 4
     M-LAG role     : Master
     Authorization : Strict checking
     ACL           : Disabled
     User profile  : Disabled
     Dual stack    : Disabled
     Max users     : Not configured
 IPv4:
     Portal status: Enabled
     PEER_SM state: M_Synced
     Portal authentication method: Direct    authentication shared mode: Disabled
     Portal web server: newpt
     Portal mac-trigger-server: Not configured
     Authentication domain: dm1
     Pre-auth domain: Not configured
     User-dhcp-only: Disabled
     Pre-auth IP pool: Not configured
```

```
    Max users: Not configured
    Bas-ip: 2.2.2.1
    User detection: Not configured
    Action for server detection:
        Server type    Server name                          Action
        --             --                                   --
    Layer3 source network:
        IP address              Mask

    Destination authentication subnet:
        IP address              Mask
IPv6:
    Portal status: Disabled
    Portal authentication method: Disabled    authentication shared mode: Disabled
    Portal web server: Not configured
    Portal mac-trigger-server: Not configured
    Authentication domain: Not configured
    Pre-auth domain: Not configured
    User-dhcp-only: Disabled
    Pre-auth IP pool: Not configured
    Max users: Not configured
    Bas-ipv6: Not configured
    User detection: Not configured
    Action for server detection:
        Server type    Server name                          Action
        --             --                                   --
    Layer3 source network:
        IP address                                          Prefix length

    Destination authentication subnet:
        IP address                                          Prefix length
```

A user can perform portal authentication by using the iNode client or through a Web browser. Before passing the authentication, the user can access only the authentication page **http://192.168.0.111:8080/portal**. All Web requests from the user will be redirected to the authentication page. After passing the authentication, the user can access other network resources.

After the user passes authentication, display information about the portal user on Device A.

```
[DeviceA] display portal user interface vlan-interface 10 verbose
Total portal users: 1
Basic:
  Current IP address: 2.2.2.2
  Original IP address: 2.2.2.2
  Username: 11
  User ID: 0x10000012
  Access interface: Vlan-interface10
  Service-VLAN/Customer-VLAN: 10/-
  MAC address: 00e0-4c68-8add
  Domain name: dm1
  VPN instance: N/A
```

```
  Status: Online
  M-LAG user state: Active
  Portal server: N/A
  Portal authentication method: Direct    authentication shared mode: Disabled
AAA:
  Realtime accounting interval: 0s, retry times: 1
  Idle cut: N/A
  Session duration: N/A, remaining: N/A
  Remaining traffic: N/A
  Login time: 2019-09-16 06:48:26 UTC
  Accounting-start fail action: Online
  Accounting-update fail action: Online
  Accounting quota-out action: Offline
  DHCP IP pool: N/A
  Microsegment ID: N/A
ACL&QoS&Multicast:
  Inbound CAR: N/A
  Outbound CAR: N/A
  ACL number: N/A
  User profile: N/A
  Session group profile: N/A
  Max multicast addresses: 4
  User group: N/A
Flow statistic:
  Uplink   packets/bytes: 5/290
  Downlink packets/bytes: 0/0
Peer flow statistic:
  Uplink   packets/bytes: 5/290
  Downlink packets/bytes: 0/0
```

Display information about the portal user on Device B.

```
[DeviceB] display portal user interface Vlan-interface 10 verbose
Total portal users: 1
Basic:
  Current IP address: 2.2.2.2
  Original IP address: 2.2.2.2
  Username: 11
  User ID: 0x1000000c
  Access interface: Vlan-interface10
  Service-VLAN/Customer-VLAN: 10/-
  MAC address: 00e0-4c68-8add
  Domain name: dm1
  VPN instance: N/A
  Status: Online
  M-LAG user state: Inactive
  Portal server: N/A
  Portal authentication method: Direct    authentication shared mode: Disabled
AAA:
  Realtime accounting interval: 0s, retry times: 1
```

```
  Idle cut: N/A
  Session duration: N/A, remaining: N/A
  Remaining traffic: N/A
  Login time: 2019-09-16 06:48:29 UTC
  Accounting-start fail action: Online
  Accounting-update fail action: Online
  Accounting quota-out action: Offline
  DHCP IP pool: N/A
  Microsegment ID: N/A
ACL&QoS&Multicast:
  Inbound CAR: N/A
  Outbound CAR: N/A
  ACL number: N/A
  User profile: N/A
  Session group profile: N/A
  Max multicast addresses: 4
  User group: N/A
Flow statistic:
  Uplink   packets/bytes: 5/290
  Downlink packets/bytes: 0/0
Peer flow statistic:
  Uplink   packets/bytes: 5/290
  Downlink packets/bytes: 0/0
```

Troubleshooting portal

No portal authentication page is pushed for users

Symptom

When a user is redirected to the IMC portal authentication server, no portal authentication page or error message is prompted for the user. The login page is blank.

Analysis

The key configured on the portal access device and that configured on the portal authentication server are inconsistent. As a result, packet verification fails, and the portal authentication server refuses to push the authentication page.

Solution

Use the **display this** command in portal authentication server view on the access device to check whether a key is configured for the portal authentication server.

- If no key is configured, configure the right key.
- If a key is configured, use the **ip** or **ipv6** command in the portal authentication server view to correct the key, or correct the key configured for the access device on the portal authentication server.

Cannot log out portal users on the access device

Symptom

You cannot use the **portal delete-user** command on the access device to log out a portal user, but the portal user can log out by clicking the **Disconnect** button on the portal authentication client.

Analysis

When you execute the **portal delete-user** command on the access device to log out a user, the access device sends an unsolicited logout notification message to the portal authentication server. The destination port number in the logout notification is the listening port number of the portal authentication server configured on the access device. If this listening port number is not the actual listening port number configured on the server, the server cannot receive the notification. As a result, the server does not log out the user.

When a user uses the **Disconnect** button on the authentication client to log out, the portal authentication server sends an unsolicited logout request message to the access device. The access device uses the source port in the logout request as the destination port in the logout ACK message. As a result, the portal authentication server can definitely receive the logout ACK message and log out the user.

Solution

1. Use the **display portal server** command to display the listening port of the portal authentication server configured on the access device.
2. Use the **portal server** command in system view to change the listening port number to the actual listening port of the portal authentication server.

Cannot log out portal users on the RADIUS server

Symptom

The access device uses the IMC server as the RADIUS server to perform identity authentication for portal users. You cannot log out the portal users on the RADIUS server.

Analysis

The IMC server uses session control packets to send disconnection requests to the access device. On the access device, the listening UDP port for session control packets is disabled by default. Therefore, the access device cannot receive the portal user logout requests from the RADIUS server.

Solution

On the access device, execute the **radius session-control enable** command in system view to enable the RADIUS session control function.

Users logged out by the access device still exist on the portal authentication server

Symptom

After you log out a portal user on the access device, the user still exists on the portal authentication server.

Analysis

When you execute the **portal delete-user** command on the access device to log out a user, the access device sends an unsolicited logout notification to the portal authentication server. If the

BAS-IP or BAS-IPv6 address carried in the logout notification is different from the portal device IP address specified on the portal authentication server, the portal authentication server discards the logout notification. When sending of the logout notifications times out, the access device logs out the user. However, the portal authentication server does not receive the logout notification successfully, and therefore it regards the user is still online.

Solution

Configure the BAS-IP or BAS-IPv6 attribute on the interface enabled with portal authentication. Make sure the attribute value is the same as the portal device IP address specified on the portal authentication server.

Re-DHCP portal authenticated users cannot log in successfully

Symptom

The device performs re-DHCP portal authentication for users. A user enters the correct username and password, and the client successfully obtains the private and public IP addresses. However, the authentication result for the user is failure.

Analysis

When the access device detects that the client IP address is changed, it sends an unsolicited portal packet to notify of the IP change to the portal authentication server. The portal authentication server notifies of the authentication success only after it receives the IP change notification from both the access device and the client.

If the BAS-IP or BAS-IPv6 address carried in the portal notification packet is different from the portal device IP address specified on the portal authentication server, the portal authentication server discards the portal notification packet. As a result, the portal authentication server considers that the user has failed the authentication.

Solution

Configure the BAS-IP or BAS-IPv6 attribute on the interface enabled with portal authentication. Make sure the attribute value is the same as the portal device IP address specified on the portal authentication server.

Contents

Configuring Web authentication 1

About Web authentication 1
Advantages of Web authentication 1
Web authentication system 1
Web authentication process 2
Web authentication support for EAP 3
Web authentication support for VLAN assignment 3
Web authentication support for authorization ACLs 4
Web authentication support for microsegment assignment 5
Restrictions and guidelines: Web authentication configuration 5
Web authentication tasks at a glance 6
Prerequisites for Web authentication 7
Configuring a remote Web server for Web authentication 7
Configuring a remote portal authentication server 8
Obtaining user information from ARP or ND entries 9
Configuring a local Web server for Web authentication 9
Configuring a local portal service 10
Enabling Web authentication 10
Specifying a Web authentication domain 11
Setting the redirection wait time 11
Configuring the aging timer for temporary MAC address entries 12
Configuring Web authentication-free subnets or host names 12
Configuring a Web authentication-free subnet 12
Configuring a Web authentication-free host name 13
Setting the maximum number of Web authentication users 13
Configuring online Web authentication user detection 14
Enabling online detection bypass for Web authentication users 14
Configuring an Auth-Fail VLAN 15
Configuring Web authentication multi-VLAN mode 15
Configuring Web authentication to support Web proxy 16
Applying a portal MAC binding server for Web authentication 16
Enabling Web authentication user logging 17
Display and maintenance commands for Web authentication 17
Web authentication configuration examples 18
Example: Configuring local Web authentication by using the AAA local authentication method 18
Example: Configuring local Web authentication by using the AAA remote authentication method 20
Example: Configuring remote Web authentication by using the AAA local authentication method 22
Example: Configuring remote Web authentication by using the AAA remote authentication method 24
Troubleshooting Web authentication 26
Failure to come online (local authentication interface using the default ISP domain 26

Configuring Web authentication

About Web authentication

Web authentication is deployed on Layer 2 interfaces of the access device to control user access to networks. The access device redirects unauthenticated users to the specified website. The users can access the resources on the website without authentication. If the users want to access other network resources, they must pass authentication.

Advantages of Web authentication

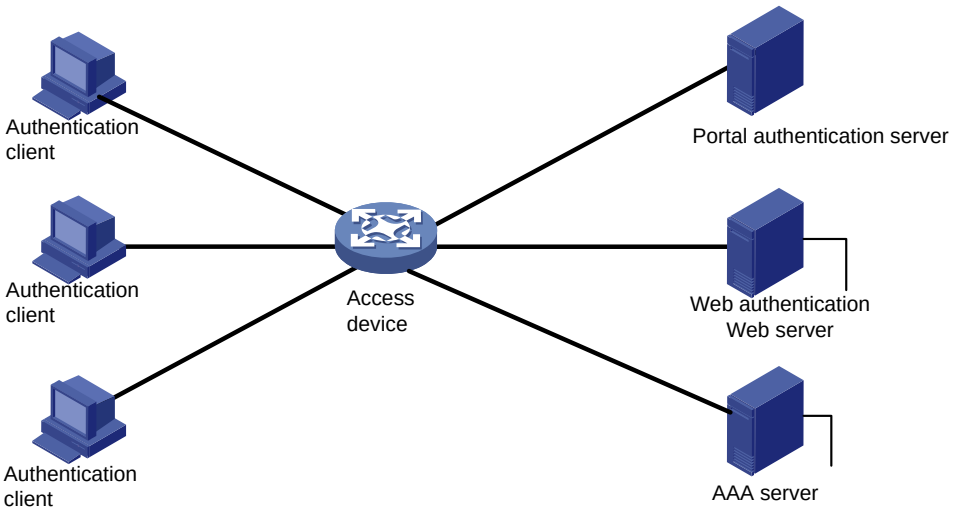
Web authentication has the following advantages:

- Allows users to perform authentication through webpages without installing client software.
- Provides ISPs with diversified management choices and extended functions. For example, the ISPs can place advertisements, provide community services, and publish information on the authentication page.

Web authentication system

A typical Web authentication system consists of the following basic components: authentication client, access device, Web server for Web authentication, portal authentication server, and AAA server.

Figure 1 Web authentication system using the local portal server



Authentication client

An authentication client is a Web browser that runs HTTP or HTTPS on the user endpoint.

Access device

An access device has the following functions:

- Redirects all user HTTP or HTTPS requests that do not match authentication-free rules to the Web authentication page before authentication.
- Interacts with the AAA server to complete authentication, authorization, and accounting. For more information about AAA, see "Configuring AAA."

- Allows users that pass authentication to access authorized network resources.

Web server for Web authentication

A Web server pushes the Web authentication page to authentication clients and obtains user authentication information (username and password).

A Web server for Web authentication can be a local Web server or a remote Web server. A local Web server is usually integrated into the access device. A remote Web server is a portal Web server.

Portal authentication server

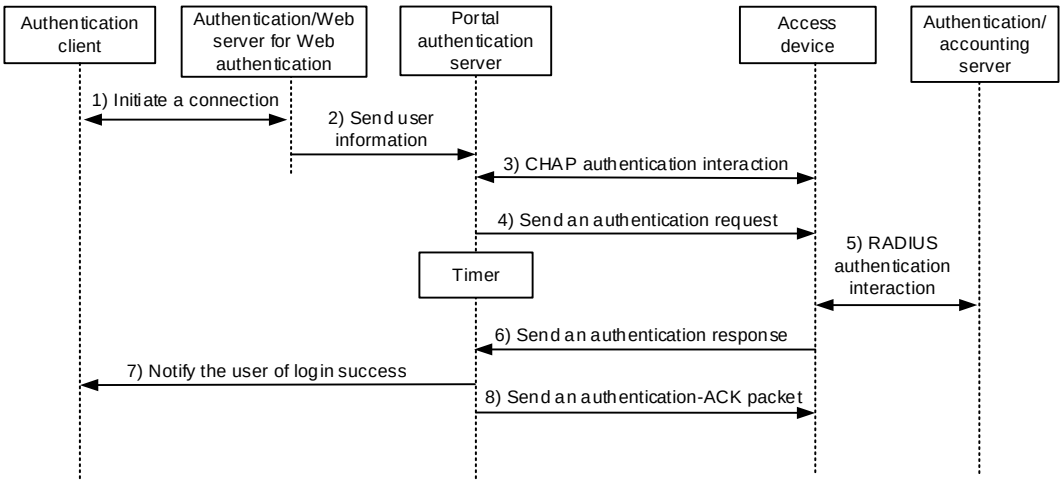
Web authentication uses a portal authentication server as the authentication server, which exchanges client authentication information with the access device.

AAA server

An AAA server interacts with the access device to implement user authentication, authorization, and accounting. A RADIUS server can perform authentication, authorization, and accounting for Web authentication users. An LDAP server can perform authentication for Web authentication users.

Web authentication process

Figure 2 Web authentication process



The Web authentication process is as follows:

1. An unauthenticated user sends an HTTP or HTTPS request. When the access device receives the HTTP or HTTPS request on a Layer 2 Ethernet interface enabled with Web authentication, the request, it redirects the request to the Web authentication page. The user enters the username and password on the Web authentication page.

   If the user requests the Web authentication page or free Web resources, the access device permits the request. No Web authentication is performed.
2. The Web server for Web authentication sends the user information to the portal authentication.
3. The portal authentication server and the access device interact with each other to perform CHAP authentication. This step is skipped if PAP authentication is used. The authentication interaction type is determined by the portal authentication server.
4. The portal authentication sends an authentication request that contains the username and password of the user to the access device and starts a timer to wait for an authentication response.
5. The access device and the AAA server exchange RADIUS packets to authenticate the user.

6. The access device sends an authentication response indicating authentication success or failure to the portal authentication server.
7. The portal authentication server notifies the authentication result to the client.
8. If the user passes RADIUS authentication, the portal authentication server also sends an authentication-ACK packet to the access device.

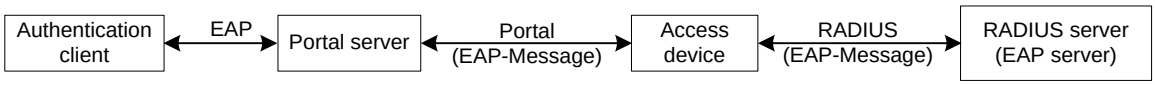
Web authentication support for EAP

This feature is supported only when Web authentication uses a remote IMC portal authentication server and an iNode portal client.

Compared with username and password based authentication, digital certificate-based authentication ensures higher security.

The Extensible Authentication Protocol (EAP) supports several digital certificate-based authentication methods, for example, EAP-TLS. Working together with EAP, Web authentication can implement digital certificate-based user authentication.

Figure 3 Web authentication support for EAP protocol exchanges



As shown in Figure 3, the authentication client and the portal authentication server exchange EAP authentication packets. The portal authentication server and the access device exchange portal authentication packets that carry the EAP-Message attributes. The access device and the RADIUS server exchange RADIUS packets that carry the EAP-Message attributes. The RADIUS server that supports the EAP server function processes the EAP packets encapsulated in the EAP-Message attributes, and provides the EAP authentication result.

The access device does not process but only transports EAP-Message attributes between the portal authentication server and the RADIUS server. Therefore, the access device requires no additional configuration to support EAP authentication.

Web authentication support for VLAN assignment

Authorization VLAN

Web authentication uses VLANs authorized by the AAA server or the access device to control network resource access of authenticated users. The authorization VLAN information includes VLAN ID and VLAN name.

After a user passes Web authentication, the AAA server or the access device authorizes the user to access a VLAN. If the authorization VLAN does not exist, the access device first creates the VLAN and then assigns the user access interface as an untagged member to the VLAN. If the authorization VLAN already exists, the access device directly assigns the user access interface as an untagged member to the VLAN. Then, the user can access resources in the authorization VLAN.

Table 1 describes the way the access device handles authorization VLANs for Web authenticated users.

Table 1 VLAN manipulation

Port type	VLAN manipulation
• Access port • Trunk port • Hybrid port with MAC-based-VLAN disabled	The device assigns the port to the first authenticated user's authorization VLAN. The authorization VLAN becomes the PVID. All Web authentication users on the port must be assigned the same authorization VLAN. If a different authorization VLAN is assigned to a subsequent user, the user cannot pass Web authentication.

Port type	VLAN manipulation
Hybrid port with MAC-based VLAN enabled	The device maps the MAC address of each user to its own authorization VLAN regardless of whether the port is a tagged member. The PVID of the port does not change.

Auth-Fail VLAN

An Auth-Fail VLAN is a VLAN assigned to users who fail authentication. The Auth-Fail VLAN provides network resources such as the patch server, virus definitions server, client software server, and anti-virus software server to the users. The users can use these resources to upgrade their client software or other programs.

Web authentication supports Auth-Fail VLAN on an interface that performs MAC-based access control. If a user on the interface fails authentication, the access devices creates a MAC VLAN entry based on the MAC address of the user and adds the user to the Auth-Fail VLAN. Then, the user can access the authentication-free IP resources in the Auth-Fail VLAN. All HTTP or HTTPS requests to non-authentication-free IP resources will be redirected to the authentication page. If the user still fails authentication, the interface remains in the Auth-Fail VLAN. If the user passes authentication, the access device removes the interface from the Auth-Fail VLAN and assigns the interface to a VLAN as follows:

- If the authentication server assigns an authorization VLAN to the user, the access device assigns the interface to the authorization VLAN.
- If the authentication server does not assign an authorization VLAN to the user, the access device assigns the interface to the default VLAN.

Web authentication multi-VLAN mode

Web authentication supports single-VLAN or multi-VLAN mode on a port.

In multi-VLAN mode, the port forwards traffic from a user in different VLANs without reauthentication if the user has been authenticated and come online in any VLAN on the port.

In single-VLAN mode, the port reauthenticates an online user when traffic received from that user contains a VLAN tag different from the VLAN in which the user was authenticated. The authentication process differs depending on the MAC move setting in port security and the authorization VLAN assignment status, as follows:

- If no authorization VLAN has been assigned to the online user, the device first logs off the user and then reauthenticates the user in the new VLAN.
- If the online user has been assigned an authorization VLAN, the device handles the user depending on the MAC move setting in port security.
  - If MAC move is disabled in port security, the user cannot pass authentication and come online from the new VLAN until after it goes offline from the port.
  - If MAC move is enabled in port security, the user can pass authentication on the new VLAN and come online without having to first go offline from the port. After the user passes authentication on the new VLAN, the original authentication session of the user is deleted from the port.

  To enable the port security MAC move feature, use the **`port-security mac-move permit`** command.

Web authentication support for authorization ACLs

Web authentication uses ACLs authorized by the AAA server or the access device to control user access to network resources and limit user access rights. When a user passes authentication, the AAA server and the access device assigns an authorization ACL to the access interface of the user. The access device filters traffic from the user on the access interface according to the authorized ACL.

You must configure the authorization ACLs on the access device if you specify authorization ACLs on the authentication server.

To change the access control criteria for the user, you can specify a different authorization ACL on the authentication server or change rules in the authorization ACL on the access device.

The device supports the following types of authorization ACLs:

- Basic ACLs (ACL 2000 to ACL 2999).
- Advanced ACLs (ACL 3000 to ACL 3999).
- Layer 2 ACLs (ACL 4000 to ACL 4999).

For an authorization ACL to take effect, make sure the ACL exists and has ACL rules excluding rules configured with the **`counting`**, **`established`**, **`fragment`**, **`source-mac`**, or **`logging`** keyword. For more information about ACL rules, see ACL commands in *ACL and QoS Command Reference*.

In a microsegment network, if a user is assigned both an authorization microsegment and ACL, the authorization ACL does not take effect.

Web authentication support for microsegment assignment

To implement fine-grained isolation of Web authentication user traffic, users are usually assigned to different groups based on their business requirements for group-based user access control. The IDs of the groups are called microsegments, which can be assigned to users by the RADIUS server. When a user comes online, the RADIUS server assigns an authorization microsegment to the user. The access device filters traffic from the user on the access port according to the authorization microsegment. The user can access the network resources in the microsegment.

In a non-microsegment network, do not configure authorization microsegments for users.

Restrictions and guidelines: Web authentication configuration

When you configure Web authentication, follow these restrictions and guidelines:

- Web authentication cannot be configured in conjunction with the following 802.1X settings:
  - Guest, Auth-Fail, and critical VLAN settings.
  - Guest, Auth-Fail, and critical VSI settings.
  - Critical microsegment settings.
- Web authentication cannot be configured in conjunction with the following MAC authentication settings:
  - Guest and critical VLAN settings.
  - Guest and critical VSI settings.
  - Critical microsegment settings.
- Auth-Fail VLAN is mutually exclusive with the port security preauthentication domain feature.

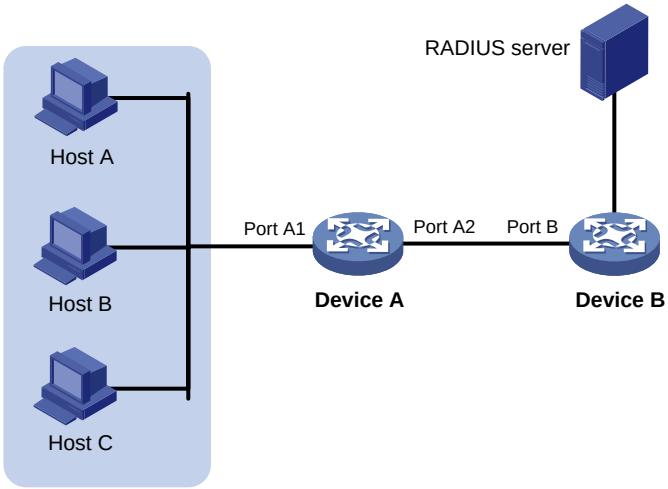
To redirect HTTPS requests for unauthenticated users correctly, make sure the packet incoming VLANs or VSIs have Layer 3 interfaces (VLAN or VSI interfaces) configured.

To access the resources in the authorization VLAN or Auth-Fail VLAN, a user must update the IP address of the client after being assigned to the authorization VLAN or Auth-Fail VLAN.

As a best practice, perform Web authentication on users directly connected to the device. As shown in Figure 4, if you enable Web authentication on Port B to authenticate non-directly connected users (the hosts), you must follow these restrictions and guidelines:

- If the RADIUS server assigns an authorization VLAN to the users, make sure the following conditions are met:
  - The link between Device A and Device B is a trunk link.
  - The PVIDs of Port A1 and Port B are the same as the authorization VLAN ID.
- If the RADIUS server does not assign an authorization VLAN to the users, make sure the PVIDs of Port A1 and Port B are the same.

Figure 4 Web authentication for non-directly connected users



Web authentication tasks at a glance

To configure Web authentication, perform the following tasks:

1. Configuring a remote Web authentication service

   Perform this task if a remote Web server is deployed on the network for Web authentication.

   a. Configuring a remote portal authentication server
   b. Configuring a remote Web server for Web authentication
   c. Obtaining user information from ARP or ND entries
2. Configuring a local Web authentication service

   Perform this task if the device is used as the local Web server for Web authentication.

   a. Configuring a local Web server for Web authentication
   b. Configuring a local portal service
3. Enabling Web authentication
4. (Optional.) Specifying a Web authentication domain
5. (Optional.) Setting the redirection wait time
6. (Optional.) Configuring the aging timer for temporary MAC address entries
7. (Optional.) Configuring Web authentication-free subnets or host names
8. (Optional.) Setting the maximum number of Web authentication users
9. (Optional.) Configuring online Web authentication user detection
10. (Optional.) Enabling online detection bypass for Web authentication users
11. (Optional.) Configuring an Auth-Fail VLAN
12. (Optional.) Configuring Web authentication to support Web proxy
13. Applying a portal MAC binding server for Web authentication

    This task is required to perform the MAC-based quick authentication.

14. (Optional.) Enabling Web authentication user logging

Prerequisites for Web authentication

The device supports two methods for Web authentication, which are local authentication and RADIUS authentication.

To use the RADIUS authentication method, you must complete the following tasks:

- Install a RADIUS server and configure the RADIUS server properly.
- Make sure the authentication client, the access device, and the RADIUS server can reach each other.
- Configure user accounts on the RADIUS server and configure the RADIUS client information on the access device.

To use the local authentication method, you must configure local users on the access device.

For more information about RADIUS clients and local users, see "Configuring AAA."

Configuring a remote Web server for Web authentication

About this task

In the view of a remote Web server, you can configure the Web server IP address, redirection URL, and the parameters to be carried in the redirection URL.

By default, the state of the Web server specified by the redirection URL is always **active**. The device cannot obtain the real server state. When primary and secondary Web servers are deployed for continuous authentication service, the device needs to quickly obtain the primary server's state changes to perform primary/secondary switchovers accordingly.

You can associate the Web server with a track entry that is associated with NQA, so the device can periodically detect the reachability of the server through NQA.

- The Track module changes the state of the track entry according to the NQA detection result. The device changes the state of the Web server according to the state of the track entry.
- In this way, if NQA detects that the Web server becomes reachable, the device will set the server state to active. If NQA detects that the Web server becomes unreachable, the device will set the server state to inactive.

For more information about the Track configuration, see *High Availability Configuration Guide*.

Restrictions and guidelines

To deploy both primary and secondary Web servers, you must associate the primary Web server with a track entry to monitor the reachability status of the server. Otherwise, the device cannot sense the reachability state changes of the primary server to perform primary/secondary switchovers.

The NQA detection for server reachability is performed according to the **`nqa schedule`** configuration, which schedules the NQA operation associated with the track entry. Configure the NQA operation time and other detection parameters as needed.

Procedure

1. Enter system view.

   `system-view`

2. Create a remote Web server for Web authentication and enter Web authentication remote Web server view.

web-auth remote server *server-name*

The local and remote Web servers cannot use the same name.

3. Specify the IP address of the remote Web server for Web authentication.

   IPv4:

   ip *ipv4-address*

   IPv6:

   ipv6 *ipv6-address*

   By default, no IP address is specified for the remote Web server for Web authentication.

   The IP address of the remote Web server must be the IP address of the portal Web server used for Web authentication.

   You can configure one IPv4 address and one IPv6 address for a remote Web server.

4. Configure the redirection URL of the remote Web server for Web authentication.

   url *url-string* [**track** *track-entry-number*]

   By default, no redirection URL is specified for a Web server for Web authentication.

   To provide Web authentication pages for both IPv4 and IPv6 Web authentication users, configure the redirection URL to carry the domain name of the Web server. For example: **https://example.com**, where **example.com** is the domain name of the Web server.

   A Web server can be associated with only one track entry. If you associate Track with the Web server multiple times, the most recent configuration takes effect.

5. (Optional.) Add parameters to the redirection URL of the remote Web server.

   url-parameter *parameter-name* { **original-url** | **source-address** | **source-mac** | **value** *expression* }

   By default, no parameters are added to the redirection URL of a Web server.

   You must use the URL parameter name supported by the Web browser. Different Web browsers support different URL parameter names.

6. (Optional.) Configure the unescaped special characters in the Web authentication URL redirected to users.

   url-unescape-chars *character-string*

   By default, the device escapes all special characters in the Web authentication URL redirected to users.

7. (Optional.) Configure the Web server to support the Oasis platform.

   server-type oauth

   By default, a Web authentication server does not support the Oasis platform.

   To use an Oasis server as a Web authentication server, execute this command to support the Oasis platform.

Configuring a remote portal authentication server

For information about configuring a remote portal authentication server, see "Configuring portal authentication."

Obtaining user information from ARP or ND entries

About this task

By default, the device obtains user information required by the portal authentication server from MIB entries. However, MIB entries do not contain the Layer 2 user information (such as MAC address and Layer 2 access interface) required for Web authentication. Users will fail the Web authentication as a result.

To resolve this issue, configure the device to obtain user information from ARP or ND entries.

Procedure

1. Enter system view.

   system-view

2. Configure the device to obtain portal user information from ARP or ND entries.

   portal access-info trust { **arp** | **nd** }

   By default, the device obtains user information from FIB entries.

   For more information about the configuration, see "Configuring portal authentication."

Configuring a local Web server for Web authentication

Restrictions and guidelines

Specify the IP address of a Layer 3 interface on the device that is routable to the Web client as the listening IP address of the local Web server for Web authentication. As a best practice, use the IP address of a loopback interface. A loopback interface has the following advantages:

- The status of a loopback interface is stable. There will be no authentication page access failures caused by interface failures.
- A loopback interface does not forward received packets to any networks, avoiding impact on system performances when there are many network access requests.

Procedure

1. Enter system view.

   system-view

2. Create a local Web server for Web authentication and enter Web authentication local Web server view.

   web-auth server *server-name*

   The local and remote Web servers cannot use the same name.

3. Specify the IP address and port number of the local Web server.

   IPv4:

   ip *ipv4-address* **port** *port-number*

   IPv6:

   ipv6 *ipv6-address* **port** *port-number*

   The port number of the local Web server must be the same as the listening port number of the local portal Web service.

   You can configure one IPv4 address and one IPv6 address for a local Web server.

4. Specify the redirection URL for the local Web server.

   url *url-string*

   By default, no redirection URL is specified for a Web server for Web authentication.

   The redirection URL must start with **http://** or **https://**. The IP address and port number in the specified redirection URL must be the same as those of the local Web server for Web authentication.

   To provide Web authentication pages for both IPv4 and IPv6 Web authentication users, configure the redirection URL to carry the domain name of the Web server.

5. (Optional.) Add parameters to the redirection URL of the local Web server.

   url-parameter *parameter-name* { **original-url** | **source-address** | **source-mac** | **value** *expression* }

   By default, no parameters are added to the redirection URL of a Web server.

   You must use the URL parameter name supported by the Web browser. Different Web browsers support different URL parameter names.

6. (Optional.) Configure the unescaped special characters in the Web authentication URL redirected to users.

   url-unescape-chars *character-string*

   By default, the device escapes all special characters in the Web authentication URL redirected to users.

7. (Optional.) Configure the Web server to support the Oasis platform.

   server-type oauth

   By default, a Web authentication server does not support the Oasis platform.

   To use an Oasis server as a Web authentication server, execute this command to support the Oasis platform.

Configuring a local portal service

For information about the local portal service configuration, see "Configuring portal authentication."

Enabling Web authentication

Restrictions and guidelines

For Web authentication to operate correctly, do not enable port security or configure the port security mode on the Layer 2 interface enabled with Web authentication. For more information about port security, see "Configuring port security."

To deploy both primary and secondary Web servers, follow these restrictions and guidelines:

- Only a remote Web server can be used as the primary Web server. A remote or local Web server can be used as the secondary Web server.
- Associate the primary Web server with a track entry (by using the **url** *url-string* **track** *track-entry-number* command) to monitor the reachability status of the server. Otherwise, the device cannot sense the reachability state changes of the primary server to perform primary/secondary switchovers.

The default HTTPS redirect listening port number on the device is 6654. To change the HTTPS redirect listening port number, see HTTP redirect configuration in *Layer 3—IP Services Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable Web authentication and specify Web servers for Web authentication.

   web-auth enable apply server *primary-server-name* [**secondary-server** *secondary-server-name*]

   By default, Web authentication is disabled.

Specifying a Web authentication domain

About this task

You can specify different authentication domains for Web authentication users on different interfaces. After you specify a Web authentication domain on an interface, the device uses the authentication domain for AAA of all Web authentication users on the interface, ignoring the domain names carried in the usernames.

The device selects the authentication domain for a Web authentication user on an interface in this order:

1. The authentication domain specified for the interface.
2. The authentication domain carried in the username.
3. The system default authentication domain.
4. The ISP domain configured to accommodate users assigned to nonexistent domains.

If the selected domain does not exist on the device, user authentication fails. For information about ISP domains, see "Configuring AAA."

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Specify an authentication domain for Web authentication users on the interface.

   web-auth domain *domain-name*

   By default, no authentication domain is specified for Web authentication users.

Setting the redirection wait time

About this task

The redirection wait time determines the length of time that the device waits to redirect a user to the specified webpage after the user passes Web authentication.

You need to change the redirection wait time in some scenarios, for example, when a user needs to update the client IP address after passing local Web authentication. To ensure that the specified webpage can be successfully opened, set the redirection wait time to be greater than the time that the user takes to update the IP address of the client.

Procedure

1. Enter system view.

   system-view

2. Enter Web authentication local Web server view.

web-auth server *server-name*

3. Set the redirection wait time.

   redirect-wait-time *period*

   By default, the redirection wait time for authenticated users is 5 seconds.

Configuring the aging timer for temporary MAC address entries

About this task

If Web authentication is enabled, the device generates a temporary MAC address entry when it detects traffic from a user for the first time. The entry records the MAC address, access interface, and VLAN ID of the user, as well as the aging time of the entry.

The aging timer works as follows:

- If the user does not initiate authentication when the aging timer expires, the device deletes the temporary entry.
- If the user passes authentication before the aging timer expires, the device deletes the aging timer and records online information for the Web authentication user.
- If the user fails authentication before the aging timer expires and an Auth-Fail VLAN is specified for Web authentication, the device binds the MAC address of the user to the Auth-Fail VLAN and reset the aging timer. If the user still fails authentication when the aging timer expires, the device deletes the temporary entry for the user.

Restrictions and guidelines

As a best practice, change the aging timer to a bigger value in the following cases:

- Web authentication users without access rights frequently send traffic in a short time. As a result, the access device continuously initiates the web authentication process, increasing the load on the device.
- When a user fails authentication, the user does not have enough time to obtain resources from the Auth-Fail VLAN, for example, it failed to download the virus patches.

Procedure

1. Enter system view.

   system-view

2. Configure the aging timer for temporary MAC address entries.

   web-auth timer temp-entry-aging *aging-time-value*

   By default, the aging timer for temporary MAC address entries is 60 seconds.

Configuring Web authentication-free subnets or host names

Configuring a Web authentication-free subnet

About this task

You can configure a Web authentication-free subnet so that users can freely access the network resources in the subnet without being authenticated.

Restrictions and guidelines

As a best practice, do not configure the same address value for a Web authentication-free subnet and a 802.1X free IP. Otherwise, when you cancel one of the configuration, the other configuration does not take effect, either.

Procedure

1. Enter system view.

   system-view

2. Configure a Web authentication-free subnet.

   web-auth free-ip *ip-address* { *mask-length* | *mask* }

   By default, no Web authentication-free subnets exist.

Configuring a Web authentication-free host name

About this task

You can configure Web authentication-free host names so that users can freely access the specified host resources without being authenticated.

Restrictions and guidelines

Web authentication-free host names support only precise matching. For example, Web authentication-free host name **abc.com.cn** matches only HTTP/HTTPS request packets that contain the host name of **abc.com.cn**. It does not match packets that contain host name **dfabc.com.cn**.

Prerequisites

Before configuring Web authentication-free destination host names, make sure a DNS server has been deployed on the network, or the host name-to-IPv4 address mappings have been configured by using the **ip host** command. For more information about the **ip host** command, see DNS commands in *Layer 3—IP Services Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Configure a Web authentication-free host name.

   web-auth free-host *host-name*

   By default, no Web authentication-free host names exist.

Setting the maximum number of Web authentication users

Restrictions and guidelines

If the maximum number of online Web authentication users you set is less than that of the current online Web authentication users, the limit can be set successfully and does not impact the online Web authentication users. However, the system does not allow new Web authentication users to log in until the number drops down below the limit.

Procedure

1. Enter system view.

system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Set the maximum number of Web authentication users on the interface.
   web-auth max-user [**preauth-domain** | **auth-fail-domain**] *max-number*
   By default, the maximum number of Web authentication users is 1024.

Configuring online Web authentication user detection

About this task

This feature enables the device to detect packets of an online user at the specified detection interval. If no packet from the user is received within the interval, the device logs out the user and notifies the RADIUS server to stop accounting for the user.

Restrictions and guidelines

To prevent the device from mistakenly logging out users, set the detection interval to be the same as the aging time of MAC address entries.

On a Layer 2 aggregate interface, the real interval for logging out a user is greater than the specified detection interval. The maximum delay is 300 seconds.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable online Web authentication user detection.
   web-auth offline-detect interval *interval*
   By default, online Web authentication user detection is disabled.

Enabling online detection bypass for Web authentication users

About this task

By default, when all RADIUS servers in the ISP domain are unreachable and online Web authentication user detection is enabled, the device logs out the Web authentication users.

Enable online detection bypass for Web authentication users if you want to use online Web authentication user detection but retain the online users when all RADIUS servers are unreachable.

With online detection bypass enabled for Web authentication users, when all the RADIUS servers are unreachable, the device automatically disables online Web authentication user detection to keep the users online.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.

interface *interface-type interface-number*

3. Enable online detection bypass for Web authentication users.

   web-auth auth-server-unavailable escape

   By default, online detection bypass for Web authentication users is disabled when no RADIUS servers are reachable.

Configuring an Auth-Fail VLAN

Restrictions and guidelines

To make the Auth-Fail VLAN take effect, you must also enable MAC-based VLAN on the interface, and set the subnet of the Auth-Fail VLAN as the Web authentication-free subnet.

Because MAC-based VLAN takes effect only on Hybrid ports, Auth-Fail VLAN also takes effect only on Hybrid ports.

If a VLAN is specified as the super VLAN, do not configure the VLAN as an Auth-Fail VLAN of an interface. If a VLAN is specified as an Auth-Fail VLAN of an interface, do not configure the VLAN as a super VLAN.

Do not delete the VLAN that has been configured as an Auth-Fail VLAN. To delete this VLAN, first cancel the Auth-Fail VLAN configuration by using **undo web-auth auth-fail vlan** command.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure an Auth-Fail VLAN.

   web-auth auth-fail vlan *authfail-vlan-id*

   By default, no Auth-Fail VLAN is configured on an interface.

Configuring Web authentication multi-VLAN mode

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable the multi-VLAN mode for Web authentication on the port.

   web-auth host-mode multi-vlan

   By default, Web authentication operates in single-VLAN mode on a port.

Configuring Web authentication to support Web proxy

About this task

By default, proxied HTTP or HTTPS requests cannot trigger Web authentication but are silently dropped. To allow such HTTP or HTTPS requests to trigger Web authentication, specify the port numbers of the Web proxy servers on the device.

Restrictions and guidelines

If a user's browser uses the Web Proxy Auto-Discovery (WPAD) protocol to discover Web proxy servers, follow these restrictions and guidelines:

- Add the port numbers of the Web proxy servers on the device.
- Configure a Web authentication-free rule on the device to allow user packets destined for the IP address of the WPAD server to pass without authentication.
- Users must add the IP address of the Web server for Web authentication as a proxy exception in their browsers. Then, HTTP or HTTPS packets that the users send to the Web server for Web authentication will not be sent to Web proxy servers.

Procedure

1. Enter system view.

   system-view

2. Add a Web proxy server port number.

   web-auth proxy [**https**] **port** *port-number*

   By default, no Web proxy server port numbers are configured on the device.

   Do not specify port number 80 or 443 because 80 and 443 are port numbers reserved for Web authentication.

   Do not specify the same Web proxy server port number for HTTP and HTTPS.

Applying a portal MAC binding server for Web authentication

About this task

In Web authentication scenarios where valid users need to frequently access the network, you can configure MAC-based quick authentication to allow the users to pass Web authentication without manually entering the authentication information.

MAC-based quick authentication is also called MAC-trigger authentication or transparent authentication. It requires a portal MAC binding server.

When a user accesses the network, the access device sends the user's MAC address in a query message to the portal MAC binding server. If the portal MAC binding server finds user authentication information bound to the MAC address, it sends the user authentication information to the access device to initiate Web authentication. The user is authenticated without entering the username and password.

Restrictions and guidelines

This feature supports only IPv4 portal MAC binding servers, and the server type must be Oasis platform.

To make the MAC-based quick authentication take effect, complete the following tasks:

- Complete the normal Layer 2 Web authentication configuration.
- Configure the IP address and port number of the portal MAC binding server to be used.
- Apply the portal MAC binding server to a Layer 2 Ethernet interface or Layer 2 aggregate interface.

Prerequisites

Before applying a portal MAC binding server for Web authentication, deploy the portal MAC binding server on the network.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Apply a portal MAC binding server to the interface.
   web-auth apply portal mac-trigger-server *server-name*
   By default, no portal MAC binding server is applied for Web authentication.

Enabling Web authentication user logging

About this task

Web authentication user logs can help troubleshoot and solve issues. This feature enables the device to generate logs about Web authentication users and send the logs to the information center. For the logs to be output correctly, you must also configure the information center on the device. With the information center, you can configure log output destinations and output rules. For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Outputting a large number of Web authentication user logs might degrade device performance. As a best practice, enable Web authentication user logging only when necessary.

Procedure

1. Enter system view.
   system-view
2. Enable Web authentication user logging.
   web-auth access-user log enable [**abnormal-logoff** | **failed-login** | **normal-logoff** | **successful-login**] *
   By default, Web authentication user logging is disabled.

Display and maintenance commands for Web authentication

Execute **display** commands in any view and the **reset** command in user view.

Task	Command
Display Web authentication configuration information on interfaces.	**display web-auth** [**interface** *interface-type interface-number*]
Display Web authentication-free subnets.	**display web-auth free-ip**

Task	Command
Display information about Web servers for Web authentication.	**display web-auth server** [*server-name*]
Display Web authentication user information.	**display web-auth user** [**m-lag** [**local** | **peer**]] [**interface** *interface-type interface-number* | **slot** *slot-number*]
Log off Web authentication users.	**reset web-auth access-user** [**interface** *interface-type interface-number* | **mac** *mac-address* | **username** *username*]

Web authentication configuration examples

Example: Configuring local Web authentication by using the AAA local authentication method

Network configuration

As shown in Figure 5, the host is directly connected to the device through GigabitEthernet 1/0/1.

Configure local Web authentication on GigabitEthernet 1/0/1, and configure the device to perform local authentication and authorization for the local Web authentication users.

Configure a local Web server on the device that uses the IP address of Loopback 0 and port 80 for Web authentication. Configure the device to use HTTP to transfer the authentication data.

Figure 5 Network diagram

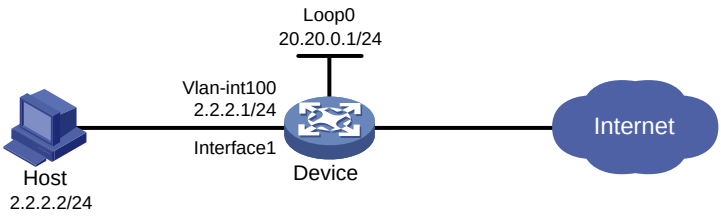


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

Create VLANs, assign IP addresses to the VLAN interfaces, and assign interfaces to the VLANs.

Procedure

1. Configure a local user:

   # Create a local network access user named **localuser**.

   ```
   <Device>system-view
   [Device] local-user localuser class network
   ```

   # Set the password for user **localuser.**

   ```
   [Device-luser-network-localuser] password simple 123456TESTplat&!
   ```

   # Authorize the user to use LAN access services.

```
[Device-luser-network-localuser] service-type lan-access
[Device-luser-network-localuser] quit
```

2. Configure an ISP domain:

Create an ISP domain named **local**.

```
[Device] domain local
```

Configure the ISP domain to perform local authentication, authorization, and accounting for LAN access users.

```
[Device-isp-local] authentication lan-access local
[Device-isp-local] authorization lan-access local
[Device-isp-local] accounting lan-access local
[Device-isp-local] quit
```

3. Configure a local portal Web service:

Create an HTTP-based local portal Web service and enter its view.

```
[Device] portal local-web-server http
```

Specify file **defaultfile.zip** as the default authentication page file for the local portal Web service. (Make sure the authentication page file already exists in the root directory of the storage medium on the device.)

```
[Device-portal-local-websvr-http] default-logon-page defaultfile.zip
```

Specify the HTTP listening port number as 80 for the portal Web service.

```
[Device–portal-local-websvr-http] tcp-port 80
[Device-portal-local-websvr-http] quit
```

4. Configure local Web authentication:

Create a local Web server named **user** for Web authentication and enter its view.

```
[Device] web-auth server user
```

Configure the redirection URL for the local Web server as **http://20.20.0.1:80/portal/**.

```
[Device-web-auth-server-user] url http://20.20.0.1:80/portal/
```

Specify the IP address of the local Web server as 20.20.0.1 (the IP address of Loopback 0) and the port number as 80.

```
[Device-web-auth-server-user] ip 20.20.0.1 port 80
[Device-web-auth-server-user] quit
```

Specify ISP domain **local** as the Web authentication domain.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] web-auth domain local
```

Enable Web authentication that uses local Web server **user**.

```
[Device-GigabitEthernet1/0/1] web-auth enable apply server user
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Display online Web authentication user information after user **localuser** passes Web authentication.

```
[Device] display web-auth user
  Total online web-auth users: 1

User Name: localuser
  MAC address: acf1-df6c-f9ad
  Access interface: GigabitEthernet1/0/1
  Initial VLAN: 100
  Authorization VLAN: N/A
```

```
Authorization ACL ID: N/A
Authorization user profile: N/A
Authorization microsegment ID: N/A
```

Example: Configuring local Web authentication by using the AAA remote authentication method

Network configuration

As shown in Figure 6, the host is directly connected to the device through GigabitEthernet 1/0/1.

Configure local Web authentication on GigabitEthernet 1/0/1, and use a remote RADIUS server to perform authentication, authorization, and accounting for the local Web authentication users.

Configure a local Web server on the device that uses the IP address of Loopback 0 and port 80 for Web authentication. Configure the device to use HTTP to transfer the authentication data.

Figure 6 Network diagram

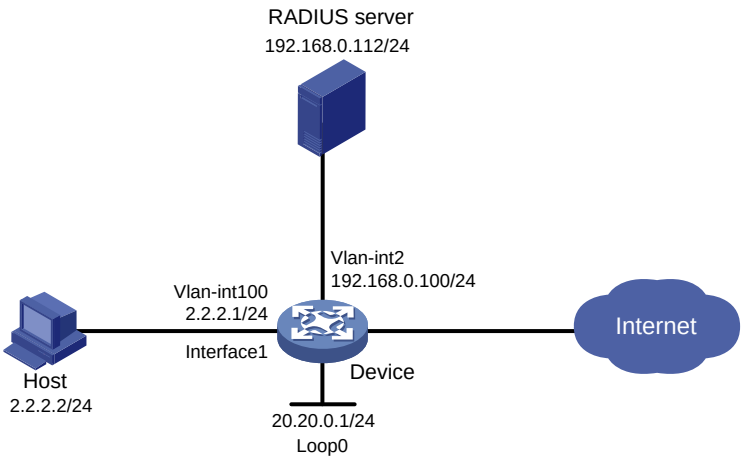


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

- Configure the RADIUS server properly to provide authentication, authorization, and accounting functions for users. In this example, the username is configured as **user1** on the RADIUS server.
- Create VLANs, assign IP addresses to the VLAN interfaces, and assign interfaces to the VLANs. Make sure the host, the RADIUS server, and the device can reach each other.

Procedure

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1**.

```
<Device> system-view
[Device] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Device-radius-rs1] primary authentication 192.168.0.112
```

```
[Device-radius-rs1] primary accounting 192.168.0.112
[Device-radius-rs1] key authentication simple radius
[Device-radius-rs1] key accounting simple radius
```

Exclude the ISP domain name from the username sent to the RADIUS server.

```
[Device-radius-rs1] user-name-format without-domain
[Device-radius-rs1] quit
```

2. Configure an authentication domain:

Create an ISP domain named **dm1**.

```
[Device] domain dm1
```

Configure AAA methods for the ISP domain

```
[Device-isp-dm1] authentication lan-access radius-scheme rs1
[Device-isp-dm1] authorization lan-access radius-scheme rs1
[Device-isp-dm1] accounting lan-access radius-scheme rs1
[Device-isp-dm1] quit
```

3. Configure a local portal Web service:

Create an HTTP-based local portal Web service.

```
[Device] portal local-web-server http
```

Specify the file **defaultfile.zip** as the default authentication page file for the local portal Web service. (Make sure the authentication page file already exists in the root directory of the storage medium on the device.)

```
[Device-portal-local-websvr-http] default-logon-page defaultfile.zip
```

Specify 80 as the port number listened by the local portal Web service.

```
[Device–portal-local-websvr-http] tcp-port 80
[Device-portal-local-websvr-http] quit
```

4. Configure local Web authentication:

Create a local Web server named **user** for Web authentication and enter its view.

```
[Device] web-auth server user
```

Specify **http://20.20.0.1:80/portal/** as the redirection URL for the local Web server.

```
[Device-web-auth-server-user] url http://20.20.0.1:80/portal/
```

Specify the IP address of the local Web server as 20.20.0.1 (the IP address of Loopback 0) and the port number as 80.

```
[Device-web-auth-server-user] ip 20.20.0.1 port 80
[Device-web-auth-server-user] quit
```

Specify domain **dml** as the Web authentication domain.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] web-auth domain dm1
```

Enable Web authentication that uses local Web server **user**.

```
[Device-GigabitEthernet1/0/1] web-auth enable apply server user
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Display Web authentication user information after user **user1** passes Web authentication.

```
[Device] display web-auth user
  Total online web-auth users: 1

User Name: user1
  MAC address: acf1-df6c-f9ad
  Access interface: GigabitEthernet1/0/1
```

```
Initial VLAN: 100
Authorization VLAN: N/A
Authorization ACL ID: N/A
Authorization user profile: N/A
Authorization microsegment ID: N/A
```

Example: Configuring remote Web authentication by using the AAA local authentication method

Network configuration

As shown in Figure 7, the host is directly connected to the device through GigabitEthernet 1/0/1.

Configure remote Web authentication on GigabitEthernet 1/0/1, and configure the device to perform local authentication and authorization for the remote Web authentication users.

The remote Web server pushes Web authentication pages to users.

Figure 7 Network diagram

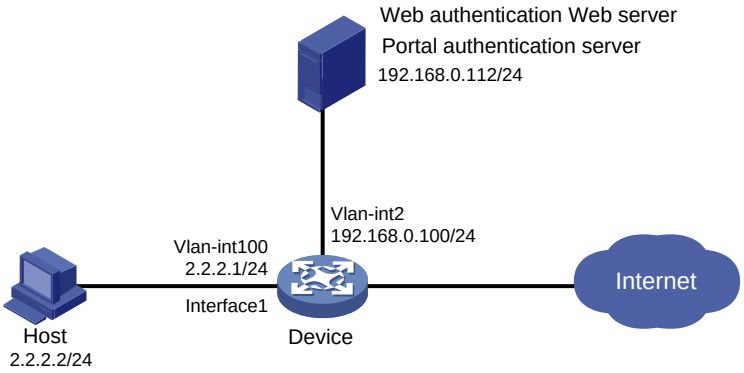


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

- Configure the portal authentication server properly to provide normal portal functions for users.
- Create VLANs, assign IP addresses to the VLAN interfaces, and assign interfaces to the VLANs. Make sure the host, the RADIUS server, and the device can reach each other.

Procedure

1. Configure a local user:

   # Create a local network access user named **localuser**.

   ```
   <Device>system-view
   [Device] local-user localuser class network
   ```

   # Set the password for user **localuser.**

   ```
   [Device-luser-network-localuser] password simple 123456TESTplat&!
   ```

   # Authorize the user to use LAN access services.

   ```
   [Device-luser-network-localuser] service-type lan-access
   [Device-luser-network-localuser] quit
   ```

2. Configure an ISP domain:

   # Create an ISP domain named **local**.

   ```
   [Device] domain local
   ```

   # Configure the ISP domain to perform local authentication, authorization, and accounting for LAN access users.

   ```
   [Device-isp-local] authentication lan-access local
   [Device-isp-local] authorization lan-access local
   [Device-isp-local] accounting lan-access local
   [Device-isp-local] quit
   ```

3. Configure the portal authentication server:

   ```
   [Device] portal server newpt
   [Device-portal-server-newpt] ip 192.168.0.112 key simple portal
   [Device-portal-server-newpt] port 50100
   [Device-portal-server-newpt] quit
   ```

4. Configure remote Web authentication:

   # Create a remote Web server named **user** for Web authentication and enter its view.

   ```
   [Device] web-auth remote server user
   ```

   # Configure the redirection URL of the remote Web server as **https://192.168.0.112:8080/portal/**.

   ```
   [Device-web-auth-remote-server-user] url https://192.168.0.112:8080/portal/
   ```

   # Configure the IP address of the remote Web server as 192.168.0.112.

   ```
   [Device-web-auth-remote-server-user] ip 192.168.0.112
   [Device-web-auth-remote-server-user] quit
   ```

   # Specify an authentication domain for Web authentication users.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] web-auth domain local
   ```

   # Enable Web authentication that uses remote Web server **user**.

   ```
   [Device-GigabitEthernet1/0/1] web-auth enable apply server user
   [Device-GigabitEthernet1/0/1] quit
   ```

5. Configure the device to obtain user information from ARP entries.

   ```
   [Device] portal access-info trust arp
   ```

Verifying the configuration

Display Web authentication user information after user **user1** passes Web authentication.

```
[Device] display web-auth user
  Total online web-auth users: 1

User Name: user1
  MAC address: acf1-df6c-f9ad
  Access interface: GigabitEthernet1/0/1
  Initial VLAN: 100
  Authorization VLAN: N/A
  Authorization ACL ID: N/A
  Authorization user profile: N/A
  Authorization microsegment ID: N/A
```

Example: Configuring remote Web authentication by using the AAA remote authentication method

Network configuration

As shown in Figure 8, the host is directly connected to the device through GigabitEthernet 1/0/1.

Configure remote Web authentication on GigabitEthernet 1/0/1, and use a remote RADIUS server to perform authentication, authorization, and accounting for the remote Web authentication users.

The remote Web server pushes Web authentication pages to users.

Figure 8 Network diagram

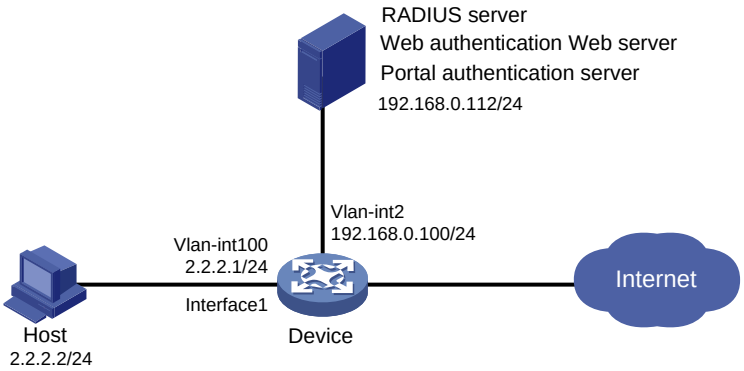


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

- Configure the RADIUS server properly to provide authentication, authorization, and accounting functions for users. In this example, the username is configured as **user1** on the RADIUS server.
- Configure the portal authentication server properly to provide normal portal functions for users.
- Create VLANs, assign IP addresses to the VLAN interfaces, and assign interfaces to the VLANs. Make sure the host, the RADIUS server, and the device can reach each other.

Procedure

1. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1**.

```
<Device> system-view
[Device] radius scheme rs1
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Device-radius-rs1] primary authentication 192.168.0.112
[Device-radius-rs1] primary accounting 192.168.0.112
[Device-radius-rs1] key authentication simple radius
[Device-radius-rs1] key accounting simple radius
```

Exclude the ISP domain name from the username sent to the RADIUS server.

```
[Device-radius-rs1] user-name-format without-domain
```

```
[Device-radius-rs1] quit
```

2. Configure an authentication domain:

   # Create an ISP domain named **dm1**.

   ```
   [Device] domain dm1
   ```

   # Configure AAA methods for the ISP domain

   ```
   [Device-isp-dm1] authentication lan-access radius-scheme rs1
   [Device-isp-dm1] authorization lan-access radius-scheme rs1
   [Device-isp-dm1] accounting lan-access radius-scheme rs1
   [Device-isp-dm1] quit
   ```

3. Configure the portal authentication server:

   ```
   [Device] portal server newpt
   [Device-portal-server-newpt] ip 192.168.0.112 key simple portal
   [Device-portal-server-newpt] port 50100
   [Device-portal-server-newpt] quit
   ```

4. Configure remote Web authentication:

   # Create a remote Web server named **user** for Web authentication and enter its view.

   ```
   [Device] web-auth remote server user
   ```

   # Configure the redirection URL of the remote Web server as **https://192.168.0.112:8080/portal/**.

   ```
   [Device-web-auth-remote-server-user] url https://192.168.0.112:8080/portal/
   ```

   # Configure the IP address of the remote Web server as 192.168.0.112.

   ```
   [Device-web-auth-remote-server-user] ip 192.168.0.112
   [Device-web-auth-remote-server-user] quit
   ```

   # Specify an authentication domain for Web authentication users.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] web-auth domain dm1
   ```

   # Enable Web authentication that uses remote Web server **user**.

   ```
   [Device-GigabitEthernet1/0/1] web-auth enable apply server user
   [Device-GigabitEthernet1/0/1] quit
   ```

5. Configure the device to obtain user information from ARP entries.

   ```
   [Device] portal access-info trust arp
   ```

Verifying the configuration

Display Web authentication user information after user **user1** passes Web authentication.

```
[Device] display web-auth user
  Total online web-auth users: 1

User Name: user1
  MAC address: acf1-df6c-f9ad
  Access interface: GigabitEthernet1/0/1
  Initial VLAN: 100
  Authorization VLAN: N/A
  Authorization ACL ID: N/A
  Authorization user profile: N/A
  Authorization microsegment ID: N/A
```

Troubleshooting Web authentication

Failure to come online (local authentication interface using the default ISP domain

Symptom

No authentication domain is specified for the local authentication interface. A user fails to pass Web authentication to come online.

Analysis

If no Web authentication domain is specified, the system default ISP domain (domain **system**) is used for Web authentication. The system default domain uses the local authentication method by default. Using these default domain settings, the local authentication should have operated correctly.

The local authentication fails might because that the authentication method of the system default domain is changed or the system default domain is changed.

Solution

To resolve the problem, perform the following tasks:

1. Use the **`display domain`** command to identify whether the AAA methods for Web users in the system default domain are local.
2. If the AAA methods for Web users in the system default domain are not local, reconfigure the AAA methods as local.

Contents

Configuring triple authentication 1

About triple authentication 1
Typical network of triple authentication 1
Triple authentication mechanism 1
Triple authentication support for VLAN assignment 2
Triple authentication support for ACL authorization 2
Triple authentication support for online user detection 3
Restrictions and guidelines: Triple authentication 3
Triple authentication tasks at a glance 3
Triple authentication configuration examples 3
Example: Configuring basic triple authentication 3
Example: Configuring triple authentication to support authorization VLAN and authentication failure VLAN 7

Configuring triple authentication

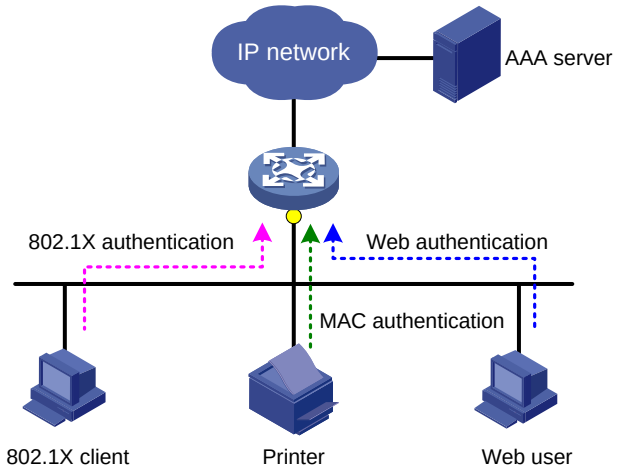
About triple authentication

Triple authentication enables an access port to perform Web, MAC, and 802.1X authentication. A terminal can access the network if it passes one type of authentication. For more information about 802.1X authentication, MAC authentication, and Web authentication, see "Configuring 802.1X authentication", "Configuring MAC authentication", and "Configuring Web authentication."

Typical network of triple authentication

Triple authentication is suitable for a LAN that comprises terminals that require different authentication services, as shown in Figure 1. The triple authentication-enabled access port can perform MAC authentication for the printer, 802.1X authentication for the PC installed with the 802.1X client, and Web authentication for the Web user.

Figure 1 Triple authentication network diagram



Triple authentication mechanism

The three types of authentication are triggered by different packets:

- The access port performs MAC authentication for a terminal when it receives an ARP or DHCP broadcast packet from the terminal for the first time. If the terminal passes MAC authentication, the terminal can access the network. If the MAC authentication fails, the access port performs 802.1X or Web authentication.
- The access port performs 802.1X authentication when it receives an EAP packet from an 802.1X client or a third-party client. If the unicast trigger feature of 802.1X is enabled on the access port, any packet from the client can trigger an 802.1X authentication.
- The access port performs Web authentication when it receives an HTTP packet from a terminal.

If a terminal triggers different types of authentication, the authentications are processed at the same time. The failure of one type of authentication does not affect the others. When a terminal passes one type of authentication, the other types of authentication are processed as follows:

- If the terminal first passes MAC authentication, Web authentication is terminated immediately, but 802.1X authentication will proceed. If the terminal also passes 802.1X authentication, the 802.1X authentication information will overwrite the MAC authentication information for the

terminal. If the terminal fails 802.1X authentication, the user stays online as a MAC authentication user, and only 802.1X authentication can be triggered again.

- If the terminal first passes 802.1X or Web authentication, the other types of authentication are terminated immediately and cannot be triggered again.

Triple authentication support for VLAN assignment

Authorization VLAN

After a user passes authentication, the authentication server assigns an authorization VLAN to the access port for the user. The user can then access the network resources in the authorized VLAN.

Authentication failure VLAN

The access port adds a user to an authentication failure VLAN configured on the port after the user fails authentication.

- For an 802.1X authentication user—Adds the user to the Auth-Fail VLAN configured for 802.1X authentication.
- For a Web authentication user—Adds the user to the Auth-Fail VLAN configured for Web authentication.
- For a MAC authentication user—Adds the user to the guest VLAN configured for MAC authentication.

If a user in the MAC authentication guest VLAN fails 802.1X authentication, the user is moved to the 802.1X Auth-Fail VLAN.

Server-unreachable VLAN

If a user fails authentication due to the unreachable server, the access port adds the user to an server-unreachable VLAN.

- For an 802.1X authentication user—Adds the user to the critical VLAN configured for 802.1X authentication.
- For a Web authentication user—Adds the user to the Auth-Fail VLAN configured for Web authentication.
- For a MAC authentication user—Adds the user to the critical VLAN configured for MAC authentication.

The access port supports configuring all types of server-unreachable VLANs at the same time. A user is added to the server-unreachable VLAN as follows:

- If the user does not undergo 802.1X authentication, the user is added to the server-unreachable VLAN configured for the last authentication.
- If the user in the Web Auth-Fail VLAN or MAC authentication critical VLAN also fails 802.1X authentication, the user is added to the 802.1X authentication critical VLAN.

Triple authentication support for ACL authorization

After a user passes authentication, the authentication server assigns an authorization ACL to the access port for the user. The access port uses the ACL to filter traffic for the user.

To use ACL authorization, you must specify authorization ACLs on the authentication server and configure the ACLs on the access device. You can change the user's access authorization by changing the authorization ACL on the authentication server or changing rules of the authorization ACL on the access device.

Triple authentication support for online user detection

You can configure the following features to detect the online status of users:

- Enable online user detection for Web authentication users.
- Enable the online user handshake or periodic online user reauthentication feature for 802.1X users.
- Enable offline detection for MAC authentication users.

Restrictions and guidelines: Triple authentication

In triple authentication, 802.1X authentication must use the MAC-based access control method.

If Web authentication is enabled on a port, configure the subnets of the server-unreachable VLANs of the port as Web authentication-free subnets. This ensures that an authentication-failed user can access the server-unreachable VLAN.

Do not configure both Web authentication-free IPs and 802.1X free IPs. If you do so, only 802.1X free IPs take effect.

Web authentication cannot be configured in conjunction with the following settings:

- Guest VLAN, Auth-Fail VLAN, and critical VLAN of 802.1 X authentication.
- Guest VLAN and critical VLAN of MAC authentication.

Triple authentication tasks at a glance

Choose the following tasks as needed:

- Configure 802.1X authentication

  For more information, see "Configuring 802.1X."
- Configure MAC authentication

  For more information, see "Configuring MAC authentication."
- Configure Web authentication

  For more information, see "Configuring Web authentication."

In triple authentication, if the above three authentication methods are configured, use the **`port-security triple-auth-order mac-dot1x-web`** command to specify the authentication order as MAC authentication, 802.1X authentication, and Web authentication. For more information about the **`port-security triple-auth-order mac-dot1x-web`** command, see port security commands in *Security Command Reference*.

Triple authentication configuration examples

Example: Configuring basic triple authentication

Network configuration

As shown in Figure 2, the terminals are connected to the device to access the IP network. Configure triple authentication on the device's Layer 2 interface that connects to the terminals. A terminal passing one of the three authentication methods, 802.1X authentication, Web authentication, and MAC authentication, can access the IP network.

- Assign IP addresses on subnet 192.168.1.0/24 to the terminals.

- Use the remote RADIUS server to perform authentication, authorization, and accounting. Configure the device to send usernames carrying no ISP domain names to the RADIUS server.
- Configure the local Web authentication server on the device to use listening IP address 4.4.4.4. Configure the device to send a default authentication page to the Web user and forward authentication data by using HTTP.

Figure 2 Network diagram

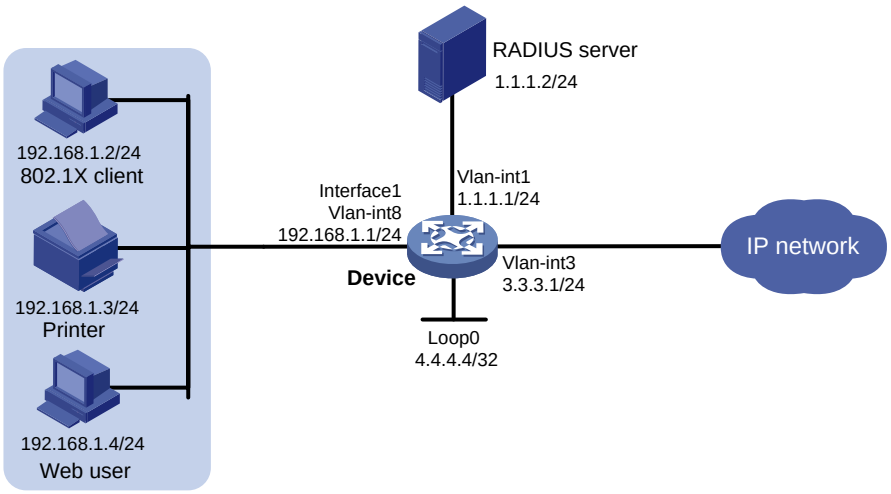


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Make sure that the terminals, the server, and the device can reach each other. (Details not shown.)
2. Configure the RADIUS server to provide normal authentication, authorization, and accounting for users. In this example, configure the following on the RADIUS server:
   - An 802.1X user with username **userdot**.
   - A Web authentication user with username **userpt**.
   - A MAC authentication user with a username and password both being the MAC address of the printer **f07d6870725f**.
3. Configure Web authentication:

   # Configure VLANs and IP addresses for the VLAN interfaces, and add ports to specific VLANs. (Details not shown.)

   # Create an HTTP-based local portal Web service, and specify file **defaultfile.zip** as the default authentication page file of the local portal Web service. (Make sure the authentication page file already exists in the root directory of the storage medium on the device.)

   ```
   <Device> system-view
   [Device] portal local-web-server http
   [Device-portal-local-websvr-http] default-logon-page defaultfile.zip
   [Device-portal-local-websvr-http] quit
   ```

   # Assign IP address 4.4.4.4 to Loopback 0.

   ```
   [Device] interface loopback 0
   [Device-LoopBack0] ip address 4.4.4.4 32
   [Device-LoopBack0] quit
   ```

Create a Web authentication server named **webserver** and enter its view.

```
[Device] web-auth server webserver
```

Configure the redirection URL for the Web authentication server as **https://4.4.4.4/portal/**.

```
[Device-web-auth-server-webserver] url https://4.4.4.4/portal/
```

Specify 4.4.4.4 as the IP address and 80 as the port number of the Web authentication server.

```
[Device-web-auth-server-webserver] ip 4.4.4.4 port 80
[Device-web-auth-server-webserver] quit
```

Enable Web authentication on GigabitEthernet 1/0/1 and specify Web authentication server **webserver** on the interface.

```
[Device] interface gigabitethernet 1/0/1
[Device–GigabitEthernet1/0/1] web-auth enable apply server webserver
[Device–GigabitEthernet1/0/1] quit
```

4. Configure 802.1X authentication:

Enable 802.1X authentication globally.

```
[Device] dot1x
```

Enable 802.1X authentication (MAC-based access control required) on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device–GigabitEthernet1/0/1] dot1x port-method macbased
[Device–GigabitEthernet1/0/1] dot1x
[Device–GigabitEthernet1/0/1] quit
```

5. Configure MAC authentication:

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Enable MAC authentication on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device–GigabitEthernet1/0/1] mac-authentication
[Device–GigabitEthernet1/0/1] quit
```

6. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1**.

```
[Device] radius scheme rs1
```

Specify the primary authentication and accounting servers and keys.

```
[Device-radius-rs1] primary authentication 1.1.1.2
[Device-radius-rs1] primary accounting 1.1.1.2
[Device-radius-rs1] key authentication simple radius
[Device-radius-rs1] key accounting simple radius
```

Specify usernames sent to the RADIUS server to carry no domain names.

```
[Device-radius-rs1] user-name-format without-domain
[Device-radius-rs1] quit
```

7. Configure an ISP domain:

Create an ISP domain named **triple**.

```
[Device] domain triple
```

Configure the domain to use RADIUS scheme **rs1** for authentication, authorization and accounting of LAN access users.

```
[Device-isp-triple] authentication lan-access radius-scheme rs1
[Device-isp-triple] authorization lan-access radius-scheme rs1
[Device-isp-triple] accounting lan-access radius-scheme rs1
```

```
[Device-isp-triple] quit
```

Configure domain **triple** as the default domain. If a username entered by a user includes no ISP domain name, the AAA method of the default domain is used.

```
[Device] domain default enable triple
```

Verifying the configuration

1. Verify that the Web user can pass Web authentication.

   # On the Web user terminal, use a Web browser to access an external network and then enter the correct username and password on the authentication page **https://4.4.4.4/portal/logon.html**. (Details not shown.)

   # Display information about online Web authentication users.

```
[Device] display web-auth user
  Total online web-auth users: 1
User Name: localuser
  MAC address: acf1-df6c-f9ad
  Access interface: GigabitEthernet1/0/1
  Initial VLAN: 8
  Authorization VLAN: N/A
  Authorization ACL ID: N/A
  Authorization user profile: N/A
```

2. Verify that the printer can pass MAC authentication.

   # Connect the printer to the network. (Details not shown.)

   # Display information about online MAC authentication users.

```
[Device] display mac-authentication connection
Total connections: 1
Slot ID: 1
User MAC address: f07d-6870-725f
Access interface: GigabitEthernet1/0/1
Username: f07d6870725f
User access state: Successful
Authentication domain: triple
Initial VLAN: 8
Authorization untagged VLAN: N/A
Authorization tagged VLAN: N/A
Authorization VSI: N/A
Authorization ACL ID: N/A
Authorization user profile: N/A
Authorization CAR: N/A
Authorization URL: N/A
Termination action: Default
Session timeout period: N/A
Online from: 2015/01/04 18:01:43
Online duration: 0h 0m 2s
```

3. Verify that the 802.1X client can pass 802.1X authentication.

   # On the 802.1X client, initiate 802.1X authentication and then enter the correct username and password. (Details not shown.)

   # Display information about online 802.1X users.

```
[Device] display dot1x connection
Total connections: 1
```

```
Slot ID: 1
User MAC address: 7446-a091-84fe
Access interface: GigabitEthernet1/0/1
Username: userdot
User access state: Successful
Authentication domain: triple
IPv4 address: 192.168.1.2
Authentication method: CHAP
Initial VLAN: 8
Authorization untagged VLAN: N/A
Authorization tagged VLAN list: N/A
Authorization VSI: N/A
Authorization ACL ID: N/A
Authorization user profile: N/A
Authorization CAR: N/A
Authorization URL: N/A
Termination action: Default
Session timeout period: N/A
Online from: 2015/01/04 18:13:01
Online duration: 0h 0m 14s
```

Example: Configuring triple authentication to support authorization VLAN and authentication failure VLAN

Network configuration

As shown in Figure 3, the terminals are connected to the device to access the IP network. Configure triple authentication on the device's Layer 2 interface connected to the terminals. A terminal passing one of the three authentication methods, 802.1X authentication, Web authentication, and MAC authentication, can access the IP network.

- The Web authentication terminal uses DHCP to get an IP address in 192.168.1.0/24 before authentication and in 3.3.3.0/24 after passing authentication. If the terminal fails authentication, it requests IP addresses in 2.2.2.0/24 through DHCP.

  You can use the access device or an attached device as the DHCP server. In this example, the access device (the device) provides the DHCP service.
- The 802.1X terminal uses DHCP to get an IP address in 192.168.1.0/24 before authentication and in 3.3.3.0/24 after passing authentication. If the terminal fails authentication, it requests IP addresses in 2.2.2.0/24 through DHCP.
- After passing authentication, the printer obtains IP address 3.3.3.111/24 that is bound with its MAC address through DHCP.
- Use the remote RADIUS server to perform authentication, authorization, and accounting. Configure the device to remove the ISP domain names from usernames sent to the RADIUS server.
- Configure the local Web authentication server on the device to use listening IP address 4.4.4.4. Configure the device to forward authentication data by using HTTP.
- Configure VLAN 3 as the authorization VLAN. Users passing authentication are added to this VLAN.
- Configure VLAN 2 as the authentication failure VLAN. Users failing authentication are added to this VLAN.

Figure 3 Network diagram

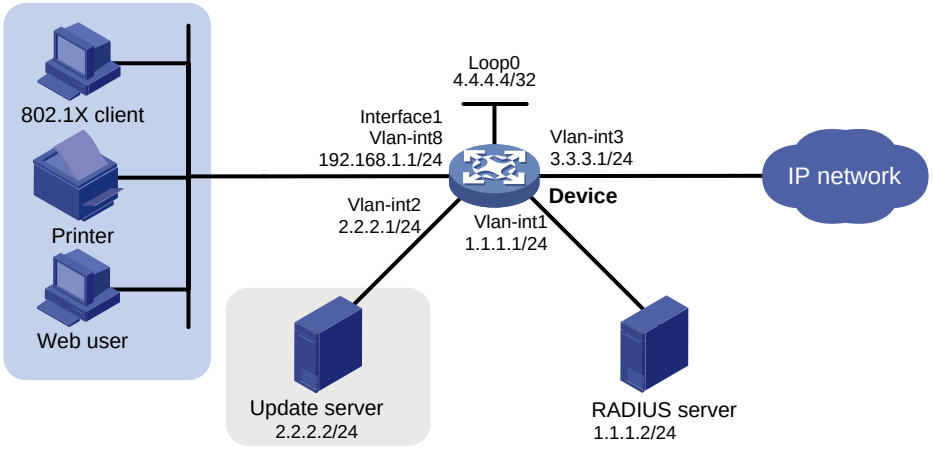


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Make sure the terminals, the servers, and the device can reach each other. (Details not shown.)
2. Configure the RADIUS server to provide normal authentication, authorization, and accounting for users. In this example, configure the following on the RADIUS server:
   - An 802.1X user with username **userdot**.
   - A Web authentication user with username **userpt**.
   - A MAC authentication user with a username and password both being the MAC address of the printer **f07d6870725f**.
   - An authorization VLAN (VLAN 3).
3. Configure the IP address of the update server as an authentication-free IP address.

   ```
   <Device> system-view
   [Device] web-auth free-ip 2.2.2.2 24
   ```
4. Configure DHCP:

   # Configure VLANs and IP addresses for the VLAN interfaces, and add ports to specific VLANs. (Details not shown.)

   # Enable DHCP.

   ```
   [Device] dhcp enable
   ```

   # Exclude the IP address of the update server from dynamic address assignment.

   ```
   [Device] dhcp server forbidden-ip 2.2.2.2
   ```

   # Configure DHCP address pool 1 to assign IP addresses and other configuration parameters to clients on subnet 192.168.1.0.

   ```
   [Device] dhcp server ip-pool 1
   [Device-dhcp-pool-1] network 192.168.1.0 mask 255.255.255.0
   [Device-dhcp-pool-1] expired day 0 hour 0 minute 1
   [Device-dhcp-pool-1] gateway-list 192.168.1.1
   [Device-dhcp-pool-1] quit
   ```

   # Configure DHCP address pool 2 to assign IP address and other configuration parameters to clients on subnet 2.2.2.0.

```
[Device] dhcp server ip-pool 2
[Device-dhcp-pool-2] network 2.2.2.0 mask 255.255.255.0
[Device-dhcp-pool-2] expired day 0 hour 0 minute 1
[Device-dhcp-pool-2] gateway-list 2.2.2.1
[Device-dhcp-pool-2] quit
```

Configure DHCP address pool 3 to assign IP address and other configuration parameters to clients on subnet 3.3.3.0.

```
[Device] dhcp server ip-pool 3
[Device-dhcp-pool-3] network 3.3.3.0 mask 255.255.255.0
[Device-dhcp-pool-3] expired day 0 hour 0 minute 1
[Device-dhcp-pool-3] gateway-list 3.3.3.1
[Device-dhcp-pool-3] quit
```

Configure DHCP address pool 4, and bind the printer's MAC address f07d-6870-725f to IP address 3.3.3.111/24 in this address pool.

```
[Device] dhcp server ip-pool 4
[Device-dhcp-pool-4] static-bind ip-address 3.3.3.111 mask 255.255.255.0
client-identifier f07d-6870-725f
[Device-dhcp-pool-4] quit
```

5. Configure Web authentication:

Create an HTTP-based local portal Web service, and specify file **defaultfile.zip** as the default authentication page file of the local portal Web service. (Make sure the authentication page file already exists in the root directory of the storage medium on the device.)

```
[Device] portal local-web-server http
[Device-portal-local-websvr-http] default-logon-page defaultfile.zip
[Device-portal-local-websvr-http] quit
```

Assign IP address 4.4.4.4 to Loopback 0.

```
[Device] interface loopback 0
[Device-LoopBack0] ip address 4.4.4.4 32
[Device-LoopBack0] quit
```

Create a Web authentication server named **webserver**.

```
[Device] web-auth server webserver
```

Configure the redirection URL of the Web authentication server as **https://4.4.4.4/portal/**.

```
[Device-web-auth-server-webserver] url https://4.4.4.4/portal/
```

Specify 4.4.4.4 as the IP address and 80 as the port number of the Web authentication server.

```
[Device-web-auth-server-webserver] ip 4.4.4.4 port 80
[Device-web-auth-server-webserver] quit
```

Configure the IP address of the update server as an authentication-free IP address.

```
[Device] web-auth free-ip 2.2.2.2 24
```

Enable Web authentication on GigabitEthernet 1/0/1 and specify VLAN 2 as the Auth-Fail VLAN.

```
[Device] interface gigabitethernet 1/0/1
[Device–GigabitEthernet1/0/1] port link-type hybrid
[Device–GigabitEthernet1/0/1] mac-vlan enable
[Device–GigabitEthernet1/0/1] web-auth enable apply server webserver
[Device–GigabitEthernet1/0/1] web-auth auth-fail vlan 2
[Device–GigabitEthernet1/0/1] quit
```

6. Configure 802.1X authentication:

Enable 802.1X authentication globally.

```
[Device] dot1x
```

Enable 802.1X authentication (MAC-based access control required) on GigabitEthernet 1/0/1, and specify VLAN 2 as the Auth-Fail VLAN.

```
[Device] interface gigabitethernet 1/0/1
[Device–GigabitEthernet1/0/1] dot1x port-method macbased
[Device–GigabitEthernet1/0/1] dot1x
[Device–GigabitEthernet1/0/1] dot1x auth-fail vlan 2
[Device–GigabitEthernet1/0/1] quit
```

7. Configure MAC authentication:

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Enable MAC authentication on GigabitEthernet 1/0/1, and specify VLAN 2 as the guest VLAN.

```
[Device] interface gigabitethernet 1/0/1
[Device–GigabitEthernet1/0/1] mac-authentication
[Device–GigabitEthernet1/0/1] mac-authentication guest-vlan 2
[Device–GigabitEthernet1/0/1] quit
```

8. Configure a RADIUS scheme:

Create a RADIUS scheme named **rs1**.

```
[Device] radius scheme rs1
```

Specify the primary authentication and accounting servers and keys.

```
[Device-radius-rs1] primary authentication 1.1.1.2
[Device-radius-rs1] primary accounting 1.1.1.2
[Device-radius-rs1] key authentication simple radius
[Device-radius-rs1] key accounting simple radius
```

Specify usernames sent to the RADIUS server to carry no domain names.

```
[Device-radius-rs1] user-name-format without-domain
[Device-radius-rs1] quit
```

9. Configure an ISP domain:

Create an ISP domain named **triple**.

```
[Device] domain triple
```

Configure the domain to use RADIUS scheme **rs1** for authentication, authorization and accounting of LAN access users.

```
[Device-isp-triple] authentication lan-access radius-scheme rs1
[Device-isp-triple] authorization lan-access radius-scheme rs1
[Device-isp-triple] accounting lan-access radius-scheme rs1
[Device-isp-triple] quit
```

Configure domain **triple** as the default domain. If a username entered by a user includes no ISP domain name, the AAA methods of the default domain is used.

```
[Device] domain default enable triple
```

Verifying the configuration

1. Verify that the Web user can pass Web authentication.

On the Web user terminal, use a Web browser to access an external network and then enter the correct username and password on the authentication page **https://4.4.4.4/portal/logon.html**. (Details not shown.)

Use the **display web-auth user** command to display information about online users.

```
[Device] display web-auth user
  Total online web-auth users: 1
```

```
User Name: userpt
  MAC address: 6805-ca17-4a0b
  Access interface: GigabitEthernet1/0/1
  Initial VLAN: 8
  Authorization VLAN: 3
  Authorization ACL ID: N/A
  Authorization user profile: N/A
```

2. Verify that the printer can pass MAC authentication.

Connect the printer to the network. (Details not shown.)

Display information about online MAC authentication users.

```
[Device] display mac-authentication connection
Total connections: 1
Slot ID: 1
User MAC address: f07d-6870-725f
Access interface: GigabitEthernet1/0/1
Username: f07d6870725f
User access state: Successful
Authentication domain: triple
Initial VLAN: 8
Authorization untagged VLAN: 3
Authorization tagged VLAN: N/A
Authorization VSI: N/A
Authorization ACL ID: N/A
Authorization user profile: N/A
Authorization CAR: N/A
Authorization URL: N/A
Termination action: Default
Session timeout period: N/A
Online from: 2015/01/04 18:01:43
Online duration: 0h 0m 2s
```

3. Verify that the 802.1X user can pass 802.1X authentication.

On the 802.1X client, initiate 802.1X authentication and enter the correct username and password. (Details not shown.)

Display information about online 802.1X users.

```
[Device] display dot1x connection
Total connections: 1
Slot ID: 1
User MAC address: 7446-a091-84fe
Access interface: GigabitEthernet1/0/1
Username: userdot
User access state: Successful
Authentication domain: triple
IPv4 address: 3.3.3.3
Authentication method: CHAP
Initial VLAN: 8
Authorization untagged VLAN: 3
Authorization tagged VLAN list: N/A
```

```
Authorization VSI: N/A
Authorization ACL ID: N/A
Authorization user profile: N/A
Authorization CAR: N/A
Authorization URL: N/A
Termination action: Default
Session timeout period: N/A
Online from: 2015/01/04 18:13:01
Online duration: 0h 0m 14s
```

4. Verify that users that pass authentication have been assigned authorization VLANs.

Display MAC-VLAN entries of online users.

```
[Device] display mac-vlan all
  The following MAC VLAN addresses exist:
  S:Static  D:Dynamic
  MAC ADDR         MASK             VLAN ID   PRIO   STATE
  ------------------------------------------------------------
  6805-ca17-4a0b   ffff-ffff-ffff   3         0      D
  f07d-6870-725f   ffff-ffff-ffff   3         0      D
  7446-a091-84fe   ffff-ffff-ffff   3         0      D
  Total MAC VLAN address count:3
```

5. Verify that online users have been assigned IP addresses.

```
[Device] display dhcp server ip-in-use
IP address       Client-identifier/    Lease expiration          Type
                  Hardware address
 3.3.3.111        01f0-7d68-7072-5f     Jan  4 18:14:17 2015      Auto:(C)
 3.3.3.2          0168-05ca-174a-0b     Jan  4 18:15:01 2015      Auto:(C)
 3.3.3.3          0174-46a0-9184-fe     Jan  4 18:15:03 2015      Auto:(C)
```

6. When a terminal fails authentication, it is added to VLAN 2. You can use the previous display commands to display the MAC-VLAN entry and IP address of the terminal. (Details not shown.)

Contents

Configuring port security 1

About port security 1
Major port security functions 1
Port security modes 2
Secure MAC addresses 6
Authentication load sharing modes in M-LAG environments 7
Port security application scenarios 8
Restrictions and guidelines: Port security configuration 9
Port security tasks at a glance 10
Enabling port security 11
Setting the port security mode 11
Setting port security's limit on the number of secure MAC addresses on a port 13
Configuring secure MAC addresses 13
About secure MAC addresses 13
Restrictions and guidelines 14
Prerequisites 14
Adding a static secure MAC address 14
Adding a sticky MAC address 14
Converting sticky MAC addresses to dynamic secure MAC addresses 15
Configuring secure MAC address aging 15
Configuring NTK 16
Configuring intrusion protection 16
Configuring an authentication load sharing mode for users attached to M-LAG interfaces 17
Ignoring authorization information from the server 18
Configuring MAC move 19
Enabling overwrite of local MAC entries with remote MAC entries for MAC addresses moved to a remote device 20
Enabling the authorization-fail-offline feature 20
Setting port security's limit on the number of MAC addresses for specific VLANs on a port 21
Enabling open authentication mode 22
Configuring free VLANs for port security 23
Applying a NAS-ID profile to port security 23
Enabling the escape critical VSI feature for 802.1X and MAC authentication users 24
Configuring the trigger order for authentication methods on a port 25
Configuring single-access authentication 26
Specifying a preauthentication domain for port security users on a port 27
Specifying a domain for port security users redirected to an unavailable URL 27
Configuring static users for port access authentication 28
Configuring a static user range for port access authentication 28
Enabling static user IP update 30
Setting the maximum number of concurrent static users allowed on a port 30
Specifying an ACL to match the MAC addresses of static users 31
Configuring a port security authentication profile 31
Configuring the port security authentication order (port security authentication profile view) 32
Configuring port security access policies 33
Configuring the action to be taken on online users when their access ports go down 34
Setting port security timers 34
Configuring the maximum number of reauthentication attempts for users in the preauthentication domain or Auth-Fail domain 35
Enabling immediate reauthentication for users in the preauthentication domain or Auth-Fail domain when the authentication server becomes reachable 36
Permitting MAC move between member ports in a TC group 36
Enabling SNMP notifications for port security 37
Enabling port security user logging 38
Display and maintenance commands for port security 38
Port security configuration examples 40
Example: Configuring port security in autoLearn mode 40

Example: Configuring port security in userLoginWithOUI mode ····42
Example: Configuring port security in macAddressElseUserLoginSecure mode ····46
Example: Configuring port security for M-LAG ····52
Troubleshooting port security ····65
Cannot set the port security mode ····65
Cannot configure secure MAC addresses ····65

Configuring port security

About port security

Port security not only provides port security modes to enhance network access security, but also provides MAC-based network access control for 802.1X, MAC authentication, Web authentication, and static users.

Major port security functions

Port security provides multiple security modes, intrusion detection and port security protection, and unified management of access authentication.

Multiple port security modes

Port security provides different security modes to control MAC address learning on ports and define combinations of authentication methods. It enables the device to learn legitimate source MAC addresses and simplifies the procedure to configure basic port authentication for desired network management.

Intrusion detection and port security protection

Typically, you use the intrusion detection and port security protection features in conjunction with port security modes to enhance security.

NTK

The need to know (NTK) feature prevents traffic interception by checking the destination MAC address in the outbound frames. The feature ensures that frames are sent only to the following hosts:

- Hosts that have passed authentication.
- Hosts whose MAC addresses have been learned or configured on the access device.

The NTK feature supports the following modes:

- **ntk-withbroadcasts**—Forwards only broadcast and unicast frames with a known destination MAC address.
- **ntk-withmulticasts**—Forwards only broadcast, multicast, and unicast frames with a known destination MAC address.
- **ntkauto**—Forwards only broadcast, multicast, and unicast frames with a known destination MAC address, and only when the port has online users.
- **ntkonly**—Forwards only unicast frames with a known destination MAC address.

Intrusion protection

The intrusion protection feature checks the source MAC address in inbound frames for illegal frames, and takes a predefined action on each detected illegal frame. The action can be disabling the port temporarily, disabling the port permanently, or blocking frames from the illegal MAC address for a period set by the MAC block timer.

A frame is illegal if its source MAC address cannot be learned in a port security mode or it is from a client that has failed 802.1X or MAC authentication.

Intrusion protection takes one of the following actions on a port in response to illegal frames:

- **blockmac**—Adds the source MAC addresses of illegal frames to the blocked MAC address list, and then discards frames sourced from blocked MAC addresses for a period set by the block timer. A blocked MAC address will be unblocked when the block timer expires.

- **disableport**—Disables the port until you bring it up manually.
- **disableport-temporarily**—Disables the port for a period of time. The period can be configured with the **`port-security timer disableport`** command.

Unified management of access authentication on ports

Port security enables centralized configuration and management of authentication methods (802.1X, MAC authentication, Web authentication, and static users) on ports, as well as centralized management of access users.

Port security modes

Port security supports the following categories of security modes:

- **MAC learning control**—Includes two modes: autoLearn and secure. MAC address learning is permitted on a port in autoLearn mode and disabled in secure mode.
- **Authentication**—Security modes in this category implement MAC authentication, 802.1X authentication, or a combination of these two authentication methods.

Controlling MAC address learning

- autoLearn.

  A port in this mode can learn MAC addresses. The automatically learned MAC addresses are not added to the MAC address table as dynamic MAC address. Instead, these MAC addresses are added to the secure MAC address table as secure MAC addresses. You can also configure secure MAC addresses by using the **`port-security mac-address security`** command.

  A port in autoLearn mode allows frames sourced from the following MAC addresses to pass:

  - Secure MAC addresses.
  - MAC addresses configured by using the **`mac-address dynamic`** and **`mac-address static`** commands.

  When the number of secure MAC addresses reaches the upper limit, the port transitions to secure mode.

- secure.

  MAC address learning is disabled on a port in secure mode. You configure MAC addresses by using the **`mac-address static`** and **`mac-address dynamic`** commands. For more information about configuring MAC address table entries, see *Layer 2—LAN Switching Configuration Guide*.

  A port in secure mode allows only frames sourced from the following MAC addresses to pass:

  - Secure MAC addresses.
  - MAC addresses configured by using the **`mac-address dynamic`** and **`mac-address static`** commands.

Performing authentication

For the ease of memory, use the following conventions to interprete the names of port security modes for authentication:

- **userLogin** specifies 802.1X authentication and port-based access control. **userLogin** with **Secure** specifies 802.1X authentication and MAC-based access control. **Ext** indicates allowing multiple 802.1X users to be authenticated and serviced at the same time. A security mode without **Ext** allows only one user to pass 802.1X authentication.
- **macAddress** specifies MAC authentication.
- **Else** specifies that the authentication method before **Else** is applied first. If the authentication fails, whether to turn to the authentication method following **Else** depends on the protocol type of the authentication request.

- **Or** specifies that the authentication method following **Or** is applied first. If the authentication fails, the authentication method before **Or** is applied.
- **And** indicates that the authentication method before **And** is applied first. After the authentication succeeds, the system turns to the authentication method that follows **And** upon receiving the request of the authentication method.

Performing 802.1X authentication

- userLogin.

  A port in this mode performs 802.1X authentication and implements port-based access control. The port can service multiple 802.1X users. Once an 802.1X user passes authentication on the port, any subsequent 802.1X users can access the network through the port without authentication.

- userLoginMab

  Perform port-based 802.1X MAC bypass authentication for users. If the device does not receive an EAP message from a user on a port in userLoginMab mode, the device authenticates the user by using the MAC address of the client as the username. After a successful authentication, the user will come online as an 802.1X user. On a port in this mode, after an 802.1X user passes authentication, the subsequent users can access the network without authentication.

- userLoginSecure.

  A port in this mode performs 802.1X authentication and implements MAC-based access control. The port services only one user passing 802.1X authentication.

- userLoginSecureExt.

  This mode is similar to the userLoginSecure mode except that this mode supports multiple online 802.1X users.

- userLoginWithOUI.

  This mode is similar to the userLoginSecure mode. The difference is that a port in this mode also permits frames from one user whose MAC address contains a specific OUI.

  In this mode, the port performs OUI check at first. If the OUI check fails, the port performs 802.1X authentication. The port permits frames that pass OUI check or 802.1X authentication.

NOTE:

An OUI is a 24-bit number that uniquely identifies a vendor, manufacturer, or organization. In MAC addresses, the first three octets are the OUI.

Performing MAC authentication

macAddressWithRadius: A port in this mode performs MAC authentication, and services multiple users.

Performing a combination of MAC authentication and 802.1X authentication

- macAddressOrUserLoginSecure.

  This mode is the combination of the macAddressWithRadius and userLoginSecure modes. The mode allows one 802.1X authentication user and multiple MAC authentication users to log in.

  In this mode, the port performs 802.1X authentication first. By default, if 802.1X authentication fails, MAC authentication is performed.

  However, the port in this mode processes authentication differently when the following conditions exist:

  - The port is enabled with parallel processing of MAC authentication and 802.1X authentication.
  - The port is enabled with the 802.1X unicast trigger.
  - The port receives a packet from an unknown MAC address.

Under such conditions, the port sends a unicast EAP-Request/Identity packet to the MAC address to initiate 802.1X authentication. After that, the port immediately processes MAC authentication without waiting for the 802.1X authentication result.

- macAddressOrUserLoginSecureExt.

  This mode is similar to the macAddressOrUserLoginSecure mode, except that this mode supports multiple 802.1X and MAC authentication users.

- macAddressElseUserLoginSecure.

  This mode is the combination of the macAddressWithRadius and userLoginSecure modes, with MAC authentication having a higher priority as the **Else** keyword implies. The mode allows one 802.1X authentication user and multiple MAC authentication users to log in.

  In this mode, the port performs MAC authentication upon receiving non-802.1X frames. Upon receiving 802.1X frames, the port performs MAC authentication and then, if the authentication fails, 802.1X authentication.

- macAddressElseUserLoginSecureExt.

  This mode is similar to the macAddressElseUserLoginSecure mode except that this mode supports multiple 802.1X and MAC authentication users as the **Ext** keyword implies.

- macAddressAndUserLoginSecureExt.

  In this mode, a user must pass both MAC authentication and 802.1X authentication to access the authorized network resources.

  The device uses the following process to handle an access user on a port operating in this mode:

  a. Performs MAC authentication for the user.

  b. Marks the user as a temporary MAC authentication user when the user passes MAC authentication. A temporary MAC authentication user can access only resources in the 802.1X guest VLAN or VSI.

  c. After receiving 802.1X protocol packets from the user on the port, the device performs 802.1X authentication for the user.

  d. After the user passes 802.1X authentication on the port, the device removes the temporary MAC authentication user entry. Then, the user comes online as an 802.1X user.

- macAddressAndUserLoginSecureExt.

  In this mode, a user must pass both MAC authentication and 802.1X authentication to access the authorized network resources.

  The device uses the following process to handle an access user on a port operating in this mode:

  a. Performs MAC authentication for the user.

  b. Marks the user as a temporary MAC authentication user when the user passes MAC authentication. A temporary MAC authentication user can access only resources in the 802.1X guest VLAN.

  c. After receiving 802.1X protocol packets from the user on the port, the device performs 802.1X authentication for the user.

  d. After the user passes 802.1X authentication on the port, the device removes the temporary MAC authentication user entry. Then, the user comes online as an 802.1X user.

- macAddressAndUserLoginSecureExt.

  In this mode, a user must pass both MAC authentication and 802.1X authentication to access the authorized network resources.

  The device uses the following process to handle an access user on a port operating in this mode:

  a. Performs MAC authentication for the user.

b. Marks the user as a temporary MAC authentication user when the user passes MAC authentication. A temporary MAC authentication user can access only resources in the 802.1X guest VLAN or guest VSI.

c. After receiving 802.1X protocol packets from the user on the port, the device performs 802.1X authentication for the user.

d. After the user passes 802.1X authentication on the port, the device removes the temporary MAC authentication user entry. Then, the user comes online as an 802.1X user.

Workflow in a port security mode

Upon receiving a frame, the port in a security mode searches the MAC address table for the source MAC address.

- If a match is found, the port forwards the frame.
- If no match is found, the port learns the MAC address or performs authentication, depending on the security mode.
  - If MAC address learning or user authentication succeeds, the device adds the MAC address to the MAC address table and forwards the frame.
  - If MAC address learning or user authentication fails, the device determines that the frame is illegal. It then takes the predefined NTK or intrusion protection action on the port or sends SNMP notifications.

Outgoing frames are not restricted by port security's NTK action unless they trigger the NTK feature.

Table 1 describes the port security modes and the security features.

Table 1 Port security modes

Purpose	Security mode		Features that can be triggered
Turning off the port security feature	noRestrictions (the default mode) In this mode, port security is disabled on the port and access to the port is not restricted.		N/A
Controlling MAC address learning	autoLearn		NTK/intrusion protection
	secure		
Performing authentication	userLogin		NTK (**ntkauto** mode)
	userLoginMab		
	userLoginSecure		NTK/intrusion protection
	userLoginSecureExt		
	userLoginWithOUI		
Performing authentication	macAddressWithRadius		NTK/intrusion protection
Performing authentication	Or	macAddressOrUserLoginSecure	NTK/intrusion protection
		macAddressOrUserLoginSecureExt	
	Else	macAddressElseUserLoginSecure	
		macAddressElseUserLoginSecureExt	
	And	macAddressAndUserLoginSecureExt	

The mode names are illustrated as follows:

- **userLogin** specifies 802.1X authentication and port-based access control. **userLogin** with **Secure** specifies 802.1X authentication and MAC-based access control. **Ext** indicates allowing

multiple 802.1X users to be authenticated and serviced at the same time. A security mode without **Ext** allows only one user to pass 802.1X authentication.

- **macAddress** specifies MAC authentication.
- **Else** specifies that the authentication method before **Else** is applied first. If the authentication fails, whether to turn to the authentication method following **Else** depends on the protocol type of the authentication request.
- **Or** specifies that the authentication method following **Or** is applied first. If the authentication fails, the authentication method before **Or** is applied.
- **And** indicates that the authentication method before **And** is applied first. After the authentication succeeds, the system turns to the authentication method that follows **And** upon receiving the request of the authentication method.

Secure MAC addresses

When port security is enabled, a port in a port security mode for MAC learning control permits only packets from the following types of MAC addresses to pass through:

- The secure MAC addresses on the port.
- The MAC addresses manually configured by using the **`mac-address dynamic`** command or the **`mac-address static`** command.

Secure MAC addresses include static, sticky, and dynamic secure MAC addresses, as shown in Table 2.

Table 2 Comparison of static, sticky, and dynamic secure MAC addresses

Type	Address sources	Can be saved and survive a device reboot?	Aging mechanism
Static	Manually added (by using the **`port-security mac-address security`** command without the **`sticky`** keyword).	Yes.	Not available. The static secure MAC addresses never age out unless you perform any of the following tasks: • Manually remove these MAC addresses. • Change the port security mode. • Disable the port security feature.
Sticky	• Manually added (by using the **`port-security mac-address security`** command with the **`sticky`** keyword). • Converted from dynamic secure MAC addresses. • Automatically learned when the dynamic secure MAC feature is disabled.	Yes. The secure MAC aging timer restarts at a reboot.	By default, sticky MAC addresses do not age out. However, you can configure an aging timer or use the aging timer together with the inactivity aging feature to remove old sticky MAC addresses. • If only the aging timer is configured, the aging timer counts up regardless of whether traffic data has been sent from the sticky MAC addresses. • If both the aging timer and the inactivity aging feature are configured, the aging timer restarts once traffic data is detected from the sticky MAC addresses.
Dynamic	• Converted from sticky MAC	No. All dynamic secure	Same as sticky MAC addresses.

Type	Address sources	Can be saved and survive a device reboot?	Aging mechanism
	addresses. • Automatically learned after the dynamic secure MAC feature is enabled.	MAC addresses are lost at reboot.	

Authentication load sharing modes in M-LAG environments

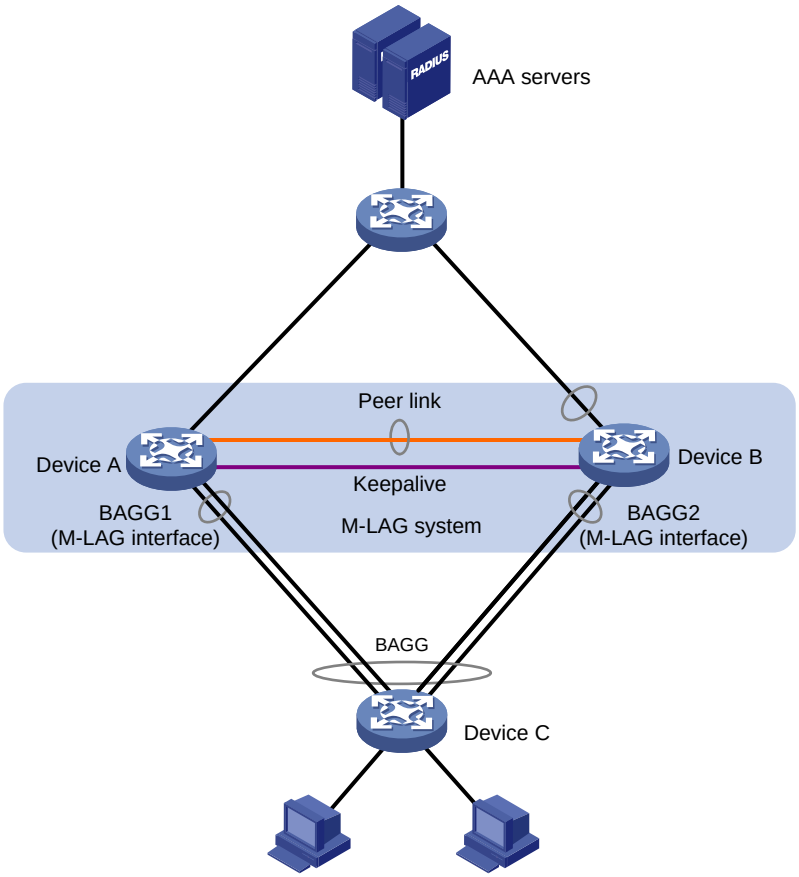
As shown in Figure 1, use M-LAG to configure Device A and Device B into an M-LAG system for load sharing and node redundancy. When one of the devices fails, traffic is switched to the other device to ensure service continuity.

For users attached to the M-LAG interfaces on an M-LAG system, port security provides the following authentication load sharing modes:

- **Centralized mode**—In this mode, the primary M-LAG member device processes authentication services for all users attached to any M-LAG interfaces in the system.
- **Distributed modes**—In a distributed mode, both M-LAG member devices provide authentication services for users attached to the M-LAG interfaces. Port security provides the following distributed authentication processing modes:
  - **Distributed local mode**—Each M-LAG member device processes authentication for users attached to their local M-LAG interfaces.
  - **Distributed even-/odd-MAC mode**—In this mode, one M-LAG member device processes authentication for packets with an even source MAC address and the other M-LAG member device processes authentication for packets with an odd source MAC address. In this mode, you must configure one M-LAG member device to process authentication for even source MAC addresses and configure the other M-LAG member device to process authentication for odd source MAC addresses.

One M-LAG member device automatically synchronizes user data to the other M-LAG member device upon each successful user authentication. This ensures that when one M-LAG member device fails, the other member device can take over to process authentication services for all users.

Figure 1 M-LAG network diagram

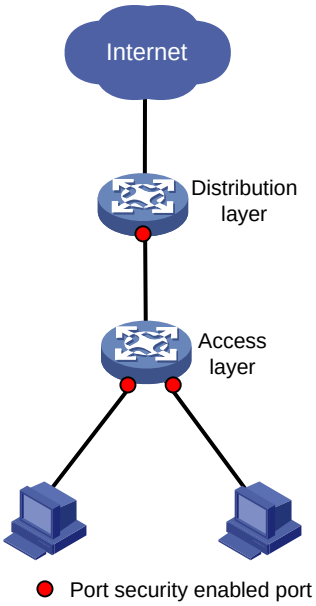


Port security application scenarios

Port security is typically used in the scenarios shown in Figure 2.

- Used on access devices to prevent fake users from launching attacks through ports other than those used by legitimate users.
  - If the access location of users is fixed but the link is unreliable, enable port security to convert their MAC addresses to sticky MAC addresses. This ensures that their MAC address bindings can be saved to the configuration file to persist throu gh device reboot or port state change between up and down.
  - If there are only a few users and their access location is fixed, manually configure their MAC addresses as static secure MAC addresses, which never age out.
  - If users frequently move between ports, configure port security to convert their MAC addresses to dynamic secure MAC addresses. This action ensures prompt removal of port and MAC bindings for inactive users, preventing unnecessary use of port and MAC binding resources. It also eliminates MAC entry conflicts to ensure that users can move between ports.
- Used on distribution devices to control the number of access users, preventing network congestion caused by excessive network accesses.

Figure 2 Port security application scenarios



Restrictions and guidelines: Port security configuration

This feature applies to networks, such as a WLAN, that require different authentication methods for different users on a port.

As a best practice, use the 802.1X authentication or MAC authentication feature rather than port security for scenarios that require only 802.1X authentication or MAC authentication. For more information about 802.1X and MAC authentication, see "Configuring 802.1X" and "Configuring MAC authentication."

The **`secure`** and **`userlogin-withoui`** port security modes are not supported on Layer 2 aggregate interfaces. Other port security settings are supported on both Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces.

After a Layer 2 Ethernet interface is added to an aggregation group, port security settings on the interface do not take effect.

Do not delete a Layer 2 aggregate interface if the interface has online 802.1X or MAC authentication users.

When RADIUS DAS is enabled, an M-LAG system does not support the following operations:

- Shuts down or reboots the access ports for 802.1X, Web authentication, MAC authentication users by sending CoA messages.
- Reauthenticates these users by sending CoA messages.

For more information about RADIUS DAS, see "Configuring AAA."

In an M-LAG scenario with load sharing in even-/odd-MAC mode, make sure the port security settings match on both devices except for the different processing modes. For other load sharing modes, keep all port security settings identical on both devices.

To ensure a successful HTTPS redirect for users in either of the following situations, make sure VLAN interfaces or VSI interfaces exist for the VLANs or VSIs that transport their packets:

- The users are assigned a redirect URL.

- EAD assistant is enabled.

Port security tasks at a glance

To configure port security, perform the following tasks:

1. Configuring basic features of port security
   - Enabling port security
   - Setting the port security mode
2. Configuring port security detection and protection

   You typically use port security detection and protection features in conjunction with port security modes for enhanced security.
   - (Optional.) Configuring NTK
   - (Optional.) Configuring intrusion protection
3. (Optional) Configuring secure MAC addresses and their control settings
   - (Optional.) Setting port security's limit on the number of secure MAC addresses on a port
   - (Optional.) Configuring secure MAC addresses
     - Adding a static secure MAC address
     - Adding a sticky MAC address
     - Converting sticky MAC addresses to dynamic secure MAC addresses
     - Configuring secure MAC address aging
4. (Optional.) Configuring extended features of port security

   The following features are applicable when port security is enabled or when 802.1X, MAC authentication or Web authentication is enabled.
   - Configuring an authentication load sharing mode for users attached to M-LAG interfaces

     Perform this task on an M-LAG system.
   - Configuring static users for port access authentication
     - Configuring a static user range for port access authentication
     - Enabling static user IP update
     - Setting the maximum number of concurrent static users allowed on a port
     - Specifying an ACL to match the MAC addresses of static users
   - Configuring port security authentication profile settings
     - Configuring a port security authentication profile
     - Configuring the port security authentication order (port security authentication profile view)
   - Configuring port security access policies
   - Ignoring authorization information from the server
   - Configuring MAC move
   - Enabling overwrite of local MAC entries with remote MAC entries for MAC addresses moved to a remote device
   - Enabling the authorization-fail-offline feature
   - Setting port security's limit on the number of MAC addresses for specific VLANs on a port
   - Enabling open authentication mode
   - Configuring free VLANs for port security
   - Applying a NAS-ID profile to port security

- Enabling the escape critical VSI feature for 802.1X and MAC authentication users
- Configuring single-access authentication
- Specifying a preauthentication domain for port security users on a port
- Specifying a domain for port security users redirected to an unavailable URL
- Setting port security timers
- Configuring the maximum number of reauthentication attempts for users in the preauthentication domain or Auth-Fail domain
- Enabling immediate reauthentication for users in the preauthentication domain or Auth-Fail domain when the authentication server becomes reachable
- Permitting MAC move between member ports in a TC group
- Enabling SNMP notifications for port security
- Enabling port security user logging

Enabling port security

Restrictions and guidelines

When you configure port security, follow these restrictions and guidelines:

- Executing the **`undo port-security enable`** command to disable port security will log off online users (if any).
- Enabling or disabling port security resets the following security settings to the default:
  - 802.1X access control mode, which is MAC-based.
  - Port authorization state, which is **auto**.

  These settings are not configurable after port security is enabled.

For more information about 802.1X authentication and MAC authentication configuration, see "Configuring 802.1X" and "Configuring MAC authentication."

Prerequisites

Before you enable port security, disable 802.1X and MAC authentication globally.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable port security.
   ```
   port-security enable
   ```
   By default, port security is disabled.

Setting the port security mode

Restrictions and guidelines

You can specify a port security mode when port security is disabled, but your configuration cannot take effect.

Changing the port security mode of a port logs off the online users of the port.

Do not enable 802.1X authentication or MAC authentication on a port where port security is enabled.

After enabling port security, you can change the port security mode of a port only when the port is operating in noRestrictions (the default) mode. To change the port security mode for a port in any

other mode, first use the **undo port-security port-mode** command to restore the default port security mode.

The device supports the URL attribute assigned by a RADIUS server in the following port security modes:

- mac-authentication.
- mac-else-userlogin-secure.
- mac-else-userlogin-secure-ext.
- userlogin-secure.
- userlogin-secure-ext.
- userlogin-secure-or-mac.
- userlogin-secure-or-mac-ext.
- userlogin-withoui.

During authentication, the HTTP or HTTPS requests of a user are redirected to the Web interface specified by the server-assigned URL attribute. After the user passes the Web authentication, the RADIUS server records the MAC address of the user and uses a DM (Disconnect Message) to log off the user. When the user initiates 802.1X or MAC authentication again, it will pass the authentication and come online successfully.

When the port security mode is **macAddressAndUserLoginSecureExt** on a port, follow these restrictions and guidelines:

- To make sure the 802.1X clients attached to the port can initiate authentication, enable unicast trigger on the port by using the **dot1x unicast-trigger** command.
- The MAC authentication guest VLAN or VSI on the port does not take effect. For the temporary MAC authentication users to access a limited set of resources, configure an 802.1X guest VLAN or VSI on the port.
- If accounting is not required for the temporary MAC authentication users, configure different ISP domains for MAC authentication users and 802.1X users. In the ISP domain for MAC authentication users, set the accounting method to none.

To perform 802.1X MAC bypass authentication for clients incapable of sending EAP messages on a port in userLoginMab mode, enable 802.1X unicast trigger on the port by executing the **dot1x unicast-trigger** command.

To redirect the HTTPS requests of port security users, specify the HTTPS redirect listening port on the device. For more information, see HTTP redirect in *Layer 3—IP Services Configuration Guide*.

Prerequisites

Before you set a port security mode for a port, complete the following tasks:

- Disable 802.1X and MAC authentication.
- If you are configuring the autoLearn mode, set port security's limit on the number of secure MAC addresses. You cannot change the setting when the port is operating in autoLearn mode.

Procedure

1. Enter system view.

   system-view

2. Set an OUI value for user authentication.

   port-security oui index *index-value* **mac-address** *oui-value*

   By default, no OUI values are configured for user authentication.

   This command is required only for the **userlogin-withoui** mode.

   You can set multiple OUIs, but when the port security mode is **userlogin-withoui**, the port allows one 802.1X user and only one user that matches one of the specified OUIs.

3. Enter interface view.

   interface *interface-type interface-number*

4. Set the port security mode.

   port-security port-mode { **autolearn** | **mac-and-userlogin-secure-ext** | **mac-authentication** | **mac-else-userlogin-secure** | **mac-else-userlogin-secure-ext** | **secure** | **userlogin** | **userlogin-mab** | **userlogin-secure** | **userlogin-secure-ext** | **userlogin-secure-or-mac** | **userlogin-secure-or-mac-ext** | **userlogin-withoui** }

   By default, a port operates in noRestrictions mode.

Setting port security's limit on the number of secure MAC addresses on a port

About this task

You can set the maximum number of secure MAC addresses that port security allows on a port for the following purposes:

- Controlling the number of concurrent users on the port.

  For a port operating in a security mode (except for autoLearn and secure), the upper limit equals the smaller of the following values:

  - The limit of the secure MAC addresses that port security allows.
  - The limit of concurrent users allowed by the authentication mode in use.

- Controlling the number of secure MAC addresses on the port in autoLearn mode.

  You can also set the maximum number of secure MAC addresses that port security allows for specific VLANs or each VLAN on a port. When the number of secure MAC addresses reaches the upper limit, the port transitions to secure mode. In secure mode, a port cannot learn MAC addresses.

Port security's limit on the number of secure MAC addresses on a port is independent of the MAC learning limit described in MAC address table configuration. For more information about MAC address table configuration, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the maximum number of secure MAC addresses allowed on a port.

   port-security max-mac-count *max-count* [**vlan** [*vlan-id-list*]]

   By default, port security does not limit the number of secure MAC addresses on a port.

Configuring secure MAC addresses

About secure MAC addresses

For more information about secure MAC addresses, see "Configuring secure MAC addresses."

Restrictions and guidelines

You can bind a secure MAC address to only one port in a VLAN.

You cannot add a MAC address to a VLAN as a static secure MAC address if it has been added as a sticky MAC address in that VLAN, and vice versa.

Prerequisites

Before you configure secure MAC addresses, complete the following tasks:

- Set port security's limit on the number of MAC addresses on the port. Perform this task before you enable autoLearn mode.
- Set the port security mode to autoLearn.
- Configure the port to permit packets of the specified VLAN to pass or add the port to the VLAN. Make sure the VLAN already exists.

Adding a static secure MAC address

1. Enter system view.

   system-view
2. Add a static secure MAC address.
   - Add a static secure MAC address in system view.

     port-security mac-address security *mac-address* **interface** *interface-type interface-number* **vlan** *vlan-id*
   - Execute the following commands in sequence to add a static secure MAC address in interface view:

     interface *interface-type interface-number*

     port-security mac-address security *mac-address* **vlan** *vlan-id*

   By default, no static secure MAC addresses exist.

Adding a sticky MAC address

About this task

When port security is enabled, the MAC addresses learned on a port in autoLearn mode are sticky MAC addresses by default. The following information describes the procedure to add a sticky MAC address manually.

Procedure

1. Enter system view.

   system-view
2. Add a sticky MAC address.
   - Add a sticky MAC address in system view.

     port-security mac-address security sticky *mac-address* **interface** *interface-type interface-number* **vlan** *vlan-id*
   - Execute the following commands in sequence to add a sticky MAC address in interface view:

     interface *interface-type interface-number*

     port-security mac-address security sticky *mac-address* **vlan** *vlan-id*

By default, no manually configured sticky MAC addresses exist.

Converting sticky MAC addresses to dynamic secure MAC addresses

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Convert sticky MAC addresses to dynamic secure MAC addresses.
   port-security mac-address dynamic

   By default, sticky MAC addresses can be saved to the configuration file. Once saved, they can survive a device reboot.

Configuring secure MAC address aging

About this task

By default, secure MAC addresses do not age out. This results in the following issues:

- If a legitimate user leaves a port and an illegitimate user uses the legitimate user's source MAC to access the network, the legitimate user cannot reconnect.
- When legitimate users move from one port to another port within the same VLAN, they cannot come online from the new port if their secure MAC entries on the original port do not age out.
- Legitimate users who have left still occupy the MAC address resources on the port, preventing other legitimate users from connecting to the network.

To enhance port security and improve port resource usage, you can have certain types of secure MAC addresses periodically age out.

Setting the secure MAC aging timer

1. Enter system view.
   system-view
2. Set the secure MAC aging timer.
   port-security timer autolearn aging [**second**] *time-value*

   By default, secure MAC addresses do not age out.

Enabling inactivity aging for secure MAC addresses

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable inactivity aging for secure MAC addresses.
   port-security mac-address aging-type inactivity

   By default, the inactivity aging feature is disabled for secure MAC addresses.

Configuring NTK

About this task

The NTK feature checks the destination MAC address in outbound frames to make sure frames are forwarded only to trustworthy devices.

The NTK feature supports the following modes:

- **ntkonly**—Forwards only unicast frames with a known destination MAC address.
- **ntk-withbroadcasts**—Forwards only broadcast and unicast frames with a known destination MAC address.
- **ntk-withmulticasts**—Forwards only broadcast, multicast, and unicast frames with a known destination MAC address.
- **ntkauto**—Forwards only broadcast, multicast, and unicast frames with a known destination MAC address, and only when the port has online users.

Restrictions and guidelines

The NTK feature drops any unicast frame with an unknown destination MAC address.

Not all port security modes support triggering the NTK feature. For more information, see the port security modes table in "Workflow in a port security mode."

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the NTK feature.

   port-security ntk-mode { **ntk-withbroadcasts** | **ntk-withmulticasts** | **ntkauto** | **ntkonly** }

   By default, NTK is disabled on a port and all frames are allowed to be sent.

Configuring intrusion protection

About this task

Intrusion protection takes one of the following actions on a port in response to illegal frames:

- **blockmac**—Adds the source MAC addresses of illegal frames to the blocked MAC address list, and then discards frames sourced from blocked MAC addresses for a period set by the block timer. A blocked MAC address will be unblocked when the block timer expires.
- **disableport**—Disables the port until you bring it up manually.
- **disableport-temporarily**—Disables the port for a period of time. The period can be configured with the **port-security timer disableport** command.

If the action is **disableport** or **disableport-temporarily**, you can also enable strict intrusion protection to enhance security. With strict intrusion protection, the port shuts down when it receives a frame whose source MAC address is a secure MAC address on another port in the same VLAN.

Restrictions and guidelines

On a port operating in either macAddressElseUserLoginSecure mode or macAddressElseUserLoginSecureExt mode, intrusion protection is triggered only after both MAC authentication and 802.1X authentication fail for the same frame.

Strict intrusion protection takes effect only when the intrusion protection action is **disableport** or **disableport-temporarily**. When strict intrusion protection is enabled on a port, you cannot change the intrusion protection action to **blockmac** on that port. To change the intrusion protection action to **blockmac**, you must first disable strict intrusion protection on that port.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the intrusion protection feature.
   port-security intrusion-mode { **blockmac** | **disableport** | **disableport-temporarily** }
   By default, intrusion protection is disabled.
4. (Optional.) Enable strict intrusion protection.
   port-security strict-intrusion-protection enable
   By default, strict intrusion protection is disabled.
5. Return to system view.
   quit
6. (Optional.) Set the silence timeout period during which a port remains disabled.
   port-security timer disableport *time-value*
   By default, the port silence timeout period is 20 seconds.
7. (Optional.) Set the block timer for blocked MAC addresses.
   port-security timer blockmac *time-value*
   By default, the block timer is 180 seconds.

Configuring an authentication load sharing mode for users attached to M-LAG interfaces

About this task

This feature takes effect only on 802.1X, Web authentication, and MAC authentication users attached to M-LAG interfaces in an M-LAG environment.

In an M-LAG system, the M-LAG member devices exchange configuration information with each other to check for configuration conflicts. If a configuration conflict exists, the M-LAG member devices do not allow new users to come online.

For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*.

Restrictions and guidelines

CAUTION:
To avoid user logoffs caused by configuration conflicts, do not change the authentication load sharing mode for users on M-LAG interfaces when port security, 802.1X, Web authentication, or MAC authentication is enabled.

Make sure the M-LAG member devices are consistent in authentication load sharing settings for users attached to M-LAG interfaces.

- **Centralized mode**—Configure both devices to operate in centralized mode for user authentication.

- **Distributed local mode**—Configure both M-LAG member devices to operate in distributed local mode for user authentication.
- **Distributed even-/odd-MAC mode**—Configure one M-LAG member device in distributed even-MAC mode and the other to operate in distributed odd-MAC mode for user authentication.

To ensure correct user data processing, follow these guidelines when you configure the peer aggregate interfaces on each remote access device connected to the M-LAG interfaces:

- If the M-LAG system uses distributed local mode for user authentication, link-aggregation load sharing on the access device can only be based on one of the following criteria:
  - Destination IP address.
  - Destination MAC address.
  - Source IP address.
  - Source MAC address.
- In an 802.1X authentication scenario, you must configure the access device to ignore all packet fields except the source MAC if it uses the default link-aggregation load sharing mode.

In centralized mode, if all member ports of an M-LAG interface belong only to one M-LAG member device and the M-LAG interface forwards authentication traffic, users attached to the M-LAG interface cannot come online. To ensure that users attached to such M-LAG interfaces can come online, do not set the authentication load sharing mode to centralized mode.

For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*. For more information about link aggregation load sharing, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Set the authentication load sharing mode for users attached to M-LAG interfaces.

   port-security m-lag load-sharing-mode { **centralized** | **distributed** { **even-mac** | **local** | **odd-mac** } }

   By default, centralized mode applies.

Ignoring authorization information from the server

About this task

You can configure a port to ignore the authorization information received from the server (local or remote) after an 802.1X or MAC authentication user passes authentication.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Ignore the authorization information received from the authentication server.

   port-security authorization ignore

   By default, a port uses the authorization information received from the authentication server.

Configuring MAC move

About this task

Port security MAC move takes effect on online users authenticated through 802.1X authentication, MAC authentication, or Web authentication in the following scenarios:

- **Inter-port move on a device**—An authenticated online user moves between ports on the device. The user VLAN or authentication method might change or stay unchanged after the move.
- **Inter-VLAN move on a port**—An authenticated online user moves between VLANs on a trunk or hybrid port. This mode takes effect only when the packets that trigger authentication are VLAN tagged.

Port security MAC move allows an authenticated online user on one port or VLAN to be reauthenticated and come online on another port or VLAN without going offline first. After the user passes authentication on the new port or VLAN, the system removes the authentication session of the user on the original port or VLAN. This action ensures that the user stays online on only one port in one VLAN.

NOTE:

For MAC authentication, the MAC move feature applies only when MAC authentication single-VLAN mode is used. The MAC move feature does not apply to MAC authentication users that move between VLANs on a port with MAC authentication multi-VLAN mode enabled.

If this feature is disabled, authenticated users must go offline first before they can be reauthenticated successfully on a new port or VLAN to come online.

For a user moving between ports, the port from which the user moves is called the source port and the port to which the user moves is called the destination port.

On the destination port, an authenticated user will reauthenticate in the VLAN authorized on the source port if the source port is enabled with MAC-based VLAN. If that VLAN is not permitted to pass through on the destination port, reauthentication will fail. To avoid this situation, enable VLAN check bypass on the destination port.

Restrictions and guidelines

As a best practice to minimize security risks, enable MAC move only if user roaming between ports is required.

Authenticated users cannot move between ports on a device or between VLANs on a port if the maximum number of online users on the authentication server has been reached.

MAC authentication multi-VLAN mode has higher priority than MAC move for users moving between VLANs on a port. If MAC authentication multi-VLAN mode is enabled, these users can come online in the new VLAN without being reauthenticated. To enable MAC authentication multi-VLAN mode, use the **`mac-authentication host-mode multi-vlan`** command. For more information about MAC authentication multi-VLAN mode, see "Configuring MAC authentication."

When you configure VLAN check bypass for users moving between ports, follow these guidelines:

- To ensure a successful reauthentication, enable VLAN check bypass on a destination port if the source port is enabled with MAC-based VLAN.
- If the destination port is an 802.1X-enabled trunk port, you must configure it to send 802.1X protocol packets without VLAN tags. For more information, see "Configuring 802.1X."

Procedure

1. Enter system view.

   `system-view`

2. Enable MAC move.

   port-security mac-move permit [**port** | **vlan**]

   By default, MAC move is disabled.
3. (Optional.) Enable VLAN check bypass for users moving between ports.
   a. Enter interface view.

      interface *interface-type interface-number*
   b. Enable VLAN check bypass for users moving to the port from other ports.

      port-security mac-move bypass-vlan-check

      By default, the VLAN check bypass feature is disabled.

Enabling overwrite of local MAC entries with remote MAC entries for MAC addresses moved to a remote device

About this task

This feature is typically used in an EVPN or M-LAG environment. It enables the device to overwrite local authenticated MAC entries with remote authenticated MAC entries received from remote devices over tunnels or the peer link immediately, without waiting for the local entries to age out. This overwrite mechanism ensures that the device can promptly update authenticated MAC entries to forward traffic correctly for port security users after they move to a remote device and reauthenticate to the network.

Procedure

1. Enter system view.

   system-view
2. Enable the device to overwrite the local MAC entry for a MAC address with the remote MAC entry for that MAC address after a MAC move

   port-security mac-move overwrite-local

   By default, the device overwrites the local MAC entry for a MAC address with the remote MAC entry for that MAC address after a MAC move.

Enabling the authorization-fail-offline feature

About the authorization-fail-offline feature

IMPORTANT:

The authorization-fail-offline feature takes effect only on port security users that have failed ACL authorization.

The authorization-fail-offline feature logs off port security users that have failed authorization.

A user fails authorization in the following situations:

- The device or server fails to assign the specified authorization attribute to the user.
- The device or server assigns authorization information that does not exist on the device to the user.

This feature does not apply to users that have failed VLAN authorization. The device logs off these users directly.

You can also enable the quiet timer feature for 802.1X or MAC authentication users that are logged off by the authorization-fail-offline feature. The device adds these users to the 802.1X or MAC authentication quiet queue. Within the quiet timer, the device does not process packets from these users or authenticate them. If you do not enable the quiet timer feature, the device immediately authenticates these users upon receiving packets from them.

Prerequisites

For the quiet timer feature to take effect, complete the following tasks:

- For 802.1X users, use the **dot1x quiet-period** command to enable the quiet timer and use the **dot1x timer quiet-period** command to set the timer.
- For MAC authentication users, use the **mac-authentication timer quiet** command to set the quiet timer for MAC authentication.

Procedure

1. Enter system view.

   system-view
2. Enable the authorization-fail-offline feature.

   port-security authorization-fail offline [**quiet-period**]

   By default, this feature is disabled, and the device does not log off users that have failed authorization.

Setting port security's limit on the number of MAC addresses for specific VLANs on a port

About this task

Typically, port security allows the access of the following types of MAC addresses on a port:

- MAC addresses that pass 802.1X authentication or MAC authentication.
- MAC addresses in the MAC authentication guest or critical VLAN, and MAC addresses in the MAC authentication guest or critical VSI, and MAC addresses in the MAC authentication critical microsegment.
- MAC addresses in the 802.1X guest, Auth-Fail, or critical VLAN and MAC addresses in the 802.1X guest, Auth-Fail, or critical VSI.
- MAC addresses that pass Web authentication and MAC addresses in the Web authentication Auth-Fail VLAN.

This feature limits the number of MAC addresses that port security allows to access a port through specific VLANs. Use this feature to prevent resource contentions among MAC addresses and ensure reliable performance for each access user on the port. When the number of MAC addresses in a VLAN on the port reaches the upper limit, the device denies any subsequent MAC addresses in the VLAN on the port.

Restrictions and guidelines

On a port, the maximum number of MAC addresses in a VLAN cannot be smaller than the number of existing MAC addresses in the VLAN. If the specified maximum number is smaller, the setting does not take effect.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

interface *interface-type interface-number*

3. Set port security's limit on the number of MAC addresses for specific VLANs on the port.
   port-security mac-limit *max-number* **per-vlan** *vlan-id-list*

   By default, no limit is set to the number of MAC addresses for specific VLANs on the port.

Enabling open authentication mode

About this task

When planning for initial network deployment, administrators must estimate the number of users and determine their authentication methods. This information prepares them to configure user information on the RADIUS server for future access control. If 802.1X or MAC address authentication is enabled on ports prematurely, some users might fail to come online due to mismatching user information on the RADIUS server, preventing administrators from gathering sufficient user data.

This feature enables access users (802.1X or MAC authentication users) of a port to come online and access the network even if they use nonexistent usernames or incorrect passwords.

Access users that come online in open authentication mode are called open users. Authorization and accounting are not available for open users. To display open user information, use the following commands:

- **display dot1x connection open**.
- **display mac-authentication connection open**.

Open users do not include those users that are normally authenticated to come online from a port in open authentication mode.

Restrictions and guidelines

When you configure open authentication mode, follow these restrictions and guidelines:

- If global open authentication mode is enabled, all ports are enabled with open authentication mode regardless of the port-specific open authentication mode setting. If global open authentication mode is disabled, whether a port is enabled with open authentication mode depends on the port-specific open authentication mode setting.
- The open authentication mode setting has lower priority than the 802.1X Auth-Fail VLAN and the MAC authentication guest VLAN. Open authentication mode does not take effect on a port if the port is also configured with the 802.1X Auth-Fail VLAN or the MAC authentication guest VLAN. For information about 802.1X authentication and MAC authentication, see "802.1X overview," "Configuring 802.1X," and "Configuring MAC authentication."
- The open authentication mode setting has lower priority than the 802.1X Auth-Fail VSI and the MAC authentication guest VSI. Open authentication mode does not take effect on a port if the port is also configured with the 802.1X Auth-Fail VSI or the MAC authentication guest VSI. For information about 802.1X authentication and MAC authentication, see "802.1X overview," "Configuring 802.1X," and "Configuring MAC authentication."

Procedure

1. Enter system view.
   system-view
2. Enable global open authentication mode.
   port-security authentication open global

   By default, global open authentication mode is disabled.
3. Enter interface view.
   interface *interface-type interface-number*

4. Enable open authentication mode on the port.

   port-security authentication open

   By default, open authentication mode is disabled on a port.

Configuring free VLANs for port security

About this task

This feature allows packets from the specified VLANs to not trigger 802.1X or MAC authentication on a port configured with any of the following features:

- 802.1X authentication.
- MAC authentication.
- One of the following port security modes:
  - userLogin.
  - userLoginSecure.
  - userLoginWithOUI.
  - userLoginSecureExt.
  - macAddressWithRadius.
  - macAddressOrUserLoginSecure.
  - macAddressElseUserLoginSecure.
  - macAddressOrUserLoginSecureExt.
  - macAddressElseUserLoginSecureExt.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure free VLANs for port security.

   port-security free-vlan *vlan-id-list*

   By default, no free VLANs for port security exist on a port.

Applying a NAS-ID profile to port security

About this task

By default, the device sends its device name in the NAS-Identifier attribute of all RADIUS requests.

A NAS-ID profile enables you to send different NAS-Identifier attribute strings in RADIUS requests from different VLANs. The strings can be organization names, service names, or any user categorization criteria, depending on the administrative requirements.

For example, map the NAS-ID **companyA** to all VLANs of company A. The device will send **companyA** in the NAS-Identifier attribute for the RADIUS server to identify requests from any Company A users.

Restrictions and guidelines

You can apply a NAS-ID profile to port security globally or on a port. On a port, the device selects a NAS-ID profile in the following order:

1. The port-specific NAS-ID profile.

2. The NAS-ID profile applied globally.

If no NAS-ID profile is applied or no matching binding is found in the selected profile, the device uses the device name as the NAS-ID.

For more information about the NAS-ID profile configuration, see "Configuring AAA."

Procedure

1. Enter system view.
   system-view
2. Apply a NAS-ID profile.
   - Apply a NAS-ID profile globally.
     port-security nas-id-profile *profile-name*
   - Execute the following commands in sequence to apply a NAS-ID profile to an interface:
     interface *interface-type interface-number*
     port-security nas-id-profile *profile-name*

   By default, no NAS-ID profile is applied in system view or in interface view.

Enabling the escape critical VSI feature for 802.1X and MAC authentication users

About this task

The escape critical VSI feature operates on VXLAN networks. It enables 802.1X and MAC authentication users to escape the authentication failure that occurs because the RADIUS server is malfunctioning.

You can enable this feature temporarily to prevent 802.1X and MAC authentication service interruption while you are troubleshooting a malfunctioning RADIUS server.

After the escape critical VSI feature is enabled on a port, the device performs the following operations when 802.1X or MAC authentication is triggered for a user:

1. Dynamically creates an Ethernet service instance that matches the user's access VLAN and MAC address on the access port.
2. Maps the Ethernet service instance to the 802.1X or MAC authentication critical VSI on the port.

The user can then come online without authentication and access resources in the VXLAN associated with the critical VSI.

The escape critical VSI feature does not affect 802.1X or MAC authentication users that have been online before this feature is enabled.

Restrictions and guidelines

For the escape critical VSI feature to work correctly on a port, make sure the port does not have the following settings:

- Web authentication.
- Guest, Auth-Fail, or critical VLAN for 802.1X authentication.
- Guest or critical VLAN for MAC authentication.

You can enable or disable this feature globally on all ports or on a per-port basis.

If the **mac-authentication critical vsi** *critical-vsi-name* **url-user-logoff** command is used in conjunction with this feature, MAC authentication users that have been assigned authorization URLs on the port will be logged off. For more information, see "Configuring MAC authentication."

When you disable the escape critical VSI both globally and on a port, the device logs off the users in the critical VSIs for 802.1X and MAC authentication on that port. The users must perform authentication to come online again on that port.

The escape critical VSI feature does not take effect on a new 802.1X or MAC authentication user if any of the following conditions exists:

- The 802.1X client and the device use different EAP message handling methods.
- 802.1X MAC address binding is enabled on the user's access port, but the MAC address of the 802.1X user is not bound to that port.
- The user's MAC address is an all-zero, all-F, or multicast MAC address.

Prerequisites

Before you enable the escape critical VSI feature on a port, configure an 802.1X critical VSI and a MAC authentication critical VSI on that port. For more information about critical VSI configuration, see "Configuring 802.1X" and "Configuring MAC authentication."

Procedure

1. Enter system view.

   system-view
2. Enable the escape critical VSI feature.
   - Enable the global escape critical VSI feature.

     port-security global escape critical-vsi
   - Execute the following commands in sequence to enable the escape critical VSI feature on a port:

     interface *interface-type interface-number*

     port-security escape critical-vsi

   By default, the escape critical VSI feature is disabled.

Configuring the trigger order for authentication methods on a port

About this task

This feature takes effect only on ports with triple authentication enabled. Triple authentication allows Web authentication, MAC authentication, and 802.1X authentication to be enabled concurrently on a Layer 2 port for user access. Different types of endpoint packets trigger different types of authentication first. For more information, see "Configuring triple authentication."

By default, if 802.1X unicast trigger is enabled, any packdets from endpoints will trigger 802.1X authentication. To enable any endpoint packets to trigger MAC authentication first, use this feature.

A port can run authentication processes concurrently for multiple authentication methods. The failure of one authentication does not affect the processes for other authentication methods. However, if an endpoint passes one authentication on a port, the device handles processes for other authentication methods on the port as follows:

- If the endpoint passes MAC authentication, the device generates a MAC authentication user entry on the port and continues to perform 802.1X authentication for the endpoint on the port. However, the device cannot continue Web authentication for the endpoint on the port.
  - If the endpoint passes 802.1X authentication after MAC authentication, the device generates an 802.1X user entry for the endpoint on the port. The 802.1X user entry overwrites the MAC authentication user entry.

    - If the endpoint does not pass 802.1X authentication after MAC authentication, the MAC authentication user entry is retained on the port. The endpoint can trigger 802.1X authentication again, but it cannot trigger Web authentication.
- If the endpoint fails MAC authentication but passes 802.1X or Web authentication, the device immediately stops all authentication methods on the port except the one the endpoint has passed. In addition, the device can no longer trigger authentication processes for the stopped authentication methods for the endpoint on the port.

Restrictions and guidelines

This feature takes effect only on ports with triple authentication enabled.

This feature causes users that are being authenticated to fail authentication. The users must retrigger authentication to come online. As a best practice to avoid users failing to come online, use this feature with caution.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the trigger order for authentication methods on the port as MAC authentication, 802.1X authentication, and Web authentication.

   port-security triple-auth-order mac-dot1x-web

   By default, the authentication that is triggered first depends on the type of packets sent from endpoints in a triple authentication environment.

Configuring single-access authentication

About this task

By default, when multiple authentication methods are configured on an interface, a user can proceed with other methods after passing one. For more information, see triple authentication configuration in Security Configuration Guide. For example, if a client sends an EAP message after the client user has come online through MAC authentication, the device performs 802.1X authentication for the user. If 802.1X authentication succeeds, the user will come online as an 802.1X user.

If single-access authentication is enabled, users will not undergo any other authentication methods after they have authenticated with one method.

Restrictions and guidelines

If multi-authentication mode is used, this feature does not take effect.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Enable single-access authentication.

   port-security single-access enable

   By default, single-access authentication is disabled. When multiple authentication methods are configured on an interface, users can proceed with other methods of authentication after passing one.

Specifying a preauthentication domain for port security users on a port

About this task

A preauthentication domain accommodates 802.1X, Web authentication, and MAC authentication users that have not performed authentication. A preauthentication domain is applicable to the following scenarios:

- A user accesses the network for the first time. This scenario is applicable only to 802.1X and Web authentication users.
- A user fails authentication, but no Auth-Fail domain is configured.
- No server is reachable, but the critical domain is not configured.

When a port is configured with a preauthentication domain, authentication users that access that port will be assigned authorization attributes (including ACL, VSI, , microsegment, and VLAN) in the preauthentication domain after they are assigned to the preauthentication domain. They can access only network resources permitted in the preauthentication domain. If they pass authentication, AAA will assign new authorization information to them.

Restrictions and guidelines

If the ACL, VSI, , microsegment, and VLAN authorization settings in the current preauthentication domain have changes, the changes take effect only on users that are assigned to the preauthentication domain after the changes are made. Users that have been assigned to the preauthentication domain before the changes are made still use the original settings.

On a port, port security preauthentication domain is mutually exclusive with the following settings:

- 802.1X guest VLAN, guest VSI, Auth-Fail VLAN, Auth-Fail VSI, critical VLAN, critical VSI, critical microsegment, and critical profile settings.
- MAC authentication guest VLAN, guest VSI, critical VLAN, critical VSI, critical microsegment, and critical profile settings.
- Web authentication Auth-Fail VLAN settings.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Specify a preauthentication domain for port security users.

   port-security pre-auth domain *isp-name*

   By default, no preauthentication domain is specified for port security users on a port.

Specifying a domain for port security users redirected to an unavailable URL

About this task

During user authentication, if the Web server specified by the redirect URL is unavailable, users cannot be redirected to the Web authentication page on the Web server. As a result, the users cannot come online. To allow users to access the resources in an ISP domain when the redirect URL is unavailable, use this feature to specify that ISP domain for the users.

Restrictions and guidelines

This feature is applicable only to MAC authentication and Web authentication users.

This feature is mutually exclusive with the following 802.1X, MAC authentication, and Web authentication settings:

- Guest VLAN and VSI settings.
- Auth-Fail VLAN and VSI settings.
- Critical VLAN, VSI, and microsegment settings.
- Critical profile settings.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify a domain for port security users redirected to an unavailable URL.
   port-security url-unavailable domain *isp-name*
   By default, no domain is specified for port security users redirected to an unavailable URL.

Configuring static users for port access authentication

Configuring a static user range for port access authentication

About this task

Typically, administrators assign static IP addresses to dumb terminals such as printers. For such users, you can configure them as static users to facilitate flexible authentication. After you configure a dumb terminal as a static user, the device can use the static user information such as the user IP address as the username to perform authentication for the dumb terminal on the user's access interface. For this purpose, make sure 802.1X, Web authentication, or MAC authentication is enabled on the user's access interface.

The device uses the following rules to handle static users:

- If two users have different IP addresses and MAC addresses, the users are different static user. They both can come online.
- If a user that has the same MAC address as an online static user attempts to come online, the device does not trigger authentication for the new user, regardless of whether the users use the same IP address. To update the IP addresses of static users, use the **port-security static-user update-ip enable** command.
- If users that have the same IP address but different MAC addresses come online, the device handles the users as follows:
  - If a MAC address has been bound to the IP address, the user that has the bound MAC address can come online. Users with other MAC addresses cannot trigger authentication to come online.
  - If no MAC address is bound to the IP address, the device allows all users that have IP addresses belonging to a static user IP range to trigger authentication to come online.

Restrictions and guidelines

Modification to a static user range does not affect static users that are already online. The modification takes effect only on static users that will come online.

In the public network or the same VPN instance, the IP address ranges for all static user ranges cannot overlap.

When the usernames of static users are their IP or MAC addresses, do not enable RESTful server-assisted automatic MAC authentication user recovery. If you enable RESTful server-assisted automatic MAC authentication user recovery, the device will recover static users as MAC authentication users after the device reboots or recovers from a failure. For more information about RESTful server-assisted automatic MAC authentication user recovery, see "Configuring MAC authentication."

Procedure

1. Enter system view.

   system-view

2. Configure a static user range for port access authentication.

   port-security static-user { **ip** | **ipv6** } *start-ip-address* [*end-ip-address*] [**vpn-instance** *vpn-instance-name*] [**domain** *isp-name* | [**interface** *interface-type interface-number* [**detect**]] **vlan** *vlan-id* | **mac** *mac-address* | **keep-online**] *

   By default, no static user ranges are configured for port access authentication.

3. (Optional.) Configure the username format used by static users when they come online.

   port-security static-user user-name-format { **ip-address** | **mac-address** | **system-name** }

   By default, the username of a static user is in the format of *SysnameIP*, in which *Sysname* is the name of the access device and *IP* is the user IP address.

4. (Optional.) Configure the user account format when MAC addresses of static users are used as their usernames.

   port-security static-user user-name-format mac-address { **one-section** | { **six-section** | **three-section** } **delimiter** { **colon** | **hyphen** } } [**uppercase**] [**password-with-mac**]

   By default, a static user does not have a password when its MAC address is used as its username. The MAC address contains three hyphen-separated sections and letters in the MAC address are in lower case.

5. (Optional.) Configure a password for static users.

   port-security static-user password { **cipher** | **simple** } *string*

   By default, no password is configured for static users.

6. (Optional.) Set the offline detect period for static users.

   port-security static-user timer offline-detect time-value

   By default, the offline detect period is 5 minutes for static users.

7. (Optional.) Set the interval at which the device actively sends ARP packets to trigger authentication for static users.

   port-security static-user timer detect-period *time-value*

   By default, the device actively sends ARP packets to trigger authentication for static users at intervals of 3 minutes.

Enabling static user IP update

About this task

After you use the **port-security static-user** command to configure an IP address range for a static user range, endpoints at IP addresses in the specified IP address range will come online as static users. If the IP address of an endpoint changes, the endpoint might send abnormal ARP packets to the access device when it comes online. The source IP address of these ARP packets does not belong to the specified IP address range. This issue triggers the device to update the IP address of the endpoint when static user IP update is enabled. After address update, the endpoint is no longer a static user. As a result, the endpoint is logged off.

By default, the device does not update IP addresses for static users when it receives ARP packets with source IP address not belonging to the specified IP address range from these users. This setting prevents the ARP packets from logging off online static users. To trace IP address changes for endpoints, you can enable static user IP update to allow the device to update the IP addresses of static users.

Restrictions and guidelines

Use static user IP update in conjunction with DHCP snooping, ARP snooping, DHCPv6 snooping, or ND snooping. To receive notifications about IP address changes from a snooping module, you must enable the corresponding snooping feature.

Procedure

1. Enter system view.
   system-view
2. Enable static user IP update.
   port-security static-user update-ip enable
   By default, static user IP update is disabled.

Setting the maximum number of concurrent static users allowed on a port

About this task

Set the maximum number of concurrent static users on a port to prevent the system resources from being overused. When the maximum number is reached, the port denies subsequent static users.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the maximum number of concurrent static users allowed on the port.
   port-security static-user max-user *max-number*
   By default, a port supports a maximum of 4294967295 concurrent static users.

Specifying an ACL to match the MAC addresses of static users

About this task

Typically, endpoints that match static user IP ranges come online as static users. However, the device recognizes the endpoints as MAC authentication users instead of static users in the following situations:

- The first packet sent by an endpoint is a Layer 2 packet that does not contain an IP address. In this situation, the packet triggers MAC authentication first.
- An endpoint has both IPv4 and IPv6 addresses and the first packet sent by the endpoint is an IPv6 packet, but only static user IPv4 ranges are configured on the device. In this situation, the packet triggers MAC authentication first.

To resolve the issues, configure this feature to use MAC address as the criterion to match static users. With this feature, the device allows users that match the specified ACL to trigger authentication and come online only as static users. The users cannot trigger other authentication processes.

Restrictions and guidelines

The specified ACL must be a Layer 2 ACL. The ACL can contain only permit rules with the source MAC range criteria.

Procedure

1. Enter system view.

   system-view

2. Specify an ACL to match the MAC addresses of static users.

   port-security static-user match-mac acl *acl-number*

   By default, no ACL is specified to match the MAC addresses of static users.

Configuring a port security authentication profile

About this task

To implement fast port security authentication on users, the device uses port security authentication profiles for unified management of the access authentication configuration. In a port security authentication profile, you can bind 802.1X and MAC authentication access profiles and configure the authentication order to control user access.

After an interface is bound to a port security authentication profile, the interface uses the configuration of the bound profile to perform authentication for access users.

- For functions available in both interface view and port security authentication profile view (commands might differ), the functions configured in interface view are deleted regardless of whether they are configured in the profile view or not. To use those functions on the port, configure them in the bound port security authentication profile.
- For the functions available only in interface view, they can take effect on the interface after being configured.
- The authentication mode used on a port is determined by the 802.1X access profile and MAC authentication access profile bound to the port security authentication profile.

After an interface is bound to a port security authentication profile, the interface supports only MAC-based authentication. If you bind a MAC authentication access profile to the port security authentication profile, the interface will use MAC authentication. If you bind an 802.1X access profile

to the port security authentication profile, the interface will use 802.1X authentication in MAC-based access control.

Restrictions and guidelines

A port security authentication profile can be bound to different interfaces. An interface can be bound to only one port security authentication profile. To change the bound port security authentication profile of an interface, you must first unbind the profile from the interface.

After the authentication profile bound to an interface takes effect, deleting the bound authentication profile will cause abnormal disconnection of online users on the interface.

To avoid authentication anomalies, do not execute the **dot1x port-method portbased** command on an interface that is bound to a port security authentication profile.

Procedure

1. Enter system view.
   system-view
2. Create a port security authentication profile and enter its view.
   port-security authentication-profile name *profile-name*
   By default, no port security authentication profiles exist.
3. Bind an 802.1X access profile to the port security authentication profile.
   dot1x access-profile *profile-name*
   By default, no 802.1X access profile is bound to a port security authentication profile.
4. Bind a MAC authentication access profile to the port security authentication profile.
   mac-authentication access-profile *profile-name*
   By default, no MAC authentication access profile is bound to a port security authentication profile.
5. Exit port security authentication profile view.
   quit
6. Enter interface view.
   interface *interface-type interface-number*
7. Bind a port security authentication profile to the interface.
   port-security authentication-profile *profile-name*
   By default, no port security authentication profile is bound to an interface.

Configuring the port security authentication order (port security authentication profile view)

About this task

You can configure the access authentication order on a port enabled with both 802.1X and MAC authentication methods.

- In **dot1x-mac** order, for a port to perform MAC authentication and assign an authorization VLAN or authorization VSI before it joins the 802.1X guest VLAN or guest VSI, enable parallel 802.1X and MAC authentication processing by using the **port-security auth-order dot1x-mac parallel** command and enable 802.1X guest VLAN or VSI assignment delay. For information about the commands for enabling 802.1X guest VLAN or VSI assignment delay, see 802.1X commands in Security Command Reference.

- In **mac-dot1x** order, to allow a user to go online only after the user passes both MAC and 802.1X authentication, enable multi-authentication mode by using the **port-security auth-order mac-dot1x multiple** command.

Restrictions and guidelines

Changing the access authentication order by using the **port-security auth-order** command will result in authentication failure for users that are currently being authenticated. The users must trigger authentication again in order to go online. To avoid such authentication failures, change the authentication order only when necessary.

If parallel 802.1X and MAC authentication processing is enabled, do not configure MAC authentication delay as a best practice.

Procedure

1. Enter system view.

   system-view

2. Enter port security authentication profile view.

   port-security authentication-profile name *profile-name*

   By default, no port security authentication profiles exist.

3. Configure the port security authentication order.

   port-security auth-order { **dot1x-mac** [**parallel**] | **mac-dot1x** [**multiple**] }

   By default, after a port receives a packet with unknown source MAC address, it performs 802.1X authentication and then MAC authentication for the user.

Configuring port security access policies

About this task

To simplify port security configuration, you can configure a port security access policy and uniformly apply it to all interfaces. By configuring match criteria and actions in port security access policy view, you can control port access users more granularly.

Restrictions and guidelines

A port security access policy only takes effect on users that come online for the first time, and does not affect the re-authentication process of existing online users.

Procedure

1. Enter system view.

   system-view

2. Create a port security access policy and enter its view.

   port-security access-policy *policy-name*

   By default, no port security access policy exists.

3. Configure the match criteria and actions for port security access users.

   if-match [**not**] **acl mac** *acl-number* **action access-type** { **dot1x** | **mac-auth** | **static** | **web-auth** } *

   By default, no match criteria and actions are configured for port security access users.

4. Apply the port security access policy globally.

   port-security apply access-policy *policy-name*

   By default, no port security access policy is applied globally.

Configuring the action to be taken on online users when their access ports go down

About this task

By default, the device immediately logs off online users when their access ports go down. When the ports come up, the users must be reauthenticated to come online. To prevent users from having to repeatedly be reauthenticated and come online in the event of frequent port flapping, you can configure the following actions for online users as needed:

- **keep-online**—Allows online users to stay online when their access ports go down. When the ports come up, the users can come online without being reauthenticated.
- **offline-delay**—Delays logging off online users when their access ports go down.
  - If the access ports come up before the delay time expires, the users can come online without being reauthenticated.
  - If the access ports do not come up before the delay time expires, the users are logged off when the delay time expires. When the ports come up, the users must be reauthenticated to come online.

Restrictions and guidelines

This feature takes effect for online users on a port only when the port automatically goes down due to link abnormalities. It does not take effect when the port goes down manually by executing the **shutdown** command.

This feature takes effect for online users on a port only if you configure it when that access port is up. It does not take effect for online users on a port if that access port is down when you configure it. In this case, any modification or deletion to the feature configuration cannot take effect on that access port.

Procedure

1. Enter system view.

   system-view
2. Enter port security authentication profile view.

   port-security authentication-profile name *profile-name*

   By default, no port security authentication profiles exist.
3. Configure the action to be taken on online users when their access ports go down.

   port-security link-down action { **keep-online** | **offline-delay** *delay-value* }

   By default, the device immediately logs off online users when their access ports go down.

Setting port security timers

About this task

Port security supports the following timers for users that have not performed authentication or that have failed authentication:

- **Periodic reauthentication timer**—Sets the interval at which the device initiates reauthentication for online 802.1X, Web authentication, or MAC authentication users in the specified domain.
- **User aging timer**—Sets the period that the device must wait before it can remove online 802.1X, Web authentication, or MAC authentication users from the specified domain.

Procedure

1. Enter system view.

   system-view

2. Set a port security timer for specific users.

   port-security timer { **reauth-period** { **auth-fail-domain** | **preauth-domain** } | **user-aging** { **auth-fail-domain** | **critical-domain** | **preauth-domain** } } *time-value*

   By default, the period for the periodic reauthentication timer is 600 seconds and the period for the user aging timer is 23 hours.

Configuring the maximum number of reauthentication attempts for users in the preauthentication domain or Auth-Fail domain

About this task

By default, the device performs periodic reauthentication for users in the preauthentication domain and Auth-Fail domain, and the number of reauthentications is not limited. Because of limited system resources, unlimited number of reauthentications without success can lead to resource contention and degraded device performance. To ensure device performance, use this feature to set the maximum number of reauthentication attempts, preventing users from occupying system resources for an extended period.

With this feature, after the maximum number of reauthentications is reached for a preauthentication or Auth-Fail domain user, the device will not trigger periodic reauthentication for that user. The user remains in the current domain and can initiate reauthentication from the client.

Restrictions and guidelines

This feature takes effect only on users who join the preauthentication domain or Auth-Fail domain after the feature command is executed.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the maximum number of reauthentication attempts for users in the preauthentication domain or Auth-Fail domain.

   port-security re-authenticate max-attempt { **auth-fail-domain | preauth-domain** } *max-attempt*

   By default, the device does not limit the number of reauthentication attempts for users in the preauthentication domain or Auth-Fail domain.

Enabling immediate reauthentication for users in the preauthentication domain or Auth-Fail domain when the authentication server becomes reachable

About this task

When all the authentication servers are unreachable or have failed, users might be assigned to the preauthentication domain or Auth-Fail domain. In normal conditions, when the periodic reauthentication feature is enabled for users in the preauthentication domain or Auth-Fail domain, the device triggers reauthentication for users only after the set reauthentication period ends. To have the device immediately trigger reauthentication for these users when any one of the authentication servers become reachable, use this feature.

When the device triggers reauthentication for preauthentication or Auth-Fail domain users, the device sets the maximum reauthentication attempts for them to the value set by using the **port-security re-authenticate max-attempt** command.

Procedure

1. Enter system view.

   system-view

2. Enable immediate reauthentication for users in the preauthentication domain or Auth-Fail domain when the authentication server becomes reachable.

   port-security reauth-trigger server-reachable { **auth-fail-domain** | **preauth-domain** }

   By default, when the authentication server changes from unreachable to reachable, the device does not immediately trigger reauthentication for users in the preauthentication domain or Auth-Fail domain.

Permitting MAC move between member ports in a TC group

About this task

Application scenarios

When the network topology changes, the STP module sends a topology change (TC) event message to notify relevant devices that the network topology has changed. TC events might cause traffic forwarding exceptions on a member port in a TC group.

To resolve this issue, use this feature to permit MAC move between member ports in a TC group when the network topology changes. MAC move allows authenticated online users on a member port in a TC group to move to the other member port in the same TC group without being authenticated again. The process is as described in "Operating mechanism."

Use this feature on the device if the device is connected to users that cannot actively send packets to trigger MAC move when the network topology changes.

Operating mechanism

The operating mechanism of this feature is as follows:

- If a member port in a TC group is up and receives a TC event message, the device searches for online authenticated users that come online from that port. In addition, the device sends ARP or

NS detection packets to these users at detection intervals through the other member port in the same TC group.

- If the other member port receives a response packet for a user, that user moves to the other member port and comes online without being authenticated.
- If the other member port does not receive any response packets for a user after the device has made the maximum number of attempts for sending a detection packet, the device determines that the network topology of that TC group does not change. It does not move the user to the other member port.

- If the network topology changes because a member port in a TC group goes down, the device does not wait to receive TC event messages sent by the STP module or actively detect online authenticated users on that port from the other member port. Instead, it immediately moves the online authenticated users on that port to the other member port without authenticating them. To detect whether the users can come online correctly on the other member port, you can enable offline detection or ARP or NS packet detection on the other member port.

For more information about TC groups, see spanning tree configuration in *Layer 2—LAN Switching Configuration Guide*.

Restrictions and guidelines

This feature takes effect only on static users, MAC authentication users, and 802.1X users.

As a best practice to ensure successful MAC move between member ports in a TC group, the member ports in that TC group must use the same settings.

Procedure

1. Enter system view.

   `system-view`

2. Allow authenticated online users to move between member ports in the same TC group without being authenticated again.

   `port-security topology-change free-mac-move`

   By default, the device does not allow authenticated online users to move between member ports in the same TC group without being authenticated again.

3. Set the interval at which the device actively sends ARP or NS detection packets when the network topology changes.

   `port-security topology-change detect-period` *`time-value`*

   By default, the device actively sends ARP or NS detection packets at intervals of 5 seconds when the network topology changes.

4. Set the maximum number of attempts for sending an ARP or NS detection packet when the network topology changes.

   `port-security topology-change detect-retry` *`retries`*

   By default, the device attempts to send an ARP or NS detection packet for a maximum of three times when the network topology changes.

Enabling SNMP notifications for port security

About this task

Use this feature to report critical port security events to an NMS. For port security event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Enable SNMP notifications for port security.

   snmp-agent trap enable port-security [**ac-creation-failure** | **acl-author-failure** | **acl-author-success** | **address-learned** | **dot1x-failure** | **dot1x-ip-change** | **dot1x-logoff** | **dot1x-logon** | **intrusion** | **intrusion-recover** | **mac-auth-failure** | **mac-auth-logoff** | **mac-auth-logon** | **mac-auth-ip-change** | **mac-auth-not-support** | **ntk-ineffective** | **port-mode-ineffective** | **url-author-failure** | **url-author-success**] *

   By default, SNMP notifications are disabled for port security.

Enabling port security user logging

About this task

This feature enables the device to generate logs about port security users and send the logs to the information center. For the logs to be output correctly, you must also configure the information center on the device. For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

To prevent excessive port security user log entries, use this feature only if you need to analyze abnormal port security user events.

Procedure

1. Enter system view.

   system-view

2. Enable port security user logging.

   port-security access-user log enable [**failed-authorization** | **mac-learning** | **violation** | **vlan-mac-limit**] *

   By default, port security user logging is disabled.

   If you do not specify any parameters, this command enables all types of port security user logs.

Display and maintenance commands for port security

Execute **display** commands in any view:

Task	Command
Display the port security configuration, operation information, and statistics.	**display port-security** [**interface** *interface-type interface-number*]
Display entries for online port security access users.	**display port-security access-user** [**m-lag** [**local** | **peer**]] [**access-type** { **dot1x** | **mac-auth** | **web-auth** | **static** } | **domain** *domain-name* | **microsegment** *microsegment-id* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **slot** *slot-number*] * [**brief**] **display port-security access-user** [**m-lag**

Task	Command
	[**local** | **peer**]] [**access-type** { **dot1x** | **mac-auth** | **web-auth** | **static** } | **domain** *domain-name* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** | **url-unavailable-domain** } | **slot** *slot-number*] *
Display configuration information for port security authentication profiles	**display port-security authentication-profile** [**name** *profile-name*]
Display information about blocked MAC addresses.	**display port-security mac-address block** [**interface** *interface-type interface-number*] [**vlan** *vlan-id*] [**count**]
Display MAC address entries synchronized from the M-LAG peer.	**display port-security mac-address m-lag sync-from-peer** [**user-mac** *mac-address*]
Display MAC address entries synchronized to the M-LAG peer.	**display port-security mac-address m-lag sync-to-peer** [**interface** *interface-type interface-number* | **user-mac** *mac-address*]
Display information about secure MAC addresses.	**display port-security mac-address security** [**interface** *interface-type interface-number*] [**vlan** *vlan-id*] [**count**]
Display static user configuration information.	**display port-security static-user**
Display online static users.	**display port-security static-user connection** [[**m-lag** [**local** | **peer**]] [**interface** *interface-type interface-number* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **slot** *slot-number* | **user-name** *user-name*] | { **ip** | **ipv6** } *ip-address* | **mac** *mac-address*]
Display port security statistics.	**display port-security statistics** [**slot** *slot-number*]
Log off online static users.	**reset port-security static-user** [**interface** *interface-type interface-number* | { **ip** | **ipv6** } *ip-address* | **mac** *mac-address* | **online-type** { **auth-fail-domain** | **critical-domain** | **preauth-domain** | **success** } | **user-name** *user-name*]
Clear port security statistics.	**reset port-security statistics**

Port security configuration examples

Example: Configuring port security in autoLearn mode

Network configuration

As shown in Figure 3, configure GigabitEthernet 1/0/1 on the device to meet the following requirements:

- Accept up to 64 users without authentication.
- Be permitted to learn and add MAC addresses as sticky MAC addresses, and set the secure MAC aging timer to 30 minutes.
- Stop learning MAC addresses after the number of secure MAC addresses reaches 64. If any frame with an unknown MAC address arrives, intrusion protection starts, and the port shuts down and stays silent for 30 seconds.

Figure 3 Network diagram

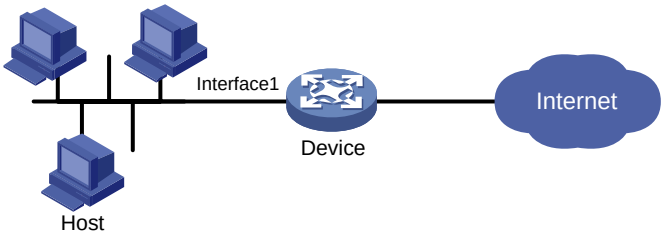


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Enable port security.

```
<Device> system-view
[Device] port-security enable
```

Set the secure MAC aging timer to 30 minutes.

```
[Device] port-security timer autolearn aging 30
```

Set port security's limit on the number of secure MAC addresses to 64 on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] port-security max-mac-count 64
```

Set the port security mode to autoLearn.

```
[Device-GigabitEthernet1/0/1] port-security port-mode autolearn
```

Configure the port to be silent for 30 seconds after the intrusion protection feature is triggered.

```
[Device-GigabitEthernet1/0/1] port-security intrusion-mode disableport-temporarily
[Device-GigabitEthernet1/0/1] quit
[Device] port-security timer disableport 30
```

Verifying the configuration

Verify the port security configuration.

```
[Device] display port-security interface gigabitethernet 1/0/1
Global port security parameters:
```

```
  Port security                              : Enabled
  M-LAG load sharing mode (criterion)        : Distributed (local)
  M-LAG member's authentication scope        : Local M-LAG interfaces
  M-LAG member configuration conflict        : Unknown
  AutoLearn aging time                       : 30 min
  Disableport timeout                        : 30 sec
  Blockmac timeout                           : 180 sec
  MAC move                                   : Denied
  Authorization fail                         : Online
  NAS-ID profile                             : Not configured
  Dot1x-failure trap                         : Disabled
  Dot1x-logon trap                           : Disabled
  Dot1x-logoff trap                          : Disabled
  Intrusion trap                             : Disabled
  Intrusion-recover trap                     : Disabled
  Address-learned trap                       : Disabled
  Mac-auth-failure trap                      : Disabled
  Mac-auth-logon trap                        : Disabled
  Mac-auth-logoff trap                       : Disabled
  Mac-auth-not-support trap                  : Disabled
  AC-creation-failure trap                   : Disabled
  ACL-author-failure trap                    : Disabled
  ACL-author-success trap                    : Disabled
  URL-author-failure trap                    : Disabled
  URL-author-success trap                    : Disabled
  NTK-ineffective trap                       : Disabled
  Port-mode-ineffective trap                 : Disabled
  Open authentication                        : Disabled
  Traffic-statistics                         : Disabled
  User aging period for preauth domain       : 82800 sec
  User aging period for Auth-Fail domain     : 82800 sec
  User aging period for critical domain      : 82800 sec
  Reauth period for preauth domain           : 600 sec
  Reauth period for Auth-Fail domain         : 600 sec
  MAC move for topology change protection    : Denied
    Topology change detection period         : 5 sec
    Max detection attempts                   : 3
  Record max-number                          : 4096
  Record aging period                        : 1440 min
  OUI value list                             :
   Index :  1           Value : 123401

GigabitEthernet1/0/1 is link-up
  Port mode                                  : autoLearn
  Pre-auth domain                            : Not configured
  URL-unavailable domain                     : Not configured
  NeedToKnow mode                            : Disabled
  Intrusion protection mode                  : DisablePortTemporarily
```

```
   Strict intrusion protection             : Disabled
   Security MAC address attribute
       Learning mode                       : Sticky
       Aging type                          : Periodical
   Max secure MAC addresses                : 64
   Current secure MAC addresses            : 0
   Authorization                           : Permitted
   NAS-ID profile                          : Not configured
   Free VLANs                              : Not configured
   Open authentication                     : Disabled
   MAC-move VLAN check bypass              : Disabled
   Server-reachable reauth
       Preauth domain                      : Disabled
       Auth-Fail domain                    : Disabled
```

The port allows for MAC address learning, and you can view the number of learned MAC addresses in the **Current secure MAC addresses** field.

Display additional information about the learned MAC addresses.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] display this
#
interface GigabitEthernet1/0/1
 port-security max-mac-count 64
 port-security port-mode autolearn
 port-security mac-address security sticky 0002-0000-0015 vlan 1
 port-security mac-address security sticky 0002-0000-0014 vlan 1
 port-security mac-address security sticky 0002-0000-0013 vlan 1
 port-security mac-address security sticky 0002-0000-0012 vlan 1
 port-security mac-address security sticky 0002-0000-0011 vlan 1
#
```

Use the **`display port-security interface`** command to verify that the port security mode changes to **secure** after the number of MAC addresses learned by the port reaches 64.

Use the **`display interface`** command to verify that the port will be disabled for 30 seconds after it receives a frame with an unknown MAC address. (Details not shown.)

After the port is re-enabled, delete several secure MAC addresses. (Details not shown.)

Verify that the port security mode of the port changes to **autoLearn**, and the port can learn MAC addresses again. (Details not shown.)

Example: Configuring port security in userLoginWithOUI mode

Network configuration

As shown in Figure 4, a client is connected to the device through GigabitEthernet 1/0/1. The device authenticates the client with a RADIUS server in ISP domain **sun**. If the authentication succeeds, the client is authorized to access the Internet.

- The RADIUS server at 192.168.1.2 acts as the primary authentication server and the secondary accounting server. The RADIUS server at 192.168.1.3 acts as the secondary authentication

server and the primary accounting server. The shared key for authentication is **name**, and the shared key for accounting is **money**.

- All users use the authentication, authorization, and accounting methods of ISP domain **sun**.
- The RADIUS server response timeout time is 5 seconds. The maximum number of RADIUS packet retransmission attempts is 5. The device sends real-time accounting packets to the RADIUS server at 15-minute intervals, and sends usernames without domain names to the RADIUS server.

Configure GigabitEthernet 1/0/1 to allow only one 802.1X user and a user that uses one of the specified OUI values to be authenticated.

Figure 4 Network diagram

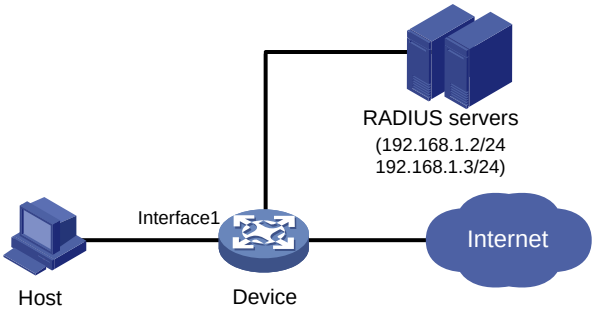


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

The following configuration steps cover some AAA/RADIUS configuration commands. For more information about the commands, see *Security Command Reference*.

Make sure the host and the RADIUS server can reach each other.

1. Configure AAA:

Configure a RADIUS scheme named **radsun**.

```
<Device> system-view
[Device] radius scheme radsun
[Device-radius-radsun] primary authentication 192.168.1.2
[Device-radius-radsun] primary accounting 192.168.1.3
[Device-radius-radsun] secondary authentication 192.168.1.3
[Device-radius-radsun] secondary accounting 192.168.1.2
[Device-radius-radsun] key authentication simple name
[Device-radius-radsun] key accounting simple money
[Device-radius-radsun] timer response-timeout 5
[Device-radius-radsun] retry 5
[Device-radius-radsun] timer realtime-accounting 15
[Device-radius-radsun] user-name-format without-domain
[Device-radius-radsun] quit
```

Configure ISP domain **sun**.

```
[Device] domain sun
[Device-isp-sun] authentication lan-access radius-scheme radsun
[Device-isp-sun] authorization lan-access radius-scheme radsun
```

```
[Device-isp-sun] accounting lan-access radius-scheme radsun
[Device-isp-sun] quit
```

2. Configure 802.1X:

Set the 802.1X authentication method to CHAP. By default, the authentication method for 802.1X is CHAP.

```
[Device] dot1x authentication-method chap
```

Specify ISP domain **sun** as the mandatory authentication domain for 802.1X users on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] dot1x mandatory-domain sun
[Device-GigabitEthernet1/0/1] quit
```

3. Configure port security:

Enable port security.

```
[Device] port-security enable
```

Add five OUI values. (You can add up to 16 OUI values. The port permits only one user matching one of the OUIs to pass authentication.)

```
[Device] port-security oui index 1 mac-address 1234-0100-1111
[Device] port-security oui index 2 mac-address 1234-0200-1111
[Device] port-security oui index 3 mac-address 1234-0300-1111
[Device] port-security oui index 4 mac-address 1234-0400-1111
[Device] port-security oui index 5 mac-address 1234-0500-1111
```

Set the port security mode to userLoginWithOUI.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] port-security port-mode userlogin-withoui
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Verify that GigabitEthernet 1/0/1 allows only one 802.1X user to be authenticated.

```
[Device] display port-security interface gigabitethernet 1/0/1
Global port security parameters:
   Port security                          : Enabled
   M-LAG load sharing mode (criterion)    : Distributed (local)
   M-LAG member's authentication scope    : Local M-LAG interfaces
   M-LAG member configuration conflict    : Unknown
   AutoLearn aging time                   : 30 min
   Disableport timeout                    : 30 sec
   Blockmac timeout                       : 180 sec
   MAC move                               : Denied
   Authorization fail                     : Online
   NAS-ID profile                         : Not configured
   Dot1x-failure trap                     : Disabled
   Dot1x-logon trap                       : Disabled
   Dot1x-logoff trap                      : Disabled
   Intrusion trap                         : Disabled
   Intrusion-recover trap                 : Disabled
   Address-learned trap                   : Disabled
   Mac-auth-failure trap                  : Disabled
   Mac-auth-logon trap                    : Disabled
   Mac-auth-logoff trap                   : Disabled
```

```
  Mac-auth-not-support trap               : Disabled
  AC-creation-failure trap                : Disabled
  ACL-author-failure trap                 : Disabled
  ACL-author-success trap                 : Disabled
  URL-author-failure trap                 : Disabled
  URL-author-success trap                 : Disabled
  NTK-ineffective trap                    : Disabled
  Port-mode-ineffective trap              : Disabled
  Open authentication                     : Disabled
  Traffic-statistics                      : Disabled
  User aging period for preauth domain    : 82800 sec
  User aging period for Auth-Fail domain  : 82800 sec
  User aging period for critical domain   : 82800 sec
  Reauth period for preauth domain        : 600 sec
  Reauth period for Auth-Fail domain      : 600 sec
  MAC move for topology change protection : Denied
    Topology change detection period      : 5 sec
    Max detection attempts                : 3
  Record max-number                       : 4096
  Record aging period                     : 1440 min
  OUI value list                          :
      Index :  1       Value : 123401
      Index :  2       Value : 123402
      Index :  3       Value : 123403
      Index :  4       Value : 123404
      Index :  5       Value : 123405

GigabitEthernet1/0/1 is link-up
  Port mode                               : userLoginWithOUI
  Pre-auth domain                         : Not configured
  URL-unavailable domain                  : Not configured
  NeedToKnow mode                         : Disabled
  Intrusion protection mode               : NoAction
  Strict intrusion protection             : Disabled
  Security MAC address attribute
      Learning mode                       : Sticky
      Aging type                          : Periodical
  Max secure MAC addresses                : Not configured
  Current secure MAC addresses            : 1
  Authorization                           :Permitted
  NAS-ID profile                          : Not configured
  Free VLANs                              : Not configured
  Open authentication                     : Disabled
  MAC-move VLAN check bypass              : Disabled
  Server-reachable reauth
      Preauth domain                      : Disabled
      Auth-Fail domain                    : Disabled
```

Use the **display dot1x** command to display information about the online 802.1X user to verify 802.1X configuration.

Verify that the port also allows one user whose MAC address has an OUI among the specified OUIs to pass authentication.

```
[Device] display mac-address interface gigabitethernet 1/0/1
MAC Address     VLAN ID   State          Port/Nickname            Aging
1234-0300-0011  1         Learned        GE1/0/1                  Y
1234-0300-0011  1         Learned        HGE1/0/1                 Y
1234-0300-0011  1         Learned        XGE1/0/1                 Y
1234-0300-0011  1         Learned        WGE1/0/1                 Y
1234-0300-0011  1         Learned        FGE1/0/1                 Y
1234-0300-0011  1         Learned        Eth1/0/1                 Y
1234-0300-0011  1         Learned        XGE3/0/1GE1/0/1                 Y
1234-0300-0011  1         Learned        XGE3/0/1GE1/0/1                 Y
```

Example: Configuring port security in macAddressElseUserLoginSecure mode

Network configuration

As shown in Figure 5, a client is connected to the device through GigabitEthernet 1/0/1. The device authenticates the client by a RADIUS server in ISP domain **sun**. If the authentication succeeds, the client is authorized to access the Internet.

Configure GigabitEthernet 1/0/1 of the device to meet the following requirements:

- Allow more than one MAC authenticated user to log on.
- For 802.1X users, perform MAC authentication first and then, if MAC authentication fails, 802.1X authentication. Allow only one 802.1X user to log on.
- Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in upper case.
- Set the total number of MAC authenticated users and 802.1X authenticated users to 64.
- Enable NTK (**ntkonly** mode) to prevent frames from being sent to unknown MAC addresses.

Figure 5 Network diagram

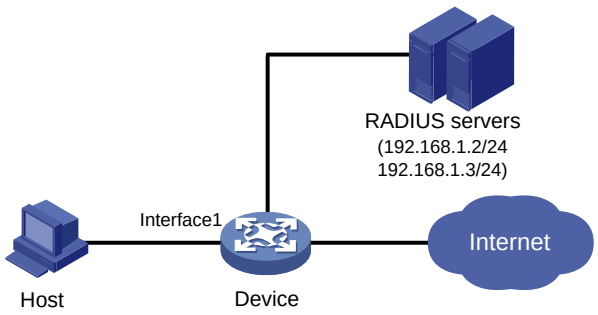


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Make sure the host and the RADIUS server can reach each other.

1. Configure RADIUS authentication/accounting and ISP domain settings.
2. Configure port security:

   # Enable port security.

   ```
   <Device> system-view
   [Device] port-security enable
   ```

   # Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hexadecimal notation with hyphens, and letters are in upper case.

   ```
   [Device] mac-authentication user-name-format mac-address with-hyphen uppercase
   ```

   # Specify the MAC authentication domain.

   ```
   [Device] mac-authentication domain sun
   ```

   # Set the 802.1X authentication method to CHAP. By default, the authentication method for 802.1X is CHAP.

   ```
   [Device] dot1x authentication-method chap
   ```

   # Set port security's limit on the number of MAC addresses to 64 on the port.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] port-security max-mac-count 64
   ```

   # Set the port security mode to macAddressElseUserLoginSecure.

   ```
   [Device-GigabitEthernet1/0/1] port-security port-mode mac-else-userlogin-secure
   ```

   # Specify ISP domain **sun** as the mandatory authentication domain for 802.1X users.

   ```
   [Device-GigabitEthernet1/0/1] dot1x mandatory-domain sun
   ```

   # Set the NTK mode of the port to ntkonly.

   ```
   [Device-GigabitEthernet1/0/1] port-security ntk-mode ntkonly
   [Device-GigabitEthernet1/0/1] quit
   ```

Verifying the configuration

Verify the port security configuration.

```
[Device] display port-security interface gigabitethernet 1/0/1
Global port security parameters:
   Port security                          : Enabled
   M-LAG load sharing mode (criterion)    : Distributed (local)
   M-LAG member's authentication scope    : Local M-LAG interfaces
   M-LAG member configuration conflict    : Unknown
   AutoLearn aging time                   : 30 min
   Disableport timeout                    : 30 sec
   Blockmac timeout                       : 180 sec
   MAC move                               : Denied
   Authorization fail                     : Online
   NAS-ID profile                         : Not configured
   Dot1x-failure trap                     : Disabled
   Dot1x-logon trap                       : Disabled
   Dot1x-logoff trap                      : Disabled
   Intrusion trap                         : Disabled
   Intrusion-recover trap                 : Disabled
   Address-learned trap                   : Disabled
   Mac-auth-failure trap                  : Disabled
```

```
  Mac-auth-logon trap                    : Disabled
  Mac-auth-logoff trap                   : Disabled
  Mac-auth-not-support trap              : Disabled
  AC-creation-failure trap               : Disabled
  ACL-author-failure trap                : Disabled
  ACL-author-success trap                : Disabled
  URL-author-failure trap                : Disabled
  URL-author-success trap                : Disabled
  NTK-ineffective trap                   : Disabled
  Port-mode-ineffective trap             : Disabled
  Open authentication                    : Disabled
  Traffic-statistics                     : Disabled
  User aging period for preauth domain   : 82800 sec
  User aging period for Auth-Fail domain : 82800 sec
  User aging period for critical domain  : 82800 sec
  Reauth period for preauth domain       : 600 sec
  Reauth period for Auth-Fail domain     : 600 sec
  MAC move for topology change protection : Denied
    Topology change detection period     : 5 sec
    Max detection attempts               : 3
  Record max-number                      : 4096
  Record aging period                    : 1440 min
  OUI value list                         :

 GigabitEthernet1/0/1 is link-up
  Port mode                              : macAddressElseUserLoginSecure
  Pre-auth domain                        : Not configured
  URL-unavailable domain                 : Not configured
  NeedToKnow mode                        : NeedToKnowOnly
  Intrusion protection mode              : NoAction
  Strict intrusion protection            : Disabled
  Security MAC address attribute
     Learning mode                       : Sticky
     Aging type                          : Periodical
  Max secure MAC addresses               : 64
  Current secure MAC addresses           : 0
  Authorization                          : Permitted
  NAS-ID profile                         : Not configured
  Free VLANs                             : Not configured
  Open authentication                    : Disabled
  MAC-move VLAN check bypass             : Disabled
  Server-reachable reauth
      Preauth domain                     : Disabled
      Auth-Fail domain                   : Disabled
```

After users pass authentication, display MAC authentication information. Verify that GigabitEthernet 1/0/1 allows multiple MAC authentication users to be authenticated.

```
[Device] display mac-authentication interface gigabitethernet 1/0/1
 Global MAC authentication parameters:
```

```
   MAC authentication                  : Enabled
   Authentication method               : PAP
   M-LAG member configuration conflict : Unknown
   User name format                    : MAC address in uppercase(XX-XX-XX-XX-XX-XX)
           Username                    : mac
           Password                    : Not configured
   MAC range accounts                  : 0
          MAC address          Mask                 Username
   Offline detect period                      : 300 s
   Quiet period                               : 180 s
   Server timeout                             : 100 s
   Reauth period                              : 3600 s
   User aging period for critical VLAN        : 1000 s
   User aging period for critical VSI         : 1000 s
   User aging period for guest VLAN           : 1000 s
   User aging period for guest VSI            : 1000 s
   User aging period for critical microsegment: 1000 s
   Temporary user aging period                : 60 s
   Authentication domain                      : sun
   HTTP proxy port list                       : Not configured
   HTTPS proxy port list                      : Not configured
 Online MAC-auth wired users                  : 3

 Silent MAC users:
          MAC address       VLAN ID  From port               Port index

GigabitEthernet1/0/1 is link-up
   MAC authentication                      : Enabled
   Carry User-IP                           : Disabled
   Authentication domain                   : Not configured
   Auth-delay timer                        : Disabled
   Periodic reauth                         : Disabled
   Re-auth server-unreachable              : Logoff
   Guest VLAN                              : Not configured
   Guest VLAN auth-period                  : 30 s
   Critical VLAN                           : Not configured
   Critical voice VLAN                     : Disabled
   Host mode                               : Single VLAN
   Offline detection                       : Enabled
   Authentication order                    : Default
   User aging                              : Enabled
   Server-recovery online-user-sync        : Enabled

   Guest VSI                               : Not configured
   Guest VSI reauthentication              : Enabled
   Guest VSI auth-period                   : 30 s
   Critical VSI                            : Not configured
   Critical microsegment ID                : Not configured
```

```
  Critical profile                          : Not configured
  URL user logoff                           : No
  Auto-tag feature                          : Disabled
  VLAN tag configuration ignoring           : Disabled
  Max online users                          : 4294967295
  Max online preauth-domain users           : 4294967295
  Max online Auth-Fail-domain users         : 4294967295
  Auth-server-unavailable escape            : Disabled
  Authentication attempts                   : successful 3, failed 7
  Current online users                      : 3
         MAC address       Auth state
         1234-0300-0011    Authenticated
         1234-0300-0012    Authenticated
         1234-0300-0013    Authenticated
```

Display 802.1X authentication information. Verify that GigabitEthernet 1/0/1 allows only one 802.1X user to be authenticated.

```
[Device] display dot1x interface gigabitethernet 1/0/1
 Global 802.1X parameters:
  802.1X authentication                     : Enabled
  M-LAG member configuration conflict       : Unknown
  CHAP authentication                       : Enabled
  Max-tx period                             : 30 s
  Handshake period                          : 15 s
  Offline detect period                     : 300 s
  Quiet timer                               : Disabled
      Quiet period                          : 60 s
  Supp timeout                              : 30 s
  Server timeout                            : 100 s
  Reauth period                             : 3600 s
  Unicast-trigger quiet period              : 0 s
  Max auth requests                         : 2
  User aging period for Auth-Fail VLAN      : 1000 s
  User aging period for Auth-Fail VSI       : 1000 s
  User aging period for critical VLAN       : 1000 s
  User aging period for critical VSI        : 1000 s
  User aging period for guest VLAN          : 1000 s
  User aging period for guest VSI           : 1000 s
  EAD assistant function                    : Disabled
      Permit authentication-escape          : Disabled
      EAD timeout                           : 30 min
  Domain delimiter                          : @
  Max-user-alarm trigger threshold          : 100%
  Max-user-alarm clear threshold            : 50%
 Online 802.1X wired users                  : 1

 GigabitEthernet1/0/1  is link-up
  802.1X authentication               : Enabled
  Access profile name                 : Not configured
```

```
Handshake                              : Enabled
Handshake reply                        : Disabled
Handshake security                     : Disabled
Unicast trigger                        : Disabled
Periodic reauth                        : Disabled
Port role                              : Authenticator
Authorization mode                     : Auto
Port access control                    : MAC-based
Multicast trigger                      : Enabled
Mandatory auth domain                  : sun
Guest VLAN                             : Not configured
Auth-Fail VLAN                         : Not configured
Critical VLAN                          : Not configured
Critical voice VLAN                    : Disabled
Add Guest VLAN delay                   : Disabled
Re-auth server-unreachable             : Logoff
Max online users                       : 4294967295
Max online preauth-domain users        : 4294967295
Max online Auth-Fail-domain users      : 4294967295
User IP freezing                       : Disabled
Send Packets Without Tag               : Disabled
Max Attempts Fail Number               : Not configured
Guest VSI                              : Not configured
Auth-Fail VSI                          : Not configured
Critical VSI                           : Not configured
Add Guest VSI delay                    : Disabled
Critical microsegment ID               : Not configured
Critical profile                       : Not configured
Unauthenticated-user aging             : Enabled
Server-recovery online-user-sync       : Enabled
Auth-Fail EAPOL                        : Disabled
Critical EAPOL                         : Disabled
Discard duplicate EAPOL-Start          : No
Auth-server-unavailable escape         : Disabled

EAPOL packets: Tx 16331, Rx 102
Sent EAP Request/Identity packets : 16316
     EAP Request/Challenge packets: 6
     EAP Success packets: 4
     EAP Failure packets: 5
Received EAPOL Start packets : 6
         EAPOL LogOff packets: 2
         EAP Response/Identity packets : 80
         EAP Response/Challenge packets: 6
         Error packets: 0
Online 802.1X users: 1
       MAC address         Auth state
       0002-0000-0011      Authenticated
```

Verify that frames with an unknown destination MAC address, multicast address, or broadcast address are discarded. (Details not shown.)

Example: Configuring port security for M-LAG

Network configuration

As shown in Figure 6, Device A and Device B authenticate the host users to the Internet. Configure Device A and Device B as follows:

- Use RADIUS servers to perform authentication, authorization, and accounting for the users.
- Form an M-LAG system for node redundancy.
- Set the port security mode to **macAddressWithRadius** on their M-LAG interfaces. Set the authentication load sharing mode to distributed local mode for users attached to the M-LAG interfaces.
- To make sure the source IP address of the RADIUS packets sent for users attached to one member device does not change after a node failover, configure the following settings on each member device:
  - Configure one source IP address for local users and one for users attached to the peer member device. On each member device, use the active and standby M-LAG virtual IP addresses as the local and peer source IP addresses of the outgoing RADIUS packets, respectively.

Figure 6 Network diagram

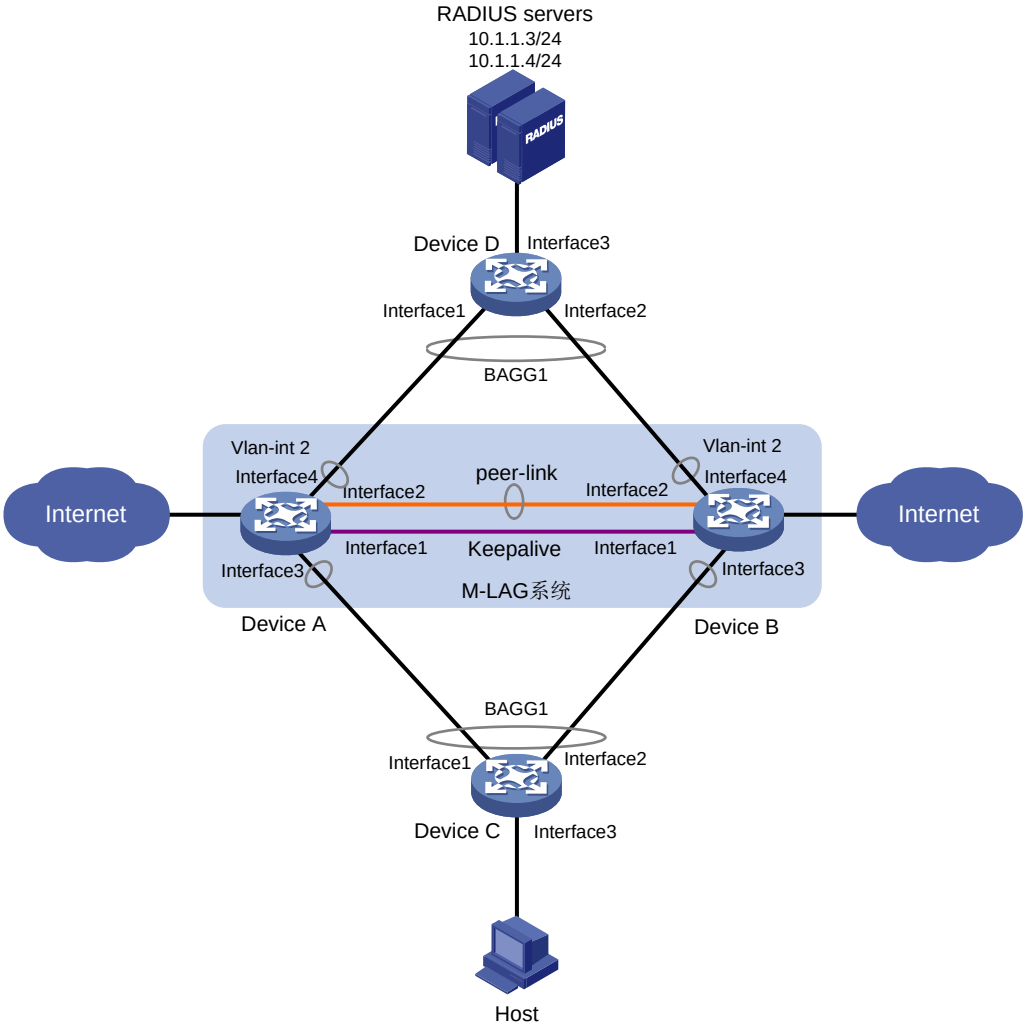


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Table 7 IP assignment scheme

Device	Interface	M-LAG interface	IP address
Device A (M-LAG1)	GE1/0/1	Keepalive interface	1.1.1.2/24
	BAGG1 (GE1/0/2)	Peer link interface	-
	BAGG2 (GE1/0/4)	M-LAG interface	-
	BAGG3 (GE1/0/3)	M-LAG interface	-
	VLAN-interface 2	-	10.1.1.1/24
Device B (M-LAG2)	GE1/0/1	Keepalive interface	1.1.1.1/24
	BAGG1 (GE1/0/2)	Peer link interface	-

Device	Interface	M-LAG interface	IP address
	BAGG2 (GE1/0/4)	M-LAG interface	-
	BAGG3 (GE1/0/3)	M-LAG interface	-
	VLAN-interface 2	-	-

Restrictions and guidelines

The configuration procedure in this example uses some AAA and RADIUS commands. For more information about these commands, see *Security Command Reference*.

The configuration procedure in this example uses some M-LAG commands. For more information about these commands, see *Layer 2—LAN Switching Command Reference*.

Prerequisites

Make sure the host and RADIUS servers can reach each other.

Configure the RADIUS servers and add user accounts. Make sure the RADIUS servers can provide authentication, authorization, and accounting services.

Procedure

1. Configure Device A:

- Configure AAA:

  # Configure RADIUS scheme **radsun**.

```
<DeviceA> system-view
[DeviceA] radius scheme radsun
[DeviceA-radius-radsun] primary authentication 10.1.1.3
[DeviceA-radius-radsun] primary accounting 10.1.1.4
[DeviceA-radius-radsun] secondary authentication 10.1.1.4
[DeviceA-radius-radsun] secondary accounting 10.1.1.3
[DeviceA-radius-radsun] key authentication simple name
[DeviceA-radius-radsun] key accounting simple money
[DeviceA-radius-radsun] timer response-timeout 5
[DeviceA-radius-radsun] retry 5
[DeviceA-radius-radsun] timer realtime-accounting 15
[DeviceA-radius-radsun] user-name-format without-domain
```

  # Specify 10.1.1.1 as the source IP address of outgoing RADIUS packets for users attached to the local M-LAG interfaces. Specify 10.1.1.2 as the source IP address of outgoing RADIUS packets for users attached to the peer M-LAG interfaces.

```
[DeviceA-radius-radsun] nas-ip m-lag local 10.1.1.1
[DeviceA-radius-radsun] nas-ip m-lag peer 10.1.1.2
[DeviceA-radius-radsun] quit
```

  # Configure ISP domain **sun**.

```
[DeviceA] domain sun
[DeviceA-isp-sun] authentication lan-access radius-scheme radsun
[DeviceA-isp-sun] authorization lan-access radius-scheme radsun
[DeviceA-isp-sun] accounting lan-access radius-scheme radsun
[DeviceA-isp-sun] quit
```

- Configure M-LAG settings:

  # Configure M-LAG system settings.

```
[DeviceA] m-lag system-mac 1-1-1
[DeviceA] m-lag system-number 1
```

```
[DeviceA] m-lag system-priority 123
```

Configure the destination and source IP addresses of keepalive packets.

```
[DeviceA] m-lag keepalive ip destination 1.1.1.1 source 1.1.1.2
```

Configure port GigabitEthernet 1/0/1 to operate in Layer 3 mode and configure the source IP address of keepalive packets as the IP address of this port.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-mode route
[DeviceA-GigabitEthernet1/0/1] ip address 1.1.1.2 24
[DeviceA-GigabitEthernet1/0/1] quit
```

Exclude GigabitEthernet 1/0/1 (keepalive interface) from being shut down upon detection of multi-active collisions.

```
[DeviceA] m-lag mad exclude interface gigabitethernet 1/0/1
```

Create Layer 2 aggregate interface **Bridge-Aggregation 1** and set the aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation1] quit
```

Assign GigabitEthernet 1/0/2 to aggregation group 1.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/2] quit
```

Configure Layer 2 aggregate interface **Bridge-Aggregation 1** as a peer link interface.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] port m-lag peer-link 1
[DeviceA-Bridge-Aggregation1] quit
```

On the VLAN interface, specify 10.1.1.1 as the M-LAG virtual IPv4 address in active state and specify 10.1.1.2 as the M-LAG virtual IPv4 address in standby state.

```
[DeviceA] vlan 2
[DeviceA-vlan2] quit
[DeviceA] interface vlan-interface 2
[DeviceA-Vlan-interface2] port m-lag virtual-ip 10.1.1.1 24 active virtual-mac
11-1-1
[DeviceA-Vlan-interface2] port m-lag virtual-ip 10.1.1.2 24 standby virtual-mac
11-1-2
[DeviceA-Vlan-interface2] quit
```

Create Layer 2 aggregate interface **Bridge-Aggregation 2** and set the aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 2
[DeviceA-Bridge-Aggregation2] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation2] quit
```

Assign GigabitEthernet 1/0/4 to aggregation group 2.

```
[DeviceA] interface gigabitethernet 1/0/4
[DeviceA-GigabitEthernet1/0/4] port link-aggregation group 2
[DeviceA-GigabitEthernet1/0/4] quit
```

Assign Bridge-Aggregation 2 to M-LAG group 2.

```
[DeviceA] interface bridge-aggregation 2
[DeviceA-Bridge-Aggregation2] port m-lag group 2
[DeviceA-Bridge-Aggregation2] port access vlan 2
[DeviceA-Bridge-Aggregation2] quit
```

Create Layer 2 aggregate interface **Bridge-Aggregation 3** and set the aggregation mode to dynamic.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/3 to aggregation group 3.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 3
[DeviceA-GigabitEthernet1/0/3] quit
```

Assign Bridge-Aggregation 3 to M-LAG group 3.

```
[DeviceA] interface bridge-aggregation 3
[DeviceA-Bridge-Aggregation3] port m-lag group 3
[DeviceA-Bridge-Aggregation3] quit
```

- Configure port security settings:

  # Enable port security.

  ```
  [DeviceA] port-security enable
  ```

  # Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hyphenated hexadecimal notation, with letters in upper case.

  ```
  [DeviceA] mac-authentication user-name-format mac-address with-hyphen uppercase
  ```

  # Configure domain **sun** as the global MAC authentication domain on the device.

  ```
  [DeviceA] mac-authentication domain sun
  ```

  # Set the port security mode to **macAddressWithRadius** on Bridge-Aggregation 3.

  ```
  [DeviceA] interface bridge-aggregation 3
  [DeviceA-Bridge-Aggregation3] port-security port-mode mac-authentication
  ```

  # Configure domain **sun** as the MAC authentication domain on Bridge-Aggregation 3.

  ```
  [DeviceA-Bridge-Aggregation3] mac-authentication domain sun
  [DeviceA-Bridge-Aggregation3] quit
  ```

2. Configure Device B:

- Configure AAA settings:

  # Configure RADIUS scheme **radsun**.

  ```
  <DeviceB> system-view
  [DeviceB] radius scheme radsun
  [DeviceB-radius-radsun] primary authentication 10.1.1.3
  [DeviceB-radius-radsun] primary accounting 10.1.1.4
  [DeviceB-radius-radsun] secondary authentication 10.1.1.4
  [DeviceB-radius-radsun] secondary accounting 10.1.1.3
  [DeviceB-radius-radsun] key authentication simple name
  [DeviceB-radius-radsun] key accounting simple money
  [DeviceB-radius-radsun] timer response-timeout 5
  [DeviceB-radius-radsun] retry 5
  [DeviceB-radius-radsun] timer realtime-accounting 15
  [DeviceB-radius-radsun] user-name-format without-domain
  ```

  # Specify 10.1.1.2 as the source IP address of outgoing RADIUS packets for users attached to the local M-LAG interfaces. Specify 10.1.1.1 as the source IP address of outgoing RADIUS packets for users attached to the peer M-LAG interfaces.

  ```
  [DeviceB-radius-radsun] nas-ip m-lag local 10.1.1.2
  [DeviceB-radius-radsun] nas-ip m-lag peer 10.1.1.1
  ```

```
[DeviceB-radius-radsun] quit
```

Configure ISP domain **sun**.

```
[DeviceB] domain sun
[DeviceB-isp-sun] authentication lan-access radius-scheme radsun
[DeviceB-isp-sun] authorization lan-access radius-scheme radsun
[DeviceB-isp-sun] accounting lan-access radius-scheme radsun
[DeviceB-isp-sun] quit
```

- Configure M-LAG settings:

Configure M-LAG system settings.

```
[DeviceB] m-lag system-mac 1-1-1
[DeviceB] m-lag system-number 2
[DeviceB] m-lag system-priority 123
```

Configure the destination and source IP addresses of keepalive packets.

```
[DeviceB] m-lag keepalive ip destination 1.1.1.2 source 1.1.1.1
```

Configure GigabitEthernet 1/0/1 to operate in Layer 3 mode and configure the source IP address of keepalive packets as the IP address of this port.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-mode route
[DeviceB-GigabitEthernet1/0/1] ip address 1.1.1.1 24
[DeviceB-GigabitEthernet1/0/1] quit
```

Exclude GigabitEthernet 1/0/1 (keepalive interface) from being shut down upon detection of multi-active collisions.

```
[DeviceB] m-lag mad exclude interface gigabitethernet 1/0/1
```

Create Layer 2 aggregate interface **Bridge-Aggregation 1** and set the aggregation mode to dynamic.

```
[DeviceB] interface bridge-aggregation 1
[DeviceB-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation1] quit
```

Assign GigabitEthernet 1/0/2 to aggregation group 1.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-aggregation group 1
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure Layer 2 aggregate interface **Bridge-Aggregation 1** as a peer link interface.

```
[DeviceB] interface bridge-aggregation 1
[DeviceB-Bridge-Aggregation1] port m-lag peer-link 1
[DeviceB-Bridge-Aggregation1] quit
```

Create VLAN-interface 2. On the VLAN interface, specify 10.1.1.2 as the M-LAG virtual IPv4 address in active state and specify 10.1.1.1 as the M-LAG virtual IPv4 address in standby state.

```
[DeviceB] vlan 2
[DeviceB-vlan2] quit
[DeviceB] interface vlan-interface 2
[DeviceB-Vlan-interface2] port m-lag virtual-ip 10.1.1.2 24 active virtual-mac
11-1-2
[DeviceB-Vlan-interface2] port m-lag virtual-ip 10.1.1.1 24 standby virtual-mac
11-1-1
[DeviceB-Vlan-interface2] quit
```

Create Layer 2 aggregate interface **Bridge-Aggregation 2** and set the aggregation mode to dynamic.

```
[DeviceB] interface bridge-aggregation 2
[DeviceB-Bridge-Aggregation2] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation2] quit
```

Assign GigabitEthernet 1/0/4 to aggregation group 2.

```
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port link-aggregation group 2
[DeviceB-GigabitEthernet1/0/4] quit
```

Assign Bridge-Aggregation 2 to M-LAG group 2.

```
[DeviceB] interface bridge-aggregation 2
[DeviceB-Bridge-Aggregation2] port m-lag group 2
[DeviceB-Bridge-Aggregation2] port access vlan 2
[DeviceB-Bridge-Aggregation2] quit
```

Create Layer 2 aggregate interface **Bridge-Aggregation 3** and set the aggregation mode to dynamic.

```
[DeviceB] interface bridge-aggregation 3
[DeviceB-Bridge-Aggregation3] link-aggregation mode dynamic
[DeviceB-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/3 to aggregation group 3.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-aggregation group 3
[DeviceB-GigabitEthernet1/0/3] quit
```

Assign Bridge-Aggregation 3 to M-LAG group 3.

```
[DeviceB] interface bridge-aggregation 3
[DeviceB-Bridge-Aggregation3] port m-lag group 3
[DeviceB-Bridge-Aggregation3] quit
```

- Configure port security settings:

  # Enable port security.

  ```
  [DeviceB] port-security enable
  ```

  # Use the MAC address of each user as both the username and password for MAC authentication. The MAC addresses are in hyphenated hexadecimal notation, with letters in upper case.

  ```
  [DeviceB] mac-authentication user-name-format mac-address with-hyphen uppercase
  ```

  # Configure domain **sun** as the global MAC authentication domain on the device.

  ```
  [DeviceB] mac-authentication domain sun
  ```

  # Set the port security mode to **macAddressWithRadius** on Layer 2 aggregate interface **Bridge-Aggregation 3**.

  ```
  [DeviceB] interface bridge-aggregation 3
  [DeviceB-Bridge-Aggregation3] port-security port-mode mac-authentication
  ```

  # Configure domain **sun** as the MAC authentication domain on Bridge-Aggregation 3.

  ```
  [DeviceB-Bridge-Aggregation3] mac-authentication domain sun
  [DeviceB-Bridge-Aggregation3] quit
  ```

3. Configure Device C:

Create Layer 2 aggregate interface **Bridge-Aggregation 1** and set the aggregation mode to dynamic.

```
<DeviceC> system-view
[DeviceC] interface bridge-aggregation 1
[DeviceC-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceC-Bridge-Aggregation1] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 1.

```
[DeviceC] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/2
[DeviceC-if-range] port link-aggregation group 1
[DeviceC-if-range] quit
```

4. Configure Device D:

Create Layer 2 aggregate interface **Bridge-Aggregation 1** and set the aggregation mode to dynamic.

```
<DeviceD> system-view
[DeviceD] interface bridge-aggregation 1
[DeviceD-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceD-Bridge-Aggregation1] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to aggregation group 1.

```
[DeviceD] interface range gigabitethernet 1/0/1 to gigabitethernet 1/0/2
[DeviceD-if-range] port link-aggregation group 1
[DeviceD-if-range] quit
```

Verifying the configuration

1. On Device A:

Display M-LAG summary information on Device A to verify that Device A and Device B have formed an M-LAG system.

```
[DeviceA] display m-lag summary
Flags: A -- Aggregate interface down, B -- No peer M-LAG interface configured
       C -- Configuration consistency check failed

Peer-link interface: BAGG1
Peer-link interface state (cause): UP
Keepalive link state (cause): UP

                    M-LAG interface information
M-LAG IF      M-LAG group  Local state (cause)  Peer state  Remaining down time(s)
BAGG2         2            UP                   UP          -
BAGG3         3            UP                   UP          -
```

Display port security settings on Device A.

```
[DeviceA] display port-security interface bridge-aggregation 3
Global port security parameters:
   Port security                          : Enabled
   M-LAG load sharing mode (criterion)    : Distributed (local)
   M-LAG member's authentication scope    : Local M-LAG interfaces
   M-LAG member configuration conflict    : Not conflicted
   AutoLearn aging time                   : 0 min
   Disableport timeout                    : 20 sec
   Blockmac timeout                       : 180 sec
   MAC move                               : Denied
   Authorization fail                     : Online
   NAS-ID profile                         : Not configured
   Dot1x-failure trap                     : Disabled
   Dot1x-logon trap                       : Disabled
   Dot1x-logoff trap                      : Disabled
   Intrusion trap                         : Disabled
```

```
  Intrusion-recover trap                 : Disabled
  Address-learned trap                   : Disabled
  Mac-auth-failure trap                  : Disabled
  Mac-auth-logon trap                    : Disabled
  Mac-auth-logoff trap                   : Disabled
  Mac-auth-not-support trap              : Disabled
  AC-creation-failure trap               : Disabled
  ACL-author-failure trap                : Disabled
  ACL-author-success trap                : Disabled
  URL-author-failure trap                : Disabled
  URL-author-success trap                : Disabled
  NTK-ineffective trap                   : Disabled
  Port-mode-ineffective trap             : Disabled
  Open authentication                    : Disabled
  Traffic-statistics                     : Enabled
  User aging period for preauth domain   : 82800 sec
  User aging period for Auth-Fail domain : 82800 sec
  User aging period for critical domain  : 82800 sec
  Reauth period for preauth domain       : 600 sec
  Reauth period for Auth-Fail domain     : 600 sec
  MAC move for topology change protection : Denied
    Topology change detection period     : 5 sec
    Max detection attempts               : 3
  Record max-number                      : 4096
  Record aging period                    : 1440 min
  OUI value list                         :

 Bridge-Aggregation3 is link-up
  Port mode                              : macAuthentication
  Pre-auth domain                        : Not configured
  URL-unavailable domain                 : Not configured
  NeedToKnow mode                        : Disabled
  Intrusion protection mode              : NoAction
  Max secure MAC addresses               : Not configured
  Authorization                          : Permitted
  NAS-ID profile                         : Not configured
  Free VLANs                             : Not configured
  Open authentication                    : Disabled
  MAC-move VLAN check bypass             : Disabled
  Server-reachable reauth
      Preauth domain                     : Disabled
      Auth-Fail domain                   : Disabled
```

After users pass authentication, display MAC authentication information on Device A.

```
[DeviceA] display mac-authentication interface bridge-aggregation 3
 Global MAC authentication parameters:
  MAC authentication                 : Enabled
  Authentication method              : CHAP
  M-LAG member configuration conflict : Not conflicted
```

```
  Username format                    : MAC address in lowercase(xxxxxxxxxxxx)
          Username                   : mac
          Password                   : Not configured
  MAC range accounts                 : 0
         MAC address          Mask                Username

  Offline detect period                      : 60 s
  Quiet period                               : 60 s
  Server timeout                             : 100 s
  Reauth period                              : 3600 s
  User aging period for critical VLAN        : 1000 s
  User aging period for critical VSI         : 1000 s
  User aging period for guest VLAN           : 1000 s
  User aging period for guest VSI            : 1000 s
  Temporary user aging period                : 60 s
  Authentication domain                      : test
  HTTP proxy port list                       : Not configured
  HTTPS proxy port list                      : Not configured
  Max number of silent MACs                  : 1024 (per slot)
Online MAC-auth wired users                  : 10

Silent MAC users:
          MAC address       VLAN ID  From port               Port index

Bridge-Aggregation3 is link-up
  MAC authentication                         : Enabled
  Carry User-IP                              : Disabled
  Authentication domain                      : test
  Auth-delay timer                           : Disabled
  Periodic reauth                            : Disabled
  Re-auth server-unreachable                 : Logoff
  Guest VLAN                                 : Not configured
  Guest VLAN reauthentication                : Enabled
    Guest VLAN auth-period                   : 30 s
  Critical VLAN                              : Not configured
  Critical voice VLAN                        : Disabled
  Host mode                                  : Single VLAN
  Offline detection                          : Enabled
  Authentication order                       : Default
  User aging                                 : Enabled
  Server-recovery online-user-sync           : Disabled

  Guest VSI                                  : Not configured
  Guest VSI reauthentication                 : Enabled
    Guest VSI auth-period                    : 30 s
  Critical VSI                               : Not configured
  Critical microsegment ID                   : Not configured
  Critical profile                           : Not configured
```

```
   URL user logoff                           : No
   Auto-tag feature                          : Disabled
   VLAN tag configuration ignoring           : Disabled
   Max online users                          : 4294967295
   Max online preauth-domain users           : 4294967295
   Max online Auth-Fail-domain users         : 4294967295
   Auth-server-unavailable escape            : Disabled
   Authentication attempts                   : successful 6, failed 3174
   Current online users                      : 10
          MAC address       Auth state
          0000-0000-0001    Authenticated
          0000-0000-0002    Authenticated
          0000-0000-0003    Authenticated
          0000-0000-0004    Authenticated
          0000-0000-0005    Authenticated
          0000-0000-0006    Authenticated
          0000-0000-0007    Authenticated
          0000-0000-0008    Authenticated
          0000-0000-0009    Authenticated
          0000-0000-000a    Authenticated
```

2. On Device B:

Display M-LAG summary information on Device B to verify that Device A and Device B have formed an M-LAG system.

```
[DeviceB] display m-lag summary
Flags: A -- Aggregate interface down, B -- No peer M-LAG interface configured
       C -- Configuration consistency check failed

Peer-link interface: BAGG1
Peer-link interface state (cause): UP
Keepalive link state (cause): UP
                    M-LAG interface information
M-LAG IF      M-LAG group  Local state (cause)  Peer state  Remaining down time(s)
BAGG2         2            UP                   UP          -
BAGG3         3            UP                   UP          -
```

Display port security settings on Device B.

```
[DeviceB] display port-security interface bridge-aggregation 3
Global port security parameters:
   Port security                          : Enabled
   M-LAG load sharing mode (criterion)    : Distributed (local)
   M-LAG member's authentication scope    : Local M-LAG interfaces
   M-LAG member configuration conflict    : Not conflicted
   AutoLearn aging time                   : 0 min
   Disableport timeout                    : 20 sec
   Blockmac timeout                       : 180 sec
   MAC move                               : Denied
   Authorization fail                     : Online
   NAS-ID profile                         : Not configured
   Dot1x-failure trap                     : Disabled
```

```
  Dot1x-logon trap                        : Disabled
  Dot1x-logoff trap                       : Disabled
  Intrusion trap                          : Disabled
  Intrusion-recover trap                  : Disabled
  Address-learned trap                    : Disabled
  Mac-auth-failure trap                   : Disabled
  Mac-auth-logon trap                     : Disabled
  Mac-auth-logoff trap                    : Disabled
  Mac-auth-not-support trap               : Disabled
  AC-creation-failure trap                : Disabled
  ACL-author-failure trap                 : Disabled
  ACL-author-success trap                 : Disabled
  URL-author-failure trap                 : Disabled
  URL-author-success trap                 : Disabled
  NTK-ineffective trap                    : Disabled
  Port-mode-ineffective trap              : Disabled
  Open authentication                     : Disabled
  Traffic-statistics                      : Enabled
  User aging period for preauth domain    : 82800 sec
  User aging period for Auth-Fail domain  : 82800 sec
  User aging period for critical domain   : 82800 sec
  Reauth period for preauth domain        : 600 sec
  Reauth period for Auth-Fail domain      : 600 sec
  MAC move for topology change protection : Denied
    Topology change detection period      : 5 sec
    Max detection attempts                : 3
  Record max-number                       : 4096
  Record aging period                     : 1440 min
  OUI value list                          :

 Bridge-Aggregation3 is link-up
  Port mode                               : macAuthentication
  Pre-auth domain                         : Not configured
  URL-unavailable domain                  : Not configured
  NeedToKnow mode                         : Disabled
  Intrusion protection mode               : NoAction
  Max secure MAC addresses                : Not configured
  Authorization                           : Permitted
  NAS-ID profile                          : Not configured
  Free VLANs                              : Not configured
  Open authentication                     : Disabled
  MAC-move VLAN check bypass              : Disabled
  Server-reachable reauth
      Preauth domain                      : Disabled
      Auth-Fail domain                    : Disabled
```

After users pass authentication, display MAC authentication information on Device B. Verify that Device B and Device A have the same user authentication information.

```
[DeviceB] display mac-authentication interface bridge-aggregation 3
 Global MAC authentication parameters:
```

```
  MAC authentication                  : Enabled
  Authentication method               : CHAP
  M-LAG member configuration conflict : Not conflicted
  Username format                     : MAC address in lowercase(xxxxxxxxxxxx)
          Username                    : mac
          Password                    : Not configured
  MAC range accounts                  : 0
         MAC address          Mask                Username
  Offline detect period                      : 60 s
  Quiet period                               : 60 s
  Server timeout                             : 100 s
  Reauth period                              : 3600 s
  User aging period for critical VLAN        : 1000 s
  User aging period for critical VSI         : 1000 s
  User aging period for guest VLAN           : 1000 s
  User aging period for guest VSI            : 1000 s
  Temporary user aging period                : 60 s
  Authentication domain                      : test
  HTTP proxy port list                       : Not configured
  HTTPS proxy port list                      : Not configured
  Max number of silent MACs                  : 1024 (per slot)
Online MAC-auth wired users                  : 10

Silent MAC users:
         MAC address       VLAN ID  From port               Port index

Bridge-Aggregation3 is link-up
  MAC authentication                         : Enabled
  Carry User-IP                              : Disabled
  Authentication domain                      : test
  Auth-delay timer                           : Disabled
  Periodic reauth                            : Disabled
  Re-auth server-unreachable                 : Logoff
  Guest VLAN                                 : Not configured
  Guest VLAN reauthentication                : Enabled
    Guest VLAN auth-period                   : 30 s
  Critical VLAN                              : Not configured
  Critical voice VLAN                        : Disabled
  Host mode                                  : Single VLAN
  Offline detection                          : Enabled
  Authentication order                       : Default
  User aging                                 : Enabled
  Server-recovery online-user-sync           : Disabled
  Guest VSI                                  : Not configured
  Guest VSI reauthentication                 : Enabled
    Guest VSI auth-period                    : 30 s
  Critical VSI                               : Not configured
  Critical microsegment ID                   : Not configured
```

```
   Critical profile                          : Not configured
   URL user logoff                           : No
   Auto-tag feature                          : Disabled
   VLAN tag configuration ignoring           : Disabled
   Max online users                          : 4294967295
   Max online preauth-domain users           : 4294967295
   Max online Auth-Fail-domain users         : 4294967295
   Auth-server-unavailable escape            : Disabled
   Authentication attempts                   : successful 20, failed 3099
   Current online users                      : 10
          MAC address       Auth state
          0000-0000-0001    Authenticated
          0000-0000-0002    Authenticated
          0000-0000-0003    Authenticated
          0000-0000-0004    Authenticated
          0000-0000-0005    Authenticated
          0000-0000-0006    Authenticated
          0000-0000-0007    Authenticated
          0000-0000-0008    Authenticated
          0000-0000-0009    Authenticated
          0000-0000-000a    Authenticated
```

Troubleshooting port security

Cannot set the port security mode

Symptom

Cannot set the port security mode for a port.

Analysis

For a port operating in a port security mode other than noRestrictions, you cannot change the port security mode by using the **`port-security port-mode`** command.

Solution

To resolve the issue:

1. Set the port security mode to noRestrictions.

   ```
   [Device-GigabitEthernet1/0/1] undo port-security port-mode
   ```

2. Set a new port security mode for the port, for example, autoLearn.

   ```
   [Device-GigabitEthernet1/0/1] port-security port-mode autolearn
   ```

3. If the issue persists, contact H3C Support.

Cannot configure secure MAC addresses

Symptom

Cannot configure secure MAC addresses.

Analysis

No secure MAC address can be configured on a port operating in a port security mode other than autoLearn.

Solution

To resolve the issue:

1. Set the port security mode to autoLearn.

```
[Device-GigabitEthernet1/0/1] undo port-security port-mode
[Device-GigabitEthernet1/0/1] port-security max-mac-count 64
[Device-GigabitEthernet1/0/1] port-security port-mode autolearn
[Device-GigabitEthernet1/0/1] port-security mac-address security 1-1-2 vlan 1
```

2. If the issue persists, contact H3C Support.

Contents

Configuring user profiles 1
About user profiles 1
Prerequisites for user profile 1
Configuring a user profile 1
Display and maintenance commands for user profiles 1
User profile configuration examples 2
Example: Configuring user profiles and QoS policies 2

Configuring user profiles

About user profiles

A user profile defines a set of parameters, such as a QoS policy, for a user or a class of users. A user profile can be reused when a user connected to the network on a different interface.

The user profile application allows flexible traffic policing on a per-user basis. Each time a user passes authentication, the server sends the device the name of the user profile specified for the user. The device applies the parameters in the user profile to the user.

User profiles are typically used for resource allocation per user. For example, the interface-based traffic policing limits the total amount of bandwidth available to a group of users. However, user-profile-based traffic policing can limit the amount of bandwidth available to a single user.

Prerequisites for user profile

A user profile works with authentication methods. You must configure authentication for a user profile. For information about supported authentication methods, see the configuration guides for the related authentication modules.

Configuring a user profile

1. Enter system view.
   system-view
2. Create a user profile and enter user profile view.
   user-profile *profile-name*
3. Apply an existing QoS policy to the user profile.
   qos apply policy *policy-name* { **inbound** | **outbound** }
   By default, no QoS policy is applied to a user profile.
   For information about QoS policies, see *ACL and QoS Configuration Guide*.

Display and maintenance commands for user profiles

Execute **display** commands in any view.

Task	Command
Display configuration and online user information for the specified user profile or all user profiles.	**display user-profile** [**name** *profile-name*] [**slot** *slot-number*]

User profile configuration examples

Example: Configuring user profiles and QoS policies

Network configuration

As shown in Figure 1, the device performs local 802.1X authentication on the users in domain **user** for authentication and authorization efficiency.

Configure user profiles and QoS policies on the device to meet the following requirements:

- User A cannot access the Internet during 8: 30 and 12:00 every day even if User A passes 802.1X authentication.
- User B has an upload speed of 2 Mbps after passing 802.1X authentication.
- User C has a download speed of 4 Mbps after passing 802.1X authentication.

Figure 1 Network diagram

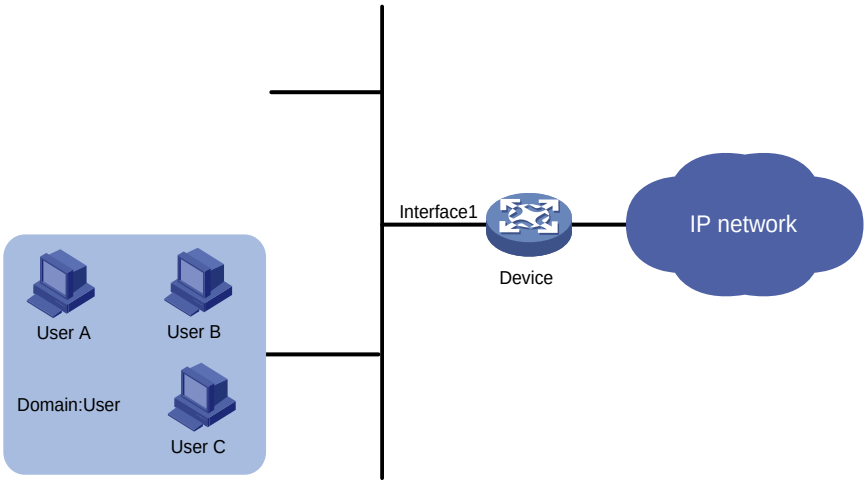


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure a QoS policy for User A:

   # Create a periodic time range from 8:30 to 12:00 every day.

   ```
   <Device> system-view
   [Device] time-range for_usera 8:30 to 12:00 daily
   ```

   # Create IPv4 basic ACL 2000, and configure a rule to match all packets during the time range **for_usera**.

   ```
   [Device] acl basic 2000
   [Device-acl-basic-2000] rule permit time-range for_usera
   [Device-acl-basic-2000] quit
   ```

   # Create a traffic class named **for_usera**, and use ACL 2000 as the match criterion.

   ```
   [Device] traffic classifier for_usera
   [Device-classifier-for_usera] if-match acl 2000
   [Device-classifier-for_usera] quit
   ```

Create a traffic behavior named **for_usera**, and configure the deny action.

```
[Device] traffic behavior for_usera
[Device-behavior-for_usera] filter deny
[Device-behavior-for_usera] quit
```

Create a QoS policy named **for_usera**, and associate traffic class **for_usera** and traffic behavior **for_usera** in the QoS policy.

```
[Device] qos policy for_usera
[Device-qospolicy-for_usera] classifier for_usera behavior for_usera
[Device-qospolicy-for_usera] quit
```

2. Create a user profile for User A and apply the QoS policy to the user profile:

Create a user profile named **usera**.

```
[Device] user-profile usera
```

Apply QoS policy **for_usera** to the inbound direction of user profile **usera**.

```
[[Device-user-profile-usera] qos apply policy for_usera inbound
[Device-user-profile-usera] quit
```

3. Configure a QoS policy for limiting the traffic rate for User B:

Create a traffic class named **class** to match all packets.

```
[Device] traffic classifier class
[Device-classifier-class] if-match any
[Device-classifier-class] quit
```

Create a traffic behavior named **for_userb**, and configure a traffic policing action (CIR 2000 kbps).

```
[Device] traffic behavior for_userb
[Device-behavior-for_userb] car cir 2000
[Device-behavior-for_userb] quit
```

Create a QoS policy named **for_userb**, and associate traffic class **class** and traffic behavior **for_userb** in the QoS policy.

```
[Device] qos policy for_userb
[Device-qospolicy-for_userb] classifier class behavior for_userb
[Device-qospolicy-for_userb] quit
```

4. Create a user profile for User B and apply the QoS policy to the user profile:

Create a user profile named **userb**.

```
[Device] user-profile userb
```

Apply QoS policy **for_userb** to the inbound direction of user profile **userb**.

```
[Device-user-profile-userb] qos apply policy for_userb inbound
[Device-user-profile-userb] quit
```

5. Configure a QoS policy for limiting the traffic rate for User C:

Create a traffic behavior named **for_userc**, and configure a traffic policing action (CIR 4000 kbps).

```
[Device] traffic behavior for_userc
[Device-behavior-for_userc] car cir 4000
[Device-behavior-for_userc] quit
```

Create a QoS policy named **for_userc**, and associate traffic class **class** and traffic behavior **for_userc** in the QoS policy.

```
[Device] qos policy for_userc
[Device-qospolicy-for_userc] classifier class behavior for_userc
[Device-qospolicy-for_userc] quit
```

6. Create a user profile for User C and apply the QoS policy to the user profile:

Create a user profile named **userc**.

```
[Device] user-profile userc
```

Apply QoS policy **for_userc** to the outbound direction of user profile **userc**.

```
[Device-user-profile-userc] qos apply policy for_userc outbound
[Device-user-profile-userc] quit
```

7. Configure local users:

Create a local user named **usera**.

```
[Device] local-user usera class network
New local user added.
```

Set the password to **123456TESTusera&!** for user **usera**.

```
[Device-luser-network-usera] password simple 123456TESTusera&!
```

Authorize user **usera** to use the LAN access service.

```
[Device-luser-network-usera] service-type lan-access
```

Specify user profile **usera** as the authorization user profile for user **usera**.

```
[Device-luser-network-usera] authorization-attribute user-profile usera
[Device-luser-network-usera] quit
```

Create a local user named **userb**.

```
[Device] local-user userb class network
New local user added.
```

Set the password to **123456TESTuserb&!** for user **userb**.

```
[Device-luser-network-userb] password simple 123456TESTuserb&!
```

Authorize user **userb** to use the LAN access service.

```
[Device-luser-network-userb] service-type lan-access
```

Specify user profile **userb** as the authorization user profile for user **userb**.

```
[Device-luser-network-userb] authorization-attribute user-profile userb
[Device-luser-network-userb] quit
```

Create a local user named **userc**.

```
[Device] local-user userc class network
New local user added.
```

Set the password to **123456TESTuserc&!** for user **userc**.

```
[Device-luser-network-userc] password simple 123456TESTuserc&!
```

Authorize user **userc** to use the LAN access service.

```
[Device-luser-network-userc] service-type lan-access
```

Specify user profile **userc** as the authorization user profile for user **userc**.

```
[Device-luser-network-userc] authorization-attribute user-profile userc
[Device-luser-network-userc] quit
```

8. Configure the authentication, authorization, and accounting methods for local users:

```
[Device] domain user
[Device-isp-user] authentication lan-access local
[Device-isp-user] authorization lan-access local
[Device-isp-user] accounting login none
[Device-isp-user] quit
```

9. Configure 802.1X:

Enable 802.1X on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] dot1x
```

Enable MAC-based access control on the port. By default, a port uses MAC-based access control.

```
[Device-GigabitEthernet1/0/1] dot1x port-method macbased
[Device-GigabitEthernet1/0/1] quit
```

Enable 802.1X globally.

```
[Device] dot1x
```

Verifying the configuration

Verify that the three users can pass 802.1X authentication and that QoS policies take effect on these users. (Details not shown.)

Display user profile information.

```
<Device> display user-profile
  User-Profile: usera
    Inbound:
      Policy: for_usera

    slot 1:
      User -:
        Authentication type: 802.1X
        Network attributes:
          Interface    : GigabitEthernet1/0/1
          MAC address  : 6805-ca06-557b
          Service VLAN : 1

  User-Profile: userb
    Inbound:
      Policy: for_userb

    slot 1:
      User -:
        Authentication type: 802.1X
        Network attributes:
          Interface    : GigabitEthernet1/0/1
          MAC address  : 80c1-6ee0-2664
          Service VLAN : 1

  User-Profile: userc
    Outbound:
      Policy: for_userc

    slot 1:
      User -:
        Authentication type: 802.1X
        Network attributes:
          Interface    : GigabitEthernet1/0/1
          MAC address  : 6805-ca05-3efa
          Service VLAN : 1
```

Contents

Configuring password control 1

About password control 1
Password setting 1
Password updating and expiration 2
User login control 3
Password not displayed in any form 5
Logging 5
FIPS compliance 5
Restrictions and guidelines: Password control configuration 5
Password control tasks at a glance 5
Enabling password control 6
Setting global password control parameters 7
Setting user group password control parameters 9
Setting local user password control parameters 10
Setting super password control parameters 11
Enabling SNMP notifications for password control 12
Display and maintenance commands for password control 12
Password control configuration examples 13
Example: Configuring password control 13

Configuring password control

About password control

Password control allows you to implement the following features:

- Manage login and super password setup, expirations, and updates for local users.
- Control user login status based on predefined policies.

For more information about local users, see "Configuring AAA." For information about super passwords, see RBAC in *Fundamentals Configuration Guide*.

Password setting

Minimum password length

You can define the minimum length of user passwords. The system rejects the setting of a password that is shorter than the configured minimum length.

Password composition policy

A password can be a combination of characters from the following types:

- Uppercase letters A to Z.
- Lowercase letters a to z.
- Digits 0 to 9.
- Special characters in Table 1.

Table 1 Special Characters

Character name	Symbol	Character name	Symbol
Ampersand sign	&	Apostrophe	'
Asterisk	*	At sign	@
Back quote	`	Back slash	\
Blank space	N/A	Caret	^
Colon	:	Comma	,
Dollar sign	$	Dot	.
Equal sign	=	Exclamation point	!
Left angle bracket	<	Left brace	{
Left bracket	[	Left parenthesis	(
Minus sign	-	Percent sign	%
Plus sign	+	Pound sign	#
Quotation marks	"	Right angle bracket	>
Right brace	}	Right bracket	]
Right parenthesis	)	Semi-colon	;
Slash	/	Tilde	~

Character name	Symbol	Character name	Symbol
Underscore	_	Vertical bar	|

Depending on the system's security requirements, you can set the minimum number of character types a password must contain and the minimum number of characters for each type, as shown in Table 2.

Table 2 Password composition policy

Password combination level	Minimum number of character types	Minimum number of characters for each type
Level 1	One	One
Level 2	Two	One
Level 3	Three	One
Level 4	Four	One

When a user sets or changes a password, the system checks if the password meets the combination requirement. If it does not, the operation fails.

Password complexity checking policy

A less complicated password is more likely to be cracked, such as a password containing the username or repeated characters. For higher security, you can configure a password complexity checking policy to ensure that all user passwords are relatively complicated. When a user configures a password, the system checks the complexity of the password. If the password is complexity-incompliant, the configuration will fail.

You can apply the following password complexity requirements:

- A password cannot contain the username, the reverse of the username, or part of the username or the reverse of the username. For example, if the username is **abc**, a password such as **abc982**, **2cba**, **ab**, or **cb** is not complex enough.
- A password cannot contain three or more identical consecutive characters. For example, password **a111** is not complex enough.
- A password cannot contain multiple horizontally-adjacent characters on the keyboard. For example, if a password can contain only two horizontally-adjacent characters on the keyboard, **QWE12** and **EWQ21** do not meet the complexity requirements.

Password updating and expiration

Password updating

This feature allows you to set the minimum interval at which users can change their passwords. A user can only change the password once within the specified interval.

The minimum interval does not apply to the following situations:

- A user is prompted to change the password at first login.
- The password aging time expires.

Password expiration

Password expiration imposes a lifecycle on a user password. After the password expires, the user needs to change the password.

The system displays an error message for a login attempt with an expired password. The user is asked to provide a new password. The new password must be valid, and the user must enter exactly the same password when confirming it.

Telnet users, SSH users, and console users can change their own passwords. FTP users must have their passwords changed by the administrator.

Early notice on pending password expiration

When a user logs in, the system checks whether the password will expire in a time equal to or less than the specified notification period. If so, the system notifies the user when the password will expire and provides a choice for the user to change the password.

- If the user sets a new valid password, the system records the new password and the setup time.
- If the user does not or fails to change the password, the system allows the user to log in by using the current password until the password expires.

Telnet users, SSH users, and console users can change their own passwords. FTP users must have their passwords changed by the administrator.

Login with an expired password

You can allow a user to log in a certain number of times within a period of time after the password expires. For example, if you set the maximum number of logins with an expired password to 3 and the time period to 15 days, a user can log in three times within 15 days after the password expires.

Password history

This feature allows the system to store passwords that a user has used.

When a network access user changes the password, the system compares the new password with the current password and those stored in the password history records. The new password must be different from the current one and those stored in the history records by a minimum of four different characters. Otherwise, the system will display an error message, and the new password is not successfully set.

The local passwords and super passwords for device management users are stored in hash form and cannot be converted to plain texts. When a device management user changes a local password or super password, follow these rules:

- If the new password is set by using the hash method, the system will not compare the new password with the current one and those stored in the history password records.
- If the new password is set in plain text, the system compares the new password with the current password and those stored in the password history records. A new password must be different from those stored in the history password records. If the current password is required, the new password must also be different from the current one by a minimum of four different characters. Otherwise, the system will display an error message, and the new password is not successfully set.

You can set the maximum number of history password records for the system to maintain for each user. When the number of history password records exceeds the setting, the most recent record overwrites the earliest one.

Current login passwords are not stored in the password history for device management users. Device management users have their passwords saved in cipher text, which cannot be recovered to plaintext passwords.

User login control

First login

By default, if the global password control feature is enabled, users must change the password at first login before they can access the system. In this situation, password changes are not subject to the minimum password update interval. If it is not necessary for users to change the password at first login, disable the password change at first login feature.

Password control blacklist

The password control blacklist prevents abnormal users from logging in by recording the login failures and maintaining the status of blacklisted user accounts.

If an FTP or VTY user fails to log in, the system adds the user account and the user's IP address to the password control blacklist. If a console user fails to log in, the system adds the user account to the password control blacklist.

If a user uses the same user account to log in to the device from different IP addresses and fails the logins, the device will create a blacklist entry for each IP address for the user account. When the maximum number of blacklist entries for the user account is reached, a blacklist entry for a new IP address of the user account will overwrite the earliest blacklist entry for the user account.

Login attempt limit

Limiting the number of consecutive login failures can effectively prevent password guessing.

When the number of consecutive failed attempts reaches the maximum limit, login attempt limit limits the user in any of the following ways:

- For an FTP or VTY user, the system prohibits the user from using the user account to log in through the user's IP address. For a console user, the system prohibits the user from logging in through console user line. The locked user can use their own account to log in to the device only after the account is manually removed from the password control blacklist.
- Allows the user to continue using the user account. The user account is removed from the password control blacklist when the user uses this account to successfully log in to the device.
- Locks the user account and the user's IP address for a period of time.

  The user can use the account to log in when either of the following conditions exists:

  - The locking timer expires.
  - The account is manually removed from the password control blacklist before the locking timer expires.

NOTE:

This account is locked only for the user at the locked IP address. A user from an unlocked IP address can still use this account, and the user at the locked IP address can use other unlocked user accounts.

Maximum account idle time

You can set the maximum account idle time for user accounts. When an account is idle for this period of time since the last successful login, the account becomes invalid.

Login control with a weak password

The system checks for weak passwords for Telnet, SSH, HTTP, or HTTPS device management users. A password is weak if it does not meet the following requirements:

- Password composition restriction.
- Minimum password length restriction.
- Password complexity username checking policy.

By default, the system displays a message about a weak password but does not force the user to change it. To improve the device security, you can enable the mandatory weak password change feature, which forces the users to change the identified weak passwords. The users can log in to the device only after their passwords meet the password requirements.

Password not displayed in any form

For security purposes, nothing is displayed when a user enters a password.

Logging

The system generates a log each time a user changes its password successfully or is added to the password control blacklist because of login failures.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see "Configuring FIPS") and non-FIPS mode.

Restrictions and guidelines: Password control configuration

IMPORTANT:

To successfully enable the global password control feature and allow device management users to log in to the device, make sure the device have sufficient storage space.

The password control features can be configured in several different views, and different views support different features. The settings configured in different views or for different objects have the following application ranges:

- Settings for super passwords apply only to super passwords.
- Settings in local user view apply only to the password of the local user.
- Settings in user group view apply to the passwords of the local users in the user group if you do not configure password policies for these users in local user view.
- Global settings in system view apply to the passwords of the local users in all user groups if you do not configure password policies for these users in both local user view and user group view.

For local user passwords, the settings with a smaller application scope have higher priority. The system prefers to use the non-default setting in local user view for a local user.

- If the default setting is configured for the local user, the system uses the non-default setting for the user group to which the local user belongs.
- If the default setting is configured for the user group, the system uses the global setting.

Password control tasks at a glance

To configure password control, perform the following tasks:

1. Enabling password control
2. (Optional.) Setting global password control parameters
3. (Optional.) Setting user group password control parameters
4. (Optional.) Setting local user password control parameters
5. (Optional.) Setting super password control parameters
6. (Optional.) Enabling SNMP notifications for password control

Enabling password control

About this task

For device management users, enabling global password control is a prerequisite for password control functions to take effect except for password composition policy, minimum password length, and username checking. For a specific password restriction setting (password expiration or password history management) to take effect, make sure you enable both the global password control feature and the specific password restriction.

For network access users, all password control functions take effect only after the global password control feature is enabled.

Restrictions and guidelines

After global password control is enabled, follow these restrictions and guidelines:

- You cannot display the password and super password configurations for device management users by using the corresponding **display** commands.
- You cannot display the password configuration for network access users by using the corresponding **display** command.
- The passwords configured for local users must contain a minimum of four different characters.
- In FIPS mode, the global password control feature is enabled for device management users and cannot be disabled for them.
- To ensure correct function of password control, configure the device to use NTP to obtain the UTC time. After global password control is enabled, password control will record the UTC time when the password is set. The recorded UTC time might not be consistent with the actual UTC time due to power failure or device reboot. The inconsistency will cause the password expiration feature to malfunction. For information about NTP, see *Network Management and Monitoring Configuration Guide*.
- The device automatically generates a .dat file and saves the file to the storage media. The file is used to record authentication and login information of the local users. Do not manually delete or modify the file.
- The global password control feature enables the system to record history passwords. When the number of history password records of a user reaches the maximum number, the newest history record overwrites the earliest one. To delete the existing history password records, use one of the following methods:
  - Use the **undo password-control enable** command to disable the password control feature globally.
  - Use the **reset password-control history-record** command to clear the passwords manually.

Procedure

1. Enter system view.

   system-view

2. Enable the global password control feature.

   In non-FIPS mode:

   password-control enable [**network-class**]

   By default, the global password control feature is disabled for device management and network access users.

   In FIPS mode:

   password-control enable [**network-class**]

By default, the global password control feature is enabled for device management users and cannot be disabled. The global password control feature is disabled for network access users.

3. (Optional.) Enable a password restriction feature.

   password-control { **aging** | **composition** | **history** | **length** } **enable**

   By default, all four password restriction features are enabled.

Setting global password control parameters

Restrictions and guidelines

The global password control parameters in system view apply to all device management and network access local users.

You can configure all password control features for device management users. The password aging time, minimum password length, password complexity policy, password composition policy, and user login attempt limit can be configured in system view, user group view, and local user view.

You can configure only the following password control features for network access users:

- Minimum password length.
- Password complexity policy.
- Password composition policy.
- Minimum password update interval.
- Maximum number of history password records for each user.

Where, the minimum password length, password complexity policy, and password composition policy can be configured in system view, user group view, and local user view.

The password settings with a smaller application scope have higher priority. For local users, password settings configured in local user view have the highest priority, and global settings in system view have the lowest priority.

The **password-control login-attempt** command takes effect immediately and can affect the users already in the password control blacklist. Other password control configurations do not take effect on users that have been logged in or passwords that have been configured.

Procedure

1. Enter system view.

   system-view

2. Configure password settings.
   - Set the minimum password length.

     In non-FIPS mode:

     password-control length *length*

     The default setting is 10 characters.

     In FIPS mode:

     password-control length *length*

     The default length is 15 characters.
   - Configure the password composition policy.

     In non-FIPS mode:

     password-control composition type-number *type-number* [**type-length** *type-length*]

     By default, a password must contain a minimum of two character types and a minimum of one character for each type.

In FIPS mode:

password-control composition type-number *type-number* [**type-length** *type-length*]

By default, a password must contain a minimum of four character types and a minimum of one character for each type.

- Configure the password complexity checking policy.

  password-control complexity { **adjacent-character** | **same-character** | **user-name** } **check**

  In non-FIPS mode:

  By default, username checking is enabled but repeated character checking and horizontally-adjacent character checking are disabled.

  In FIPS mode:

  By default, the system does not perform password complexity checking.

- Specify the maximum number of horizontally-adjacent characters on the keyboard that can be included in a password.

  password-control adjacent-character max-number *number*

  By default, a maximum of three horizontally-adjacent characters on the keyboard can be included in a password.

- Set the maximum number of history password records for each user.

  password-control history *max-record-number*

  The default setting is 4.

3. Configure password updating and expiration.

- Set the minimum password update interval.

  password-control update interval *interval*

  The default setting is 24 hours.

- Set the password aging time.

  password-control aging *aging-time*

  The default setting is 90 days.

- Set the number of days during which a user is notified of the pending password expiration.

  password-control alert-before-expire *alert-time*

  The default setting is 7 days.

- Set the maximum number of days and maximum number of times that a user can log in after the password expires.

  password-control expired-user-login delay *delay* **times** *times*

  By default, a user can log in three times within 30 days after the password expires.

4. Configure user login control.

- Enable the password control blacklist feature for all user line types.

  password-control blacklist all-line

  By default, the password control blacklist feature is disabled for all user line types. The password control blacklist feature is enabled only for FTP and VTY users when you enable the global password control feature.

- Add only usernames to the password control blacklist for users failing authentication.

  password-control blacklist user-info username-only

  By default, both usernames and IP addresses are added to the password control blacklist when the users fail authentication.

- Set the maximum number of blacklist entries for a user account.

password-control per-user blacklist-limit *max-number*

By default, the maximum number of blacklist entries for a user account is 32.

- Configure the login attempt limit.

  password-control login-attempt *login-times* [**exceed** { **lock** | **lock-time** *time* | **unlock** }]

  By default, the maximum number of login attempts is 3 and a user failing to log in after the specified number of attempts must wait for 1 minute before trying again.

- Set the maximum account idle time.

  password-control login idle-time *idle-time*

  The default setting is 90 days.

  If a user account is idle for this period of time, the account becomes invalid and can no longer be used to log in to the device. To disable the account idle time restriction, set the idle time value to 0.

- Set the user authentication timeout time.

  password-control authentication-timeout *timeout*

  The default setting is 600 seconds.

  This command takes effect only on Telnet and terminal users.

- Disable password change at first login.

  undo password-control change-password first-login enable

  By default, the password change at first login feature is enabled.

  In FIPS mode, the password change at first login feature cannot be disabled.

- Enable mandatory weak password change.

  password-control change-password weak-password enable

  In non-FIPS mode:

  By default, the mandatory weak password change feature is disabled.

  In FIPS mode:

  By default, the mandatory weak password change feature is enabled, and cannot be disabled.

Setting user group password control parameters

1. Enter system view.

   system-view

2. Create a user group and enter its view.

   user-group *group-name*

   For information about how to configure a user group, see "Configuring AAA."

3. Configure the password aging time for the user group.

   password-control aging *aging-time*

   By default, the password aging time of the user group equals the global password aging time.

4. Configure the minimum password length for the user group.

   password-control length *length*

   By default, the minimum password length of the user group equals the global minimum password length.

5. Configure the password composition policy for the user group.

password-control composition type-number *type-number* [**type-length** *type-length*]

By default, the password composition policy of the user group equals the global password composition policy.

6. Configure the password complexity checking policy for the user group.

   password-control complexity { **adjacent-character** | **same-character** | **user-name** } **check**

   By default, the password complexity checking policy of the user group equals the global password complexity checking policy.

7. Configure the login attempt limit.

   password-control login-attempt *login-times* [**exceed** { **lock** | **lock-time** *time* | **unlock** }]

   By default, the login-attempt policy of the user group equals the global login-attempt policy.

Setting local user password control parameters

1. Enter system view.

   system-view

2. Create a device management or network access user and enter its view.
   - Create a device management user and enter its view.

     local-user *user-name* **class manage**
   - Create a network access user and enter its view.

     local-user *user-name* **class network**

   By default, no local user exists.

   For information about local user configuration, see "Configuring AAA."

3. Configure the password aging time for the local user.

   password-control aging *aging-time*

   By default, the setting equals that for the user group to which the local user belongs. If no aging time is configured for the user group, the global setting applies to the local user.

   This command is available only for device management users.

4. Configure the minimum password length for the local user.

   password-control length *length*

   By default, the setting equals that for the user group to which the local user belongs. If no minimum password length is configured for the user group, the global setting applies to the local user.

5. Configure the password composition policy for the local user.

   password-control composition type-number *type-number* [**type-length** *type-length*]

   By default, the settings equal those for the user group to which the local user belongs. If no password composition policy is configured for the user group, the global settings apply to the local user.

6. Configure the password complexity checking policy for the local user.

   password-control complexity { **adjacent-character** | **same-character** | **user-name** } **check**

   By default, the settings equal those for the user group to which the local user belongs. If no password complexity checking policy is configured for the user group, the global settings apply to the local user.

7. Configure the login attempt limit.

   password-control login-attempt *login-times* [**exceed** { **lock** | **lock-time** *time* | **unlock** }]

   By default, the settings equal those for the user group to which the local user belongs. If no login-attempt policy is configured for the user group, the global settings apply to the local user.

   This command is available only for device management users.

Setting super password control parameters

1. Enter system view.

   system-view

2. Set the password aging time for super passwords.

   password-control super aging *aging-time*

   The default setting is 90 days.

3. Configure the minimum length for super passwords.

   In non-FIPS mode:

   password-control super length *length*

   The default setting is 10 characters.

   In FIPS mode:

   password-control super length *length*

   The default setting is 15 characters.

4. Configure the password composition policy for super passwords.

   In non-FIPS mode:

   password-control super composition type-number *type-number* [**type-length** *type-length*]

   By default, a super password must contain a minimum of two character types and a minimum of one character for each type.

   In FIPS mode:

   password-control super composition type-number *type-number* [**type-length** *type-length*]

   By default, a super password must contain a minimum of four character types and a minimum of one character for each type.

5. Configure the super password complexity checking policy.

   password-control super complexity { **adjacent-character** | **same-character** } **check**

   By default, the system does not perform password complexity checking on super passwords. That is, a super password can contain consecutive identical characters and horizontally-adjacent characters on the keyboard.

6. Specify the maximum number of horizontally-adjacent characters on the keyboard that can be included in a super password.

   password-control super adjacent-character max-number *number*

   By default, a maximum of three horizontally-adjacent characters on the keyboard can be included in a super password.

Enabling SNMP notifications for password control

About this task

Password control supports the following SNMP notifications:

- **convert-format-failed**: Sends SNMP notifications when converting memory data to file storage fails before writing user password records to the password record file.
- **copy-file-failed**: Sends SNMP notifications when the password records failed to be copied to the record file on the subordinate device.
- **notify-write-file-failed**: Sends SNMP notifications when a request message to write user passwords to the password record file failed to be issued.
- **open-file-failed**: Sends SNMP notifications when a .dat file failed to be created or opened due to a file system error.
- **space-not-enough**: Sends SNMP notifications when the space of the storage medium (flash or CF card) where a .dat file is saved is insufficient.
- **write-password-failed**: Sends SNMP notifications when user passwords failed to be written to the password record file.

If you enable SNMP notifications for the password control module, the module will generate notifications to report key events on the module. The generated notifications will be sent to the SNMP module on the device. You can configure SNMP notification settings as needed.

For more information about notifications, see SNMP configuration in Network Management and Monitoring Configuration Guide.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for password control.

   snmp-agent trap enable password-control [**convert-format-failed** | **copy-file-failed** | **notify-write-file-failed** | **open-file-failed** | **space-not-enough** | **write-password-failed**] *

   By default, SNMP notifications are disabled for password control.

Display and maintenance commands for password control

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display password control configuration.	**display password-control** [**super**]
Display information about users in the password control blacklist.	**display password-control blacklist** [**user-name** *user-name* | **ip** *ipv4-address* | **ipv6** *ipv6-address*]
Delete users from the password control blacklist.	**reset password-control blacklist** [**user-name** *user-name*]
Clear history password records.	**reset password-control history-record** [**user-name** *user-name* | **super** [**role**

	role-name] | **network-class** [**user-name** *user-name*]]

Password control configuration examples

Example: Configuring password control

Network configuration

Configure a global password control policy to meet the following requirements:

- A password must contain a minimum of 16 characters.
- A password must contain a minimum of four character types and a minimum of four characters for each type.
- An FTP or VTY user failing to provide the correct password in two successive login attempts is permanently prohibited from logging in with the current IP address.
- A user can log in five times within 60 days after the password expires.
- A password expires after 30 days.
- The minimum password update interval is 36 hours.
- The maximum account idle time is 30 days.
- A password cannot contain the username or the reverse of the username, or part of the username or the reverse of the username.
- A password cannot contain three or more identical consecutive characters.
- A password cannot contain more than four horizontally-adjacent characters on the keyboard.

Configure a super password control policy for user role **network-operator** to meet the following requirements:

- A super password must contain a minimum of 24 characters.
- A super password must contain a minimum of four character types and a minimum of five characters for each type.
- A super password cannot contain more than four horizontally-adjacent characters on the keyboard.

Configure a password control policy for local Telnet user **test** to meet the following requirements:

- The password must contain a minimum of 24 characters.
- The password must contain a minimum of four character types and a minimum of five characters for each type.
- The password for the local user expires after 20 days.

Procedure

Enable the password control feature globally.

```
<Sysname> system-view
[Sysname] password-control enable
```

Allow a maximum of two consecutive login failures on a user account, and lock the user account and the user's IP address permanently if the limit is reached.

```
[Sysname] password-control login-attempt 2 exceed lock
```

Set all passwords to expire after 30 days.

```
[Sysname] password-control aging 30
```

Globally set the minimum password length to 16 characters.

```
[Sysname] password-control length 16
```

Set the minimum password update interval to 36 hours.

```
[Sysname] password-control update-interval 36
```

Specify that a user can log in five times within 60 days after the password expires.

```
[Sysname] password-control expired-user-login delay 60 times 5
```

Set the maximum account idle time to 30 days.

```
[Sysname] password-control login idle-time 30
```

Enable username checking.

```
[Sysname] password-control complexity user-name check
```

Enable repeated character checking.

```
[Sysname] password-control complexity same-character check
```

Enable horizontally-adjacent character checking in the password complexity checking policy.

```
[Sysname] password-control complexity adjacent-character check
```

Refuse a password that contains more than four horizontally-adjacent characters on the keyboard.

```
[Sysname] password-control adjacent-character max-number 4
```

Globally specify that all passwords must each contain a minimum of four character types and a minimum of four characters for each type.

```
[Sysname] password-control composition type-number 4 type-length 4
```

Set the minimum super password length to 24 characters.

```
[Sysname] password-control super length 24
```

Specify that a super password must contain a minimum of four character types and a minimum of five characters for each type.

```
[Sysname] password-control super composition type-number 4 type-length 5
```

Enable horizontally-adjacent character checking in the super password complexity checking policy.

```
[Sysname] password-control super complexity adjacent-character check
```

Refuse a super password that contains more than four horizontally-adjacent characters on the keyboard.

```
[Sysname] password-control super adjacent-character max-number 4
```

Configure a super password used for switching to user role **network-operator** as **123456789ABGFTweuix@#$%!** in plain text.

```
[Sysname] super password role network-operator simple 123456789ABGFTweuix@#$%!
```

Create a device management user named **test**.

```
[Sysname] local-user test class manage
```

Set the service type of the user to **Telnet**.

```
[Sysname-luser-manage-test] service-type telnet
```

Set the minimum password length to 24 for the local user.

```
[Sysname-luser-manage-test] password-control length 24
```

Specify that the password of the local user must contain a minimum of four character types and a minimum of five characters for each type.

```
[Sysname-luser-manage-test] password-control composition type-number 4 type-length 5
```

Set the password for the local user to expire after 20 days.

```
[Sysname-luser-manage-test] password-control aging 20
```

Configure the password of the local user in interactive mode.

```
[Sysname-luser-manage-test] password
```

```
Password:
Confirm :
Updating user information. Please wait ... ...
[Sysname-luser-manage-test] quit
```

Verifying the configuration

Display the global password control configuration.

```
<Sysname> display password-control
 Global password control configurations:
 Password control:                    Enabled(device management users)
                                      Disabled (network access users)
 Password aging:                      Enabled (30 days)
 Password length:                     Enabled (16 characters)
 Password composition:                Enabled (4 types, 4 characters per type)
 Password history:                    Enabled (max history record:4)
 Early notice on password expiration:  7 days
 User authentication timeout:         600 seconds
 Maximum login attempts:              2
 Action for exceeding login attempts:  Lock
 Minimum interval between two updates: 36 hours
 User account idle time:              30 days
 Logins with aged password:           5 times in 60 days
 Password complexity:                 Enabled (username checking)
                                      Enabled (repeated characters checking)
 Password change:                     Enabled (first login)
                                      Disabled (mandatory weak password change)
                                      Enabled (max adjacent characters:4)
 All line:                            Disabled (all line blacklist)
```

Display the password control configuration for super passwords.

```
<Sysname> display password-control super
 Super password control configurations:
 Password aging:                      Enabled (90 days)
 Password length:                     Enabled (24 characters)
 Password composition:                Enabled (4 types, 5 characters per type)
Password complexity:                 Disabled (repeated characters checking)
                                      Enabled (max adjacent characters:4)
```

Display the password control configuration for local user **test**.

```
<Sysname> display local-user user-name test class manage
Total 1 local users matched.

Device management user test:
  State:                     Active
  Service type:              Telnet
  User group:                system
  Bind attributes:
  Authorization attributes:
    Work directory:          flash:
```

```
  User role list:          network-operator
Password control configurations:
  Password aging:          20 days
  Password length:         24 characters
  Password composition:    4 types, 5 characters per type
  Password complexity:     username checking
```

Contents

Configuring keychains 1

About keychains 1
Restrictions and guidelines: Keychain configuration 1
Configuring a keychain 1
Display and maintenance commands for keychain 2
Keychain configuration example 2
Example: Configuring keychains 2

Configuring keychains

About keychains

A keychain, a sequence of keys, provides dynamic authentication to ensure secure communication by periodically changing the key and authentication algorithm without service interruption.

A keychain operates in absolute time mode. In this mode, each time point during a key's lifetime is the UTC time and is not affected by the system's time zone or daylight saving time.

Each key in a keychain has a key string, authentication algorithm, sending lifetime, and receiving lifetime. When the system time is within the lifetime of a key in a keychain, an application uses the key to authenticate incoming and outgoing packets. The keys in the keychain take effect one by one according to the sequence of the configured lifetimes. In this way, the authentication algorithms and keys are dynamically changed to implement dynamic authentication.

Restrictions and guidelines: Keychain configuration

To make sure only one key in a keychain is used at a time to authenticate packets to a peer, set non-overlapping sending lifetimes for the keys in the keychain.

The keys used by the local device and the peer device must have the same authentication algorithm and key string.

Configuring a keychain

1. Enter system view.

   system-view

2. Create a keychain and enter keychain view.

   keychain *keychain-name* **mode absolute**

3. (Optional.) Configure TCP authentication.
   - Set the kind value in the TCP Enhanced Authentication Option.

     tcp-kind *kind-value*

     By default, the kind value is 254.
   - Set an algorithm ID for a TCP authentication algorithm.

     tcp-algorithm-id { **hmac-md5** | **hmac-sha-256** | **hamc-sm3** | **md5** | **sm3** } *algorithm-id*

     By default, the algorithm ID is 3 for the MD5 authentication algorithm, 51 for the SM3 authentication algorithm, 52 for the HMAC-SM3 authentication algorithm, 7 for the HMAC-SHA-256 authentication algorithm, and 5 for the HMAC-MD5 authentication algorithm.

   When the local device uses TCP to communicate with a peer device from another vendor, make sure both devices have the same kind value and algorithm ID settings. If they do not, modify the settings on the local device.

4. (Optional.) Set a tolerance time for accept keys in the keychain.

   accept-tolerance { *value* | **infinite** }

   By default, no tolerance time is configured for accept keys in a keychain.

If authentication information is changed, information mismatch occurs on the local and peer devices, and the service might be interrupted. Use this command to ensure continuous packet authentication.

5. Create a key and enter key view.

   key *key-id*

6. Configure the key.

   - Specify an authentication algorithm for the key.

     authentication-algorithm { **aes-128-cmac** | **aes-256-cmac** | **hmac-md5** | **hmac-sha-256** | **hmac-sm3** | **md5** | **sm3** }

     By default, no authentication algorithm is specified for a key.

   - Configure a key string for the key.

     key-string { **cipher** | **plain** } *string*

     By default, no key string is configured.

   - Set the sending lifetime in UTC mode for the key.

     send-lifetime utc *start-time start-date* { **duration** { *duration-value* | **infinite** } | **to** *end-time end-date* }

     By default, the sending lifetime is not configured for a key.

   - Set the receiving lifetime in UTC mode for the key.

     accept-lifetime utc *start-time start-date* { **duration** { *duration-value* | **infinite** } | **to** *end-time end-date* }

     By default, the receiving lifetime is not configured for a key.

   - (Optional.) Specify the key as the default send key.

     default-send-key

     By default, no default send key is configured for a keychain.

     You can specify only one key as the default send key in a keychain.

Display and maintenance commands for keychain

Execute **display** commands in any view.

Task	Command
Display keychain information.	**display keychain** [**name** *keychain-name* [**key** *key-id*]]

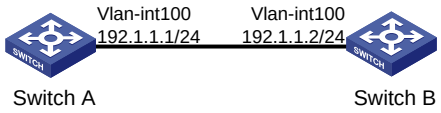
Keychain configuration example

Example: Configuring keychains

Network configuration

As shown in Figure 1, establish an OSPF neighbor relationship between Switch A and Switch B, and use a keychain to authenticate packets between the switches. Configure key 1 and key 2 for the keychain and make sure key 2 is used immediately when key 1 expires.

Figure 1 Network diagram



Procedure

1. Configure Switch A:

Configure IP addresses for interfaces. (Details not shown.)

Configure OSPF.

```
<SwitchA> system-view
[SwitchA] ospf 1 router-id 1.1.1.1
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Create a keychain named **abc**, and specify the absolute time mode for it.

```
[SwitchA] keychain abc mode absolute
```

Create key **1** for keychain **abc**, specify an authentication algorithm, and configure a key string and the sending and receiving lifetimes for the key.

```
[SwitchA-keychain-abc] key 1
[SwitchA-keychain-abc-key-1] authentication-algorithm md5
[SwitchA-keychain-abc-key-1] key-string plain 123456
[SwitchA-keychain-abc-key-1] send-lifetime utc 10:00:00 2015/02/06 to 11:00:00
2015/02/06
[SwitchA-keychain-abc-key-1] accept-lifetime utc 10:00:00 2015/02/06 to 11:00:00
2015/02/06
[SwitchA-keychain-abc-key-1] quit
```

Create key **2** for keychain **abc**, specify an authentication algorithm, and configure a key string and the sending and receiving lifetimes for the key.

```
[SwitchA-keychain-abc] key 2
[SwitchA-keychain-abc-key-2] authentication-algorithm hmac-md5
[SwitchA-keychain-abc-key-2] key-string plain pwd123
[SwitchA-keychain-abc-key-2] send-lifetime utc 11:00:00 2015/02/06 to 12:00:00
2015/02/06
[SwitchA-keychain-abc-key-2] accept-lifetime utc 11:00:00 2015/02/06 to 12:00:00
2015/02/06
[SwitchA-keychain-abc-key-2] quit
[SwitchA-keychain-abc] quit
```

Configure VLAN-interface 100 to use keychain **abc** for authentication.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ospf authentication-mode keychain abc
[SwitchA-Vlan-interface100] quit
```

2. Configure Switch B:

Configure IP addresses for interfaces. (Details not shown.)

Configure OSPF.

```
[SwitchB] ospf 1 router-id 2.2.2.2
[SwitchB-ospf-1] area 0
```

```
[SwitchB-ospf-1-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Create a keychain named **abc**, and specify the absolute time mode for it.

```
[SwitchB] keychain abc mode absolute
```

Create key **1** for keychain **abc**, specify an authentication algorithm, and configure a key string and the sending and receiving lifetimes for the key.

```
[SwitchB-keychain-abc] key 1
[SwitchB-keychain-abc-key-1] authentication-algorithm md5
[SwitchB-keychain-abc-key-1] key-string plain 123456
[SwitchB-keychain-abc-key-1] send-lifetime utc 10:00:00 2015/02/06 to 11:00:00
2015/02/06
[SwitchB-keychain-abc-key-1] accept-lifetime utc 10:00:00 2015/02/06 to 11:00:00
2015/02/06
[SwitchB-keychain-abc-key-1] quit
```

Create key **2** for keychain **abc**, specify an authentication algorithm, and configure a key string and the sending and receiving lifetimes for the key.

```
[SwitchB-keychain-abc] key 2
[SwitchB-keychain-abc-key-2] authentication-algorithm hmac-md5
[SwitchB-keychain-abc-key-2] key-string plain pwd123
[SwitchB-keychain-abc-key-2] send-lifetime utc 11:00:00 2015/02/06 to 12:00:00
2015/02/06
[SwitchB-keychain-abc-key-2] accept-lifetime utc 11:00:00 2015/02/06 to 12:00:00
2015/02/06
[SwitchB-keychain-abc-key-2] quit
[SwitchB-keychain-abc] quit
```

Configure VLAN-interface 100 to use keychain **abc** for authentication.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ospf authentication-mode keychain abc
[SwitchB-Vlan-interface100] quit
```

Verifying the configuration

1. When the system time is within the lifetime from 10:00:00 to 11:00:00 on the day 2015/02/06, verify the status of the keys in keychain **abc**.

   # Display keychain information on Switch A. The output shows that key 1 is the valid key.

```
[SwitchA] display keychain

 Keychain name          : abc
   Mode                 : absolute
   Accept tolerance     : 0
   TCP kind value       : 254
   TCP algorithm value
     HMAC-MD5           : 5
     HMAC-SHA-256       : 7
     MD5                : 3
     HMAC-SM3           : 52
     SM3                : 51
   Default send key ID  : None
   Active send key ID   : 1
```

```
  Active accept key IDs: 1

  Key ID              : 1
    Key string        : $c$3$dYTC8QeOKJkwFwP2k/rWL+1p6uMTw3MqNg==
    Algorithm         : md5
    Send lifetime     : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
    Send status       : Active
    Accept lifetime   : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
    Accept status     : Active

  Key ID              : 2
    Key string        : $c$3$7TSPbUxoP1ytOqkdcJ3K3x0BnXEWl4mOEw==
    Algorithm         : hmac-md5
    Send lifetime     : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
    Send status       : Inactive
    Accept lifetime   : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
    Accept status     : Inactive
```

Display keychain information on Switch B. The output shows that key 1 is the valid key.

```
[SwitchB] display keychain

 Keychain name          : abc
   Mode                 : absolute
   Accept tolerance     : 0
   TCP kind value       : 254
   TCP algorithm value
     HMAC-MD5           : 5
     HMAC-SHA-256       : 7
     MD5                : 3
     HMAC-SM3           : 52
     SM3                : 51
   Default send key ID  : None
   Active send key ID   : 1
   Active accept key IDs: 1

   Key ID               : 1
     Key string         : $c$3$/G/Shnh6heXWprlSQy/XDmftHa2JZJBSgg==
     Algorithm          : md5
     Send lifetime      : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
     Send status        : Active
     Accept lifetime    : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
     Accept status      : Active

   Key ID               : 2
     Key string         : $c$3$t4qHAw1hpZYN0JKIEpXPcMFMVT81u0hiOw==
     Algorithm          : hmac-md5
     Send lifetime      : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
     Send status        : Inactive
     Accept lifetime    : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
```

```
    Accept status      : Inactive
```

2. When the system time is within the lifetime from 11:00:00 to 12:00:00 on the day 2015/02/06, verify the status of the keys in keychain **abc**.

Display keychain information on Switch A. The output shows that key 2 becomes the valid key.

```
[SwitchA] display keychain

 Keychain name          : abc
   Mode                 : absolute
   Accept tolerance     : 0
   TCP kind value       : 254
   TCP algorithm value
     HMAC-MD5           : 5
     HMAC-SHA-256       : 7
     MD5                : 3
     HMAC-SM3           : 52
     SM3                : 51
   Default send key ID  : None
   Active send key ID   : 2
   Active accept key IDs: 2

   Key ID               : 1
     Key string         : $c$3$dYTC8QeOKJkwFwP2k/rWL+1p6uMTw3MqNg==
     Algorithm          : md5
     Send lifetime      : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
     Send status        : Inactive
     Accept lifetime    : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
     Accept status      : Inactive

   Key ID               : 2
     Key string         : $c$3$7TSPbUxoP1ytOqkdcJ3K3x0BnXEWl4mOEw==
     Algorithm          : hmac-md5
     Send lifetime      : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
     Send status        : Active
     Accept lifetime    : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
     Accept status      : Active
```

Display keychain information on Switch B. The output shows that key 2 becomes the valid key.

```
[SwitchB] display keychain

 Keychain name          : abc
   Mode                 : absolute
   Accept tolerance     : 0
   TCP kind value       : 254
   TCP algorithm value
     HMAC-MD5           : 5
     HMAC-SHA-256       : 7
     MD5                : 3
     HMAC-SM3           : 52
```

```
  SM3                : 51
Default send key ID  : None
Active send key ID   : 1
Active accept key IDs: 1

Key ID               : 1
  Key string         : $c$3$/G/Shnh6heXWprlSQy/XDmftHa2JZJBSgg==
  Algorithm          : md5
  Send lifetime      : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
  Send status        : Inactive
  Accept lifetime    : 10:00:00 2015/02/06 to 11:00:00 2015/02/06
  Accept status      : Inactive

Key ID               : 2
  Key string         : $c$3$t4qHAw1hpZYN0JKIEpXPcMFMVT81u0hiOw==
  Algorithm          : hmac-md5
  Send lifetime      : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
  Send status        : Active
  Accept lifetime    : 11:00:00 2015/02/06 to 12:00:00 2015/02/06
  Accept status      : Active
```

Contents

Managing public keys 1

About public key management 1
Asymmetric key algorithm overview 1
Usage of asymmetric key algorithms 1
FIPS compliance 2
Public key management tasks at a glance 2
Creating a local key pair 2
Distributing a local host public key 3
About distribution of local host public keys 3
Exporting a host public key 4
Displaying a host public key 4
Configuring a peer host public key 5
About peer host public key configuration 5
Restrictions and guidelines for peer host public key configuration 5
Importing a peer host public key from a public key file 5
Entering a peer host public key 5
Destroying a local key pair 6
Configuring master keys 6
Setting a master key 6
Clearing master keys 7
Display and maintenance commands for public keys 8
Examples of public key management 8
Example: Entering a peer host public key 8
Example: Importing a public key from a public key file 10

Managing public keys

About public key management

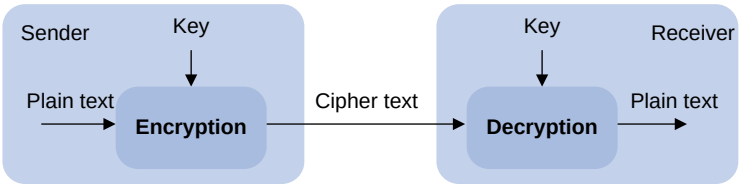
This chapter describes public key management for the following asymmetric key algorithms:

- Revest-Shamir-Adleman Algorithm (RSA).
- Digital Signature Algorithm (DSA).
- Elliptic Curve Digital Signature Algorithm (ECDSA).
- SM2 algorithm.

Asymmetric key algorithm overview

Asymmetric key algorithms are used by security applications to secure communications between two parties, as shown in Figure 1. Asymmetric key algorithms use two separate keys (one public and one private) for encryption and decryption. Symmetric key algorithms use only one key.

Figure 1 Encryption and decryption



A key owner can distribute the public key in plain text on the network but must keep the private key in privacy. It is mathematically infeasible to calculate the private key even if an attacker knows the algorithm and the public key.

The following asymmetric key algorithms are supported:

- Revest-Shamir-Adleman Algorithm (RSA).
- Digital Signature Algorithm (DSA).
- Elliptic Curve Digital Signature Algorithm (ECDSA).

Usage of asymmetric key algorithms

Security applications (such as SSH, SSL, and PKI) use the asymmetric key algorithms for the following purposes:

- **Encryption and decryption**—Any public key receiver can use the public key to encrypt information, but only the private key owner can decrypt the information.
- **Digital signature**—The key owner uses the private key to digitally sign information to be sent. The receiver decrypts the information with the sender's public key to verify information authenticity.

RSA, DSA, and ECDSA can all perform digital signature, but only RSA can perform encryption and decryption.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see "Configuring FIPS") and non-FIPS mode.

Public key management tasks at a glance

To manage public keys, perform the following tasks:

1. Creating a local key pair
2. Distributing a local host public key

   Choose one of the following tasks:

   - Exporting a host public key
   - Displaying a host public key

   To enable the peer device to authenticate the local device, you must distribute the local device's public key to the peer device.
3. Configuring a peer host public key

   Choose one of the following tasks:

   - Importing a peer host public key from a public key file
   - Entering a peer host public key

   To encrypt information sent to a peer device or authenticate the digital signature of the peer device, you must configure the peer device's public key on the local device.
4. (Optional.) Destroying a local key pair
5. (Optional.) Configuring master keys

Creating a local key pair

Restrictions and guidelines

When you create a local key pair, follow these guidelines:

- The key algorithm must be the same as required by the security application.
- When you create an RSA or DSA key pair, enter an appropriate key modulus length at the prompt. The longer the key modulus length, the higher the security, and the longer the key generation time.

  When you create an ECDSA key pair, choose the appropriate elliptic curve. The elliptic curve determines the ECDSA key length. The longer the key length, the higher the security, and the longer the key generation time.

  See Table 1 for more information about key modulus lengths and key lengths.
- If you do not assign the key pair a name, the system assigns the default name to the key pair and marks the key pair as **default**. You can also assign the default name to another key pair, but the system does not mark the key pair as **default**. The key pair name must be unique among all manually named key pairs that use the same key algorithm. If a name conflict occurs, the system asks whether you want to overwrite the existing key pair.
- The key pairs are automatically saved and can survive system reboots.

Table 1 A comparison of different types of asymmetric key algorithms

Type	Generated key pairs	Modulus/key length
RSA	• In non-FIPS mode: ○ One host key pair, if you specify a key pair name. ○ One server key pair and one host key pair, if you do not specify a key pair name. Both key pairs use their default names. • In FIPS mode: One host key pair. **NOTE:** Only SSH 1.5 uses the RSA server key pair.	RSA key modulus length: • In non-FIPS mode: 512 to 4096 bits, 1024 bits by default. To ensure security, use a minimum of 768 bits. • In FIPS mode: A multiple of 256 bits in the range of 2048 to 4096 bits, 2048 bits by default.
DSA	One host key pair.	DSA key modulus length: • In non-FIPS mode: 512 to 2048 bits, 1024 bits by default. To ensure security, use a minimum of 768 bits. • In FIPS mode: The device does not support generating DSA host key pairs.
ECDSA	One host key pair.	ECDSA key length: • In non-FIPS mode: 192, 256, 384, or 521 bits. • In FIPS mode: 256, 384, or 521 bits.

Procedure

1. Enter system view.

   system-view

2. Create a local key pair.

   In non-FIPS mode:

   public-key local create { **dsa** | **ecdsa** [**secp192r1** | **secp256r1** | **secp384r1** | **secp521r1**] | **rsa** } [**name** *key-name*]

   In FIPS mode:

   public-key local create { **ecdsa** [**secp256r1** | **secp384r1** | **secp521r1**] | **rsa** } [**name** *key-name*]

   By default, no local key pairs exist.

Distributing a local host public key

About distribution of local host public keys

You must distribute a local host public key to a peer device so the peer device can perform the following operations:

- Use the public key to encrypt information sent to the local device.
- Authenticate the digital signature signed by the local device.

To distribute a local host public key, you must first export or display the key.

- Export a host public key:
  - Export a host public key to a file.

- Export a host public key to the monitor screen, and then save it to a file.

  After the key is exported to a file, transfer the file to the peer device. On the peer device, import the key from the file.

- Display a host public key.

  After the key is displayed, record the key, for example, copy it to an unformatted file. On the peer device, you must literally enter the key.

Exporting a host public key

Restrictions and guidelines

When you export a host public key, follow these restrictions and guidelines:

- If you specify a file name in the command, the command exports the key to the specified file.
- If you do not specify a file name, the command exports the key to the monitor screen. You must manually save the exported key to a file.

Procedure

1. Enter system view.

   system-view

2. Export a local host public key.
   - Export an RSA host public key:

     In non-FIPS mode:

     public-key local export rsa [**name** *key-name*] { **openssh** | **ssh1** | **ssh2** } [*filename*]

     In FIPS mode:

     public-key local export rsa [**name** *key-name*] { **openssh** | **ssh2** } [*filename*]
   - Export an ECDSA host public key.

     public-key local export ecdsa [**name** *key-name*] { **openssh** | **ssh2** } [*filename*]
   - Export a DSA host public key.

     public-key local export dsa [**name** *key-name*] { **openssh** | **ssh2** } [*filename*]

     This command is supported only in non-FIPS mode.

Displaying a host public key

Perform the following tasks in any view:

- Display local RSA public keys.

  display public-key local rsa public [**name** *key-name*]

  Do not distribute the RSA server public key **serverkey (default)** to a peer device.

- Display local ECDSA public keys.

  display public-key local ecdsa public [**name** *key-name*]

- Display local DSA public keys.

  display public-key local dsa public [**name** *key-name*]

Configuring a peer host public key

About peer host public key configuration

To encrypt information sent to a peer device or authenticate the digital signature of the peer device, you must configure the peer device's public key on the local device.

You can configure the peer host public key by using the following methods:

- Import the peer host public key from a public key file (recommended).
- Manually enter (type or copy) the peer host public key.

For information about how to obtain the host public key of a device, see "Distributing a local host public key."

Restrictions and guidelines for peer host public key configuration

When you configure a peer host public key, follow these restrictions and guidelines:

- When you manually enter the peer host public key, make sure the entered key is in the correct format. To obtain the peer host public key in the correct format, use the **display public-key local public** command to display the public key on the peer device and record the key. The format of the public key displayed in any other way might be incorrect. If the key is not in the correct format, the system discards the key and displays an error message.
- Always import rather than enter the peer host public key if you are not sure whether the device supports the format of the recorded peer host public key.

Importing a peer host public key from a public key file

About this task

Before you perform this task, make sure you have exported the host public key to a file on the peer device and obtained the file from the peer device. For information about exporting a host public key, see "Exporting a host public key."

After you import the key, the system automatically converts the imported public key to a string in the Public Key Cryptography Standards (PKCS) format.

Procedure

1. Enter system view.

   system-view

2. Import a peer host public key from a public key file.

   public-key peer *keyname* **import sshkey** *filename*

   By default, no peer host public keys exist.

Entering a peer host public key

About this task

Before you perform this task, make sure you have displayed the key on the peer device and recorded the key. For information about displaying a host public key, see "Displaying a host public key."

Procedure

1. Enter system view.
   system-view
2. Specify a name for the peer host public key and enter public key view.
   public-key peer *keyname*
3. Type or copy the key.
   You can use spaces and carriage returns, but the system does not save them.
4. Exit public key view.
   peer-public-key end
   When you exit public key view, the system automatically saves the peer host public key.

Destroying a local key pair

About this task

To ensure security, destroy the local key pair and generate a new key pair in any of the following situations:

- The local key has leaked. An intrusion event might occur.
- The storage media of the device is replaced.
- The local certificate has expired. For more information about local certificates, see "Configuring PKI."

Procedure

1. Enter system view.
   system-view
2. Destroy a local key pair.
   In non-FIPS mode:
   public-key local destroy { **dsa** | **ecdsa** | **rsa** } [**name** *key-name*]
   In FIPS mode:
   public-key local destroy { **ecdsa** | **rsa** } [**name** *key-name*]

Configuring master keys

Setting a master key

About this task

By default, a device uses the unified default master key to encrypt the key strings configured for each service module on the device. After you load the exported configuration file of a device on another device, you can use the default master key to decrypt and obtain the configured key strings. In this way, keys can be copied between devices, bringing security risks.

To improve the security and reliability of data, you can set a master key for the device. After you configure this feature on the device, you need to configure the relevant settings as prompted:

1. Validate your identity:
   - If the login authentication mode is **password** or **scheme**, the CLI prompts you to enter your password for login. After you pass the authentication, you go to step 2.

- If the login authentication mode is **none**, you do not need to perform identity authentication but go to step 2 directly.

  For more information about login authentication modes, see login management in *Fundamentals Configuration Guide*.

2. Select whether to enter the master key:
   - If you enter **N** or **n**, the device will automatically and randomly generate a master key.
   - If you enter **Y** or **y**, manually enter the prepared master key and confirm it.

The device stores the master key in the hard disk. Power-off or reboot of the device will not affect the set master key. The device uses the master key to encrypt the key strings configured for each service module on the device.

You can only use the **display master-key type** command to view the type of the master key. You cannot view or obtain the value of the master key by any means. This prevents copying the keys on the device to another device.

Restrictions and guidelines

When the storage space for master keys is insufficient, the device prompts whether to overwrite the oldest master key.

When the login authentication mode is **password** or **scheme**, you will be logged off after three consecutive authentication failures and thus cannot perform master key-related operations.

The minimum interval between setting master keys is 3 minutes.

Prerequisites

To manually enter a master key, prepare the master key and keep it safe before configuring this feature.

Procedure

1. Enter system view.
   system-view
2. Set a master key.
   set master-key

Clearing master keys

About this task

You can configure this feature to clear master keys in the following situations:

- You do not need to use the master key feature.
- The storage space for master keys is insufficient, and you want to clear existing master keys to release storage space.

The default master key of the device uses the fixed factory default value. With the **clear master-key** command executed, all set master keys will be cleared and the default fixed master key will be restored.

Procedure

1. Enter system view.
   system-view
2. Clear master keys.
   clear master-key

Display and maintenance commands for public keys

Execute **display** commands in any view.

Task	Command
Display local public keys.	In non-FIPS mode: **display public-key local** { **dsa** | **ecdsa** | **rsa** } **public** [**name** *key-name*] In FIPS mode: **display public-key local** { **ecdsa** | **rsa** } **public** [**name** *key-name*]
Display peer host public keys.	**display public-key peer** [**brief** | **name** *publickey-name*]
Display the type of the current master key.	**display master-key type**

Examples of public key management

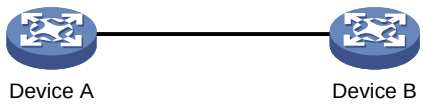
Example: Entering a peer host public key

Network configuration

As shown in Figure 2, to prevent illegal access, Device B authenticates Device A through a digital signature. Before configuring authentication parameters on Device B, use the following procedure to configure the public key of Device A on Device B:

- Create RSA key pairs on Device A and display the public keys of the RSA key pairs.
- Manually specify the RSA host public key of Device A on Device B.

Figure 2 Network diagram



Procedure

1. Configure Device A:

   # Create local RSA key pairs with the default names on Device A, and use the default key modulus length (1024 bits).

```
<DeviceA> system-view
[DeviceA] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
...
```

```
Create the key pair successfully.
```

Display all local RSA public keys.

```
[DeviceA] display public-key local rsa public

=========================================
Key name: hostkey(default)
Key type: RSA
Key length: 1024
Time when key pair created: 16:48:31 2011/05/12
Key code:
   30819F300D06092A864886F70D010101050003818D003081890281810 0DA3B90F59237347B
   8D41B58F8143512880139EC9111BFD31EB84B6B7C7A1470027AC8F04A827B30C2CAF79242E
   45FDFF51A9C7E917DB818D54CB7AEF538AB261557524A7441D288EC54A5D31EFAE4F681257
   6D7796490AF87A8C78F4A7E31F0793D8BA06FB95D54EBB9F94EB1F2D561BF66EA27DFD4788
   CB47440AF6BB25ACA50203010001

=========================================
Key name: serverkey(default)
Key type: RSA
Key length: 768
Time when key pair created: 16:48:31 2011/05/12
Key code:
   307C300D06092A864886F70D0101010500036B003068026100C9451A80F7F0A9BA1A90C7BC
   1C02522D194A2B19F19A75D9EF02219068BD7FD90FCC2AF3634EEB9FA060478DD0A1A49ACE
   E1362A4371549ECD85BA04DEE4D6BB8BE53B6AED7F1401EE88733CA3C4CED391BAE633028A
   AC41C80A15953FB22AA30203010001
```

2. Configure Device B:

Enter the host public key of Device A in public key view. The key must be literally the same as displayed on Device A.

```
<DeviceB> system-view
[DeviceB] public-key peer devicea
Enter public key view. Return to system view with "peer-public-key end" command.
[DeviceB-pkey-public-key-devicea]30819F300D06092A864886F70D010101050003818D003081
8902818100DA3B90F59237347B
[DeviceB-pkey-public-key-devicea]8D41B58F8143512880139EC9111BFD31EB84B6B7C7A14700
27AC8F04A827B30C2CAF79242E
[DeviceB-pkey-public-key-devicea]45FDFF51A9C7E917DB818D54CB7AEF538AB261557524A744
1D288EC54A5D31EFAE4F681257
[DeviceB-pkey-public-key-devicea]6D7796490AF87A8C78F4A7E31F0793D8BA06FB95D54EBB9F
94EB1F2D561BF66EA27DFD4788
[DeviceB-pkey-public-key-devicea]CB47440AF6BB25ACA50203010001
```

Save the public key and return to system view.

```
[DeviceB-pkey-public-key-devicea] peer-public-key end
```

Verifying the configuration

Verify that the peer host public key configured on Device B is the same as the key displayed on Device A.

```
[DeviceB] display public-key peer name devicea
```

```
=============================================
Key name: devicea
Key type: RSA
Key length: 1024
Key code:
    30819F300D06092A864886F70D010101050003818D003081890281810 0DA3B90F59237347B
    8D41B58F8143512880139EC9111BFD31EB84B6B7C7A1470027AC8F04A827B30C2CAF79242E
    45FDFF51A9C7E917DB818D54CB7AEF538AB261557524A7441D288EC54A5D31EFAE4F681257
    6D7796490AF87A8C78F4A7E31F0793D8BA06FB95D54EBB9F94EB1F2D561BF66EA27DFD4788
    CB47440AF6BB25ACA50203010001
```

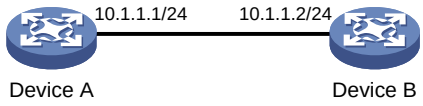
Example: Importing a public key from a public key file

Network configuration

As shown in Figure 3, Device B authenticates Device A through a digital signature. Before configuring authentication parameters on Device B, use the following procedure to configure the public key of Device A on Device B:

- Create RSA key pairs on Device A and export the RSA host public key to a file.
- Import the RSA host public key of Device A from the public key file to Device B.

Figure 3 Network diagram



Procedure

1. Configure Device A:

   # Create local RSA key pairs with the default names on Device A, and use the default key modulus length (1024 bits).

```
<DeviceA> system-view
[DeviceA] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
...
Create the key pair successfully.
```

   # Display all local RSA public keys.

```
[DeviceA] display public-key local rsa public

=============================================
Key name: hostkey(default)
Key type: RSA
Key length: 1024
Time when key pair created: 16:48:31 2011/05/12
Key code:
    30819F300D06092A864886F70D010101050003818D003081890281810 0DA3B90F59237347B
```

```
   8D41B58F8143512880139EC9111BFD31EB84B6B7C7A1470027AC8F04A827B30C2CAF79242E
   45FDFF51A9C7E917DB818D54CB7AEF538AB261557524A7441D288EC54A5D31EFAE4F681257
   6D7796490AF87A8C78F4A7E31F0793D8BA06FB95D54EBB9F94EB1F2D561BF66EA27DFD4788
   CB47440AF6BB25ACA50203010001

=========================================
Key name: serverkey(default)
Key type: RSA
Key length: 768
Time when key pair created: 16:48:31 2011/05/12
Key code:
   307C300D06092A864886F70D0101010500036B003068026100C9451A80F7F0A9BA1A90C7BC
   1C02522D194A2B19F19A75D9EF02219068BD7FD90FCC2AF3634EEB9FA060478DD0A1A49ACE
   E1362A4371549ECD85BA04DEE4D6BB8BE53B6AED7F1401EE88733CA3C4CED391BAE633028A
   AC41C80A15953FB22AA30203010001
```

Export the RSA host public key to file **devicea.pub**.

```
[DeviceA] public-key local export rsa ssh2 devicea.pub
```

Enable the FTP server, create an FTP user with username **ftp** and password **123456TESTplat&!**, and configure the FTP user role as **network-admin**.

```
[DeviceA] ftp server enable
[DeviceA] local-user ftp
[DeviceA-luser-manage-ftp] password simple 123456TESTplat&!
[DeviceA-luser-manage-ftp] service-type ftp
[DeviceA-luser-manage-ftp] authorization-attribute user-role network-admin
[DeviceA-luser-manage-ftp] quit
```

2. Configure Device B:

Use FTP in binary mode to get public key file **devicea.pub** from Device A.

```
<DeviceB> ftp 10.1.1.1
Connected to 10.1.1.1 (10.1.1.1).
220 FTP service ready.
User(10.1.1.1:(none)):ftp
331 Password required for ftp.
Password:
230 User logged in.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> binary
200 TYPE is now 8-bit binary
ftp> get devicea.pub
227 Entering Passive Mode (10,1,1,1,118,252)
150 Accepted data connection
226 File successfully transferred
301 bytes received in 0.003 seconds (98.0 kbyte/s)
ftp> quit
221-Goodbye. You uploaded 0 and downloaded 1 kbytes.
221 Logout.
```

Import the host public key from key file **devicea.pub**.

```
<DeviceB> system-view
```

```
[DeviceB] public-key peer devicea import sshkey devicea.pub
```

Verifying the configuration

Verify that the peer host public key configured on Device B is the same as the key displayed on Device A.

```
[DeviceB] display public-key peer name devicea

=============================================
Key name: devicea
Key type: RSA
Key length: 1024
Key code:
   30819F300D06092A864886F70D010101050003818D003081890281810 0DA3B90F59237347B
   8D41B58F8143512880139EC9111BFD31EB84B6B7C7A1470027AC8F04A827B30C2CAF79242E
   45FDFF51A9C7E917DB818D54CB7AEF538AB261557524A7441D288EC54A5D31EFAE4F681257
   6D7796490AF87A8C78F4A7E31F0793D8BA06FB95D54EBB9F94EB1F2D561BF66EA27DFD4788
   CB47440AF6BB25ACA50203010001
```

Contents

Configuring PKI 1

About PKI 1
PKI terminology 1
PKI architecture 2
Retrieval, usage, and maintenance of a digital certificate 3
PKI applications 3
FIPS compliance 3
PKI tasks at a glance 4
Configuring a PKI entity 4
Configuring a PKI domain 5
About PKI domain 5
PKI domain tasks at a glance 5
Creating a PKI domain 6
Specifying the trusted CA 6
Specifying the PKI entity name 6
Specifying the certificate request reception authority 7
Specifying the certificate request URL 7
Setting the SCEP polling interval and maximum polling attempts 7
Specifying the LDAP server 7
Specifying the fingerprint for root CA certificate verification 8
Specifying the key pair for certificate request 8
Specifying the intended purpose for the certificate 9
Specifying the source IP address for PKI protocol packets 9
Specifying the storage path for certificates and CRLs 10
Requesting a certificate 10
About certificate request configuration 10
Restrictions and guidelines for certificate request configuration 10
Prerequisites for certificate request configuration 11
Enabling the automatic online certificate request mode 11
Manually submitting an online certificate request 11
Manually submitting a certificate request in offline mode 12
Aborting a certificate request 13
Obtaining certificates 13
Verifying PKI certificates 14
About certification verification 14
Restrictions and guidelines for certificate verification 15
Verifying certificates with CRL checking 15
Verifying certificates without CRL checking 16
Exporting certificates 16
Removing a certificate 17
Configuring a certificate-based access control policy 17
About certificate-based access control policies 17
Procedure 18
Enabling local certificate expiration notification 18
Enabling SNMP notifications for PKI 19
Enabling PKI logging 19
Display and maintenance commands for PKI 20
PKI configuration examples 20
Example: Requesting a certificate from an RSA Keon CA server 20
Example: Requesting a certificate from a Windows Server 2003 CA server 23
Example: Requesting a certificate from an OpenCA server 26
Example: Configuring IKE negotiation with RSA digital signature from a Windows Server 2003 CA server 29
Example: Importing and exporting certificates 32
Troubleshooting PKI configuration 37
Failed to obtain the CA certificate 37
Failed to obtain local certificates 38

Failed to request local certificates ····· 38
Failed to obtain CRLs ····· 39
Failed to import the CA certificate ····· 40
Failed to import the local certificate ····· 40
Failed to export certificates ····· 41
Failed to set the storage path ····· 41

Configuring PKI

About PKI

Public Key Infrastructure (PKI) is an asymmetric key infrastructure to encrypt and decrypt data for securing network services.

PKI uses digital certificates to distribute and employ public keys, and provides network communication and e-commerce with security services such as user authentication, data confidentiality, and data integrity. For more information about public keys, see "Managing public keys."

PKI terminology

Digital certificate

A digital certificate is an electronic document signed by a CA that binds a public key with the identity of its owner.

A digital certificate includes the following information:

- Issuer name (name of the CA that issued the certificate).
- Subject name (name of the individual or group to which the certificate is issued).
- Identity information of the subject.
- Subject's public key.
- Signature of the CA.
- Validity period.

A digital certificate must comply with the international standards of ITU-T X.509, of which X.509 v3 is the most commonly used.

This chapter covers the following types of certificates:

- **CA certificate**—Certificate of a CA. Multiple CAs in a PKI system form a CA tree, with the root CA at the top. The root CA generates a self-signed certificate, and each lower level CA holds a CA certificate issued by the CA immediately above it. The chain of these certificates forms a chain of trust.
- **Registration authority (RA) certificate**—Certificate issued by a CA to an RA. RAs act as proxies for CAs to process enrollment requests in a PKI system.
- **Local certificate**—Certificate of a local PKI entity, which contains the entity's public key. Local certificates include the following types:
  - **CA-signed certificate**—Digital certificate signed and issued by a CA to the local PKI entity.
  - **Self-signed certificate**—Digital certificate signed and issued by the local PKI entity (the device) itself. When the device cannot request a local certificate from the CA, it can sign and issue a local certificate for communication with the peer to provide services. Using this type of certificate is simple in configuration but has low security. You can use self-signed certificates in scenarios with low security requirements.
- **Peer certificate**—CA-signed digital certificate of a peer, which contains the peer's public key.

Fingerprint of root CA certificate

Each root CA certificate has a unique fingerprint, which is the hash value of the certificate content. The fingerprint of a root CA certificate can be used to authenticate the validity of the root CA.

Certificate revocation list

A certificate revocation list (CRL) is a list of serial numbers for certificates that have been revoked. A CRL is created and signed by the CA that originally issued the certificates.

The CA publishes CRLs periodically to revoke certificates. Entities that are associated with the revoked certificates should not be trusted.

The CA must revoke a certificate when any of the following conditions occurs:

- The certificate subject name is changed.
- The private key is compromised.
- The association between the subject and CA is changed. For example, when an employee terminates employment with an organization.

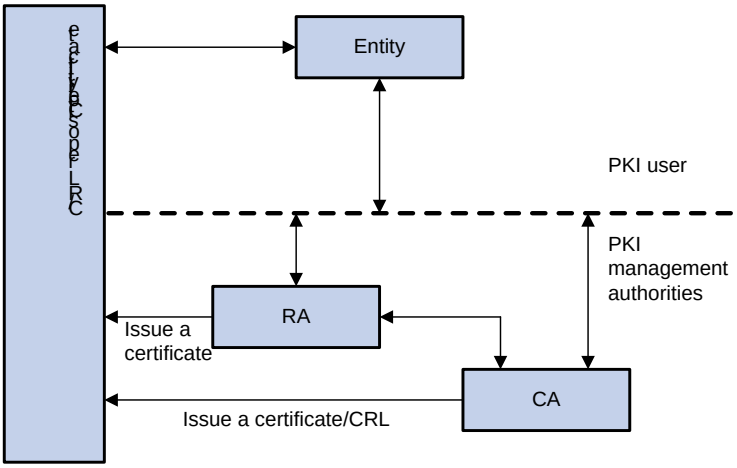
CA policy

A CA policy is a set of criteria that a CA follows to process certificate requests, to issue and revoke certificates, and to publish CRLs. Typically, a CA advertises its policy in a certification practice statement (CPS). You can obtain a CA policy through out-of-band means such as phone, disk, and email. Make sure you understand the CA policy before you select a trusted CA for certificate request because different CAs might use different policies.

PKI architecture

A PKI system consists of PKI entities, CAs, RAs and a certificate/CRL repository, as shown in Figure 1.

Figure 1 PKI architecture



PKI entity

A PKI entity is an end user using PKI certificates. The PKI entity can be an operator, an organization, a device like a router or a switch, or a process running on a computer. PKI entities use SCEP to communicate with the CA or RA.

CA

Certification authority that grants and manages certificates. A CA issues certificates, defines the certificate validity periods, and revokes certificates by publishing CRLs.

RA

Registration authority, which offloads the CA by processing certificate enrollment requests. The RA accepts certificate requests, verifies user identity, and determines whether to ask the CA to issue certificates.

The RA is optional in a PKI system. In cases when there is security concern over exposing the CA to direct network access, it is advisable to delegate some of the tasks to an RA. Then, the CA can concentrate on its primary tasks of signing certificates and CRLs.

Certificate/CRL repository

A certificate distribution point that stores certificates and CRLs, and distributes these certificates and CRLs to PKI entities. It also provides the query function. A PKI repository can be a directory server using the LDAP or HTTP protocol, of which LDAP is commonly used.

Retrieval, usage, and maintenance of a digital certificate

The following workflow describes the retrieval, usage, and maintenance of a digital certificate. This example uses a CA which has an RA to process certificate enrollment requests.

1. A PKI entity generates an asymmetric key pair and submits a certificate request to the RA.

   The certificate request contains the public key and its identity information.
2. The RA verifies the identity of the entity and sends a digital signature containing the identity information and the public key to the CA.
3. The CA verifies the digital signature, approves the request, and issues a certificate.
4. After receiving the certificate from the CA, the RA sends the certificate to the certificate repository and notifies the PKI entity that the certificate has been issued.
5. The PKI entity obtains the certificate from the certificate repository.
6. To establish a secure connection for communication, two PKI entities exchange local certificates to authenticate each other. The connection can be established only if both entities verify that the peer's certificate is valid.
7. You can remove the local certificate of a PKI entity and request a new one when any of the following conditions occur:
   - The local certificate is about to expire.
   - The certificate's private key is compromised.

PKI applications

The PKI technology can meet security requirements of online transactions. As an infrastructure, PKI has a wide range of applications. H3C's PKI system can provide certificate management for IPsec and SSL.

The following are some application examples.

VPN

A VPN is a private data communication network built on the public communication infrastructure. A VPN can use network layer security protocols (for example, IPsec) in conjunction with PKI-based encryption and digital signature technologies for confidentiality.

Secure emails

PKI can address the email requirements for confidentiality, integrity, authentication, and non-repudiation. A common secure email protocol is Secure/Multipurpose Internet Mail Extensions (S/MIME), which is based on PKI and allows for transfer of encrypted mails with signature.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see "Configuring FIPS") and non-FIPS mode.

PKI tasks at a glance

To configure PKI, perform the following tasks:

1. Configuring a PKI entity
2. Configuring a PKI domain
3. (Optional.) Specifying the storage path for certificates and CRLs
4. Requesting a certificate

   Choose one of the following tasks:

   - Enabling the automatic online certificate request mode
   - Manually submitting an online certificate request
   - Manually submitting a certificate request in offline mode
5. (Optional.) Aborting a certificate request
6. (Optional.) Obtaining certificates

   You can obtain the CA certificate, local certificates, and peer certificates related to a PKI domain from a CA and save them locally for higher lookup efficiency.
7. (Optional.) Verifying PKI certificates
8. (Optional.) Exporting certificates
9. (Optional.) Removing a certificate
10. (Optional.) Configuring a certificate-based access control policy

    Certificate-based access control policies allow you to authorize access to a device (for example, an HTTPS server) based on the attributes of an authenticated client's certificate.
11. (Optional.) Enabling local certificate expiration notification
12. (Optional.) Enabling SNMP notifications for PKI
13. (Optional.) Enabling PKI logging

Configuring a PKI entity

About this task

A certificate applicant uses an entity to provide its identity information to a CA. A valid PKI entity must include one or more of following identity categories:

- Distinguished name (DN) of the entity, which further includes the common name, country code, locality, organization, unit in the organization, and state. If you configure the DN for an entity, a common name is required.
- FQDN of the entity.
- IP address of the entity.

Restrictions and guidelines

Follow these restrictions and guidelines when you configure a PKI entity:

- Whether the identity categories are required or optional depends on the CA policy. Follow the CA policy to configure the entity settings. For example, if the CA policy requires the entity DN, but you configure only the IP address, the CA rejects the certificate request from the entity.
- The SCEP add-on on the Windows 2000 CA server has restrictions on the data length of a certificate request. If a request from a PKI entity exceeds the data length limit, the CA server does not respond to the certificate request. In this case, you can use an out-of-band means to submit the request. Other types of CA servers, such as RSA servers and OpenCA servers, do not have such restrictions.

Procedure

1. Enter system view.
   system-view
2. Create a PKI entity and enter its view.
   pki entity *entity-name*
3. Set a common name for the entity.
   common-name *common-name-sting*
   By default, the common name is not set.
4. Set the country code of the entity.
   country *country-code-string*
   By default, the country code is not set.
5. Set the locality of the entity.
   locality *locality-name*
   By default, the locality is not set.
6. Set the organization of the entity.
   organization *org-name*
   By default, the organization is not set.
7. Set the unit of the entity in the organization.
   organization-unit *org-unit-name*
   By default, the unit is not set.
8. Set the state where the entity resides.
   state *state-name*
   By default, the state is not set.
9. Set the FQDN of the entity.
   fqdn *fqdn-name-string*
   By default, the FQDN is not set.
10. Configure the IP address of the entity.
    ip { *ip-address* | **interface** *interface-type interface-number* }
    By default, the IP address is not configured.
11. Configure the title of the entity.
    title *title*
    By default, the title is not configured.

Configuring a PKI domain

About PKI domain

A PKI domain contains enrollment information for a PKI entity. It is locally significant and is intended only for use by other applications like IKE and SSL.

PKI domain tasks at a glance

To configure a PKI domain, perform the following tasks:

1. Creating a PKI domain

2. Specifying the trusted CA
3. Specifying the PKI entity name
4. Specifying the certificate request reception authority
5. Specifying the certificate request URL
6. (Optional.) Setting the SCEP polling interval and maximum polling attempts
7. Specifying the LDAP server

   This task is required when either of the following conditions is met:

   - The device must obtain certificates from the CA by using the LDAP protocol.
   - An LDAP URL which does not contain the host name of the LDAP server is specified as the CRL repository URL.
8. Specifying the fingerprint for root CA certificate verification

   This step is required if the auto certificate request mode is configured in the PKI domain.

   If the manual certificate request mode is configured, you can skip this step and manually verify the fingerprint displayed during verification of the root CA certificate.
9. Specifying the key pair for certificate request
10. (Optional.) Specifying the intended purpose for the certificate
11. (Optional.) Specifying the source IP address for PKI protocol packets

Creating a PKI domain

1. Enter system view.

   system-view
2. Create a PKI domain and enter its view.

   pki domain *domain-name*

Specifying the trusted CA

About this task

The PKI domain must have a CA certificate before you can request a local certificate. To obtain a CA certificate, the trusted CA name must be specified. The trusted CA name uniquely identifies the CA to be used if multiple CAs exist on the CA server.

Procedure

1. Enter system view.

   system-view
2. Enter PKI domain view.

   pki domain *domain-name*
3. Specify the trusted CA name.

   ca identifier *name*

   By default, no trusted CA name is specified.

Specifying the PKI entity name

1. Enter system view.

   system-view
2. Enter PKI domain view.

pki domain *domain-name*

3. Specify the PKI entity name.
   certificate request entity *entity-name*
   By default, no PKI entity name is specified.

Specifying the certificate request reception authority

1. Enter system view.
   system-view
2. Enter PKI domain view.
   pki domain *domain-name*
3. Specify the certificate request reception authority.
   certificate request from { **ca** | **ra** }
   By default, no certificate request reception authority is specified.

Specifying the certificate request URL

1. Enter system view.
   system-view
2. Enter PKI domain view.
   pki domain *domain-name*
3. Specify the URL of the certificate request reception authority to which the device sends certificate requests.
   certificate request url *url-string* [**vpn-instance** *vpn-instance-name*]
   By default, the certificate request URL is not specified.

Setting the SCEP polling interval and maximum polling attempts

1. Enter system view.
   system-view
2. Enter PKI domain view.
   pki domain *domain-name*
3. Set the SCEP polling interval and maximum number of polling attempts.
   certificate request polling { **count** *count* | **interval** *interval* }
   By default, the device polls the CA server for the certificate request status every 20 minutes. The maximum number of polling attempts is 50.

Specifying the LDAP server

1. Enter system view.
   system-view
2. Enter PKI domain view.
   pki domain *domain-name*
3. Specify the LDAP server.

ldap-server host *hostname* [**port** *port-number*] [**vpn-instance** *vpn-instance-name*]

By default, no LDAP server is specified.

Specifying the fingerprint for root CA certificate verification

1. Enter system view.

   system-view
2. Enter PKI domain view.

   pki domain *domain-name*
3. Configure the fingerprint for verifying the root CA certificate.

   In non-FIPS mode:

   root-certificate fingerprint { **md5** | **sha1** } *string*

   In FIPS mode:

   root-certificate fingerprint sha1 *string*

   By default, no fingerprint is configured.

Specifying the key pair for certificate request

Restrictions and guidelines

You can specify a nonexistent key pair for certificate request. The PKI entity automatically creates the key pair before submitting a certificate request.

Procedure

1. Enter system view.

   system-view
2. Enter PKI domain view.

   pki domain *domain-name*
3. Specify the key pair for certificate request.
   - Specify an RSA key pair.

     public-key rsa { { **encryption name** *encryption-key-name* [**length** *key-length*] | **signature name** *signature-key-name* [**length** *key-length*] } * | **general name** *key-name* [**length** *key-length*] }
   - Specify an ECDSA key pair.

     In non-FIPS mode:

     public-key ecdsa name *key-name* [**secp192r1** | **secp256r1** | **secp384r1** | **secp521r1**]

     In FIPS mode:

     public-key ecdsa name *key-name* [**secp256r1** | **secp384r1** | **secp521r1**]
   - Specify a DSA key pair.

     public-key dsa name *key-name* [**length** *key-length*]

     This command is supported only in non-FIPS mode.

   By default, no key pair is specified.

Specifying the intended purpose for the certificate

About this task

An issued certificate contains the extensions which restrict the usage of the certificate to specific purposes. You can specify the intended purposes for a certificate, which will be included in the certificate request sent to the CA. However, the actual extensions contained in an issued certificate depend on the CA policy, and they might be different from those specified in the PKI domain. Whether an application will use the certificate during authentication depends on the application's policy.

Supported certificate extensions include:

- **ike**—Certificates carrying this extension can be used by IKE peers.
- **ssl-client**—Certificates carrying this extension can be used by SSL clients.
- **ssl-server**—Certificates carrying this extension can be used by SSL servers.

Procedure

1. Enter system view.

   system-view
2. Enter PKI domain view.

   pki domain *domain-name*
3. Specify the intended use for the certificate.

   usage { **ike** | **ssl-client** | **ssl-server** } *

   By default, the certificate can be used by all supported applications, including IKE, SSL client, and SSL server.

Specifying the source IP address for PKI protocol packets

About this task

This task is required if the CA policy requires that the CA server accept certificate requests from a specific IP address or subnet.

Procedure

1. Enter system view.

   system-view
2. Enter PKI domain view.

   pki domain *domain-name*
3. Specify a source IP address for the PKI protocol packets.

   IPv4:

   source ip { *ip-address* | **interface** *interface-type interface-number* }

   IPv6:

   source ipv6 { *ipv6-address* | **interface** *interface-type interface-number* }

   By default, the source IP address of PKI protocol packets is the IP address of their outgoing interface.

Specifying the storage path for certificates and CRLs

About this task

The device has a default storage path for certificates and CRLs. You can change the storage path and specify different paths for the certificates and CRLs.

After you change the storage path for certificates or CRLs, the certificate files and CRL files in the original path are moved to the new path. Certificate files use the .cer or .p12 file extension and CRL files use the .crl file extension.

Restrictions and guidelines

If you change the storage path, save the configuration before you reboot or shut down the device to avoid loss of certificates or CRLs.

Procedure

1. Enter system view.

   system-view

2. Specify the storage path for certificates and CRLs.

   pki storage { **certificates** | **crls** } *dir-path*

   By default, the device stores certificates and CRLs in the PKI directory on the storage media of the device.

Requesting a certificate

About certificate request configuration

To request a certificate, a PKI entity must provide its identity information and public key to a CA.

A certificate request can be submitted to a CA in offline or online mode.

- **Offline mode**—A certificate request is submitted by using an out-of-band method, such as phone, disk, or email.
- **Online mode**—A certificate request can be automatically or manually submitted to a CA through the Simple Certificate Enrollment Protocol (SCEP).

Restrictions and guidelines for certificate request configuration

When you request a local certificate in a PKI domain, follow these restrictions and guidelines:

- To prevent an existing local certificate from becoming invalid, do not perform the following tasks:
  - Create a key pair with the same name as the key pair contained in the certificate.

    To create a key pair, use the **public-key local create** command.
  - Destroy the key pair contained in the certificate.

    To destroy a key pair, use the **public-key local destroy** command.
- To manually request a new certificate in a PKI domain that already has a local certificate, use the following procedure:

  a. Use the **pki delete-certificate** command to delete the existing local certificate.

b. Use the **public-key local create** command to generate a new key pair.

c. Manually submit a certificate request.

- A PKI domain can have local certificates using only one type of cryptographic algorithms (DSA, ECDSA, or RSA). If DSA or ECDSA is used, a PKI domain can have only one local certificate. If RSA is used, a PKI domain can have one local certificate for signature, and one local certificate for encryption.

Prerequisites for certificate request configuration

Make sure the device is time synchronized with the CA server. If the device is not time synchronized with the CA server, the certificate request might fail because the certificate might be considered to be outside of the validity period. For information about configuring the system time, see *Fundamentals Configuration Guide*.

Enabling the automatic online certificate request mode

About this task

In auto request mode, a PKI entity with no local certificates automatically submits a certificate request to the CA when an application works with the PKI entity. For example, when IKE negotiation uses a digital signature for identity authentication, but no local certificate is available, the entity automatically submits a certificate request. It saves the certificate locally after obtaining the certificate from the CA.

A CA certificate must be present before you request a local certificate. If no CA certificate exists in the PKI domain, the PKI entity automatically obtains a CA certificate before sending a certificate request.

Restrictions and guidelines

In auto request mode, the device does not automatically request a new certificate if the current certificate is about to expire or has expired, which might cause service interruptions.

Procedure

1. Enter system view.

system-view

2. Enter PKI domain view.

pki domain *domain-name*

3. Enable the automatic online certificate request mode.

certificate request mode auto [**password** { **cipher** | **simple** } *string*]

By default, the manual request mode applies.

If the CA policy requires a password for certificate revocation, specify the password in this command.

Manually submitting an online certificate request

About this task

In manual request mode, you must execute the **pki request-certificate domain** command to request a local certificate in a PKI domain. The certificate will be saved in the domain after it is obtained from the CA.

Procedure

1. Enter system view.

system-view

2. Enter PKI domain view.

   pki domain *domain-name*

3. Set the certificate request mode to manual.

   certificate request mode manual

   By default, the manual request mode applies.

4. Return to system view.

   quit

5. Obtain a CA certificate.

   See "Obtaining certificates."

   This step is required if the PKI domain does not have a CA certificate. The CA certificate is used to verify the authenticity and validity of the obtained local certificate.

6. Manually submit an SCEP certificate request.

   pki request-certificate domain *domain-name* [**password** *password*]

   This command is not saved in the configuration file.

   If the CA policy requires a password for certificate revocation, specify the password in this command.

Manually submitting a certificate request in offline mode

About this task

Use this method if the CA does not support SCEP or if a network connection between the device and CA is not possible.

Procedure

1. Enter system view.

   system-view

2. Enter PKI domain view.

   pki domain *domain-name*

3. Set the certificate request mode to manual.

   certificate request mode manual

   By default, the manual request mode applies.

4. Return to system view.

   quit

5. Obtain the CA certificate.

   See "Obtaining certificates."

   This step is required if the PKI domain does not have a CA certificate. The CA certificate is used to verify the authenticity and validity of the obtained local certificate.

6. Print the certificate request in PKCS10 format on the terminal or save the certificate request to a PKCS10 file.

   pki request-certificate domain *domain-name* **pkcs10** [**filename** *filename*]

   This command is not saved in the configuration file.

7. Transfer certificate request information to the CA by using an out-of-band method.

8. Transfer the issued local certificate from the CA to the local device by using an out-of-band method.

9. Import the local certificate to the PKI domain.

   pki import domain *domain-name* { **der local filename** *filename* | **p12 local filename** *filename* | **pem local** [**filename** *filename*] }

Aborting a certificate request

About this task

Before the CA issues a certificate, you can abort a certificate request and change its parameters, such as the common name, country code, or FQDN. You can use the **display pki certificate request-status** command to display the status of a certificate request.

Alternatively, you also can remove a PKI domain to abort the associated certificate request.

Procedure

1. Enter system view.

   system-view

2. Abort a certificate request.

   pki abort-certificate-request domain *domain-name*

   This command is not saved in the configuration file.

Obtaining certificates

About this task

You can obtain the CA certificate, local certificates, and peer certificates related to a PKI domain from a CA and save them locally for higher lookup efficiency. To do so, use either the offline mode or the online mode:

- In offline mode, obtain the certificates by an out-of-band means like FTP, disk, or email, and then import them locally. Use this mode when the CRL repository is not specified, the CA server does not support SCEP, or the CA server generates the key pair for the certificates.
- In online mode, you can obtain the CA certificate through SCEP and obtain local certificates or peer certificates through LDAP.

Restrictions and guidelines

Follow these restrictions and guidelines when obtain certificates from a CA

- If a CA certificate already exists locally, you cannot obtain it again in online mode. If you want to obtain a new CA certificate, use the **pki delete-certificate** command to delete the existing CA certificate and local certificates first.
- If local or peer certificates already exist, you can obtain new local or peer certificates to overwrite the existing ones. If RSA is used, a PKI domain can have two local certificates, one for signature and the other for encryption.
- If CRL checking is enabled, obtaining a certificate triggers CRL checking. If the certificate to be obtained has been revoked, the certificate cannot be obtained.
- The device compares the validity period of a certificate with the local system time to determine whether the certificate is valid. Make sure the system time of the device is synchronized with the CA server.

Prerequisites

- Before you obtain local or peer certificates in online mode, make sure an LDAP server is correctly configured in the PKI domain.
- Before you import certificates in offline mode, complete the following tasks:

- Use FTP or TFTP to upload the certificate files to the storage media of the device.

  If FTP or TFTP is not available, display and copy the contents of a certificate to a file on the device. Make sure the certificate is in PEM format because only certificates in PEM format can be imported.
- Before you import a local certificate or peer certificate, obtain the CA certificate chain that signs the certificate.

  This step is required only if the CA certificate chain is neither available in the PKI domain nor contained in the certificate to be imported.
- Before you import a local certificate that contains an encrypted key pair, contact the CA administrator to obtain the password required for importing the certificate.

Procedure

1. Enter system view.

   system-view
2. Obtain certificates.
   - Import certificates in offline mode.

     pki import domain *domain-name* { **der** { **ca** | **local** | **peer** } **filename** *filename* | **p12 local filename** *filename* | **pem** { **ca** | **local** | **peer** } [**filename** *filename*] }
   - Obtain certificates in online mode.

     pki retrieve-certificate domain *domain-name* { **ca** | **local** | **peer** *entity-name* }

     This command is not saved in the configuration file.

Verifying PKI certificates

About certification verification

A certificate is automatically verified when it is requested, obtained, or used by an application. If the certificate expires, if it is not issued by a trusted CA, or if it is revoked, the certificate cannot be used You can also manually verify a certificate.

You can enable or disable CRL checking in a PKI domain. CRL checking checks whether a certificate is in the CRL. If it is, the certificate has been revoked and its home entity is not trusted.

To use CRL checking, a CRL must be obtained from a CRL repository. The device selects a CRL repository in the following order:

1. CRL repository specified in the PKI domain by using the **crl url** command.
2. CRL repository in the certificate that is being verified.
3. CRL repository in the CA certificate or CRL repository in the upper-level CA certificate if the certificate being verified is a CA certificate

If no CRL repository is found after the selection process, the device obtains the CRL through SCEP. In this scenario, the CA certificate and the local certificates must have been obtained.

A certificate fails CRL checking in the following situations:

- A CRL cannot be obtained during CRL checking of the certificate.
- CRL checking verifies that the certificate has been revoked.

Restrictions and guidelines for certificate verification

When verifying the CA certificate of a PKI domain, the system needs to verify all the certificates in the CA certificate chain. To ensure a successful certificate verification process, the device must have all the PKI domains to which the CA certificates in the certificate chain belong.

The system verifies the CA certificates in the CA certificate chain as follows:

1. Identifies the parent certificate of the lowest-level certificate.

   Each CA certificate contains an issuer field that identifies the parent CA that issued the certificate.
2. Locates the PKI domain to which the parent certificate belongs.
3. Performs CRL checking in the PKI domain to check whether the parent certificate has been revoked. If it has been revoked, the certificate cannot be used.

   This step will not be performed when CRL checking is disabled in the PKI domain.
4. Repeats the previous steps for upper-level certificates in the CA certificate chain until the root CA certificate is reached.
5. Verifies that each CA certificate in the certificate chain is issued by the named parent CA, starting from the root CA.

Verifying certificates with CRL checking

1. Enter system view.

   system-view
2. Enter PKI domain view.

   pki domain *domain-name*
3. (Optional.) Specify the URL of the CRL repository.

   crl url *url-string* [**vpn-instance** *vpn-instance-name*]

   By default, the URL of the CRL repository is not specified.
4. Enable CRL checking.

   crl check enable

   By default, CRL checking is enabled.
5. Return to system view.

   quit
6. Obtain the CA certificate.

   See "Obtaining certificates."

   The PKI domain must have a CA certificate before you can verify certificates in it.
7. (Optional.) Obtain the CRL.
   - Obtain the CRL from the CRL repository and save it locally.

     pki retrieve-crl domain *domain-name*

     To verify a non-root CA certificate and local certificates, the device automatically retrieves the CRL if the PKI domain has no CRL.

     The newly obtained CRL overwrites the old one, if any.

     The obtained CRL is issued by a CA in the CA certificate chain stored in the PKI domain.
   - Import a CRL file in the specified local path.

     pki retrieve-crl domain *domain-name*

     Use this method if no CRL repository exists in the environment.
8. Manually verify the validity of the certificates.

pki validate-certificate domain *domain-name* { **ca** | **local** }

Verifying certificates without CRL checking

1. Enter system view.
   system-view
2. Enter PKI domain view.
   pki domain *domain-name*
3. Disable CRL checking.
   undo crl check enable
   By default, CRL checking is enabled.
4. Return to system view.
   quit
5. Obtain a CA certificate for the PKI domain.
   See "Obtaining certificates."
   The PKI domain must have a CA certificate before you can verify certificates in it.
6. Manually verify the certificate validity.
   pki validate-certificate domain *domain-name* { **ca** | **local** }
   This command is not saved in the configuration file.

Exporting certificates

About this task

You can export the CA certificate and the local certificates in a PKI domain to certificate files. The exported certificate files can then be imported back to the device or other PKI applications.

Restrictions and guidelines

To export all certificates in PKCS12 format, the PKI domain must have a minimum of one local certificate. If the PKI domain does not have any local certificates, the certificates in the PKI domain cannot be exported.

If you do not specify a file name when you export a certificate in PEM format, this command displays the certificate content on the terminal.

When you export a local certificate with RSA key pairs to a file, the certificate file name might be different from the file name specified in the command. The actual certificate file name depends on the purpose of the key pair contained in the certificate. For more information about the file naming rule, see the **pki export** command in *Security Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Export certificates.
   - Export certificates in DER format.
     pki export domain *domain-name* **der** { **all** | **ca** | **local** } **filename** *filename*
   - Export certificates in PKCS12 format.
     pki export domain *domain-name* **p12** { **all** | **local** } **passphrase** *p12-key* **filename** *filename*

- Export certificates in PEM format.

  pki export domain *domain-name* **pem** { { **all** | **local** } [{ **3des-cbc** | **aes-128-cbc** | **aes-192-cbc** | **aes-256-cbc** | **des-cbc** } *pem-key*] | **ca** } [**filename** *filename*]

Removing a certificate

About this task

You can remove certificates from a PKI domain in the following situations:

- Remove a CA certificate, local certificate, or peer certificate if the certificate has expired or is about to expire.
- Remove a local certificate if the certificate's private key is compromised, or if you want to request a new local certificate to replace the existing one.

Restrictions and guidelines

After you remove the CA certificate, the system automatically removes the local certificates, peer certificates, and CRLs from the domain.

To remove a local certificate and request a new certificate, perform the following tasks:

1. Remove the local certificate.
2. Use the **public-key local destroy** command to destroy the existing local key pair.
3. Use the **public-key local create** command to generate a new key pair.
4. Request a new certificate.

For more information about the **public-key local destroy** and **public-key local create** commands, see *Security Command Reference*.

Procedure

1. Enter system view.

   system-view
2. Remove a certificate.

   pki delete-certificate domain *domain-name* { **ca** | **local** | **peer** [**serial** *serial-num*] }

   If you use the **peer** keyword without specifying a serial number, this command removes all peer certificates.

Configuring a certificate-based access control policy

About certificate-based access control policies

Certificate-based access control policies allow you to authorize access to a device (for example, an HTTPS server) based on the attributes of an authenticated client's certificate.

Access control rules and certificate attribute groups

A certificate-based access control policy is a set of access control rules (permit or deny statements), each associated with a certificate attribute group. A certificate attribute group contains multiple attribute rules, each defining a matching criterion for an attribute in the certificate issuer name, subject name, or alternative subject name field.

Certificate matching mechanism

If a certificate matches all attribute rules in a certificate attribute group associated with an access control rule, the system determines that the certificate matches the access control rule. In this scenario, the match process stops, and the system performs the access control action defined in the access control rule.

The following conditions describe how a certificate-based access control policy verifies the validity of a certificate:

- If a certificate matches a permit statement, the certificate passes the verification.
- If a certificate matches a deny statement or does not match any statements in the policy, the certificate is regarded invalid.
- If a statement is associated with a non-existing attribute group, or the attribute group does not have attribute rules, the certificate matches the statement.
- If the certificate-based access control policy specified for a security application (for example, HTTPS) does not exist, all certificates in the application pass the verification.

Procedure

1. Enter system view.

   system-view
2. Create a certificate attribute group and enter its view.

   pki certificate attribute-group *group-name*
3. Configure an attribute rule for issuer name, subject name, or alternative subject name.

   attribute *id* { **alt-subject-name** { **fqdn** | **ip** } | { **issuer-name** | **subject-name** } { **dn** | **fqdn** | **ip** } } { **ctn** | **equ** | **nctn** | **nequ**} *attribute-value*

   By default, not attribute rules are configured.
4. Return to system view.

   quit
5. Create a certificate-based access control policy and enter its view.

   pki certificate access-control-policy *policy-name*

   By default, no certificate-based access control policies exist.
6. Create a certificate access control rule.

   rule [*id*] { **deny** | **permit** } *group-name*

   By default, no certificate access control rules are configured, and all certificates can pass the verification.

   You can create multiple certificate access control rules for a certificate-based access control policy.

Enabling local certificate expiration notification

About this task

After this feature is enabled, the system checks the validity date for local certificates every other hour. When a local certificate is about to expire in 30 days (included) or has expired, the system sends a notification log message for the certificate every other day.

Procedure

1. Enter system view.

system-view

2. Enable local certificate expiration notification.

   pki certificate logging { **local-will-expire** | **local-has-expired** } **enable**

   By default, local certificate expiration notification is disabled.

Enabling SNMP notifications for PKI

About this task

To report critical PKI events to an NMS, enable SNMP notifications for PKI. For PKI notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP notifications, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

For the device to generate and output certain types of PKI notifications, enable global PKI notifications and those types of PKI notifications. For the device to stop generating certain types of PKI notifications, disable those types of PKI notifications.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for PKI.

   snmp-agent trap enable pki [**ca-cert-invalid** | **ca-cert-valid** | **crl-http-failure** | **crl-http-success** | **crl-invalid** | **crl-ldap-failure** | **crl-ldap-success** | **crl-scep-failure** | **crl-scep-success** | **crl-valid** | **get-ca-failure** | **get-crl-failure** | **global** | **local-cert-has-expired** | **local-cert-invalid** | **local-cert-ldap-failure**| **local-cert-ldap-success** | **local-cert-nearly-expired** | **local-cert-scep-failure** | **local-cert-scep-success** | **local-cert-valid**] *

   By default, SNMP notifications for PKI are disabled.

Enabling PKI logging

About this task

PKI log messages help administrators easily locate and resolve issues. PKI logging allows the device to record the events and errors on the PKI module and output the corresponding log messages.

These PKI log messages will be sent to the information center. The configuration for the information center determines the output rules and destinations for the PKI log messages. For more information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

Recording PKI logs consumes memory resources. As a best practice to prevent excessive logging, do not enable this feature unless auditing or troubleshooting is needed.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for PKI.
   pki logging enable
   By default, PKI logging is disabled.

Display and maintenance commands for PKI

Execute **display** commands in any view.

Task	Command
Display certificate-based access control policy information.	**display pki certificate access-control-policy** [*policy-name*]
Display certificate attribute group information.	**display pki certificate attribute-group** [*group-name*]
Display the contents of a certificate.	**display pki certificate domain** *domain-name* { **ca** | **local** | **peer** [**serial** *serial-num*] }
Display certificate request status.	**display pki certificate request-status** [**domain** *domain-name*]
Display locally stored CRLs in a PKI domain.	**display pki crl domain** *domain-name*

PKI configuration examples

You can use different software applications, such as Windows server, RSA Keon, and OpenCA, to act as the CA server.

If you use Windows server or OpenCA, you must install the SCEP add-on for Windows server or enable SCEP for OpenCA. In either case, when you configure a PKI domain, you must use the **certificate request from ra** command to specify the RA to accept certificate requests.

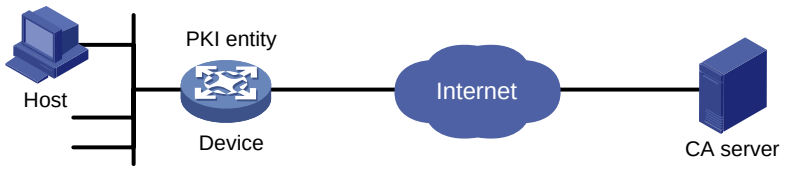
If you use RSA Keon, the SCEP add-on is not required. When you configure a PKI domain, you must use the **certificate request from ca** command to specify the CA to accept certificate requests.

Example: Requesting a certificate from an RSA Keon CA server

Network configuration

Configure the PKI entity (the device) to request a local certificate from the CA server.

Figure 2 Network diagram



Configuring the RSA Keon CA server

1. Create a CA server named **myca**:

In this example, you must configure these basic attributes on the CA server:

- **Nickname**—Name of the trusted CA.
- **Subject DN**—DN attributes of the CA, including the common name (CN), organization unit (OU), organization (O), and country (C).

You can use the default values for other attributes.

2. Configure extended attributes:

   Configure parameters in the **Jurisdiction Configuration** section on the management page of the CA server:

   - Select the correct extension profiles.
   - Enable the SCEP autovetting function to enable the CA server to automatically approve certificate requests without manual intervention.
   - Specify the IP address list for SCEP autovetting.

Configuring the device

1. Synchronize the system time of the device with the CA server for the device to correctly request certificates or obtain CRLs. (Details not shown.)
2. Create an entity named **aaa** and set the common name to **Device**.

```
<Device> system-view
[Device] pki entity aaa
[Device-pki-entity-aaa] common-name Device
[Device-pki-entity-aaa] quit
```

3. Configure a PKI domain:

   # Create a PKI domain named **torsa** and enter its view.

```
[Device] pki domain torsa
```

   # Specify the name of the trusted CA. The setting must be the same as CA name configured on the CA server. This example uses **myca**.

```
[Device-pki-domain-torsa] ca identifier myca
```

   # Configure the URL of the CA server. The URL format is **http://***host:port***/***Issuing Jurisdiction ID*, where *Issuing Jurisdiction ID* is a hexadecimal string generated on the CA server.

```
[Device-pki-domain-torsa] certificate request url
http://1.1.2.22:446/80f6214aa8865301d07929ae481c7ceed99f95bd
```

   # Configure the device to send certificate requests to **ca**.

```
[Device-pki-domain-torsa] certificate request from ca
```

   # Set the PKI entity name to **aaa**.

```
[Device-pki-domain-torsa] certificate request entity aaa
```

   # Specify the URL of the CRL repository.

```
[Device-pki-domain-torsa] crl url ldap://1.1.2.22:389/CN=myca
```

   # Specify a 1024-bit general-purpose RSA key pair named **abc** for certificate request.

```
[Device-pki-domain-torsa] public-key rsa general name abc length 1024
[Device-pki-domain-torsa] quit
```

4. Generate the RSA key pair.

```
[Device] public-key local create rsa name abc
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512,it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
```

```
.........................++++++
.....................................++++++
Create the key pair successfully.
```

5. Request a local certificate:

Obtain the CA certificate and save it locally.

```
[Device] pki retrieve-certificate domain torsa ca
The trusted CA's finger print is:
    MD5  fingerprint:EDE9 0394 A273 B61A F1B3 0072 A0B1 F9AB
    SHA1 fingerprint: 77F9 A077 2FB8 088C 550B A33C 2410 D354 23B2 73A8
Is the finger print correct?(Y/N):y
Retrieved the certificates successfully.
```

Submit a certificate request manually and set the certificate revocation password to **1111**. The certificate revocation password is required when an RSA Keon CA server is used.

```
[Device] pki request-certificate domain torsa password 1111
Start to request general certificate ...
……
Request certificate of domain torsa successfully
```

Verifying the configuration

Display information about the local certificate in PKI domain **torsa**.

```
[Device] display pki certificate domain torsa local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            15:79:75:ec:d2:33:af:5e:46:35:83:bc:bd:6e:e3:b8
        Signature Algorithm: sha1WithRSAEncryption
        Issuer: CN=myca
        Validity
            Not Before: Jan  6 03:10:58 2013 GMT
            Not After : Jan  6 03:10:58 2014 GMT
        Subject: CN=Device
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
                Public-Key: (1024 bit)
                Modulus:
                    00:ab:45:64:a8:6c:10:70:3b:b9:46:34:8d:eb:1a:
                    a1:b3:64:b2:37:27:37:9d:15:bd:1a:69:1d:22:0f:
                    3a:5a:64:0c:8f:93:e5:f0:70:67:dc:cd:c1:6f:7a:
                    0c:b1:57:48:55:81:35:d7:36:d5:3c:37:1f:ce:16:
                    7e:f8:18:30:f6:6b:00:d6:50:48:23:5c:8c:05:30:
                    6f:35:04:37:1a:95:56:96:21:95:85:53:6f:f2:5a:
                    dc:f8:ec:42:4a:6d:5c:c8:43:08:bb:f1:f7:46:d5:
                    f1:9c:22:be:f3:1b:37:73:44:f5:2d:2c:5e:8f:40:
                    3e:36:36:0d:c8:33:90:f3:9b
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 CRL Distribution Points:
```

```
                Full Name:
                  DirName: CN = myca

    Signature Algorithm: sha1WithRSAEncryption
        b0:9d:d9:ac:a0:9b:83:99:bf:9d:0a:ca:12:99:58:60:d8:aa:
        73:54:61:4b:a2:4c:09:bb:9f:f9:70:c7:f8:81:82:f5:6c:af:
        25:64:a5:99:d1:f6:ec:4f:22:e8:6a:96:58:6c:c9:47:46:8c:
        f1:ba:89:b8:af:fa:63:c6:c9:77:10:45:0d:8f:a6:7f:b9:e8:
        25:90:4a:8e:c6:cc:b8:1a:f8:e0:bc:17:e0:6a:11:ae:e7:36:
        87:c4:b0:49:83:1c:79:ce:e2:a3:4b:15:40:dd:fe:e0:35:52:
        ed:6d:83:31:2c:c2:de:7c:e0:a7:92:61:bc:03:ab:40:bd:69:
        1b:f5
```

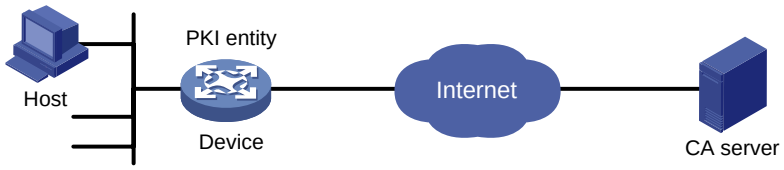
To display detailed information about the CA certificate, use the **display pki certificate domain** command.

Example: Requesting a certificate from a Windows Server 2003 CA server

Network configuration

Configure the PKI entity (the device) to request a local certificate from a Windows Server 2003 CA server.

Figure 3 Network diagram



Configuring the Windows Server 2003 CA server

1. Install the certificate service component:
   a. Select **Control Panel** > **Add or Remove Programs** from the start menu.
   b. Select **Add/Remove Windows Components** > **Certificate Services**.
   c. Click **Next** to begin the installation.
   d. Set the CA name. In this example, set the CA name to **myca**.
2. Install the SCEP add-on:

   By default, Windows Server 2003 does not support SCEP. You must install the SCEP add-on on the server for a PKI entity to register and obtain a certificate from the server. After the SCEP add-on installation is complete, you will see a URL. Specify this URL as the certificate request URL on the device.
3. Modify the certificate service attributes:
   a. Select **Control Panel** > **Administrative Tools** > **Certificate Authority** from the start menu.

      If the certificate service component and SCEP add-on have been installed successfully, there should be two certificates issued by the CA to the RA.
   b. Right-click the CA server in the navigation tree and select **Properties** > **Policy Module**.
   c. Click **Properties**, and then select **Follow the settings in the certificate template, if applicable. Otherwise, automatically issue the certificate**.
4. Modify the Internet information services attributes:

a. Select **Control Panel** > **Administrative Tools** > **Internet Information Services (IIS) Manager** from the start menu.
b. Select **Web Sites** from the navigation tree.
c. Right-click **Default Web Site** and select **Properties** > **Home Directory**.
d. Specify the path for certificate service in the **Local path** field.
e. Specify a unique TCP port number for the default website to avoid conflict with existing services. This example uses port 8080.

Configuring the device

1. Synchronize the device's system time with the CA server for the device to correctly request certificates. (Details not shown.)
2. Create an entity named **aaa** and set the common name to **test**.

```
<Device> system-view
[Device] pki entity aaa
[Device-pki-entity-aaa] common-name test
[Device-pki-entity-aaa] quit
```

3. Configure a PKI domain:

Create a PKI domain named **winserver** and enter its view.

```
[Device] pki domain winserver
```

Set the name of the trusted CA to **myca**.

```
[Device-pki-domain-winserver] ca identifier myca
```

Configure the certificate request URL. The URL format is **http://***host:port***/certsrv/mscep/mscep.dll**, where *host:port* is the host IP address and port number of the CA server.

```
[Device-pki-domain-winserver] certificate request url
http://4.4.4.1:8080/certsrv/mscep/mscep.dll
```

Configure the device to send certificate requests to **ra**.

```
[Device-pki-domain-winserver] certificate request from ra
```

Set the PKI entity name to **aaa**.

```
[Device-pki-domain-winserver] certificate request entity aaa
```

Configure a 1024-bit general-purpose RSA key pair named **abc** for certificate request.

```
[Device-pki-domain-winserver] public-key rsa general name abc length 1024
[Device-pki-domain-winserver] quit
```

4. Generate RSA key pair **abc**.

```
[Device] public-key local create rsa name abc
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512,it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.........................++++++
.....................................++++++
Create the key pair successfully.
```

5. Request a local certificate:

Obtain the CA certificate and save it locally.

```
[Device] pki retrieve-certificate domain winserver ca
The trusted CA's finger print is:
    MD5  fingerprint:766C D2C8 9E46 845B 4DCE 439C 1C1F 83AB
```

```
    SHA1 fingerprint:97E5 DDED AB39 3141 75FB DB5C E7F8 D7D7 7C9B 97B4
Is the finger print correct?(Y/N):y
Retrieved the certificates successfully.
```

Submit a certificate request manually.

```
[Device] pki request-certificate domain winserver
Start to request general certificate ...
…
Request certificate of domain winserver successfully
```

Verifying the configuration

Display information about the local certificate in PKI domain **winserver**.

```
[Device] display pki certificate domain winserver local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
             (Negative)01:03:99:ff:ff:ff:ff:fd:11
        Signature Algorithm: sha1WithRSAEncryption
        Issuer: CN=sec
        Validity
            Not Before: Dec 24 07:09:42 2012 GMT
            Not After : Dec 24 07:19:42 2013 GMT
        Subject: CN=test
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
                Public-Key: (2048 bit)
                Modulus:
                    00:c3:b5:23:a0:2d:46:0b:68:2f:71:d2:14:e1:5a:
                    55:6e:c5:5e:26:86:c1:5a:d6:24:68:02:bf:29:ac:
                    dc:31:41:3f:5d:5b:36:9e:53:dc:3a:bc:0d:11:fb:
                    d6:7d:4f:94:3c:c1:90:4a:50:ce:db:54:e0:b3:27:
                    a9:6a:8e:97:fb:20:c7:44:70:8f:f0:b9:ca:5b:94:
                    f0:56:a5:2b:87:ac:80:c5:cc:04:07:65:02:39:fc:
                    db:61:f7:07:c6:65:4c:e4:5c:57:30:35:b4:2e:ed:
                    9c:ca:0b:c1:5e:8d:2e:91:89:2f:11:e3:1e:12:8a:
                    f8:dd:f8:a7:2a:94:58:d9:c7:f8:1a:78:bd:f5:42:
                    51:3b:31:5d:ac:3e:c3:af:fa:33:2c:fc:c2:ed:b9:
                    ee:60:83:b3:d3:e5:8e:e5:02:cf:b0:c8:f0:3a:a4:
                    b7:ac:a0:2c:4d:47:5f:39:4b:2c:87:f2:ee:ea:d0:
                    c3:d0:8e:2c:80:83:6f:39:86:92:98:1f:d2:56:3b:
                    d7:94:d2:22:f4:df:e3:f8:d1:b8:92:27:9c:50:57:
                    f3:a1:18:8b:1c:41:ba:db:69:07:52:c1:9a:3d:b1:
                    2d:78:ab:e3:97:47:e2:70:14:30:88:af:f8:8e:cb:
                    68:f9:6f:07:6e:34:b6:38:6a:a2:a8:29:47:91:0e:
                    25:39
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 Key Usage:
                Digital Signature, Non Repudiation, Key Encipherment, Data Encipherment
```

```
            X509v3 Subject Key Identifier:
                C9:BB:D5:8B:02:1D:20:5B:40:94:15:EC:9C:16:E8:9D:6D:FD:9F:34
            X509v3 Authority Key Identifier:
                keyid:32:F1:40:BA:9E:F1:09:81:BD:A8:49:66:FF:F8:AB:99:4A:30:21:9B

            X509v3 CRL Distribution Points:

                Full Name:
                  URI:file://\\g07904c\CertEnroll\sec.crl

            Authority Information Access:
                CA Issuers - URI:http://gc/CertEnroll/gc_sec.crt
                CA Issuers - URI:file://\\gc\CertEnroll\gc_sec.crt

            1.3.6.1.4.1.311.20.2:
                .0.I.P.S.E.C.I.n.t.e.r.m.e.d.i.a.t.e.O.f.f.l.i.n.e
    Signature Algorithm: sha1WithRSAEncryption
        76:f0:6c:2c:4d:bc:22:59:a7:39:88:0b:5c:50:2e:7a:5c:9d:
        6c:28:3c:c0:32:07:5a:9c:4c:b6:31:32:62:a9:45:51:d5:f5:
        36:8f:47:3d:47:ae:74:6c:54:92:f2:54:9f:1a:80:8a:3f:b2:
        14:47:fa:dc:1e:4d:03:d5:d3:f5:9d:ad:9b:8d:03:7f:be:1e:
        29:28:87:f7:ad:88:1c:8f:98:41:9a:db:59:ba:0a:eb:33:ec:
        cf:aa:9b:fc:0f:69:3a:70:f2:fa:73:ab:c1:3e:4d:12:fb:99:
        31:51:ab:c2:84:c0:2f:e5:f6:a7:c3:20:3c:9a:b0:ce:5a:bc:
        0f:d9:34:56:bc:1e:6f:ee:11:3f:7c:b2:52:f9:45:77:52:fb:
        46:8a:ca:b7:9d:02:0d:4e:c3:19:8f:81:46:4e:03:1f:58:03:
        bf:53:c6:c4:85:95:fb:32:70:e6:1b:f3:e4:10:ed:7f:93:27:
        90:6b:30:e7:81:36:bb:e2:ec:f2:dd:2b:bb:b9:03:1c:54:0a:
        00:3f:14:88:de:b8:92:63:1e:f5:b3:c2:cf:0a:d5:f4:80:47:
        6f:fa:7e:2d:e3:a7:38:46:f6:9e:c7:57:9d:7f:82:c7:46:06:
        7d:7c:39:c4:94:41:bd:9e:5c:97:86:c8:48:de:35:1e:80:14:
        02:09:ad:08
```

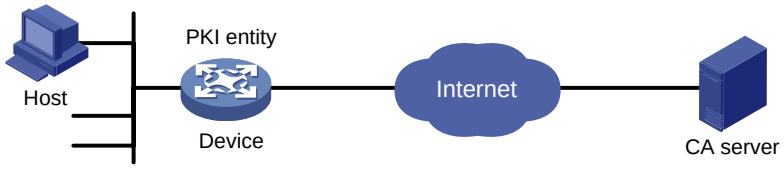
To display detailed information about the CA certificate, use the **display pki certificate domain** command.

Example: Requesting a certificate from an OpenCA server

Network configuration

Configure the PKI entity (the device) to request a local certificate from the CA server.

Figure 4 Network diagram



Configuring the OpenCA server

Configure the OpenCA server as instructed in related manuals. (Details not shown.)

Make sure the version of the OpenCA server is later than version 0.9.2 because the earlier versions do not support SCEP.

Configuring the device

1. Synchronize the device's system time with the CA server for the device to correctly request certificates. (Details not shown.)
2. Create a PKI entity named **aaa** and configure the common name, country code, organization name, and OU for the entity.

```
<Device> system-view
[Device] pki entity aaa
[Device-pki-entity-aaa] common-name rnd
[Device-pki-entity-aaa] country CN
[Device-pki-entity-aaa] organization test
[Device-pki-entity-aaa] organization-unit software
[Device-pki-entity-aaa] quit
```

3. Configure a PKI domain:

   # Create a PKI domain named **openca** and enter its view.

```
[Device] pki domain openca
```

   # Set the name of the trusted CA to **myca**.

```
[Device-pki-domain-openca] ca identifier myca
```

   # Configure the certificate request URL. The URL is in the format **http://***host***/cgi-bin/pki/scep**, where *host* is the host IP address of the OpenCA server.

```
[Device-pki-domain-openca] certificate request url
http://192.168.222.218/cgi-bin/pki/scep
```

   # Configure the device to send certificate requests to the RA.

```
[Device-pki-domain-openca] certificate request from ra
```

   # Specify PKI entity **aaa** for certificate request.

```
[Device-pki-domain-openca] certificate request entity aaa
```

   # Configure a 1024-bit general-purpose RSA key pair named **abc** for certificate request.

```
[Device-pki-domain-openca] public-key rsa general name abc length 1024
[Device-pki-domain-openca] quit
```

4. Generate RSA key pair **abc**.

```
[Device] public-key local create rsa name abc
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512,it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.........................++++++
.....................................++++++
Create the key pair successfully.
```

5. Request a local certificate:

   # Obtain the CA certificate and save it locally.

```
[Device] pki retrieve-certificate domain openca ca
The trusted CA's finger print is:
    MD5  fingerprint:5AA3 DEFD 7B23 2A25 16A3 14F4 C81C C0FA
```

```
    SHA1 fingerprint:9668 4E63 D742 4B09 90E0 4C78 E213 F15F DC8E 9122
Is the finger print correct?(Y/N):y
Retrieved the certificates successfully.
```

Submit a certificate request manually.

```
[Device] pki request-certificate domain openca
Start to request general certificate ...
…
Request certificate of domain openca successfully
```

Verifying the configuration

Display information about the local certificate in PKI domain **openca**.

```
[Device] display pki certificate domain openca local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            21:1d:b8:d2:e4:a9:21:28:e4:de
        Signature Algorithm: sha256WithRSAEncryption
        Issuer: C=CN, L=shangdi, ST=pukras, O=OpenCA Labs, OU=mysubUnit, CN=sub-ca,
DC=pki-subdomain, DC=mydomain-sub, DC=com
        Validity
            Not Before: Jun 30 09:09:09 2011 GMT
            Not After : May  1 09:09:09 2012 GMT
        Subject: CN=rnd, O=test, OU=software, C=CN
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
                Public-Key: (1024 bit)
                Modulus:
                    00:b8:7a:9a:b8:59:eb:fc:70:3e:bf:19:54:0c:7e:
                    c3:90:a5:d3:fd:ee:ff:c6:28:c6:32:fb:04:6e:9c:
                    d6:5a:4f:aa:bb:50:c4:10:5c:eb:97:1d:a7:9e:7d:
                    53:d5:31:ff:99:ab:b6:41:f7:6d:71:61:58:97:84:
                    37:98:c7:7c:79:02:ac:a6:85:f3:21:4d:3c:8e:63:
                    8d:f8:71:7d:28:a1:15:23:99:ed:f9:a1:c3:be:74:
                    0d:f7:64:cf:0a:dd:39:49:d7:3f:25:35:18:f4:1c:
                    59:46:2b:ec:0d:21:1d:00:05:8a:bf:ee:ac:61:03:
                    6c:1f:35:b5:b4:cd:86:9f:45
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Cert Type:
                SSL Client, S/MIME
            X509v3 Key Usage:
                Digital Signature, Non Repudiation, Key Encipherment
            X509v3 Extended Key Usage:
                TLS Web Client Authentication, E-mail Protection, Microsoft
Smartcardlogin
Netscape Comment:
```

```
                User Certificate of OpenCA Labs
            X509v3 Subject Key Identifier:
                24:71:C9:B8:AD:E1:FE:54:9A:EA:E9:14:1B:CD:D9:45:F4:B2:7A:1B
            X509v3 Authority Key Identifier:
                keyid:85:EB:D5:F7:C9:97:2F:4B:7A:6D:DD:1B:4D:DD:00:EE:53:CF:FD:5B

            X509v3 Issuer Alternative Name:
                DNS:root@docm.com, DNS:, IP Address:192.168.154.145, IP
Address:192.168.154.138
            Authority Information Access:
                CA Issuers - URI:http://192.168.222.218/pki/pub/cacert/cacert.crt
                OCSP - URI:http://192.168.222.218:2560/
                1.3.6.1.5.5.7.48.12 - URI:http://192.168.222.218:830/

            X509v3 CRL Distribution Points:

                Full Name:
                  URI:http://192.168.222.218/pki/pub/crl/cacrl.crl

    Signature Algorithm: sha256WithRSAEncryption
        5c:4c:ba:d0:a1:35:79:e6:e5:98:69:91:f6:66:2a:4f:7f:8b:
        0e:80:de:79:45:b9:d9:12:5e:13:28:17:36:42:d5:ae:fc:4e:
        ba:b9:61:f1:0a:76:42:e7:a6:34:43:3e:2d:02:5e:c7:32:f7:
        6b:64:bb:2d:f5:10:6c:68:4d:e7:69:f7:47:25:f5:dc:97:af:
        ae:33:40:44:f3:ab:e4:5a:a0:06:8f:af:22:a9:05:74:43:b6:
        e4:96:a5:d4:52:32:c2:a8:53:37:58:c7:2f:75:cf:3e:8e:ed:
        46:c9:5a:24:b1:f5:51:1d:0f:5a:07:e6:15:7a:02:31:05:8c:
        03:72:52:7c:ff:28:37:1e:7e:14:97:80:0b:4e:b9:51:2d:50:
        98:f2:e4:5a:60:be:25:06:f6:ea:7c:aa:df:7b:8d:59:79:57:
        8f:d4:3e:4f:51:c1:34:e6:c1:1e:71:b5:0d:85:86:a5:ed:63:
        1e:08:7f:d2:50:ac:a0:a3:9e:88:48:10:0b:4a:7d:ed:c1:03:
        9f:87:97:a3:5e:7d:75:1d:ac:7b:6f:bb:43:4d:12:17:9a:76:
        b0:bf:2f:6a:cc:4b:cd:3d:a1:dd:e0:dc:5a:f3:7c:fb:c3:29:
        b0:12:49:5c:12:4c:51:6e:62:43:8b:73:b9:26:2a:f9:3d:a4:
        81:99:31:89
```

To display detailed information about the CA certificate, use the **display pki certificate domain** command.

Example: Configuring IKE negotiation with RSA digital signature from a Windows Server 2003 CA server

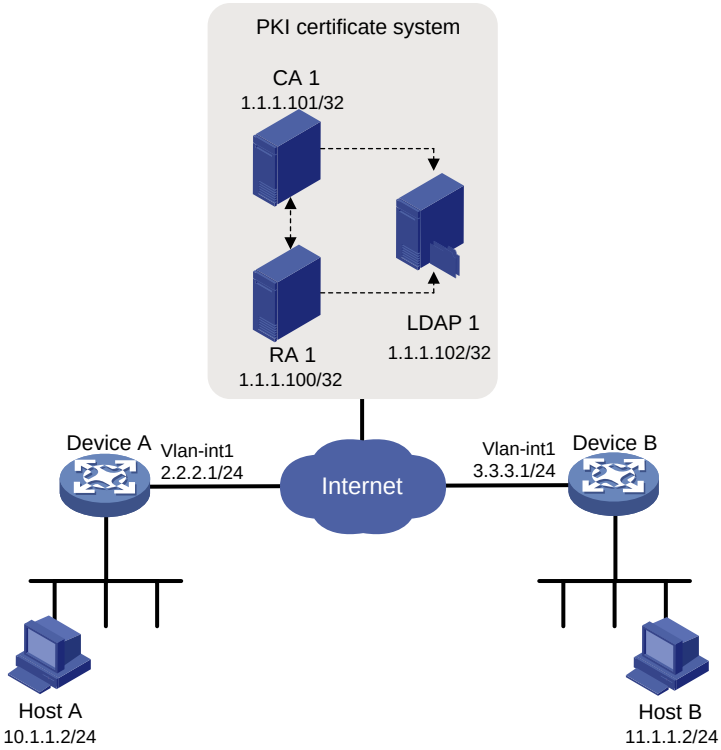
Network configuration

As shown in the network diagram, an IPsec tunnel is required to be established between Device A and Device B. The IPsec tunnel protects the traffic between Host A on subnet 10.1.1.0/24 and Host B on subnet 1.1.1.0/24.

Device A and Device B use IKE to set up SAs, and the IKE proposal uses RSA digital signature for identity authentication.

Device A and Device B use the same CA.

Figure 5 Network diagram



Configuring the Windows Server 2003 CA server

See "Example: Requesting a certificate from a Windows Server 2003 CA server."

Configuring Device A

Configure a PKI entity.

```
<DeviceA> system-view
[DeviceA] pki entity en
[DeviceA-pki-entity-en] ip 2.2.2.1
[DeviceA-pki-entity-en] common-name devicea
[DeviceA-pki-entity-en] quit
```

Configure a PKI domain.

```
[DeviceA] pki domain 1
[DeviceA-pki-domain-1] ca identifier CA1
[DeviceA-pki-domain-1] certificate request url http://1.1.1.100/certsrv/mscep/mscep.dll
[DeviceA-pki-domain-1] certificate request entity en
[DeviceA-pki-domain-1] ldap-server host 1.1.1.102
```

Configure the device to send certificate requests to **ra**.

```
[DeviceA-pki-domain-1] certificate request from ra
```

Configure a 1024-bit general-purpose RSA key pair named **abc** for certificate request.

```
[DeviceA-pki-domain-1] public-key rsa general name abc length 1024
[DeviceA-pki-domain-1] quit
```

Generate the RSA key pair.

```
[DeviceA] public-key local create rsa name abc
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512,it will take a few minutes.
```

```
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.........................++++++
.....................................++++++
Create the key pair successfully.
```

Obtain the CA certificate and save it locally.

```
[DeviceA] pki retrieve-certificate domain 1 ca
```

Submit a certificate request manually.

```
[DeviceA] pki request-certificate domain 1
```

Create IKE proposal 1, and configure the authentication method as RSA digital signature.

```
[DeviceA] ike proposal 1
[DeviceA-ike-proposal-1] authentication-method rsa-signature
[DeviceA-ike-proposal-1] quit
```

Specify the PKI domain used in IKE negotiation for IKE profile **peer**.

```
[DeviceA] ike profile peer
[DeviceA-ike-profile-peer] certificate domain 1
```

Configuring Device B

Configure a PKI entity.

```
<DeviceB> system-view
[DeviceB] pki entity en
[DeviceB-pki-entity-en] ip 3.3.3.1
[DeviceB-pki-entity-en] common-name deviceb
[DeviceB-pki-entity-en] quit
```

Configure a PKI domain.

```
[DeviceB] pki domain 1
[DeviceB-pki-domain-1] ca identifier CA1
[DeviceB-pki-domain-1] certificate request url http://1.1.1.100/certsrv/mscep/mscep.dll
[DeviceB-pki-domain-1] certificate request entity en
[DeviceB-pki-domain-1] ldap-server host 1.1.1.102
```

Configure the device to send certificate requests to **ra**.

```
[DeviceB-pki-domain-1] certificate request from ra
```

Configure a 1024-bit general-purpose RSA key pair named **abc** for certificate request.

```
[DeviceB-pki-domain-1] public-key rsa general name abc length 1024
[DeviceB-pki-domain-1] quit
```

Generate the RSA key pair.

```
[DeviceB] public-key local create rsa name abc
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512,it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.........................++++++
.....................................++++++
Create the key pair successfully.
```

Obtain the CA certificate and save it locally.

```
[DeviceB] pki retrieve-certificate domain 1 ca
The trusted CA's finger print is:
    MD5  fingerprint:5C41 E657 A0D6 ECB4 6BD6 1823 7473 AABC
    SHA1 fingerprint:1616 E7A5 D89A 2A99 9419 1C12 D696 8228 87BC C266
Is the finger print correct?(Y/N):y
Retrieved the certificates successfully.
```

Submit a certificate request manually.

```
[DeviceB] pki request-certificate domain 1
Start to request general certificate ...
...
Certificate requested successfully.
```

Create IKE proposal 1, and configure the authentication method as RSA digital signature.

```
[DeviceB] ike proposal 1
[DeviceB-ike-proposal-1] authentication-method rsa-signature
[DeviceB-ike-proposal-1] quit
```

Specify the PKI domain used in IKE negotiation for IKE profile **peer**.

```
[DeviceB] ike profile peer
[DeviceB-ike-profile-peer] certificate domain 1
```

The configurations are for IKE negotiation with RSA digital signature. For information about how to configure IPsec SAs to be set up, see "Configuring IPsec."

Example: Importing and exporting certificates

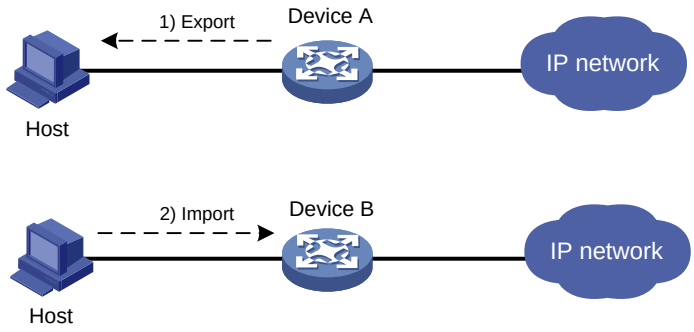
Network configuration

As shown in Figure 6, Device B will replace Device A in the network. PKI domain **exportdomain** on Device A has two local certificates containing the private key and one CA certificate. To make sure the certificates are still valid after Device B replaces Device A, copy the certificates on Device A to Device B as follows:

1. Export the certificates in PKI domain **exportdomain** on Device A to .pem certificate files.

   During the export, encrypt the private key in the local certificates using 3DES_CBC with the password 11111.
2. Transfer the certificate files from Device A to Device B through the FTP host.
3. Import the certificate files to PKI domain **importdomain** on Device B.

Figure 6 Network diagram



Procedure

1. Export the certificates on Device A:

Export the CA certificate to a .pem file.

```
<DeviceA> system-view
[DeviceA] pki export domain exportdomain pem ca filename pkicachain.pem
```

Export the local certificate to a file named **pkilocal.pem** in PEM format, and use 3DES_CBC to encrypt the private key with the password **111111**.

```
[DeviceA] pki export domain exportdomain pem local 3des-cbc 111111 filename
pkilocal.pem
```

Now, Device A has three certificate files in PEM format:

- A CA certificate file named **pkicachain.pem**.
- A local certificate file named **pkilocal.pem-signature**, which contains the private key for signature.
- A local certificate file named **pkilocal.pem-encryption**, which contains the private key for encryption.

Display local certificate file **pkilocal.pem-signature**.

```
[DeviceA] quit
<DeviceA> more pkilocal.pem-signature
Bag Attributes
    friendlyName:
    localKeyID: 90 C6 DC 1D 20 49 4F 24 70 F5 17 17 20 2B 9E AC 20 F3 99 89
subject=/C=CN/O=OpenCA Labs/OU=Users/CN=subsign 11
issuer=/C=CN/L=shangdi/ST=pukras/O=OpenCA Labs/OU=docm/CN=subca1
-----BEGIN CERTIFICATE-----
MIIEgjCCA2qgAwIBAgILAJgsebpejZc5UwAwDQYJKoZIhvcNAQELBQAwZjELMAkG
…
-----END CERTIFICATE-----
Bag Attributes
    friendlyName:
    localKeyID: 90 C6 DC 1D 20 49 4F 24 70 F5 17 17 20 2B 9E AC 20 F3 99 89
Key Attributes: <No Attributes>
-----BEGIN ENCRYPTED PRIVATE KEY-----
MIICxjBABgkqhkiG9w0BBQ0wMzAbBgkqhkiG9w0BBQwwDgQIZtjSjfslJCoCAggA
…
-----END ENCRYPTED PRIVATE KEY-----
```

Display local certificate file **pkilocal.pem-encryption**.

```
<DeviceA> more pkilocal.pem-encryption
Bag Attributes
    friendlyName:
    localKeyID: D5 DF 29 28 C8 B9 D9 49 6C B5 44 4B C2 BC 66 75 FE D6 6C C8
subject=/C=CN/O=OpenCA Labs/OU=Users/CN=subencr 11
issuer=/C=CN/L=shangdi/ST=pukras/O=OpenCA Labs/OU=docm/CN=subca1
-----BEGIN CERTIFICATE-----
MIIEUDCCAzigAwIBAgIKCHxnAVyzWhIPLzANBgkqhkiG9w0BAQsFADBmMQswCQYD
…
-----END CERTIFICATE-----
Bag Attributes
    friendlyName:
    localKeyID: D5 DF 29 28 C8 B9 D9 49 6C B5 44 4B C2 BC 66 75 FE D6 6C C8
Key Attributes: <No Attributes>
```

```
-----BEGIN ENCRYPTED PRIVATE KEY-----
MIICxjBABgkqhkiG9w0BBQ0wMzAbBgkqhkiG9w0BBQwwDgQI7H0mb4O7/GACAggA
…
-----END ENCRYPTED PRIVATE KEY-----
```

2. Download certificate files **pkicachain.pem**, **pkilocal.pem-signature**, and **pkilocal.pem-encryption** from Device A to the host through FTP. (Details not shown.)
3. Upload certificate files **pkicachain.pem**, **pkilocal.pem-signature**, and **pkilocal.pem-encryption** from the host to Device B through FTP. (Details not shown.)
4. Import the certificate files to Device B:

   # Disable CRL checking. (You can configure CRL checking as required. This example assumes CRL checking is not required.)

```
<DeviceB> system-view
[DeviceB] pki domain importdomain
[DeviceB-pki-domain-importdomain] undo crl check enable
```

   # Specify RSA key pair **sign** for signature and RSA key pair **encr** for encryption.

```
[DeviceB-pki-domain-importdomain] public-key rsa signature name sign encryption name encr
[DeviceB-pki-domain-importdomain] quit
```

   # Import CA certificate file **pkicachain.pem** in PEM format to the PKI domain.

```
[DeviceB] pki import domain importdomain pem ca filename pkicachain.pem
```

   # Import local certificate file **pkilocal.pem-signature** in PEM format to the PKI domain. The certificate file contains a key pair.

```
[DeviceB] pki import domain importdomain pem local filename pkilocal.pem-signature
Please input the password:
```

   # Import local certificate file **pkilocal.pem-encryption** in PEM format to the PKI domain. The certificate file contains a key pair.

```
[DeviceB] pki import domain importdomain pem local filename pkilocal.pem-encryption
Please input the password:
```

   # Display the imported local certificate information on Device B.

```
[DeviceB] display pki certificate domain importdomain local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            98:2c:79:ba:5e:8d:97:39:53:00
        Signature Algorithm: sha256WithRSAEncryption
        Issuer: C=CN, L=shangdi, ST=pukras, O=OpenCA Labs, OU=docm, CN=subca1
        Validity
            Not Before: May 26 05:56:49 2011 GMT
            Not After : Nov 22 05:56:49 2012 GMT
        Subject: C=CN, O=OpenCA Labs, OU=Users, CN=subsign 11
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
                Public-Key: (1024 bit)
                Modulus:
                    00:9f:6e:2f:f6:cb:3d:08:19:9a:4a:ac:b4:ac:63:
                    ce:8d:6a:4c:3a:30:19:3c:14:ff:a9:50:04:f5:00:
                    ee:a3:aa:03:cb:b3:49:c4:f8:ae:55:ee:43:93:69:
                    6c:bf:0d:8c:f4:4e:ca:69:e5:3f:37:5c:83:ea:83:
```

```
                    ad:16:b8:99:37:cb:86:10:6b:a0:4d:03:95:06:42:
                    ef:ef:0d:4e:53:08:0a:c9:29:dd:94:28:02:6e:e2:
                    9b:87:c1:38:2d:a4:90:a2:13:5f:a4:e3:24:d3:2c:
                    bf:98:db:a7:c2:36:e2:86:90:55:c7:8c:c5:ea:12:
                    01:31:69:bf:e3:91:71:ec:21
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Cert Type:
                SSL Client, S/MIME
            X509v3 Key Usage:
                Digital Signature, Non Repudiation
            X509v3 Extended Key Usage:
                TLS Web Client Authentication, E-mail Protection, Microsoft
Smartcardlogin
            Netscape Comment:
                User Certificate of OpenCA Labs
            X509v3 Subject Key Identifier:
                AA:45:54:29:5A:50:2B:89:AB:06:E5:BD:0D:07:8C:D9:79:35:B1:F5
            X509v3 Authority Key Identifier:
                keyid:70:54:40:61:71:31:02:06:8C:62:11:0A:CC:A5:DB:0E:7E:74:DE:DD

            X509v3 Subject Alternative Name:
                email:subsign@docm.com
            X509v3 Issuer Alternative Name:
                DNS:subca1@docm.com, DNS:, IP Address:1.1.2.2, IP Address:2.2.1.1
            Authority Information Access:
                CA Issuers - URI:http://titan/pki/pub/cacert/cacert.crt
                OCSP - URI:http://titan:2560/
                1.3.6.1.5.5.7.48.12 - URI:http://titan:830/

            X509v3 CRL Distribution Points:

                Full Name:
                  URI:http://192.168.40.130/pki/pub/crl/cacrl.crl

    Signature Algorithm: sha256WithRSAEncryption
        18:e7:39:9a:ad:84:64:7b:a3:85:62:49:e5:c9:12:56:a6:d2:
        46:91:53:8e:84:ba:4a:0a:6f:28:b9:43:bc:e7:b0:ca:9e:d4:
        1f:d2:6f:48:c4:b9:ba:c5:69:4d:90:f3:15:c4:4e:4b:1e:ef:
        2b:1b:2d:cb:47:1e:60:a9:0f:81:dc:f2:65:6b:5f:7a:e2:36:
        29:5d:d4:52:32:ef:87:50:7c:9f:30:4a:83:de:98:8b:6a:c9:
        3e:9d:54:ee:61:a4:26:f3:9a:40:8f:a6:6b:2b:06:53:df:b6:
        5f:67:5e:34:c8:c3:b5:9b:30:ee:01:b5:a9:51:f9:b1:29:37:
        02:1a:05:02:e7:cc:1c:fe:73:d3:3e:fa:7e:91:63:da:1d:f1:
        db:28:6b:6c:94:84:ad:fc:63:1b:ba:53:af:b3:5d:eb:08:b3:
        5b:d7:22:3a:86:c3:97:ef:ac:25:eb:4a:60:f8:2b:a3:3b:da:
        5d:6f:a5:cf:cb:5a:0b:c5:2b:45:b7:3e:6e:39:e9:d9:66:6d:
```

```
        ef:d3:a0:f6:2a:2d:86:a3:01:c4:94:09:c0:99:ce:22:19:84:
        2b:f0:db:3e:1e:18:fb:df:56:cb:6f:a2:56:35:0d:39:94:34:
        6d:19:1d:46:d7:bf:1a:86:22:78:87:3e:67:fe:4b:ed:37:3d:
        d6:0a:1c:0b

Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            08:7c:67:01:5c:b3:5a:12:0f:2f
        Signature Algorithm: sha256WithRSAEncryption
        Issuer: C=CN, L=shangdi, ST=pukras, O=OpenCA Labs, OU=docm, CN=subca1
        Validity
            Not Before: May 26 05:58:26 2011 GMT
            Not After : Nov 22 05:58:26 2012 GMT
        Subject: C=CN, O=OpenCA Labs, OU=Users, CN=subencr 11
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
                Public-Key: (1024 bit)
                Modulus:
                    00:db:26:13:d3:d1:a4:af:11:f3:6d:37:cf:d0:d4:
                    48:50:4e:0f:7d:54:76:ed:50:28:c6:71:d4:48:ae:
                    4d:e7:3d:23:78:70:63:18:33:f6:94:98:aa:fa:f6:
                    62:ed:8a:50:c6:fd:2e:f4:20:0c:14:f7:54:88:36:
                    2f:e6:e2:88:3f:c2:88:1d:bf:8d:9f:45:6c:5a:f5:
                    94:71:f3:10:e9:ec:81:00:28:60:a9:02:bb:35:8b:
                    bf:85:75:6f:24:ab:26:de:47:6c:ba:1d:ee:0d:35:
                    75:58:10:e5:e8:55:d1:43:ae:85:f8:ff:75:81:03:
                    8c:2e:00:d1:e9:a4:5b:18:39
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Cert Type:
                SSL Server
            X509v3 Key Usage:
                Key Encipherment, Data Encipherment
            Netscape Comment:
                VPN Server of OpenCA Labs
            X509v3 Subject Key Identifier:
                CC:96:03:2F:FC:74:74:45:61:38:1F:48:C0:E8:AA:18:24:F0:2B:AB
            X509v3 Authority Key Identifier:
                keyid:70:54:40:61:71:31:02:06:8C:62:11:0A:CC:A5:DB:0E:7E:74:DE:DD

            X509v3 Subject Alternative Name:
                email:subencr@docm.com
            X509v3 Issuer Alternative Name:
                DNS:subca1@docm.com, DNS:, IP Address:1.1.2.2, IP Address:2.2.1.1
```

```
            Authority Information Access:
                CA Issuers - URI:http://titan/pki/pub/cacert/cacert.crt
                OCSP - URI:http://titan:2560/
                1.3.6.1.5.5.7.48.12 - URI:http://titan:830/

            X509v3 CRL Distribution Points:

                Full Name:
                  URI:http://192.168.40.130/pki/pub/crl/cacrl.crl

    Signature Algorithm: sha256WithRSAEncryption
        53:69:66:5f:93:f0:2f:8c:54:24:8f:a2:f2:f1:29:fa:15:16:
        90:71:e2:98:e3:5c:c6:e3:d4:5f:7a:f6:a9:4f:a2:7f:ca:af:
        c4:c8:c7:2c:c0:51:0a:45:d4:56:e2:81:30:41:be:9f:67:a1:
        23:a6:09:50:99:a1:40:5f:44:6f:be:ff:00:67:9d:64:98:fb:
        72:77:9e:fd:f2:4c:3a:b2:43:d8:50:5c:48:08:e7:77:df:fb:
        25:9f:4a:ea:de:37:1e:fb:bc:42:12:0a:98:11:f2:d9:5b:60:
        bc:59:72:04:48:59:cc:50:39:a5:40:12:ff:9d:d0:69:3a:5e:
        3a:09:5a:79:e0:54:67:a0:32:df:bf:72:a0:74:63:f9:05:6f:
        5e:28:d2:e8:65:49:e6:c7:b5:48:7d:95:47:46:c1:61:5a:29:
        90:65:45:4a:88:96:e4:88:bd:59:25:44:3f:61:c6:b1:08:5b:
        86:d2:4f:61:4c:20:38:1c:f4:a1:0b:ea:65:87:7d:1c:22:be:
        b6:17:17:8a:5a:0f:35:4c:b8:b3:73:03:03:63:b1:fc:c4:f5:
        e9:6e:7c:11:e8:17:5a:fb:39:e7:33:93:5b:2b:54:72:57:72:
        5e:78:d6:97:ef:b8:d8:6d:0c:05:28:ea:81:3a:06:a0:2e:c3:
        79:05:cd:c3
```

To display detailed information about the CA certificate, use the **`display pki certificate domain`** command.

Troubleshooting PKI configuration

This section provides troubleshooting information for common problems with PKI.

Failed to obtain the CA certificate

Symptom

The CA certificate cannot be obtained.

Analysis

- The network connection is down, for example, because the network cable is damaged or the connectors have bad contact.
- No trusted CA is specified.
- The certificate request URL is incorrect or not specified.
- The system time of the device is not synchronized with the CA server.
- The CA server does not accept the source IP address specified in the PKI domain, or no source IP address is specified.
- The fingerprint of the root CA certificate is illegal.

Solution

1. Fix the network connection problems, if any.
2. Configure the trusted CA and all other required parameters in the PKI domain.
3. Use the **ping** command to verify that the CA server is reachable.
4. Synchronize the system time of the device with the CA server.
5. Specify the correct source IP address that the CA server can accept. For the correct settings, contact the CA administrator.
6. Verify the fingerprint of the CA certificate on the CA server.
7. If the problem persists, contact H3C Support.

Failed to obtain local certificates

Symptom

The local certificates can be obtained.

Analysis

- The network connection is down.
- The PKI domain does not have a CA certificate before you submit the local certificate request.
- The LDAP server is not configured or is incorrectly configured.
- No key pair is specified for certificate request in the PKI domain, or the specified key pair does not match the one contained in the local certificates to the obtained.
- No PKI entity is configured in the PKI domain, or the PKI entity configuration is incorrect.
- CRL checking is enabled, but the PKI domain does not have a CRL and cannot obtain one.
- The CA server does not accept the source IP address specified in the PKI domain, or no source IP address is specified.
- The system time of the device is not synchronized with the CA server.

Solution

1. Fix the network connection problems, if any..
2. Obtain or import the CA certificate.
3. Configure the correct LDAP server parameters.
4. Specify the key pair for certificate request, or remove the existing key pair, specify a new key pair, and submit a local certificate request again.
5. Check the registration policy on the CA or RA, and make sure the attributes of the PKI entity meet the policy requirements.
6. Obtain the CRL from the CRL repository.
7. Specify the correct source IP address that the CA server can accept. For the correct settings, contact the CA administrator.
8. Synchronize the system time of the device with the CA server.
9. If the problem persists, contact H3C Support.

Failed to request local certificates

Symptom

Local certificate requests cannot be submitted.

Analysis

- The network connection is down, for example, because the network cable is damaged or the connectors have bad contact.
- The PKI domain does not have a CA certificate before the local certificate request is submitted.
- The certificate request URL is incorrect or is not specified.
- The certificate request reception authority is incorrect or is not specified.
- Required PKI entity parameters are not configured or are incorrectly configured.
- No key pair is specified in the PKI domain for certificate request, or the key pair is changed during a certificate request process.
- Exclusive certificate request applications are running in the PKI domain.
- The CA server does not accept the source IP address specified in the PKI domain, or no source IP address is specified.
- The system time of the device is not synchronized with the CA server.

Solution

1. Fix the network connection problems, if any.
2. Obtain or import the CA certificate.
3. Use the **ping** command to verify that the registration server is reachable.
4. Use the **certificate request from** command to specify the correct certificate request reception authority.
5. Configure the PKI entity parameters as required by the registration policy on the CA or RA.
6. Specify the key pair for certificate request, or remove the existing key pair, specify a new key pair, and submit a local certificate request again.
7. Use the **pki abort-certificate-request domain** command to abort the certificate request.
8. Specify the correct source IP address that the CA server can accept. For the correct settings, contact the CA administrator.
9. Synchronize the system time of the device with the CA server.
10. If the problem persists, contact H3C Support.

Failed to obtain CRLs

Symptom

CRLs cannot be obtained.

Analysis

- The network connection is down, for example, because the network cable is damaged or the connectors have bad contact.
- The PKI domain does not have a CA certificate before you try to obtain CRLs.
- The URL of the CRL repository is not configured and cannot be obtained from the CA certificate or local certificates in the PKI domain.
- The specified URL of the CRL repository is incorrect.
- The device tries to obtain CRLs through SCEP, but it experiences the following problems:
  - The PKI domain does not have local certificates.
  - The key pairs in the certificates have been changed.
  - The PKI domain has incorrect URL for certificate request.

- The CRL repository uses LDAP for CRL distribution. However, the IP address or host name of the LDAP server is neither contained in the CRL repository URL nor configured in the PKI domain.
- The CA does not issue CRLs.
- The CA server does not accept the source IP address specified in the PKI domain, or no source IP address is specified.

Solution

1. Fix the network connection problems, if any.
2. Obtain or import the CA certificate.
3. If the URL of the CRL repository cannot be obtained, verify that the following conditions exist:
   - The URL for certificate request is valid.
   - A local certificate has been successfully obtained.
   - The local certificate contains a public key that matches the locally stored key pair.
4. Make sure the LDAP server address is contained in the CRL repository URL, or is configured in the PKI domain.
5. Make sure the CA server support publishing CRLs.
6. Specify a correct source IP address that the CA server can accept. For the correct settings, contact the CA administrator.
7. If the problem persists, contact H3C Support.

Failed to import the CA certificate

Symptom

The CA certificate cannot be imported.

Analysis

- CRL checking is enabled, but the device does not have a CRL in the PKI domain and cannot obtain one.
- The specified format in which the CA certificate file is to be imported does not match actual certificate file format.

Solution

1. Use the **`undo crl check enable`** command to disable CRL checking in the PKI domain.
2. Make sure the format of the imported file is correct.
3. If the problem persists, contact H3C Support.

Failed to import the local certificate

Symptom

The local certificate cannot be imported.

Analysis

- The PKI domain does not have a CA certificate, and the local certificate file to be imported does not contain the CA certificate chain.
- CRL checking is enabled, but the device does not have a CRL in the PKI domain and cannot obtain one.
- The specified format in which the local certificate file is to be imported does not match actual certificate file format.

- The device and the certificate do not have the local key pair.
- The certificate has been revoked.
- The certificate is out of the validity period.
- The system time is incorrect.

Solution

1. Obtain or import the CA certificate.
2. Use the **undo crl check enable** command to disable CRL checking, or obtain the correct CRL before you import certificates.
3. Make sure the format of the file to be imported is correct.
4. Make sure the certificate file contains the private key.
5. Make sure the certificate is not revoked.
6. Make sure the certificate is valid.
7. Configure the correct system time for the device.
8. If the problem persists, contact H3C Support.

Failed to export certificates

Symptom

Certificates cannot be exported.

Analysis

- The PKI domain does not have local certificates when you export all certificates in PKCS12 format.
- The specified export path does not exist.
- The specified export path is illegal.
- The public key of the local certificate to be exported does not match the public key of the key pair configured in the PKI domain.
- The storage space of the device is full.

Solution

1. Obtain or request local certificates first.
2. Use the **mkdir** command to create the required path.
3. Specify a correct export path.
4. Configure the correct key pair in the PKI domain.
5. Clear up the storage space of the device.
6. If the problem persists, contact H3C Support.

Failed to set the storage path

Symptom

The storage path for certificates or CRLs cannot be set.

Analysis

- The specified storage path does not exist.
- The specified storage path is illegal.
- The storage space of the device is full.

Solution

1. Use the **mkdir** command to create the path.
2. Specify a valid storage path for certificates or CRLs.
3. Clear up the storage space of the device.
4. If the problem persists, contact H3C Support.

Contents

Configuring IPsec 3

About IPsec 3
IPsec framework 3
IPsec security services 3
Benefits of IPsec 3
Security protocols 3
Encapsulation modes 4
Security association 5
Authentication and encryption 6
IPsec-protected traffic 6
ACL-based IPsec 6
IPv6 routing protocol-based IPsec 7
IPsec policy and IPsec profile 7
Protocols and standards 8
FIPS compliance 8
Restrictions and guidelines: IPsec configuration 8
Implementing ACL-based IPsec 8
ACL-based IPsec tasks at a glance 8
Configuring an ACL 9
Configuring an IPsec transform set 11
Configuring a manual IPsec policy 14
Configuring an IKE-based IPsec policy 15
Applying an IPsec policy to an interface 18
Enabling ACL checking for de-encapsulated packets 19
Configuring IPsec anti-replay 19
Configuring IPsec anti-replay redundancy 20
Binding a source interface to an IPsec policy 20
Enabling QoS pre-classify 21
Configuring the DF bit of IPsec packets 21
Configuring IPsec RRI 22
Configuring IPsec for IPv6 routing protocols 23
IPsec protection for IPv6 routing protocols tasks at a glance 23
Configuring a manual IPsec profile 23
Applying the IPsec profile to an IPv6 routing protocol 24
Configuring the global IPsec SA lifetime and idle timeout 25
Configuring IPsec fragmentation 25
Setting the maximum number of IPsec tunnels 26
Configuring IPsec logging 26
Configuring SNMP notifications for IPsec 26
Display and maintenance commands for IPsec 27
IPsec configuration examples 28
Example: Configuring a manual mode IPsec tunnel for IPv4 packets 28
Example: Configuring an IKE-based IPsec tunnel for IPv4 packets 32
Example: Configuring IPsec for RIPng 37
Example: Configuring IPsec RRI 43

Configuring IKE 51

About IKE 51
Benefits of IKE 51
Relationship between IPsec and IKE 51
IKE negotiation process 51
IKE security mechanism 53
Protocols and standards 53
FIPS compliance 54
IKE tasks at a glance 54
Prerequisites for IKE configuration 54
Configuring an IKE profile 55

Creating an IKE profile ····· 55
Configuring peer IDs for the IKE profile ····· 55
Specifying the IKE keychain or PKI domain ····· 55
Configuring the IKE phase 1 negotiation mode ····· 56
Specifying IKE proposals for the IKE profile ····· 56
Configuring the local ID for the IKE profile ····· 57
Configuring optional features for the IKE profile ····· 57
Configuring an IKE proposal ····· 58
Configuring an IKE keychain ····· 59
Configuring the global identity information ····· 60
Configuring the IKE keepalive feature ····· 61
Configuring the IKE NAT keepalive feature ····· 61
Configuring global IKE DPD ····· 62
Enabling invalid SPI recovery ····· 62
Setting the maximum number of IKE SAs ····· 63
Configuring an IKE IPv4 address pool ····· 63
Configuring IKE logging ····· 64
Configuring SNMP notifications for IKE ····· 64
Display and maintenance commands for IKE ····· 65
IKE configuration examples ····· 65
Example: Configuring main-mode IKE with preshared key authentication ····· 65
Example: Configuring an IKE-based IPsec tunnel for IPv4 packets ····· 70
Troubleshooting IKE ····· 74
IKE negotiation failed because no matching IKE proposals were found ····· 74
IKE negotiation failed because no IKE proposals or IKE keychains are specified correctly ····· 75
IPsec SA negotiation failed because no matching IPsec transform sets were found ····· 75
IPsec SA negotiation failed due to invalid identity information ····· 76
Configuring IKEv2 ····· 79
About IKEv2 ····· 79
IKEv2 negotiation process ····· 79
New features in IKEv2 ····· 80
Protocols and standards ····· 80
IKEv2 tasks at a glance ····· 80
Prerequisites for IKEv2 configuration ····· 81
Configuring an IKEv2 profile ····· 81
Creating an IKEv2 profile ····· 81
Specifying the local and remote identity authentication methods ····· 82
Configuring the IKEv2 keychain or PKI domain ····· 82
Configuring the local ID for the IKEv2 profile ····· 83
Configuring peer IDs for the IKEv2 profile ····· 83
Specifying a VPN instance for the IKEv2 profile ····· 83
Configuring optional features for the IKEv2 profile ····· 84
Configuring an IKEv2 policy ····· 85
Configuring an IKEv2 proposal ····· 86
Configuring an IKEv2 keychain ····· 87
Configure global IKEv2 parameters ····· 88
Enabling the cookie challenging feature ····· 88
Configuring the IKEv2 DPD feature ····· 88
Configuring the IKEv2 NAT keepalive feature ····· 89
Configuring IKEv2 address pools ····· 89
Display and maintenance commands for IKEv2 ····· 89
Troubleshooting IKEv2 ····· 90
IKEv2 negotiation failed because no matching IKEv2 proposals were found ····· 90
IPsec SA negotiation failed because no matching IPsec transform sets were found ····· 91
IPsec tunnel establishment failed ····· 91

Configuring IPsec

About IPsec

IP Security (IPsec) is defined by the IETF to provide interoperable, high-quality, cryptography-based security for IP communications. It is a Layer 3 VPN technology that transmits data in a secure channel established between two endpoints (such as two security gateways). Such a secure channel is usually called an IPsec tunnel.

IPsec framework

IPsec is a security framework that has the following protocols and algorithms:

- Authentication Header (AH).
- Encapsulating Security Payload (ESP).
- Internet Key Exchange (IKE).
- Algorithms for authentication and encryption.

AH and ESP are security protocols that provide security services. IKE performs automatic key exchange. For more information about IKE, see "Configuring IKE."

IPsec security services

IPsec provides the following security services for data packets in the IP layer:

- **Confidentiality**—The sender encrypts packets before transmitting them over the Internet, protecting the packets from being eavesdropped en route.
- **Data integrity**—The receiver verifies the packets received from the sender to make sure they are not tampered with during transmission.
- **Data origin authentication**—The receiver verifies the authenticity of the sender.
- **Anti-replay**—The receiver examines packets and drops outdated and duplicate packets.

Benefits of IPsec

IPsec delivers the following benefits:

- Reduced key negotiation overhead and simplified maintenance by supporting the IKE protocol. IKE provides automatic key negotiation and automatic IPsec security association (SA) setup and maintenance.
- Good compatibility. You can apply IPsec to all IP-based application systems and services without modifying them.
- Encryption on a per-packet rather than per-flow basis. Per-packet encryption allows for flexibility and greatly enhances IP security.

Security protocols

IPsec comes with two security protocols, AH and ESP. They define how to encapsulate IP packets and the security services that they can provide.

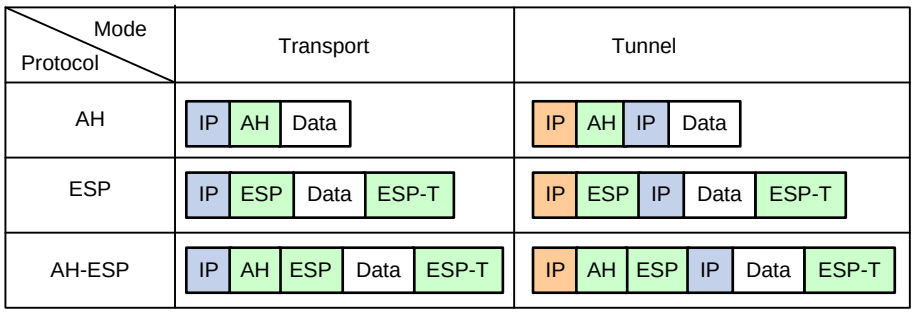
- AH (protocol 51) defines the encapsulation of the AH header in an IP packet, as shown in Figure 1. AH can provide data origin authentication, data integrity, and anti-replay services to

prevent data tampering, but it cannot prevent eavesdropping. Therefore, it is suitable for transmitting non-confidential data. The authentication algorithms supported by AH include HMAC-MD5 and HMAC-SHA1. AH does not support NAT traversal.

- ESP (protocol 50) defines the encapsulation of the ESP header and trailer in an IP packet, as shown in Figure 1. ESP can provide data encryption, data origin authentication, data integrity, and anti-replay services. Unlike AH, ESP can guarantee data confidentiality because it can encrypt the data before encapsulating the data to IP packets. ESP-supported encryption algorithms include DES, 3DES, and AES, and authentication algorithms include HMAC-MD5 and HMAC-SHA1.

Both AH and ESP provide authentication services, but the authentication service provided by AH is stronger. In practice, you can choose either or both security protocols. When both AH and ESP are used, an IP packet is encapsulated first by ESP and then by AH.

Figure 1 Security protocol encapsulations in different modes



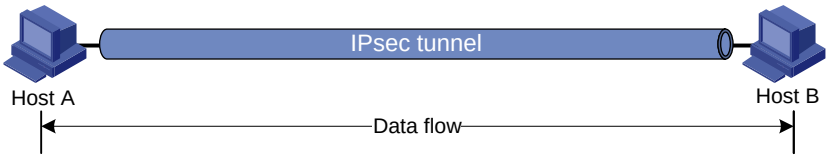
Encapsulation modes

IPsec supports the following encapsulation modes: transport mode and tunnel mode.

Transport mode

The security protocols protect the upper layer data of an IP packet. Only the transport layer data is used to calculate the security protocol headers. The calculated security protocol headers and the encrypted data (only for ESP encapsulation) are placed after the original IP header. You can use the transport mode when end-to-end security protection is required (the secured transmission start and end points are the actual start and end points of the data). The transport mode is typically used for protecting host-to-host communications, as shown in Figure 2.

Figure 2 IPsec protection in transport mode



Tunnel mode

The security protocols protect the entire IP packet. The entire IP packet is used to calculate the security protocol headers. The calculated security protocol headers and the encrypted data (only for ESP encapsulation) are encapsulated in a new IP packet. In this mode, the encapsulated packet has two IP headers. The inner IP header is the original IP header. The outer IP header is added by the network device that provides the IPsec service. You must use the tunnel mode when the secured transmission start and end points are not the actual start and end points of the data packets (for example, when two gateways provide IPsec but the data start and end points are two hosts behind the gateways). The tunnel mode is typically used for protecting gateway-to-gateway communications, as shown in Figure 3.

Figure 3 IPsec protection in tunnel mode

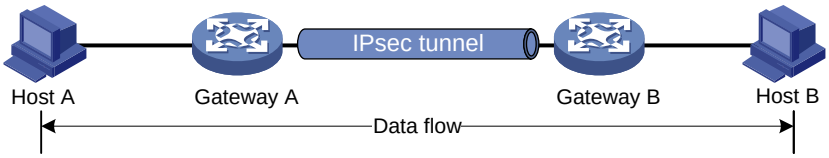


Figure 4 shows how the security protocols encapsulate an IP packet in different encapsulation modes.

Figure 4 Security protocol encapsulations in different modes

Mode / Protocol	Transport	Tunnel
AH	IP AH Data	IP AH IP Data
ESP	IP ESP Data ESP-T	IP ESP IP Data ESP-T
AH-ESP	IP AH ESP Data ESP-T	IP AH ESP IP Data ESP-T

Security association

A security association (SA) is an agreement negotiated between two communicating parties called IPsec peers. An SA includes the following parameters for data protection:

- Security protocols (AH, ESP, or both).
- Encapsulation mode (transport mode or tunnel mode).
- Authentication algorithm (HMAC-MD5 or HMAC-SHA1).
- Encryption algorithm (DES, 3DES, or AES).
- Shared keys and their lifetimes.

An SA is unidirectional. At least two SAs are needed to protect data flows in a bidirectional communication. If two peers want to use both AH and ESP to protect data flows between them, they construct an independent SA for each protocol in each direction.

An SA is uniquely identified by a triplet, which consists of the security parameter index (SPI), destination IP address, and security protocol identifier. An SPI is a 32-bit number. It is transmitted in the AH/ESP header.

An SA can be set up manually or through IKE.

- **Manual mode**—Configure all parameters for the SA through commands. This configuration mode is complex and does not support some advanced features (such as periodic key update), but it can implement IPsec without IKE. This mode is mainly used in small and static networks or when the number of IPsec peers in the network is small.
- **IKE negotiation mode**—The peers negotiate and maintain the SA through IKE. This configuration mode is simple and has good expansibility. As a best practice, set up SAs through IKE negotiations in medium- and large-scale dynamic networks.

A manually configured SA never ages out. An IKE-created SA has a lifetime, which comes in two types:

- **Time-based lifetime**—Defines how long the SA can be valid after it is created.
- **Traffic-based lifetime**—Defines the maximum traffic that the SA can process.

If both lifetime timers are configured for an SA, the SA becomes invalid when either of the lifetime timers expires. Before the SA expires, IKE negotiates a new SA, which takes over immediately after its creation.

Authentication and encryption

Authentication algorithms

IPsec uses hash algorithms to perform authentication. A hash algorithm produces a fixed-length digest for an arbitrary-length message. IPsec peers respectively calculate message digests for each packet. The receiver compares the local digest with that received from the sender. If the digests are identical, the receiver considers the packet intact and the sender's identity valid. IPsec uses the Hash-based Message Authentication Code (HMAC) based authentication algorithms, including HMAC-MD5 and HMAC-SHA1. Compared with HMAC-SHA1, HMAC-MD5 is faster but less secure.

Encryption algorithms

IPsec uses symmetric encryption algorithms, which encrypt and decrypt data by using the same keys. The following encryption algorithms are available for IPsec on the device:

- **DES**—Encrypts a 64-bit plaintext block with a 56-bit key. DES is the least secure but the fastest algorithm.
- **3DES**—Encrypts plaintext data with three 56-bit DES keys. The key length totals up to 168 bits. It provides moderate security strength and is slower than DES.
- **AES**—Encrypts plaintext data with a 128-bit, 192-bit, or 256-bit key. AES provides the highest security strength and is slower than 3DES.

IPsec-protected traffic

IPsec tunnels can protect the following types of traffic:

- Packets that match specific ACLs.
- Packets routed to a tunnel interface.
- Packets of IPv6 routing protocols.

Two peers use security policies (IPsec policies or IPsec profiles) to protect packets between them. A security policy defines the range of packets to be protected by IPsec and the security parameters used for the protection. For more information about IPsec policies and IPsec profiles, see "IPsec policy and IPsec profile."

The following information describes how IPsec protects packets:

- When an IPsec peer identifies the packets to be protected according to the security policy, it sets up an IPsec tunnel and sends the packet to the remote peer through the tunnel. The IPsec tunnel can be manually configured beforehand, or it can be set up through IKE negotiation triggered by the packet. The IPsec tunnels are actually the IPsec SAs. The inbound packets are protected by the inbound SA, and the outbound packets are protected by the outbound SA.
- When the remote IPsec peer receives the packet, it drops, de-encapsulates, or directly forwards the packet according to the configured security policy.

ACL-based IPsec

To implement ACL-based IPsec, configure an ACL to define the data flows to be protected, specify the ACL in an IPsec policy, and then apply the IPsec policy to an interface.

ACL-based IPsec works as follows:

- When packets sent by the interface match a permit rule of the ACL, the packets are protected by the outbound IPsec SA and encapsulated with IPsec.

- When the interface receives an IPsec packet destined for the local device, it searches for the inbound IPsec SA according to the SPI in the IPsec packet header for de-encapsulation. If the de-encapsulated packet matches a permit rule of the ACL, the device processes the packet. If the de-encapsulated packet does not match any permit rule of the ACL, the device drops the packet.

The device supports the following data flow protection modes:

- **Standard mode**—One IPsec tunnel protects one data flow. The data flow permitted by an ACL rule is protected by one IPsec tunnel that is established solely for it.
- **Aggregation mode**—One IPsec tunnel protects all data flows permitted by all the rules of an ACL. This mode is only used to communicate with old-version devices.
- **Per-host mode**—One IPsec tunnel protects one host-to-host data flow. One host-to-host data flow is identified by one ACL rule and protected by one IPsec tunnel established solely for it. This mode consumes more system resources when multiple data flows exist between two subnets to be protected.

IPv6 routing protocol-based IPsec

You can implement IPv6 routing protocol-based IPsec by binding an IPsec profile to an IPv6 routing protocol. All packets of the protocol are encapsulated with IPsec. Supported IPv6 routing protocols include OSPFv3, IPv6 BGP, and RIPng.

All packets of the applications that are not bound to IPsec and the IPsec packets that failed to be de-encapsulated are dropped.

In one-to-many communication scenarios, you must configure the IPsec SAs for an IPv6 routing protocol in manual mode because of the following reasons:

- The automatic key exchange mechanism protects communications between two points. In one-to-many communication scenarios, automatic key exchange cannot be implemented.
- One-to-many communication scenarios require that all the devices use the same SA parameters (SPI and key) to receive and send packets. IKE negotiated SAs cannot meet this requirement.

IPsec policy and IPsec profile

IPsec policies and IPsec profiles define the parameters used to establish IPsec tunnels between two peers and the range of packets to be protected.

IPsec policy

An IPsec policy is a set of IPsec policy entries that have the same name but different sequence numbers.

An IPsec policy contains the following settings:

- An ACL that defines the range of data flows to be protected.
- An IPsec transform set that defines the security parameters used for IPsec protection.
- IPsec SA establishment mode.

  IPsec SAs can be established through manual configuration or IKE negotiation.
- Local and remote IP addresses that define the start and end points of the IPsec tunnel.

In the same IPsec policy, an IPsec policy entry with a smaller sequence number has a higher priority. When sending a packet, the interface applied with an IPsec policy looks through the IPsec policy's entries in ascending order of sequence numbers. If the packet matches the ACL of an IPsec policy entry, the interface encapsulates the packet according to the IPsec policy entry. If no match is found, the interface sends the packet out without IPsec protection.

When the interface receives an IPsec packet destined for the local device, it searches for the inbound IPsec SA according to the SPI in the IPsec packet header for de-encapsulation. If the de-encapsulated packet matches a permit rule of the ACL, the device processes the packet. If the de-encapsulated packet does not match a permit rule of the ACL, the device drops the packet.

IPsec profile

An IPsec profile has similar settings as an IPsec policy. It is uniquely identified by a name and does not support ACL configuration.

IPsec profiles can be classified into the following types:

- **Manual IPsec profile**—A manual IPsec profile is used to protect IPv6 routing protocols. It specifies the IPsec transform set used for protecting data flows, and the SPIs and keys used by the SAs.
- **IKE-based IPsec profile**—An IKE-based IPsec profile is applied to tunnel interfaces to protect tunneled traffic. It specifies the IPsec transform sets used for protecting data flows, and the IKE profile used for IKE negotiation.

Protocols and standards

- RFC 2401, *Security Architecture for the Internet Protocol*
- RFC 2402, *IP Authentication Header*
- RFC 2406, *IP Encapsulating Security Payload*
- RFC 4552, *Authentication/Confidentiality for OSPFv3*

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see "Configuring FIPS") and non-FIPS mode.

Restrictions and guidelines: IPsec configuration

IPsec supports encrypting and decrypting only software forwarded packets on the control plane. It does not handle the data packets on the user plane. For more information about the control plane and user plane, see QoS configuration in *ACL and QoS Configuration Guide*.

Typically, IKE uses UDP port 500 for communication, and AH and ESP use the protocol numbers 51 and 50, respectively. Make sure traffic of these protocols is not denied on the interfaces with IKE or IPsec configured.

Implementing ACL-based IPsec

ACLs for IPsec take effect only on traffic that is generated by the device and traffic that is destined for the device. They do not take effect on traffic forwarded through the device. For example, an ACL-based IPsec tunnel can protect log messages the device sends to a log server, but it does not protect data flows and voice flows that are forwarded by the device.

ACL-based IPsec tasks at a glance

To configure ACL-based IPsec, perform the following tasks:

1. Configuring an ACL

2. Configuring an IPsec transform set
3. Configuring an IPsec policy

   Choose one of the following tasks:
   - Configuring a manual IPsec policy
   - Configuring an IKE-based IPsec policy
4. Applying an IPsec policy to an interface
5. (Optional.) Configuring accessibility features for ACL-based IPsec
   - Enabling ACL checking for de-encapsulated packets
   - Configuring IPsec anti-replay
   - Configuring IPsec anti-replay redundancy
   - Binding a source interface to an IPsec policy
   - Enabling QoS pre-classify
   - Configuring the DF bit of IPsec packets
   - Configuring IPsec RRI
   - Configuring the global IPsec SA lifetime and idle timeout
   - Configuring IPsec fragmentation
6. (Optional.) Configuring logging and SNMP notification for IPsec.
   - Configuring IPsec logging
   - Configuring SNMP notifications for IPsec

Configuring an ACL

IPsec uses ACLs to identify the traffic to be protected.

Keywords in ACL rules

An ACL is a collection of ACL rules. Each ACL rule is a deny or permit statement. A permit statement identifies a data flow protected by IPsec, and a deny statement identifies a data flow that is not protected by IPsec. IPsec compares a packet against the ACL rules and processes the packet according to the first rule it matches.

- Each ACL rule matches both the outbound traffic and the returned inbound traffic. Suppose there is a rule **rule 0 permit ip source 1.1.1.0 0.0.0.255 destination 2.2.2.0 0.0.0.255**. This rule matches both traffic from 1.1.1.0 to 2.2.2.0 and traffic from 2.2.2.0 to 1.1.1.0.
- In the outbound direction, if a permit statement is matched, IPsec considers that the packet requires protection and continues to process it. If a deny statement is matched or no match is found, IPsec considers that the packet does not require protection and delivers it to the next module.
- In the inbound direction:
  - Non-IPsec packets that match a permit statement are dropped.
  - IPsec packets destined for the device itself are de-encapsulated. By default, the de-encapsulated packets are compared against the ACL rules. Only those that match a permit statement are processed. Other packets are dropped. If ACL checking for de-encapsulated IPsec packets is disabled, the de-encapsulated packets are not compared against the ACL rules and are directly processed by other modules.

When defining ACL rules for IPsec, follow these guidelines:

- Permit only data flows that need to be protected and use the **any** keyword with caution. With the **any** keyword specified in a permit statement, all outbound traffic matching the permit statement will be protected by IPsec. All inbound IPsec packets matching the permit statement will be received and processed, but all inbound non-IPsec packets will be dropped. This will cause all the inbound traffic that does not need IPsec protection to be dropped.

- Avoid statement conflicts in the scope of IPsec policy entries. When creating a deny statement, be careful with its match scope and match order relative to permit statements. The policy entries in an IPsec policy have different match priorities. ACL rule conflicts between them are prone to cause mistreatment of packets. For example, when configuring a permit statement for an IPsec policy entry to protect an outbound traffic flow, you must avoid the situation that the traffic flow matches a deny statement in a higher priority IPsec policy entry. Otherwise, the packets will be sent out as normal packets. If they match a permit statement at the receiving end, they will be dropped by IPsec.

The following example shows how an improper statement causes unexpected packet dropping. Only the ACL-related configuration is presented.

Assume Device A is connected to subnet 1.1.2.0/24 and Device B is connected to subnet 3.3.3.0/24, and the IPsec policy configuration on Device A and Device B is as follows:

- IPsec configuration on Device A:

```
acl advanced 3000
 rule 0 permit ip source 1.1.1.0 0.0.0.255 destination 2.2.2.0 0.0.0.255
 rule 1 deny ip
acl advanced 3001
 rule 0 permit ip source 1.1.2.0 0.0.0.255 destination 3.3.3.0 0.0.0.255
 rule 1 deny ip
#
ipsec policy testa 1 isakmp <---IPsec policy entry with a higher priority
 security acl 3000
 ike-profile aa
 transform-set 1
#
ipsec policy testa 2 isakmp <---IPsec policy entry with a lower priority
 security acl 3001
 ike-profile bb
 transform-set 1
```

- IPsec configuration on Device B:

```
acl advanced 3001
 rule 0 permit ip source 3.3.3.0 0.0.0.255 destination 1.1.2.0 0.0.0.255
 rule 1 deny ip
#
ipsec policy testb 1 isakmp
 security acl 3001
 ike-profile aa
 transform-set 1
```

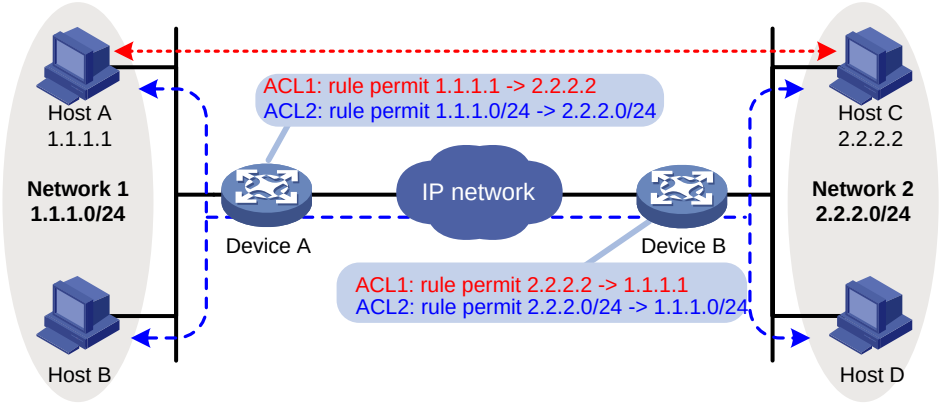
On Device A, apply the IPsec policy **testa** to the outbound interface of Device A. The IPsec policy contains two policy entries, **testa 1** and **testa 2**. The ACLs used by the two policy entries each contain a rule that matches traffic from 1.1.2.0/24 to 3.3.3.0/24. The one used in the policy entry **testa 1** is a deny statement and the one used in the policy entry **testa 2** is a permit statement. Because **testa 1** is matched prior to **testa 2**, traffic from 1.1.2.0/24 to 3.3.3.0/24 will match the deny statement and be sent as normal traffic. When the traffic arrives at Device B, the traffic matches rule 0 (a permit statement) in ACL 3001 used in the applied IPsec policy **testb**. Because non-IPsec traffic that matches a permit statement must be dropped on the inbound interface, Device B drops the traffic.

To make sure subnet 1.1.2.0/24 can access subnet 3.3.3.0/24, you can delete the deny rule in ACL 3000 on Device A.

Mirror image ACLs

To make sure SAs can be set up and the traffic protected by IPsec can be processed correctly between two IPsec peers, create mirror image ACLs on the IPsec peers. As shown in Figure 5, ACL rules on Device B are mirror images of the rules on Device A. In this way, SAs can be created successfully for the traffic between Host A and Host C and for the traffic between Network 1 and Network 2.

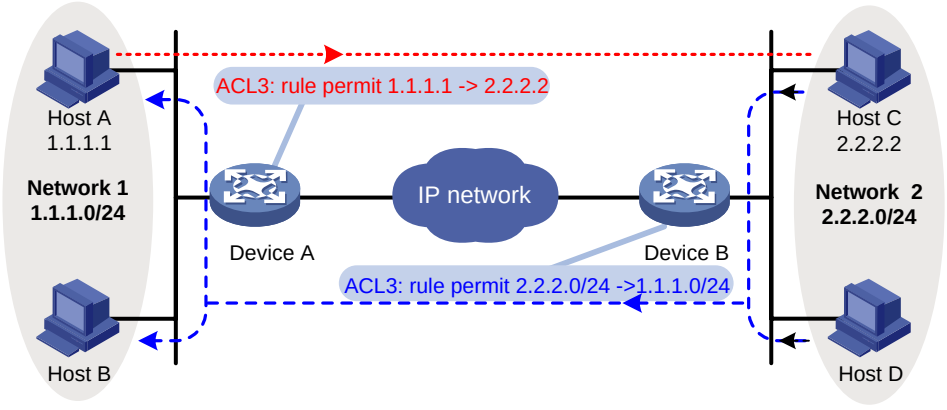
Figure 5 Mirror image ACLs



If the ACL rules on IPsec peers do not form mirror images of each other, SAs can be set up only when both of the following requirements are met:

- The range specified by an ACL rule on one peer is covered by its counterpart ACL rule on the other peer. As shown in Figure 6, the range specified by the ACL rule configured on Device A is covered by its counterpart on Device B.
- The peer with the narrower rule initiates SA negotiation. If a wider ACL rule is used by the SA initiator, the negotiation request might be rejected because the matching traffic is beyond the scope of the responder. As shown in Figure 6, the SA negotiation initiated by Host A to Host C is accepted but the SA negotiations from Host C to Host A, from Host C to Host B, and from Host D to Host A are rejected.

Figure 6 Non-mirror image ACLs



Configuring an IPsec transform set

About this task

An IPsec transform set, part of an IPsec policy, defines the security parameters for IPsec SA negotiation, including the security protocol, encryption algorithms, and authentication algorithms.

Restrictions and guidelines

Changes to an IPsec transform set affect only SAs negotiated after the changes. To apply the changes to existing SAs, execute the reset ipsec sa command to clear the SAs so that they can be set up by using the updated parameters.

In FIPS mode, you must specify both the ESP encryption algorithm and the ESP authentication algorithm for an IPsec transform set that uses the ESP security protocol.

When you set the packet encapsulation mode (tunnel or transport) for an IPsec transform set, follow these guidelines:

- The transport mode applies only when the source and destination IP addresses of data flows match those of the IPsec tunnel.
- IPsec for IPv6 routing protocols supports only the transport mode.

When you configure the Perfect Forward Secrecy (PFS) feature in an IPsec transform set, follow these guidelines:

- In IKEv1, the security level of the DH group of the initiator must be higher than or equal to that of the responder. This restriction does not apply to IKEv2.
- The end without the PFS feature performs SA negotiation according to the PFS requirements of the peer end.

You can specify multiple authentication or encryption algorithms for the same security protocol. The algorithm specified earlier has a higher priority.

Some algorithms are available only for IKEv2. See Table 1.

Table 1 Algorithms available only for IKEv2

Type	Algorithms
Encryption algorithm	aes-ctr-128 aes-ctr-192 aes-ctr-256 camellia-cbc-128 camellia-cbc-192 camellia-cbc-256 gmac-128 gmac-192 gmac-256 gcm-128 gcm-192 gcm-256
Authentication algorithm	aes-xcbc-mac

Procedure

1. Enter system view.
   system-view
2. Create an IPsec transform set and enter its view.
   ipsec transform-set *transform-set-name*
3. Specify the security protocol for the IPsec transform set.
   protocol { **ah** | **ah-esp** | **esp** }
   By default, the ESP security protocol is used.

4. Specify the encryption algorithms for ESP. Skip this step if the **protocol ah** command is configured.

   In non-FIPS mode:

   esp encryption-algorithm { **3des-cbc** | **aes-cbc-128** | **aes-cbc-192** | **aes-cbc-256** | **aes-ctr-128** | **aes-ctr-192** | **aes-ctr-256** | **camellia-cbc-128** | **camellia-cbc-192** | **camellia-cbc-256** | **des-cbc** | **gmac-128** | **gmac-192** | **gmac-256** | **gcm-128** | **gcm-192** | **gcm-256** | **null** | **sm1-cbc-128** | **sm1-cbc-192** | **sm1-cbc-256** | **sm4-cbc** } *

   By default, no encryption algorithm is specified for ESP.

   In FIPS mode:

   esp encryption-algorithm { **gcm-128** | **gcm-192** | **gcm-256** } *

   By default, no encryption algorithm is specified for ESP.

5. Specify the authentication algorithms for ESP. Skip this step if the **protocol ah** command is configured.

   esp authentication-algorithm { **aes-xcbc-mac** | **md5** | **sha1** | **sha256** | **sha384** | **sha512** | **sm3** } *

   By default, no authentication algorithm is specified for ESP.

   The **aes-xcbc-mac** algorithm is available only for IKEv2.

6. Specify the authentication algorithms for AH. Skip this step if the **protocol esp** command is configured.

   In non-FIPS mode:

   ah authentication-algorithm { **aes-xcbc-mac** | **md5** | **sha1** | **sha256** | **sha384** | **sha512** | **sm3** } *

   By default, no authentication algorithm is specified for AH.

   The **aes-xcbc-mac** algorithm is available only for IKEv2.

   In FIPS mode:

   ah authentication-algorithm { **sha256** | **sha384** | **sha512** } *

   By default, no authentication algorithm is specified for AH.

7. Specify the packet encapsulation mode.

   encapsulation-mode { **transport** | **tunnel** }

   By default, the security protocol encapsulates IP packets in tunnel mode.

8. (Optional.) Enable the PFS feature.

   In non-FIPS mode:

   pfs { **dh-group1** | **dh-group2** | **dh-group5** | **dh-group14** | **dh-group24** | **dh-group19** | **dh-group20** }

   In FIPS mode:

   pfs { **dh-group19** | **dh-group20** | **dh-group24** }

   By default, the PFS feature is disabled.

   For more information about PFS, see "Configuring IKE."

9. (Optional.) Enable the Extended Sequence Number (ESN) feature.

   esn enable [**both**]

   By default, the ESN feature is disabled.

Configuring a manual IPsec policy

In a manual IPsec policy, the parameters are configured manually, such as the keys, the SPIs, and the IP addresses of the two ends in tunnel mode.

Restrictions and guidelines

When you configure a manual IPsec policy, make sure the IPsec configuration at both ends of the IPsec tunnel meets the following requirements:

- The IPsec policies at the two ends must have IPsec transform sets that use the same security protocols, security algorithms, and encapsulation mode.
- The remote IPv4 address configured on the local end must be the same as the primary IPv4 address of the interface applied with the IPsec policy at the remote end. The remote IPv6 address configured on the local end must be the same as the first IPv6 address of the interface applied with the IPsec policy at the remote end.
- At each end, configure parameters for both the inbound SA and the outbound SA, and make sure the SAs in each direction are unique: For an outbound SA, make sure its triplet (remote IP address, security protocol, and SPI) is unique. For an inbound SA, make sure its SPI is unique.
- The local inbound SA must use the same SPI and keys as the remote outbound SA. The same is true of the local outbound SA and remote inbound SA.
- The keys for the IPsec SAs at the two tunnel ends must be configured in the same format. For example, if the local end uses a key in hexadecimal format, the remote end must also use a key in hexadecimal format. If you configure a key in both the character and the hexadecimal formats, only the most recent configuration takes effect.
- If you configure a key in character format for ESP, the device automatically generates an authentication key and an encryption key for ESP.

Procedure

1. Enter system view.

   system-view

2. Create a manual IPsec policy entry and enter its view.

   ipsec { **ipv6-policy** | **policy** } *policy-name seq-number* **manual**

3. (Optional.) Configure a description for the IPsec policy.

   description *text*

   By default, no description is configured.

4. Specify an ACL for the IPsec policy.

   security acl [**ipv6**] { *acl-number* | **name** *acl-name* }

   By default, no ACL is specified for an IPsec policy.

   You can specify only one ACL for an IPsec policy.

5. Specify an IPsec transform set for the IPsec policy.

   transform-set *transform-set-name*

   By default, no IPsec transform set is specified for an IPsec policy.

   You can specify only one IPsec transform set for a manual IPsec policy.

6. Specify the remote IP address of the IPsec tunnel.

   remote-address { *ipv4-address* | **ipv6** *ipv6-address* }

   By default, the remote IP address of the IPsec tunnel is not specified.

7. Configure an SPI for the inbound IPsec SA.

   sa spi inbound { **ah** | **esp** } *spi-number*

   By default, no SPI is configured for the inbound IPsec SA.

8. Configure an SPI for the outbound IPsec SA.

   sa spi outbound { **ah** | **esp** } *spi-number*

   By default, no SPI is configured for the outbound IPsec SA.

9. Configure keys for the IPsec SA.
   - Configure an authentication key in hexadecimal format for AH.

     sa hex-key authentication { **inbound** | **outbound** } **ah** { **cipher** | **simple** } *string*
   - Configure an authentication key in character format for AH.

     sa string-key { **inbound** | **outbound** } **ah** { **cipher** | **simple** } *string*
   - Configure a key in character format for ESP.

     sa string-key { **inbound** | **outbound** } **esp** { **cipher** | **simple** } *string*
   - Configure an authentication key in hexadecimal format for ESP.

     sa hex-key authentication { **inbound** | **outbound** } **esp** { **cipher** | **simple** }
   - Configure an encryption key in hexadecimal format for ESP.

     sa hex-key encryption { **inbound** | **outbound** } **esp** { **cipher** | **simple** } *string*

   By default, no keys are configured for the IPsec SA.

   Configure keys correctly for the security protocol (AH, ESP, or both) you have specified in the IPsec transform set used by the IPsec policy.

Configuring an IKE-based IPsec policy

About this task

In an IKE-based IPsec policy, the parameters are automatically negotiated through IKE.

To configure an IKE-based IPsec policy, use one of the following methods:

- Directly configure it by configuring the parameters in IPsec policy view.
- Configure it by using an existing IPsec policy template with the parameters to be negotiated configured.

  A device using an IPsec policy that is configured in this way cannot initiate an SA negotiation, but it can respond to a negotiation request. The parameters not defined in the template are determined by the initiator. For example, in an IPsec policy template, the ACL is optional. If you do not specify an ACL, the IPsec protection range has no limit. So the device accepts all ACL settings of the negotiation initiator.

  When the remote end's information (such as the IP address) is unknown, this method allows the remote end to initiate negotiations with the local end.

The configurable parameters for an IPsec policy template are the same as those when you directly configure an IKE-based IPsec policy. The difference is that more parameters are optional for an IPsec policy template. Except the IPsec transform sets and the IKE profile, all other parameters are optional.

Restrictions and guidelines for IKE-based IPsec policy configuration

The IPsec policies at the two tunnel ends must have IPsec transform sets that use the same security protocols, security algorithms, and encapsulation mode.

The IPsec policies at the two tunnel ends must have the same IKE profile parameters.

An IKE-based IPsec policy can use a maximum of six IPsec transform sets. During an IKE negotiation, IKE searches for a fully matched IPsec transform set at the two ends of the IPsec tunnel. If no match is found, no SA can be set up, and the packets expecting to be protected will be dropped.

The remote IP address of the IPsec tunnel is required on an IKE negotiation initiator and is optional on the responder. The remote IP address specified on the local end must be the same as the local IP address specified on the remote end.

The IPsec SA uses the local lifetime settings or those proposed by the peer, whichever are smaller.

The IPsec SA can have both a time-based lifetime and a traffic-based lifetime. The IPsec SA expires when either lifetime expires.

If you specify both an IKEv1 profile and an IKEv2 profile for an IPsec policy, the IKEv2 profile is used preferentially. For more information about IKEv1 and IKEv2 profiles, see "Configuring IKE" and "Configuring IKEv2."

Directly configuring an IKE-based IPsec policy

1. Enter system view.

   system-view

2. Create an IKE-based IPsec policy entry and enter its view.

   ipsec { **ipv6-policy** | **policy** } *policy-name seq-number* **isakmp**

3. (Optional.) Configure a description for the IPsec policy.

   description *text*

   By default, no description is configured.

4. Specify an ACL for the IPsec policy.

   security acl [**ipv6**] { *acl-number* | **name** *acl-name* } [**aggregation** | **per-host**]

   By default, no ACL is specified for an IPsec policy.

   You can specify only one ACL for an IPsec policy.

5. Specify IPsec transform sets for the IPsec policy.

   transform-set *transform-set-name*&<1-6>

   By default, no IPsec transform sets are specified for an IPsec policy.

6. Specify an IKE profile or IKEv2 profile for the IPsec policy.
   - **Specify an IKE profile.**

     ike-profile *profile-name*

     By default, no IKE profile is specified for an IPsec policy.
   - **Specify an IKEv2 profile.**

     ikev2-profile *profile-name*

     By default, no IKEv2 profile is specified for an IPsec policy.

7. Specify the local IP address of the IPsec tunnel.

   local-address { *ipv4-address* | **ipv6** *ipv6-address* }

   By default, the local IPv4 address of the IPsec tunnel is the primary IPv4 address of the interface to which the IPsec policy is applied. The local IPv6 address of the IPsec tunnel is the first IPv6 address of the interface to which the IPsec policy is applied.

   The local IP address specified by this command must be the same as the IP address used as the local IKE identity.

   In a VRRP network, the local IP address must be the virtual IP address of the VRRP group to which the IPsec-applied interface belongs.

8. Specify the remote IP address of the IPsec tunnel.

   remote-address { [**ipv6**] *host-name* | *ipv4-address* | **ipv6** *ipv6-address* }

   By default, the remote IP address of the IPsec tunnel is not specified.

9. (Optional.) Set the lifetime or idle timeout for the IPsec SA.

- Set the IPsec SA lifetime.

  sa duration { **time-based** *seconds* | **traffic-based** *kilobytes* }

  By default, the global SA lifetime is used.
- Set the IPsec SA idle timeout.

  sa idle-time *seconds*

  By default, the global IPsec SA idle timeout is used.

10. (Optional.) Enable the Traffic Flow Confidentiality (TFC) padding feature.

tfc enable

By default, the TFC padding feature is disabled.

TFC padding applies only to IPsec SAs negotiated by IKEv2.

Configuring an IKE-based IPsec policy by using an IPsec policy template

1. Enter system view.

system-view

2. Create an IPsec policy template and enter its view.

ipsec { **ipv6-policy-template** | **policy-template** } *template-name seq-number*

3. (Optional.) Configure a description for the IPsec policy template.

description *text*

By default, no description is configured.

4. (Optional.) Specify an ACL for the IPsec policy template.

security acl [**ipv6**] { *acl-number* | **name** *acl-name* } [**aggregation** | **per-host**]

By default, no ACL is specified for an IPsec policy template.

You can specify only one ACL for an IPsec policy template.

5. Specify IPsec transform sets for the IPsec policy template.

transform-set *transform-set-name*&<1-6>

By default, no IPsec transform sets are specified for an IPsec policy template.

6. Specify an IKE profile or IKEv2 profile for the IPsec policy template.

- Specify an IKE profile.

  ike-profile *profile-name*

  By default, no IKE profile is specified for an IPsec policy template.

  Make sure the specified IKE profile is not used by another IPsec policy or IPsec policy template.
- Specify an IKEv2 profile.

  ikev2-profile *profile-name*

  By default, no IKEv2 profile is specified for an IPsec policy template.

7. Specify the local IP address of the IPsec tunnel.

local-address { *ipv4-address* | **ipv6** *ipv6-address* }

The default local IPv4 address and IPv6 address is the primary IPv4 address and first IPv6 address of the interface where the IPsec policy is applied.

The local IP address specified by this command must be the same as the IP address used as the local IKE identity.

In a VRRP network, the local IP address must be the virtual IP address of the VRRP group to which the IPsec-applied interface belongs.

8. Specify the remote IP address of the IPsec tunnel.

remote-address { [**ipv6**] *host-name* | *ipv4-address* | **ipv6** *ipv6-address* }

By default, the remote IP address of the IPsec tunnel is not specified.

9. (Optional.) Set the lifetime and idle timeout for the IPsec SA.
   - Set the IPsec SA lifetime.

     sa duration { **time-based** *seconds* | **traffic-based** *kilobytes* }

     By default, the global SA lifetime is used.
   - Set the IPsec SA idle timeout.

     sa idle-time *seconds*

     By default, the global IPsec SA idle timeout is used.
10. (Optional.) Enable the Traffic Flow Confidentiality (TFC) padding feature.

    tfc enable

    By default, the TFC padding feature is disabled.

    TFC padding applies only to IPsec SAs negotiated by IKEv2.
11. Return to system view.

    quit
12. Create an IPsec policy by using the IPsec policy template.

    ipsec { **ipv6-policy** | **policy** } *policy-name seq-number* **isakmp template** *template-name*

Applying an IPsec policy to an interface

Restrictions and guidelines

You can apply an IPsec policy to interfaces to protect data flows.

- An IKE-based IPsec policy can be applied to multiple interfaces. As a best practice, apply an IKE-based IPsec policy to only one interface.
- A manual IPsec policy can be applied to only one interface.

To cancel the IPsec protection, remove the application of the IPsec policy.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Apply an IPsec policy to the interface.

   ipsec apply { **ipv6-policy** | **policy** } *policy-name*

   By default, no IPsec policy is applied to an interface.

   On one interface, you can apply a maximum of two IPsec policies: one IPv4 IPsec policy and one IPv6 IPsec policy.

   This feature is available only on Layer 3 interfaces.

   ○

Enabling ACL checking for de-encapsulated packets

About this task

This feature compares the de-encapsulated incoming IPsec packets against the ACL in the IPsec policy and discards those that do not match any permit rule of the ACL. This feature can protect networks against attacks using forged IPsec packets.

This feature applies only to tunnel-mode IPsec.

Procedure

1. Enter system view.
   system-view
2. Enable ACL checking for de-encapsulated packets.
   ipsec decrypt-check enable
   By default, ACL checking for de-encapsulated packets is enabled.

Configuring IPsec anti-replay

About this task

IPsec anti-replay protects networks against anti-replay attacks by using a sliding window mechanism called anti-replay window. This feature checks the sequence number of each received IPsec packet against the current IPsec packet sequence number range of the sliding window. If the sequence number is not in the current sequence number range, the packet is considered a replayed packet and is discarded.

IPsec packet de-encapsulation involves complicated calculation. De-encapsulation of replayed packets is not required, and the de-encapsulation process consumes large amounts of resources and degrades performance, resulting in DoS. IPsec anti-replay can check and discard replayed packets before de-encapsulation.

In some situations, service data packets are received in a different order than their original order. The IPsec anti-replay feature drops them as replayed packets, which impacts communications. If this happens, disable IPsec anti-replay checking or adjust the size of the anti-replay window as required.

Restrictions and guidelines

IPsec anti-replay does not affect manually created IPsec SAs. According to the IPsec protocol, only IKE-based IPsec SAs support anti-replay.

Set the anti-replay window size as small as possible to reduce the impact on system performance.

Failure to detect anti-replay attacks might result in denial of services. If you want to disable IPsec anti-replay, make sure you understand the impact of the operation on network security.

Procedure

1. Enter system view.
   system-view
2. Enable IPsec anti-replay.
   ipsec anti-replay check
   By default, IPsec anti-replay is enabled.
3. Set the size of the IPsec anti-replay window.
   ipsec anti-replay window *width*
   The default size is 64.

Configuring IPsec anti-replay redundancy

About this task

This feature synchronizes the following information from the active device to the standby device at configurable packet-based intervals:

- Lower bound values of the IPsec anti-replay window for inbound packets.
- IPsec anti-replay sequence numbers for outbound packets.

This feature, used together with IPsec redundancy, ensures uninterrupted IPsec traffic forwarding and anti-replay protection when the active device fails.

Procedure

1. Enter system view.

   system-view

2. Enable IPsec redundancy.

   ipsec redundancy enable

   By default, IPsec redundancy is disabled.

3. Enter IPsec policy view or IPsec policy template view.
   - Enter IPsec policy view.

     ipsec { **ipv6-policy** | **policy** } *policy-name seq-number* [**isakmp** | **manual**]

   - Enter IPsec policy template view.

     ipsec { **ipv6-policy-template** | **policy-template** } *template-name seq-number*

4. Set the anti-replay window synchronization interval for inbound packets and the sequence number synchronization interval for outbound packets.

   redundancy replay-interval inbound *inbound-interval* **outbound** *outbound-interval*

   By default, the active device synchronizes the anti-replay window every time it receives 1000 packets and synchronizes the sequence number every time it sends 100000 packets.

Binding a source interface to an IPsec policy

About this task

For high availability, a core device is usually connected to an ISP through two links, which operate in backup or load sharing mode. The two interfaces negotiate with their peers to establish IPsec SAs respectively. When one interface fails and a link failover occurs, the other interface needs to take some time to renegotiate SAs, resulting in service interruption.

To solve these problems, bind a source interface to an IPsec policy and apply the policy to both interfaces. This enables the two physical interfaces to use the same source interface to negotiate IPsec SAs. As long as the source interface is up, the negotiated IPsec SAs will not be removed and will keep working, regardless of link failover.

Restrictions and guidelines

Only the IKE-based IPsec policies can be bound to a source interface.

An IPsec policy can be bound to only one source interface.

A source interface can be bound to multiple IPsec policies.

If the source interface bound to an IPsec policy is removed, the IPsec policy becomes a common IPsec policy.

If no local address is specified for an IPsec policy that has been bound to a source interface, the IPsec policy uses the IP address of the bound source interface to perform IKE negotiation. If a local address is specified, the IPsec policy uses the local address to perform IKE negotiation.

Procedure

1. Enter system view.

   system-view

2. Bind a source interface to an IPsec policy.

   ipsec { **ipv6-policy** | **policy** } *policy-name* **local-address** *interface-type interface-number*

   By default, no source interface is bound to an IPsec policy.

Enabling QoS pre-classify

About this task

When both an IPsec policy and a QoS policy are applied to an interface, QoS classifies packets by using the new headers added by IPsec. If you want QoS to classify packets by using the headers of the original IP packets, enable the QoS pre-classify feature.

Restrictions and guidelines

If you configure both IPsec and QoS on an interface, make sure the IPsec traffic classification rules match the QoS traffic classification rules. If the rules do not match, QoS might classify the packets of one IPsec SA to different queues, causing packets to be sent out of order. When IPsec anti-replay is enabled, IPsec will drop the incoming packets that are out of the anti-replay window, resulting in packet loss.

IPsec traffic classification rules are determined by the rules of the specified ACL. For more information about QoS policy and classification, see *ACL and QoS Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter IPsec policy view or IPsec policy template view.

   - Enter IPsec policy view.

     ipsec { **ipv6-policy** | **policy** } *policy-name seq-number* [**isakmp** | **manual**]

   - Enter IPsec policy template view.

     ipsec { **ipv6-policy-template** | **policy-template** } *template-name seq-number*

3. Enable QoS pre-classify.

   qos pre-classify

   By default, QoS pre-classify is disabled.

Configuring the DF bit of IPsec packets

About this task

Perform this task to configure the Don't Fragment (DF) bit in the new IP header of IPsec packets in one of the following ways:

- **clear**—Clears the DF bit in the new header.
- **set**—Sets the DF bit in the new header.

- **copy**—Copies the DF bit in the original IP header to the new IP header.

You can configure the DF bit in system view and interface view. The interface-view DF bit setting takes precedence over the system-view DF bit setting. If the interface-view DF bit setting is not configured, the interface uses the system-view DF bit setting.

Restrictions and guidelines for DF bit configuration for IPsec packets

The DF bit setting takes effect only in tunnel mode, and it changes the DF bit in the new IP header rather than the original IP header.

This feature is available only on Layer 3 interfaces.

Configure the same DF bit setting on the interfaces where the same IPsec policy bound to a source interface is applied.

If the DF bit is set, the devices on the path cannot fragment the IPsec packets. To prevent IPsec packets from being discarded, make sure the path MTU is larger than the IPsec packet size. As a best practice, clear the DF bit if you cannot make sure the path MTU is larger than the IPsec packet size.

Configuring the DF bit of IPsec packets on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the DF bit of IPsec packets on the interface.
   ipsec df-bit { **clear** | **copy** | **set** }
   By default, the interface uses the global DF bit setting.

Configuring the DF bit of IPsec packets globally

1. Enter system view.
   system-view
2. Configure the DF bit of IPsec packets globally.
   ipsec global-df-bit { **clear** | **copy** | **set** }
   By default, IPsec copies the DF bit in the original IP header to the new IP header.

Configuring IPsec RRI

Restrictions and guidelines

Enabling IPsec RRI for an IPsec policy deletes all existing IPsec SAs created by this IPsec policy. IPsec RRI creates static routes according to new IPsec SAs.

Disabling IPsec RRI for an IPsec policy deletes all existing IPsec SAs created by this IPsec policy and the associated static routes.

IPsec RRI is supported in both tunnel mode and transport mode.

If you change the preference value or tag value for an IPsec policy, the device deletes all IPsec SAs created by this IPsec policy, and the associated static routes. The change takes effect for future IPsec RRI-created static routes.

IPsec RRI does not generate a static route to a destination address to be protected if the destination address is not defined in the ACL used by an IPsec policy or an IPsec policy template. You must manually configure a static route to the destination address.

Procedure

1. Enter system view.

system-view

2. Enter IPsec policy view or IPsec policy template view.
   - Enter IPsec policy view.
     ipsec { **policy** | **ipv6-policy** } *policy-name seq-number* **isakmp**
   - Enter IPsec policy template view.
     ipsec { **ipv6-policy-template** | **policy-template** } *template-name seq-number*
3. Enable IPsec RRI.
   reverse-route dynamic
   By default, IPsec RRI is disabled.
4. (Optional.) Set the preference value for the static routes created by IPsec RRI.
   reverse-route preference *number*
   The default value is 60.
5. (Optional.) Set the tag value for the static routes created by IPsec RRI.
   reverse-route tag *tag-value*
   The default value is 0.

Configuring IPsec for IPv6 routing protocols

IPsec protection for IPv6 routing protocols tasks at a glance

To configure IPsec protection for IPv6 routing protocols, perform the following tasks:

1. Configuring an IPsec transform set
2. Configuring a manual IPsec profile
3. Applying the IPsec profile to an IPv6 routing protocol
4. (Optional.) Configuring IPsec fragmentation
5. (Optional.) Configuring SNMP notifications for IPsec

Configuring a manual IPsec profile

About this task

A manual IPsec profile specifies the IPsec transform set used for protecting data flows, and the SPIs and keys used by the SAs.

Restrictions and guidelines

When you configure a manual IPsec profile, make sure the IPsec profile configuration at both tunnel ends meets the following requirements:

- The IPsec transform set specified in the IPsec profile at the two tunnel ends must have the same security protocol, encryption and authentication algorithms, and packet encapsulation mode.
- The local inbound and outbound IPsec SAs must have the same SPI and key.
- The IPsec SAs on the devices in the same scope must have the same key. The scope is defined by protocols. For OSPFv3, the scope consists of OSPFv3 neighbors or an OSPFv3 area. For RIPng, the scope consists of directly-connected neighbors or a RIPng process. For BGP, the scope consists of BGP peers or a BGP peer group.
- The keys for the IPsec SAs at the two tunnel ends must be configured in the same format. For example, if the local end uses a key in hexadecimal format, the remote end must also use a key

in hexadecimal format. If you configure a key in both the character and the hexadecimal formats, only the most recent configuration takes effect.

- If you configure a key in character format for ESP, the device automatically generates an authentication key and an encryption key for ESP.

Procedure

1. Enter system view.
   system-view
2. Create a manual IPsec profile and enter its view.
   ipsec profile *profile-name* **manual**
   The **manual** keyword is not needed if you enter the view of an existing IPsec profile.
3. (Optional.) Configure a description for the IPsec profile.
   description *text*
   By default, no description is configured.
4. Specify an IPsec transform set.
   transform-set *transform-set-name*
   By default, no IPsec transform set is specified in an IPsec profile.
   The specified IPsec transform set must use the transport mode.
5. Configure an SPI for an SA.
   sa spi { **inbound** | **outbound** } { **ah** | **esp** } *spi-number*
   By default, no SPI is configured for an SA.
6. Configure keys for the IPsec SA.
   - Configure an authentication key in hexadecimal format for AH.
     sa hex-key authentication { **inbound** | **outbound** } **ah** { **cipher** | **simple** } *string*
   - Configure an authentication key in character format for AH.
     sa string-key { **inbound** | **outbound** } **ah** { **cipher** | **simple** } *string*
   - Configure a key in character format for ESP.
     sa string-key { **inbound** | **outbound** } **esp** { **cipher** | **simple** } *string*
   - Configure an authentication key in hexadecimal format for ESP.
     sa hex-key authentication { **inbound** | **outbound** } **esp** { **cipher** | **simple** }
   - Configure an encryption key in hexadecimal format for ESP.
     sa hex-key encryption { **inbound** | **outbound** } **esp** { **cipher** | **simple** } *string*

   By default, no keys are configured for the IPsec SA.
   Configure a key for the security protocol (AH, ESP, or both) you have specified.

Applying the IPsec profile to an IPv6 routing protocol

For information about the configuration procedure, see IPv6 BGP, OSPFv3, and RIPng configuration in *Layer 3—IP Routing Configuration Guide*.

Configuring the global IPsec SA lifetime and idle timeout

About this task

If the IPsec SA lifetime and idle timeout are not configured in an IPsec policy, IPsec policy template, or IPsec profile, the global settings are used.

When IKE negotiates IPsec SAs, it uses the local lifetime settings or those proposed by the peer, whichever are smaller.

An IPsec SA can have both a time-based lifetime and a traffic-based lifetime. The IPsec SA expires when either lifetime expires.

Procedure

1. Enter system view.

   system-view
2. Set the global IPsec SA lifetime or idle timeout.
   - Set the global IPsec SA lifetime.

     ipsec sa global-duration { **time-based** *seconds* | **traffic-based** *kilobytes* }

     By default, the time-based SA lifetime is 3600 seconds, and the traffic-based SA lifetime is 1843200 kilobytes.
   - Set the global SA idle timeout.

     ipsec sa idle-time *seconds*

     By default, the global IPsec SA idle timeout feature is disabled.

Configuring IPsec fragmentation

About this task

Perform this task to configure the device to fragment packets before or after IPsec encapsulation.

If you configure the device to fragment packets before IPsec encapsulation, the device predetermines the encapsulated packet size before the actual encapsulation. If the encapsulated packet size exceeds the MTU of the output interface, the device fragments the packets before encapsulation. If a packet's DF bit is set, the device drops the packet and sends an ICMP error message.

If you configure the device to fragment packets after IPsec encapsulation, the device directly encapsulates the packets and fragments the encapsulated packets in subsequent service modules.

Restrictions and guidelines

This feature takes effect on IPsec protected IPv4 packets.

Procedure

1. Enter system view.

   system-view
2. Configure IPsec fragmentation.

   ipsec fragmentation { **after-encryption** | **before-encryption** }

   By default, the device fragments packets before IPsec encapsulation.

Setting the maximum number of IPsec tunnels

Restrictions and guidelines

To maximize concurrent performance of IPsec when memory is sufficient, increase the maximum number of IPsec tunnels. To ensure service availability when memory is insufficient, decrease the maximum number of IPsec tunnels.

Procedure

1. Enter system view.
   system-view
2. Set the maximum number of IPsec tunnels.
   ipsec limit max-tunnel *tunnel-limit*
   The number of IPsec tunnels is not limited.

Configuring IPsec logging

About this task

IPsec logging provides you with records of important IPsec configuration, negotiation, and packet events.

- After you enable configuration event logging, the device generates log entries when configuration events such as adding or deleting IPsec policies occur. A configuration event log entry primarily includes the name of the IPsec policy being configured, template number, and the number of IPsec policies that have been configured.
- After you enable negotiation logging, the device generates a log entry when negotiation succeeds or fails, for example, when IPsec SA creation succeeds or fails. A negotiation log entry primarily includes the source and destination IP addresses, SPI value, and sequence number of the negotiation packets. If the device drops a negotiation packet, the log entry also includes the reason for the packet drop.
- After you enable IPsec packet logging, the device generates a log entry when it drops a packet because of failure to find an inbound SA, AH or ESP authentication failure, ESP encryption failure, or any other issue. An IPsec packet log entry primarily includes the source and destination IP addresses, SPI value, and sequence number of the negotiation packets. If the device drops a negotiation packet, the log entry also includes the reason for the packet drop.

If you execute this command multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Enable logging for IPsec packets.
   ipsec logging { **config** | **negotiation** | **packet** } **enable**
   By default, logging for IPsec packets is disabled.

Configuring SNMP notifications for IPsec

About this task

After you enable SNMP notifications for IPsec, the IPsec module notifies the NMS of important module events. The notifications are sent to the device's SNMP module. For the notifications to be

sent correctly, you must also configure SNMP on the device. For more information about SNMP notifications, see *Network Management and Monitoring Configuration Guide*.

To generate and output SNMP notifications for a specific IPsec failure or event type, perform the following tasks:

1. Enable SNMP notifications for IPsec globally.
2. Enable SNMP notifications for the failure or event type.

Procedure

1. Enter system view.

   system-view
2. Enable SNMP notifications for IPsec globally.

   snmp-agent trap enable ipsec global

   By default, SNMP notifications for IPsec are disabled.
3. Enable SNMP notifications for the specified failure or event types.

   snmp-agent trap enable ipsec [**auth-failure** | **decrypt-failure** | **encrypt-failure** | **invalid-sa-failure** | **no-sa-failure** | **policy-add** | **policy-attach** | **policy-delete** | **policy-detach** | **sa-establish** | **sa-establish-failure** | **sa-initiate** | **sa-terminate** | **tunnel-start** | **tunnel-stop**] *

   By default, SNMP notifications for all failure and event types are disabled.

Display and maintenance commands for IPsec

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display IPsec policy information.	**display ipsec** { **ipv6-policy** | **policy** } [*policy-name* [*seq-number*]]
Display IPsec policy template information.	**display ipsec** { **ipv6-policy-template** | **policy-template** } [*template-name* [*seq-number*]]
Display IPsec profile information.	**display ipsec profile** [*profile-name*]
Display IPsec SA information.	**display ipsec sa** [**brief** | **count** | **interface** *interface-type interface-number* | { **ipv6-policy** | **policy** } *policy-name* [*seq-number*] | **profile** *profile-name* | **remote** [**ipv6**] *ip-address*]
Display IPsec statistics.	**display ipsec statistics** [**tunnel-id** *tunnel-id*]
Display IPsec transform set information.	**display ipsec transform-set** [*transform-set-name*]
Display IPsec tunnel information.	**display ipsec tunnel** { **brief** | **count** | **tunnel-id** *tunnel-id* }

Task	Command
Clear IPsec SAs.	**reset ipsec sa** [{ **ipv6-policy** | **policy** } *policy-name* [*seq-number*] | **profile** *policy-name* | **remote** { *ipv4-address* | **ipv6** *ipv6-address* } | **spi** { *ipv4-address* | **ipv6** *ipv6-address* } { **ah** | **esp** } *spi-num*]
Clear IPsec statistics.	**reset ipsec statistics** [**tunnel-id** *tunnel-id*]

IPsec configuration examples

Example: Configuring a manual mode IPsec tunnel for IPv4 packets

Network configuration

As shown in Figure 7, establish an IPsec tunnel between Switch A and Switch B to protect data flows between the switches. Configure the tunnel as follows:

- Specify the encapsulation mode as tunnel, the security protocol as ESP, the encryption algorithm as AES-CBC-192, and the authentication algorithm as HMAC-SHA1.
- Manually set up IPsec SAs.

Figure 7 Network diagram



Procedure

1. Configure Switch A:

   # Assign IP addresses to interfaces, as shown in Figure 7.

   ```
   <Sysname> system-view
   [Sysname] sysname SwitchA
   [SwitchA] vlan 1
   [SwitchA-vlan1] quit
   [SwitchA] interface vlan-interface 1
   [SwitchA-Vlan-interface1] ip address 2.2.2.1 255.255.255.0
   [SwitchA-Vlan-interface1] quit
   [SwitchA] interface gigabitethernet 1/0/2
   [SwitchA-GigabitEthernet1/0/2] port access vlan 1
   [SwitchA-GigabitEthernet1/0/2] quit
   ```

   # Configure an IPv4 advanced ACL to identify the data flows between Switch A and Switch B.

   ```
   [SwitchA] acl advanced 3101
   [SwitchA-acl-ipv4-adv-3101] rule 0 permit ip source 2.2.2.1 0 destination 2.2.3.1 0
   [SwitchA-acl-ipv4-adv-3101] quit
   ```

   # Create an IPsec transform set named **tran1**.

```
[SwitchA] ipsec transform-set tran1
```

Specify the encapsulation mode as **tunnel**.

```
[SwitchA-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Specify the security protocol as **ESP**.

```
[SwitchA-ipsec-transform-set-tran1] protocol esp
```

Specify the ESP encryption and authentication algorithms.

```
[SwitchA-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchA-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchA-ipsec-transform-set-tran1] quit
```

Create a manual IPsec policy entry. Specify the policy name as **map1** and set the sequence number to 10.

```
[SwitchA] ipsec policy map1 10 manual
```

Specify ACL 3101.

```
[SwitchA-ipsec-policy-manual-map1-10] security acl 3101
```

Specify IPsec transform set **tran1**.

```
[SwitchA-ipsec-policy-manual-map1-10] transform-set tran1
```

Specify the remote IP address of the IPsec tunnel as 2.2.3.1.

```
[SwitchA-ipsec-policy-manual-map1-10] remote-address 2.2.3.1
```

Configure inbound and outbound SPIs for ESP.

```
[SwitchA-ipsec-policy-manual-map1-10] sa spi outbound esp 12345
[SwitchA-ipsec-policy-manual-map1-10] sa spi inbound esp 54321
```

Configure the inbound and outbound SA keys for ESP.

```
[SwitchA-ipsec-policy-manual-map1-10] sa string-key outbound esp simple abcdefg
[SwitchA-ipsec-policy-manual-map1-10] sa string-key inbound esp simple gfedcba
[SwitchA-ipsec-policy-manual-map1-10] quit
```

Apply IPsec policy **map1** to VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ipsec apply policy map1
```

2. Configure Switch B:

Assign IP addresses to interfaces, as shown in Figure 7.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 1
[SwitchB-vlan1] quit
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 2.2.3.1 255.255.255.0
[SwitchB-Vlan-interface1] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port access vlan 1
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure an IPv4 advanced ACL to identify the data flows between Switch B and Switch A.

```
[SwitchB] acl advanced 3101
[SwitchB-acl-ipv4-adv-3101] rule 0 permit ip source 2.2.3.1 0 destination 2.2.2.1 0
[SwitchB-acl-ipv4-adv-3101] quit
```

Create an IPsec transform set named **tran1**.

```
[SwitchB] ipsec transform-set tran1
```

Specify the encapsulation mode as **tunnel**.

```
[SwitchB-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Specify the security protocol as **ESP**.

```
[SwitchB-ipsec-transform-set-tran1] protocol esp
```

Specify the ESP encryption and authentication algorithms.

```
[SwitchB-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchB-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran1] quit
```

Create a manual IPsec policy entry. Specify the policy name as **use1** and set the sequence number to 10.

```
[SwitchB] ipsec policy use1 10 manual
```

Specify ACL 3101.

```
[SwitchB-ipsec-policy-manual-use1-10] security acl 3101
```

Specify IPsec transform set **tran1**.

```
[SwitchB-ipsec-policy-manual-use1-10] transform-set tran1
```

Specify the remote IP address of the IPsec tunnel as 2.2.2.1.

```
[SwitchB-ipsec-policy-manual-use1-10] remote-address 2.2.2.1
```

Configure the inbound and outbound SPIs for ESP.

```
[SwitchB-ipsec-policy-manual-use1-10] sa spi outbound esp 54321
[SwitchB-ipsec-policy-manual-use1-10] sa spi inbound esp 12345
```

Configure the inbound and outbound SA keys for ESP.

```
[SwitchB-ipsec-policy-manual-use1-10] sa string-key outbound esp simple gfedcba
[SwitchB-ipsec-policy-manual-use1-10] sa string-key inbound esp simple abcdefg
[SwitchB-ipsec-policy-manual-use1-10] quit
```

Apply IPsec policy **use1** to VLAN-interface 1.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ipsec apply policy use1
```

Verifying the configuration

After the configuration is completed, an IPsec tunnel between Switch A and Switch B is established, and the traffic between the switches is IPsec-protected. This example uses Switch A to verify the configuration.

Use the **display ipsec sa** command to display IPsec SAs on Switch A.

```
[SwitchA] display ipsec sa
-------------------------------
Interface: Vlan-interface 1
-------------------------------

  -----------------------------
  IPsec policy: map1
  Sequence number: 10
  Mode: manual
  -----------------------------
    Tunnel id: 549
    Encapsulation mode: tunnel
    Path MTU: 1443
    Tunnel:
        local  address: 2.2.2.1
        remote address: 2.2.3.1
```

```
    Flow:
        as defined in ACL 3101
    [Inbound ESP SA]
      SPI: 54321 (0x0000d431)
      Transform set: ESP-ENCRYPT-AES-CBC-192 ESP-AUTH-SHA1
      No duration limit for this SA
    [Outbound ESP SA]
      SPI: 12345 (0x00003039)
      Transform set: ESP-ENCRYPT-AES-CBC-192 ESP-AUTH-SHA1
      No duration limit for this SA
```

Configuration files

Switch A

```
#
sysname SwitchA
#
vlan 1
#
interface vlan-interface 1
ip address 2.2.2.1 255.255.255.0
#
interface gigabitethernet 1/0/2
port access vlan 1
#
acl advanced 3101
rule 0 permit ip source 2.2.2.1 0 destination 2.2.3.1 0
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
#
ipsec policy map1 10 manual
security acl 3101
transform-set tran1
remote-address 2.2.3.1
sa spi outbound esp 12345
sa spi inbound esp 54321
sa string-key outbound esp simple abcdefg
sa string-key inbound esp simple gfedcba
#
interface vlan-interface 1
ipsec apply policy map1
#
```

Switch B

```
#
sysname SwitchB
#
```

```
vlan 1
#
interface vlan-interface 1
ip address 2.2.3.1 255.255.255.0
#
interface gigabitethernet 1/0/2
port access vlan 1
#
acl advanced 3101
rule 0 permit ip source 2.2.3.1 0 destination 2.2.2.1 0
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
#
ipsec policy use1 10 manual
security acl 3101
transform-set tran1
remote-address 2.2.2.1
sa spi outbound esp 54321
sa spi inbound esp 12345
sa string-key outbound esp simple gfedcba
sa string-key inbound esp simple abcdefg
#
interface vlan-interface 1
ipsec apply policy use1
#
```

Example: Configuring an IKE-based IPsec tunnel for IPv4 packets

Network configuration

As shown in Figure 8, establish an IPsec tunnel between Switch A and Switch B to protect data flows between them. Configure the IPsec tunnel as follows:

- Specify the encapsulation mode as tunnel, the security protocol as ESP, the encryption algorithm as AES-CBC-192, and the authentication algorithm as HMAC-SHA1.
- Set up SAs through IKE negotiation.

Figure 8 Network diagram



Procedure

1. Configure Switch A:

Assign IP addresses to interfaces, as shown in Figure 8.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 1
[SwitchA-vlan1] quit
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 2.2.2.1 255.255.255.0
[SwitchA-Vlan-interface1] quit
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port access vlan 1
[SwitchA-GigabitEthernet1/0/2] quit
```

Configure an IPv4 advanced ACL to identify data flows from Switch A to Switch B.

```
[SwitchA] acl number 3101
[SwitchA-acl-adv-3101] rule 0 permit ip source 2.2.2.1 0 destination 2.2.3.1 0
[SwitchA-acl-adv-3101] quit
```

Create an IPsec transform set named **tran1**.

```
[SwitchA] ipsec transform-set tran1
```

Specify the encapsulation mode as **tunnel**.

```
[SwitchA-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Specify the security protocol as **ESP**.

```
[SwitchA-ipsec-transform-set-tran1] protocol esp
```

Specify the ESP encryption and authentication algorithms.

```
[SwitchA-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchA-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchA-ipsec-transform-set-tran1] quit
```

Create an IKE keychain named **keychain1**.

```
[SwitchA] ike keychain keychain1
```

Specify **12345zxcvb!@#$%ZXCVB** in plain text as the preshared key to be used with the remote peer at 2.2.3.1.

```
[SwitchA-ike-keychain-keychain1] pre-shared-key address 2.2.3.1 255.255.255.0 key
simple 12345zxcvb!@#$%ZXCVB
[SwitchA-ike-keychain-keychain1] quit
```

Create and configure the IKE profile named **profile1**.

```
[SwitchA] ike profile profile1
[SwitchA-ike-profile-profile1] keychain keychain1
[SwitchA-ike-profile-profile1] match remote identity address 2.2.3.1 255.255.255.0
[SwitchA-ike-profile-profile1] quit
```

Create an IKE-based IPsec policy entry. Specify the policy name as **map1** and set the sequence number to 10.

```
[SwitchA] ipsec policy map1 10 isakmp
```

Apply ACL 3101.

```
[SwitchA-ipsec-policy-isakmp-map1-10] security acl 3101
```

Apply IPsec transform set **tran1**.

```
[SwitchA-ipsec-policy-isakmp-map1-10] transform-set tran1
```

Specify the local and remote IP addresses of the IPsec tunnel as 2.2.2.1 and 2.2.3.1.

```
[SwitchA-ipsec-policy-isakmp-map1-10] local-address 2.2.2.1
[SwitchA-ipsec-policy-isakmp-map1-10] remote-address 2.2.3.1
```

Apply IKE profile **profile1**.

```
[SwitchA-ipsec-policy-isakmp-map1-10] ike-profile profile1
[SwitchA-ipsec-policy-isakmp-map1-10] quit
```

Apply IPsec policy **map1** to VLAN interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ipsec apply policy map1
```

2. Configure Switch B:

Assign IP addresses to interfaces, as shown in Figure 8.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 1
[SwitchB-vlan1] quit
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 2.2.3.1 255.255.255.0
[SwitchB-Vlan-interface1] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port access vlan 1
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure an IPv4 advanced ACL to identify data flows from Switch B to Switch A.

```
[SwitchB] acl number 3101
[SwitchB-acl-adv-3101] rule 0 permit ip source 2.2.3.1 0 destination 2.2.2.1 0
[SwitchB-acl-adv-3101] quit
```

Create an IPsec transform set named **tran1**.

```
[SwitchB] ipsec transform-set tran1
```

Specify the encapsulation mode as **tunnel**.

```
[SwitchB-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Specify the security protocol as **ESP**.

```
[SwitchB-ipsec-transform-set-tran1] protocol esp
```

Specify the ESP encryption and authentication algorithms.

```
[SwitchB-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchB-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran1] quit
```

Create an IKE keychain named **keychain1**.

```
[SwitchB] ike keychain keychain1
```

Specify **12345zxcvb!@#$%ZXCVB** in plain text as the preshared key to be used with the remote peer at 2.2.2.1.

```
[SwitchB-ike-keychain-keychain1] pre-shared-key address 2.2.2.1 255.255.255.0 key
simple 12345zxcvb!@#$%ZXCVB
[SwitchB-ike-keychain-keychain1] quit
```

Create and configure the IKE profile named **profile1**.

```
[SwitchB] ike profile profile1
[SwitchB-ike-profile-profile1] keychain keychain1
[SwitchB-ike-profile-profile1] match remote identity address 2.2.2.1 255.255.255.0
[SwitchB-ike-profile-profile1] quit
```

Create an IKE-based IPsec policy entry. Specify the policy name as **use1** and set the sequence number to 10.

```
[SwitchB] ipsec policy use1 10 isakmp
```

Apply ACL 3101.

```
[SwitchB-ipsec-policy-isakmp-use1-10] security acl 3101
```

Apply IPsec transform set **tran1**.

```
[SwitchB-ipsec-policy-isakmp-use1-10] transform-set tran1
```

Specify the local and remote IP addresses of the IPsec tunnel as 2.2.3.1 and 2.2.2.1.

```
[SwitchB-ipsec-policy-isakmp-use1-10] local-address 2.2.3.1
[SwitchB-ipsec-policy-isakmp-use1-10] remote-address 2.2.2.1
```

Apply IKE profile **profile1**.

```
[SwitchB-ipsec-policy-isakmp-use1-10] ike-profile profile1
[SwitchB-ipsec-policy-isakmp-use1-10] quit
```

Apply IPsec policy **use1** to VLAN-interface 1.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ipsec apply policy use1
```

Verifying the configuration

Initiate a connection between Switch A and Switch B to trigger IKE negotiation. After IPsec SAs are successfully negotiated by IKE, the traffic between the two switches is IPsec-protected.

Configuration files

Switch A

```
#
sysname SwitchA
#
vlan 1
quit
interface vlan-interface 1
ip address 2.2.2.1 255.255.255.0
quit
interface gigabitethernet 1/0/2
port access vlan 1
quit
acl number 3101
rule 0 permit ip source 2.2.2.1 0 destination 2.2.3.1 0
quit
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
quit
ike keychain keychain1
pre-shared-key address 2.2.3.1 255.255.255.0 key simple 12345zxcvb!@#$%ZXCVB
quit
ike profile profile1
keychain keychain1
match remote identity address 2.2.3.1 255.255.255.0
quit
ipsec policy map1 10 isakmp
security acl 3101
transform-set tran1
local-address 2.2.2.1
```

```
remote-address 2.2.3.1
ike-profile profile1
quit
interface vlan-interface 1
ipsec apply policy map1
#
```

Switch B

```
#
sysname SwitchB
#
vlan 1
#
interface vlan-interface 1
ip address 2.2.3.1 255.255.255.0
quit
interface gigabitethernet 1/0/2
port access vlan 1
#
acl number 3101
rule 0 permit ip source 2.2.3.1 0 destination 2.2.2.1 0
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
#
ike keychain keychain1
pre-shared-key address 2.2.2.1 255.255.255.0 key simple 12345zxcvb!@#$%ZXCVB
#
ike profile profile1
keychain keychain1
match remote identity address 2.2.2.1 255.255.255.0
#
ipsec policy use1 10 isakmp
security acl 3101
transform-set tran1
local-address 2.2.3.1
remote-address 2.2.2.1
ike-profile profile1
#
interface vlan-interface 1
ipsec apply policy use1
#
```

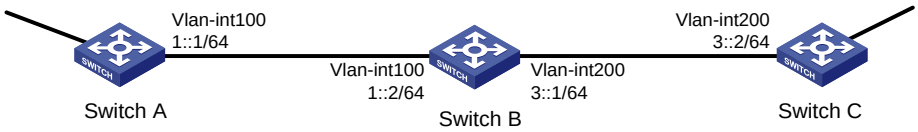
Example: Configuring IPsec for RIPng

Network configuration

As shown in Figure 9, Switch A, Switch B, and Switch C learn IPv6 routes through RIPng.

Establish an IPsec tunnel between the switches to protect the RIPng packets transmitted in between. Specify the security protocol as ESP, the encryption algorithm as 128-bit AES, and the authentication algorithm as HMAC-SHA1 for the IPsec tunnel.

Figure 9 Network diagram



Requirements analysis

To meet the network configuration requirements, perform the following tasks:

1. Configure basic RIPng.

   For more information about RIPng configurations, see *Layer 3—IP Routing Configuration Guide*.

2. Configure an IPsec profile.
   - The IPsec profiles on all the switches must have IPsec transform sets that use the same security protocol, authentication and encryption algorithms, and encapsulation mode.
   - The SPI and key configured for the inbound SA and those for the outbound SA must be the same on each switch.
   - The SPI and key configured for the SAs on all the switches must be the same.
3. Apply the IPsec profile to a RIPng process or to an interface.

Procedure

1. Configure Switch A:

   # Assign IP addresses to interfaces,as shown in Figure 9.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 100
[SwitchA-vlan100] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipv6 address 1::1/64
[SwitchA-Vlan-interface100] quit
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port access vlan 100
[SwitchA-GigabitEthernet1/0/1] quit
```

   # Configure basic RIPng.

```
[SwitchA] ripng 1
[SwitchA-ripng-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ripng 1 enable
[SwitchA-Vlan-interface100] quit
```

   # Create and configure the IPsec transform set named **tran1**.

```
[SwitchA] ipsec transform-set tran1
```

```
[SwitchA-ipsec-transform-set-tran1] encapsulation-mode transport
[SwitchA-ipsec-transform-set-tran1] protocol esp
[SwitchA-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-128
[SwitchA-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchA-ipsec-transform-set-tran1] quit
```

Create and configure the IPsec profile named **profile001**.

```
[SwitchA] ipsec profile profile001 manual
[SwitchA-ipsec-profile-manual-profile001] transform-set tran1
[SwitchA-ipsec-profile-manual-profile001] sa spi outbound esp 123456
[SwitchA-ipsec-profile-manual-profile001] sa spi inbound esp 123456
[SwitchA-ipsec-profile-manual-profile001] sa string-key outbound esp simple abcdefg
[SwitchA-ipsec-profile-manual-profile001] sa string-key inbound esp simple abcdefg
[SwitchA-ipsec-profile-manual-profile001] quit
```

Apply the IPsec profile to RIPng process 1.

```
[SwitchA] ripng 1
[SwitchA-ripng-1] enable ipsec-profile profile001
[SwitchA-ripng-1] quit
```

2. Configure Switch B:

Assign IP addresses to interfaces,as shown in Figure 9.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 100
[SwitchB-vlan100] quit
[SwitchB] vlan 200
[SwitchB-vlan200] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ipv6 address 3::1/64
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ipv6 address 1::2/64
[SwitchB-Vlan-interface100] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port access vlan 100
[SwitchB-GigabitEthernet1/0/1] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port access vlan 200
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure basic RIPng.

```
[SwitchB] ripng 1
[SwitchB-ripng-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ripng 1 enable
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit
```

Create and configure the IPsec transform set named **tran1**.

```
[SwitchB] ipsec transform-set tran1
```

```
[SwitchB-ipsec-transform-set-tran1] encapsulation-mode transport
[SwitchB-ipsec-transform-set-tran1] protocol esp
[SwitchB-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-128
[SwitchB-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran1] quit
```

Create and configure the IPsec profile named **profile001**.

```
[SwitchB] ipsec profile profile001 manual
[SwitchB-ipsec-profile-manual-profile001] transform-set tran1
[SwitchB-ipsec-profile-manual-profile001] sa spi outbound esp 123456
[SwitchB-ipsec-profile-manual-profile001] sa spi inbound esp 123456
[SwitchB-ipsec-profile-manual-profile001] sa string-key outbound esp simple abcdefg
[SwitchB-ipsec-profile-manual-profile001] sa string-key inbound esp simple abcdefg
[SwitchB-ipsec-profile-manual-profile001] quit
```

Apply the IPsec profile to RIPng process 1.

```
[SwitchB] ripng 1
[SwitchB-ripng-1] enable ipsec-profile profile001
[SwitchB-ripng-1] quit
```

3. Configure Switch C:

Assign IP addresses to interfaces,as shown in Figure 9.

```
<Sysname> system-view
[Sysname] sysname SwitchC
[SwitchC] vlan 100
[SwitchC-vlan100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] ipv6 address 3::2/64
[SwitchC-Vlan-interface200] quit
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] port access vlan 200
[SwitchC-GigabitEthernet1/0/1] quit
```

Configure basic RIPng.

```
[SwitchC] ripng 1
[SwitchC-ripng-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] ripng 1 enable
[SwitchC-Vlan-interface200] quit
```

Create and configure the IPsec transform set named **tran1**.

```
[SwitchC] ipsec transform-set tran1
[SwitchC-ipsec-transform-set-tran1] encapsulation-mode transport
[SwitchC-ipsec-transform-set-tran1] protocol esp
[SwitchC-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-128
[SwitchC-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchC-ipsec-transform-set-tran1] quit
```

Create and configure the IPsec profile named **profile001**.

```
[SwitchC] ipsec profile profile001 manual
[SwitchC-ipsec-profile-manual-profile001] transform-set tran1
[SwitchC-ipsec-profile-manual-profile001] sa spi outbound esp 123456
[SwitchC-ipsec-profile-manual-profile001] sa spi inbound esp 123456
[SwitchC-ipsec-profile-manual-profile001] sa string-key outbound esp simple abcdefg
```

```
[SwitchC-ipsec-profile-manual-profile001] sa string-key inbound esp simple abcdefg
[SwitchC-ipsec-profile-manual-profile001] quit
```

Apply the IPsec profile to RIPng process 1.

```
[SwitchC] ripng 1
[SwitchC-ripng-1] enable ipsec-profile profile001
[SwitchC-ripng-1] quit
```

Verifying the configuration

After the configuration is completed, Switch A, Switch B, and Switch C learn IPv6 routing information through RIPng. IPsec SAs are set up successfully on the switches to protect RIPng packets. This example uses Switch A to verify the configuration.

Display the RIPng configuration. The output shows that IPsec profile **profile001** has been applied to RIPng process 1.

```
[SwitchA] display ripng 1
    RIPng process : 1
       Preference : 100
       Checkzero : Enabled
       Default Cost : 0
       Maximum number of load balanced routes : 8
       Update time   :   30 secs  Timeout time         :  180 secs
       Suppress time :  120 secs  Garbage-Collect time :  120 secs
       Update output delay:   20(ms)  Output count:    3
       Graceful-restart interval:   60 secs
       Triggered Interval : 5 50 200
       Number of periodic updates sent : 186
       Number of triggered updates sent : 1
       IPsec profile name: profile001
```

Display the established IPsec SAs.

```
[SwitchA] display ipsec sa
------------------------------
Global IPsec SA
------------------------------

  ----------------------------
  IPsec profile: profile001
  Mode: Manual
  ----------------------------
    Encapsulation mode: transport
    [Inbound ESP SA]
      SPI: 123456 (0x3039)
      Connection ID: 90194313219
      Transform set: ESP-ENCRYPT-AES-CBC-128 ESP-AUTH-SHA1
      No duration limit for this SA
    [Outbound ESP SA]
      SPI: 123456 (0x3039)
      Connection ID: 64424509441
      Transform set: ESP-ENCRYPT-AES-CBC-128ESP-AUTH-SHA1
      No duration limit for this SA
```

Configuration files

Switch A

```
#
sysname SwitchA
#
vlan 100
#
interface vlan-interface 100
 ipv6 address 1::1/64
#
interface gigabitethernet 1/0/1
 port access vlan 100
#
ripng 1
#
interface gigabitethernet 1/0/1
ripng 1 enable
#
ipsec transform-set tran1
encapsulation-mode transport
protocol esp
esp encryption-algorithm aes-cbc-128
esp authentication-algorithm sha1
#
ipsec profile profile001 manual
transform-set tran1
sa spi outbound esp 123456
sa spi inbound esp 123456
sa string-key outbound esp simple abcdefg
sa string-key inbound esp simple abcdefg
#
ripng 1
enable ipsec-profile profile001
#
```

Switch B

```
#
sysname SwitchB
#
vlan 100
#
vlan 200
#
interface vlan-interface200
 ipv6 address 3::1/64
#
interface vlan-interface100
 ipv6 address 1::2/64
#
```

```
interface GigabitEthernet1/0/1
 port access vlan 100
#
interface GigabitEthernet1/0/2
 port access vlan 200
#
ripng 1
#
interface gigabitethernet 1/0/1
ripng 1 enable
#
interface gigabitethernet 1/0/2
ripng 1 enable
#
ipsec transform-set tran1
encapsulation-mode transport
protocol esp
esp encryption-algorithm aes-cbc-128
esp authentication-algorithm sha1
#
ipsec profile profile001 manual
transform-set tran1
sa spi outbound esp 123456
sa spi inbound esp 123456
sa string-key outbound esp simple abcdefg
sa string-key inbound esp simple abcdefg
#
ripng 1
enable ipsec-profile profile001
#
```

Switch C

```
#
sysname SwitchC
#
vlan 100
#
interface vlan-interface 200
 ipv6 address 3::2/64
#
interface gigabitethernet 1/0/1
 port access vlan 200
#
ripng 1
#
interface gigabitethernet 1/0/1
ripng 1 enable
#
ipsec transform-set tran1
```

```
encapsulation-mode transport
protocol esp
esp encryption-algorithm aes-cbc-128
esp authentication-algorithm sha1
#
ipsec profile profile001 manual
transform-set tran1
sa spi outbound esp 123456
sa spi inbound esp 123456
sa string-key outbound esp simple abcdefg
sa string-key inbound esp simple abcdefg
#
ripng 1
enable ipsec-profile profile001
#
```

Example: Configuring IPsec RRI

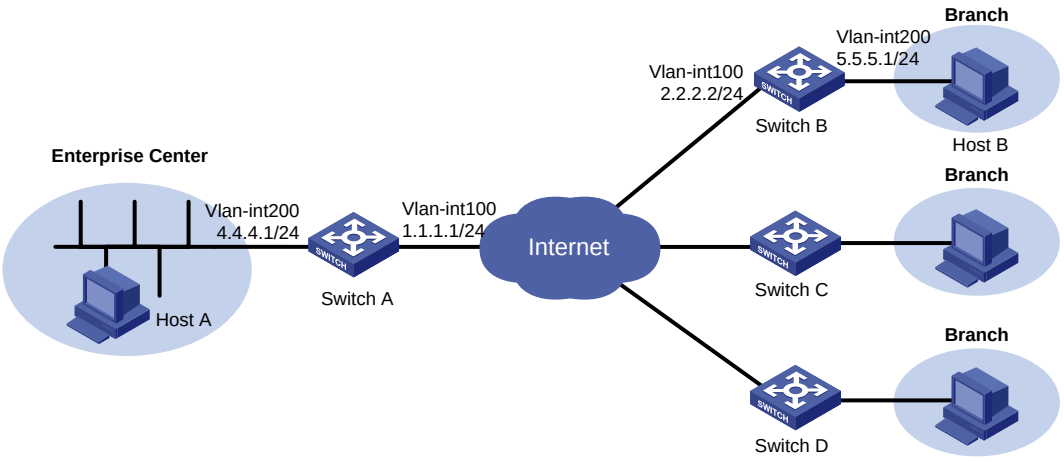
Network configuration

As shown in Figure 10, branches access the enterprise center through an IPsec VPN.

Configure the IPsec VPN as follows:

- Configure an IPsec tunnel between Switch A and each branch gateway (Switch B, Switch C, and Switch D) to protect traffic between subnets 4.4.4.0/24 and 5.5.5.0/24.
- Configure the tunnels to use security protocol ESP, encryption algorithm DES, and authentication algorithm SHA1-HMAC-96. Use IKE for IPsec SA negotiation.
- Configure IKE proposal to use the preshared key authentication method, encryption algorithm 3DES, and authentication algorithm HMAC-SHA1.
- Configure IPsec RRI on Switch A to automatically create static routes to the branches based on the established IPsec SAs.

Figure 10 Network diagram



Procedure

1. Configure Switch A:

   # Assign IP addresses to interfaces, as shown in Figure 10.

```
<Sysname> system-view
[Sysname] sysname SwitchA
[SwitchA] vlan 100
[SwitchA-vlan100] quit
[SwitchA] vlan 200
[SwitchA-vlan200] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ip address 4.4.4.1 255.255.255.0
[SwitchA-Vlan-interface200] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 1.1.1.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port access vlan 100
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port access vlan 200
[SwitchA-GigabitEthernet1/0/2] quit
```

Configure settings for routing. This example configures a static route, and the next hop in the route is 1.1.1.2.

```
[SwitchA] ip route-static 2.2.2.0 24 1.1.1.2
```

Create an IPsec transform set named **tran1**, and specify **ESP** as the security protocol, **DES** as the encryption algorithm, and **HMAC-SHA-1-96** as the authentication algorithm.

```
[SwitchA] ipsec transform-set tran1
[SwitchA-ipsec-transform-set-tran1] encapsulation-mode tunnel
[SwitchA-ipsec-transform-set-tran1] protocol esp
[SwitchA-ipsec-transform-set-tran1] esp encryption-algorithm des
[SwitchA-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchA-ipsec-transform-set-tran1] quit
```

Create and configure the IKE profile named **profile1**.

```
[SwitchA] ike profile profile1
[SwitchA-ike-profile-profile1] keychain key1
[SwitchA-ike-profile-profile1] match remote identity address 2.2.2.2 255.255.255.0
[SwitchA-ike-profile-profile1] quit
```

Create an IPsec policy template named **temp1**. Specify IPsec transform set **tran1** and IKE profile **profile1** for the IPsec policy template.

```
[SwitchA] ipsec policy-template temp1 1
[SwitchA-ipsec-policy-template-temp1-1] transform-set tran1
[SwitchA-ipsec-policy-template-temp1-1] ike-profile profile1
```

Enable IPsec RRI, set the preference to 100 and the tag to 1000 for the static routes created by IPsec RRI.

```
[SwitchA-ipsec-policy-template-temp1-1] reverse-route dynamic
[SwitchA-ipsec-policy-template-temp1-1] reverse-route preference 100
[SwitchA-ipsec-policy-template-temp1-1] reverse-route tag 1000
[SwitchA-ipsec-policy-template-temp1-1] quit
```

Create an IKE-based IPsec policy entry by using IPsec policy template **temp1**. Specify the policy name as **map1** and set the sequence number to 10.

```
[SwitchA] ipsec policy map1 10 isakmp template temp1
```

Create an IKE proposal named **1**, and specify **3DES** as the encryption algorithm, **HMAC-SHA1** as the authentication algorithm, and **pre-share** as the authentication method.

```
[SwitchA] ike proposal 1
[SwitchA-ike-proposal-1] encryption-algorithm 3des-cbc
[SwitchA-ike-proposal-1] authentication-algorithm sha
[SwitchA-ike-proposal-1] authentication-method pre-share
[SwitchA-ike-proposal-1] quit
```

Create an IKE keychain named **key1** and specify **123** in plain text as the preshared key to be used with the remote peer at 2.2.2.2.

```
[SwitchA] ike keychain key1
[SwitchA-ike-keychain-key1] pre-shared-key address 2.2.2.2 key simple 123
[SwitchA-ike-keychain-key1] quit
```

Apply IPsec policy **map1** to VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipsec apply policy map1
[SwitchA-Vlan-interface100] quit
```

2. Configure Switch B:

Assign IP addresses to interfaces, as shown in Figure 10.

```
<Sysname> system-view
[Sysname] sysname SwitchB
[SwitchB] vlan 100
[SwitchB-vlan100] quit
[SwitchB] vlan 200
[SwitchB-vlan200] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ip address 5.5.5.1 255.255.255.0
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 2.2.2.2 255.255.255.0
[SwitchB-Vlan-interface100] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port access vlan 100
[SwitchB-GigabitEthernet1/0/1] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port access vlan 200
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure static routes. In this example, the next hop in the routes is 2.2.2.3.

```
[SwitchB] ip route-static 1.1.1.0 24 2.2.2.3
[SwitchB] ip route-static 4.4.4.0 24 2.2.2.3
```

Create an IPsec transform set named **tran1**, and specify **ESP** as the security protocol, **DES** as the encryption algorithm, and **HMAC-SHA-1-96** as the authentication algorithm.

```
[SwitchB] ipsec transform-set tran1
[SwitchB-ipsec-transform-set-tran1] encapsulation-mode tunnel
[SwitchB-ipsec-transform-set-tran1] protocol esp
[SwitchB-ipsec-transform-set-tran1] esp encryption-algorithm des
[SwitchB-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran1] quit
```

Configure IPv4 advanced ACL 3000 to identify traffic from subnet 5.5.5.0/24 to subnet 4.4.4.0/24.

```
[SwitchB] acl advanced 3000
[SwitchB-acl-ipv4-adv-3000] rule permit ip source 5.5.5.0 0.0.0.255 destination 4.4.4.0 0.0.0.255
[SwitchB-acl-ipv4-adv-3000] quit
```

Create and configure the IKE profile named **profile1**.

```
[SwitchB] ike profile profile1
[SwitchB-ike-profile-profile1] keychain key1
[SwitchB-ike-profile-profile1] match remote identity address 1.1.1.1 255.255.255.0
[SwitchB-ike-profile-profile1] quit
```

Create an IKE-based IPsec policy entry named **map1** and configure the following settings for the policy entry:

- Set the sequence number to 10.
- Specify transform set **tran1** and ACL 3000.
- Specify the remote IP address for the tunnel as 1.1.1.1.
- Specify IKE profile **profile1**.

```
[SwitchB] ipsec policy map1 10 isakmp
[SwitchB-ipsec-policy-isakmp-map1-10] transform-set tran1
[SwitchB-ipsec-policy-isakmp-map1-10] security acl 3000
[SwitchB-ipsec-policy-isakmp-map1-10] remote-address 1.1.1.1
[SwitchB-ipsec-policy-isakmp-map1-10] ike-profile profile1
[SwitchB-ipsec-policy-isakmp-map1-10] quit
```

Create an IKE proposal named **1**, and specify **3DES** as the encryption algorithm, **HMAC-SHA1** as the authentication algorithm, and **pre-share** as the authentication method.

```
[SwitchB] ike proposal 1
[SwitchB-ike-proposal-1] encryption-algorithm 3des-cbc
[SwitchB-ike-proposal-1] authentication-algorithm sha
[SwitchB-ike-proposal-1] authentication-method pre-share
[SwitchB-ike-proposal-1] quit
```

Create an IKE keychain named **key1** and specify **123** in plain text as the preshared key to be used with the remote peer at 1.1.1.1.

```
[SwitchB] ike keychain key1
[SwitchB-ike-keychain-key1] pre-shared-key address 1.1.1.1 key simple 123
[SwitchB-ike-keychain-key1] quit
```

Apply IPsec policy **map1** to VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ipsec apply policy map1
[SwitchB-Vlan-interface100] quit
```

Make sure Switch B has a route to the peer private network, with the outgoing interface as VLAN-interface 100.

3. Configure Switch C and Switch D in the same way Switch B is configured.

Verifying the configuration

1. Verify that IPsec RRI can automatically create a static route from Switch A to Switch B:

   # Initiate a connection from subnet 5.5.5.0/24 to subnet 4.4.4.0/24. IKE negotiation is triggered to establish IPsec SAs between Switch A and Switch B. (Details not shown.)

   # Verify that IPsec SAs are established on Switch A.

```
[SwitchA] display ipsec sa
```

```
------------------------------
Interface: Vlan-interface100
------------------------------

  ----------------------------
  IPsec policy: map1
  Sequence number: 10
  Mode: Template
  ----------------------------
    Tunnel id: 0
    Encapsulation mode: tunnel
    Perfect Forward Secrecy:
    Inside VPN:
    Extended Sequence Numbers enable: N
    Traffic Flow Confidentiality enable: N
    Path MTU: 1463
    Tunnel:
        local  address: 1.1.1.1
        remote address: 2.2.2.2
    Flow:
        sour addr: 4.4.4.0/255.255.255.0  port: 0  protocol: ip
        dest addr: 5.5.5.0/255.255.255.0  port: 0  protocol: ip

    [Inbound ESP SAs]
      SPI: 1014286405 (0x3c74c845)
      Connection ID: 90194313219
      Transform set: ESP-ENCRYPT-DES-CBC ESP-AUTH-SHA1
      SA duration (kilobytes/sec): 1843200/3600
      SA remaining duration (kilobytes/sec): 1843199/3590
      Max received sequence-number: 4
      Anti-replay check enable: Y
      Anti-replay window size: 64
      UDP encapsulation used for NAT traversal: N
      Status: Active

    [Outbound ESP SAs]
      SPI: 4011716027 (0xef1dedbb)
      Connection ID: 64424509441
      Transform set: ESP-ENCRYPT-DES-CBC ESP-AUTH-SHA1
      SA duration (kilobytes/sec): 1843200/3600
      SA remaining duration (kilobytes/sec): 1843199/3590
      Max sent sequence-number: 4
      UDP encapsulation used for NAT traversal: N
      Status: Active
```

Verify that IPsec RRI has created a static route to reach Switch B.

```
[SwitchA] display ip routing-table verbose
```

2. Verify that Switch A can automatically create static routes to Switch C and Switch D in the same way that you verify the IPsec RRI feature by using Switch A and Switch B. (Details not shown.)

Configuration files

Switch A

```
#
sysname SwitchA
#
vlan 100
#
vlan 200
#
interface vlan-interface 200
ip address 4.4.4.1 255.255.255.0
#
interface vlan-interface 100
ip address 1.1.1.1 255.255.255.0
#
interface gigabitethernet 1/0/1
port access vlan 100
#
interface gigabitethernet 1/0/2
port access vlan 200
#
ip route-static 2.2.2.0 24 1.1.1.2
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm des
esp authentication-algorithm sha1
#
ike profile profile1
keychain key1
match remote identity address 2.2.2.2 255.255.255.0
#
ipsec policy-template temp1 1
transform-set tran1
ike-profile profile1
reverse-route dynamic
reverse-route preference 100
reverse-route tag 1000
#
ipsec  policy map1 10 isakmp template temp1
ike proposal 1
encryption-algorithm 3des-cbc
authentication-algorithm sha
authentication-method pre-share
#
ike keychain key1
pre-shared-key address 2.2.2.2 key simple 123
```

```
#
interface gigabitethernet 1/0/1
ipsec apply policy map1
#
```

Switch B

```
#
sysname SwitchB
#
vlan 100
#
vlan 200
#
interface vlan-interface 200
ip address 5.5.5.1 255.255.255.0
#
interface vlan-interface 100
ip address 2.2.2.2 255.255.255.0
#
interface gigabitethernet 1/0/1
port access vlan 100
#
interface gigabitethernet 1/0/2
port access vlan 200
#
ip route-static 1.1.1.0 24 2.2.2.3
ip route-static 4.4.4.0 24 2.2.2.3
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm des
esp authentication-algorithm sha1
#
acl advanced 3000
rule permit ip source 5.5.5.0 0.0.0.255 destination 4.4.4.0 0.0.0.255
#
ike profile profile1
keychain key1
match remote identity address 1.1.1.1 255.255.255.0
#
ipsec policy map1 10 isakmp
transform-set tran1
security acl 3000
remote-address 1.1.1.1
ike-profile profile1
#
ike proposal 1
encryption-algorithm 3des-cbc
```

```
authentication-algorithm sha
authentication-method pre-share
#
ike keychain key1
pre-shared-key address 1.1.1.1 key simple 123
#
interface gigabitethernet 1/0/1
ipsec apply policy map1
#
```

Configuring IKE

Unless otherwise specified, the term "IKE" in this chapter refers to IKEv1.

About IKE

Built on a framework defined by ISAKMP, Internet Key Exchange (IKE) provides automatic key negotiation and SA establishment services for IPsec.

Benefits of IKE

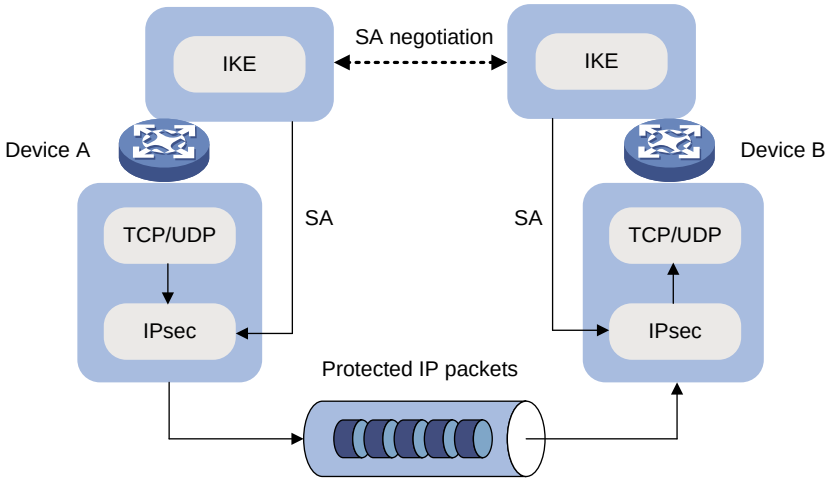
IKE provides the following benefits for IPsec:

- Automatically negotiates IPsec parameters.
- Performs DH exchanges to calculate shared keys, making sure each SA has a key that is independent of other keys.
- Automatically negotiates SAs when the sequence number in the AH or ESP header overflows, making sure IPsec can provide the anti-replay service by using the sequence number.

Relationship between IPsec and IKE

As shown in Figure 11, IKE negotiates SAs for IPsec and transfers the SAs to IPsec, and IPsec uses the SAs to protect IP packets.

Figure 11 Relationship between IKE and IPsec



IKE negotiation process

IKE negotiates keys and SAs for IPsec in two phases:

1. **Phase 1**—The two peers establish an IKE SA, a secure, authenticated channel for communication.
2. **Phase 2**—Using the IKE SA established in phase 1, the two peers negotiate to establish IPsec SAs.

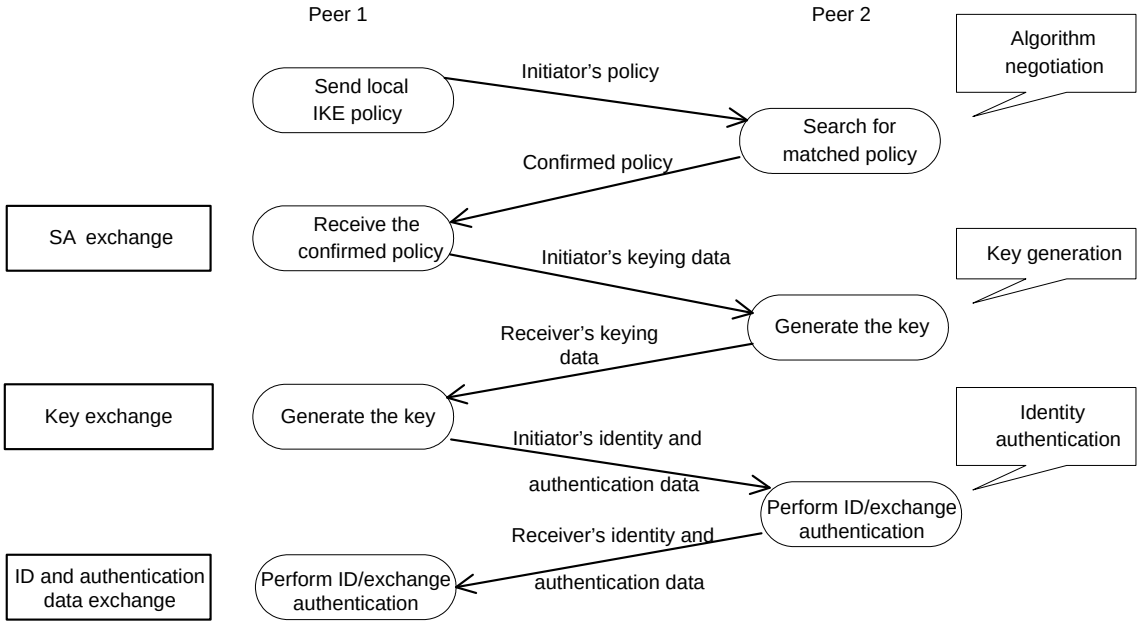
Phase 1 negotiation can use either main mode or aggressive mode.

IKE exchange process in main mode

As shown in Figure 12, the main mode of IKE negotiation in phase 1 involves three pairs of messages:

- **SA exchange**—Used for negotiating the IKE security policy.
- **Key exchange**—Used for exchanging the DH public value and other values, such as the random number. The two peers use the exchanged data to generate key data and use the encryption key and authentication key to ensure the security of IP packets.
- **ID and authentication data exchange**—Used for identity authentication.

Figure 12 IKE exchange process in main mode



IKE exchange process in aggressive mode

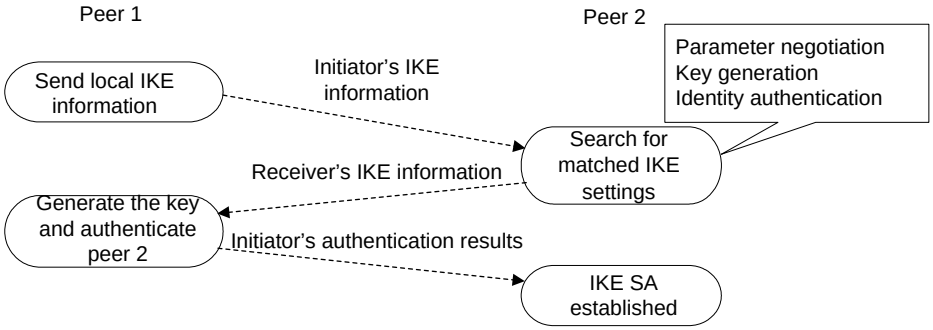
As shown in Figure 13, the process of phase 1 IKE negotiation in aggressive mode is as follows:

1. The initiator (peer 1) sends a message containing the local IKE information to peer 2. The message includes parameters used for IKE SA establishment, keying data, and peer 1's identity information.
2. Peer 2 chooses the IKE establishment parameters to use, generate the key, and authenticate peer 1's identity. Then it sends the IKE data to peer 1.
3. Peer 1 generates the key, authenticates peer 2's identity, and sends the results to peer 1.

After the preceding process, an IKE SA is established between peer 1 and peer 2.

The aggressive mode is faster than the main mode but it does not provide identity information protection. The main mode provides identity information protection but is slower. Choose the appropriate negotiation mode according to your requirements.

Figure 13 IKE exchange process in aggressive mode



IKE security mechanism

IKE has a series of self-protection mechanisms and supports secure identity authentication, key distribution, and IPsec SA establishment on insecure networks.

Identity authentication

The IKE identity authentication mechanism is used to authenticate the identity of the communicating peers. The device supports the following identity authentication methods:

- **Preshared key authentication**—Two communicating peers use the pre-configured shared key for identity authentication.
- **RSA signature authentication** and **DSA signature authentication**—Two communicating peers use the digital certificates issued by the CA for identity authentication.

The preshared key authentication method does not require certificates and is easy to configure. It is usually deployed in small networks.

The signature authentication methods provide higher security and are usually deployed in networks with the headquarters and some branches. When deployed in a network with many branches, a signature authentication method can simplify the configuration because only one PKI domain is required. If you use the preshared key authentication method, you must configure a preshared key for each branch on the Headquarters node.

DH algorithm

The DH algorithm is a public key algorithm. With this algorithm, two peers can exchange keying material and then use the material to calculate the shared keys. Due to the decryption complexity, a third party cannot decrypt the keys even after intercepting all keying materials.

PFS

The Perfect Forward Secrecy (PFS) feature is a security feature based on the DH algorithm. After PFS is enabled, an additional DH exchange is performed in IKE phase 2 to make sure IPsec keys have no derivative relations with IKE keys and a broken key brings no threats to other keys.

Protocols and standards

- RFC 2408, *Internet Security Association and Key Management Protocol (ISAKMP)*
- RFC 2409, *The Internet Key Exchange (IKE)*
- RFC 2412, *The OAKLEY Key Determination Protocol*
- Internet Draft, *draft-ietf-ipsec-isakmp-xauth-06*
- Internet Draft, *draft-dukes-ike-mode-cfg-02*

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see "Configuring FIPS") and non-FIPS mode.

IKE tasks at a glance

To configure IKE, perform the following tasks:

1. (Optional.) Configuring an IKE profile
   a. Configuring an IKE profile
   b. Configuring peer IDs for the IKE profile
   c. Specifying the IKE keychain or PKI domain
   d. Configuring the IKE phase 1 negotiation mode
   e. Specifying IKE proposals for the IKE profile
   f. Configuring the local ID for the IKE profile
   g. Configuring optional features for the IKE profile
2. Configuring an IKE proposal
3. Configuring an IKE keychain
4. (Optional.) Configuring the global identity information
5. (Optional.) Configuring the IKE keepalive feature
6. (Optional.) Configuring the IKE NAT keepalive feature
7. (Optional.) Configuring global IKE DPD
8. (Optional.) Enabling invalid SPI recovery
9. (Optional.) Setting the maximum number of IKE SAs
10. (Optional.) Configuring an IKE IPv4 address pool
11. (Optional.) Configuring SNMP notifications for IKE

Prerequisites for IKE configuration

Determine the following parameters prior to IKE configuration:

- The algorithms to be used during IKE negotiation, including the identity authentication method, encryption algorithm, authentication algorithm, and DH group.
  - Different algorithms provide different levels of protection. A stronger algorithm provides more resistance to decryption but uses more resources.
  - A DH group that uses more bits provides higher security but needs more time for processing.
- The preshared key or PKI domain for IKE negotiation. For more information about PKI, see "Configuring PKI."
- The IKE-based IPsec policies for the communicating peers. If you do not specify an IKE profile in an IPsec policy, the device selects an IKE profile for the IPsec policy. If no IKE profile is configured, the globally configured IKE settings are used. For more information about IPsec, see "Configuring IPsec."

Configuring an IKE profile

Creating an IKE profile

About this task

Perform this task to create an IKE profile.

An IKE profile is intended to provide a set of parameters for IKE negotiation.

Procedure

1. Enter system view.
   system-view
2. Create an IKE profile and enter its view.
   ike profile *profile-name*

Configuring peer IDs for the IKE profile

About this task

Perform this task to configure the peer IDs for IKE profile matching. When the device needs to select an IKE profile for IKE negotiation with a peer, it compares the received peer ID with the peer IDs of its local IKE profiles. If a match is found, it uses the IKE profile with the matching peer ID for IKE negotiation.

Restrictions and guidelines

For an IKE profile, you can configure multiple peer IDs. A peer ID configured earlier has a higher priority.

Two IKE peers must both have or both not have peer IDs configured.

Procedure

1. Enter system view.
   system-view
2. Enter IKE profile view.
   ike profile *profile-name*
3. Configure a peer ID for the IKE profile.
   match remote { **certificate** *policy-name* | **identity** { **address** { { *ipv4-address* [*mask* | *mask-length*] | **range** *low-ipv4-address high-ipv4-address* } | **ipv6** { *ipv6-address* [*prefix-length*] | **range** *low-ipv6-address high-ipv6-address* } } [**vpn-instance** *vpn-instance-name*] | **fqdn** *fqdn-name* | **user-fqdn** *user-fqdn-name* } }

Specifying the IKE keychain or PKI domain

Restrictions and guidelines

Configure the IKE keychain or PKI domain for the IKE proposals to use. To use digital signature authentication, configure a PKI domain. To use preshared key authentication, configure an IKE keychain.

Procedure

1. Enter system view.

system-view

2. Enter IKE profile view.

   ike profile *profile-name*

3. Specify the keychain for preshared key authentication or the PKI domain used to request a certificate for digital signature authentication.

   - Specify the keychain.

     keychain *keychain-name*

   - Specify the PKI domain.

     certificate domain *domain-name*

   By default, no IKE keychain or PKI domain is specified in an IKE profile.

Configuring the IKE phase 1 negotiation mode

Restrictions and guidelines

Specify the IKE phase 1 negotiation mode (main or aggressive) that the device uses as the initiator. When the device acts as the responder, it uses the IKE negotiation mode of the initiator.

Procedure

1. Enter system view.

   system-view

2. Enter IKE profile view.

   ike profile *profile-name*

3. Specify the IKE negotiation mode for phase 1.

   In non-FIPS mode:

   exchange-mode { **aggressive** | **main** }

   In FIPS mode:

   exchange-mode main

   By default, IKE negotiation in phase 1 uses the main mode.

Specifying IKE proposals for the IKE profile

Restrictions and guidelines

Specify the IKE proposals that the device can use as the initiator. An IKE proposal specified earlier has a higher priority. When the device acts as the responder, it uses the IKE proposals configured in system view to match the IKE proposals received from the initiator. If no matching proposal is found, the negotiation fails.

Procedure

1. Enter system view.

   system-view

2. Enter IKE profile view.

   ike profile *profile-name*

3. Specify IKE proposals for the IKE profile.

   proposal *proposal-number*&<1-6>

   By default, no IKE proposals are specified for an IKE profile and the IKE proposals configured in system view are used for IKE negotiation.

Configuring the local ID for the IKE profile

Restrictions and guidelines

For digital signature authentication, the device can use an ID of any type. If the local ID is an IP address that is different from the IP address in the local certificate, the device uses the FQDN (the device name configured by using the **sysname** command) instead.

For preshared key authentication, the device can use an ID of any type other than the DN.

Procedure

1. Enter system view.

   system-view

2. Enter IKE profile view.

   ike profile *profile-name*

3. Configure the local ID.

   local-identity { **address** { *ipv4-address* | **ipv6** *ipv6-address* } | **dn** | **fqdn** [*fqdn-name*] | **user-fqdn** [*user-fqdn-name*] }

   By default, no local ID is configured for an IKE profile, and an IKE profile uses the local ID configured in system view. If the local ID is not configured in system view, the IKE profile uses the IP address of the interface to which the IPsec policy or IPsec policy template is applied as the local ID.

Configuring optional features for the IKE profile

1. Enter system view.

   system-view

2. Enter IKE profile view.

   ike profile *profile-name*

3. Configure optional features as needed.

   - Configure IKE DPD.

     dpd interval *interval* [**retry** *seconds*] { **on-demand** | **periodic** }

     By default, IKE DPD is not configured for an IKE profile and an IKE profile uses the DPD settings configured in system view. If IKE DPD is not configured in system view either, the device does not perform dead IKE peer detection.

     The IKE DPD settings configured in the IKE profile view take precedence over those configured in system view.

   - Specify the local interface or IP address to which the IKE profile can be applied.

     match local address { *interface-type interface-number* | { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] }

     By default, an IKE profile can be applied to any local interface or IP address.

     An IKE profile configured with this command has a higher priority over those not configured with this command.

   - Specify a priority for the IKE profile.

     priority *priority*

     By default, the priority of an IKE profile is 100.

     The device selects a local IKE profile for IKE negotiation as follows:

     – First, it selects an IKE profile with the **match local address** command configured.

- If a tie exists, it selects the IKE profile with a smaller priority number.
- If a tie still exists, it selects the IKE profile configured earlier.

- Enable client authentication.

  client-authentication xauth

  By default, client authentication is disabled.

  This feature is typically used on the IPsec gateway at the enterprise center. It enables the IPsec gateway to perform extended (XAUTH) authentication on remote users through AAA after IKE phase 1 negotiation. AAA configuration is also required on the IPsec gateway for client authentication.

- Enable AAA authorization.

  aaa authorization domain *domain-name* **username** *user-name*

  By default, AAA authorization is disabled.

  The AAA authorization feature enables IKE to request authorization attributes, such as the IKE IPv4 address pool, from AAA. IKE uses the address pool to assign IPv4 addresses to remote users. For more information about AAA authorization, see "Configuring AAA."

Configuring an IKE proposal

About this task

An IKE proposal defines a set of attributes describing how IKE negotiation in phase 1 should take place. You can create multiple IKE proposals with different priorities. The priority of an IKE proposal is represented by its sequence number. The lower the sequence number, the higher the priority.

Two peers must have at least one matching IKE proposal for successful IKE negotiation. During IKE negotiation:

- The initiator sends its IKE proposals to the peer.
  - If the initiator is using an IPsec policy with an IKE profile, the initiator sends all IKE proposals specified in the IKE profile to the peer. An IKE proposal specified earlier for the IKE profile has a higher priority.
  - If the initiator is using an IPsec policy with no IKE profile, the initiator sends all its IKE proposals to the peer. An IKE proposal with a smaller number has a higher priority.
- The peer searches its own IKE proposals for a match. The search starts from the IKE proposal with the highest priority and proceeds in descending order of priority until a match is found. The matching IKE proposals are used to establish the IKE SA. If all user-defined IKE proposals are found mismatching, the two peers use their default IKE proposals to establish the IKE SA.

Two matching IKE proposals have the same encryption algorithm, authentication method, authentication algorithm, and DH group. The SA lifetime takes the smaller one of the two proposals' SA lifetime settings.

Procedure

1. Enter system view.

   system-view

2. Create an IKE proposal and enter its view.

   ike proposal *proposal-number*

   By default, a default IKE proposal exists.

3. Configure a description for the IKE proposal.

   description

   By default, an IKE proposal does not have a description.

4. Specify an encryption algorithm for the IKE proposal.

In non-FIPS mode:

encryption-algorithm { **3des-cbc** | **aes-cbc-128** | **aes-cbc-192** | **aes-cbc-256** | **des-cbc** | **sm1-cbc-128** | **sm1-cbc-192** | **sm1-cbc-256** | **sm4-cbc** }

By default, the 56-bit DES encryption algorithm in CBC mode is used .

In FIPS mode:

encryption-algorithm { **aes-cbc-128** | **aes-cbc-192** | **aes-cbc-256** }

By default, the 128-bit AES encryption algorithm in CBC mode is used.

5. Specify an authentication method for the IKE proposal.

   In non-FIPS mode:

   authentication-method { **dsa-signature** | **ecdsa-signature** | **pre-share** | **rsa-signature** }

   In FIPS mode:

   authentication-method { **ecdsa-signature** | **pre-share** | **rsa-signature** }

   By default, the preshared key authentication method is used.

6. Specify an authentication algorithm for the IKE proposal.

   authentication-algorithm { **md5** | **sha** | **sha256** | **sha384** | **sha512** | **sm3** }

   By default, the HMAC-SHA1 authentication algorithm is used.

7. Specify a DH group for key negotiation in phase 1.

   In non-FIPS mode:

   dh { **group1** | **group14** | **group19** | **group2** | **group20** | **group24** | **group5** }

   DH group 1 (the 768-bit DH group) is used by default.

   In FIPS mode:

   dh { **group14** | **group19** | **group20** | **group24** }

   DH group 14 (the 2048-bit DH group) is used by default.

8. (Optional.) Set the IKE SA lifetime for the IKE proposal.

   sa duration *seconds*

   By default, the IKE SA lifetime is 86400 seconds.

Configuring an IKE keychain

About this task

Perform this task when you configure the IKE to use the preshared key for authentication.

Follow these guidelines when you configure an IKE keychain:

- Two peers must be configured with the same preshared key to pass preshared key authentication.
- You can specify the local address configured in IPsec policy or IPsec policy template view (using the **local-address** command) for the IKE keychain to be applied. If no local address is configured, specify the IP address of the interface that uses the IPsec policy.
- The device determines the priority of an IKE keychain as follows:
  a. The device examines the existence of the **match local address** command. An IKE keychain with the **match local address** command configured has a higher priority.
  b. If a tie exists, the device compares the priority numbers. An IKE keychain with a smaller priority number has a higher priority.
  c. If a tie still exists, the device prefers an IKE keychain configured earlier.

Procedure

1. Enter system view.

   system-view

2. Create an IKE keychain and enter its view.

   ike keychain *keychain-name* [**vpn-instance** *vpn-instance-name*]

3. Configure a preshared key.

   In non-FIPS mode:

   pre-shared-key { **address** { *ipv4-address* [*mask* | *mask-length*] | **ipv6** *ipv6-address* [*prefix-length*] } | **hostname** *host-name* } **key** { **cipher** | **simple** } *string*

   In FIPS mode:

   pre-shared-key { **address** { *ipv4-address* [*mask* | *mask-length*] | **ipv6** *ipv6-address* [*prefix-length*] } | **hostname** *host-name* } **key** [**cipher** *string*]

   By default, no preshared key is configured.

4. (Optional.) Specify a local interface or IP address to which the IKE keychain can be applied.

   match local address { *interface-type interface-number* | { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] }

   By default, an IKE keychain can be applied to any local interface or IP address.

5. (Optional.) Specify a priority for the IKE keychain.

   priority *priority*

   The default priority is 100.

Configuring the global identity information

Restrictions and guidelines

The global identity can be used by the device for all IKE SA negotiations, and the local identity (set by the **local-identity** command) can be used only by the device that uses the IKE profile.

When signature authentication is used, you can set any type of the identity information.

When preshared key authentication is used, you cannot set the DN as the identity.

Procedure

1. Enter system view.

   system-view

2. Configure the global identity to be used by the local end.

   ike identity { **address** { *ipv4-address* | **ipv6** *ipv6-address* }| **dn** | **fqdn** [*fqdn-name*] | **user-fqdn** [*user-fqdn-name*] }

   By default, the IP address of the interface to which the IPsec policy or IPsec policy template is applied is used as the IKE identity.

3. (Optional.) Configure the local device to always obtain the identity information from the local certificate for signature authentication.

   ike signature-identity from-certificate

   By default, the local end uses the identity information specified by **local-identity** or **ike identity** for signature authentication.

Configure this command when the aggressive mode and signature authentication are used and the device interconnects with a Comware 5-based peer device. Comware 5 supports only DN for signature authentication.

Configuring the IKE keepalive feature

About this task

IKE sends keepalive packets to query the liveness of the peer. If the peer is configured with the keepalive timeout time, you must configure the keepalive interval on the local device. If the peer receives no keepalive packets during the timeout time, the IKE SA is deleted along with the IPsec SAs it negotiated.

Restrictions and guidelines

Configure IKE DPD instead of IKE keepalive unless IKE DPD is not supported on the peer. The IKE keepalive feature sends keepalives at regular intervals, which consumes network bandwidth and resources.

The keepalive timeout time configured on the local device must be longer than the keepalive interval configured at the peer. Since it seldom occurs that more than three consecutive packets are lost on a network, you can set the keepalive timeout three times as long as the keepalive interval.

Procedure

1. Enter system view.
   `system-view`
2. Set the IKE SA keepalive interval.
   `ike keepalive interval` *`interval`*
   By default, no keepalives are sent to the peer.
3. Set the IKE SA keepalive timeout time.
   `ike keepalive timeout` *`seconds`*
   By default, IKE SA keepalive never times out.

Configuring the IKE NAT keepalive feature

About this task

If IPsec traffic passes through a NAT device, you must configure the NAT traversal feature. If no packet travels across an IPsec tunnel in a period of time, the NAT sessions are aged and deleted, disabling the tunnel from transmitting data to the intended end. To prevent NAT sessions from being aged, configure the NAT keepalive feature on the IKE gateway behind the NAT device to send NAT keepalive packets to its peer periodically to keep the NAT session alive.

Procedure

1. Enter system view.
   `system-view`
2. Set the IKE NAT keepalive interval.
   `ike nat-keepalive` *`seconds`*
   The default interval is 20 seconds.

Configuring global IKE DPD

About this task

DPD detects dead peers. It can operate in periodic mode or on-demand mode.

- **Periodic DPD**—Sends a DPD message at regular intervals. It features an earlier detection of dead peers, but consumes more bandwidth and CPU.
- **On-demand DPD**—Sends a DPD message based on traffic. When the device has traffic to send and is not aware of the liveness of the peer, it sends a DPD message to query the status of the peer. If the device has no traffic to send, it never sends DPD messages. As a best practice, use the on-demand mode.

The IKE DPD works as follows:

1. The local device sends a DPD message to the peer, and waits for a response from the peer.
2. If the peer does not respond within the retry interval specified by the **retry** *seconds* parameter, the local device resends the message.
3. If still no response is received within the retry interval, the local end sends the DPD message again. The system allows a maximum of four retries.
4. If the local device receives no response after two retries, the device considers the peer to be dead, and deletes the IKE SA along with the IPsec SAs it negotiated.
5. If the local device receives a response from the peer during the detection process, the peer is considered alive. The local device performs a DPD detection again when the triggering interval is reached or it has traffic to send, depending on the DPD mode.

Restrictions and guidelines

When DPD settings are configured in both IKE profile view and system view, the DPD settings in IKE profile view apply. If DPD is not configured in IKE profile view, the DPD settings in system view apply.

It is a good practice to set the triggering interval longer than the retry interval so that a DPD detection is not triggered during a DPD retry.

Procedure

1. Enter system view.

   system-view

2. Enable sending IKE DPD messages.

   ike dpd interval *interval* [**retry** *seconds*] { **on-demand** | **periodic** }

   By default, IKE DPD is disabled.

Enabling invalid SPI recovery

About this task

An IPsec "black hole" occurs when one IPsec peer fails (for example, a peer can fail if a reboot occurs). One peer fails and loses its SAs with the other peer. When an IPsec peer receives a data packet for which it cannot find an SA, an invalid SPI is encountered. The peer drops the data packet and tries to send an SPI invalid notification to the data originator. This notification is sent by using the IKE SA. Because no IKE SA is available, the notification is not sent. The originating peer continues sending the data by using the IPsec SA that has the invalid SPI, and the receiving peer keeps dropping the traffic.

The invalid SPI recovery feature enables the receiving peer to set up an IKE SA with the originator so that an SPI invalid notification can be sent. Upon receiving the notification, the originating peer deletes the IPsec SA that has the invalid SPI. If the originator has data to send, new SAs will be set up.

Restrictions and guidelines

Use caution when you enable the invalid SPI recovery feature because using this feature can result in a DoS attack. Attackers can make a great number of invalid SPI notifications to the same peer.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable invalid SPI recovery.
   ```
   ike invalid-spi-recovery enable
   ```
   By default, the invalid SPI recovery is disabled.

Setting the maximum number of IKE SAs

About this task

You can set the maximum number of half-open IKE SAs and the maximum number of established IKE SAs.

- The supported maximum number of half-open IKE SAs depends on the device's processing capability. Adjust the maximum number of half-open IKE SAs to make full use of the device's processing capability without affecting the IKE SA negotiation efficiency.
- The supported maximum number of established IKE SAs depends on the device's memory space. Adjust the maximum number of established IKE SAs to make full use of the device's memory space without affecting other applications in the system.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Set the maximum number of half-open IKE SAs and the maximum number of established IKE SAs.
   ```
   ike limit { max-negotiating-sa negotiation-limit | max-sa sa-limit }
   ```
   By default, there is no limit to the maximum number of IKE SAs.

Configuring an IKE IPv4 address pool

About this task

To perform centralized management on remote users, an IPsec gateway can use an IPv4 address pool to assign private IPv4 addresses to remote users.

You must use an IKE IPv4 address pool together with AAA authorization by specifying the IKE IPv4 address pool as an AAA authorization attribute. For more information about AAA authorization, see "Configuring AAA."

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Configure an IKE IPv4 address pool.
   ```
   ike address-group group-name start-ipv4-address end-ipv4-address
   [ mask | mask-length ]
   ```

Configuring IKE logging

About this task

IKE logging provides you with records of important IKE configuration and negotiation events.

- IKE negotiation event logging generates log entries at the beginning and end of IKE negotiation phases and during key exchange and security parameters negotiation. An IKE negotiation event log entry primarily includes the source and destination IP addresses, source and destination ports, and initiator and responder. If the negotiation fails, the log entry also includes the reason for negotiation failure.
- IKE configuration event logging primarily records IKE proposal creation and deletion events. An IKE configuration log entry primarily includes proposal number information.

If you execute this command multiple times, the most recent configuration takes effect.

Restrictions and guidelines

This command is not available in FIPS mode.

Procedure

1. Enter system view.

   system-view
2. Enable IKE logging for event types of interest.

   ike logging { **config** | **negotiation** } * **enable**

   By default, IKE logging is disabled.

Configuring SNMP notifications for IKE

About this task

After you enable SNMP notifications for IKE, the IKE module notifies the NMS of important module events. The notifications are sent to the device's SNMP module. For the notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP notifications, see *Network Management and Monitoring Configuration Guide*.

To generate and output SNMP notifications for a specific IKE failure or event type, perform the following tasks:

1. Enable SNMP notifications for IKE globally.
2. Enable SNMP notifications for the failure or event type.

Procedure

1. Enter system view

   system-view
2. Enable SNMP notifications for IKE globally.

   snmp-agent trap enable ike global

   By default, SNMP notifications for IKE are disabled.
3. Enable SNMP notifications for the specified failure or event types.

   snmp-agent trap enable ike [**attr-not-support** | **auth-failure** | **cert-type-unsupport** | **cert-unavailable** | **cert-verify-failure** | **decrypt-failure** | **encrypt-failure** | **invalid-cert-auth** | **invalid-cookie** | **invalid-id** | **invalid-proposal** | **invalid-protocol** | **invalid-sign** | **no-sa-failure** | **p1sa-failure** | **p2sa-failure** |

p2sa-terminate | **proposal-add** | **proposal-delete** | **tunnel-start** | **tunnel-stop** | **unsupport-exch-type**] *

By default, SNMP notifications for all failure and event types are disabled.

Display and maintenance commands for IKE

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display configuration information about all IKE proposals.	**display ike proposal**
Display information about the current IKE SAs.	**display ike sa** [**verbose** [**connection-id** *connection-id*	**remote-address** [**ipv6**] *remote-address* [**vpn-instance** *vpn-instance-name*]]]
Display IKE statistics.	**display ike statistics**
Delete IKE SAs.	**reset ike sa** [**connection-id** *connection-id*]
Clear IKE MIB statistics.	**reset ike statistics**

IKE configuration examples

Example: Configuring main-mode IKE with preshared key authentication

Network configuration

As shown in Figure 14, configure an IKE-based IPsec tunnel between Switch A and Switch B to secure the communication between the switches.

- Configure the two switches to use the default IKE proposal for the IKE negotiation.
- Configure the two switches to use the preshared key authentication method for the IKE negotiation.

Figure 14 Network diagram



Procedure

1. Make sure Switch A and Switch B can reach each other.
2. Configure Switch A:

   # Assign IP addresses to interfaces.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 1
   ```

```
[SwitchA-vlan1] quit
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port access vlan 1
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface vlan-interface 1
[SwitchA-vlan-interface1] ip address 1.1.1.1 255.255.0.0
[SwitchA-vlan-interface1] quit
```

Configure a static route to Switch B. In this example, the next hop in the route is 1.1.1.2.

```
[SwitchA] ip route-static 2.2.0.0 255.255.0.0 1.1.1.2
```

Configure IPv4 advanced ACL 3101 to identify the traffic from Switch A to Switch B.

```
[SwitchA] acl advanced 3101
[SwitchA-acl-ipv4-adv-3101] rule 0 permit ip source 1.1.1.1 0 destination 2.2.2.2 0
[SwitchA-acl-ipv4-adv-3101] quit
```

Create an IPsec transform set named **tran1**.

```
[SwitchA] ipsec transform-set tran1
```

Set the packet encapsulation mode to **tunnel**.

```
[SwitchA-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Use the ESP protocol for the IPsec transform set.

```
[SwitchA-ipsec-transform-set-tran1] protocol esp
```

Specify the encryption and authentication algorithms.

```
[SwitchA-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchA-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchA-ipsec-transform-set-tran1] quit
```

Create an IKE keychain named **keychain1**.

```
[SwitchA] ike keychain keychain1
```

Specify **12345zxcvb!@#$%ZXCVB** in plain text as the preshared key to be used with the remote peer at 2.2.2.2.

```
[SwitchA-ike-keychain-keychain1] pre-shared-key address 2.2.2.2 255.255.0.0 key
simple 12345zxcvb!@#$%ZXCVB
[SwitchA-ike-keychain-keychain1] quit
```

Create an IKE profile named **profile1**.

```
[SwitchA] ike profile profile1
```

Specify IKE keychain **keychain1**.

```
[SwitchA-ike-profile-profile1] keychain keychain1
```

Configure a peer ID with the identity type as IP address and the value as **2.2.2.2/16**.

```
[SwitchA-ike-profile-profile1] match remote identity address 2.2.2.2 255.255.0.0
[SwitchA-ike-profile-profile1] quit
```

Create an IKE-based IPsec policy entry. Specify the policy name as **map1** and set the sequence number to 10.

```
[SwitchA] ipsec policy map1 10 isakmp
```

Specify the remote IP address 2.2.2.2 for the IPsec tunnel.

```
[SwitchA-ipsec-policy-isakmp-map1-10] remote-address 2.2.2.2
```

Specify ACL 3101 to identify the traffic to be protected.

```
[SwitchA-ipsec-policy-isakmp-map1-10] security acl 3101
```

Specify IPsec transform set **tran1** for the IPsec policy.

```
[SwitchA-ipsec-policy-isakmp-map1-10] transform-set tran1
```

Specify IKE profile **profile1** for the IPsec policy.

```
[SwitchA-ipsec-policy-isakmp-map1-10] ike-profile profile1
[SwitchA-ipsec-policy-isakmp-map1-10] quit
```

Apply IPsec policy **map1** to VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ipsec apply policy map1
```

3. Configure Switch B:

Assign IP addresses to interfaces.

```
<SwitchB> system-view
[SwitchB] vlan 1
[SwitchB-vlan1] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port access vlan 1
[SwitchB-GigabitEthernet1/0/1] quit
[SwitchB] interface Vlan-interface1
[SwitchB-Vlan-interface1] ip address 2.2.2.2 255.255.0.0
[SwitchB-Vlan-interface1] quit
```

Configure a static route to Switch A. In this example, the next hop in the route is 2.2.2.3.

```
[SwitchB] ip route-static 1.1.0.0 255.255.0.0 2.2.2.3
```

Configure IPv4 advanced ACL 3101 to identify the traffic from Switch B to Switch A.

```
[SwitchB] acl number 3101
[SwitchB-acl-adv-3101] rule 0 permit ip source 2.2.2.2 0 destination 1.1.1.1 0
[SwitchB-acl-adv-3101] quit
```

Create an IPsec transform set named **tran1**.

```
[SwitchB] ipsec transform-set tran1
```

Set the packet encapsulation mode to **tunnel**.

```
[SwitchB-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Use the ESP protocol for the IPsec transform set.

```
[SwitchB-ipsec-transform-set-tran1] protocol esp
```

Specify the encryption and authentication algorithms.

```
[SwitchB-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchB-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran1] quit
```

Create an IKE keychain named **keychain1**.

```
[SwitchB]ike keychain keychain1
```

Specify **12345zxcvb!@#$%ZXCVB** in plain text as the preshared key to be used with the remote peer at 1.1.1.1.

```
[SwitchB-ike-keychain-keychain1] pre-shared-key address 1.1.1.1 255.255.0.0 key
simple 12345zxcvb!@#$%ZXCVB
[SwitchB-ike-keychain-keychain1] quit
```

Create an IKE profile named **profile1**.

```
[SwitchB] ike profile profile1
```

Specify IKE keychain **keychain1**

```
[SwitchB-ike-profile-profile1] keychain keychain1
```

Configure a peer ID with the identity type as IP address and the value as **1.1.1.1/16**.

```
[SwitchB-ike-profile-profile1] match remote identity address 1.1.1.1 255.255.0.0
[SwitchB-ike-profile-profile1] quit
```

Create an IKE-based IPsec policy entry. Specify the policy name as **use1** and set the sequence number to 10.

```
[SwitchB] ipsec policy use1 10 isakmp
```

Specify the remote IP address 1.1.1.1 for the IPsec tunnel.

```
[SwitchB-ipsec-policy-isakmp-use1-10] remote-address 1.1.1.1
```

Specify ACL 3101 to identify the traffic to be protected.

```
[SwitchB-ipsec-policy-isakmp-use1-10] security acl 3101
```

Specify IPsec transform set **tran1** for the IPsec policy.

```
[SwitchB-ipsec-policy-isakmp-use1-10] transform-set tran1
```

Specify IKE profile **profile1** for the IPsec policy.

```
[SwitchB-ipsec-policy-isakmp-use1-10] ike-profile profile1
[SwitchB-ipsec-policy-isakmp-use1-10] quit
```

Apply IPsec policy **use1** to VLAN-interface 1.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ipsec apply policy use1
```

Verifying the configuration

Initiate a connection between Switch A and Switch B to trigger IKE negotiation. After IPsec SAs are successfully negotiated by IKE, the traffic between the two switches is IPsec-protected.

Configuration files

Switch A

```
#
sysname SwitchA
#
vlan 1
#
interface gigabitethernet 1/0/1
port access vlan 1
#
interface vlan-interface 1
ip address 1.1.1.1 255.255.0.0
#
ip route-static 2.2.0.0 255.255.0.0 1.1.1.2
acl advanced 3101
rule 0 permit ip source 1.1.1.1 0 destination 2.2.2.2 0
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
#
ike keychain keychain1
pre-shared-key address 2.2.2.2 255.255.0.0 key simple 12345zxcvb!@#$%ZXCVB
#
ike profile profile1
keychain keychain1
match remote identity address 2.2.2.2 255.255.0.0
```

```
#
ipsec policy map1 10 isakmp
remote-address 2.2.2.2
security acl 3101
transform-set tran1
ike-profile profile1
#
interface vlan-interface 1
ipsec apply policy map1
#
```

Switch B

```
#
sysname SwitchB
#
vlan 1
#
interface gigabitethernet 1/0/1
port access vlan 1
#
interface Vlan-interface1
ip address 2.2.2.2 255.255.0.0
#
ip route-static 1.1.0.0 255.255.0.0 2.2.2.3
acl number 3101
rule 0 permit ip source 2.2.2.2 0 destination 1.1.1.1 0
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
#
ike keychain keychain1
pre-shared-key address 1.1.1.1 255.255.0.0 key simple 12345zxcvb!@#$%ZXCVB
#
ike profile profile1
keychain keychain1
match remote identity address 1.1.1.1 255.255.0.0
#
ipsec policy use1 10 isakmp
remote-address 1.1.1.1
security acl 3101
transform-set tran1
ike-profile profile1
#
interface vlan-interface 1
ipsec apply policy use1
#
```

Example: Configuring an IKE-based IPsec tunnel for IPv4 packets

Network configuration

As shown in Figure 15, establish an IPsec tunnel between Switch A and Switch B to protect data flows between the switches. Configure the IPsec tunnel as follows:

- Specify the encapsulation mode as tunnel, the security protocol as ESP, the encryption algorithm as AES-CBC-192, and the authentication algorithm as HMAC-SHA1.
- Set up SAs through IKE negotiation.

Figure 15 Network diagram



Procedure

1. Configure Switch A:

Assign IP addresses to interfaces.

```
<SwitchA> system-view
[SwitchA] vlan 1
[SwitchA-vlan1] quit
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port access vlan 1
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 2.2.2.1 255.255.255.0
[SwitchA-Vlan-interface1] quit
```

Configure a static route to Switch B. In this example, the next hop in the route is 2.2.2.3.

```
[SwitchA] ip route-static 2.2.3.0 255.255.255.0 2.2.2.3
```

Configure an IPv4 advanced ACL to identify the data flows from Switch A to Switch B.

```
[SwitchA] acl advanced 3101
[SwitchA-acl-ipv4-adv-3101] rule 0 permit ip source 2.2.2.1 0 destination 2.2.3.1 0
[SwitchA-acl-ipv4-adv-3101] quit
```

Create an IPsec transform set named **tran1**.

```
[SwitchA] ipsec transform-set tran1
```

Specify the encapsulation mode as **tunnel**.

```
[SwitchA-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Specify the security protocol as **ESP**.

```
[SwitchA-ipsec-transform-set-tran1] protocol esp
```

Specify the ESP encryption and authentication algorithms.

```
[SwitchA-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchA-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchA-ipsec-transform-set-tran1] quit
```

Create an IKE keychain named **keychain1**.

```
[SwitchA] ike keychain keychain1
```

Specify **12345zxcvb!@#$%ZXCVB** in plain text as the preshared key to be used with the remote peer at 2.2.3.1.

```
[SwitchA-ike-keychain-keychain1] pre-shared-key address 2.2.3.1 255.255.255.0 key simple 12345zxcvb!@#$%ZXCVB
[SwitchA-ike-keychain-keychain1] quit
```

Create and configure an IKE profile named **profile1**.

```
[SwitchA] ike profile profile1
[SwitchA-ike-profile-profile1] keychain keychain1
[SwitchA-ike-profile-profile1] match remote identity address 2.2.3.1 255.255.255.0
[SwitchA-ike-profile-profile1] quit
```

Create an IKE-based IPsec policy entry. Specify the policy name as **map1** and set the sequence number to 10.

```
[SwitchA] ipsec policy map1 10 isakmp
```

Specify ACL 3101.

```
[SwitchA-ipsec-policy-isakmp-map1-10] security acl 3101
```

Specify IPsec transform set **tran1**.

```
[SwitchA-ipsec-policy-isakmp-map1-10] transform-set tran1
```

Specify the local and remote IP addresses of the IPsec tunnel as 2.2.2.1 and 2.2.3.1.

```
[SwitchA-ipsec-policy-isakmp-map1-10] local-address 2.2.2.1
[SwitchA-ipsec-policy-isakmp-map1-10] remote-address 2.2.3.1
```

Specify IKE profile **profile1**.

```
[SwitchA-ipsec-policy-isakmp-map1-10] ike-profile profile1
[SwitchA-ipsec-policy-isakmp-map1-10] quit
```

Apply IPsec policy **map1** to VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ipsec apply policy map1
```

2. Configure Switch B:

Assign IP addresses to interfaces.

```
<SwitchB> system-view
[SwitchB] vlan 1
[SwitchB-vlan1] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port access vlan 1
[SwitchB-GigabitEthernet1/0/1] quit
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 2.2.3.1 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

Configure a static route to Switch A. In this example, the next hop in the route is 3.3.3.2.

```
[SwitchB] ip route-static 2.2.2.0 255.255.255.0 3.3.3.2
```

Configure an IPv4 advanced ACL to identify the data flows from Switch B to Switch A.

```
[SwitchB] acl advanced 3101
[SwitchB-acl-ipv4-adv-3101] rule 0 permit ip source 2.2.3.1 0 destination 2.2.2.1 0
[SwitchB-acl-ipv4-adv-3101] quit
```

Create an IPsec transform set named **tran1**.

```
[SwitchB] ipsec transform-set tran1
```

Specify the encapsulation mode as **tunnel**.

```
[SwitchB-ipsec-transform-set-tran1] encapsulation-mode tunnel
```

Specify the security protocol as **ESP**.

```
[SwitchB-ipsec-transform-set-tran1] protocol esp
```

Specify the ESP encryption and authentication algorithms.

```
[SwitchB-ipsec-transform-set-tran1] esp encryption-algorithm aes-cbc-192
[SwitchB-ipsec-transform-set-tran1] esp authentication-algorithm sha1
[SwitchB-ipsec-transform-set-tran1] quit
```

Create an IKE keychain named **keychain1**.

```
[SwitchB] ike keychain keychain1
```

Specify **12345zxcvb!@#$%ZXCVB** in plain text as the preshared key to be used with the remote peer at 2.2.2.1.

```
[SwitchB-ike-keychain-keychain1] pre-shared-key address 2.2.2.1 255.255.255.0 key
simple 12345zxcvb!@#$%ZXCVB
[SwitchB-ike-keychain-keychain1] quit
```

Create and configure an IKE profile named **profile1**.

```
[SwitchB] ike profile profile1
[SwitchB-ike-profile-profile1] keychain keychain1
[SwitchB-ike-profile-profile1] match remote identity address 2.2.2.1 255.255.255.0
[SwitchB-ike-profile-profile1] quit
```

Create an IKE-based IPsec policy entry. Specify the policy name as **use1** and set the sequence number to 10.

```
[SwitchB] ipsec policy use1 10 isakmp
```

Specify ACL 3101.

```
[SwitchB-ipsec-policy-isakmp-use1-10] security acl 3101
```

Specify IPsec transform set **tran1**.

```
[SwitchB-ipsec-policy-isakmp-use1-10] transform-set tran1
```

Specify the local and remote IP addresses of the IPsec tunnel as 2.2.3.1 and 2.2.2.1.

```
[SwitchB-ipsec-policy-isakmp-use1-10] local-address 2.2.3.1
[SwitchB-ipsec-policy-isakmp-use1-10] remote-address 2.2.2.1
```

Specify IKE profile **profile1**.

```
[SwitchB-ipsec-policy-isakmp-use1-10] ike-profile profile1
[SwitchB-ipsec-policy-isakmp-use1-10] quit
```

Apply IPsec policy **use1** to VLAN-interface 1.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ipsec apply policy use1
```

Verifying the configuration

Initiate a connection between Switch A and Switch B to trigger IKE negotiation. After IPsec SAs are successfully negotiated by IKE, the traffic between the two switches is IPsec-protected.

Configuration files

Switch A

```
#
sysname SwitchA
#
vlan 1
#
interface gigabitethernet 1/0/1
port access vlan 1
#
```

```
interface vlan-interface 1
ip address 2.2.2.1 255.255.255.0
#
ip route-static 2.2.3.0 255.255.255.0 2.2.2.3
acl advanced 3101
rule 0 permit ip source 2.2.2.1 0 destination 2.2.3.1 0
#
ipsec transform-set tran1
encapsulation-mode tunnel
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
#
ike keychain keychain1
pre-shared-key address 2.2.3.1 255.255.255.0 key simple 12345zxcvb!@#$%ZXCVB
#
ike profile profile1
keychain keychain1
match remote identity address 2.2.3.1 255.255.255.0
#
ipsec policy map1 10 isakmp
security acl 3101
transform-set tran1
local-address 2.2.2.1
remote-address 2.2.3.1
ike-profile profile1
#
interface vlan-interface 1
ipsec apply policy map1
#
```

Switch B

```
#
sysname SwitchB
#
vlan 1
#
interface gigabitethernet 1/0/1
port access vlan 1
#
interface vlan-interface 1
ip address 2.2.3.1 255.255.255.0
#
ip route-static 2.2.2.0 255.255.255.0 3.3.3.2
acl advanced 3101
rule 0 permit ip source 2.2.3.1 0 destination 2.2.2.1 0
#
ipsec transform-set tran1
encapsulation-mode tunnel
```

```
protocol esp
esp encryption-algorithm aes-cbc-192
esp authentication-algorithm sha1
#
ike keychain keychain1
pre-shared-key address 2.2.2.1 255.255.255.0 key simple 12345zxcvb!@#$%ZXCVB
#
ike profile profile1
keychain keychain1
match remote identity address 2.2.2.1 255.255.255.0
#
ipsec policy use1 10 isakmp
security acl 3101
transform-set tran1
local-address 2.2.3.1
remote-address 2.2.2.1
ike-profile profile1
#
interface vlan-interface 1
ipsec apply policy use1
#
```

Troubleshooting IKE

IKE negotiation failed because no matching IKE proposals were found

Symptom

1. The IKE SA is in Unknown state.

```
<Sysname> display ike sa
    Connection-ID   Remote                Flag         DOI
-----------------------------------------------------------------
    1               192.168.222.5         Unknown      IPsec
Flags:
RD--READY RL--REPLACED FD-FADING RK-REKEY
```

2. When IKE event debugging and packet debugging are enabled, the following messages appear:

   IKE event debugging message:

   ```
   The attributes are unacceptable.
   ```

   IKE packet debugging message:

   ```
   Construct notification packet: NO_PROPOSAL_CHOSEN.
   ```

Analysis

Certain IKE proposal settings are incorrect.

Solution

1. Examine the IKE proposal configuration to see whether the two ends have matching IKE proposals.

2. Modify the IKE proposal configuration to make sure the two ends have matching IKE proposals.

IKE negotiation failed because no IKE proposals or IKE keychains are specified correctly

Symptom

1. The IKE SA is in Unknown state.

```
<Sysname> display ike sa
    Connection-ID   Remote                Flag         DOI
-----------------------------------------------------------------
    1               192.168.222.5         Unknown      IPsec
Flags:
RD--READY RL--REPLACED FD-FADING RK-REKEY
```

2. The following IKE event debugging or packet debugging message appeared:

   IKE event debugging message:

```
Notification PAYLOAD_MALFORMED is received.
```

   IKE packet debugging message:

```
Construct notification packet: PAYLOAD_MALFORMED.
```

Analysis

- If the following debugging information appeared, the matched IKE profile is not using the matched IKE proposal:

```
Failed to find proposal 1 in profile profile1.
```

- If the following debugging information appeared, the matched IKE profile is not using the matched IKE keychain:

```
Failed to find keychain keychain1 in profile profile1.
```

Solution

- Verify that the matched IKE proposal (IKE proposal 1 in this debugging message example) is specified for the IKE profile (IKE profile 1 in the example).
- Verify that the matched IKE keychain (IKE keychain 1 in this debugging message example) is specified for the IKE profile (IKE profile 1 in the example).

IPsec SA negotiation failed because no matching IPsec transform sets were found

Symptom

1. The **`display ike sa`** command shows that the IKE SA negotiation succeeded and the IKE SA is in RD state, but the **`display ipsec sa`** command shows that the expected IPsec SA has not been negotiated yet.
2. The following IKE debugging message appeared:

```
The attributes are unacceptable.
```

   Or:

```
Construct notification packet: NO_PROPOSAL_CHOSEN.
```

Analysis

Certain IPsec policy settings are incorrect.

Solution

1. Examine the IPsec configuration to see whether the two ends have matching IPsec transform sets.
2. Modify the IPsec configuration to make sure the two ends have matching IPsec transform sets.

IPsec SA negotiation failed due to invalid identity information

Symptom

1. The **display ike sa** command shows that the IKE SA negotiation succeeded and the IKE SA is in RD state, but the **display ipsec sa** command shows that the expected IPsec SA has not been negotiated yet.
2. The following IKE debugging message appeared:

   ```
   Notification INVALID_ID_INFORMATION is received.
   ```

   Or:

   ```
   Failed to get IPsec policy when renegotiating IPsec SA. Delete IPsec SA.
   Construct notification packet: INVALID_ID_INFORMATION.
   ```

Analysis

Certain IPsec policy settings of the responder are incorrect. Verify the settings as follows:

1. Use the **display ike sa verbose** command to verify that matching IKE profiles were found in IKE negotiation phase 1. If no matching IKE profiles were found and the IPsec policy is using an IKE profile, the IPsec SA negotiation fails.

   # Identify whether matching IKE profiles were found in IKE negotiation phase 1. The following output shows that no matching IKE profile was found:

   ```
   <Sysname> display ike sa verbose
      ---------------------------------------------
      Connection ID: 3
      Outside VPN:
      Inside VPN:
      Profile:
      Transmitting entity: Responder
      ---------------------------------------------
      Local IP: 192.168.222.5
      Local ID type: IPV4_ADDR
      Local ID: 192.168.222.5

      Remote IP: 192.168.222.71
      Remote ID type: IPV4_ADDR
      Remote ID: 192.168.222.71

      Authentication-method: PRE-SHARED-KEY
      Authentication-algorithm: MD5
      Encryption-algorithm: 3DES-CBC

      Life duration(sec): 86400
      Remaining key duration(sec): 85847
      Exchange-mode: Main
      Diffie-Hellman group: Group 1
   ```

```
   NAT traversal: Not detected
```

Identify whether the IPsec policy is using an IKE profile. The following output shows that the IPsec policy is using an IKE profile.

```
[Sysname] display ipsec policy
-----------------------------------------
IPsec Policy: policy1
Interface: GigabitEthernet1/0/1
-----------------------------------------

  ----------------------------
  Sequence number: 1
  Mode: ISAKMP
  ----------------------------
  Description:
  Security data flow: 3000
  Selector mode: aggregation
  Local address: 192.168.222.5
  Remote address: 192.168.222.71
  Transform set:  transform1
  IKE profile: profile1
  SA duration(time based):
  SA duration(traffic based):
  SA idle time:
```

2. Verify that the ACL specified for the IPsec policy is correctly configured. If the flow range defined by the responder's ACL is smaller than that defined by the initiator's ACL, IPsec proposal matching will fail.

For example, if the initiator's ACL defines a flow from one network segment to another but the responder's ACL defines a flow from one host to another host, IPsec proposal matching will fail.

On the initiator:

```
[Sysname] display acl 3000
Advanced IPv4 ACL 3000, 1 rule,
ACL's step is 5
 rule 0 permit ip source 192.168.222.0 0.0.0.255 destination 192.168.222.0 0.0.0.255
```

On the responder:

```
[Sysname] display acl 3000
Advanced IPv4 ACL 3000, 1 rule,
ACL's step is 5
 rule 0 permit ip source 192.168.222.71 0 destination 192.168.222.5 0
```

3. Verify that the IPsec policy has a remote address and an IPsec transform set configured and that the IPsec transform set has all necessary settings configured.

If, for example, the IPsec policy has no remote address configured, the IPsec SA negotiation will fail:

```
[Sysname] display ipsec policy
-----------------------------------------
IPsec Policy: policy1
Interface: GigabitEthernet1/0/1
-----------------------------------------

  ----------------------------
```

```
Sequence number: 1
Mode: ISAKMP
-----------------------------
Security data flow: 3000
Selector mode: aggregation
Local address: 192.168.222.5
Remote address:
Transform set:  transform1
IKE profile: profile1
SA duration(time based):
SA duration(traffic based):
SA idle time:
```

Solution

1. If the IPsec policy specifies an IKE profile but no matching IKE profiles was found in IKE negotiation, perform one of the following tasks on the responder:
   - Remove the specified IKE profile from the IPsec policy.
   - Modify the specified IKE profile to match the IKE profile of the initiator.
2. If the flow range defined by the responder's ACL is smaller than that defined by the initiator's ACL, modify the responder's ACL so the ACL defines a flow range equal to or greater than that of the initiator's ACL.

   For example:

   ```
   [Sysname] display acl 3000
   Advanced IPv4 ACL 3000, 2 rules,
   ACL's step is 5
    rule 0 permit ip source 192.168.222.0 0.0.0.255 destination 192.168.222.0 0.0.0.255
   ```
3. Configure the missing settings (for example, the remote address).

Configuring IKEv2

About IKEv2

Internet Key Exchange version 2 (IKEv2) is an enhanced version of IKEv1. The same as IKEv1, IKEv2 has a set of self-protection mechanisms and can be used on insecure networks for reliable identity authentication, key distribution, and IPsec SA negotiation. IKEv2 provides stronger protection against attacks and higher key exchange ability and needs fewer message exchanges than IKEv1.

IKEv2 negotiation process

Compared with IKEv1, IKEv2 simplifies the negotiation process and is much more efficient.

IKEv2 defines three types of exchanges: initial exchanges, CREATE_CHILD_SA exchange, and INFORMATIONAL exchange.

As shown in Figure 16, IKEv2 uses two exchanges during the initial exchange process: IKE_SA_INIT and IKE_AUTH, each with two messages.

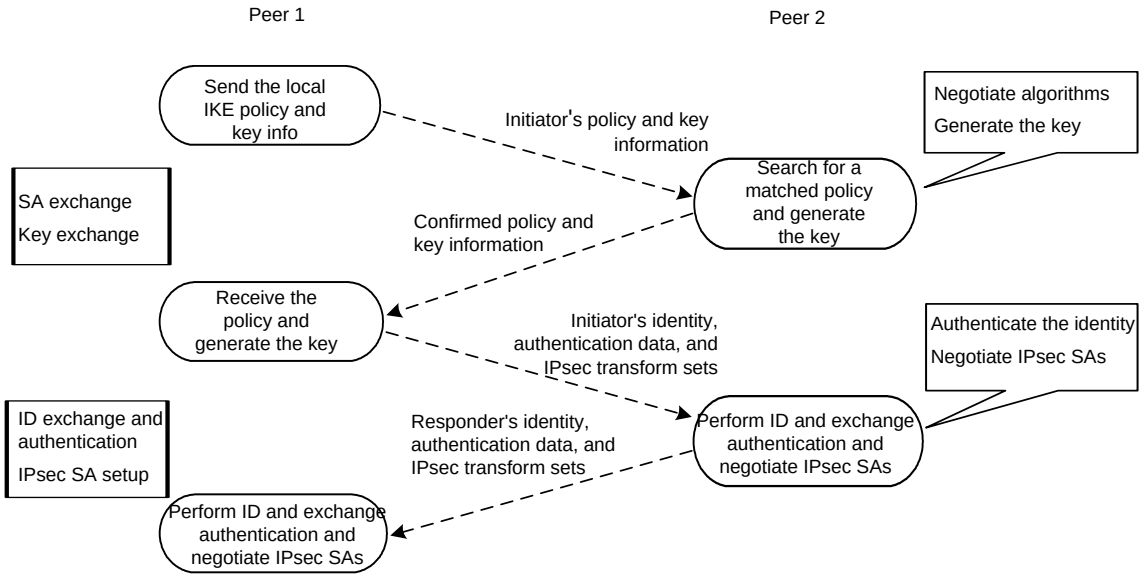
- **IKE_SA_INIT exchange**—Negotiates IKE SA parameters and exchanges keys.
- **IKE_AUTH exchange**—Authenticates the identity of the peer and establishes IPsec SAs.

After the four-message initial exchanges, IKEv2 sets up one IKE SA and one pair of IPsec SAs. For IKEv1 to set up one IKE SA and one pair of IPsec SAs, it must go through two phases that use a minimum of six messages.

To set up one more pair of IPsec SAs within the IKE SA, IKEv2 goes on to perform an additional two-message exchange—the CREATE_CHILD_SA exchange. One CREATE_CHILD_SA exchange creates one pair of IPsec SAs. IKEv2 also uses the CREATE_CHILD_SA exchange to rekey IKE SAs and Child SAs.

IKEv2 uses the INFORMATIONAL exchange to convey control messages about errors and notifications.

Figure 16 IKEv2 Initial exchange process



New features in IKEv2

DH guessing

In the IKE_SA_INIT exchange, the initiator guesses the DH group that the responder is most likely to use and sends it in an IKE_SA_INIT request message. If the initiator's guess is correct, the responder responds with an IKE_SA_INIT response message and the IKE_SA_INIT exchange is finished. If the guess is wrong, the responder responds with an INVALID_KE_PAYLOAD message that contains the DH group that it wants to use. The initiator then uses the DH group selected by the responder to reinitiate the IKE_SA_INIT exchange. The DH guessing mechanism allows for more flexible DH group configuration and enables the initiator to adapt to different responders.

Cookie challenging

Messages for the IKE_SA_INIT exchange are in plain text. An IKEv1 responder cannot confirm the validity of the initiators and must maintain half-open IKE SAs, which makes the responder susceptible to DoS attacks. An attacker can send a large number of IKE_SA_INIT requests with forged source IP addresses to the responder, exhausting the responder's system resources.

IKEv2 introduces the cookie challenging mechanism to prevent such DoS attacks. When an IKEv2 responder maintains a threshold number of half-open IKE SAs, it starts the cookie challenging mechanism. The responder generates a cookie and includes it in the response sent to the initiator. If the initiator initiates a new IKE_SA_INIT request that carries the correct cookie, the responder considers the initiator valid and proceeds with the negotiation. If the carried cookie is incorrect, the responder terminates the negotiation.

The cookie challenging mechanism automatically stops working when the number of half-open IKE SAs drops below the threshold.

IKEv2 SA rekeying

For security purposes, both IKE SAs and IPsec SAs have a lifetime and must be rekeyed when the lifetime expires. An IKEv1 SA lifetime is negotiated. An IKEv2 SA lifetime, in contrast, is configured. If two peers are configured with different lifetimes, the peer with the shorter lifetime always initiates the SA rekeying. This mechanism reduces the possibility that two peers will simultaneously initiate a rekeying. Simultaneous rekeying results in redundant SAs and SA status inconsistency on the two peers.

IKEv2 message retransmission

Unlike IKEv1 messages, IKEv2 messages appear in request/response pairs. IKEv2 uses the Message ID field in the message header to identify the request/response pair. If an initiator sends a request but receives no response with the same Message ID value within a specific period of time, the initiator retransmits the request.

It is always the IKEv2 initiator that initiates the retransmission, and the retransmitted message must use the same Message ID value.

Protocols and standards

- RFC 2408, *Internet Security Association and Key Management Protocol (ISAKMP)*
- RFC 4306, *Internet Key Exchange (IKEv2) Protocol*
- RFC 4718, *IKEv2 Clarifications and Implementation Guidelines*
- RFC 2412, *The OAKLEY Key Determination Protocol*
- RFC 5996, *Internet Key Exchange Protocol Version 2 (IKEv2)*

IKEv2 tasks at a glance

To configure IKEv2, perform the following tasks:

1. Configuring an IKEv2 profile
   a. Creating an IKEv2 profileSpecifying the local and remote identity authentication methodsConfiguring the IKEv2 keychain or PKI domain
   b. Configuring the local ID for the IKEv2 profile
   c. Configuring peer IDs for the IKEv2 profile
   d. Specifying a VPN instance for the IKEv2 profile
   e. Configuring optional features for the IKEv2 profile
2. Configuring an IKEv2 policy
3. Configuring an IKEv2 proposal

   If you specify an IKEv2 proposal in an IKEv2 policy, you must configure the IKEv2 proposal.
4. Configuring an IKEv2 keychain

   This task is required when either end or both ends use the preshared key authentication method.
5. (Optional.) Enabling the cookie challenging feature

   The cookie challenging feature takes effect only on IKEv2 responders.
6. (Optional.) Configuring the IKEv2 DPD feature
7. (Optional.) Configuring the IKEv2 NAT keepalive feature
8. (Optional.) Configuring IKEv2 address pools

Prerequisites for IKEv2 configuration

Determine the following parameters prior to IKEv2 configuration:

- The strength of the algorithms for IKEv2 negotiation, including the encryption algorithms, integrity protection algorithms, PRF algorithms, and DH groups. Different algorithms provide different levels of protection. A stronger algorithm means better resistance to decryption of protected data but requires more resources. Typically, the longer the key, the stronger the algorithm.
- The local and remote identity authentication methods.
  - To use the preshared key authentication method, you must determine the preshared key.
  - To use the RSA digital signature authentication method, you must determine the PKI domain for the local end to use. For information about PKI, see "Configuring PKI."

Configuring an IKEv2 profile

Creating an IKEv2 profile

About this task

An IKEv2 profile is intended to provide a set of parameters for IKEv2 negotiation.

Procedure

1. Enter system view.

   `system-view`
2. Create an IKEv2 profile and enter its view.

   `ikev2 profile` *`profile-name`*

Specifying the local and remote identity authentication methods

Restrictions and guidelines

The local and remote identity authentication methods must both be specified and they can be different. You can specify only one local identity authentication method and multiple remote identity authentication methods.

Procedure

1. Enter system view.

   system-view

2. Enter IKEv2 profile view.

   ikev2 profile *profile-name*

3. Specify the local and remote identity authentication methods.

   In non-FIPS mode:

   authentication-method { **local** | **remote** } { **dsa-signature** | **ecdsa-signature** | **pre-share** | **rsa-signature** }

   In FIPS mode:

   authentication-method { **local** | **remote** } { **ecdsa-signature** | **pre-share** | **rsa-signature** }

   By default, no local or remote identity authentication method is configured.

Configuring the IKEv2 keychain or PKI domain

Restrictions and guidelines

Configure the IKEv2 keychain or PKI domain for the IKEv2 profile to use. To use digital signature authentication, configure a PKI domain. To use preshared key authentication, configure an IKEv2 keychain.

Procedure

1. Enter system view.

   system-view

2. Enter IKEv2 profile view.

   ikev2 profile *profile-name*

3. Specify the keychain for preshared key authentication or the PKI domain used to request a certificate for digital signature authentication.

   - Specify the keychain.

     keychain *keychain-name*

   - Specify the PKI domain.

     certificate domain *domain-name* [**sign** | **verify**]

   By default, no IKEv2 keychain or PKI domain is specified for an IKEv2 profile.

Configuring the local ID for the IKEv2 profile

Restrictions and guidelines

For digital signature authentication, the device can use an ID of any type. If the local ID is an IP address that is different from the IP address in the local certificate, the device uses the FQDN as the local ID. The FQDN is the device name configured by using the **sysname** command.

For preshared key authentication, the device can use an ID of any type other than the DN.

Procedure

1. Enter system view.

   system-view

2. Enter IKEv2 profile view.

   ikev2 profile *profile-name*

3. Configure the local ID.

   identity local { **address** { *ipv4-address* | **ipv6** *ipv6-address* } | **dn** | **email** *email-string* | **fqdn** *fqdn-name* | **key-id** *key-id-string* }

   By default, no local ID is configured, and the device uses the IP address of the interface where the IPsec policy applies as the local ID.

Configuring peer IDs for the IKEv2 profile

About this task

Perform this task to configure the peer ID for IKEv2 profile matching. When the device needs to select an IKEv2 profile for IKEv2 negotiation with a peer, it compares the received peer ID with the peer IDs of its local IKE profiles. If a match is found, it uses the IKEv2 profile with the matching peer ID for negotiation. IKEv2 profiles will be compared in descending order of their priorities.

Procedure

1. Enter system view.

   system-view

2. Enter IKEv2 profile view.

   ikev2 profile *profile-name*

3. Configure a peer ID.

   match remote { **certificate** *policy-name* | **identity** { **address** { { *ipv4-address* [*mask* | *mask-length*] | **range** *low-ipv4-address high-ipv4-address* } | **ipv6** { *ipv6-address* [*prefix-length*] | **range** *low-ipv6-address high-ipv6-address* } } | **fqdn** *fqdn-name* | **email** *email-string* | **key-id** *key-id-string* } }

   You must configure a minimum of one peer ID on each of the two peers.

Specifying a VPN instance for the IKEv2 profile

About this task

After you specify a VPN instance for an IKEv2 profile, the IKEv2 profile can be used for IKEv2 negotiation only on the interfaces that belong to the VPN instance.

Procedure

1. Enter system view.

system-view

2. Enter IKEv2 profile view.
   ikev2 profile *profile-name*
3. Specify a VPN instance for the IKEv2 profile.
   match vrf { **name** *vrf-name* | **any** }
   By default, an IKEv2 profile belongs to the public network.

Configuring optional features for the IKEv2 profile

1. Enter system view.
   system-view
2. Enter IKEv2 profile view.
   ikev2 profile *profile-name*
3. Configure optional features as needed.
   - Configure IKEv2 DPD.
     dpd interval *interval* [**retry** *seconds*] { **on-demand** | **periodic** }
     By default, IKEv2 DPD is not configured for an IKEv2 profile and an IKEv2 profile uses the DPD settings configured in system view. If IKEv2 DPD is not configured in system view either, the device does not perform dead IKEv2 peer detection.
   - Specify the local interface or IP address to which the IKEv2 profile can be applied.
     match local address { *interface-type interface-number* | *ipv4-address* | **ipv6** *ipv6-address* }
     By default, an IKEv2 profile can be applied to any local interface or local IP address.
     Use this command to specify which address or interface can use the IKEv2 profile for IKEv2 negotiation. Specify the local address configured in IPsec policy or IPsec policy template view (using the **local-address** command) for this command. If no local address is configured, specify the IP address of the interface that uses the IPsec policy.
   - Specify a priority for the IKEv2 profile.
     priority *priority*
     By default, the priority of an IKEv2 profile is 100.
     When the device needs to select an IKEv2 profile for IKEv2 negotiation with a peer, it compares the received peer ID with the peer ID of its local IKEv2 profiles in descending order of their priorities
   - Set the IKEv2 SA lifetime for the IKEv2 profile.
     sa duration *seconds*
     By default, the IKEv2 SA lifetime is 86400 seconds.
     The local and remote ends can use different IKEv2 SA lifetimes and they do not negotiate the lifetime. The end with a smaller SA lifetime will initiate an SA negotiation when the lifetime expires.
   - Set the IKEv2 NAT keepalive interval.
     nat-keepalive *seconds*
     By default, the global IKEv2 NAT keepalive setting is used.
     Configure this command when the device is behind a NAT gateway. The device sends NAT keepalive packets regularly to its peer to prevent the NAT session from being aged because of no matching traffic.
   - Enable the configuration exchange feature.
     config-exchange { **request** | **set** { **accept** | **send** } }

By default, all configuration exchange options are disabled.

This feature applies to scenarios where the headquarters and branches communicate through virtual tunnels. It enables exchanges of IP address request and set messages between the IPsec gateway at a branch and the IPsec gateway at the headquarters.

Table 2 Parameter descriptions

Parameter	Description
request	Enables the IPsec gateway at a branch to submit IP address request messages to the IPsec gateway at the headquarters.
set accept	Enables the IPsec gateway at a branch to accept the IP addresses pushed by the IPsec gateway at the headquarters.
set send	Enables the IPsec gateway at the headquarters to push IP addresses to IPsec gateways at branches.

- Enable AAA authorization.

  aaa authorization domain *domain-name* **username** *user-name*

  By default, AAA authorization is disabled.

  The AAA authorization feature enables IKEv2 to request authorization attributes, such as the IKEv2 address pool, from AAA. IKEv2 uses the address pool to assign IP addresses to remote users. For more information about AAA authorization, see "Configuring AAA."

Configuring an IKEv2 policy

About this task

During the IKE_SA_INIT exchange, each end tries to find a matching IKEv2 policy, using the IP address of the local security gateway as the matching criterion.

- If IKEv2 policies are configured, IKEv2 searches for an IKEv2 policy that uses the IP address of the local security gateway. If no IKEv2 policy uses the IP address or the policy is using an incomplete proposal, the IKE_SA_INIT exchange fails.
- If no IKEv2 policy is configured, IKEv2 uses the system default IKEv2 policy **default**.

The device matches IKEv2 policies in the descending order of their priorities. To determine the priority of an IKEv2 policy:

1. First, the device examines the existence of the **match local address** command. An IKEv2 policy with the **match local address** command configured has a higher priority.
2. If a tie exists, the device compares the priority numbers. An IKEv2 policy with a smaller priority number has a higher priority.
3. If a tie still exists, the device prefers an IKEv2 policy configured earlier.

Procedure

1. Enter system view.

   system-view
2. Create an IKEv2 policy and enter its view.

   ikev2 policy *policy-name*

   By default, an IKEv2 policy named **default** exists.
3. Specify the local interface or address used for IKEv2 policy matching.

   match local address { *interface-type interface-number* | *ipv4-address* | **ipv6** *ipv6-address* }

By default, no local interface or address is used for IKEv2 policy matching, and the policy matches any local interface or address.

4. Specify a VPN instance for IKEv2 policy matching.

   match vrf { **name** *vrf-name* | **any** }

   By default, no VPN instance is specified for IKEv2 policy matching. The IKEv2 policy matches all local addresses in the public network.

5. Specify an IKEv2 proposal for the IKEv2 policy.

   proposal *proposal-name*

   By default, no IKEv2 proposal is specified for an IKEv2 policy.

6. Specify a priority for the IKEv2 policy.

   priority *priority*

   By default, the priority of an IKEv2 policy is 100.

Configuring an IKEv2 proposal

About this task

An IKEv2 proposal contains security parameters used in IKE_SA_INIT exchanges, including the encryption algorithms, integrity protection algorithms, PRF algorithms, and DH groups. An algorithm specified earlier has a higher priority.

Restrictions and guidelines

A complete IKEv2 proposal must have at least one set of security parameters, including one encryption algorithm, one integrity protection algorithm, one PRF algorithm, and one DH group.

You can specify multiple IKEv2 proposals for an IKEv2 policy. A proposal specified earlier has a higher priority.

Procedure

1. Enter system view.

   system-view

2. Create an IKEv2 proposal and enter its view.

   ikev2 proposal *proposal-name*

   By default, an IKEv2 proposal named **default** exists.

   In non-FIPS mode, the default proposal uses the following settings:

   - Encryption algorithms AES-CBC-128 and 3DES.
   - Integrity protection algorithms HMAC-SHA1 and HMAC-MD5.
   - PRF algorithms HMAC-SHA1 and HMAC-MD5.
   - DH groups 2 and 5.

   In FIPS mode, the default proposal uses the following settings:

   - Encryption algorithms AES-CBC-128 and AES-CTR-128.
   - Integrity protection algorithms HMAC-SHA1 and HMAC-SHA256.
   - PRF algorithms HMAC-SHA1 and HMAC-SHA256.
   - DH groups 14 and 19.

3. Specify the encryption algorithms.

   In non-FIPS mode:

   encryption { **3des-cbc** | **aes-cbc-128** | **aes-cbc-192** | **aes-cbc-256** | **aes-ctr-128** | **aes-ctr-192** | **aes-ctr-256** | **camellia-cbc-128** |

camellia-cbc-192 | **camellia-cbc-256** | **des-cbc** | **aes-gcm-128** | **aes-gcm-192** | **aes-gcm-256** } *

In FIPS mode:

encryption { **aes-gcm-128** | **aes-gcm-192** | **aes-gcm-256** } *

By default, an IKEv2 proposal does not have any encryption algorithms.

4. Specify the integrity protection algorithms.

   In non-FIPS mode:

   integrity { **aes-xcbc-mac** | **md5** | **sha1** | **sha256** | **sha384** | **sha512** } *

   By default, an IKEv2 proposal does not have any integrity protection algorithms.

5. Specify the DH groups.

   In non-FIPS mode:

   dh { **group1** | **group14** | **group19** | **group2** | **group20** | **group24** | **group5** } *

   In FIPS mode:

   dh { **group19** | **group20** | **group24** } *

   By default, an IKEv2 proposal does not have any DH groups.

6. Specify the PRF algorithms.

   In non-FIPS mode:

   prf { **aes-xcbc-mac** | **md5** | **sha1** | **sha256** | **sha384** | **sha512** } *

   In FIPS mode:

   prf { **sha256** | **sha384** | **sha512** } *

   By default, an IKEv2 proposal uses the integrity protection algorithms as the PRF algorithms.

Configuring an IKEv2 keychain

About this task

An IKEv2 keychain specifies the preshared keys used for IKEv2 negotiation.

An IKEv2 keychain can have multiple IKEv2 peers. Each peer has a symmetric preshared key or an asymmetric preshared key pair, and information for identifying the peer (such as the peer's host name, IP address or address range, or ID).

An IKEv2 negotiation initiator uses the peer host name or IP address/address range as the matching criterion to search for a peer. A responder uses the peer host IP address/address range or ID as the matching criterion to search for a peer.

Procedure

1. Enter system view.

   system-view

2. Create an IKEv2 keychain and enter its view.

   ikev2 keychain *keychain-name*

3. Create an IKEv2 peer and enter its view.

   peer *name*

4. Configure a host name for the peer:

   hostname *name*

   By default, no host name is configured for an IKEv2 peer.

5. Configure a host IP address or address range for the peer:

address { *ipv4-address* [*mask* | *mask-length*] | **ipv6** *ipv6-address* [*prefix-length*] }

By default, no host IP address or address range is configured for an IKEv2 peer.

You must configure different host IP addresses/address ranges for different peers.

6. Configure an ID for the peer:

   identity { **address** { *ipv4-address* | **ipv6** { *ipv6-address* } } | **fqdn** *fqdn-name* | **email** *email-string* | **key-id** *key-id-string* }

   By default, no identity information is configured for an IKEv2 peer.

7. Configure a preshared key for the peer.

   pre-shared-key [**local** | **remote**] { **ciphertext** | **plaintext** } *string*

   By default, an IKEv2 peer does not have a preshared key.

Configure global IKEv2 parameters

Enabling the cookie challenging feature

About this task

Enable cookie challenging on responders to protect them against DoS attacks that use a large number of source IP addresses to forge IKE_SA_INIT requests.

Procedure

1. Enter system view.

   system-view

2. Enable IKEv2 cookie challenging.

   ikev2 cookie-challenge *number*

   By default, IKEv2 cookie challenging is disabled.

Configuring the IKEv2 DPD feature

About this task

IKEv2 DPD detects dead IKEv2 peers in periodic or on-demand mode.

- **Periodic IKEv2 DPD**—Verifies the liveness of an IKEv2 peer by sending DPD messages at regular intervals.
- **On-demand IKEv2 DPD**—Verifies the liveness of an IKEv2 peer by sending DPD messages before sending data.
  - Before the device sends data, it identifies the time interval for which the last IPsec packet has been received from the peer. If the time interval exceeds the DPD interval, it sends a DPD message to the peer to detect its liveliness.
  - If the device has no data to send, it never sends DPD messages.

Restrictions and guidelines

If you configure IKEv2 DPD in both IKEv2 profile view and system view, the IKEv2 DPD settings in IKEv2 profile view apply. If you do not configure IKEv2 DPD in IKEv2 profile view, the IKEv2 DPD settings in system view apply.

Procedure

1. Enter system view.

   system-view

2. Configure global IKEv2 DPD.

   ikev2 dpd interval *interval* [**retry** *seconds*] { **on-demand** | **periodic** }

   By default, global DPD is disabled.

Configuring the IKEv2 NAT keepalive feature

About this task

Configure this feature on the IKEv2 gateway behind the NAT device. The gateway then sends NAT keepalive packets regularly to its peer to keep the NAT session alive, so that the peer can access the device.

The NAT keepalive interval must be shorter than the NAT session lifetime.

This feature takes effect after the device detects the NAT device.

Procedure

1. Enter system view.

   system-view

2. Set the IKEv2 NAT keepalive interval.

   ikev2 nat-keepalive *seconds*

   By default, the IKEv2 NAT keepalive interval is 10 seconds.

Configuring IKEv2 address pools

About this task

To perform centralized management on remote users, an IPsec gateway can use an address pool to assign private IP addresses to remote users.

You must use an IKEv2 address pool together with AAA authorization by specifying the IKEv2 address pool as an AAA authorization attribute. For more information about AAA authorization, see "Configuring AAA."

Procedure

1. Enter system view.

   system-view

2. Configure an IKEv2 IPv4 address pool.

   ikev2 address-group *group-name start-ipv4-address end-ipv4-address* [*mask* | *mask-length*]

3. Configure an IKEv2 IPv6 address pool.

   ikev2 ipv6-address-group *group-name* **prefix** *prefix/prefix-len* **assign-len** *assign-len*

Display and maintenance commands for IKEv2

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the IKEv2 policy configuration.	**display ikev2 policy** [*policy-name* | **default**]

Task	Command
Display the IKEv2 profile configuration.	**display ikev2 profile** [*profile-name*]
Display the IKEv2 proposal configuration.	**display ikev2 proposal** [*name* | **default**]
Display the IKEv2 SA information.	**display ikev2 sa** [**count** | [{ **local** | **remote** } { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]] [**verbose** [**tunnel** *tunnel-id*]]]
Display IKEv2 statistics.	**display ikev2 statistics**
Delete IKEv2 SAs and the child SAs negotiated through the IKEv2 SAs.	**reset ikev2 sa** [[{ **local** | **remote** } { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]] | **tunnel** *tunnel-id*] [**fast**]
Clear IKEv2 statistics.	**reset ikev2 statistics**

Troubleshooting IKEv2

IKEv2 negotiation failed because no matching IKEv2 proposals were found

Symptom

The IKEv2 SA is in IN-NEGO status.

```
<Sysname> display ikev2 sa
Tunnel ID   Local                       Remote                      Status
  ------------------------------------------------------------------------------
  5           123.234.234.124/500         123.234.234.123/500         IN-NEGO
Status:
IN-NEGO: Negotiating, EST: Established, DEL:Deleting
```

Analysis

Certain IKEv2 proposal settings are incorrect.

Solution

1. Examine the IKEv2 proposal configuration to see whether the two ends have matching IKEv2 proposals.
2. Modify the IKEv2 proposal configuration to make sure the two ends have matching IKEv2 proposals.

IPsec SA negotiation failed because no matching IPsec transform sets were found

Symptom

The **display ikev2 sa** command shows that the IKEv2 SA negotiation succeeded and the IKEv2 SA is in EST status. The **display ipsec sa** command shows that the expected IPsec SAs have not been negotiated yet.

Analysis

Certain IPsec policy settings are incorrect.

Solution

1. Examine the IPsec configuration to see whether the two ends have matching IPsec transform sets.
2. Modify the IPsec configuration to make sure the two ends have matching IPsec transform sets.

IPsec tunnel establishment failed

Symptom

The ACLs and IKEv2 proposals are correctly configured on both ends. The two ends cannot establish an IPsec tunnel or cannot communicate through the established IPsec tunnel.

Analysis

The IKEv2 SA or IPsec SAs on either end are lost. The reason might be that the network is unstable and the device reboots.

Solution

1. Use the **display ikev2 sa** command to examine whether an IKEv2 SA exists on both ends. If the IKEv2 SA on one end is lost, delete the IKEv2 SA on the other end by using the **reset ikev2 sa** command and trigger new negotiation. If an IKEv2 SA exists on both ends, go to the next step.
2. Use the **display ipsec sa** command to examine whether IPsec SAs exist on both ends. If the IPsec SAs on one end are lost, delete the IPsec SAs on the other end by using the **reset ipsec sa** command and trigger new negotiation.

Contents

Configuring SSH 1

About SSH 1
SSH applications 1
How SSH works 1
SSH authentication methods 2
SSH support for Suite B 3
FIPS compliance 4
Configuring the device as an SSH server 4
SSH server tasks at a glance 4
Generating local key pairs 5
Specifying the SSH service port 5
Enabling the Stelnet server 6
Enabling the SFTP server 6
Enabling the SCP server 6
Enabling NETCONF over SSH 7
Configuring the user lines for SSH login 7
Configuring a client's host public key 7
Configuring an SSH user 9
Configuring the SSH management parameters 10
Specifying a PKI domain for the SSH server 13
Disconnecting SSH sessions 13
Configuring the device as an Stelnet client 14
Stelnet client tasks at a glance 14
Generating local key pairs 14
Specifying the source IP address for outgoing SSH packets 14
Establishing a connection to an Stelnet server 15
Deleting server public keys saved in the public key file on the Stelnet client 17
Establishing a connection to an Stelnet server based on Suite B 17
Configuring the device as an SFTP client 17
SFTP client tasks at a glance 17
Generating local key pairs 18
Specifying the source IP address for outgoing SFTP packets 18
Establishing a connection to an SFTP server 19
Deleting server public keys saved in the public key file on the SFTP client 20
Establishing a connection to an SFTP server based on Suite B 21
Working with SFTP directories 21
Working with SFTP files 22
Displaying help information 23
Terminating the connection with the SFTP server 23
Configuring the device as an SCP client 23
SCP client tasks at a glance 23
Generating local key pairs 24
Specifying the source IP address for outgoing SCP packets 24
Establishing a connection to an SCP server 25
Deleting server public keys saved in the public key file on the SCP client 26
Establishing a connection to an SCP server based on Suite B 27
Specifying algorithms for SSH2 27
About algorithms for SSH2 27
Specifying key exchange algorithms for SSH2 27
Specifying public key algorithms for SSH2 28
Specifying encryption algorithms for SSH2 28
Specifying MAC algorithms for SSH2 28
Display and maintenance commands for SSH 29
Stelnet configuration examples 29
Example: Configuring the device as an Stelnet server (password authentication) 30
Example: Configuring the device as an Stelnet server (publickey authentication) 32
Example: Configuring the device as an Stelnet client (password authentication) 38

Example: Configuring the device as an Stelnet client (publickey authentication) ····42
Example: Configuring Stelnet based on 128-bit Suite B algorithms ····44
SFTP configuration examples ····48
Example: Configuring the device as an SFTP server (password authentication) ····48
Example: Configuring the device as an SFTP client (publickey authentication) ····50
Example: Configuring SFTP based on 192-bit Suite B algorithms ····54
SCP configuration examples ····57
Example: Configuring SCP with password authentication ····58
Example: Configuring SCP file transfer with a Linux SCP client ····59
Example: Configuring SCP based on Suite B algorithms ····61
NETCONF over SSH configuration examples ····68
Example: Configuring NETCONF over SSH with password authentication ····68

Configuring SSH

About SSH

Secure Shell (SSH) is a network security protocol. Using encryption and authentication, SSH can implement secure remote access and file transfer over an insecure network.

SSH uses the typical client-server model to establish a channel for secure data transfer based on TCP.

SSH includes two versions: SSH1.x and SSH2.0 (hereinafter referred to as SSH1 and SSH2), which are not compatible. SSH2 is better than SSH1 in performance and security.

SSH applications

The device supports the following SSH applications:

- **Secure Telnet**—Stelnet provides secure and reliable network terminal access services. Through Stelnet, a user can securely log in to a remote server. Stelnet can protect devices against attacks, such as IP spoofing and plain text password interception. The device can act as an Stelnet server or an Stelnet client.
- **Secure File Transfer Protocol**—Based on SSH2, SFTP uses SSH connections to provide secure file transfer. The device can act as an SFTP server, allowing a remote user to log in to the SFTP server for secure file management and transfer. The device can also act as an SFTP client, enabling a user to log in from the device to a remote device for secure file transfer.
- **Secure Copy**—Based on SSH2, SCP offers a secure method to copy files. The device can act as an SCP server, allowing a user to log in to the device for file upload and download. The device can also act as an SCP client, enabling a user to log in from the device to a remote device for secure file transfer.
- **NETCONF over SSH**—Based on SSH2, it enables users to securely log in to the device through SSH and perform NETCONF operations on the device through the NETCONF-over-SSH connections. The device can act only as a NETCONF-over-SSH server. For more information about NETCONF, see *Network Management and Monitoring Configuration Guide*.

The device supports different SSH versions depending on its role:

- When acting as an Stelnet server, the device supports both SSH2 and SSH1 in non-FIPS mode and only SSH2 in FIPS mode.
- When acting as an SSH client, the device supports only SSH2.
- When acting as an SFTP, SCP, or NETCONF-over-SSH server, the device supports only SSH2.

How SSH works

This section uses SSH2 as an example to describe the stages to establish an SSH session.

Table 1 Stages to establish an SSH session

Stages	Description
Connection establishment	The SSH server listens to connection requests on port 22. After a client initiates a connection request, the server and the client establish a TCP connection.
Version negotiation	The two parties determine a version to use.

Stages	Description
Algorithm negotiation	SSH supports multiple algorithms. Based on the local algorithms, the two parties negotiate the following algorithms: • Key exchange algorithm for generating session keys. • Encryption algorithm for encrypting data. • Public key algorithm for the digital signature and authentication. • HMAC algorithm for protecting data integrity.
Key exchange	The two parties use the DH exchange algorithm to dynamically generate the session keys and session ID. • The session keys are used for protecting data transfer. • The session ID is used for identifying the SSH connection. In this stage, the client also authenticates the server.
Authentication	The SSH server authenticates the client in response to the client's authentication request.
Session request	After passing the authentication, the client sends a session request to the server to request the establishment of a session for the Stelnet, NETCONF, SCP, or SFTP service.
Interaction	After the server grants the request, the client and the server start to communicate with each other in the session. In this stage, you can paste commands in text format and execute them at the CLI. The text pasted at one time must be no more than 2000 bytes. As a best practice to ensure the correct execution of commands, paste commands that are in the same view. To execute commands of more than 2000 bytes, save the commands in a configuration file, upload the file to the server through SFTP, and use it to restart the server.

SSH authentication methods

This section describes authentication methods that are supported by the device when it acts as an SSH server.

Password authentication

The SSH server authenticates a client through the AAA mechanism. The password authentication process is as follows:

1. The client sends the server an authentication request that includes the encrypted username and password.
2. The server performs the following operations:
   a. Decrypts the request to get the username and password in plain text.
   b. Verifies the username and password locally or through remote AAA authentication.
   c. Informs the client of the authentication result.

If the remote AAA server requires the user to enter a password for secondary authentication, it send the SSH server an authentication response carrying a prompt. The prompt is transparently transmitted to the client to notify the user to enter a specific password. When the user enters the correct password, the AAA sever examines the password validity. If the password is valid, the SSH server returns an authentication success message to the client.

SSH1 clients do not support secondary password authentication initiated by the AAA server.

For more information about AAA, see "Configuring AAA."

Keyboard-interactive authentication

In keyboard-interactive authentication, the remote authentication server and user exchanges information for authentication as follows:

1. The remote authentication server sends a prompt to the SSH server in an authentication response.
2. The prompt indicates the information required to be provided by the user.
3. The SSH server transparently transmits the prompt to the client terminal.
4. The user enters the required information as prompted.

This process repeats multiple times if the remote authentication server requires more interactive information. The remote authentication server returns an authentication success message after the user provides all required interactive information.

If the remote authentication server does not require interactive information, the keyboard-interactive authentication process is the same as the password authentication.

Publickey authentication

The server authenticates a client by verifying the digital signature of the client. The publickey authentication process is as follows:

1. The client sends the server a publickey authentication request that includes the username, public key, and public key algorithm name.

   If the digital certificate of the client is required in authentication, the client also encapsulates the digital certificate in the authentication request. The digital certificate carries the public key information of the client.
2. The server verifies the client's public key.
   - If the public key is invalid, the server informs the client of the authentication failure.
   - If the public key is valid, the server requests the digital signature of the client. After receiving the signature, the server uses the public key to verify the signature and informs the client of the authentication result.

When acting as an SSH server, the device supports using the public key algorithms DSA, ECDSA, and RSA to verify digital signatures.

When acting as an SSH client, the device supports using the public key algorithms DSA, ECDSA, and RSA to generate digital signatures.

For more information about public key configuration, see "Managing public keys."

Password-publickey authentication

The server requires SSH2 clients to pass both password authentication and publickey authentication. However, an SSH1 client only needs to pass either authentication.

Any authentication

The server requires clients to pass keyboard-interactive authentication, password authentication, or publickey authentication. Success with any one authentication method is sufficient to connect to the server.

SSH support for Suite B

Suite B contains a set of encryption and authentication algorithms that meet high security requirements. Table 2 lists all algorithms in Suite B.

The SSH server and client support using the X.509v3 certificate for identity authentication in compliance with the algorithm, negotiation, and authentication specifications defined in RFC 6239.

Table 2 Suite B algorithms

Security level	Key exchange algorithm	Encryption algorithm and HMAC algorithm	Public key algorithm
128-bit	ecdh-sha2-nistp256	AES128-GCM	x509v3-ecdsa-sha2-nistp256
192-bit	ecdh-sha2-nistp384	AES256-GCM	x509v3-ecdsa-sha2-nistp384
Both	ecdh-sha2-nistp256 ecdh-sha2-nistp384	AES128-GCM AES256-GCM	x509v3-ecdsa-sha2-nistp256 x509v3-ecdsa-sha2-nistp384

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see "Configuring FIPS."

Configuring the device as an SSH server

SSH server tasks at a glance

To configure an SSH server, perform the following tasks:

1. Generating local key pairs
2. (Optional.) Specifying the SSH service port
3. Enabling the SSH server
   - Enabling the Stelnet server
   - Enabling the SFTP server
   - Enabling the SCP server
   - Enabling NETCONF over SSH
4. Configuring the user lines for SSH login

   Required only for Stelnet and NETCONF-over-SSH servers.
5. Configuring a client's host public key

   Required for authentication method **publickey**, **password-publickey,** or **any**.
6. Configuring an SSH user
   - Required for authentication method **keyboard-interactive**, **publickey**, **password-publickey,** or **any**.
   - Optional for the **password** authentication method.
7. (Optional.) Configuring the SSH management parameters

   SSH management settings, such as authentication and connection control settings, help improve security of SSH connections.
8. (Optional.) Specifying a PKI domain for the SSH server
9. (Optional.) Disconnecting SSH sessions

Generating local key pairs

About this task

The DSA, ECDSA, or RSA key pairs on the SSH server are required for generating the session keys and session ID in the key exchange stage. They can also be used by a client to authenticate the server. When a client authenticates the server, it compares the public key received from the server with the server's public key that the client saved locally. If the keys are consistent, the client uses the locally saved server's public key to decrypt the digital signature received from the server. If the decryption succeeds, the server passes the authentication.

To support SSH clients that use different types of key pairs, generate DSA, ECDSA, and RSA key pairs on the SSH server.

- **RSA key pairs**—The SSH server generates a server key pair and a host key pair for RSA. The RSA server key pair is only used in SSH1 to encrypt the session key for secure transmission of the session key. It is not used in SSH2, because no session key transmission is required in SSH2.
- **DSA key pair**—The SSH server generates only one DSA host key pair. SSH1 does not support the DSA algorithm.
- **ECDSA key pair**—The SSH server generates only one ECDSA host key pair.

Restrictions and guidelines

Local DSA, ECDSA, and RSA key pairs for SSH use default names. You cannot assign names to the key pairs. For more information about the command for generating key pairs, see public key management commands in *Security Command Reference*.

If the device does not have RSA key pairs with default names, it automatically generates one RSA server key pair and one RSA host key pair when SSH starts. Both key pairs use their default names. The SSH application starts when you execute an SSH server command on the device.

The key modulus length must be less than 2048 bits when you generate the DSA key pair on the SSH server.

When you generate an ECDSA key pair, you can generate only a **secp256r1** or **secp384r1** ECDSA key pair.

The SSH server operating in FIPS mode supports only ECDSA and RSA key pairs. Do not generate a DSA key pair on the SSH server in FIPS mode.

Procedure

1. Enter system view.

   system-view
2. Generate local key pairs.

   public-key local create { **dsa** | **ecdsa** { **secp256r1** | **secp384r1** } | **rsa** }

Specifying the SSH service port

About this task

The default port of the SSH service is 22. You can specify another port for the SSH service to improve security of SSH connections.

Restrictions and guidelines

If you modify the SSH port number when the SSH server is enabled, the SSH service is restarted and all SSH connections are terminated after the modification. SSH users must reconnect to the SSH server to access the server.

If you set the SSH port to a well-known port number, the service that uses the well-known port number might fail to start. Well-known port numbers are in the range of 1 to 1024.

Procedure

1. Enter system view.
   system-view
2. Specify the SSH service port.
   ssh server port *port-number*
   By default, the SSH service port is 22.

Enabling the Stelnet server

About this task

After you enable the Stelnet server on the device, a client can log in to the device through Stelnet.

Procedure

1. Enter system view.
   system-view
2. Enable the Stelnet server.
   ssh [**ipv4** | **ipv6**] **server enable**
   By default, the Stelnet server is disabled.

Enabling the SFTP server

About this task

After you enable the SFTP server on the device, a client can log in to the device through SFTP.

Restrictions and guidelines

When acting as an SFTP server, the device does not support SFTP connections initiated by SSH1 clients.

Procedure

1. Enter system view.
   system-view
2. Enable the SFTP server.
   sftp [**ipv4** | **ipv6**] **server enable**
   By default, the SFTP server is disabled.

Enabling the SCP server

About this task

After you enable the SCP server on the device, a client can log in to the device through SCP.

Restrictions and guidelines

When acting as an SCP server, the device does not support SCP connections initiated by SSH1 clients.

Procedure

1. Enter system view.

system-view

2. Enable the SCP server.

   scp [**ipv4** | **ipv6**] **server enable**

   By default, the SCP server is disabled.

Enabling NETCONF over SSH

About this task

After you enable NETCONF over SSH on the device, a client can perform NETCONF operations on the device through a NETCONF-over-SSH connection.

Restrictions and guidelines

When acting as a server in the NETCONF-over-SSH connection, the device does not support connection requests initiated by SSH1 clients.

Procedure

1. Enter system view.

   system-view

2. Enable NETCONF over SSH.

   netconf ssh server enable

   By default, NETCONF over SSH is disabled.

   For more information about NETCONF over SSH commands, see *Network Management and Monitoring Command Reference*.

Configuring the user lines for SSH login

About this task

Depending on the SSH application, an SSH client can be an Stelnet client, SFTP client, SCP client, or NETCONF-over-SSH client.

Only Stelnet and NETCONF-over-SSH clients require the user line configuration. The user line configuration takes effect on the clients at the next login.

Procedure

1. Enter system view.

   system-view

2. Enter VTY user line view.

   line vty *number* [*ending-number*]

3. Set the login authentication mode to scheme.

   authentication-mode scheme

   By default, the authentication mode is **password**.

   For more information about this command, see *Fundamentals Command Reference*.

Configuring a client's host public key

About this task

In publickey authentication, the server compares the SSH username and the client's host public key received from the client with the locally saved SSH username and the client's host public key. If they

are the same, the server checks the digital signature that the client sends. The client generates the digital signature by using the private key that is paired with the client's host public key.

For publickey authentication, password-publickey authentication, or any authentication, you must perform the following tasks:

1. Configure the client's DSA, ECDSA, or RSA host public key on the server.
2. Specify the associated host private key on the client to generate the digital signature.

   If the device acts as an SSH client, specify the public key algorithm on the client. The algorithm determines the associated host private key for generating the digital signature.

Client public key configuration methods

You can configure the client host public key by using the following methods:

- Manually enter the content of a client's host public key on the server.
  a. Display the host public key on the client and record the key.
  b. Type the client's host public key character by character on the server, or use the copy and paste method.

  The manually entered key must be in DER format without being converted. For the displayed key to meet the requirement when the client is an H3C device, use the **display public-key local public** command. The format of the public key displayed in any other way (for example, by using the **public-key local export** command) might be incorrect. If the key is not in correct format, the system discards the key.
- Import the client host public key from a public key file.
  c. Save the client public key file to the server. For example, transfer the client public key file to the server in binary mode through FTP or TFTP.
  d. Import the client public key from the locally saved public key file.

  During the import process, the server automatically converts the host public key to a string in PKCS format.

Restrictions and guidelines

As a best practice, configure no more than 20 SSH client's host public keys on an SSH server.

Import the client's host public key as a best practice.

Entering a client's host public key

1. Enter system view.

   system-view
2. Enter public key view.

   public-key peer *keyname*
3. Configure a client's host public key.

   Enter the content of the client's host public key character by character, or use the copy and paste method.

   When you enter the content of a client's host public key, you can use spaces and carriage returns between characters but the system does not save them. For more information, see "Managing public keys."
4. Exit public key view and save the key.

   peer-public-key end

Importing a client's host public key from the public key file

1. Enter system view.

   system-view
2. Import a client's public key from the public key file.

public-key peer *keyname* **import sshkey** *filename*

Configuring an SSH user

About this task

Configure an SSH user and a local user depending on the authentication method.

- If the authentication method is **publickey**, you must create an SSH user and a local user on the SSH server. The two users must have the same username, so that the SSH user can be assigned the correct working directory and user role.
- If the authentication method is **password**, you must perform one of the following tasks:
  - For local authentication, configure a local user on the SSH server.
  - For remote authentication, configure an SSH user on a remote authentication server, for example, a RADIUS server.

  You do not need to create an SSH user by using the **ssh user** command. However, if you want to display all SSH users, including the password-only SSH users, for centralized management, you can use this command to create them. If such an SSH user has been created, make sure you have specified the correct service type and authentication method.
- If the authentication method is **keyboard-interactive**, **password-publickey**, or **any**, you must create an SSH user on the SSH server and perform one of the following tasks:
  - For local authentication, configure a local user on the SSH server.
  - For remote authentication, configure an SSH user on a remote authentication server, for example, a RADIUS server.

  In either case, the local user or the SSH user configured on the remote authentication server must have the same username as the SSH user.

For information about configuring local users and remote authentication, see "Configuring AAA."

Restrictions and guidelines

If you change the authentication parameters for a logged-in SSH user, the change takes effect on the user at the next login.

When the device operates as an SSH server in FIPS mode, the device does not support authentication method **any** or **publickey**.

For an SFTP or SCP user, the working directory depends on the authentication method.

- If the authentication method is **publickey** or **password-publickey**, the working folder is specified by the **authorization-attribute** command in the associated local user view.
- If the authentication method is **keyboard-interactive** or **password**, the working directory is authorized by AAA.

For an SSH user, the user role also depends on the authentication method.

- If the authentication method is **publickey** or **password-publickey**, the user role is specified by the **authorization-attribute** command in the associated local user view.
- If the authentication method is **keyboard-interactive** or **password**, the user role is authorized by AAA.

For all authentication methods except keyboard-interactive authentication and password authentication, you must specify a client's host public key or digital certificate.

- For a client that sends the user's public key information directly to the server, specify the client's host public key on the server. The specified public key must already exist. For more information about public keys, see "Configuring a client's host public key." If you specify multiple client public keys, the device verifies the user identity by using the public keys in the order they are specified. The user is valid if the user passes one public key check.

- For a client that sends the user's public key information to the server through a digital certificate, specify the PKI domain on the server. This PKI domain verifies the client's digital certificate. For successful verification, the specified PKI domain must have the correct CA certificate. To specify the PKI domain, use the **ssh user** or **ssh server pki-domain** command. For more information about configuring a PKI domain, see "Configuring PKI."

Procedure

1. Enter system view.

   system-view

2. Create an SSH user, and specify the service type and authentication method.

   In non-FIPS mode:

   ssh user *username* **service-type** { **all** | **netconf** | **scp** | **sftp** | **stelnet** } **authentication-type** { **keyboard-interactive** | **password** | { **any** | **password-publickey** | **publickey** } [**assign** { **pki-domain** *domain-name* | **publickey** *keyname*&<1-6> }] }

   In FIPS mode:

   ssh user *username* **service-type** { **all** | **netconf** | **scp** | **sftp** | **stelnet** } **authentication-type** { **keyboard-interactive** | **password** | **password-publickey** [**assign** { **pki-domain** *domain-name* | **publickey** *keyname*&<1-6> }] }

   An SSH server supports up to 1024 SSH users.

Configuring the SSH management parameters

Enabling the SSH server to support SSH1 clients

1. Enter system view.

   system-view

2. Enable the SSH server to support SSH1 clients.

   ssh server compatible-ssh1x enable

   By default, the SSH server does not support SSH1 clients.

   This command is not available in FIPS mode.

Enabling SSH algorithm renegotiation and key re-exchange

1. Enter system view.

   system-view

2. Use one of the following methods or both at the same time to trigger SSH algorithm renegotiation and key re-exchange:

   - Enable SSH algorithm renegotiation and key re-exchange.
     - In non-FIPS mode:

       ssh server key-re-exchange enable [**interval** *interval*]

     - In FIPS mode:

       ssh server key-re-exchange enable [**interval-minutes** *interval*]

     By default, SSH algorithm renegotiation and key re-exchange are disabled.

     The command takes effect only on new SSH connections that are established after the command is executed, and it does not affect existing SSH connections.

   - Configure the traffic threshold that triggers SSH algorithm renegotiation and key re-exchange

     ssh server key-re-exchange traffic-threshold *traffic-threshold*

By default, the traffic threshold that enables SSH algorithm renegotiation and key re-exchange depends on the encryption algorithm for the negotiation:

- For DES and 3DESC encryption algorithms, the traffic threshold that triggers SSH algorithm renegotiation and key re-exchange is 1 GB.
- For AES and SM4 encryption algorithms, the traffic threshold that triggers SSH algorithm renegotiation and key re-exchange is 64 GB.

This feature does not affect the existing SSH connections.

Setting the minimum interval for updating the RSA server key pair

1. Enter system view.

   system-view

2. Set the minimum interval for updating the RSA server key pair.

   ssh server rekey-interval *interval*

   By default, the device does not update the RSA server key pair.

   This command is not available in FIPS mode.

   This configuration takes effect only on SSH1 clients.

Setting the SSH user authentication timeout timer

1. Enter system view.

   system-view

2. Set the SSH user authentication timeout timer.

   ssh server authentication-timeout *time-out-value*

   The default setting is 60 seconds.

   Perform this task to prevent malicious occupation of TCP connections. If a user does not finish the authentication when the timeout timer expires, the connection cannot be established.

Setting the maximum number of SSH authentication attempts

1. Enter system view.

   system-view

2. Set the maximum number of SSH authentication attempts.

   ssh server authentication-retries *retries*

   The default setting is 3.

   Perform this task to prevent malicious hacking of usernames and passwords. If the authentication method is **any**, the total number of publickey authentication attempts and password authentication attempts cannot exceed the upper limit.

Specifying an SSH login control ACL

1. Enter system view.

   system-view

2. Specify an SSH login control ACL.

   IPv4:

   ssh server acl { *advanced-acl-number* | *basic-acl-number* | **mac** *mac-acl-number* }

   IPv6:

   ssh server ipv6 acl { **ipv6** { *advanced-acl-number* | *basic-acl-number* } | **mac** *mac-acl-number* }

   This feature uses an ACL to filter SSH clients that initiate SSH connections to the server. By default, no ACLs are specified and all SSH users can initiate SSH connections to the server.

Only SSH users that match the ACL's permit rules can initiate SSH connections to the server. If the specified ACL does not exist or the ACL contains no rules, no SSH users can initiate SSH connections to the server.

This configuration takes effect only on new SSH connections. It does not affect the existing SSH connections.

Enabling logging for SSH login attempts that are denied by the SSH login control ACL

1. Enter system view.

   system-view

2. Enable logging for SSH login attempts that are denied by the SSH login control ACL.

   ssh server acl-deny-log enable

   By default, logging is disabled for login attempts that are denied by the SSH login control ACL.

   This command enables SSH to generate log messages for SSH login attempts that are denied by the SSH login control ACL and send the messages to the information center.

Setting the DSCP value in the packets that the SSH server sends to SSH clients

1. Enter system view.

   system-view

2. Set the DSCP value in the packets that the SSH server sends to the SSH clients.

   IPv4:

   ssh server dscp *dscp-value*

   IPv6:

   ssh server ipv6 dscp *dscp-value*

   By default, the DSCP value of SSH packets is 48.

   The DSCP value is carried in the ToS field of an IPv4 packet and in the Traffic Class field of an IPv6 packet. It defines the priority of a packet and affects the transmission priority of the packet. A bigger DSCP value represents a higher priority.

Setting the SFTP connection idle timeout timer

1. Enter system view.

   system-view

2. Set the SFTP connection idle timeout timer.

   sftp server idle-timeout *time-out-value*

   By default, the SFTP connection idle timeout is 10 minutes.

   When the SFTP connection idle timeout timer expires, the system automatically tears the connection down and releases the connection resources.

Setting the maximum number of online SSH users

1. Enter system view.

   system-view

2. Set the maximum number of online SSH users.

   aaa session-limit ssh *max-sessions*

   The default setting is 32.

   When the number of online SSH users reaches the upper limit, the system denies new SSH connection requests. Changing the upper limit does not affect online SSH users.

   For more information about this command, see AAA commands in *Security Command Reference*.

Setting the maximum number of records for SSH user login exceptions

1. Enter system view.

   system-view

2. Set the maximum number of records for SSH user login exceptions.

   ssh exception-record max-number *max-number*

   By default, the maximum number of records for SSH user login exceptions is 10.

   To view more records for SSH user login exceptions, you can use this command to set the maximum number of records.

Setting alarm and recovery thresholds for SSH user login failures in the specified statistics period

1. Enter system view.

   system-view

2. Set alarm and recovery thresholds for SSH user login failures in the specified statistics period.

   ssh server login-failed threshold-alarm upper-limit *report-time* **lower-limit** *resume-time* **period** *period-time*

   By default, the statistics period is 5 minutes, and the SSH user login failure alarm threshold and alarm recovery threshold are 30 and 20, respectively.

   Configure this feature to monitor the SSH user login situation. With this feature configured, the device generates alarms as follows:

   - Generates an alarm when the number of login failures in the specified statistics period reaches or exceeds the alarm threshold.
   - Generates a recovery alarm when the number of login failures in the specified statistics period drops below the alarm recovery threshold from the alarm threshold or a higher value.

Specifying a PKI domain for the SSH server

About this task

The PKI domain specified for the SSH server has the following functions:

- The SSH server uses the PKI domain to send its certificate to the client in the key exchange stage.
- The SSH server uses the PKI domain to authenticate the client's certificate if no PKI domain is specified for the client authentication by using the **ssh user** command.

Procedure

1. Enter system view.

   system-view

2. Specify a PKI domain for the SSH server.

   ssh server pki-domain *domain-name*

   By default, no PKI domain is specified for the SSH server.

Disconnecting SSH sessions

About this task

The device supports concurrent login sessions. To avoid an SSH login user interfering with your configuration, you can disconnect that SSH login user.

Procedure

Execute the following command in user view to disconnect SSH sessions:

free ssh { **user-ip** { *ip-address* | **ipv6** *ipv6-address* } [**port** *port-number*] | **user-pid** *pid-number* | **username** *username* }

Configuring the device as an Stelnet client

Stelnet client tasks at a glance

To configure an Stelnet client, perform the following tasks:

1. Generating local key pairs

   Only required for authentication method **publickey**, **password-publickey**, or **any**.
2. (Optional.) Specifying the source IP address for outgoing SSH packets
3. Establishing a connection to an Stelnet server
4. (Optional.) Deleting server public keys saved in the public key file on the Stelnet client
5. (Optional.) Establishing a connection to an Stelnet server based on Suite B

Generating local key pairs

About this task

You must generate local key pairs on Stelnet clients when the Stelnet server uses the **publickey**, **password-publickey**, or **any** authentication method.

Restrictions and guidelines

Local DSA, ECDSA, and RSA key pairs for SSH use default names. You cannot assign names to the key pairs. For more information about the command for generating key pairs, see public key management commands in *Security Command Reference*.

The key modulus length must be less than 2048 bits when you generate a DSA key pair.

When you generate an ECDSA key pair, you can generate only a **secp256r1** or **secp384r1** ECDSA key pair.

The Stelnet client operating in FIPS mode supports only ECDSA and RSA key pairs.

Procedure

1. Enter system view.

   system-view
2. Generate local key pairs.

   public-key local create { **dsa** | **ecdsa** { **secp256r1** | **secp384r1** } | **rsa** }

Specifying the source IP address for outgoing SSH packets

About this task

After you specify the source IP address for outgoing SSH packets on an Stelnet client, the client uses the specified IP address to communicate with the Stelnet server.

Restrictions and guidelines

As a best practice, specify the IP address of a loopback interface as the source address of outgoing SSH packets for the following purposes:

- Ensuring the communication between the Stelnet client and the Stelnet server.
- Improving the manageability of Stelnet clients in authentication service.

Procedure

1. Enter system view.

   system-view

2. Specify the source address for outgoing SSH packets.

   IPv4:

   ssh client source { **interface** *interface-type interface-number* | **ip** *ip-address* }

   By default, an IPv4 Stelnet client uses the primary IPv4 address of the output interface in the matching route as the source address of the outgoing SSH packets.

   IPv6:

   ssh client ipv6 source { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }

   By default, an IPv6 Stelnet client automatically selects a source IPv6 address for outgoing SSH packets in compliance with RFC 3484.

Establishing a connection to an Stelnet server

About this task

Perform this task to enable the Stelnet client feature on the device and establish a connection to the Stelnet server. You can specify the public key algorithm and the preferred encryption, HMAC, and key exchange algorithms to be used during the connection.

To access the server, a client must use the server's host public key to authenticate the server. As a best practice, configure the server's host public key on the device in an insecure network. If the server's host public key is not configured on the client, the client will notify you to confirm whether to continue with the access.

- If you choose to continue, the client accesses the server and downloads the server's host public key. The downloaded public key will be used to authenticate the server in subsequent accesses.

  If the server public key is not specified the when you connect to the server, the device saves the server public key to the public key file. It does not save the server public key to the configuration file.
- If you choose to not continue, the connection cannot be established.

Restrictions and guidelines for establishing a connection to an Stelnet server

An Stelnet client cannot establish connections to both IPv4 and IPv6 Stelnet servers.

Establishing a connection to an IPv4 Stelnet server

Execute the following command in user view to establish a connection with an IPv4 Stelnet server:

In non-FIPS mode:

ssh2 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**identity-key** { **dsa** | **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** } | **prefer-kex** { **dh-group-exchange-sha1** | **dh-group1-sha1** | **dh-group14-sha1** | **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **3des-cbc** | **aes128-cbc** |

aes128-ctr | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** }] * [**dscp** *dscp-value* | **escape** *character* | { **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] *

In FIPS mode:

ssh2 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**identity-key** { **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-ctos-hmac** { **sha2-256** | **sha2-512** } | **prefer-kex** { **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-stoc-hmac** { **sha2-256** | **sha2-512** }] * [**dscp** *dscp-value* | **escape** *character* | { **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] *

Establishing a connection to an IPv6 Stelnet server

Execute the following command in user view to establish a connection to an IPv6 Stelnet server:

In non-FIPS mode:

ssh2 ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] [**identity-key** { **dsa** | **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** } | **prefer-kex** { **dh-group-exchange-sha1** | **dh-group1-sha1** | **dh-group14-sha1** | **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** }] * [**dscp** *dscp-value* | **escape** *character* | { **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] *

In FIPS mode:

ssh2 ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] [**identity-key** { **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-ctos-hmac** { **sha2-256** | **sha2-512** } | **prefer-kex** { **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-stoc-hmac** { **sha2-256** | **sha2-512** }] * [**dscp** *dscp-value* | **escape** *character* | { **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] *

Deleting server public keys saved in the public key file on the Stelnet client

About this task

When the Stelnet client switches to FIPS mode but the locally saved server public key does not comply with FIPS, the client cannot connect to the server. To connect to the server, delete the server public key saved on the client and make sure a FIPS-compliant public key has been generated on the server.

Procedure

1. Enter system view.
   system-view
2. Delete server public keys saved in the public key file on the Stelnet client.
   delete ssh client server-public-key [**server-ip** *ip-address*]

Establishing a connection to an Stelnet server based on Suite B

Execute the following command in user view to establish a connection to an Stelnet server based on Suite B:

IPv4:

ssh2 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] **suite-b** [**128-bit** | **192-bit**] **pki-domain** *domain-name* [**server-pki-domain** *domain-name*] [**prefer-compress zlib**] [**dscp** *dscp-value* | **escape** *character* | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] *

IPv6:

ssh2 ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] **suite-b** [**128-bit** | **192-bit**] **pki-domain** *domain-name* [**server-pki-domain** *domain-name*] [**prefer-compress zlib**] [**dscp** *dscp-value* | **escape** *character* | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] *

Configuring the device as an SFTP client

SFTP client tasks at a glance

To configure an SFTP client, perform the following tasks:

1. Generating local key pairs
   Only required for authentication method **publickey**, **password-publickey**, or **any**.
2. (Optional.) Specifying the source IP address for outgoing SFTP packets
3. Establishing a connection to an SFTP server
4. (Optional.) Deleting server public keys saved in the public key file on the SFTP client
5. (Optional.) Establishing a connection to an SFTP server based on Suite B
6. (Optional.) Working with SFTP directories
7. (Optional.) Working with SFTP files

8. (Optional.) Displaying help information
9. (Optional.) Terminating the connection with the SFTP server

Generating local key pairs

About this task

You must generate local key pairs on SFTP clients when the SFTP server uses the **publickey**, **password-publickey**, or **any** authentication method.

Restrictions and guidelines

Local DSA, ECDSA, and RSA key pairs for SSH use default names. You cannot assign names to the key pairs. For more information about the command for generating key pairs, see public key management commands in *Security Command Reference*.

The key modulus length must be less than 2048 bits when you generate a DSA key pair.

When you generate an ECDSA key pair, you can generate only a **secp256r1** or **secp384r1** ECDSA key pair.

The SFTP client operating in FIPS mode supports only ECDSA and RSA key pairs.

Procedure

1. Enter system view.

   system-view
2. Generate local key pairs.

   public-key local create { **dsa** | **ecdsa** { **secp256r1** | **secp384r1** } | **rsa** }

Specifying the source IP address for outgoing SFTP packets

About this task

After you specify the source IP address for outgoing SFTP packets on an SFTP client, the client uses the specified IP address to communicate with the SFTP server.

Restrictions and guidelines

As a best practice, specify the IP address of a loopback interface as the source address of outgoing SFTP packets for the following purposes:

- Ensuring the communication between the SFTP client and the SFTP server.
- Improving the manageability of SFTP clients in authentication service.

Procedure

1. Enter system view.

   system-view
2. Specify the source address for outgoing SFTP packets.

   IPv4:

   sftp client source { **ip** *ip-address* | **interface** *interface-type interface-number* }

   By default, an SFTP client uses the primary IPv4 address of the output interface in the matching route as the source address of the outgoing SFTP packets.

   IPv6:

   sftp client ipv6 source { **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }

By default, an IPv6 SFTP client automatically selects a source IPv6 address for the outgoing SFTP packets in compliance with RFC 3484.

Establishing a connection to an SFTP server

About this task

Perform this task to enable the SFTP client feature on the device and establish a connection to the SFTP server. You can specify the public key algorithm and the preferred encryption, HMAC, and key exchange algorithms to be used during the connection.

To access the server, a client must use the server's host public key to authenticate the server. As a best practice, configure the server's host public key on the device in an insecure network. If the server's host public key is not configured on the client, the client will notify you to confirm whether to continue with the access.

- If you choose to continue, the client accesses the server and downloads the server's host public key. The downloaded public key will be used to authenticate the server in subsequent accesses.

  If the server public key is not specified when you connect to the server, the device saves the server public key to the public key file. It does not save the server public key to the configuration file.
- If you choose to not continue, the connection cannot be established.

Restrictions and guidelines for establishing a connection to an SFTP server

An SFTP client cannot establish connections to both IPv4 and IPv6 SFTP servers.

Establishing a connection to an IPv4 SFTP server

Execute the following command in user view to establish a connection to an IPv4 SFTP server:

In non-FIPS mode:

sftp *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**identity-key** { **dsa** | **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** } | **prefer-kex** { **dh-group-exchange-sha1** | **dh-group1-sha1** | **dh-group14-sha1** | **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** }] * [**dscp** *dscp-value* | { **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] *

In FIPS mode:

sftp *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**identity-key** { **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-ctos-hmac** { **sha2-256** | **sha2-512** } | **prefer-kex** { **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-stoc-hmac** { **sha2-256** | **sha2-512** }] * [**dscp** *dscp-value* | { **public-key** *keyname* |

server-pki-domain *domain-name* } | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] *

Establishing a connection to an IPv6 SFTP server

Execute the following command in user view to establish a connection to an IPv6 SFTP server:

In non-FIPS mode:

sftp ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] [**identity-key** { **dsa** | **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** } | **prefer-kex** { **dh-group-exchange-sha1** | **dh-group1-sha1** | **dh-group14-sha1** | **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** }] * [**dscp** *dscp-value* | { **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] *

In FIPS mode:

sftp ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] [**identity-key** { **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-ctos-hmac** { **sha2-256** | **sha2-512** } | **prefer-kex** { **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-stoc-hmac** { **sha2-256** | **sha2-512** }] * [**dscp** *dscp-value* | { **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] *

Deleting server public keys saved in the public key file on the SFTP client

About this task

When the SFTP client switches to FIPS mode but the locally saved server public key does not comply with FIPS, the client cannot connect to the server. To connect to the server, delete the server public key saved on the client and make sure a FIPS-compliant public key has been generated on the server.

Procedure

1. Enter system view.

   system-view

2. Delete server public keys saved in the public key file on the SFTP client.

   delete ssh client server-public-key [**server-ip** *ip-address*]

Establishing a connection to an SFTP server based on Suite B

Execute the following command in user view to establish a connection to an SFTP server based on Suite B:

IPv4:

sftp *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] **suite-b** [**128-bit** | **192-bit**] **pki-domain** *domain-name* [**server-pki-domain** *domain-name*] [**prefer-compress** **zlib**] [**dscp** *dscp-value* | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] *

IPv6:

sftp ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] **suite-b** [**128-bit** | **192-bit**] **pki-domain** *domain-name* [**server-pki-domain** *domain-name*] [**prefer-compress** **zlib**] [**dscp** *dscp-value* | **escape** *character* | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] *

Working with SFTP directories

About this task

After you establish a connection to an SFTP server, you can operate directories of the SFTP server.

Changing the working directory on the SFTP server

1. Enter SFTP client view.

   For more information, see "Establishing a connection to an SFTP server."
2. Change the working directory on the SFTP server.

   cd [*remote-path*]
3. (Optional.) Return to the upper-level directory.

   cdup

Displaying the current working directory on the SFTP server

1. Enter SFTP client view.

   For more information, see "Establishing a connection to an SFTP server."
2. Display the current working directory on the SFTP server.

   pwd

Displaying files under a directory

1. Enter SFTP client view.

   For more information, see "Establishing a connection to an SFTP server."
2. Display files under a directory.
   - **dir** [**-a** | **-l**] [*remote-path*]
   - **ls** [**-a** | **-l**] [*remote-path*]

   The **dir** command has the same function as the **ls** command.

Changing the name of a directory on the SFTP server

1. Enter SFTP client view.

For more information, see "Establishing a connection to an SFTP server."

2. Change the name of a directory on the SFTP server.
   rename *oldname newname*

Creating a new directory on the SFTP server

1. Enter SFTP client view.
   For more information, see "Establishing a connection to an SFTP server."
2. Create a new directory on the SFTP server.
   mkdir *remote-path*

Deleting directories on the SFTP server

1. Enter SFTP client view.
   For more information, see "Establishing a connection to an SFTP server."
2. Delete one or more directories from the SFTP server.
   rmdir *remote-path*

Working with SFTP files

About this task

After you establish a connection to an SFTP server, you can operate files on the SFTP server.

Changing the name of a file on the SFTP server

1. Enter SFTP client view.
   For more information, see "Establishing a connection to an SFTP server."
2. Change the name of a file on the SFTP server.
   rename *old-name new-name*

Downloading a file from the SFTP server and save it locally

1. Enter SFTP client view.
   For more information, see "Establishing a connection to an SFTP server."
2. Download a file from the SFTP server and save it locally.
   get *remote-file* [*local-file*]

Uploading a local file to the SFTP server

1. Enter SFTP client view.
   For more information, see "Establishing a connection to an SFTP server."
2. Upload a local file to the SFTP server.
   put *local-file* [*remote-file*]

Display files under a directory

1. Enter SFTP client view.
   For more information, see "Establishing a connection to an SFTP server."
2. Display files under a directory.
   - **dir** [**-a** | **-l**] [*remote-path*]
   - **ls** [**-a** | **-l**] [*remote-path*]

   The **dir** command has the same function as the **ls** command.

Deleting a file from the SFTP server

1. Enter SFTP client view.

   For more information, see "Establishing a connection to an SFTP server."
2. Delete a file from the SFTP server.
   - **`delete`** *`remote-file`*
   - **`remove`** *`remote-file`*

   The **`delete`** command has the same function as the **`remove`** command.

Displaying help information

About this task

After you establish a connection to the SFTP server, you can display the help information of SFTP client commands, including the command syntax and parameter configuration.

Procedure

1. Enter SFTP client view.

   For more information, see "Establishing a connection to an SFTP server."
2. Display SFTP client command help information.
   - **`help`**
   - **`?`**

   The **`help`** command has the same function as the **`?`** command.

Terminating the connection with the SFTP server

1. Enter SFTP client view.

   For more information, see "Establishing a connection to an SFTP server."
2. Terminate the connection with the SFTP server and return to user view.
   - **`bye`**
   - **`exit`**
   - **`quit`**

   The three commands have the same function.

Configuring the device as an SCP client

SCP client tasks at a glance

To configure an SCP client, perform the following tasks:

1. Generating local key pairs

   Only required for the **publickey**, **password-publickey**, or **any** authentication method.
2. (Optional.) Specifying the source IP address for outgoing SCP packets
3. Establishing a connection to an SCP server
4. (Optional.) Deleting server public keys saved in the public key file on the SCP client
5. (Optional.) Establishing a connection to an SCP server based on Suite B

Generating local key pairs

About this task

You must generate local key pairs on SCP clients when the SCP server uses the **publickey**, **password-publickey**, or **any** authentication method.

Restrictions and guidelines

Local DSA, ECDSA, and RSA key pairs for SSH use default names. You cannot assign names to the key pairs. For more information about the command for generating key pairs, see public key management commands in *Security Command Reference*.

The key modulus length must be less than 2048 bits when you generate a DSA key pair.

When you generate an ECDSA key pair, you can generate only a **secp256r1** or **secp384r1** ECDSA key pair.

The SCP client operating in FIPS mode supports only ECDSA and RSA key pairs.

Procedure

1. Enter system view.

   system-view

2. Generate local key pairs.

   public-key local create { **dsa** | **ecdsa** { **secp256r1** | **secp384r1** } | **rsa** }

Specifying the source IP address for outgoing SCP packets

About this task

After you specify the source IP address for outgoing SCP packets on an SCP client, the client uses the specified IP address to communicate with the SCP server.

Restrictions and guidelines

As a best practice, specify the IP address of a loopback interface as the source address of outgoing SCP packets for the following purposes:

- Ensuring the communication between the SCP client and the SCP server.
- Improving the manageability of SCP clients in authentication service.

Procedure

1. Enter system view.

   system-view

2. Specify the source address for outgoing SCP packets.

   IPv4:

   scp client source { **interface** *interface-type interface-number* | **ip** *ip-address* }

   By default, an SCP client uses the primary IPv4 address of the output interface in the matching route as the source address of the outgoing SCP packets.

   IPv6:

   scp client ipv6 source { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }

   By default, an SCP client automatically selects an IPv6 address as the source address of the outgoing packets in compliance with RFC 3484.

Establishing a connection to an SCP server

About this task

Perform this task to enable the SCP client feature on the device, establish a connection to the SCP server, and transfer files with the server. You can specify the public key algorithm and the preferred encryption, HMAC, and key exchange algorithms to be used during the connection.

To access the server, a client must use the server's host public key to authenticate the server. As a best practice, configure the server's host public key on the device in an insecure network. If the server's host public key is not configured on the client, the client will notify you to confirm whether to continue with the access.

- If you choose to continue, the client accesses the server and downloads the server's host public key. The downloaded public key will be used to authenticate the server in subsequent accesses.

  If the server public key is not specified when you connect to the server, the device saves the server public key to the public key file. It does not save the server public key to the configuration file.
- If you choose to not continue, the connection cannot be established.

Restrictions and guidelines for establishing a connection to an SCP server

An SCP client cannot establish connections to both IPv4 and IPv6 SCP servers.

Establishing a connection to an IPv4 SCP server

Execute the following command in user view to connect to an IPv4 SCP server, and transfer files with the server:

In non-FIPS mode:

scp *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] { **put** | **get** } *source-file-name* [*destination-file-name*] [**identity-key** { **dsa** | **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** } | **prefer-kex** { **dh-group-exchange-sha1** | **dh-group1-sha1** | **dh-group14-sha1** | **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** }] * [{ **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] * [**user** *username* [**password** *password* [**no-more-input**]]]

In FIPS mode:

scp *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] { **put** | **get** } *source-file-name* [*destination-file-name*] [**identity-key** { **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress zlib** | **prefer-ctos-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-ctos-hmac** { **sha2-256** | **sha2-512** } | **prefer-kex** { **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-stoc-hmac** { **sha2-256** | **sha2-512** }] * [{ **public-key** *keyname* | **server-pki-domain**

domain-name } | **source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] * [**user** *username* [**password** *password* [**no-more-input**]]]

Establishing a connection to an IPv6 SCP server

Execute the following command in user view to connect to an IPv6 SCP server, and transfer files with the server.

In non-FIPS mode:

scp ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] { **put** | **get** } *source-file-name* [*destination-file-name*] [**identity-key** { **dsa** | **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress** **zlib** | **prefer-ctos-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** } | **prefer-kex** { **dh-group-exchange-sha1** | **dh-group1-sha1** | **dh-group14-sha1** | **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **des-cbc** | **sm4-cbc** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** }] * [{ **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] * [**user** *username* [**password** *password* [**no-more-input**]]]

In FIPS mode:

scp ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] { **put** | **get** } *source-file-name* [*destination-file-name*] [**identity-key** { **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | { **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } **pki-domain** *domain-name* } | **prefer-compress** **zlib** | **prefer-ctos-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-ctos-hmac** { **sha2-256** | **sha2-512** } | **prefer-kex** { **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } | **prefer-stoc-cipher** { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } | **prefer-stoc-hmac** { **sha2-256** | **sha2-512** }] * [{ **public-key** *keyname* | **server-pki-domain** *domain-name* } | **source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] * [**user** *username* [**password** *password* [**no-more-input**]]]

Deleting server public keys saved in the public key file on the SCP client

About this task

When the SCP client switches to FIPS mode but the locally saved server public key does not comply with FIPS, the client cannot connect to the server. To connect to the server, delete the server public key saved on the client and make sure a FIPS-compliant public key has been generated on the server.

Procedure

1. Enter system view.

   system-view

2. Delete server public keys saved in the public key file on the SCP client.

   delete ssh client server-public-key [**server-ip** *ip-address*]

Establishing a connection to an SCP server based on Suite B

Execute the following command in user view to establish a connection to an SCP server based on Suite B:

IPv4:

scp *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] { **put** | **get** } *source-file-name* [*destination-file-name*] **suite-b** [**128-bit** | **192-bit**] **pki-domain** *domain-name* [**server-pki-domain** *domain-name*] [**prefer-compress zlib**] [**source** { **interface** *interface-type interface-number* | **ip** *ip-address* }] * [**user** *username* [**password** *password* [**no-more-input**]]]

IPv6:

scp ipv6 *server* [*port-number*] [**vpn-instance** *vpn-instance-name*] [**-i** *interface-type interface-number*] { **put** | **get** } *source-file-name* [*destination-file-name*] **suite-b** [**128-bit** | **192-bit**] **pki-domain** *domain-name* [**server-pki-domain** *domain-name*] [**prefer-compress zlib**] [**source** { **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }] * [**user** *username* [**password** *password* [**no-more-input**]]]

Specifying algorithms for SSH2

About algorithms for SSH2

The SSH2 client and server use the following types of algorithms for algorithm negotiation during the Stelnet, SFTP, or SCP session establishment:

- Key exchange algorithms.
- Public key algorithms.
- Encryption algorithms.
- MAC algorithms.

If you specify algorithms, SSH2 uses only the specified algorithms for algorithm negotiation. The client uses the specified algorithms to initiate the negotiation, and the server uses the matching algorithms to negotiate with the client. If multiple algorithms of the same type are specified, the algorithm specified earlier has a higher priority during negotiation.

Specifying key exchange algorithms for SSH2

1. Enter system view.

   system-view

2. Specify key exchange algorithms for SSH2.

   In non-FIPS mode:

   ssh2 algorithm key-exchange { **dh-group-exchange-sha1** | **dh-group1-sha1** | **dh-group14-sha1** | **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } *

   By default, SSH2 uses the **ecdh-sha2-nistp256**, **ecdh-sha2-nistp384**, **dh-group-exchange-sha1**, **dh-group14-sha1**, and **dh-group1-sha1** key exchange algorithms in descending order of priority for algorithm negotiation.

   In FIPS mode:

ssh2 algorithm key-exchange { **ecdh-sha2-nistp256** | **ecdh-sha2-nistp384** } *

By default, SSH2 uses the **ecdh-sha2-nistp256** and **ecdh-sha2-nistp384** key exchange algorithms in descending order of priority for algorithm negotiation.

Specifying public key algorithms for SSH2

1. Enter system view.

   system-view

2. Specify public key algorithms for SSH2.

   In non-FIPS mode:

   ssh2 algorithm public-key { **dsa** | **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } *

   By default, SSH2 uses the **x509v3-ecdsa-sha2-nistp256**, **x509v3-ecdsa-sha2-nistp384**, **ecdsa-sha2-nistp256**, **ecdsa-sha2-nistp384**, **rsa**, and **dsa** public key algorithms in descending order of priority for algorithm negotiation.

   In FIPS mode:

   ssh2 algorithm public-key { **ecdsa-sha2-nistp256** | **ecdsa-sha2-nistp384** | **rsa** | **x509v3-ecdsa-sha2-nistp256** | **x509v3-ecdsa-sha2-nistp384** } *

   By default, SSH2 uses the **x509v3-ecdsa-sha2-nistp256**, **x509v3-ecdsa-sha2-nistp384**, **ecdsa-sha2-nistp256**, **ecdsa-sha2-nistp384**, and **rsa** public key algorithms in descending order of priority for algorithm negotiation.

Specifying encryption algorithms for SSH2

1. Enter system view.

   system-view

2. Specify encryption algorithms for SSH2.

   In non-FIPS mode:

   ssh2 algorithm cipher { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes128-gcm@openssh** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **aes256-gcm@openssh** | **des-cbc** | **sm4-cbc** } *

   By default, SSH2 uses the **aes128-ctr**, **aes192-ctr**, **aes256-ctr**, **aes128-gcm@openssh**, **aes256-gcm@openssh**, **aes128-gcm**, **aes256-gcm**, **aes128-cbc**, **3des-cbc**, **aes256-cbc**, **des-cbc**, and **sm4-cbc** encryption algorithms in descending order of priority for algorithm negotiation.

   In FIPS mode:

   ssh2 algorithm cipher { **aes128-gcm** | **aes128-gcm@openssh** | **aes256-gcm** | **aes256-gcm@openssh** } *

   By default, SSH2 uses the **aes128-gcm@openssh**, **aes256-gcm@openssh**, **aes128-gcm**, and **aes256-gcm** encryption algorithms in descending order of priority for algorithm negotiation.

Specifying MAC algorithms for SSH2

1. Enter system view.

   system-view

2. Specify MAC algorithms for SSH2.

   In non-FIPS mode:

   ssh2 algorithm mac { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sm3** } *

   By default, SSH2 uses the **sha2-256**, **sha2-512**, **sha1**, **md5**, **sha1-96**, **md5-96**, and **sm3** MAC algorithms in descending order of priority for algorithm negotiation.

   In FIPS mode:

   ssh2 algorithm mac { **sha2-256** | **sha2-512** } *

   By default, SSH2 uses the **sha2-256** and **sha2-512** MAC algorithms in descending order of priority for algorithm negotiation.

Display and maintenance commands for SSH

Execute **display** commands in any view.

Task	Command
Display the public keys of the local key pairs.	**display public-key local** { **dsa** | **ecdsa** | **rsa** } **public** [**name** *publickey-name*]
Display information about peer public keys.	**display public-key peer** [**brief** | **name** *publickey-name*]
Display the source IP address configuration of the SCP client.	**display scp client source**
Display the source IP address configuration of the SFTP client.	**display sftp client source**
Display server public key information saved in the public key file on the SSH client.	**display ssh client server-public-key** [**server-ip** *ip-address*]
Display the source IP address configuration of the Stelnet client.	**display ssh client source**
Display SSH server status or sessions.	**display ssh server** { **session** | **status** }
Display SSH user information on the SSH server.	**display ssh user-information** [*username*]
Display algorithms used by SSH2 in the algorithm negotiation stage.	**display ssh2 algorithm**
Display records for SSH user login exceptions.	**display ssh exception-record**

For more information about the **display public-key local** and **display public-key peer** commands, see public key management commands in *Security Command Reference*.

Stelnet configuration examples

Unless otherwise noted, devices in the configuration examples operate in non-FIPS mode.

When the device acts as an Stelnet server operating in FIPS mode, only ECDSA and RSA key pairs are supported. Do not generate a DSA key pair on the Stelnet server.

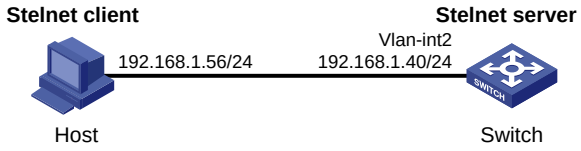
Example: Configuring the device as an Stelnet server (password authentication)

Network configuration

As shown in Figure 1:

- The switch acts as the Stelnet server and uses password authentication to authenticate the Stelnet client. The username and password of the client are saved on the switch.
- The host acts as the Stelnet client, using Stelnet client software (SSH2). After the user on the host logs in to the switch through Stelnet, the user can configure and manage the switch as a network administrator.

Figure 1 Network diagram



Procedure

1. Configure the Stelnet server:

   # Generate RSA key pairs.

```
<Switch> system-view
[Switch] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

   # Generate a DSA key pair.

```
[Switch] public-key local create dsa
The range of public key modulus is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.....................................+
...+.................+..........+...+.
Create the key pair successfully.
```

   # Generate an ECDSA key pair.

```
[Switch] public-key local create ecdsa secp256r1
Generating Keys...
.
```

```
Create the key pair successfully.
```

Enable the Stelnet server.

```
[Switch] ssh server enable
```

Assign an IP address to VLAN-interface 2. The Stelnet client uses this address as the destination for SSH connection.

```
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.1.40 255.255.255.0
[Switch-Vlan-interface2] quit
```

Set the authentication mode to AAA for user lines.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Create a local device management user named **client001**.

```
[Switch] local-user client001 class manage
```

Set a password for local user **client001**.

```
[Switch-luser-manage-client001] password simple 123456TESTplat&!
```

Authorize local user **client001** to use the **SSH** service.

```
[Switch-luser-manage-client001] service-type ssh
```

Assign the **network-admin** user role to local user **client001**.

```
[Switch-luser-manage-client001] authorization-attribute user-role network-admin
[Switch-luser-manage-client001] quit
```

Create an SSH user named **client001**. Specify the service type as **stelnet** and the authentication method as **password** for the user.

```
[Switch] ssh user client001 service-type stelnet authentication-type password
```

2. Establish a connection to the Stelnet server:

There are different types of Stelnet client software, such as PuTTY and OpenSSH. This example uses an Stelnet client that runs PuTTY version 0.58.

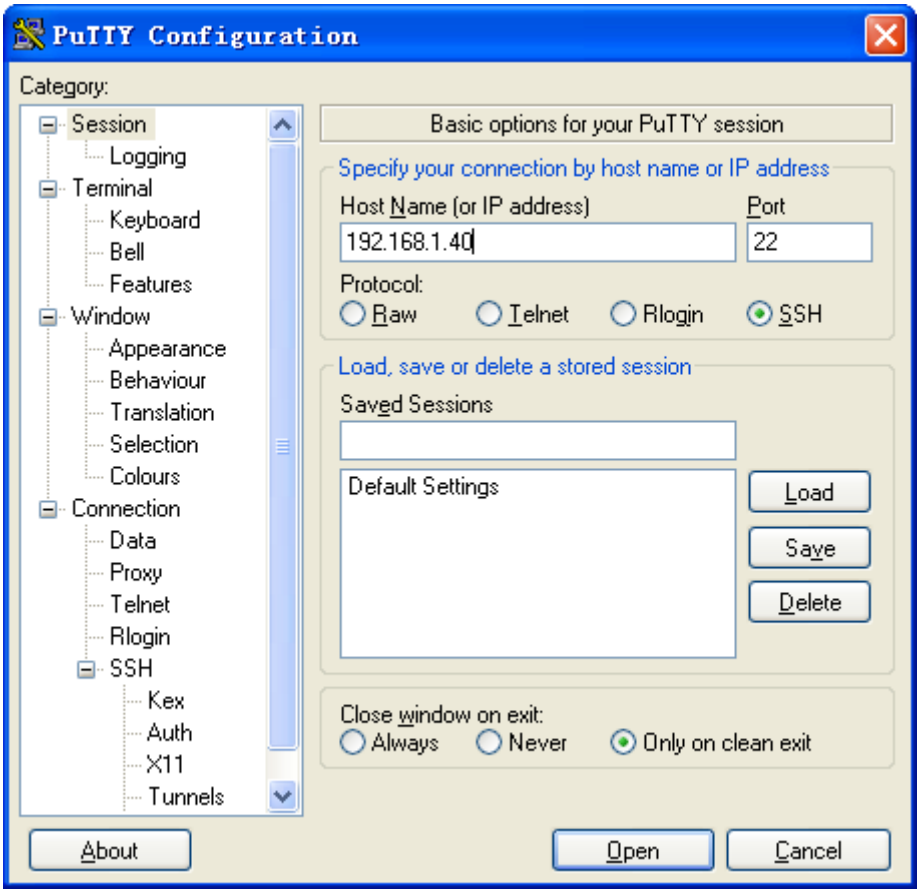
To establish a connection to the Stelnet server:

a. Launch PuTTY.exe to enter the interface shown in Figure 2.

b. In the **Host Name (or IP address)** field, enter IP address **192.168.1.40** of the Stelnet server.

c. Click **Open**.

Figure 2 Specifying the host name (or IP address)



a. Enter username **client001** and the password to log in to the Stelnet server.

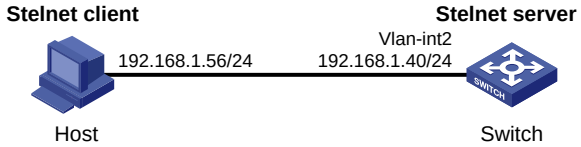
Example: Configuring the device as an Stelnet server (publickey authentication)

Network configuration

As shown in Figure 3:

- The switch acts as the Stelnet server, and it uses publickey authentication and the RSA public key algorithm.
- The host acts as the Stelnet client, using Stelnet client software (SSH2). After the user on the host logs in to the switch through Stelnet, the user can configure and manage the switch as a network administrator.

Figure 3 Network diagram



Procedure

In the server configuration, the client's host public key is required. Use the client software to generate RSA key pairs on the client before configuring the Stelnet server.

There are different types of Stelnet client software, such as PuTTY and OpenSSH. This example uses an Stelnet client that runs PuTTY version 0.58.

The configuration procedure is as follows:

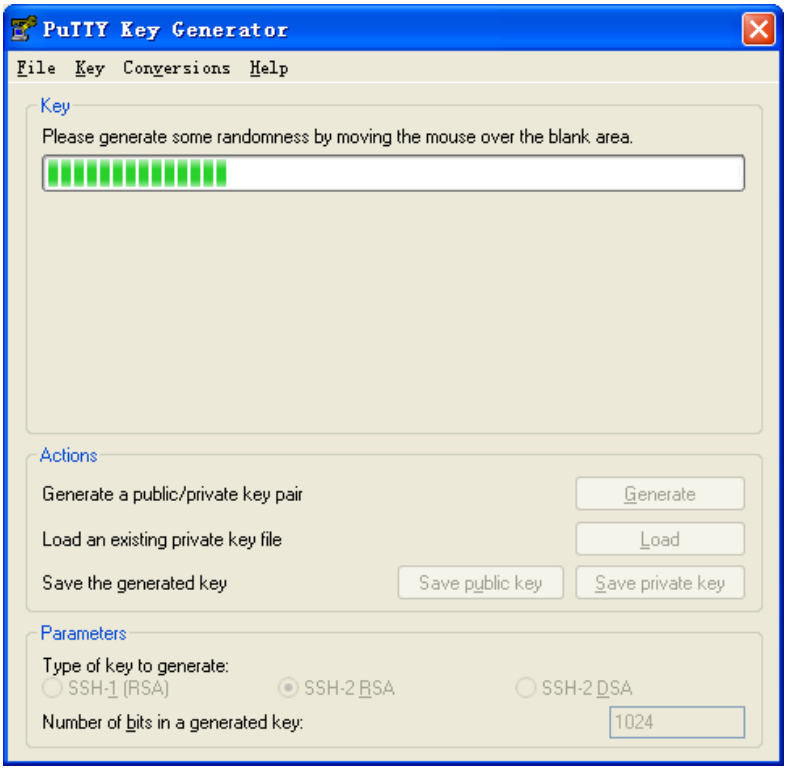
1. Generate RSA key pairs on the Stelnet client:
   a. Run PuTTYGen.exe on the client, select **SSH-2 RSA** and click **Generate**.

Figure 4 Generating a key pair on the client



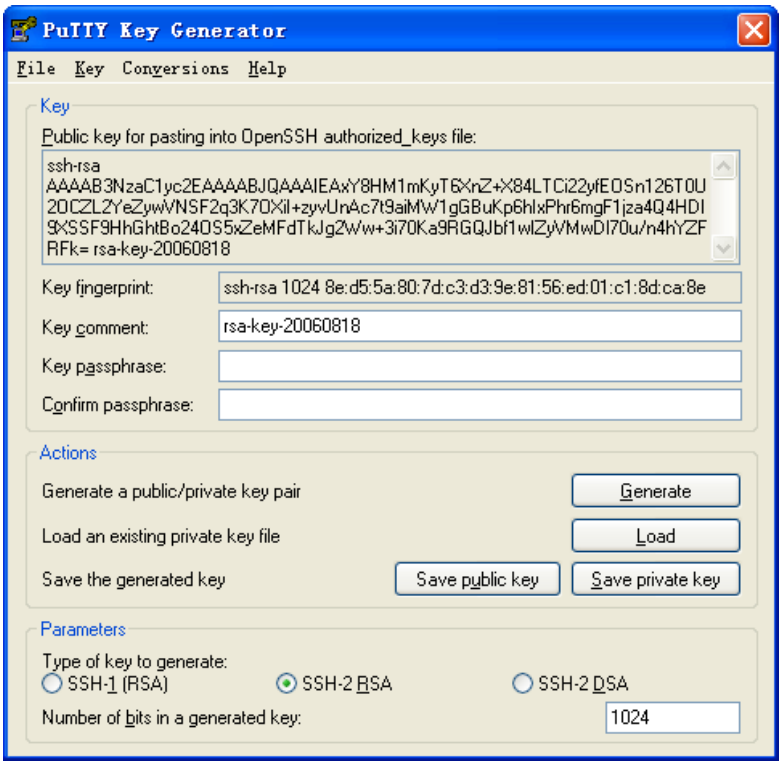
   a. Continue moving the mouse during the key generating process, but do not place the mouse over the green progress bar shown in Figure 5. Otherwise, the progress bar stops moving and the key pair generating progress stops.

Figure 5 Generating process



a. After the key pair is generated, click **Save public key** to save the public key. A file saving window appears.

Figure 6 Saving a key pair on the client



a. Enter a file name (**key.pub** in this example), and click **Save**.

b. On the page shown in Figure 6, click **Save private key** to save the private key.

A confirmation dialog box appears.

c. Click **Yes**.

A file saving window appears.

d. Enter a file name (**private.ppk** in this example), and click **Save**.

e. Transmit the public key file to the server through FTP or TFTP. (Details not shown.)

2. Configure the Stelnet server:

Generate RSA key pairs.

```
<Switch> system-view
[Switch] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

Generate a DSA key pair.

```
[Switch] public-key local create dsa
The range of public key modulus is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.................................+
...+.................+..........+...+
Create the key pair successfully.
```

Generate an ECDSA key pair.

```
[Switch] public-key local create ecdsa secp256r1
Generating Keys...
.
Create the key pair successfully.
```

Enable the Stelnet server.

```
[Switch] ssh server enable
```

Assign an IP address to VLAN-interface 2. The Stelnet client uses this IP address as the destination for SSH connection.

```
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.1.40 255.255.255.0
[Switch-Vlan-interface2] quit
```

Set the authentication mode to AAA for user lines.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
```

```
[Switch-line-vty0-63] quit
```

Import the client's public key from the public key file **key.pub** and name it **switchkey**.

```
[Switch] public-key peer switchkey import sshkey key.pub
```

Create an SSH user named **client002**. Specify the authentication method as **publickey** for the user, and assign the public key **switchkey** to the user.

```
[Switch] ssh user client002 service-type stelnet authentication-type publickey assign publickey switchkey
```

Create a local device management user named **client002**.

```
[Switch] local-user client002 class manage
```

Authorize local user **client002** to use the **SSH** service.

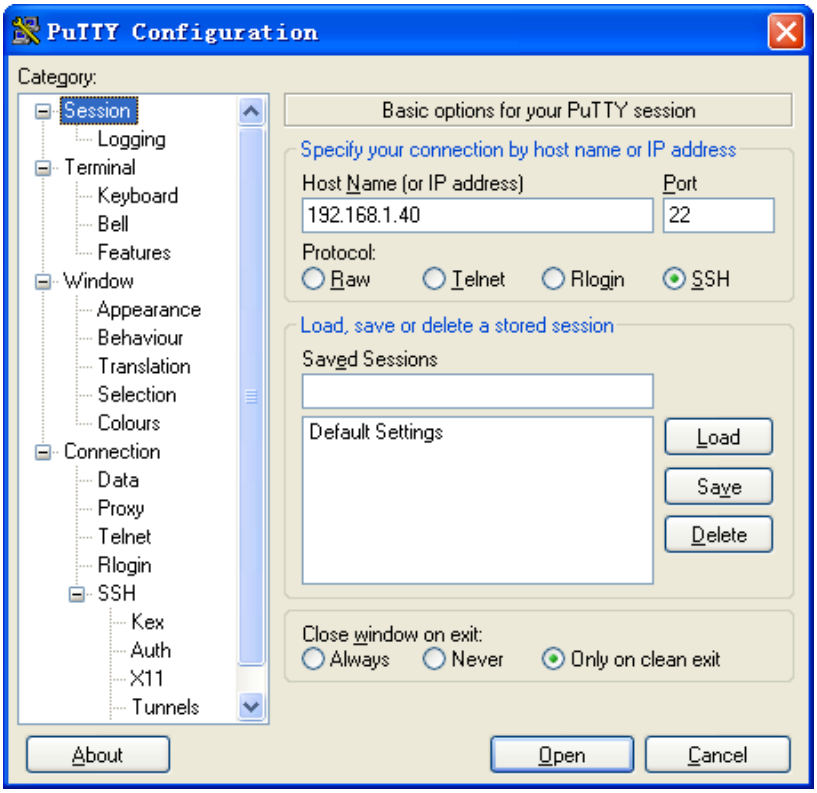
```
[Switch-luser-manage-client002] service-type ssh
```

Assign the **network-admin** user role to local user **client002**.

```
[Switch-luser-manage-client002] authorization-attribute user-role network-admin
[Switch-luser-manage-client002] quit
```

3. Specify the private key file and establish a connection to the Stelnet server:
   a. Launch PuTTY.exe on the Stelnet client to enter the interface shown in Figure 7.
   b. In the **Host Name (or IP address)** field, enter IP address **192.168.1.40** of the Stelnet server.

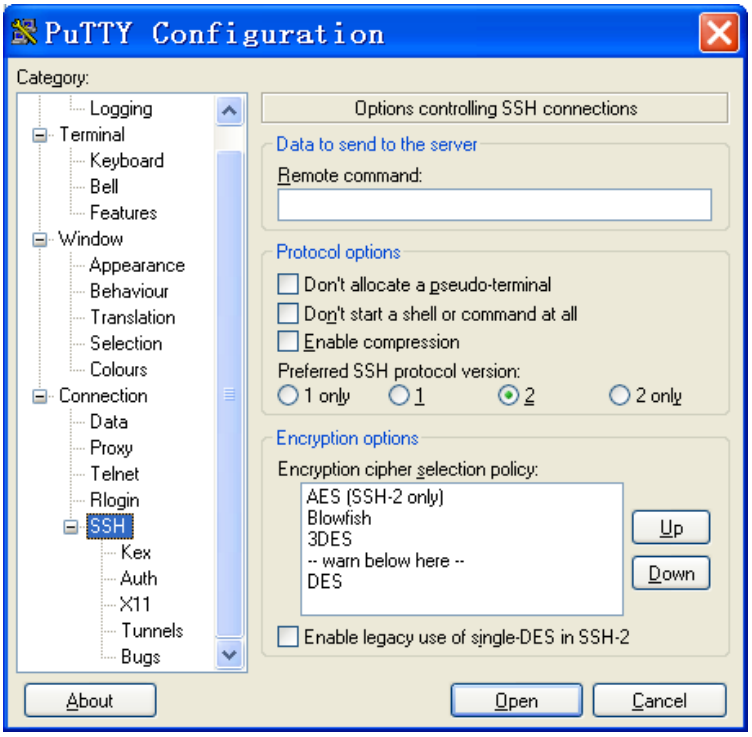
Figure 7 Specifying the host name (or IP address)



   a. From the navigation tree, select **Connection** > **SSH**.

      The window shown in Figure 8 appears.
   b. Set **Preferred SSH protocol version** to **2**.

Figure 8 Setting the preferred SSH version



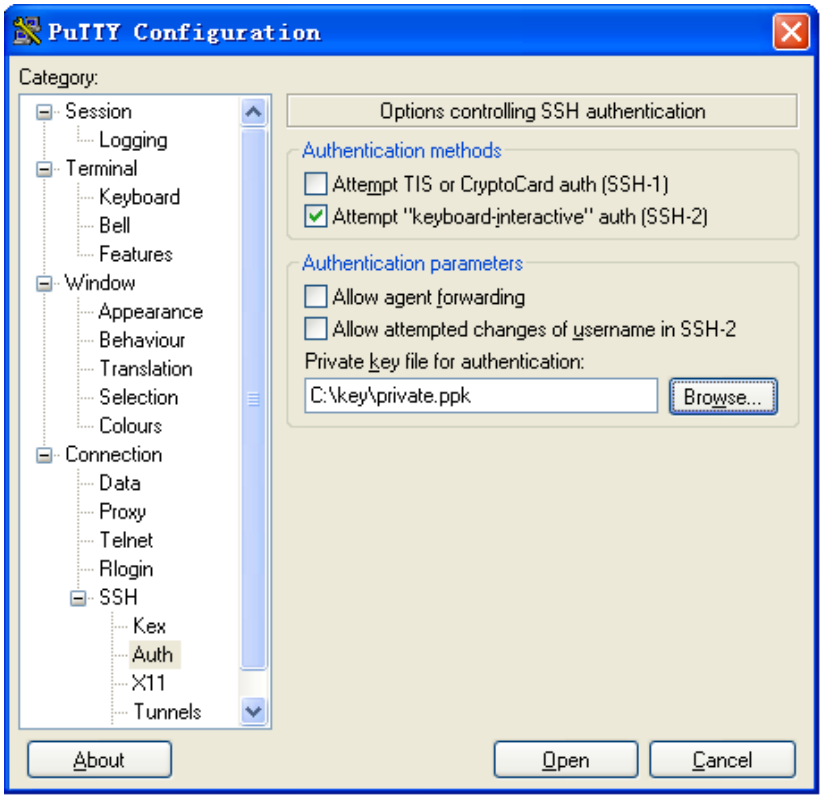
a. From the navigation tree, select **Connection** > **SSH** > **Auth**.

The window shown in Figure 9 appears.

b. Click **Browse…** to open the file selection window, and then select the private key file (**private.ppk** in this example).

c. Click **Open**.

Figure 9 Specifying the private key file



a. Entering username **client002** to log in to the Stelnet server.

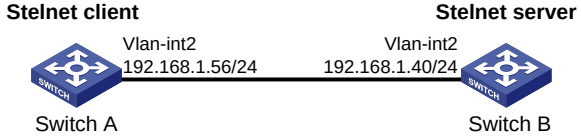
Example: Configuring the device as an Stelnet client (password authentication)

Network configuration

As shown in Figure 10:

- Switch B acts as the Stelnet server and uses password authentication to authenticate the Stelnet client. The username and password of the client are saved on Switch B.
- Switch A acts as the Stelnet client. After the user on Switch A logs in to Switch B through Stelnet, the user can configure and manage Switch B as a network administrator.

Figure 10 Network diagram



Procedure

1. Configure the Stelnet server:

Generate RSA key pairs.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
```

```
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

Generate a DSA key pair.

```
[SwitchB] public-key local create dsa
The range of public key modulus is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.................................+
...+.............+.........+...+
Create the key pair successfully.
```

Generate an ECDSA key pair.

```
[SwitchB] public-key local create ecdsa secp256r1
Generating Keys...
.
Create the key pair successfully.
```

Enable the Stelnet server.

```
[SwitchB] ssh server enable
```

Assign an IP address to VLAN-interface 2. The Stelnet client uses this address as the destination address of the SSH connection.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 192.168.1.40 255.255.255.0
[SwitchB-Vlan-interface2] quit
```

Set the authentication mode to AAA for user lines.

```
[SwitchB] line vty 0 63
[SwitchB-line-vty0-63] authentication-mode scheme
[SwitchB-line-vty0-63] quit
```

Create a local device management user named **client001**.

```
[SwitchB] local-user client001 class manage
```

Set a password for local user **client001**.

```
[SwitchB-luser-manage-client001] password simple 123456TESTplat&!
```

Authorize local user **client001** to use the **SSH** service.

```
[SwitchB-luser-manage-client001] service-type ssh
```

Assign the **network-admin** user role to local user **client001**.

```
[SwitchB-luser-manage-client001] authorization-attribute user-role network-admin
[SwitchB-luser-manage-client001] quit
```

Create an SSH user named **client001**. Specify the service type as **stelnet** and the authentication method as **password** for the user.

```
[SwitchB] ssh user client001 service-type stelnet authentication-type password
```

Specify public key algorithm **dsa** for SSH2.

```
[SwitchB] ssh2 algorithm public-key dsa
```

2. Establish a connection to the Stelnet server:

Assign an IP address to VLAN-interface 2.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.1.56 255.255.255.0
[SwitchA-Vlan-interface2] quit
[SwitchA] quit
```

- If the client does not have the server's host public key, and is connected to the server for the first time:

  # Establish an SSH connection to server 192.168.1.40. Enter username **client001**, and then enter **y** to continue accessing the server without authenticating the server, and enter **y** to download and save the server's host public key.

```
<SwitchA> ssh2 192.168.1.40
Username: client001
Press CTRL+C to abort.
Connecting to 192.168.1.40 port 22.
The server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:y
client001@192.168.1.40's password:

Enter a character ~ and a dot to abort.

******************************************************************************
* Copyright (c) 2004-2017 New H3C Technologies Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                                 *
* no decompiling or reverse-engineering shall be allowed.                    *
******************************************************************************
*

<SwitchB>
```

  After you enter the correct password, you can access Switch B successfully. At the next connection attempt, the client automatically authenticates the server by using the server's host public key that is saved on the client, and you only need to enter the password to log in to Switch B.

- If you configure the server's host public key on the client before establishing a connection to the server:

  # Use the **display public-key local dsa public** command on the server to display the server's host public key. (Details not shown.)

  # Enter public key view of the client and copy the host public key of the server to the client.

```
[SwitchA] public-key peer key1
Enter public key view. Return to system view with "peer-public-key end" command.
[SwitchA-pkey-public-key-key1]308201B73082012C06072A8648CE3804013082011F028181
0
0D757262C4584C44C211F18BD96E5F0
[SwitchA-pkey-public-key-key1]61C4F0A423F7FE6B6B85B34CEF72CE14A0D3A5222FE08CEC
E
65BE6C265854889DC1EDBD13EC8B274
```

```
[SwitchA-pkey-public-key-key1]DA9F75BA26CCB987723602787E922BA84421F22C3C89CB9B
0
6FD60FE01941DDD77FE6B12893DA76E
[SwitchA-pkey-public-key-key1]EBC1D128D97F0678D7722B5341C8506F358214B16A2FAC4B
3
68950387811C7DA33021500C773218C
[SwitchA-pkey-public-key-key1]737EC8EE993B4F2DED30F48EDACE915F0281810082269009
E
14EC474BAF2932E69D3B1F18517AD95
[SwitchA-pkey-public-key-key1]94184CCDFCEAE96EC4D5EF93133E84B47093C52B20CD35D0
2
492B3959EC6499625BC4FA5082E22C5
[SwitchA-pkey-public-key-key1]B374E16DD00132CE71B020217091AC717B612391C76C1FB2
E
88317C1BD8171D41ECB83E210C03CC9
[SwitchA-pkey-public-key-key1]B32E810561C21621C73D6DAAC028F4B1585DA7F42519718C
C
9B09EEF03818400028180000AF995917
[SwitchA-pkey-public-key-key1]E1E570A3F6B1C2411948B3B4FFA256699B3BF871221CC9C5
D
F257523777D033BEE77FC378145F2AD
[SwitchA-pkey-public-key-key1]D716D7DB9FCABB4ADBF6FB4FDB0CA25C761B308EF53009F7
1
01F7C62621216D5A572C379A32AC290
[SwitchA-pkey-public-key-key1]E55B394A217DA38B65B77F0185C8DB8095522D1EF044B465
E
8716261214A5A3B493E866991113B2D
[SwitchA-pkey-public-key-key1]485348
[SwitchA-pkey-public-key-key1] peer-public-key end
```

Specify public key algorithm **dsa** for SSH2.

```
[SwitchA] ssh2 algorithm public-key dsa
[SwitchA] quit
```

Establish an SSH connection to the server, and specify the host public key of the server.

```
<SwitchA> ssh2 192.168.1.40 public-key key1
Username: client001
Press CTRL+C to abort.
Connecting to 192.168.1.40 port 22.
client001@192.168.1.40's password:
Enter a character ~ and a dot to abort.

**************************************************************************
* Copyright (c) 2004-2017 New H3C Technologies Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                             *
* no decompiling or reverse-engineering shall be allowed.                *
**************************************************************************

<SwitchB>
```

After you enter the correct username and password, you can successfully log in to Switch B.

- If the client already has the server's host public key:

Establish an SSH connection to server 192.168.1.40.

```
<SwitchA> ssh2 192.168.1.40
Username: client001
Press CTRL+C to abort.
Connecting to 192.168.1.40 port 22.
client001@192.168.1.40's password:
Enter a character ~ and a dot to abort.
************************************************************************
* Copyright (c) 2004-2017 New H3C Technologies Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                           *
* no decompiling or reverse-engineering shall be allowed.              *
************************************************************************
<SwitchB>
```

After you enter the correct username and password, you can successfully log in to Switch B.

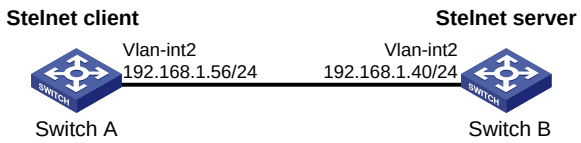
Example: Configuring the device as an Stelnet client (publickey authentication)

Network configuration

As shown in Figure 11:

- Switch B acts as the Stelnet server, and it uses publickey authentication and the DSA public key algorithm.
- Switch A acts as the Stelnet client. After the user on Switch A logs in to Switch B through Stelnet, the user can configure and manage Switch B as a network administrator.

Figure 11 Network diagram



Procedure

In the server configuration, the client's host public key is required. Generate a DSA key pair on the client before configuring the Stelnet server.

1. Configure the Stelnet client:

   # Assign an IP address to VLAN-interface 2.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ip address 192.168.1.56 255.255.255.0
   [SwitchA-Vlan-interface2] quit
   ```

   # Generate a DSA key pair.

   ```
   [SwitchA] public-key local create dsa
   The range of public key modulus is (512 ~ 2048).
   If the key modulus is greater than 512, it will take a few minutes.
   Press CTRL+C to abort.
   Input the modulus length [default = 1024]:
   Generating Keys...
   ```

```
.++++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.....................................+
...+................+.........+...+
Create the key pair successfully.
```

Export the DSA host public key to a public key file named **key.pub**.

```
[SwitchA] public-key local export dsa ssh2 key.pub
[SwitchA] quit
```

Transmit the public key file **key.pub** to the server through FTP or TFTP. (Details not shown.)

2. Configure the Stelnet server:

Generate RSA key pairs.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
The range of public key modulus is (512 ~ 4096)
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.......................++++++
...................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

Generate a DSA key pair.

```
[SwitchB] public-key local create dsa
The range of public key modulus is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.....................................+
...+................+.........+...+
Create the key pair successfully.
```

Generate an ECDSA key pair.

```
[SwitchB] public-key local create ecdsa secp256r1
Generating Keys...
.
Create the key pair successfully.
```

Enable the Stelnet server.

```
[SwitchB] ssh server enable
```

Assign an IP address to VLAN-interface 2. The Stelnet client uses this address as the destination address for SSH connection.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 192.168.1.40 255.255.255.0
[SwitchB-Vlan-interface2] quit
```

Set the authentication mode to AAA for user lines.

```
[SwitchB] line vty 0 63
[SwitchB-line-vty0-63] authentication-mode scheme
```

```
[SwitchB-line-vty0-63] quit
```

Import the peer public key from the public key file **key.pub**, and name it **switchkey**.

```
[SwitchB] public-key peer switchkey import sshkey key.pub
```

Create an SSH user named **client002**. Specify the authentication method as **publickey** for the user. Assign the public key **switchkey** to the user.

```
[SwitchB] ssh user client002 service-type stelnet authentication-type publickey
assign publickey switchkey
```

Create a local device management user named **client002**.

```
[SwitchB] local-user client002 class manage
```

Authorize local user **client002** to use the **SSH** service.

```
[SwitchB-luser-manage-client002] service-type ssh
```

Assign the **network-admin** user role to local user **client002**.

```
[SwitchB-luser-manage-client002] authorization-attribute user-role network-admin
[SwitchB-luser-manage-client002] quit
```

3. Establish an SSH connection to the Stelnet server.

```
<SwitchA> ssh2 192.168.1.40 identity-key dsa
Username: client002
Press CTRL+C to abort.
Connecting to 192.168.1.40 port 22.
The server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
Enter a character ~ and a dot to abort.

*****************************************************************************
* Copyright (c) 2004-2017 New H3C Technologies Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                                *
* no decompiling or reverse-engineering shall be allowed.                   *
*****************************************************************************

<SwitchB>
```

After you enter username **client002** and then enter **y** to continue accessing the server, you can log in to the server successfully.

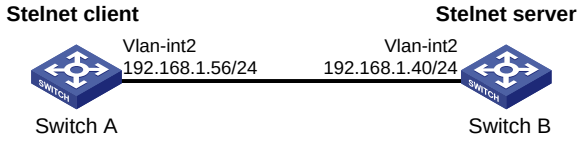
Example: Configuring Stelnet based on 128-bit Suite B algorithms

Network configuration

As shown in Figure 12:

- Switch B acts as the Stelnet Suite B server (SSH2), and it uses publickey authentication to authenticate the Stelnet client.
- Switch A acts as an Stelnet Suite B client (SSH2). After the user on Switch A logs in to Switch B through the Stelnet Suite B client software, the user can configure and manage Switch B as an administrator.

Figure 12 Network diagram



Procedure

1. Generate the client's certificate and the server's certificate. (Details not shown.)

   You must first configure the certificates of the server and the client because they are required for identity authentication between the two parties.

   In this example, the server's certificate file is **ssh-server-ecdsa256.p12** and the client's certificate file is **ssh-client-ecdsa256.p12**.

2. Configure the Stelnet client:

   You can modify the pkix version of the client software OpenSSH to support Suite B. This example uses an H3C switch as an Stelnet client.

   # Upload the server's certificate file **ssh-server-ecdsa256.p12** and the client's certificate file **ssh-client-ecdsa256.p12** to the Stelnet client through FTP or TFTP. (Details not shown.)

   # Create a PKI domain named **server256** for verifying the server's certificate and enter its view.

   ```
   <SwitchA> system-view
   [SwitchA] pki domain server256
   ```

   # Disable CRL checking.

   ```
   [SwitchA-pki-domain-server256] undo crl check enable
   [SwitchA-pki-domain-server256] quit
   ```

   # Import local certificate file **ssh-server-ecdsa256.p12** to PKI domain **server256**.

   ```
   [SwitchA] pki import domain server256 p12 local filename ssh-server-ecdsa256.p12
   The system is going to save the key pair. You must specify a key pair name, which is
   a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
   to Z, 0 to 9, and hyphens (-).
   Please enter the key pair name[default name: server256]:
   ```

   # Display information about local certificates in PKI domain **server256**.

   ```
   [SwitchA] display pki certificate domain server256 local
   Certificate:
       Data:
           Version: 3 (0x2)
           Serial Number: 3 (0x3)
       Signature Algorithm: ecdsa-with-SHA256
           Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
           Validity
               Not Before: Aug 21 08:39:51 2015 GMT
               Not After : Aug 20 08:39:51 2016 GMT
           Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=SSH Server secp256
           Subject Public Key Info:
               Public Key Algorithm: id-ecPublicKey
                   Public-Key: (256 bit)
                   pub:
                       04:a2:b4:b4:66:1e:3b:d5:50:50:0e:55:19:8d:52:
                       6d:47:8c:3d:3d:96:75:88:2f:9a:ba:a2:a7:f9:ef:
                       0a:a9:20:b7:b6:6a:90:0e:f8:c6:de:15:a2:23:81:
   ```

```
                    3c:9e:a2:b7:83:87:b9:ad:28:c8:2a:5e:58:11:8e:
                    c7:61:4a:52:51
                ASN1 OID: prime256v1
                NIST CURVE: P-256
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
            X509v3 Subject Key Identifier:
                08:C1:F1:AA:97:45:19:6A:DA:4A:F2:87:A1:1A:E8:30:BD:31:30:D7
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA256
         30:65:02:31:00:a9:16:e9:c1:76:f0:32:fc:4b:f9:8f:b6:7f:
         31:a0:9f:de:a7:cc:33:29:27:2c:71:2e:f9:0d:74:cb:25:c9:
         00:d2:52:18:7f:58:3f:cc:7e:8b:d3:42:65:00:cb:63:f8:02:
         30:01:a2:f6:a1:51:04:1c:61:78:f6:6b:7e:f9:f9:42:8d:7c:
         a7:bb:47:7c:2a:85:67:0d:81:12:0b:02:98:bc:06:1f:c1:3c:
         9b:c2:1b:4c:44:38:5a:14:b2:48:63:02:2b
```

Create a PKI domain named **client256** for the client's certificate and enter its view.

```
[SwitchA] pki domain client256
```

Disable CRL checking.

```
[SwitchA-pki-domain-client256] undo crl check enable
[SwitchA-pki-domain-client256] quit
```

Import local certificate file **ssh-client-ecdsa256.p12** to PKI domain **client256**.

```
[SwitchA] pki import domain client256 p12 local filename ssh-client-ecdsa256.p12
The system is going to save the key pair. You must specify a key pair name, which is
a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
to Z, 0 to 9, and hyphens (-).
Please enter the key pair name[default name: client256]:
```

Display information about local certificates in PKI domain **client256**.

```
[SwitchA] display pki certificate domain client256 local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 4 (0x4)
    Signature Algorithm: ecdsa-with-SHA256
        Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
        Validity
            Not Before: Aug 21 08:41:09 2015 GMT
            Not After : Aug 20 08:41:09 2016 GMT
        Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=SSH Client secp256
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (256 bit)
                pub:
```

```
                    04:da:e2:26:45:87:7a:63:20:e7:ca:7f:82:19:f5:
                    96:88:3e:25:46:f8:2f:9a:4c:70:61:35:db:e4:39:
                    b8:38:c4:60:4a:65:28:49:14:32:3c:cc:6d:cd:34:
                    29:83:84:74:a7:2d:0e:75:1c:c2:52:58:1e:22:16:
                    12:d0:b4:8a:92
                ASN1 OID: prime256v1
                NIST CURVE: P-256
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
            X509v3 Subject Key Identifier:
                1A:61:60:4D:76:40:B8:BA:5D:A1:3C:60:BC:57:98:35:20:79:80:FC
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA256
         30:66:02:31:00:9a:6d:fd:7d:ab:ae:54:9a:81:71:e6:bb:ad:
         5a:2e:dc:1d:b3:8a:bf:ce:ee:71:4e:8f:d9:93:7f:a3:48:a1:
         5c:17:cb:22:fa:8f:b3:e5:76:89:06:9f:96:47:dc:34:87:02:
         31:00:e3:af:2a:8f:d6:8d:1f:3a:2b:ae:2f:97:b3:52:63:b6:
         18:67:70:2c:93:2a:41:c0:e7:fa:93:20:09:4d:f4:bf:d0:11:
         66:0f:48:56:01:1e:c3:be:37:4e:49:19:cf:c6
```

Assign an IP address to VLAN-interface 2.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.1.56 255.255.255.0
[SwitchA-Vlan-interface2] quit
```

3. Configure the Stelnet server:

Upload the server's certificate file **ssh-server-ecdsa256.p12** and the client's certificate file **ssh-client-ecdsa256.p12** to the Stelnet server through FTP or TFTP. (Details not shown.)

Create a PKI domain named **client256** for verifying the client's certificate and import the file of the client's certificate to this domain. (Details not shown.)

Create a PKI domain named **server256** for the server's certificate and import the file of the server's certificate to this domain. (Details not shown.)

Specify Suite B algorithms for algorithm negotiation.

```
<SwitchB> system-view
[SwitchB] ssh2 algorithm key-exchange ecdh-sha2-nistp256
[SwitchB] ssh2 algorithm cipher aes128-gcm
[SwitchB] ssh2 algorithm public-key x509v3-ecdsa-sha2-nistp256
x509v3-ecdsa-sha2-nistp384
```

Specify **server256** as the PKI domain of the server's certificate.

```
[SwitchB] ssh server pki-domain server256
```

Enable the Stelnet server.

```
[SwitchB] ssh server enable
```

Assign an IP address to VLAN-interface 2.

```
[SwitchB] interface vlan-interface 2
```

```
[SwitchB-Vlan-interface2] ip address 192.168.1.40 255.255.255.0
[SwitchB-Vlan-interface2] quit
```

Set the authentication mode to AAA for user lines.

```
[SwitchB] line vty 0 63
[SwitchB-line-vty0-63] authentication-mode scheme
[SwitchB-line-vty0-63] quit
```

Create a local device management user named **client001**. Authorize the user to use the **SSH** service and assign the **network-admin** user role to the user.

```
[SwitchB] local-user client001 class manage
[SwitchB-luser-manage-client001] service-type ssh
[SwitchB-luser-manage-client001] authorization-attribute user-role network-admin
[SwitchB-luser-manage-client001] quit
```

Create an SSH user named **client001**. Specify the **publickey** authentication method for the user and specify **client256** as the PKI domain for verifying the client's certificate.

```
[SwitchB] ssh user client001 service-type stelnet authentication-type publickey
assign pki-domain client256
```

4. Establish an SSH connection to the Stelnet server based on the 128-bit Suite B algorithms:

Establish an SSH connection to the server at 192.168.1.40.

```
<SwitchA>  ssh2 192.168.1.40 suite-b 128-bit pki-domain client256 server-pki-domain
server256
Username: client001
Press CTRL+C to abort.
Connecting to 192.168.1.40 port 22.
Enter a character ~ and a dot to abort.

******************************************************************************
* Copyright (c) 2004-2017 New H3C Technologies Co., Ltd. All rights reserved.*
* Without the owner's prior written consent,                                 *
* no decompiling or reverse-engineering shall be allowed.                    *
******************************************************************************

<SwitchB>
```

SFTP configuration examples

Unless otherwise noted, devices in the configuration examples operate in non-FIPS mode.

When the device acts as an SFTP server operating in FIPS mode, only ECDSA and RSA key pairs are supported. Do not generate a DSA key pair on the SFTP server.

Example: Configuring the device as an SFTP server (password authentication)

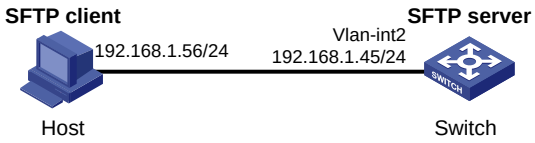
Network configuration

As shown in Figure 13:

- The switch acts as the SFTP server and uses password authentication to authenticate the SFTP client. The username and password of the client are saved on the switch.

- The host acts as the SFTP client. After the user on the client logs in to the switch through SFTP, the user can perform file management and transfer operations on the switch as a network administrator.

  Figure 13 Network diagram

Procedure

1. Configure the SFTP server:

   # Generate RSA key pairs.

   ```
   <Switch> system-view
   [Switch] public-key local create rsa
   The range of public key modulus is (512 ~ 4096).
   If the key modulus is greater than 512, it will take a few minutes.
   Press CTRL+C to abort.
   Input the modulus length [default = 1024]:
   Generating Keys...
   ........................++++++
   ...................++++++
   ..++++++++
   ............++++++++
   Create the key pair successfully.
   ```

   # Generate a DSA key pair.

   ```
   [Switch] public-key local create dsa
   The range of public key modulus is (512 ~ 2048).
   If the key modulus is greater than 512, it will take a few minutes.
   Press CTRL+C to abort.
   Input the modulus length [default = 1024]:
   Generating Keys...
   .++++++++++++++++++++++++++++++++++++++++++++++++++*
   ........+......+.....+.....................................+
   ...+................+..........+...+
   Create the key pair successfully.
   ```

   # Generate an ECDSA key pair.

   ```
   [Switch] public-key local create ecdsa secp256r1
   Generating Keys...
   .
   Create the key pair successfully.
   ```

   # Enable the SFTP server.

   ```
   [Switch] sftp server enable
   ```

   # Assign an IP address to VLAN-interface 2. The client uses this address as the destination for SSH connection.

   ```
   [Switch] interface vlan-interface 2
   [Switch-Vlan-interface2] ip address 192.168.1.45 255.255.255.0
   [Switch-Vlan-interface2] quit
   ```

Create a local device management user named **client002**.

```
[Switch] local-user client002 class manage
```

Set a password for local user **client002**.

```
[Switch-luser-manage-client002] password simple 123456TESTplat&!
```

Authorize local user **client002** to use the **SSH** service.

```
[Switch-luser-manage-client002] service-type ssh
```

Assign the **network-admin** user role and working directory **flash:/** to local user **client002**.

```
[Switch-luser-manage-client002] authorization-attribute user-role network-admin
work-directory flash:/
[Switch-luser-manage-client002] quit
```

Create an SSH user named **client002**. Specify the authentication method as **password** and service type as **sftp** for the user.

```
[Switch] ssh user client002 service-type sftp authentication-type password
```

2. Establish a connection between the SFTP client and the SFTP server:

This example uses an SFTP client that runs PSFTP of PuTTy version 0.58. PSFTP supports only password authentication.

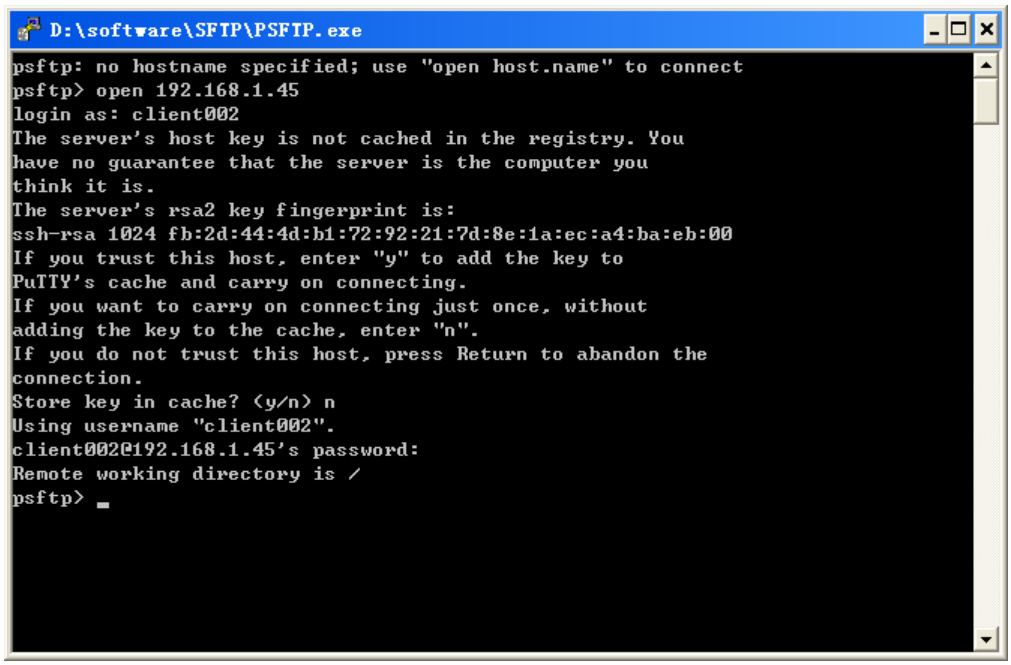
To establish a connection to the SFTP server:

a. Run the **psftp.exe** to launch the client interface shown in Figure 14, and enter the following command:

```
open 192.168.1.45
```

b. Enter username **client002** and the password to log in to the SFTP server.

Figure 14 SFTP client interface



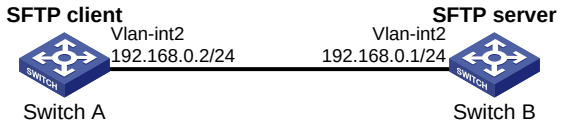
Example: Configuring the device as an SFTP client (publickey authentication)

Network configuration

As shown in Figure 15:

- Switch B acts as the SFTP server, and it uses publickey authentication and the RSA public key algorithm.
- Switch A acts as the SFTP client. After the user on Switch A logs in to Switch B through SFTP, the user can perform file management and transfer operations on Switch B as a network administrator.

Figure 15 Network diagram



Procedure

In the server configuration, the client's host public key is required. Generate RSA key pairs on the client before configuring the SFTP server.

1. Configure the SFTP client:

   # Assign an IP address to VLAN-interface 2.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.0.2 255.255.255.0
[SwitchA-Vlan-interface2] quit
```

   # Generate RSA key pairs.

```
[SwitchA] public-key local create rsa
The range of public key size is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

   # Export the host public key to a public key file named **pubkey**.

```
[SwitchA] public-key local export rsa ssh2 pubkey
[SwitchA] quit
```

   # Transmit the public key file **pubkey** to the server through FTP or TFTP. (Details not shown.)

2. Configure the SFTP server:

   # Generate RSA key pairs.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
The range of public key size is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
```

```
.............++++++++
Create the key pair successfully.
```

Generate a DSA key pair.

```
[SwitchB] public-key local create dsa
The range of public key size is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.....................................+
...+.................+..........+...+
Create the key pair successfully.
```

Generate an ECDSA key pair.

```
[SwitchB] public-key local create ecdsa secp256r1
Generating Keys...
.
Create the key pair successfully.
```

Enable the SFTP server.

```
[SwitchB] sftp server enable
```

Assign an IP address to VLAN-interface 2. The SSH client uses this address as the destination for SSH connection.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 192.168.0.1 255.255.255.0
[SwitchB-Vlan-interface2] quit
```

Import the peer public key from the public key file **pubkey**, and name it **switchkey**.

```
[SwitchB] public-key peer switchkey import sshkey pubkey
```

Create an SSH user named **client001**. Specify the service type as **sftp** and the authentication method as **publickey** for the user. Assign the public key **switchkey** to the user.

```
[SwitchB] ssh user client001 service-type sftp authentication-type publickey assign
publickey switchkey
```

Create a local device management user named **client001**.

```
[SwitchB] local-user client001 class manage
```

Authorize local user **client001** to use the **SSH** service.

```
[SwitchB-luser-manage-client001] service-type ssh
```

Assign the **network-admin** user role and working directory **flash:/** to local user **client001**.

```
[SwitchB-luser-manage-client001] authorization-attribute user-role network-admin
work-directory flash:/
[SwitchB-luser-manage-client001] quit
```

3. Establish a connection to the SFTP server:

Establish a connection to the SFTP server and enter SFTP client view.

```
<SwitchA> sftp 192.168.0.1 identity-key rsa
Username: client001
Press CTRL+C to abort.
Connecting to 192.168.0.1 port 22.
The server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
sftp>
```

Display files under the current directory of the server, delete file **z**, and verify the result.

```
sftp> dir -l
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
-rwxrwxrwx   1 noone    nogroup         0 Sep 01 08:00 z
sftp> delete z
Removing /z
sftp> dir -l
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
```

Add a directory named **new1** and verify the result.

```
sftp> mkdir new1
sftp> dir -l
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:30 new1
```

Change the name of directory **new1** to **new2** and verify the result.

```
sftp> rename new1 new2
sftp> dir -l
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:33 new2
```

Download file **pubkey2** from the server and save it as a local file named **public**.

```
sftp> get pubkey2 public
Fetching / pubkey2 to public
/pubkey2                                      100% 225     1.4KB/s   00:00
```

Upload the local file **pu** to the server, save it as **puk**, and verify the result.

```
sftp> put pu puk
Uploading pu to / puk
sftp> dir -l
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:33 new2
-rwxrwxrwx   1 noone    nogroup       283 Sep 02 06:35 pub
```

```
-rwxrwxrwx   1 noone    nogroup       283 Sep 02 06:36 puk
sftp>
```

Exit SFTP client view.

```
sftp> quit
<SwitchA>
```

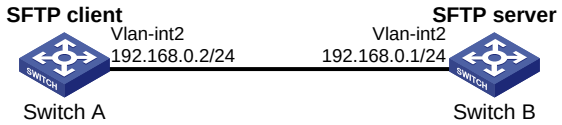
Example: Configuring SFTP based on 192-bit Suite B algorithms

Network configuration

As shown in Figure 16:

- Switch B acts as the SFTP Suite B server (SSH2), and it uses publickey authentication to authenticate the SFTP client.
- Switch A acts as an SFTP Suite B client (SSH2). After the user on Switch A logs in to Switch B based on the SFTP Suite B client software, the user can manage and transfer files on Switch B as an administrator.

Figure 16 Network diagram



Procedure

1. Generate the client's certificate and the server's certificate. (Details not shown.)

   You must first configure the certificates of the server and the client because they are required for identity authentication between the two parties.

   In this example, the server's certificate file is **ssh-server-ecdsa384.p12** and the client's certificate file is **ssh-client-ecdsa384.p12**.

2. Configure the SFTP client:

   You can modify the pkix version of the client software OpenSSH to support Suite B. This example uses an H3C switch as an SFTP client.

   # Upload the server's certificate file **ssh-server-ecdsa384.p12** and the client's certificate file **ssh-client-ecdsa384.p12** to the SFTP client through FTP or TFTP. (Details not shown.)

   # Create a PKI domain named **server384** for verifying the server's certificate and enter its view.

   ```
   <SwitchA> system-view
   [SwitchA] pki domain server384
   ```

   # Disable CRL checking.

   ```
   [SwitchA-pki-domain-server384] undo crl check enable
   [SwitchA-pki-domain-server384] quit
   ```

   # Import local certificate file **ssh-server-ecdsa384.p12** to PKI domain **server384**.

   ```
   [SwitchA] pki import domain server384 p12 local filename ssh-server-ecdsa384.p12
   The system is going to save the key pair. You must specify a key pair name, which is
   a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
   to Z, 0 to 9, and hyphens (-).
   Please enter the key pair name[default name: server384]:
   ```

   # Display information about local certificates in PKI domain **server384**.

   ```
   [SwitchA] display pki certificate domain server384 local
   ```

```
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 1 (0x1)
    Signature Algorithm: ecdsa-with-SHA384
        Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
        Validity
            Not Before: Aug 20 10:08:41 2015 GMT
            Not After : Aug 19 10:08:41 2016 GMT
        Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=ssh server
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (384 bit)
                pub:
                    04:4a:33:e5:99:8d:49:45:a7:a3:24:7b:32:6a:ed:
                    b6:36:e1:4d:cc:8c:05:22:f4:3a:7c:5d:b7:be:d1:
                    e6:9e:f0:ce:95:39:ca:fd:a0:86:cd:54:ab:49:60:
                    10:be:67:9f:90:3a:18:e2:7d:d9:5f:72:27:09:e7:
                    bf:7e:64:0a:59:bb:b3:7d:ae:88:14:94:45:b9:34:
                    d2:f3:93:e1:ba:b4:50:15:eb:e5:45:24:31:10:c7:
                    07:01:f9:dc:a5:6f:81
                ASN1 OID: secp384r1
                NIST CURVE: P-384
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
            X509v3 Subject Key Identifier:
                10:16:64:2C:DA:C1:D1:29:CD:C0:74:40:A9:70:BD:62:8A:BB:F4:D5
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA384
         30:65:02:31:00:80:50:7a:4f:c5:cd:6a:c3:57:13:7f:e9:da:
         c1:72:7f:45:30:17:c2:a7:d3:ec:73:3d:5f:4d:e3:96:f6:a3:
         33:fb:e4:b9:ff:47:f1:af:9d:e3:03:d2:24:53:40:09:5b:02:
         30:45:d1:bf:51:fd:da:22:11:90:03:f9:d4:05:ec:d6:7c:41:
         fc:9d:a1:fd:5b:8c:73:f8:b6:4c:c3:41:f7:c6:7f:2f:05:2d:
         37:f8:52:52:26:99:28:97:ac:6e:f9:c7:01
```

Create a PKI domain named **client384** for the client's certificate and enter its view.

```
[SwitchA] pki domain client384
```

Disable CRL checking.

```
[SwitchA-pki-domain-client384] undo crl check enable
[SwitchA-pki-domain-client384] quit
```

Import local certificate file **ssh-client-ecdsa384.p12** to PKI domain **client384**.

```
[SwitchA] pki import domain client384 p12 local filename ssh-client-ecdsa384.p12
```

```
The system is going to save the key pair. You must specify a key pair name, which is
a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
to Z, 0 to 9, and hyphens (-).
Please enter the key pair name[default name: client384]:
```

Display information about local certificates in PKI domain **client384**.

```
[SwitchA]display pki certificate domain client384 local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 2 (0x2)
    Signature Algorithm: ecdsa-with-SHA384
        Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
        Validity
            Not Before: Aug 20 10:10:59 2015 GMT
            Not After : Aug 19 10:10:59 2016 GMT
        Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=ssh client
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (384 bit)
                pub:
                    04:85:7c:8b:f4:7a:36:bf:74:f6:7c:72:f9:08:69:
                    d0:b9:ac:89:98:17:c9:fc:89:94:43:da:9a:a6:89:
                    41:d3:72:24:9b:9a:29:a8:d1:ba:b4:e5:77:ba:fc:
                    df:ae:c6:dd:46:72:ab:bc:d1:7f:18:7d:54:88:f6:
                    b4:06:54:7e:e7:4d:49:b4:07:dc:30:54:4b:b6:5b:
                    01:10:51:6b:0c:6d:a3:b1:4b:c9:d9:6c:d6:be:13:
                    91:70:31:2a:92:00:76
                ASN1 OID: secp384r1
                NIST CURVE: P-384
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
            X509v3 Subject Key Identifier:
                BD:5F:8E:4F:7B:FE:74:03:5A:D1:94:DB:CA:A7:82:D6:F7:78:A1:B0
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA384
         30:66:02:31:00:d2:06:fa:2c:0b:0d:f0:81:90:01:c3:3d:bf:
         97:b3:79:d8:25:a0:e2:0e:ed:00:c9:48:3e:c9:71:43:c9:b4:
         2a:a6:0a:27:80:9e:d4:0f:f2:db:db:5b:40:b1:a9:0a:e4:02:
         31:00:ee:00:e1:07:c0:2f:12:3f:88:ea:fe:19:05:ef:56:ca:
         33:71:75:5e:11:c9:a6:51:4b:3e:7c:eb:2a:4d:87:2b:71:7c:
         30:64:fe:14:ce:06:d5:0a:e2:cf:9a:69:19:ff
```

Assign an IP address to VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.0.2 255.255.255.0
```

```
[SwitchA-Vlan-interface2] quit
[SwitchA] quit
```

3. Configure the SFTP server:

   # Upload the server's certificate file **ssh-server-ecdsa384.p12** and the client's certificate file **ssh-client-ecdsa384.p12** to the SFTP server through FTP or TFTP. (Details not shown.)

   # Create a PKI domain named **client384** for verifying the client's certificate and import the file of the client's certificate to this domain. (Details not shown.)

   # Create a PKI domain named **server384** for the server's certificate and import the file of the server's certificate to this domain. (Details not shown.)

   # Specify Suite B algorithms for algorithm negotiation.

   ```
   [SwitchB] ssh2 algorithm key-exchange ecdh-sha2-nistp384
   [SwitchB] ssh2 algorithm cipher aes256-gcm
   [SwitchB] ssh2 algorithm public-key x509v3-ecdsa-sha2-nistp384
   ```

   # Specify **server384** as the PKI domain of the server's certificate.

   ```
   [SwitchB] ssh server pki-domain server384
   ```

   # Enable the SFTP server.

   ```
   [SwitchB] sftp server enable
   ```

   # Assign an IP address to VLAN-interface 2.

   ```
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] ip address 192.168.0.1 255.255.255.0
   [SwitchB-Vlan-interface2] quit
   ```

   # Set the authentication mode to AAA for user lines.

   ```
   [SwitchB] line vty 0 63
   [SwitchB-line-vty0-63] authentication-mode scheme
   [SwitchB-line-vty0-63] quit
   ```

   # Create a local device management user named **client001**. Authorize the user to use the **SSH** service and assign the **network-admin** user role to the user.

   ```
   [SwitchB] local-user client001 class manage
   [SwitchB-luser-manage-client001] service-type ssh
   [SwitchB-luser-manage-client001] authorization-attribute user-role network-admin
   [SwitchB-luser-manage-client001] quit
   ```

   # Create an SSH user named **client001**. Specify the **publickey** authentication method for the user and specify **client384** as the PKI domain for verifying the client's certificate.

   ```
   [SwitchB] ssh user client001 service-type sftp authentication-type publickey assign
   pki-domain client384
   ```

4. Establish an SFTP connection to the SFTP server based on the 192-bit Suite B algorithms:

   # Establish an SFTP connection to the server at 192.168.0.1.

   ```
   <SwitchA> sftp 192.168.0.1 suite-b 192-bit pki-domain client384 server-pki-domain
   server384
   Username: client001
   Press CTRL+C to abort.
   Connecting to 192.168.0.1 port 22.
   sftp>
   ```

SCP configuration examples

Unless otherwise noted, devices in the configuration examples operate in non-FIPS mode.

When the device acts as an SCP server operating in FIPS mode, only ECDSA and RSA key pairs are supported. Do not generate a DSA key pair on the SCP server.

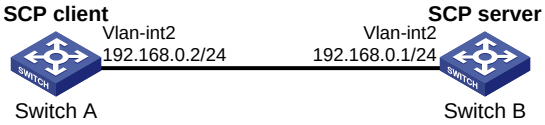
Example: Configuring SCP with password authentication

Network configuration

As shown in Figure 17:

- Switch B acts as the SCP server and uses password authentication to authenticate the SCP client. The client's username and password are saved on Switch B.
- Switch A acts as the SCP client. After the user on Switch A logs in to Switch B through SCP, the user can transfer files between switches as a network administrator.

Figure 17 Network diagram



Procedure

1. Configure the SCP server:

   # Generate RSA key pairs.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

   # Generate a DSA key pair.

```
[SwitchB] public-key local create dsa
The range of public key modulus is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.....................................+
...+.................+..........+...+.
Create the key pair successfully.
```

   # Generate an ECDSA key pair.

```
[SwitchB] public-key local create ecdsa secp256r1
Generating Keys...
.
```

```
Create the key pair successfully.
```

Enable the SCP server.

```
[SwitchB] scp server enable
```

Configure an IP address for VLAN-interface 2. The client uses this address as the destination for SCP connection.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 192.168.0.1 255.255.255.0
[SwitchB-Vlan-interface2] quit
```

Create a local device management user named **client001**.

```
[SwitchB] local-user client001 class manage
```

Set a password for local user **client001**.

```
[SwitchB-luser-manage-client001] password simple 123456TESTplat&!
```

Authorize local user **client001** to use the **SSH** service.

```
[SwitchB-luser-manage-client001] service-type ssh
```

Assign the **network-admin** user role to local user **client001**.

```
[SwitchB-luser-manage-client001] authorization-attribute user-role network-admin
[SwitchB-luser-manage-client001] quit
```

Create an SSH user named **client001**. Specify the service type as **scp** and the authentication method as **password** for the user.

```
[SwitchB] ssh user client001 service-type scp authentication-type password
```

2. Configure an IP address for VLAN-interface 2 on the SCP client.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.0.2 255.255.255.0
[SwitchA-Vlan-interface2] quit
[SwitchA] quit
```

3. Connect to the SCP server, download file **remote.bin** from the server, and save it as a local file named **local.bin**.

```
<SwitchA> scp 192.168.0.1 get remote.bin local.bin
Username: client001
Press CTRL+C to abort.
Connecting to 192.168.0.1 port 22.
The server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
client001@192.168.0.1's password:
remote.bin                                    100% 2875     2.8KB/s   00:00
```

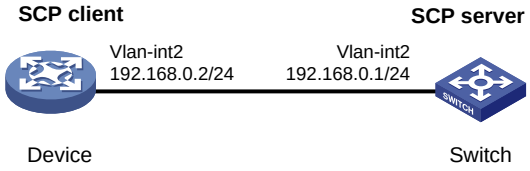
Example: Configuring SCP file transfer with a Linux SCP client

Network configuration

As shown in Figure 18, the switch acts as the SCP server and uses password authentication to authenticate the SCP client. The client's username and password are saved on the switch.

The device acts as the SCP client. After the user on the device logs in to the switch through SCP, the user can transfer files with the switch as a network administrator.

Figure 18 Network diagram



Procedure

1. Configure the SCP server:

   # Generate an RSA key pair.

```
<Switch> system-view
[Switch] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

   # Generate a DSA key pair.

```
[Switch] public-key local create dsa
The range of public key modulus is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.................................+
...+.................+..........+...+.
Create the key pair successfully.
```

   # Generate an ECDSA key pair.

```
[Switch] public-key local create ecdsa secp256r1
Generating Keys...
.
Create the key pair successfully.
```

   # Enable the SCP server.

```
[Switch] scp server enable
```

   # Configure an IP address for VLAN interface 2. The SCP client uses this IP address as the destination address of the SCP connection.

```
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.0.1 255.255.255.0
[Switch-Vlan-interface2] quit
```

   # Create a local device management user named **client001**. Set a password for the user, authorize the user to use the SSH service, assign the **network-admin** user role to the user, and specify the working directory as **flash:/**.

```
[Switch] local-user client001 class manage
[Switch-luser-manage-client001] password simple 123456TESTplat&!
[Switch-luser-manage-client001] service-type ssh
[Switch-luser-manage-client001] authorization-attribute user-role network-admin
work-directory flash:/
[Switch-luser-manage-client001] quit
```

Configure an SSH user named **client001**, specify the SCP service for the SSH user, and set the authentication method as password. (This step is optional.)

```
[Switch] ssh user client001 service-type scp authentication-type password
```

2. Configure the SCP client:

   The following uses the SCP client running on a Linux system as an example:

   # Connect to the SCP server, download the **remote.bin** file from the server, and save it as a local file named **local.bin**.

```
admin@device:~# scp client001@192.168.0.1:remote.bin local.bin
client001@192.168.0.1's password:
remote.bin                                                    100%   15
     0.0KB/s   00:00
```

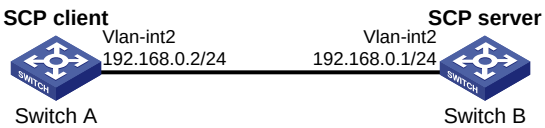
Example: Configuring SCP based on Suite B algorithms

Network configuration

As shown in Figure 19:

- Switch B acts as the SCP Suite B server (SSH2), and it uses publickey authentication to authenticate the SCP client.
- Switch A acts as an SCP Suite B client (SSH2). After the user on Switch A logs in to Switch B through the SCP Suite B client software, the user can transfer files between switches as a network administrator.

Figure 19 Network diagram



Procedure

1. Generate the client's certificates and the server's certificates. (Details not shown.)

   You must first configure the certificates of the server and the client because they are required for identity authentication between the two parties.

   In this example, the server's certificate files are **ssh-server-ecdsa256.p12** and **ssh-server-ecdsa384.p12**. The client's certificate files are **ssh-client-ecdsa256.p12** and **ssh-client-ecdsa384.p12**.

2. Configure the SCP client:

   You can modify the pkix version of the client software OpenSSH to support Suite B. This example uses an H3C switch as an SCP client.

   # Upload the server's certificate files (**ssh-server-ecdsa256.p12** and **ssh-server-ecdsa384.p12**) and the client's certificate files (**ssh-client-ecdsa256.p12** and **ssh-client-ecdsa384.p12**) to the SCP client through FTP or TFTP. (Details not shown.)

   # Create a PKI domain named **server256** for verifying the server's certificate **ecdsa256** and enter its view.

```
<SwitchA> system-view
[SwitchA] pki domain server256
```

Disable CRL checking.

```
[SwitchA-pki-domain-server256] undo crl check enable
[SwitchA-pki-domain-server256] quit
```

Import local certificate file **ssh-server-ecdsa256.p12** to PKI domain **server256**.

```
[SwitchA] pki import domain server256 p12 local filename ssh-server-ecdsa256.p12
The system is going to save the key pair. You must specify a key pair name, which is
a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
to Z, 0 to 9, and hyphens (-).
Please enter the key pair name[default name: server256]:
```

Display information about local certificates in PKI domain **server256**.

```
[SwitchA] display pki certificate domain server256 local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 3 (0x3)
    Signature Algorithm: ecdsa-with-SHA256
        Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
        Validity
            Not Before: Aug 21 08:39:51 2015 GMT
            Not After : Aug 20 08:39:51 2016 GMT
        Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=SSH Server secp256
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (256 bit)
                pub:
                    04:a2:b4:b4:66:1e:3b:d5:50:50:0e:55:19:8d:52:
                    6d:47:8c:3d:3d:96:75:88:2f:9a:ba:a2:a7:f9:ef:
                    0a:a9:20:b7:b6:6a:90:0e:f8:c6:de:15:a2:23:81:
                    3c:9e:a2:b7:83:87:b9:ad:28:c8:2a:5e:58:11:8e:
                    c7:61:4a:52:51
                ASN1 OID: prime256v1
                NIST CURVE: P-256
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
            X509v3 Subject Key Identifier:
                08:C1:F1:AA:97:45:19:6A:DA:4A:F2:87:A1:1A:E8:30:BD:31:30:D7
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA256
         30:65:02:31:00:a9:16:e9:c1:76:f0:32:fc:4b:f9:8f:b6:7f:
         31:a0:9f:de:a7:cc:33:29:27:2c:71:2e:f9:0d:74:cb:25:c9:
         00:d2:52:18:7f:58:3f:cc:7e:8b:d3:42:65:00:cb:63:f8:02:
```

```
        30:01:a2:f6:a1:51:04:1c:61:78:f6:6b:7e:f9:f9:42:8d:7c:
        a7:bb:47:7c:2a:85:67:0d:81:12:0b:02:98:bc:06:1f:c1:3c:
        9b:c2:1b:4c:44:38:5a:14:b2:48:63:02:2b
```

Create a PKI domain named **client256** for the client's certificate **ecdsa256** and enter its view.

```
[SwitchA] pki domain client256
```

Disable CRL checking.

```
[SwitchA-pki-domain-client256] undo crl check enable
[SwitchA-pki-domain-client256] quit
```

Import local certificate file **ssh-client-ecdsa256.p12** to PKI domain **client256**.

```
[SwitchA] pki import domain client256 p12 local filename ssh-client-ecdsa256.p12
The system is going to save the key pair. You must specify a key pair name, which is
a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
to Z, 0 to 9, and hyphens (-).
Please enter the key pair name[default name: client256]:
```

Display information about local certificates in PKI domain **client256**.

```
[SwitchA] display pki certificate domain client256 local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 4 (0x4)
    Signature Algorithm: ecdsa-with-SHA256
        Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
        Validity
            Not Before: Aug 21 08:41:09 2015 GMT
            Not After : Aug 20 08:41:09 2016 GMT
        Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=SSH Client secp256
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (256 bit)
                pub:
                    04:da:e2:26:45:87:7a:63:20:e7:ca:7f:82:19:f5:
                    96:88:3e:25:46:f8:2f:9a:4c:70:61:35:db:e4:39:
                    b8:38:c4:60:4a:65:28:49:14:32:3c:cc:6d:cd:34:
                    29:83:84:74:a7:2d:0e:75:1c:c2:52:58:1e:22:16:
                    12:d0:b4:8a:92
                ASN1 OID: prime256v1
                NIST CURVE: P-256
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
            X509v3 Subject Key Identifier:
                1A:61:60:4D:76:40:B8:BA:5D:A1:3C:60:BC:57:98:35:20:79:80:FC
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA256
```

```
        30:66:02:31:00:9a:6d:fd:7d:ab:ae:54:9a:81:71:e6:bb:ad:
        5a:2e:dc:1d:b3:8a:bf:ce:ee:71:4e:8f:d9:93:7f:a3:48:a1:
        5c:17:cb:22:fa:8f:b3:e5:76:89:06:9f:96:47:dc:34:87:02:
        31:00:e3:af:2a:8f:d6:8d:1f:3a:2b:ae:2f:97:b3:52:63:b6:
        18:67:70:2c:93:2a:41:c0:e7:fa:93:20:09:4d:f4:bf:d0:11:
        66:0f:48:56:01:1e:c3:be:37:4e:49:19:cf:c6
```

Create a PKI domain named **server384** for verifying the server's certificate **ecdsa384** and enter its view.

```
[SwitchA] pki domain server384
```

Disable CRL checking.

```
[SwitchA-pki-domain-server384] undo crl check enable
[SwitchA-pki-domain-server384] quit
```

Import local certificate file **ssh-server-ecdsa384.p12** to PKI domain **server384**.

```
[SwitchA] pki import domain server384 p12 local filename ssh-server-ecdsa384.p12
The system is going to save the key pair. You must specify a key pair name, which is
a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
to Z, 0 to 9, and hyphens (-).
Please enter the key pair name[default name: server384]:
```

Display information about local certificates in PKI domain **server384**.

```
[SwitchA] display pki certificate domain server384 local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 1 (0x1)
    Signature Algorithm: ecdsa-with-SHA384
        Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
        Validity
            Not Before: Aug 20 10:08:41 2015 GMT
            Not After : Aug 19 10:08:41 2016 GMT
        Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=ssh server
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (384 bit)
                pub:
                    04:4a:33:e5:99:8d:49:45:a7:a3:24:7b:32:6a:ed:
                    b6:36:e1:4d:cc:8c:05:22:f4:3a:7c:5d:b7:be:d1:
                    e6:9e:f0:ce:95:39:ca:fd:a0:86:cd:54:ab:49:60:
                    10:be:67:9f:90:3a:18:e2:7d:d9:5f:72:27:09:e7:
                    bf:7e:64:0a:59:bb:b3:7d:ae:88:14:94:45:b9:34:
                    d2:f3:93:e1:ba:b4:50:15:eb:e5:45:24:31:10:c7:
                    07:01:f9:dc:a5:6f:81
                ASN1 OID: secp384r1
                NIST CURVE: P-384
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
```

```
            X509v3 Subject Key Identifier:
                10:16:64:2C:DA:C1:D1:29:CD:C0:74:40:A9:70:BD:62:8A:BB:F4:D5
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA384
         30:65:02:31:00:80:50:7a:4f:c5:cd:6a:c3:57:13:7f:e9:da:
         c1:72:7f:45:30:17:c2:a7:d3:ec:73:3d:5f:4d:e3:96:f6:a3:
         33:fb:e4:b9:ff:47:f1:af:9d:e3:03:d2:24:53:40:09:5b:02:
         30:45:d1:bf:51:fd:da:22:11:90:03:f9:d4:05:ec:d6:7c:41:
         fc:9d:a1:fd:5b:8c:73:f8:b6:4c:c3:41:f7:c6:7f:2f:05:2d:
         37:f8:52:52:26:99:28:97:ac:6e:f9:c7:01
```

Create a PKI domain named **client384** for the client's certificate **ecdsa384** and enter its view.

```
[SwitchA] pki domain client384
```

Disable CRL checking.

```
[SwitchA-pki-domain-client384] undo crl check enable
[SwitchA-pki-domain-client384] quit
```

Import local certificate file **ssh-client-ecdsa384.p12** to PKI domain **client384**.

```
[SwitchA] pki import domain client384 p12 local filename ssh-client-ecdsa384.p12
The system is going to save the key pair. You must specify a key pair name, which is
a case-insensitive string of 1 to 64 characters. Valid characters include a to z, A
to Z, 0 to 9, and hyphens (-).
Please enter the key pair name[default name: client384]:
```

Display information about local certificates in PKI domain **client384**.

```
[SwitchA] display pki certificate domain client384 local
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 2 (0x2)
    Signature Algorithm: ecdsa-with-SHA384
        Issuer: C=CN, ST=Beijing, L=Beijing, O=H3C, OU=Software, CN=SuiteB CA
        Validity
            Not Before: Aug 20 10:10:59 2015 GMT
            Not After : Aug 19 10:10:59 2016 GMT
        Subject: C=CN, ST=Beijing, O=H3C, OU=Software, CN=ssh client
        Subject Public Key Info:
            Public Key Algorithm: id-ecPublicKey
                Public-Key: (384 bit)
                pub:
                    04:85:7c:8b:f4:7a:36:bf:74:f6:7c:72:f9:08:69:
                    d0:b9:ac:89:98:17:c9:fc:89:94:43:da:9a:a6:89:
                    41:d3:72:24:9b:9a:29:a8:d1:ba:b4:e5:77:ba:fc:
                    df:ae:c6:dd:46:72:ab:bc:d1:7f:18:7d:54:88:f6:
                    b4:06:54:7e:e7:4d:49:b4:07:dc:30:54:4b:b6:5b:
                    01:10:51:6b:0c:6d:a3:b1:4b:c9:d9:6c:d6:be:13:
                    91:70:31:2a:92:00:76
                ASN1 OID: secp384r1
                NIST CURVE: P-384
```

```
        X509v3 extensions:
            X509v3 Basic Constraints:
                CA:FALSE
            Netscape Comment:
                OpenSSL Generated Certificate
            X509v3 Subject Key Identifier:
                BD:5F:8E:4F:7B:FE:74:03:5A:D1:94:DB:CA:A7:82:D6:F7:78:A1:B0
            X509v3 Authority Key Identifier:
                keyid:5A:BE:85:49:16:E5:EB:33:80:25:EB:D8:91:50:B4:E6:3E:4F:B8:22

    Signature Algorithm: ecdsa-with-SHA384
         30:66:02:31:00:d2:06:fa:2c:0b:0d:f0:81:90:01:c3:3d:bf:
         97:b3:79:d8:25:a0:e2:0e:ed:00:c9:48:3e:c9:71:43:c9:b4:
         2a:a6:0a:27:80:9e:d4:0f:f2:db:db:5b:40:b1:a9:0a:e4:02:
         31:00:ee:00:e1:07:c0:2f:12:3f:88:ea:fe:19:05:ef:56:ca:
         33:71:75:5e:11:c9:a6:51:4b:3e:7c:eb:2a:4d:87:2b:71:7c:
         30:64:fe:14:ce:06:d5:0a:e2:cf:9a:69:19:ff
```

Assign an IP address to VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.0.2 255.255.255.0
[SwitchA-Vlan-interface2] quit
```

3. Configure the SCP server:

Upload the server's certificate files (**ssh-server-ecdsa256.p12** and **ssh-server-ecdsa384.p12**) and the client's certificate files (**ssh-client-ecdsa256.p12** and **ssh-client-ecdsa384.p12**) to the SCP server through FTP or TFTP. (Details not shown.)

Create a PKI domain named **client256** for verifying the client's certificate **ecdsa256** and import the file of this certificate to this domain. Create a PKI domain named **server256** for the server's certificate **ecdsa256** and import the file of this certificate to this domain. (Details not shown.)

Create a PKI domain named **client384** for verifying the client's certificate **ecdsa384** and import the file of this certificate to this domain. Create a PKI domain named **server384** for the server's certificate **ecdsa384** and import the file of this certificate to this domain. (Details not shown.)

Specify Suite B algorithms for algorithm negotiation.

```
<SwitchB> system-view
[SwitchB] ssh2 algorithm key-exchange ecdh-sha2-nistp256 ecdh-sha2-nistp384
[SwitchB] ssh2 algorithm cipher aes128-gcm aes256-gcm
[SwitchB] ssh2 algorithm public-key x509v3-ecdsa-sha2-nistp256
x509v3-ecdsa-sha2-nistp384
```

Enable the SCP server.

```
[SwitchB] scp server enable
```

Assign an IP address to VLAN-interface 2.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 192.168.0.1 255.255.255.0
[SwitchB-Vlan-interface2] quit
```

Set the authentication mode to AAA for user lines.

```
[SwitchB] line vty 0 63
[SwitchB-line-vty0-63] authentication-mode scheme
[SwitchB-line-vty0-63] quit
```

Create a local device management user named **client001**. Authorize the user to use the **SSH** service and assign the **network-admin** user role to the user.

```
[SwitchB] local-user client001 class manage
[SwitchB-luser-manage-client001] service-type ssh
[SwitchB-luser-manage-client001] authorization-attribute user-role network-admin
[SwitchB-luser-manage-client001] quit
```

Create a local device management user named **client002**. Authorize the user to use the **SSH** service and assign the **network-admin** user role to the user.

```
[SwitchB] local-user client002 class manage
[SwitchB-luser-manage-client002] service-type ssh
[SwitchB-luser-manage-client002] authorization-attribute user-role network-admin
[SwitchB-luser-manage-client002] quit
```

4. Establish an SCP connection to the SCP server:

- Based on the 128-bit Suite B algorithms:

  # Specify **server256** as the PKI domain of the server's certificate.

  ```
  [SwitchB]ssh server pki-domain server256
  ```

  # Create an SSH user **client001**. Specify the **publickey** authentication method for the user and specify **client256** as the PKI domain for verifying the client's certificate.

  ```
  [SwitchB] ssh user client001 service-type scp authentication-type publickey assign pki-domain client256
  ```

  # Establish an SCP connection to the SCP server at 192.168.0.1 based on the 128-bit Suite B algorithms.

  ```
  <SwitchA> scp 192.168.0.1 get src.cfg suite-b 128-bit pki-domain client256 server-pki
  -domain server256
  Username: client001
  Press CTRL+C to abort.
  Connecting to 192.168.0.1 port 22.
  src.cfg                                       100% 4814     4.7KB/s   00:00
  <SwitchA>
  ```

- Based on the 192-bit Suite B algorithms:

  # Specify **server384** as the PKI domain of the server's certificate.

  ```
  [SwitchB] ssh server pki-domain server384
  ```

  # Create an SSH user **client002**. Specify the **publickey** authentication method for the user and specify **client384** as the PKI domain for verifying the client's certificate.

  ```
  [Switch] ssh user client002 service-type scp authentication-type publickey assign pki-domain client384
  ```

  # Establish an SCP connection to the SCP server at 192.168.0.1 based on the 192-bit Suite B algorithms.

  ```
  <SwitchA> scp 192.168.0.1 get src.cfg suite-b 192-bit pki-domain client384 server-pki
  -domain server384
  Username: client002
  Press CTRL+C to abort.
  Connecting to 192.168.0.1 port 22.
  src.cfg                                       100% 4814     4.7KB/s   00:00
  <SwitchA>
  ```

NETCONF over SSH configuration examples

Unless otherwise noted, devices in the configuration examples are in non-FIPS mode.

When the device acts as a NETCONF-over-SSH server operating in FIPS mode, only ECDSA and RSA key pairs are supported. Do not generate a DSA key pair on the NETCONF-over-SSH server.

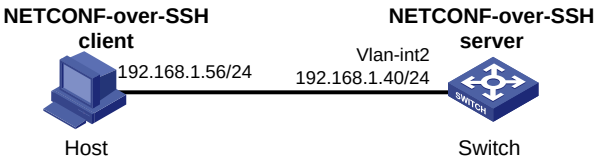
Example: Configuring NETCONF over SSH with password authentication

Network configuration

As shown in Figure 20:

- The switch acts as the NETCONF-over-SSH server and uses password authentication to authenticate the client. The client's username and password are saved on the switch.
- The host acts as the NETCONF-over-SSH client, using SSH2 client software. After the user on the host logs in to the switch through NETCONF over SSH, the user can perform NETCONF operations on the switch as a network administrator.

Figure 20 Network diagram



Procedure

Generate RSA key pairs.

```
<Switch> system-view
[Switch] public-key local create rsa
The range of public key modulus is (512 ~ 4096).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.....................++++++
.................++++++
..++++++++
............++++++++
Create the key pair successfully.
```

Generate a DSA key pair.

```
[Switch] public-key local create dsa
The range of public key modulus is (512 ~ 2048).
If the key modulus is greater than 512, it will take a few minutes.
Press CTRL+C to abort.
Input the modulus length [default = 1024]:
Generating Keys...
.++++++++++++++++++++++++++++++++++++++++++++++++*
........+......+.....+.....................................+
```

```
...+.................+.........+...+.
Create the key pair successfully.
```

Generate an ECDSA key pair.

```
[Switch] public-key local create ecdsa secp256r1
Generating Keys...
.
Create the key pair successfully.
```

Enable NETCONF over SSH.

```
[Switch] netconf ssh server enable
```

Configure an IP address for VLAN-interface 2. The client uses this address as the destination for NETCONF-over-SSH connection.

```
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.1.40 255.255.255.0
[Switch-Vlan-interface2] quit
```

Set the authentication mode to AAA for user lines.

```
[Switch] line vty 0 63
[Switch-line-vty0-63] authentication-mode scheme
[Switch-line-vty0-63] quit
```

Create a local device management user named **client001**.

```
[Switch] local-user client001 class manage
```

Set a password for local user **client001**.

```
[Switch-luser-manage-client001] password simple 123456TESTplat&!
```

Authorize local user **client001** to use the **SSH** service.

```
[Switch-luser-manage-client001] service-type ssh
```

Assign the **network-admin** user role to local user **client001**.

```
[Switch-luser-manage-client001] authorization-attribute user-role network-admin
[Switch-luser-manage-client001] quit
```

Create an SSH user named **client001**. Specify the service type as **NETCONF** and the authentication method as **password** for the user.

```
[Switch] ssh user client001 service-type netconf authentication-type password
```

Verifying the configuration

Verify that you can perform NETCONF operations after logging in to the switch. (Details not shown.)

Contents

Configuring SSL 1

About SSL 1
SSL security services 1
SSL protocol stack 1
SSL protocol versions 2
FIPS compliance 2
Restrictions and guidelines: SSL configuration 2
SSL tasks at a glance 2
Configuring an SSL server policy 3
Configuring an SSL client policy 4
Disabling SSL protocol versions for the SSL server 5
Disabling SSL session renegotiation 6
Display and maintenance commands for SSL 6

Configuring SSL

About SSL

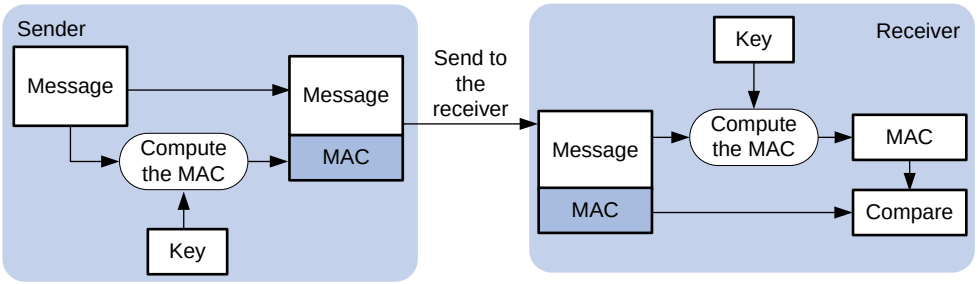
Secure Sockets Layer (SSL) is a cryptographic protocol that provides communication security for TCP-based application layer protocols such as HTTP. SSL has been widely used in applications such as e-business and online banking to provide secure data transmission over the Internet.

SSL security services

SSL provides the following security services:

- **Privacy**—SSL uses a symmetric encryption algorithm to encrypt data. It uses the asymmetric key algorithm of RSA to encrypt the key used by the symmetric encryption algorithm. For more information about RSA, see "Managing public keys."
- **Authentication**—SSL uses certificate-based digital signatures to authenticate the SSL server and client. The SSL server and client obtain digital certificates through PKI. For more information about PKI and digital certificates, see "Configuring PKI."
- **Integrity**—SSL uses the message authentication code (MAC) to verify message integrity. It uses a MAC algorithm and a key to transform a message of any length to a fixed-length message. Any change to the original message will result in a change to the calculated fixed-length message. As shown in Figure 1, the message integrity verification process is as follows:
  - **a.** The sender uses a MAC algorithm and a key to calculate a MAC value for a message. Then, it appends the MAC value to the message and sends the message to the receiver.
  - **b.** The receiver uses the same key and MAC algorithm to calculate a MAC value for the received message, and compares it with the MAC value appended to the message.
  - **c.** If the two MAC values match, the receiver considers the message intact. Otherwise, the receiver considers that the message was tampered with and it discards the message.

Figure 1 MAC algorithm diagram

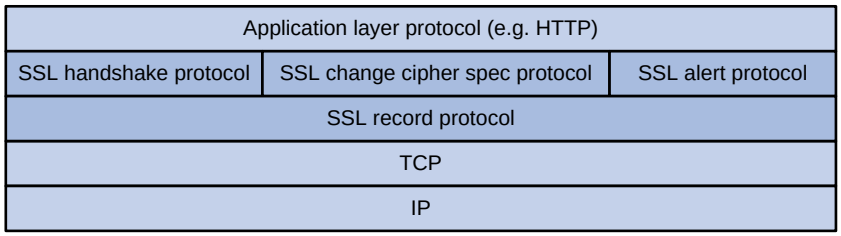


SSL protocol stack

The SSL protocol stack includes the following protocols:

- SSL record protocol at the lower layer.
- SSL handshake protocol, SSL change cipher spec protocol, and SSL alert protocol at the upper layer.

Figure 2 SSL protocol stack



The following describes the major functions of SSL protocols:

- **SSL record protocol**—Fragments data received from the upper layer, computes and adds MAC to the data, and encrypts the data.
- **SSL handshake protocol**—Negotiates the cipher suite used for secure communication, authenticates the server and client, and securely exchanges the keys between the server and client. The cipher suite that needs to be negotiated includes the symmetric encryption algorithm, key exchange algorithm, and MAC algorithm.
- **SSL change cipher spec protocol**—Notifies the receiver that subsequent packets are to be protected based on the negotiated cipher suite and key.
- **SSL alert protocol**—Sends alert messages to the receiving party. An alert message contains the alert severity level and a description.

SSL protocol versions

SSL protocol versions include SSL 2.0, SSL 3.0, TLS 1.0 (or SSL 3.1), TLS 1.1, and TLS 1.2. Because SSL 3.0 is known to be insecure, you can disable SSL 3.0 for the SSL server to ensure security.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see "Configuring FIPS") and non-FIPS mode.

Restrictions and guidelines: SSL configuration

By default, the SSL server can communicate with clients running all SSL protocol versions. When the server receives an SSL 2.0 Client Hello message from a client, it notifies the client to use a later version for communication.

SSL tasks at a glance

To configure SSL, perform the following tasks:

1. Configuring SSL server settings
   - Configuring an SSL server policy
   - (Optional.) Disabling SSL protocol versions for the SSL server
   - (Optional.) Disabling SSL session renegotiation
2. Configuring SSL client settings

   Configuring an SSL client policy

Configuring an SSL server policy

About this task

An SSL server policy is a set of SSL parameters used by the device when the device acts as the SSL server. An SSL server policy takes effect only after it is associated with an application such as HTTPS.

Procedure

1. Enter system view.

   system-view

2. Create an SSL server policy and enter its view.

   ssl server-policy *policy-name*

3. Specify a PKI domain for the SSL server policy.

   pki-domain *domain-name*

   By default, no PKI domain is specified for an SSL server policy.

   If SSL server authentication is required, you must specify a PKI domain and request a local certificate for the SSL server in the domain.

   For information about configuring a PKI domain, see "Configuring PKI."

4. Specify the cipher suites that the SSL server policy supports.

   In non-FIPS mode:

   ciphersuite { **dhe_rsa_aes_128_cbc_sha** | **dhe_rsa_aes_128_cbc_sha256** | **dhe_rsa_aes_256_cbc_sha** | **dhe_rsa_aes_256_cbc_sha256** | **ecdhe_ecdsa_aes_128_cbc_sha256** | **ecdhe_ecdsa_aes_128_gcm_sha256** | **ecdhe_ecdsa_aes_256_cbc_sha384** | **ecdhe_ecdsa_aes_256_gcm_sha384** | **ecdhe_rsa_aes_128_cbc_sha256** | **ecdhe_rsa_aes_128_gcm_sha256** | **ecdhe_rsa_aes_256_cbc_sha384** | **ecdhe_rsa_aes_256_gcm_sha384** | **exp_rsa_des_cbc_sha** | **exp_rsa_rc2_md5** | **exp_rsa_rc4_md5** | **rsa_3des_ede_cbc_sha** | **rsa_aes_128_cbc_sha** | **rsa_aes_128_cbc_sha256** | **rsa_aes_128_gcm_sha256** | **rsa_aes_256_cbc_sha** | **rsa_aes_256_cbc_sha256** | **rsa_aes_256_gcm_sha384** | **rsa_des_cbc_sha** | **rsa_rc4_128_md5** | **rsa_rc4_128_sha** } *

   In FIPS mode:

   ciphersuite { **ecdhe_ecdsa_aes_128_cbc_sha256** | **ecdhe_ecdsa_aes_256_cbc_sha384** | **ecdhe_ecdsa_aes_128_gcm_sha256** | **ecdhe_ecdsa_aes_256_gcm_sha384** | **ecdhe_rsa_aes_128_cbc_sha256** | **ecdhe_rsa_aes_128_gcm_sha256** | **ecdhe_rsa_aes_256_cbc_sha384** | **ecdhe_rsa_aes_256_gcm_sha384** | **rsa_aes_128_cbc_sha** | **rsa_aes_128_cbc_sha256**| **rsa_aes_128_gcm_sha256** | **rsa_aes_256_cbc_sha** | **rsa_aes_256_cbc_sha256** | **rsa_aes_256_gcm_sha384** } *

   By default, an SSL server policy supports all cipher suites.

5. (Optional.) Set the maximum number of sessions that the SSL server can cache and the session cache timeout time.

   session { **cachesize** *size* | **timeout** *time* } *

   By default, the SSL server can cache a maximum of 500 sessions, and the session cache timeout time is 3600 seconds.

6. Enable mandatory or optional SSL client authentication.

   client-verify { **enable** | **optional** }

By default, SSL client authentication is disabled. The SSL server does not perform digital certificate-based authentication on SSL clients.

When authenticating a client by using the digital certificate, the SSL server verifies the certificate chain presented by the client. It also verifies that the certificates in the certificate chain (except the root CA certificate) are not revoked.

7. (Optional.) Enable the SSL server to send the complete certificate chain to the client during SSL negotiation.

   certificate-chain-sending enable

   By default, the SSL server sends the server certificate rather than the complete certificate chain to the client during negotiation.

Configuring an SSL client policy

About this task

An SSL client policy is a set of SSL parameters used by the device when the device acts as the SSL client. The SSL client uses the settings in the client policy to establish a connection to the server. An SSL client policy takes effect only after it is associated with an application such as DDNS. For information about DDNS, see the DNS configuration in *Layer 3—IP Services Configuration Guide*.

Restrictions and guidelines

As a best practice to enhance system security, do not specify SSL 3.0 for the SSL client policy.

Procedure

1. Enter system view.

   system-view

2. Create an SSL client policy and enter its view.

   ssl client-policy *policy-name*

3. Specify a PKI domain for the SSL client policy.

   pki-domain *domain-name*

   By default, no PKI domain is specified for an SSL client policy.

   If SSL client authentication is required, you must specify a PKI domain and request a local certificate for the SSL client in the PKI domain.

   For information about configuring a PKI domain, see "Configuring PKI."

4. Specify the preferred cipher suite for the SSL client policy.

   In non-FIPS mode:

   prefer-cipher { **dhe_rsa_aes_128_cbc_sha** | **dhe_rsa_aes_128_cbc_sha256** | **dhe_rsa_aes_256_cbc_sha** | **dhe_rsa_aes_256_cbc_sha256** | **ecdhe_ecdsa_aes_128_cbc_sha256** | **ecdhe_ecdsa_aes_128_gcm_sha256** | **ecdhe_ecdsa_aes_256_cbc_sha384** | **ecdhe_ecdsa_aes_256_gcm_sha384** | **ecdhe_rsa_aes_128_cbc_sha256** | **ecdhe_rsa_aes_128_gcm_sha256** | **ecdhe_rsa_aes_256_cbc_sha384** | **ecdhe_rsa_aes_256_gcm_sha384** | **exp_rsa_des_cbc_sha** | **exp_rsa_rc2_md5** | **exp_rsa_rc4_md5** | **rsa_3des_ede_cbc_sha** | **rsa_aes_128_cbc_sha** | **rsa_aes_128_cbc_sha256** | **rsa_aes_128_gcm_sha256** | **rsa_aes_256_cbc_sha** | **rsa_aes_256_cbc_sha256** | **rsa_aes_256_gcm_sha384** | **rsa_des_cbc_sha** | **rsa_rc4_128_md5** | **rsa_rc4_128_sha** }

   In FIPS mode:

   prefer-cipher { **ecdhe_ecdsa_aes_128_cbc_sha256** | **ecdhe_ecdsa_aes_128_gcm_sha256** | **ecdhe_ecdsa_aes_256_cbc_sha384** | **ecdhe_ecdsa_aes_256_gcm_sha384** | **ecdhe_rsa_aes_128_cbc_sha256** |

ecdhe_rsa_aes_128_gcm_sha256 | **ecdhe_rsa_aes_256_cbc_sha384** | **ecdhe_rsa_aes_256_gcm_sha384** | **rsa_aes_128_cbc_sha** | **rsa_aes_128_cbc_sha256** | **rsa_aes_128_gcm_sha256** | **rsa_aes_256_cbc_sha** | **rsa_aes_256_cbc_sha256** | **rsa_aes_256_gcm_sha384** }

By default:

- In non-FIPS mode, the preferred cipher suite is **rsa_rc4_128_md5**.
- In FIPS mode, the preferred cipher suite is **rsa_aes_128_cbc_sha**.

5. Specify the SSL protocol version for the SSL client policy.

   In non-FIPS mode:

   version { **ssl3.0** | **tls1.0** | **tls1.1** | **tls1.2** }

   In FIPS mode:

   version { **tls1.0** | **tls1.1** | **tls1.2** }

   By default, an SSL client policy uses TLS 1.0.

6. Enable the SSL client to authenticate servers through digital certificates.

   server-verify enable

   By default, SSL server authentication is enabled.

Disabling SSL protocol versions for the SSL server

About this task

To enhance system security, you can disable the SSL server from using specific SSL protocol versions (SSL 3.0, TLS 1.0, and TLS 1.1) for session negotiation.

Restrictions and guidelines

Disabling an SSL protocol version does not affect the availability of earlier SSL protocol versions. For example, if you execute the **ssl version tls1.1 disable** command, TLS 1.1 is disabled but TLS 1.0 is still available for the SSL server.

Procedure

1. Enter system view.

   system-view

2. Disable SSL protocol versions for the SSL server.

   In non-FIPS mode:

   ssl version { **ssl3.0** | **tls1.0** | **tls1.1** } * **disable**

   In FIPS mode:

   ssl version { **tls1.0** | **tls1.1** } * **disable**

   By default:

   - In non-FIPS mode, the SSL server supports SSL 3.0, TLS 1.0, TLS 1.1, and TLS 1.2.
   - In FIPS mode, the SSL server supports TLS 1.0, TLS 1.1, and TLS 1.2.

Disabling SSL session renegotiation

About this task

The SSL session renegotiation feature enables the SSL client and server to reuse a previously negotiated SSL session for an abbreviated handshake.

Disabling session renegotiation causes more computational overhead to the system but it can avoid potential risks.

Restrictions and guidelines

Disable SSL session renegotiation only when explicitly required.

Procedure

1. Enter system view.
   `system-view`
2. Disable SSL session renegotiation.
   `ssl renegotiation disable`
   By default, SSL session renegotiation is enabled.

Display and maintenance commands for SSL

Execute **`display`** commands in any view.

Task	Command
Display SSL client policy information.	**`display ssl client-policy`** [*`policy-name`*]
Display SSL server policy information.	**`display ssl server-policy`** [*`policy-name`*]

Contents

Configuring attack detection and prevention 1

About attack detection and prevention 1
Attacks that the device can prevent 1
Single-packet attacks 1
Scanning attacks 2
Flood attacks 3
TCP fragment attack 4
Login dictionary attack 4
Blacklist feature 4
IP blacklist 4
User blacklist 5
Attack detection and prevention tasks at a glance 5
Configuring and applying an attack defense policy 5
Creating an attack defense policy 5
Configuring a single-packet attack defense policy 6
Configuring a scanning attack defense policy 7
Configuring a flood attack defense policy 8
Configuring attack detection exemption 12
Applying an attack defense policy to an interface 13
Applying an attack defense policy to the device 13
Enabling log non-aggregation for single-packet attack events 14
Configuring TCP fragment attack prevention 14
Configuring the IP blacklist feature 15
Configuring the user blacklist feature 15
Configuring login attack prevention 16
Enabling the login delay 16
Display and maintenance commands for attack detection and prevention 17
Attack detection and prevention configuration examples 22
Example: Configuring attack detection and prevention on the device 22
Example: Configuring IP blacklist 26
Example: Configuring user blacklist 27

Configuring attack detection and prevention

About attack detection and prevention

Attack detection and prevention enables a device to detect attacks by inspecting arriving packets, and to take prevention actions to protect a private network. Prevention actions include:

- Logging.
- Packet dropping.
- Client verification.

Attacks that the device can prevent

This section describes the attacks that the device can detect and prevent.

Single-packet attacks

Single-packet attacks are also known as malformed packet attacks. An attacker typically launches single-packet attacks by using the following methods:

- An attacker sends defective packets to a device, which causes the device to malfunction or crash.
- An attacker sends normal packets to a device, which interrupts connections or probes network topologies.
- An attacker sends a large number of forged packets to a target device, which consumes network bandwidth and causes denial of service (DoS).

Table 1 lists the single-packet attack types that the device can detect and prevent.

Table 1 Types of single-packet attacks

Single-packet attack	Description
ICMP redirect	An attacker sends ICMP redirect messages to modify the victim's routing table. The victim cannot forward packets correctly.
ICMP destination unreachable	An attacker sends ICMP destination unreachable messages to cut off the connections between the victim and its destinations.
ICMP type	A receiver responds to an ICMP packet according to its type. An attacker sends forged ICMP packets of a specific type to affect the packet processing of the victim.
ICMPv6 type	A receiver responds to an ICMPv6 packet according to its type. An attacker sends forged ICMPv6 packets of specific types to affect the packet processing of the victim.
Land	An attacker sends the victim a large number of TCP SYN packets, which contain the victim's IP address as the source and destination IP addresses. This attack exhausts the half-open connection resources on the victim, and locks the victim's system.
Large ICMP packet	An attacker sends large ICMP packets to crash the victim. Large ICMP packets can cause memory allocation error and crash the protocol stack.

Single-packet attack	Description
Large ICMPv6 packet	An attacker sends large ICMPv6 packets to crash the victim. Large ICMPv6 packets can cause memory allocation error and crash the protocol stack.
IP options	An attacker sends IP datagrams in which the IP options are abnormal. This attack intends to probe the network topology. The target system will break down if it is incapable of processing error packets.
IP fragment	An attacker sends the victim an IP datagram with an offset smaller than 5, which causes the victim to malfunction or crash.
IP impossible packet	An attacker sends IP packets whose source IP address is the same as the destination IP address, which causes the victim to malfunction.
Tiny fragment	An attacker makes the fragment size small enough to force Layer 4 header fields into the second fragment. These fragments can pass the packet filtering because they do not hit any match.
Smurf	An attacker sends an ICMP echo request to target networks. In these requests, the destination IP address is a network or broadcast address of a Class A, B, or C subnet, and the source IP address is the victim's IP address. Every receiver on the target networks will send an ICMP echo reply to the victim. The victim will be flooded with replies, and will be unable to provide services. Network congestion might occur.
TCP flag	An attacker sends packets with defective TCP flags to probe the operating system of the target host. Different operating systems process unconventional TCP flags differently. The target system will break down if it processes this type of packets incorrectly.
Traceroute	An attacker uses traceroute tools to probe the topology of the victim network.
WinNuke	An attacker sends Out-Of-Band (OOB) data to the TCP port 139 (NetBIOS) on the victim that runs Windows system. The malicious packets contain an illegal Urgent Pointer, which causes the victim's operating system to crash.
UDP bomb	An attacker sends a malformed UDP packet. The length value in the IP header is larger than the IP header length plus the length value in the UDP header. When the target system processes the packet, a buffer overflow can occur, which causes a system crash.
UDP Snork	An attacker sends a UDP packet with destination port 135 (the Microsoft location service) and source port 135, 7, or 19. This attack causes an NT system to exhaust its CPU.
UDP Fraggle	An attacker sends a large number of packets with source UDP port 7 and destination UDP port 19 (UDP chargen port) to a network. These packets use the victim's IP address as the source IP address. Replies will flood the victim, resulting in DoS.
Teardrop	An attacker sends a stream of overlapping fragments. The victim will crash when it tries to reassemble the overlapping fragments.
Ping of death	An attacker sends the victim an ICMP echo request larger than 65535 bytes that violates the IP protocol. When the victim reassembles the packet, a buffer overflow can occur, which causes a system crash.

Scanning attacks

Scanning is a preintrusion activity used to prepare for intrusion into a network. The scanning allows the attacker to find a way into the target network and to disguise the attacker's identity.

Attackers use scanning tools to probe a network, find vulnerable hosts, and discover services that are running on the hosts. Attackers can use the information to launch attacks.

The device can detect and prevent the IP sweep and port scan attacks. If an attacker performs port scanning from multiple hosts to the target host, distributed port scan attacks occur.

Flood attacks

An attacker launches a flood attack by sending a large number of forged requests to the victim in a short period of time. The victim is too busy responding to these forged requests to provide services for legal users, and a DoS attack occurs.

The device can detect and prevent the following types of flood attacks.

SYN flood attack

A SYN flood attacker exploits the TCP three-way handshake characteristics and makes the victim unresponsive to legal users. An attacker sends a large number of SYN packets with forged source addresses to a server. This causes the server to open a large number of half-open connections and respond to the requests. However, the server will never receive the expected ACK packets. The server is unable to accept new incoming connection requests because all of its resources are bound to half-open connections.

ACK flood attack

An ACK packet is a TCP packet only with the ACK flag set. Upon receiving an ACK packet from a client, the server must search half-open connections for a match.

An ACK flood attacker sends a large number of ACK packets to the server. This causes the server to be busy searching for half-open connections, and the server is unable to process packets for normal services.

SYN-ACK flood attack

Upon receiving a SYN-ACK packet, the server must search for the matching SYN packet it has sent. A SYN-ACK flood attacker sends a large number of SYN-ACK packets to the server. This causes the server to be busy searching for SYN packets, and the server is unable to process packets for normal services.

FIN flood attack

FIN packets are used to shut down TCP connections.

A FIN flood attacker sends a large number of forged FIN packets to a server. The victim might shut down correct connections, or be unable to provide services because it is busy searching for matching connections.

RST flood attack

RST packets are used to abort TCP connections when TCP connection errors occur.

An RST flood attacker sends a large number of forged RST packets to a server. The victim might shut down correct connections, or be unable to provide services because it is busy searching for matching connections.

DNS flood attack

The DNS server processes and replies all DNS queries that it receives.

A DNS flood attacker sends a large number of forged DNS queries. This attack consumes the bandwidth and resources of the DNS server, which prevents the server from processing and replying legal DNS queries.

HTTP flood attack

Upon receiving an HTTP GET request, the HTTP server performs complex operations, including character string searching, database traversal, data reassembly, and format switching. These operations consume a large amount of system resources.

An HTTP flood attacker sends a large number of HTTP GET requests that exceed the processing capacity of the HTTP server, which causes the server to crash.

ICMP flood attack

An ICMP flood attacker sends ICMP request packets, such as ping packets, to a host at a fast rate. Because the target host is busy replying to these requests, it is unable to provide services.

ICMPv6 flood attack

An ICMPv6 flood attacker sends ICMPv6 request packets, such as ping packets, to a host at a fast rate. Because the target host is busy replying to these requests, it is unable to provide services.

UDP flood attack

A UDP flood attacker sends UDP packets to a host at a fast rate. These packets consume a large amount of the target host's bandwidth, so the host cannot provide other services.

TCP fragment attack

An attacker launches TCP fragment attacks by sending attack TCP fragments defined in RFC 1858:

- First fragments in which the TCP header is smaller than 20 bytes.
- Non-first fragments with a fragment offset of 8 bytes (FO=1).

Typically, packet filter detects the source and destination IP addresses, source and destination ports, and transport layer protocol of the first fragment of a TCP packet. If the first fragment passes the detection, all subsequent fragments of the TCP packet are allowed to pass through.

Because the first fragment of attack TCP packets does not hit any match in the packet filter, the subsequent fragments can all pass through. After the receiving host reassembles the fragments, a TCP fragment attack occurs.

To prevent TCP fragment attacks, enable TCP fragment attack prevention to drop attack TCP fragments.

Login dictionary attack

The login dictionary attack is an automated process to attempt to log in by trying all possible passwords from a pre-arranged list of values (the dictionary). Multiple login attempts can occur in a short period of time.

You can configure the login delay feature to slow down the login dictionary attacks. This feature enables the device to delay accepting another login request after detecting a failed login attempt for a user.

Blacklist feature

IP blacklist

The IP blacklist feature is an attack prevention method that filters packets by source IP addresses in blacklist entries. Compared with ACL-based packet filtering, IP blacklist filtering is simpler and provides effective screening at a faster speed.

User blacklist

The user blacklist feature is an attack prevention method that filters packets by source users in blacklist entries. Compared with IP blacklist filtering, user blacklist filtering performs access control on the user level and makes the filtering easy to use.

The user blacklist feature must be used together with the user identification feature. User identification provides the mappings between usernames and IP addresses for the user blacklist.

Attack detection and prevention tasks at a glance

To configure attack detection and prevention, perform the following tasks:

1. Configuring and applying an attack defense policy
   a. Creating an attack defense policy
   b. Configuring an attack defense policy
      Choose the following tasks as needed:
      - Configuring a single-packet attack defense policy
      - Configuring a scanning attack defense policy
      - Configuring a flood attack defense policy
   c. (Optional.) Configuring attack detection exemption
   d. Applying an attack defense policy
      Choose the following tasks as needed:
      - Applying an attack defense policy to an interface
      - Applying an attack defense policy to the device
2. (Optional.) Enabling log non-aggregation for single-packet attack events
3. (Optional.) Configuring TCP fragment attack prevention
   Typically, this feature is separately used.
4. (Optional.) Configuring the blacklist feature
   Use this feature separately or jointly with a scanning attack defense policy.
   - Configuring the IP blacklist feature
   - Configuring the user blacklist feature
5. (Optional.) Configuring the login attack prevention feature
   Typically, this feature is separately used.
   - Configuring login attack prevention
   - Enabling the login delay

Configuring and applying an attack defense policy

Creating an attack defense policy

About this task

An attack defense policy contains a set of attack detection and prevention configuration.

To configure attack defense configuration such as detection signatures and protection actions, you must first create an attack defense policy and enter its view.

Restrictions and guidelines

CAUTION:

The default thresholds for triggering attack prevention might not be appropriate for your network. Set appropriate according to the actual application scenarios. Small thresholds might affect the Internet or webpage access speed. Large thresholds might make your network vulnerable to attacks.

Procedure

1. Enter system view.

   system-view

2. Create an attack defense policy and enter its view.

   attack-defense policy *policy-name*

Configuring a single-packet attack defense policy

About this task

Apply the single-packet attack defense policy to the interface that is connected to the external network.

Single-packet attack detection inspects incoming packets based on the packet signature. If an attack packet is detected, the device can take the following actions:

- Output logs (the default action).
- Drop attack packets.

You can also configure the device to not take any actions.

Procedure

1. Enter system view.

   system-view

2. Enter attack defense policy view.

   attack-defense policy *policy-name*

3. Configure signature detection for specific single-packet attack types, and specify the actions against the attacks.
   - Configure signature detection for well-known single-packet attacks, and specify the actions against the attacks.

     signature detect { **fraggle** | **fragment** | **impossible** | **land** | **large-icmp** | **large-icmpv6** | **smurf** | **snork** | **tcp-all-flags** | **tcp-fin-only** | **tcp-invalid-flags** | **tcp-null-flag** | **tcp-syn-fin** | **tiny-fragment** | **traceroute** | **udp-bomb** | **winnuke** } [**action** { { **drop** | **logging** } * | **none** }]

     signature detect { **ip-option-abnormal** | **ping-of-death** | **teardrop** } **action** { **drop** | **logging** } *

   - Configure signature detection for ICMP packet attacks, and specify the actions against the attacks.

     signature detect icmp-type { *icmp-type-value* | **address-mask-reply** | **address-mask-request** | **destination-unreachable** | **echo-reply** | **echo-request** | **information-reply** | **information-request** | **parameter-problem** | **redirect** | **source-quench** | **time-exceeded** | **timestamp-reply** | **timestamp-request** } [**action** { { **drop** | **logging** } * | **none** }]

- Configure signature detection for ICMPv6 packet attacks, and specify the actions against the attacks.

  signature detect icmpv6-type { *icmpv6-type-value* | **destination-unreachable** | **echo-reply** | **echo-request** | **group-query** | **group-reduction** | **group-report** | **packet-too-big** | **parameter-problem** | **time-exceeded** } [**action** { { **drop** | **logging** } * | **none** }]

- Configure signature detection for IP option attacks, and specify the actions against the attacks.

  signature detect ip-option { *option-code* | **internet-timestamp** | **loose-source-routing** | **record-route** | **route-alert** | **security** | **stream-id** | **strict-source-routing** } [**action** { { **drop** | **logging** } * | **none** }]

- Configure signature detection for abnormal IPv6 extension header attacks, and specify the actions against the attacks.

  signature detect ipv6-ext-header *ext-header-value* [**action** { { **drop** | **logging** } * | **none** }]

By default, signature detection is not configured for single-packet attacks.

4. (Optional.) Set the maximum length of safe ICMP or ICMPv6 packets.

   signature { **large-icmp** | **large-icmpv6** } **max-length** *length*

   By default, the maximum length of safe ICMP or ICMPv6 packets is 4000 bytes.

5. (Optional.) Specify the actions against single-packet attacks of a specific level.

   signature level { **high** | **info** | **low** | **medium** } **action** { { **drop** | **logging** } * | **none** }

   The default action is **logging** for single-packet attacks of the informational and low levels.

   The default actions are **logging** and **drop** for single-packet attacks of the medium and high levels.

6. (Optional.) Enable signature detection for single-packet attacks of a specific level.

   signature level { **high** | **info** | **low** | **medium** } **detect**

   By default, signature detection is disabled for all levels of single-packet attacks.

Configuring a scanning attack defense policy

About this task

Apply a scanning attack defense policy to the interface that is connected to the external network.

Scanning attack detection inspects the incoming packet rate of connections to the target system. If a source initiates connections at a rate equal to or exceeding the pre-defined threshold, the device can take the following actions:

- Output logs.
- Drop subsequent packets from the IP address of the attacker.

Procedure

1. Enter system view.

   system-view

2. Enter attack defense policy view.

   attack-defense policy *policy-name*

3. Configure scanning attack detection.

   scan detect level { **high** | **low** | **medium** } **action** { **drop** | **logging** } *

   By default, scanning attack detection is not configured.

Configuring a flood attack defense policy

About this task

Apply a flood attack defense policy to the interface that is connected to the external network to protect internal servers.

Flood attack detection monitors the rate at which connections are initiated to the internal servers.

With flood attack detection enabled, the device is in attack detection state. When the packet sending rate to an IP address reaches the threshold, the device enters prevention state and takes defensive actions, such as logging and dropping subsequent packets. When the rate is below the silence threshold (three-fourths of the threshold), the device returns to the attack detection state.

Restrictions and guidelines for flood attack detection and prevention

If a device has multiple service cards, the global trigger threshold you set takes effect on each service card. The global trigger threshold of the device is the product of multiplying the value you set by the service card quantity.

You can configure flood attack detection and prevention for a specific IP address. For non-specific IP addresses, the device uses the global attack prevention settings.

Configuring a SYN flood attack defense policy

1. Enter system view.

   system-view
2. Enter attack defense policy view.

   attack-defense policy *policy-name*
3. Enable global SYN flood attack detection.

   syn-flood detect non-specific

   By default, global SYN flood attack detection is disabled.
4. Set the global trigger threshold for SYN flood attack prevention.

   syn-flood threshold *threshold-value*

   The default setting is 1000.
5. Specify global actions against SYN flood attacks.

   syn-flood action { **drop** | **logging** } *

   By default, no global action is specified for SYN flood attacks.
6. Configure IP address-specific SYN flood attack detection.

   syn-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

   By default, IP address-specific SYN flood attack detection is not configured.

Configuring an ACK flood attack defense policy

1. Enter system view.

   system-view
2. Enter attack defense policy view.

   attack-defense policy *policy-name*
3. Enable global ACK flood attack detection.

   ack-flood detect non-specific

   By default, global ACK flood attack detection is disabled.
4. Set the global trigger threshold for ACK flood attack prevention.

ack-flood threshold *threshold-value*

The default setting is 1000.

5. Specify global actions against ACK flood attacks.

ack-flood action { **drop** | **logging** } *

By default, no global action is specified for ACK flood attacks.

6. Configure IP address-specific ACK flood attack detection.

ack-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

By default, IP address-specific ACK flood attack detection is not configured.

Configuring a SYN-ACK flood attack defense policy

1. Enter system view.

system-view

2. Enter attack defense policy view.

attack-defense policy *policy-name*

3. Enable global SYN-ACK flood attack detection.

syn-ack-flood detect non-specific

By default, global SYN-ACK flood attack detection is disabled.

4. Set the global trigger threshold for SYN-ACK flood attack prevention.

syn-ack-flood threshold *threshold-value*

The default setting is 1000.

5. Specify global actions against SYN-ACK flood attacks.

syn-ack-flood action { **drop** | **logging** }*

By default, no global action is specified for SYN-ACK flood attacks.

6. Configure IP address-specific SYN-ACK flood attack detection.

syn-ack-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

By default, IP address-specific SYN-ACK flood attack detection is not configured.

Configuring a FIN flood attack defense policy

1. Enter system view.

system-view

2. Enter attack defense policy view.

attack-defense policy *policy-name*

3. Enable global FIN flood attack detection.

fin-flood detect non-specific

By default, global FIN flood attack detection is disabled.

4. Set the global trigger threshold for FIN flood attack prevention.

fin-flood threshold *threshold-value*

The default setting is 1000.

5. Specify global actions against FIN flood attacks.

fin-flood action { **drop** | **logging** } *

By default, no global action is specified for FIN flood attacks.

6. Configure IP address-specific FIN flood attack detection.

fin-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

By default, IP address-specific FIN flood attack detection is not configured.

Configuring an RST flood attack defense policy

1. Enter system view.

   system-view

2. Enter attack defense policy view.

   attack-defense policy *policy-name*

3. Enable global RST flood attack detection.

   rst-flood detect non-specific

   By default, global RST flood attack detection is disabled.

4. Set the global trigger threshold for RST flood attack prevention.

   rst-flood threshold *threshold-value*

   The default setting is 1000.

5. Specify global actions against RST flood attacks.

   rst-flood action { **drop** | **logging** } *

   By default, no global action is specified for RST flood attacks.

6. Configure IP address-specific RST flood attack detection.

   rst-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

   By default, IP address-specific RST flood attack detection is not configured.

Configuring an ICMP flood attack defense policy

1. Enter system view.

   system-view

2. Enter attack defense policy view.

   attack-defense policy *policy-name*

3. Enable global ICMP flood attack detection.

   icmp-flood detect non-specific

   By default, global ICMP flood attack detection is disabled.

4. Set the global trigger threshold for ICMP flood attack prevention.

   icmp-flood threshold *threshold-value*

   The default setting is 1000.

5. Specify global actions against ICMP flood attacks.

   icmp-flood action { **drop** | **logging** } *

   By default, no global action is specified for ICMP flood attacks.

6. Configure IP address-specific ICMP flood attack detection.

   icmp-flood detect ip *ip-address* [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

   By default, IP address-specific ICMP flood attack detection is not configured.

Configuring an ICMPv6 flood attack defense policy

1. Enter system view.

   system-view

2. Enter attack defense policy view.
   attack-defense policy *policy-name*
3. Enable global ICMPv6 flood attack detection.
   icmpv6-flood detect non-specific
   By default, global ICMPv6 flood attack detection is disabled.
4. Set the global trigger threshold for ICMPv6 flood attack prevention.
   icmpv6-flood threshold *threshold-value*
   The default setting is 1000.
5. Specify global actions against ICMPv6 flood attacks.
   icmpv6-flood action { **drop** | **logging** } *
   By default, no global action is specified for ICMPv6 flood attacks.
6. Configure IP address-specific ICMPv6 flood attack detection.
   icmpv6-flood detect ipv6 *ipv6-address* [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]
   By default, IP address-specific ICMPv6 flood attack detection is not configured.

Configuring a UDP flood attack defense policy

1. Enter system view.
   system-view
2. Enter attack defense policy view.
   attack-defense policy *policy-name*
3. Enable global UDP flood attack detection.
   udp-flood detect non-specific
   By default, global UDP flood attack detection is disabled.
4. Set the global trigger threshold for UDP flood attack prevention.
   udp-flood threshold *threshold-value*
   The default setting is 1000.
5. Specify global actions against UDP flood attacks.
   udp-flood action { **drop** | **logging** } *
   By default, no global action is specified for UDP flood attacks.
6. Configure IP address-specific UDP flood attack detection.
   udp-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]
   By default, IP address-specific UDP flood attack detection is not configured.

Configuring a DNS flood attack defense policy

1. Enter system view.
   system-view
2. Enter attack defense policy view.
   attack-defense policy *policy-name*
3. Enable global DNS flood attack detection.
   dns-flood detect non-specific
   By default, global DNS flood attack detection is disabled.
4. Set the global trigger threshold for DNS flood attack prevention.

dns-flood threshold *threshold-value*

The default setting is 1000.

5. (Optional.) Specify the global ports to be protected against DNS flood attacks.

   dns-flood port *port-list*

   By default, DNS flood attack prevention protects port 53.

6. Specify global actions against DNS flood attacks.

   dns-flood action { **drop** | **logging** } *

   By default, no global action is specified for DNS flood attacks.

7. Configure IP address-specific DNS flood attack detection.

   dns-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**port** *port-list*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

   By default, IP address-specific DNS flood attack detection is not configured.

Configuring an HTTP flood attack defense policy

1. Enter system view.

   system-view

2. Enter attack defense policy view.

   attack-defense policy *policy-name*

3. Enable global HTTP flood attack detection.

   http-flood detect non-specific

   By default, global HTTP flood attack detection is disabled.

4. Set the global trigger threshold for HTTP flood attack prevention.

   http-flood threshold *threshold-value*

   The default setting is 1000.

5. (Optional.) Specify the global ports to be protected against HTTP flood attacks.

   http-flood port *port-list*

   By default, HTTP flood attack prevention protects port 80.

6. Specify global actions against HTTP flood attacks.

   http-flood action { **drop** | **logging** } *

   By default, no global action is specified for HTTP flood attacks.

7. Configure IP address-specific HTTP flood attack detection.

   http-flood detect { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**port** *port-list*] [**threshold** *threshold-value*] [**action** { { **drop** | **logging** } * | **none** }]

   By default, IP address-specific HTTP flood attack detection is not configured.

Configuring attack detection exemption

About this task

The attack defense policy uses the ACL to identify exempted packets. The policy does not check the packets permitted by the ACL. You can configure the ACL to identify packets from trusted servers. The exemption feature reduces the false alarm rate and improves packet processing efficiency. For example, the attack defense policy identifies multicast packets with the same source addresses and different destination addresses as scanning attack packets (for example, OSPF or PIM packets). You can configure an ACL to exempt such packets from attack detection.

Restrictions and guidelines

If an ACL is used for attack detection exemption, only the following match criteria in the ACL permit rules take effect:

- Source IP address.
- Destination IP address.
- Source port.
- Destination port.
- Protocol.
- L3VPN instance.
- The **fragment** keyword for matching non-first fragments.

Procedure

1. Enter system view.
   system-view
2. Enter attack defense policy view.
   attack-defense policy *policy-name*
3. Configure attack detection exemption.
   exempt acl [**ipv6**] { *acl-number* | **name** *acl-name* }
   By default, attack detection exemption is not configured.

Applying an attack defense policy to an interface

1. Enter system view.
   system-view
2. Enter system view.
   interface *interface-type interface-number*
3. Apply an attack defense policy to the interface.
   attack-defense apply policy *policy-name*
   By default, no attack defense policy is applied to the interface.

Applying an attack defense policy to the device

About this task

An attack defense policy applied to the device itself rather than the interfaces detects packets destined for the device and prevents attacks targeted at the device.

A switch uses hardware to implement packet forwarding and uses software to process packets if the packets are destined for the switch. The software does not provide any attack defense features, so you must apply an attack defense policy to the switch to prevent attacks aimed at the switch.

If a device and its interfaces have attack defense policies applied, a packet destined for the device is processed as follows:

1. The policy applied to the receiving interface processes the packet.
2. If the packet is not dropped by the receiving interface, the policy applied to the device processes the packet.

Procedure

1. Enter system view.

```
system-view
```

2. Apply an attack defense policy to the device.
   ```
   attack-defense local apply policy policy-name
   ```
   By default, no attack defense policy is applied to the device.

Enabling log non-aggregation for single-packet attack events

About this task

Log aggregation aggregates multiple logs generated during a period of time and sends one log. Logs that are aggregated must have the following attributes in common:

- Attacks are detected on the same interface or are destined for the device.
- Attack type.
- Attack defense action.
- Source and destination IP addresses.
- VPN instance to which the victim IP address belongs.

Restrictions and guidelines

As a best practice, do not disable log aggregation. A large number of logs will consume the display resources of the console.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable log non-aggregation for single-packet attack events.
   ```
   attack-defense signature log non-aggregate
   ```
   By default, log non-aggregation is disabled for single-packet attack events.

Configuring TCP fragment attack prevention

About this task

The TCP fragment attack prevention feature detects the length and fragment offset of received TCP fragments and drops attack TCP fragments.

Restrictions and guidelines

TCP fragment attack prevention takes precedence over single-packet attack prevention. When both are used, incoming TCP packets are processed first by TCP fragment attack prevention and then by the single-packet attack defense policy.

Procedure

1. Enter system view.
   ```
   system-view
   ```
2. Enable TCP fragment attack prevention.
   ```
   attack-defense tcp fragment enable
   ```
   By default, TCP fragment attack prevention is enabled.

Configuring the IP blacklist feature

About this task

The IP blacklist feature filters packets sourced from IP addresses in blacklist entries. If the global blacklist feature is enabled, the blacklist feature is enabled on all interfaces or security zones.

IP blacklist entries can be manually added or dynamically learned:

- You can manually add an IP blacklist entry by using the **blacklist ip** or **blacklist ipv6** command. These entries do not age out by default. You can set an aging time for each entry.
- The device can automatically add IP blacklist entries when collaborating with scanning attack detection. Each dynamically learned IP blacklist entry has an aging time, which is user configurable. Make sure the **block-source** keyword is specified as the scanning attack prevention action. For more information about the scanning attack detection and prevention, see "Configuring a scanning attack defense policy."

Restrictions and guidelines

If the global blacklist feature is enabled, the blacklist feature is enabled on all interfaces.

If the global blacklist feature is disabled, you can enable the blacklist feature on specific interfaces to meet your network requirements.

Procedure

1. Enter system view.

   system-view
2. (Optional.) Enable the global blacklist feature.

   blacklist global enable

   By default, the global blacklist feature is disabled.
3. (Optional.) Add an IP blacklist entry.
   - Add an IPv4 blacklist entry.

     blacklist ip *source-ip-address* [**vpn-instance** *vpn-instance-name*] [**ds-lite-peer** *ds-lite-peer-address*] [**timeout** *minutes*]
   - Add an IPv6 blacklist entry.

     blacklist ipv6 *source-ipv6-address* [**vpn-instance** *vpn-instance-name*] [**timeout** *minutes*]
4. (Optional.) Enable logging for the blacklist feature.

   blacklist logging enable

   By default, logging is disabled for the blacklist feature.
5. Enter interface view.

   interface *interface-type interface-number*
6. Enable the blacklist feature on the interface.

   blacklist enable

   By default, the blacklist feature is disabled on the interface.

Configuring the user blacklist feature

About this task

The user blacklist feature filters packets sourced from users in blacklist entries.

A user blacklist entry can only be manually added by using the **blacklist user** command. A user blacklist entry does not age out by default. You can set an aging time for the entry.

Procedure

1. Enter system view.
   system-view
2. Enable the global blacklist feature.
   blacklist global enable
   By default, the global blacklist feature is disabled.
3. Add a user blacklist entry.
   blacklist user *user-name* [**timeout** *minutes*]
4. (Optional.) Enable logging for the blacklist feature.
   blacklist logging enable
   By default, logging is disabled for the blacklist feature.

Configuring login attack prevention

About this task

The login attack prevention feature detects a login DoS attack if a user fails the maximum number of successive login attempts. The feature triggers the blacklist feature to add the user's IP to the blacklist. Following login attempts from the user is blocked for the block period. For login attack prevention to take effect, you must enable the global blacklist feature.

This feature can effectively prevent login DoS attacks.

Procedure

1. Enter system view.
   system-view
2. Enable login attack prevention.
   attack-defense login enable
   By default, login attack prevention is disabled.
3. Set the maximum number of successive login failures.
   attack-defense login max-attempt *max-attempt*
   The default value is three.
4. Set the block period during which a login attempt is blocked.
   attack-defense login block-timeout *minutes*
   The default value is 60 minutes.
5. Enable the global blacklist feature.
   blacklist global enable
   By default, the global blacklist feature is disabled.

Enabling the login delay

About this task

The login delay feature delays the device from accepting a login request from a user after the user fails a login attempt. This feature can slow down login dictionary attacks.

The login delay feature is independent of the login attack prevention feature.

Procedure

1. Enter system view.
   system-view
2. Enable the login delay feature.
   attack-defense login reauthentication-delay *seconds*
   By default, the login delay feature is disabled. The device does not delay accepting a login request from a user who has failed a login attempt.

Display and maintenance commands for attack detection and prevention

Use the **display** commands in any view and the **reset** commands in user view.

To display and maintain attack detection and prevention:

Task	Command
Display flood attack detection and prevention statistics for an IPv4 address.	**display attack-defense** { **ack-flood** | **dns-flood** | **fin-flood** | **flood** | **http-flood** | **icmp-flood** | **rst-flood** | **syn-ack-flood** | **syn-flood** | **udp-flood** } **statistics ip** [*ip-address* [**vpn** *vpn-instance-name*]] [[**interface** *interface-type interface-number* | **local**] [**slot** *slot-number*]] [**count**]

Task	Command
Display flood attack detection and prevention statistics for an IPv6 address.	**display attack-defense** { **ack-flood** | **dns-flood** | **fin-flood** | **flood** | **http-flood** | **icmpv6-flood** | **rst-flood** | **syn-flood** | **syn-ack-flood** | **udp-flood** } **statistics ipv6** [*ipv6-address* [**vpn** *vpn-instance-name*]] [[**interface** *interface-type interface-number* | **local**] [**slot** *slot-number*]] [**count**]
Display attack defense policy configuration.	**display attack-defense policy** [*policy-name*]

Task	Command
Display information about IPv4 addresses protected by flood attack detection and prevention.	**display attack-defense policy** *policy-name* { **ack-flood** | **dns-flood** | **fin-flood** | **flood** | **http-flood** | **icmp-flood** | **rst-flood** | **syn-ack-flood** | **syn-flood** | **udp-flood** } **ip** [*ip-address* [**vpn** *vpn-instance-name*]] [**slot** *slot-number*] [**count**]
Display information about IPv6 addresses protected by flood attack detection and prevention.	**display attack-defense policy** *policy-name* { **ack-flood** | **dns-flood** | **fin-flood** | **flood** | **http-flood** | **icmpv6-flood** | **rst-flood** | **syn-ack-flood** | **syn-flood** | **udp-flood** } **ipv6** [*ipv6-address* [**vpn** *vpn-instance-name*]] [**slot** *slot-number*] [**count**]

Task	Command
Display information about IPv4 scanning attackers.	**display attack-defense scan attacker ip** [[**interface** *interface-type interface-number* | **local**] [**slot** *slot-number*]] [**count**]
Display information about IPv6 scanning attackers.	**display attack-defense scan attacker ipv6** [[**interface** *interface-type interface-number* | **local**] [**slot** *slot-number*]] [**count**]
Display information about IPv4 scanning attack victims.	**display attack-defense scan victim ip** [[**interface** *interface-type interface-number* | **local**] [**slot** *slot-number*]] [**count**]
Display information about IPv6 scanning attack victims.	**display attack-defense scan victim ipv6** [[**interface** *interface-type interface-number* | **local**] [**slot** *slot-number*]] [**count**]

Task	Command
Display attack detection and prevention statistics on an interface.	**display attack-defense statistics interface** *interface-type interface-number* [**slot** *slot-number*]
Display attack detection and prevention statistics for the device.	**display attack-defense statistics local** [**slot** *slot-number*]
Display manually added IPv4 blacklist entries.	**display blacklist ip** [*source-ip-address* [**vpn-instance** *vpn-instance-name*] [**ds-lite-peer** *ds-lite-peer-address*] | **count**]
Display manually added IPv6 blacklist entries.	**display blacklist ipv6** [*source-ipv6-address* [**vpn-instance** *vpn-instance-name*] | **count**]
Clear flood attack detection and prevention statistics.	**reset attack-defense policy** *policy-name* **flood protected** { **ip** | **ipv6** } **statistics**
Clear attack detection and prevention statistics for an interface.	**reset attack-defense statistics interface** *interface-type interface-number*
Clear attack detection and prevention statistics for the device.	**reset attack-defense statistics local**
Clear dynamic IPv4 blacklist entries.	**reset blacklist ip** { *source-ip-address* [**vpn-instance** *vpn-instance-name*] [**ds-lite-peer** *ds-lite-peer-address*] | **all** }
Clear dynamic IPv6 blacklist entries.	**reset blacklist ipv6** { *source-ipv6-address* [**vpn-instance** *vpn-instance-name*] | **all** }
Clear blacklist statistics.	**reset blacklist statistics**

Attack detection and prevention configuration examples

Example: Configuring attack detection and prevention on the device

Network configuration

Configure attack detection and prevention on the switch (the gateway) to protect against network attacks from the user side or the network side.

- To prevent TCP flag attacks and low level scanning attacks that aim at the switch, enable TCP flag attack prevention and scanning attack prevention. Configure the device to output logs if it detects such attacks.
- To prevent the SYN flood attacks that aim at the external interface of the switch, enable IP address-specific SYN flood attack detection for 192.168.2.1/24. When the device receives 5000 or more SYN packets sent to the protected IP address per second, it outputs logs and drops the packets.
- To prevent the SYN flood attacks that aim at the internal interface of the switch, enable global SYN flood attack detection. When the device receives 2000 or more SYN packets that are destined to the switch but not to the protected IP address per second, it outputs logs.

Figure 1 Network diagram

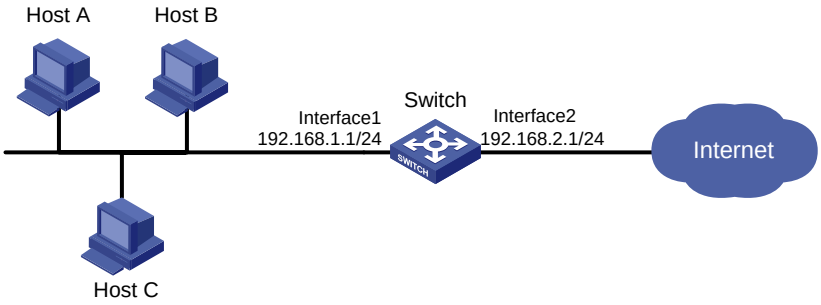


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

Assign IP addresses to interfaces according to the network diagram.

```
<Sysname> system-view
[Sysname] sysname Switch
[Switch] interface gigabitethernet 1/0/1
[Switch-GigabitEthernet1/0/1] ip address 192.168.1.1 255.255.255.0
[Switch-GigabitEthernet1/0/1] quit
[Switch] interface gigabitethernet 1/0/2
[Switch-GigabitEthernet1/0/2] ip address 192.168.2.1 255.255.255.0
[Switch-GigabitEthernet1/0/2] quit
```

Create attack defense policy **a1**.

```
[Switch] attack-defense policy a1
```

Enable signature detection for TCP single packet attacks and specify **logging** as the attack prevention action. A TCP packet is identified as an attack packet if it has all flags set, only FIN flag set, invalid flags, no TCP flags set, or both SYN and FIN flags set.

```
[Switch-attack-defense-policy-a1] signature detect tcp-all-flags action logging
[Switch-attack-defense-policy-a1] signature detect tcp-fin-only action logging
[Switch-attack-defense-policy-a1] signature detect tcp-invalid-flags action logging
[Switch-attack-defense-policy-a1] signature detect tcp-null-flag action logging
[Switch-attack-defense-policy-a1] signature detect tcp-syn-fin action logging
```

Enable low level scanning attack detection and specify **logging** as the attack prevention action.

```
[Switch-attack-defense-policy-a1] scan detect level low action logging
```

Enable SYN flood attack detection for 192.168.2.1. Set the threshold for triggering SYN flood attack prevention to 5000 and specify **logging** and **drop** as the attack prevention actions.

```
[Switch-attack-defense-policy-a1] syn-flood detect ip 192.168.2.1 threshold 5000 action
logging drop
```

Enable global SYN flood attack detection, set the global threshold for triggering SYN flood attack prevention to 2000, and specify **logging** as the global attack prevention action.

```
[Switch-attack-defense-policy-a1] syn-flood detect non-specific
[Switch-attack-defense-policy-a1] syn-flood threshold 2000
[Switch-attack-defense-policy-a1] syn-flood action logging
[Switch-attack-defense-policy-a1] quit
```

Apply the attack defense policy to the device.

```
[Switch] attack-defense local apply policy a1
```

Verifying the configuration

Verify that attack defense policy **a1** is correctly configured.

```
[Switch] display attack-defense policy a1
          Attack-defense Policy Information
------------------------------------------------------------------------------
Policy name                       : a1
Applied list                      : Local
------------------------------------------------------------------------------
Exempt IPv4 ACL                   : Not configured
Exempt IPv6 ACL                   : Not configured
------------------------------------------------------------------------------
  Actions: CV-Client verify  BS-Block source  L-Logging  D-Drop  N-None

Signature attack defense configuration:
Signature name                    Defense     Level            Actions
Fragment                          Disabled    low              L
Impossible                        Disabled    medium           L,D
Teardrop                          Disabled    medium           L,D
Tiny fragment                     Disabled    low              L
IP option abnormal                Disabled    medium           L,D
Smurf                             Enabled     medium           L,D
Traceroute                        Disabled    low              L
Ping of death                     Disabled    medium           L,D
Large ICMP                        Disabled    info             L
```

```
  Max length                      4000 bytes
Large ICMPv6                      Disabled     info            L
  Max length                      4000 bytes
TCP invalid flags                 Disabled     medium          L
TCP null flag                     Disabled     medium          L
TCP all flags                     Disabled     medium          L
TCP SYN-FIN flags                 Disabled     medium          L
TCP FIN only flag                 Disabled     medium          L
TCP Land                          Disabled     medium          L,D
Winnuke                           Disabled     medium          L,D
UDP Bomb                          Disabled     medium          L,D
UDP Snork                         Disabled     medium          L,D
UDP Fraggle                       Disabled     medium          L,D
IP option record route            Disabled     info            L
IP option internet timestamp      Disabled     info            L
IP option security                Disabled     info            L
IP option loose source routing    Disabled     info            L
IP option stream ID               Disabled     info            L
IP option strict source routing   Disabled     info            L
IP option route alert             Disabled     info            L
ICMP echo request                 Disabled     info            L
ICMP echo reply                   Disabled     info            L
ICMP source quench                Disabled     info            L
ICMP destination unreachable      Disabled     info            L
ICMP redirect                     Disabled     info            L
ICMP time exceeded                Disabled     info            L
ICMP parameter problem            Disabled     info            L
ICMP timestamp request            Disabled     info            L
ICMP timestamp reply              Disabled     info            L
ICMP information request          Disabled     info            L
ICMP information reply            Disabled     info            L
ICMP address mask request         Disabled     info            L
ICMP address mask reply           Disabled     info            L
ICMPv6 echo request               Disabled     info            L
ICMPv6 echo reply                 Disabled     info            L
ICMPv6 group membership query     Disabled     info            L
ICMPv6 group membership report    Disabled     info            L
ICMPv6 group membership reduction Disabled     info            L
ICMPv6 destination unreachable    Disabled     info            L
ICMPv6 time exceeded              Disabled     info            L
ICMPv6 parameter problem          Disabled     info            L
ICMPv6 packet too big             Disabled     info            L

Scan attack defense configuration:
 Defense : Enabled
 Level   : low
 Actions : L
```

```
Flood attack defense configuration:
Flood type      Global thres(pps)  Global actions  Service ports   Non-specific
SYN flood       2000               L               -               Enabled
ACK flood       1000(default)      -               -               Disabled
SYN-ACK flood   1000(default)      -               -               Disabled
RST flood       1000(default)      -               -               Disabled
FIN flood       1000(default)      -               -               Disabled
UDP flood       1000(default)      -               -               Disabled
ICMP flood      1000(default)      -               -               Disabled
ICMPv6 flood    1000(default)      -               -               Disabled
DNS flood       1000(default)      -               53              Disabled
HTTP flood      1000(default)      -               80              Disabled

Flood attack defense for protected IP addresses:
 Address                VPN instance Flood type    Thres(pps) Actions Ports
 192.168.2.1            --           SYN-FLOOD     5000       L,D     -
```

If the device receives TCP flag attack packets or scanning attack packets that are destined for the device, the device outputs logs. If the device receives TCP SYN flood attack packets that are destined for the protected IP address, the device outputs logs and drops the attack packets. If the device receives TCP SYN flood attack packets that are destined for the device but not to the protected IP address, the device outputs logs.

Display the attack detection and prevention statistics.

```
[Swtich] display attack-defense statistics local
Attack policy name: a1
Slot 1:
Scan attack defense statistics:
 AttackType                      AttackTimes Dropped
 Port scan                       4           0
Flood attack defense statistics:
 AttackType                      AttackTimes Dropped
 No flood attacks detected.
Signature attack defense statistics:
 AttackType                      AttackTimes Dropped
 TCP invalid flags               116         0
 TCP null flag                   709         0
 TCP all flags                   251         0
 TCP SYN-FIN flags               46          0
 TCP FIN only flag               130         0
```

Configuration files

```
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.255.0
#
interface GigabitEthernet1/0/2
 port link-mode route
 ip address 192.168.2.1 255.255.255.0
#
```

```
attack-defense local apply policy a1
#
attack-defense policy a1
 syn-flood detect ip 192.168.2.1 threshold 5000 action logging drop
 scan detect level low action logging
 signature detect tcp-all-flags action logging
 signature detect tcp-fin-only action logging
 signature detect tcp-invalid-flags action logging
 signature detect tcp-null-flag action logging
 signature detect tcp-syn-fin action logging
 syn-flood detect non-specific
 syn-flood threshold 2000
 syn-flood action logging
#
```

Example: Configuring IP blacklist

Network configuration

As shown in Figure 2, configure the IP blacklist feature on the router to block packets from the attacker Host D permanently and from Host C for 50 minutes.

Figure 2 Network diagram

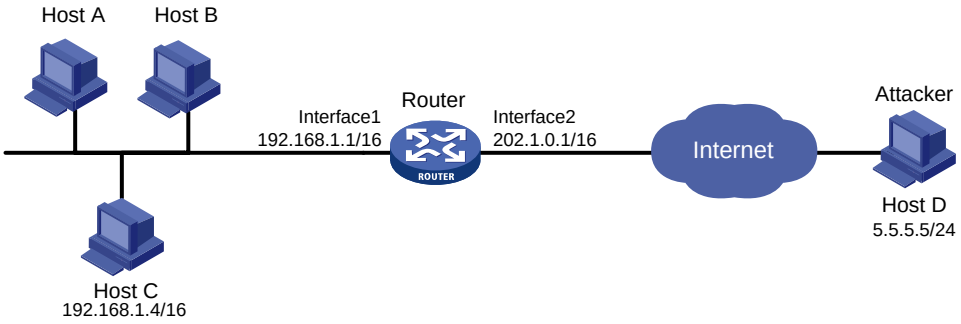


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

Assign IP addresses to interfaces according to the network diagram.

```
<Sysname> system-view
[Sysname] sysname Router
[Router] interface gigabitethernet 1/0/1
[Router-GigabitEthernet1/0/1] ip address 192.168.1.1 255.255.0.0
[Router-GigabitEthernet1/0/1] quit
[Router] interface gigabitethernet 1/0/2
[Router-GigabitEthernet1/0/2] ip address 202.1.0.1 255.255.0.0
[Router-GigabitEthernet1/0/2] quit
```

Enable the global blacklist feature.

```
<Router> system-view
[Router] blacklist global enable
```

Add an IPv4 blacklist entry for Host D.

```
[Router] blacklist ip 5.5.5.5
```

Add an IPv4 blacklist entry for Host C and set the blacklist entry aging time to 50 minutes.

```
[Router] blacklist ip 192.168.1.4 timeout 50
```

Verifying the configuration

Verify that the IPv4 blacklist entries are successfully added.

```
<Router> display blacklist ip
IP address      VPN instance   DS-Lite tunnel peer  Type    TTL(sec) Dropped
5.5.5.5         --             --                   Manual  Never    0
192.168.1.4     --             --                   Manual  2989     0
```

Verify that the router drops packets from Host D. (Details not shown.)

Execute the **`undo blacklist ip 5.5.5.5`** command and verify that the router forwards packets from Host D. (Details not shown.)

Verify that the router drops packets from Host C for 50 minutes and forwards packets from Host C after 50 minutes. (Details not shown.)

Configuration files

```
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.0.0
#
interface GigabitEthernet1/0/2
 port link-mode route
 ip address 202.1.0.1 255.255.0.0
#
 blacklist ip 5.5.5.5
 blacklist global enable
#
```

Example: Configuring user blacklist

Network configuration

As shown in Figure 3, configure the user blacklist feature on the router to block packets from User C for 50 minutes. The IP address of User C is 1.2.3.4 and the MAC address of User C is 0001-0001-0001.

Figure 3 Network diagram

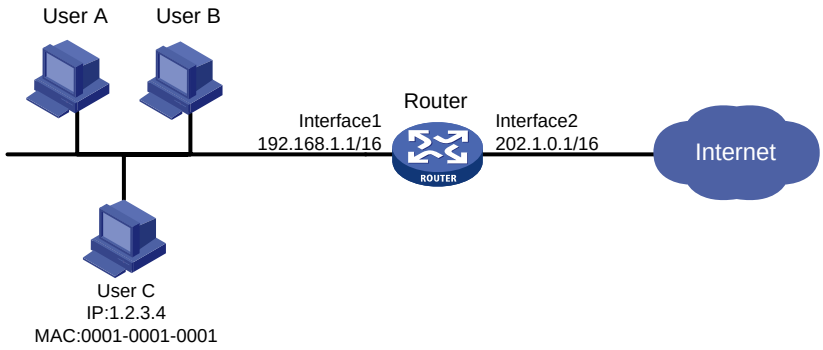


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Assign IP addresses to interfaces according to the network diagram.

```
<Sysname> system-view
[Sysname] sysname Router
[Router] interface gigabitethernet 1/0/1
[Router-GigabitEthernet1/0/1] ip address 192.168.1.1 255.255.0.0
[Router-GigabitEthernet1/0/1] quit
[Router] interface gigabitethernet 1/0/2
[Router-GigabitEthernet1/0/2] ip address 202.1.0.1 255.255.0.0
[Router-GigabitEthernet1/0/2] quit
```

2. Configure user identification:

   # Add a network access user named **userc**.

```
<Router> system-view
[Router] local-user userc class network
[Router-luser-network-userc] quit
```

   # Configure a static identity user with username **userc**, IP address 1.2.3.4, and MAC address 0001-0001-0001.

```
[Router] user-identity static-user userc bind ipv4 1.2.3.4 mac 0001-0001-0001
```

   # Enable user identification.

```
[Router] user-identity enable
```

3. Configure the user blacklist feature:

   # Enable the global blacklist feature.

```
[Router] blacklist global enable
```

   # Add a user blacklist entry for user **userc** and set the blacklist entry aging time to 50 minutes.

```
[Router] blacklist user userc timeout 50
```

Verifying the configuration

Verify that the user blacklist entry is successfully added.

```
[Router] display blacklist user
User name      Type    TTL(sec) Dropped
```

```
userc          Manual  2987     0
```

Verify that the router drops packets from User C for 50 minutes and forwards packets from User C after 50 minutes. (Details not shown.)

Configuration files

```
#
interface GigabitEthernet1/0/1
 port link-mode route
 ip address 192.168.1.1 255.255.0.0
#
interface GigabitEthernet1/0/2
 port link-mode route
 ip address 202.1.0.1 255.255.0.0
#
local-user userc class network
 authorization-attribute user-role network-operator
#
 blacklist global enable
#
 user-identity enable
 user-identity static-user userc bind ipv4 1.2.3.4 mac 0001-0001-0001
#
```

Contents

Configuring TCP attack prevention 1
About TCP attack prevention 1
Configuring Naptha attack prevention 1

Configuring TCP attack prevention

About TCP attack prevention

TCP attack prevention can detect and prevent attacks that exploit the TCP connection establishment process.

Configuring Naptha attack prevention

About this task

Naptha is a DDoS attack that targets operating systems. It exploits the resources consuming vulnerability in TCP/IP stack and network application process. The attacker establishes a large number of TCP connections in a short period of time and leaves them in certain states without requesting any data. These TCP connections starve the victim of system resources, resulting in a system breakdown.

After you enable Naptha attack prevention, the device periodically checks the number of TCP connections in each state (CLOSING, ESTABLISHED, FIN_WAIT_1, FIN_WAIT_2, and LAST_ACK). If the number of TCP connections in a state exceeds the limit, the device will accelerate the aging of the TCP connections in that state to mitigate the Naptha attack.

Procedure

1. Enter system view.

   system-view

2. Enable Naptha attack prevention.

   tcp anti-naptha enable

   By default, Naptha attack prevention is disabled.

3. (Optional.) Set the maximum number of TCP connections in a state.

   tcp state { **closing** | **established** | **fin-wait-1** | **fin-wait-2** | **last-ack** } **connection-limit** *number*

   By default, the maximum number of TCP connections in each state (CLOSING, ESTABLISHED, FIN_WAIT_1, FIN_WAIT_2, and LAST_ACK) is 50.

   To disable the device from accelerating the aging of the TCP connections in a state, set the value to 0.

4. (Optional.) Set the interval for checking the number of TCP connections in each state.

   tcp check-state interval *interval*

   By default, the interval for checking the number of TCP connections in each state is 30 seconds.

Contents

Configuring IP source guard 1

About IPSG 1
IPSG operating mechanism 1
Static IPSG bindings 1
Dynamic IPSG bindings 2
IPSG bindings synchronized by routing protocols 3
IPSG tasks at a glance 3
Configuring the IPv4SG feature 3
Enabling IPv4SG 3
Configuring a static IPv4SG binding 4
Excluding IPv4 packets from IPSG filtering 5
Enabling IPv4SG alarming 5
Configuring the IPv6SG feature 5
Enabling IPv6SG 5
Configuring a static IPv6SG binding 6
Setting the maximum number of IPv6SG bindings on an interface 7
Enabling IPv6SG alarming 7
Enabling SNMP notifications for IPSG 8
Display and maintenance commands for IPSG 8
IPSG configuration examples 9
Example: Configuring static IPv4SG 9
Example: Configuring DHCP snooping-based dynamic IPv4SG 10
Example: Configuring DHCP relay agent-based dynamic IPv4SG 12
Example: Configuring static IPv6SG 13
Example: Configuring DHCPv6 snooping-based dynamic IPv6SG address bindings 13
Example: Configuring DHCPv6 snooping-based dynamic IPv6SG prefix bindings 14
Example: Configuring DHCPv6 relay agent-based dynamic IPv6SG 16

Configuring IP source guard

About IPSG

IP source guard (IPSG) prevents spoofing attacks by using an IPSG binding table to filter out illegitimate packets. This feature is typically configured on user-side interfaces.

IPSG operating mechanism

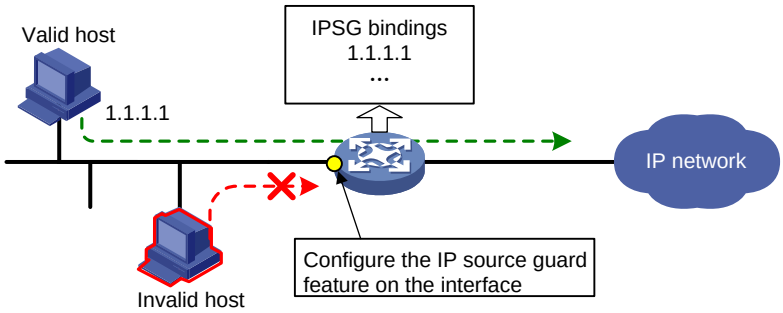
The IPSG binding table contains bindings that bind IP address, MAC address, VLAN, or any combinations. IPSG uses the bindings to match an incoming packet. If a match is found, the packet is forwarded. If no match is found, the packet is discarded.

IPSG is a per-interface packet filter. Configuring this feature on one interface does not affect packet forwarding on another interface.

IPSG bindings can be static or dynamic.

As shown in Figure 1, IPSG forwards only the packets that match an IPSG binding.

Figure 1 IPSG application



Static IPSG bindings

Static IPSG bindings are configured manually. They are suitable for scenarios where few hosts exist on a LAN and their IP addresses are manually configured. For example, you can configure a static IPSG binding on an interface that connects to a server. This binding allows the interface to receive packets only from the server.

Static IPSG bindings on an interface implement the following functions:

- Filter incoming IPv4 or IPv6 packets on the interface.
- Cooperate with ARP attack detection in IPv4 for user validity checking. For information about ARP attack detection, see "Configuring ARP attack protection."
- Cooperate with ND attack detection in IPv6 for user validity checking. For information about ND attack detection, see "Configuring ND attack defense."

Static IPSG bindings can be global or interface-specific.

- **Global static binding**—Binds the IP address and MAC address in system view. The binding takes effect on all interfaces to filter packets for user spoofing attack prevention.
- **Interface-specific static binding**—Binds the IP address, MAC address, VLAN, or any combination of the items in interface view. The binding takes effect only on the interface to check the validity of users who are attempting to access the interface.

Dynamic IPSG bindings

IPSG automatically obtains user information from other modules to generate dynamic bindings. A dynamic IPSG binding can contain MAC address, IPv4 or IPv6 address, VLAN tag, ingress interface, and binding type. The binding type identifies the source module for the binding, such as DHCP snooping, DHCPv6 snooping, DHCP relay agent, or DHCPv6 relay agent.

For example, DHCP-based IPSG bindings are suitable for scenarios where hosts on a LAN obtain IP addresses through DHCP. IPSG is configured on the DHCP server, the DHCP snooping device, or the DHCP relay agent. It generates dynamic bindings based on the client bindings on the DHCP server, the DHCP snooping entries, or the DHCP relay entries. IPSG allows only packets from the DHCP clients to pass through.

Dynamic IPv4SG

Dynamic bindings generated based on different source modules are for different usages:

Interface types	Source modules	Binding usage
Layer 2 Ethernet interface	DHCP snooping 802.1X	Packet filtering.
	ARP snooping	For cooperation with modules (such as the MFF module) to provide security services.
Layer 3 Ethernet interface VLAN interface	DHCP relay agent	Packet filtering.
	DHCP server	For cooperation with modules (such as the authorized ARP module) to provide security services.

For more information about 802.1X, see "Configuring 802.1X." For information about ARP snooping, DHCP snooping, DHCP relay, and DHCP server, see *Layer 3—IP Services Configuration Guide*.

Dynamic IPv6SG

Dynamic IPv6SG bindings generated based on different source modules are for different usages:

Interface types	Source modules	Binding usage
Layer 2 Ethernet interface	DHCPv6 snooping ND snooping 802.1X	Packet filtering.
Layer 3 Ethernet interface VLAN interface	DHCPv6 relay agent ND RA prefix entry recording	Packet filtering.
	DHCPv6 server	Reporting bindings to the controller to provide online and offline user information.

For more information about DHCPv6 snooping, see *Layer 3—IP Services Configuration Guide*. For more information about ND snooping, see IPv6 basics configuration in *Layer 3—IP Services Configuration Guide*. For more information about ND RA prefix entry recording, see IPv6 neighbor discovery in *Layer 3—IP Services Configuration Guide.* For more information about DHCPv6 relay agent, see *Layer 3—IP Services Configuration Guide*. For more information about DHCPv6 server, see *Layer 3—IP Services Configuration Guide.*

IPSG bindings synchronized by routing protocols

In addition to static and dynamic IPSG bindings, the device also supports IPSG bindings synchronized by routing protocols (such as BGP) from remote devices. These bindings are called remote IPSG bindings. By default, the remote IPSG bindings do not have interface information and cannot be used for packet filtering. When users roam from remote devices to the local device, the device learns interface information for the remote IPSG bindings through ARP or ND entries of the users. Then, the device converts the remote IPSG bindings into local IPSG bindings and uses these bindings for packet filtering.

For more information about BGP, see BGP overview in *Layer 3—IP Routing Configuration Guide*.

IPSG tasks at a glance

To configure IPSG, perform the following tasks:

1. To configure IPv4SG, perform the following tasks:
   a. Enabling IPv4SG
   b. (Optional.) Configuring a static IPv4SG binding
   c. (Optional.) Excluding IPv4 packets from IPSG filtering
   d. (Optional.) Enabling IPv4SG alarming
2. To configure IPv6SG, perform the following tasks:
   a. Enabling IPv6SG
   b. (Optional.) Configuring a static IPv6SG binding
   c. (Optional.) Enabling IPv6SG alarming
3. Enabling SNMP notifications for IPSG

Configuring the IPv4SG feature

Enabling IPv4SG

About this task

When you enable IPSG on an interface, the static and dynamic IPSG are both enabled.

- Static IPv4SG uses static bindings configured by using the **ip source binding** command. For more information, see "Configuring a static IPv4SG binding."
- Dynamic IPv4SG generates dynamic bindings from related source modules. IPv4SG uses the bindings to filter incoming IPv4 packets based on the matching criteria specified in the **ip verify source** command.

Restrictions and guidelines

To implement dynamic IPv4SG, make sure 802.1X, ARP snooping, DHCP snooping, DHCP relay agent, or DHCP server operates correctly on the network.

Enabling IPv4SG on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   The following interface types are supported:

- Layer 2 Ethernet interface.
- Layer 3 Ethernet interface.
- VLAN interface.

3. Enable the IPv4SG feature.

   ip verify source { **ip-address** | **ip-address** **mac-address** | **mac-address** }

   By default, the IPv4SG feature is disabled on an interface.

Configuring a static IPv4SG binding

About this task

You can configure global static and interface-specific static IPv4SG bindings. Interface-specific static and dynamic bindings take priority over global static bindings. An interface first uses the static and dynamic bindings on the interface to match packets. If no match is found, the interface uses the global bindings.

Restrictions and guidelines

Global static bindings take effect on all interfaces on the device.

To configure a static IPv4SG binding for the ARP attack detection feature, make sure the following conditions are met:

- The **ip-address** *ip-address* option, the **mac-address** *mac-address* option, and the **vlan** *vlan-id* option must be specified.
- ARP attack detection must be enabled for the specified VLAN.

Configuring a global static IPv4SG binding

1. Enter system view.

   system-view

2. Configure a global static IPv4SG binding.

   ip source binding ip-address *ip-address* **mac-address** *mac-address*

Configuring a static IPv4SG binding on an interface

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

   The following interface types are supported:

   - Layer 2 Ethernet interface.
   - Layer 3 Ethernet interface.
   - VLAN interface.

3. Configure a static IPv4SG binding.

   ip source binding { **ip-address** *ip-address* | **ip-address** *ip-address* **mac-address** *mac-address* | **mac-address** *mac-address* } [**vlan** *vlan-id*]

   You can configure the same static IPv4SG binding on different interfaces.

Excluding IPv4 packets from IPSG filtering

About this task

By default, IPv4SG processes all incoming IPv4 packets on an interface and discards the packets that do not match IPSG bindings. To allow specific IPv4 packets that do not match any IPSG binding to pass through the interface, you can specify source items of the packets for IPSG filtering exemption. All IPv4 packets with the specified source items are forwarded without being processed by IPSG.

Procedure

1. Enter system view.

   system-view

2. Exclude IPv4 packets with the specified source items from IPSG filtering.

   ip verify source exclude vlan *start-vlan-id* [to *end-vlan-id*]

   By default, no excluded source items are configured.

   You can execute this command multiple times to specify multiple excluded VLANs. The specified excluded VLANs cannot overlap.

Enabling IPv4SG alarming

About this task

This feature monitors the number of packets dropped by IPv4SG per second. When the packet dropping rate reaches or exceeds the alarm threshold, the device generates an alarm log. When the packet dropping rate drops below the alarm threshold, the device generates an alarm-cleared log.

Restrictions and guidelines

Do not enable this feature in a VLAN and an Ethernet interface belonging to the VLAN at the same time.

Procedure

1. Enter system view.

   system-view

2. Enter interface view or VLAN view.

   - Enter interface view.

     interface *interface-type interface-number*

   - Enter VLAN view.

     vlan *vlan-id*

3. Enable IPv4SG alarming on the interface.

   ip verify source alarm [*alarm-threshold*]

   By default, IPv4SG alarming is disabled on an interface.

Configuring the IPv6SG feature

Enabling IPv6SG

About this task

When you enable IPv6SG on an interface, the static and dynamic IPv6SG are both enabled.

- Static IPv6SG uses static bindings configured by using the **ipv6 source binding** command. For more information, see "Configuring a static IPv6SG binding."
- Dynamic IPv6SG generates dynamic bindings from related source modules. IPv6SG uses the bindings to filter incoming IPv6 packets based on the matching criteria specified in the **ipv6 verify source** command.

Restrictions and guidelines

To implement dynamic IPv6SG, make sure DHCPv6 snooping, DHCPv6 relay agent, or ND snooping operates correctly on the network.

Enabling IPv6SG on an interface

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*

   The following interface types are supported:

   - Layer 2 Ethernet interface.
   - Layer 3 Ethernet interface.
   - VLAN interface.
3. Enable the IPv6SG feature.

   ipv6 verify source { **ip-address** | **ip-address mac-address** | **mac-address** }

   By default, the IPv6SG feature is disabled on an interface.

Configuring a static IPv6SG binding

About this task

You can configure global static and interface-specific static IPv6SG bindings. Interface-specific static and dynamic bindings take priority over global static bindings. An interface first uses the static and dynamic bindings on the interface to match packets. If no match is found, the interface uses the global bindings.

Restrictions and guidelines

Global static bindings take effect on all interfaces on the device.

To configure a static IPv6SG binding for the ND attack detection feature, the **vlan** *vlan-id* option must be specified, and ND attack detection must be enabled for the specified VLAN.

If you configure both a static IPv6SG binding and VLAN mapping on a Layer 2 Ethernet interface, the interface forwards packets as follows:

- If you configure VLAN mapping and then configure a static IPv6SG binding and specify the original VLAN for the binding, the interface forwards packets based on the original VLAN.
- If you configure a static IPv6SG binding and specify a VLAN for the binding and then configure VLAN mapping, the interface forwards packets without considering the VLAN.

Configuring a global static IPv6SG binding

1. Enter system view.

   system-view
2. Configure a global static IPv6SG binding.

   ipv6 source binding ip-address *ipv6-address* **mac-address** *mac-address*

Configuring a static IPv6SG binding on an interface

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   The following interface types are supported:
   - Layer 2 Ethernet interface.
   - Layer 3 Ethernet interface.
   - VLAN interface.
3. Configure a static IPv6SG binding.
   ipv6 source binding { **ip-address** *ipv6-address* | **ip-address** *ipv6-address* **mac-address** *mac-address* | **mac-address** *mac-address* } [**vlan** *vlan-id*]
   You can configure the same static IPv6SG binding on different interfaces.

Setting the maximum number of IPv6SG bindings on an interface

About this task

You can set the maximum number of IPv6SG bindings on an interface to limit the total number of static and dynamic IPv6SG bindings on the interface. If the upper limit is reached, no more IPv6SG bindings can be added on the interface.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view.
   interface *interface-type interface-number*
3. Set the maximum number of IPv6SG bindings on the interface.
   ipv6 verify source max-entries *number*
   By default, the maximum number of IPv6SG bindings is not limited on an interface.

Enabling IPv6SG alarming

About this task

This feature enables the device to monitor the number of packets dropped by IPv6SG per second. When the packet dropping rate reaches or exceeds the alarm threshold, the device generates an alarm log. When the packet dropping rate drops below the alarm threshold, the device generates an alarm-cleared log.

Restrictions and guidelines

Do not enable this feature in a VLAN and an Ethernet interface belonging to the VLAN at the same time.

Procedure

1. Enter system view.
   system-view
2. Enter interface view or VLAN view.

- Enter interface view.
  interface *interface-type interface-number*
- *Enter VLAN view.*
  vlan *vlan-id*

3. Enable IPv6SG alarming on the interface.
   ipv6 verify source alarm [*alarm-threshold*]
   By default, IPv6SG alarming is disabled on an interface.

Enabling SNMP notifications for IPSG

About this task

To report critical IPSG events to an NMS, enable SNMP notifications for IPSG. For example, If you enable the device to generate notifications when the number of packets dropped by IPSG per second on an interface has reached, exceeded, or dropped below the alarm threshold, the device works as follows:

- When the number of packets dropped by IPSG per second on an interface is equal to or greater than the IPv4SG or IPv6SG alarm threshold, the device sends a notification.
- When the number of packets dropped by IPSG per second on an interface drops below the IPv4SG or IPv6SG alarm threshold, the device reports a threshold recovery event.

To set the IPv4 or IPv6SG alarm threshold, execute the **ip verify source alarm** or **ipv6 verify source alarm** command, respectively.

For SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP and information center configuration, see *Network Management and Monitoring Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for IPSG.
   snmp-agent trap enable ipsg [**drop-threshold**]
   By default, SNMP notifications for IPSG are enabled.

Display and maintenance commands for IPSG

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display IPv4SG bindings.	**display ip source binding** [**static** | [**vpn-instance** *vpn-instance-name*] [**arp-snooping-vlan** | **dhcp-relay** | **dhcp-server** | **dhcp-snooping** | **dot1x** | **remote**]] [**ip-address** *ip-address*] [**mac-address** *mac-address*] [**vlan** *vlan-id*] [**interface** *interface-type interface-number*] [**slot** *slot-number*]
Display local IPv4SG bindings that can be synchronized by routing protocols.	**display ip source binding-local** [**interface** *interface-type interface-number*] [**dhcp-relay**] [**ip-address** *ip-address*] [**mac-address** *mac-address*] [**vlan** *vlan-id*] [**slot** *slot-number*]

Task	Command
Display remote IPv4SG bindings synchronized by routing protocols.	**display ip source binding-remote** [**router-id** *router-id*] [**dhcp-relay**] [**ip-address** *ip-address*] [**mac-address** *mac-address*] [**vlan** *vlan-id*] [**slot** *slot-number*]
Display statistics about local and remote IPv4SG bindings that routing protocols synchronize.	**display ip source binding statistics**
Display source items that have been configured to be excluded from IPSG filtering.	**display ip verify source excluded** [**vlan** *start-vlan-id* [**to** *end-vlan-id*]] [**slot** *slot-number*]
Display IPv6SG address bindings.	**display ipv6 source binding** [**static** | [**vpn-instance** *vpn-instance-name*] [**dhcpv6-relay** | **dhcpv6-server** | **dhcpv6-snooping** | **dot1x** | **nd-snooping-vlan** | **remote**]] [**ip-address** *ipv6-address*] [**mac-address** *mac-address*] [**vlan** *vlan-id*] [**interface** *interface-type interface-number*] [**slot** *slot-number*]
Display local IPv6SG bindings that can be synchronized by routing protocols.	**display ipv6 source binding-local** [**interface** *interface-type interface-number*] [**dhcpv6-relay** | **nd-snooping-vlan**] [**ip-address** *ipv6-address*] [**mac-address** *mac-address*] [**vlan** *vlan-id*] [**slot** *slot-number*]
Display remote IPv6SG bindings synchronized by routing protocols.	**display ipv6 source binding-remote** [**router-id** *router-id*] [**dhcpv6-relay** | **nd-snooping-vlan**] [**ip-address** *ipv6-address*] [**mac-address** *mac-address*] [**vlan** *vlan-id*] [**slot** *slot-number*]
Display IPv6SG prefix bindings.	**display ipv6 source binding pd** [**vpn-instance** *vpn-instance-name*] [**prefix** *prefix/prefix-length*] [**mac-address** *mac-address*] [**vlan** *vlan-id*] [**interface** *interface-type interface-number*] [**slot** *slot-number*]
Display statistics about local and remote IPv6SG bindings that routing protocols synchronize.	**display ipv6 source binding statistics**

IPSG configuration examples

Example: Configuring static IPv4SG

Network configuration

As shown in Figure 2, all hosts use static IP addresses.

Configure static IPv4SG bindings on Device A and Device B to meet the following requirements:

- All interfaces of Device A allow IP packets from Host A to pass.
- GigabitEthernet 1/0/1 of Device A allows IP packets from Host B to pass.

Figure 2 Network diagram

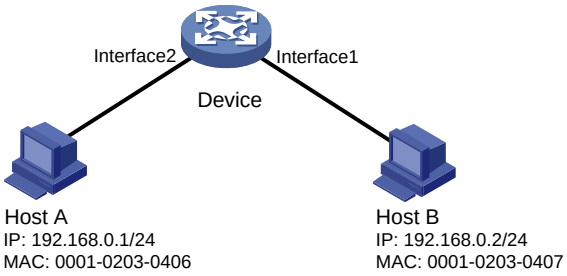


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

Configure IP addresses for the interfaces. (Details not shown.)

Enable IPv4SG on GigabitEthernet 1/0/2.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] ip verify source ip-address mac-address
[DeviceA-GigabitEthernet1/0/2] quit
```

Configure a static IPv4SG binding for Host A.

```
[DeviceA] ip source binding ip-address 192.168.0.1 mac-address 0001-0203-0406
```

Enable IPv4SG on GigabitEthernet 1/0/1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] ip verify source ip-address mac-address
```

On GigabitEthernet 1/0/1, configure a static IPv4SG binding for Host B.

```
[DeviceA-GigabitEthernet1/0/1] ip source binding mac-address 0001-0203-0407
[DeviceA-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Verify that the static IPv4SG bindings are configured successfully on Device A.

```
<DeviceA> display ip source binding static
Total entries found: 2
IP address      MAC address    Interface              VLAN Type
192.168.0.1     0001-0203-0406 N/A                    N/A  Static
N/A             0001-0203-0407 GE1/0/1                N/A  Static
```

Example: Configuring DHCP snooping-based dynamic IPv4SG

Network configuration

As shown in Figure 3, the host (the DHCP client) obtains an IP address from the DHCP server. Perform the following tasks:

- Enable DHCP snooping on the device to make sure the DHCP client obtains an IP address from the authorized DHCP server. To generate a DHCP snooping entry for the DHCP client, enable recording of client information in DHCP snooping entries.
- Enable dynamic IPv4SG on GigabitEthernet 1/0/1 to filter incoming packets by using the IPv4SG bindings generated based on DHCP snooping entries. Only packets from the DHCP client are allowed to pass.

Figure 3 Network diagram

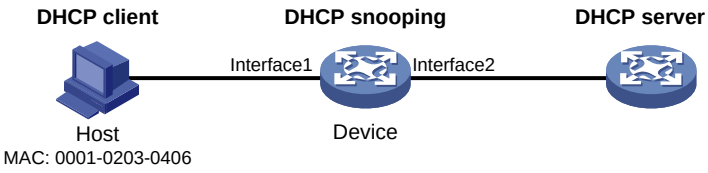


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure the DHCP server.

   For information about DHCP server configuration, see *Layer 3—IP Services Configuration Guide*.

2. Configure the device:

   # Configure IP addresses for the interfaces. (Details not shown.)

   # Enable DHCP snooping.

   ```
   <Device> system-view
   [Device] dhcp snooping enable
   ```

   # Configure GigabitEthernet 1/0/2 as a trusted interface.

   ```
   [Device] interface gigabitethernet 1/0/2
   [Device-GigabitEthernet1/0/2] dhcp snooping trust
   [Device-GigabitEthernet1/0/2] quit
   ```

   # Enable IPv4SG on GigabitEthernet 1/0/1 and verify the source IP address and MAC address for dynamic IPSG.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] ip verify source ip-address mac-address
   ```

   # Enable recording of client information in DHCP snooping entries on GigabitEthernet 1/0/1.

   ```
   [Device-GigabitEthernet1/0/1] dhcp snooping binding record
   [Device-GigabitEthernet1/0/1] quit
   ```

Verifying the configuration

Display dynamic IPv4SG bindings generated based on DHCP snooping entries.

```
[Device] display ip source binding dhcp-snooping
Total entries found: 1
IP address      MAC address    Interface               VLAN Type
192.168.0.1     0001-0203-0406 GE1/0/1                 1    DHCP snooping
```

GigabitEthernet 1/0/1 will filter packets based on the IPv4SG binding.

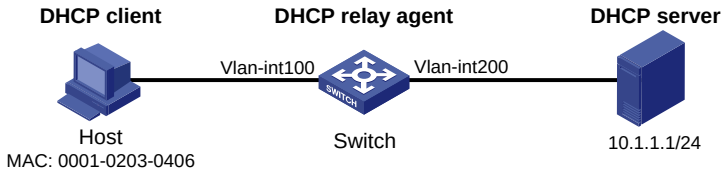
Example: Configuring DHCP relay agent-based dynamic IPv4SG

Network configuration

As shown in Figure 4, DHCP relay agent is enabled on the switch. The host obtains an IP address from the DHCP server through the DHCP relay agent.

Enable dynamic IPv4SG on VLAN-interface 100 to filter incoming packets by using the IPv4SG bindings generated based on DHCP relay entries.

Figure 4 Network diagram



Procedure

1. Configure dynamic IPv4SG:

   # Configure IP addresses for the interfaces. (Details not shown.)

   # Enable IPv4SG on VLAN-interface 100 and verify the source IP address and MAC address for dynamic IPSG.

   ```
   <Switch> system-view
   [Switch] interface vlan-interface 100
   [Switch-Vlan-interface100] ip verify source ip-address mac-address
   [Switch-Vlan-interface100] quit
   ```

2. Configure the DHCP relay agent:

   # Enable the DHCP service.

   ```
   [Switch] dhcp enable
   ```

   # Enable recording DHCP relay entries.

   ```
   [Switch] dhcp relay client-information record
   ```

   # Configure VLAN-interface 100 to operate in DHCP relay mode.

   ```
   [Switch] interface vlan-interface 100
   [Switch-Vlan-interface100] dhcp select relay
   ```

   # Specify the IP address of the DHCP server.

   ```
   [Switch-Vlan-interface100] dhcp relay server-address 10.1.1.1
   [Switch-Vlan-interface100] quit
   ```

Verifying the configuration

Display dynamic IPv4SG bindings generated based on DHCP relay entries.

```
[Switch] display ip source binding dhcp-relay
Total entries found: 1
IP address      MAC address    Interface               VLAN Type
192.168.0.1     0001-0203-0406 Vlan100                 100  DHCP relay
```

VLAN-interface 100 will filter packets based on the IPv4SG binding.

Example: Configuring static IPv6SG

Network configuration

As shown in Figure 5, configure a static IPv6SG binding on GigabitEthernet 1/0/1 of the device to allow only IPv6 packets from the host to pass.

Figure 5 Network diagram

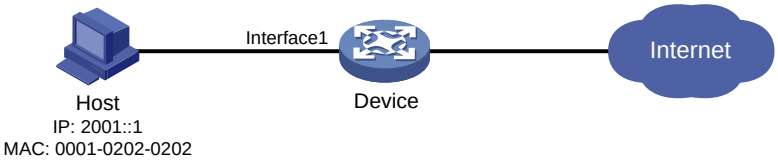


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Enable IPv6SG on GigabitEthernet 1/0/1.

```
<Device> system-view
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] ipv6 verify source ip-address mac-address
```

On GigabitEthernet 1/0/1, configure a static IPv6SG binding for the host.

```
[Device-GigabitEthernet1/0/1] ipv6 source binding ip-address 2001::1 mac-address
0001-0202-0202
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Verify that the static IPv6SG binding is configured successfully on the device.

```
[Device] display ipv6 source binding static
Total entries found: 1
IPv6 address         MAC address    Interface             VLAN Type
2001::1              0001-0202-0202 GE1/0/1               N/A  Static
```

Example: Configuring DHCPv6 snooping-based dynamic IPv6SG address bindings

Network configuration

As shown in Figure 6, the host (the DHCPv6 client) obtains an IP address from the DHCPv6 server. Perform the following tasks:

- Enable DHCPv6 snooping on the device to make sure the DHCPv6 client obtains an IPv6 address from the authorized DHCPv6 server. To generate a DHCPv6 snooping entry for the DHCPv6 client, enable recording of client information in DHCPv6 snooping entries.
- Enable dynamic IPv6SG on GigabitEthernet 1/0/1 to filter incoming packets by using the IPv6SG bindings generated based on DHCPv6 snooping entries. Only packets from the DHCPv6 client are allowed to pass.

Figure 6 Network diagram

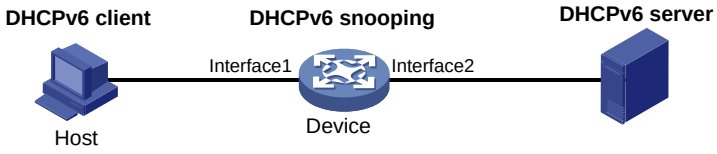


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure DHCPv6 snooping:

   # Enable DHCPv6 snooping globally.

   ```
   <Device> system-view
   [Device] ipv6 dhcp snooping enable
   ```

   # Configure GigabitEthernet 1/0/2 as a trusted interface.

   ```
   [Device] interface gigabitethernet 1/0/2
   [Device-GigabitEthernet1/0/2] ipv6 dhcp snooping trust
   [Device-GigabitEthernet1/0/2] quit
   ```

2. Enable IPv6SG:

   # Enable IPv6SG on GigabitEthernet 1/0/1 and verify the source IP address and MAC address for dynamic IPv6SG.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] ipv6 verify source ip-address mac-address
   ```

   # Enable recording of client information in DHCPv6 snooping entries on GigabitEthernet 1/0/1.

   ```
   [Device-GigabitEthernet1/0/1] ipv6 dhcp snooping binding record
   [Device-GigabitEthernet1/0/1] quit
   ```

Verifying the configuration

Display dynamic IPv6SG bindings generated based on DHCPv6 snooping entries.

```
[Device] display ipv6 source binding dhcpv6-snooping
Total entries found: 1
IPv6 address         MAC address    Interface             VLAN Type
2001::1              040a-0000-0001 GE1/0/1               1    DHCPv6 snooping
```

GigabitEthernet 1/0/1 will filter packets based on the IPv6SG binding.

Example: Configuring DHCPv6 snooping-based dynamic IPv6SG prefix bindings

Network configuration

As shown in Figure 7, the host (the DHCPv6 client) obtains an IPv6 prefix from the DHCPv6 server. Perform the following tasks:

- Enable DHCPv6 snooping on the device to make sure the DHCPv6 client obtains an IPv6 prefix from the authorized DHCPv6 server. To generate a DHCPv6 snooping prefix entry for the DHCPv6 client, enable recording IPv6 prefix information in DHCPv6 snooping entries.

- Enable dynamic IPv6SG on GigabitEthernet 1/0/1 to filter incoming packets by using the IPv6SG bindings generated based on DHCPv6 snooping prefix entries. Only packets from the DHCPv6 client are allowed to pass.

Figure 7 Network diagram

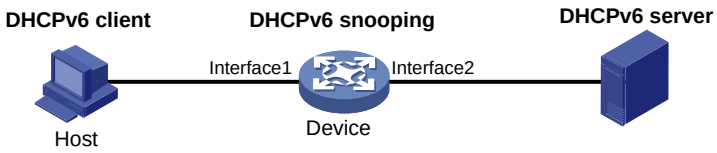


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure DHCPv6 snooping.

   # Enable DHCPv6 snooping globally.

   ```
   <Device> system-view
   [Device] ipv6 dhcp snooping enable
   ```

   # Configure GigabitEthernet 1/0/2 as a trusted interface.

   ```
   [Device] interface gigabitethernet 1/0/2
   [Device-GigabitEthernet1/0/2] ipv6 dhcp snooping trust
   [Device-GigabitEthernet1/0/2] quit
   ```

   # Enable recording DHCPv6 snooping prefix entries on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] ipv6 dhcp snooping pd binding record
   ```

2. Enable IPv6SG.

   # Enable IPv6SG on GigabitEthernet 1/0/1 and verify the source IP address and MAC address for dynamic IPv6SG.

   ```
   [Device-GigabitEthernet1/0/1] ipv6 verify source ip-address mac-address
   [Device-GigabitEthernet1/0/1] quit
   ```

Verifying the configuration

Display dynamic IPv6SG bindings generated based on DHCPv6 snooping entries.

```
[Device] display ipv6 source binding pd
Total entries found: 1
IPv6 address         MAC address    Interface              VLAN Type
2001:410:1::/48      0010-9400-0004 GE1/0/1                1
```

GigabitEthernet 1/0/1 will filter packets based on the IPv6SG binding.

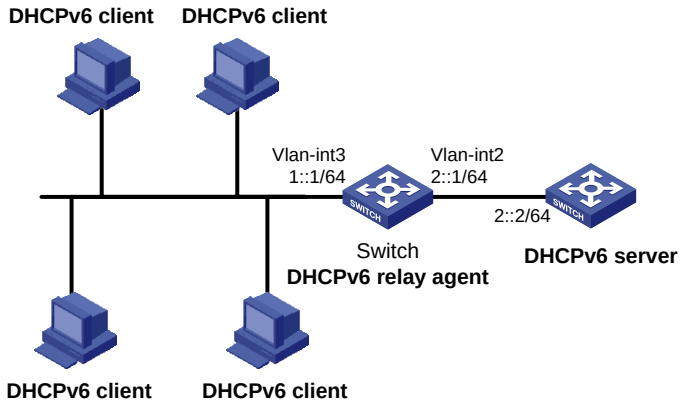
Example: Configuring DHCPv6 relay agent-based dynamic IPv6SG

Network configuration

As shown in Figure 8, DHCPv6 relay agent is enabled on the switch. The clients obtain IPv6 addresses from the DHCPv6 server through the DHCPv6 relay agent.

Enable dynamic IPv6SG on VLAN-interface 3 to filter incoming packets by using the IPv6SG bindings generated based on DHCPv6 relay entries.

Figure 8 Network diagram



Procedure

1. Configure the DHCPv6 relay agent:

   # Create VLAN 2 and VLAN 3, assign interfaces to the VLANs, and specify IP addresses for VLAN-interface 2 and VLAN-interface 3. (Details not shown.)

   # Enable the DHCPv6 relay agent on VLAN-interface 3.

   ```
   <Switch> system-view
   [Switch] interface vlan-interface 3
   [Switch-Vlan-interface3] ipv6 dhcp select relay
   ```

   # Enable recording of DHCPv6 relay entries on the interface.

   ```
   [Switch-Vlan-interface3] ipv6 dhcp relay client-information record
   ```

   # Specify the DHCPv6 server address 2::2 on the relay agent.

   ```
   [Switch-Vlan-interface3] ipv6 dhcp relay server-address 2::2
   ```

2. Enable IPv6SG on VLAN-interface 3 and verify the source IP address and MAC address for dynamic IPv6SG.

   ```
   [Switch-Vlan-interface3] ipv6 verify source ip-address mac-address
   [Switch-Vlan-interface3] quit
   ```

Verifying the configuration

Display dynamic IPv6SG bindings generated based on DHCPv6 relay entries.

```
[Switch] display ipv6 source binding dhcpv6-relay
Total entries found: 1
IP Address      MAC Address    Interface               VLAN Type
1::2            0001-0203-0406 Vlan3                   3    DHCPv6 relay
```

VLAN-interface 3 will filter packets based on the IPv6SG binding.

Contents

Configuring ARP attack protection 1

About ARP attack protection 1
ARP attack protection tasks at a glance 1
Configuring unresolvable IP attack protection 1
About unresolvable IP attack protection 1
Configuring ARP source suppression 2
Configuring ARP blackhole routing 2
Display and maintenance commands for unresolvable IP attack protection 3
Example: Configuring unresolvable IP attack protection 3
Configuring source MAC-based ARP packet rate limiting 4
Configuring source MAC-based ARP attack detection 5
About this task 5
Restrictions and guidelines 5
Procedure 5
Display and maintenance commands for source MAC-based ARP attack detection 6
Example: Configuring source MAC-based ARP attack detection 6
Configuring ARP packet source MAC consistency check 7
About ARP packet source MAC consistency check 7
Procedure 7
Configuring ARP active acknowledgement 8
Configuring authorized ARP 8
About authorized ARP 8
Procedure 9
Example: Configuring authorized ARP on a DHCP server 9
Example: Configuring authorized ARP on a DHCP relay agent 10
Configuring ARP attack detection 11
About ARP attack detection 11
Configuring user validity check 12
Configuring ARP packet validity check 13
Configuring ARP restricted forwarding 14
Ignoring ingress ports of ARP packets during user validity check 15
Configuring ARP attack detection for a VSI 15
Enabling ARP attack detection logging 17
Display and maintenance commands for ARP attack detection 17
Example: Configuring user validity check 17
Example: Configuring user validity check and ARP packet validity check 19
Example: Configuring ARP restricted forwarding 20
Configuring ARP scanning and fixed ARP 22
About this task 22
Restrictions and guidelines 23
Triggering ARP scanning 23
Configuring automatic ARP scanning 23
Configuring fixed ARP 24
Configuring ARP keepalive entry scanning 24
About ARP keepalive entry scanning 24
Configuring ARP keepalive entry scanning 24
Display and maintenance commands for ARP keepalive entry scanning 24
Configuring ARP gateway protection 25
About ARP gateway protection 25
Restrictions and guidelines 25
Procedure 25
Example: Configuring ARP gateway protection 26
Configuring ARP filtering 27
About ARP filtering 27
Restrictions and guidelines 27
Procedure 27
Example: Configuring ARP filtering 27

Configuring ARP sender IP address checking 28
About ARP sender IP address checking 28
Procedure 29
Configuring ARP entry fixing 29
Enabling SNMP notifications for ARP 30

Configuring ARP attack protection

About ARP attack protection

The device can provide multiple features to detect and prevent ARP attacks and viruses in the LAN. An attacker can exploit ARP vulnerabilities to attack network devices in the following ways:

- Sends a large number of unresolvable IP packets to have the receiving device busy with resolving IP addresses until its CPU is overloaded. Unresolvable IP packets refer to IP packets for which ARP cannot find corresponding MAC addresses.
- Sends a large number of ARP packets to overload the CPU of the receiving device.
- Acts as a trusted user or gateway to send ARP packets so the receiving devices obtain incorrect ARP entries.

ARP attack protection tasks at a glance

All ARP attack protection tasks are optional.

- Preventing flood attacks
  - Configuring unresolvable IP attack protection
  - Configuring source MAC-based ARP packet rate limiting
  - Configuring source MAC-based ARP attack detection
- Preventing user and gateway spoofing attacks
  - Configuring ARP packet source MAC consistency check
  - Configuring ARP active acknowledgement
  - Configuring authorized ARP
  - Configuring ARP attack detection
  - Configuring ARP scanning and fixed ARP
  - Configuring ARP keepalive entry scanning
  - Configuring ARP gateway protection
  - Configuring ARP filtering
  - Configuring ARP sender IP address checking
  - Configuring ARP entry fixing
- Enabling SNMP notifications for ARP

Configuring unresolvable IP attack protection

About unresolvable IP attack protection

If a device receives a large number of unresolvable IP packets from a host, the following situations can occur:

- The device sends a large number of ARP requests, overloading the target subnets.
- The device keeps trying to resolve the destination IP addresses, overloading its CPU.

To protect the device from such IP attacks, you can configure the following features:

- **ARP source suppression**—This feature is applicable if the attack packets have the same source addresses.
- **ARP blackhole routing**—This feature is applicable regardless of whether the attack packets have the same source addresses.

Configuring ARP source suppression

About this task

This feature stops resolving packets from an IP address if the number of unresolvable IP packets from the IP address exceeds the upper limit within 5 seconds. The device continues ARP resolution when the interval elapses.

For a gateway or server trusted by the device, you can configure an ARP source suppression rule so that the device always permits IP packets sourced from the IP address of the gateway or server.

Restrictions and guidelines

To have the configured ARP source suppression rules take effect, you must enable ARP source suppression by using the **arp source-suppression enable** command.

Procedure

1. Enter system view.

   system-view
2. Enable ARP source suppression.

   arp source-suppression enable

   By default, ARP source suppression is disabled.
3. Set the maximum number of unresolvable packets that the device can process per source IP address within 5 seconds.

   arp source-suppression limit *limit-value*

   By default, the maximum number is 10.
4. Configure ARP source suppression rules.

   arp source-suppression rule *rule-id* **permit** *ip-address* [*mask*]

   By default, no ARP source suppression rules exist.

Configuring ARP blackhole routing

About this task

This feature creates a blackhole route destined for an unresolved IP address. The device drops all matching packets until the blackhole route is deleted. A blackhole route is deleted when its aging timer is reached or the route becomes reachable.

After a blackhole route is created for an unresolved IP address, the device immediately starts the first ARP blackhole route probe by sending an ARP request. If the resolution fails, the device continues probing according to the probe settings. If the IP address resolution succeeds in a probe, the device converts the blackhole route to a normal route. If an ARP blackhole route ages out before the device finishes all probes, the device deletes the blackhole route and does not perform the remaining probes.

Restrictions and guidelines

Set the ARP blackhole route probe count to a big value, for example, 25. If the device fails to reach the destination IP address temporarily and the probe count is too small, all probes might finish before the problem is resolved. As a result, non-attack packets will be dropped. This setting can avoid such situation.

Procedure

1. Enter system view.
   system-view
2. Enable ARP blackhole routing.
   arp resolving-route enable
   By default, ARP blackhole routing is enabled.
3. (Optional.) Set the number of ARP blackhole route probes for each unresolved IP address.
   arp resolving-route probe-count *count*
   The default setting is three probes.
4. (Optional.) Set the interval at which the device probes ARP blackhole routes.
   arp resolving-route probe-interval *interval*
   The default setting is 1 second.

Display and maintenance commands for unresolvable IP attack protection

Execute **display** commands in any view.

Task	Command
Display ARP source suppression configuration information.	**display arp source-suppression**
Display ARP source suppression entries.	**display arp source-suppression cache** [**slot** *slot-number*]

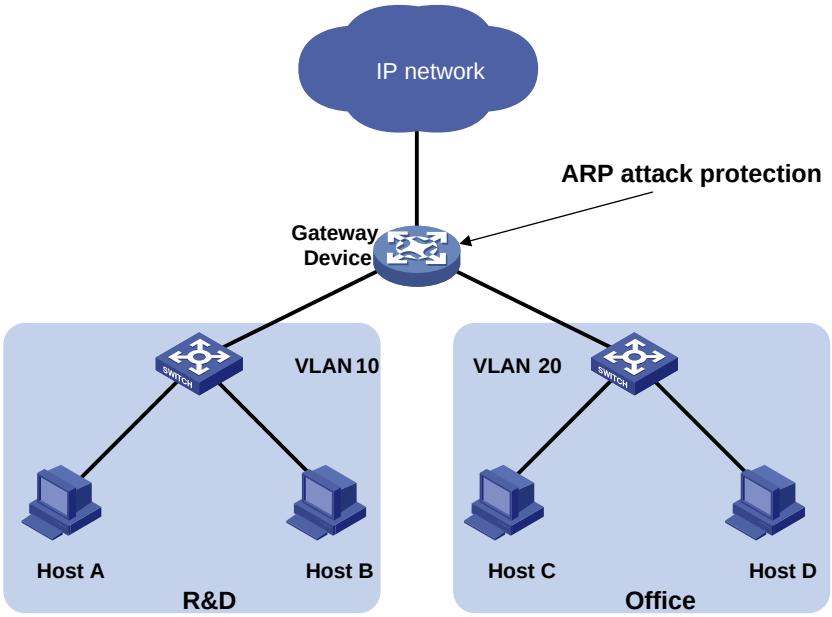
Example: Configuring unresolvable IP attack protection

Network configuration

As shown in Figure 1, a LAN contains two areas: an R&D area in VLAN 10 and an office area in VLAN 20. Each area connects to the gateway (Device) through an access switch.

A large number of ARP requests are detected in the office area and are considered an attack caused by unresolvable IP packets. To prevent the attack, configure ARP source suppression or ARP blackhole routing.

Figure 1 Network diagram



Procedure

- If the attack packets have the same source address, configure ARP source suppression:

  # Enable ARP source suppression.

  ```
  <Device> system-view
  [Device] arp source-suppression enable
  ```

  # Configure the device to process a maximum of 100 unresolvable packets per source IP address within 5 seconds.

  ```
  [Device] arp source-suppression limit 100
  ```

- If the attack packets have different source addresses, configure ARP blackhole routing:

  # Enable ARP blackhole routing.

  ```
  [Device] arp resolving-route enable
  ```

Configuring source MAC-based ARP packet rate limiting

About this task

A client might send a large number of ARP requests, which might cause high CPU usage on the gateway, adversely affect other services, and even crash the device. To resolve the issue, use the source MAC-based ARP packet rate limit to limit the receiving rate of ARP packets from the client. When the receiving rate of ARP requests from the client exceeds the limit, the device will discard subsequent ARP requests.

You can use the **`debugging arp packet`** command to view the source MAC addresses from which the rate of ARP packets is high.

Restrictions and guidelines

You can configure both global rate limit for all ARP packets and MAC-specific rate limit for ARP packets from specific MAC addresses. For the ARP packets from a specific MAC address, the global limit rate or MAC-specific limit rate whichever is smaller applies. If you execute the **`undo arp`**

rate-limit source-mac command without specifying the *mac-address* argument, the command disables only global rate limit. MAC-specific rate limit for ARP packets still takes effect.

If you set both ARP packet rate limits on interfaces and source MAC-based ARP packet rate limit, the rate limit works as follows:

- For the system, the rate limit that takes effect is the one with a smaller value between the sum of rate limits that take effect on the interfaces receiving ARP packets from the MAC address and the source MAC-based ARP packet rate limit.
- For an interface, the rate limit that takes effect is the one with a smaller value between the ARP packet rate limit and the actual ARP packet traffic received by the interface.

Procedure

1. Enter system view.
   system-view
2. Enable source MAC-based ARP packet rate limiting and set the limit rate.
   arp rate-limit source-mac [*mac-address*] *pps*
   By default, source MAC-based ARP packet rate limiting is disabled.

Configuring source MAC-based ARP attack detection

About this task

This feature checks the number of ARP packets delivered to the CPU. If the number of packets from the same MAC address within 5 seconds exceeds a threshold, the device generates an ARP attack detection entry for the MAC address. If logging is enabled for source MAC-based ARP attack detection, the device handles the attack by using either of the following methods before the ARP attack detection entry ages out:

- **Monitor**—Only generates log messages.
- **Filter**—Generates log messages and filters out subsequent ARP packets from the MAC address and data packets sourced from or destined for the MAC address.

When an ARP attack detection entry ages out, ARP packets sourced from the MAC address in the entry can be processed correctly.

Some security devices or servers might send a large number of ARP packets for security check. To prevent such ARP packets from being filtered out, exclude the MAC addresses of these devices from source MAC-based ARP attack detection. The device will directly permit the packets sourced from the MAC addresses.

Restrictions and guidelines

When you change the handling method from monitor to filter, the configuration takes effect immediately. When you change the handling method from filter to monitor, the device continues filtering packets that match existing ARP attack detection entries, and uses the monitor method to process packets that match the new ARP attack detection entries.

Procedure

1. Enter system view.
   system-view

2. Enable source MAC-based ARP attack detection and specify the handling method.
   arp source-mac { **filter** | **monitor** }
   By default, this feature is disabled.
3. Set the threshold.
   arp source-mac threshold *threshold-value*
   By default, the threshold is 30.
4. Set the aging timer for ARP attack detection entries.
   arp source-mac aging-time *time*
   By default, the lifetime is 300 seconds.
5. (Optional.) Exclude specific MAC addresses from this detection.
   arp source-mac exclude-mac *mac-address*&<1-10>
   By default, no MAC address is excluded.
6. Enable logging for source MAC-based ARP attack detection.
   arp source-mac log enable
   By default, logging for source MAC-based ARP attack detection is disabled.

Display and maintenance commands for source MAC-based ARP attack detection

Execute **display** commands in any view.

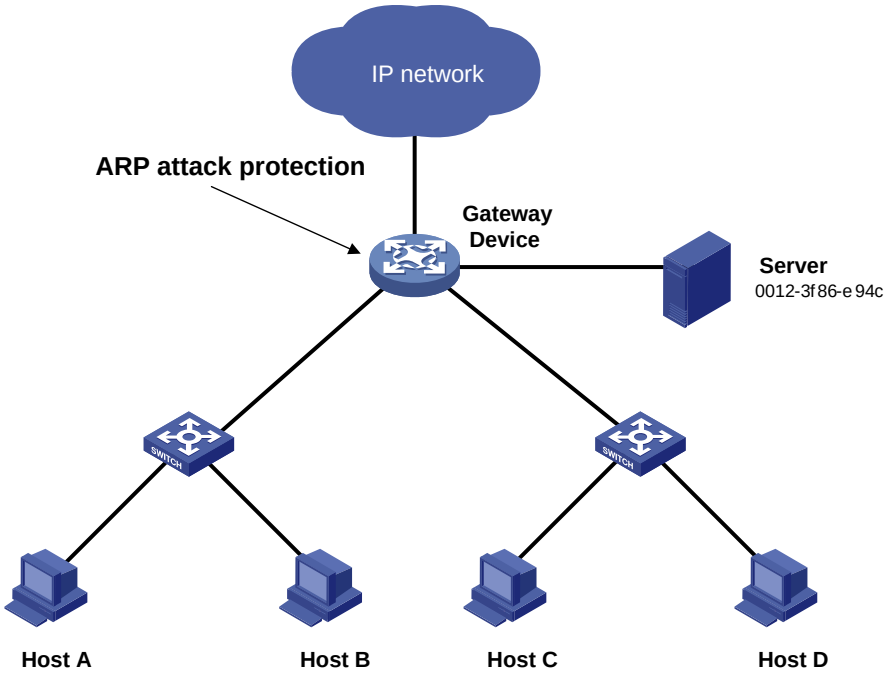
Task	Command
Display ARP attack detection entries generated by source MAC-based ARP attack detection.	**display arp source-mac** { **interface** *interface-type interface-number* [**slot** *slot-number*] | **slot** *slot-number* }

Example: Configuring source MAC-based ARP attack detection

Network configuration

As shown in Figure 2, the hosts access the Internet through a gateway (Device). If malicious users send a large number of ARP requests to the gateway, the gateway might crash and cannot process requests from the clients. To solve this problem, configure source MAC-based ARP attack detection on the gateway.

Figure 2 Network diagram



Procedure

Enable source MAC-based ARP attack detection, and specify the handling method as filter.

```
<Device> system-view
[Device] arp source-mac filter
```

Set the threshold to 30.

```
[Device] arp source-mac threshold 30
```

Set the lifetime for source MAC-based ARP attack detection entries to 60 seconds.

```
[Device] arp source-mac aging-time 60
```

Exclude MAC address 0012-3f86-e94c from this detection.

```
[Device] arp source-mac exclude-mac 0012-3f86-e94c
```

Configuring ARP packet source MAC consistency check

About ARP packet source MAC consistency check

This feature enables a gateway to filter ARP packets whose source MAC address in the Ethernet header is different from the sender MAC address in the message body. This feature allows the gateway to learn correct ARP entries.

Procedure

1. Enter system view.
   `system-view`
2. Enable ARP packet source MAC address consistency check.

arp valid-check enable

By default, ARP packet source MAC address consistency check is disabled.

Configuring ARP active acknowledgement

About this task

Use the ARP active acknowledgement feature on gateways to prevent user spoofing.

This feature enables the device to perform active acknowledgement before creating an ARP entry.

- Upon receiving an ARP request that requests the MAC address of the device, the device sends an ARP reply. Then, it sends an ARP request for the sender IP address in the received ARP request to determine whether to create an ARP entry for the sender IP address.
  - If the device receives an ARP reply within the probe interval, it creates the ARP entry.
  - If the device does not receive an ARP reply within the probe interval, it does not create the ARP entry.
- Upon receiving an ARP reply, the device examines whether it was the reply to the request that the device has sent.
  - If it was, the device creates an ARP entry for the sender IP address in the ARP reply.
  - If it was not, the device sends an ARP request for the sender IP address to determine whether to create an ARP entry for the sender IP address.
    - If the device receives an ARP reply within the probe interval, it creates the ARP entry.
    - If the device does not receive an ARP reply within the probe interval, it does not create the ARP entry.

To improve validity and reliability of ARP entries, you can enable ARP active acknowledgement in strict mode. In this mode, the device creates ARP entries only for the IP addresses that the device actively initiates the ARP resolution.

Procedure

1. Enter system view.

   system-view
2. Enable ARP active acknowledgement.

   arp active-ack [**strict**] **enable**

   By default, ARP active acknowledgement is disabled.

   For ARP active acknowledgement to take effect in strict mode, make sure ARP blackhole routing is enabled.

Configuring authorized ARP

About authorized ARP

Authorized ARP entries are generated based on the DHCP clients' address leases on the DHCP server or dynamic client entries on the DHCP relay agent. For more information about DHCP server and DHCP relay agent, see *Layer 3—IP Services Configuration Guide*.

Use this feature to prevent user spoofing and to allow only authorized clients to access network resources.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

   Supported interface types include Layer 3 Ethernet interface, Layer 3 Ethernet subinterface, Layer 3 aggregate interface, Layer 3 aggregate subinterface, and VLAN interface.

3. Enable authorized ARP on the interface.

   arp authorized enable

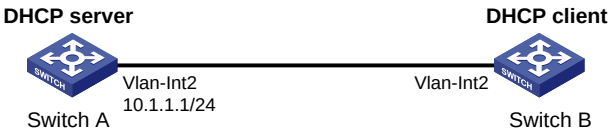
   By default, authorized ARP is disabled.

Example: Configuring authorized ARP on a DHCP server

Network configuration

As shown in Figure 3, configure authorized ARP on Vlan-interface2 of Switch A (a DHCP server) to ensure user validity.

Figure 3 Network diagram



Procedure

1. Configure Switch A:

   # Specify the IP address for Vlan-interface2.

   ```
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ip address 10.1.1.1 24
   [SwitchA-Vlan-interface2] quit
   ```

   # Configure DHCP.

   ```
   [SwitchA] dhcp enable
   [SwitchA] dhcp server ip-pool 1
   [SwitchA-dhcp-pool-1] network 10.1.1.0 mask 255.255.255.0
   [SwitchA-dhcp-pool-1] quit
   ```

   # Enter VLAN interface view.

   ```
   [SwitchA] interface vlan-interface 2
   ```

   # Enable authorized ARP.

   ```
   [SwitchA-Vlan-interface2] arp authorized enable
   [SwitchA-Vlan-interface2] quit
   ```

2. Configure Switch B:

   ```
   <SwitchB> system-view
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] ip address dhcp-alloc
   [SwitchB-Vlan-interface2] quit
   ```

Verifying the configuration

Display authorized ARP entry information on Switch A.

```
[SwitchA] display arp all
  Type: S-Static   D-Dynamic   O-Openflow   R-Rule   M-Multiport  I-Invalid
IP address      MAC address    VLAN/VSI name Interface               Aging Type
10.1.1.2        0012-3f86-e94c 2             GE1/0/1                 960   D
```

The output shows that IP address 10.1.1.2 has been assigned to Switch B.

Switch B must use the IP address and MAC address in the authorized ARP entry to communicate with Switch A. Otherwise, the communication fails. Thus user validity is ensured.

Example: Configuring authorized ARP on a DHCP relay agent

Network configuration

As shown in Figure 4, configure authorized ARP on GigabitEthernet 1/0/2 of Device B (a DHCP relay agent) to ensure user validity.

Figure 4 Network diagram

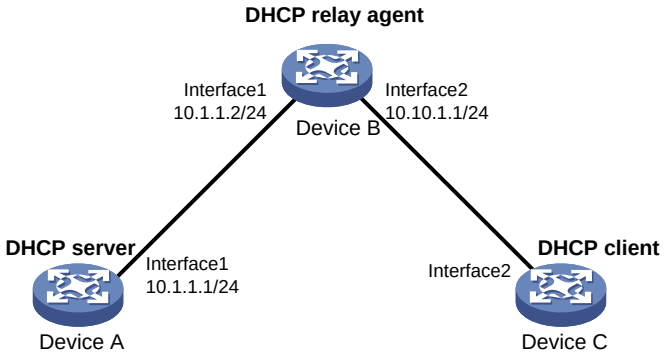


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

   # Specify the IP address for GigabitEthernet 1/0/1.

   ```
   <DeviceA> system-view
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ip address 10.1.1.1 24
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

   # Configure DHCP.

   ```
   [DeviceA] dhcp enable
   [DeviceA] dhcp server ip-pool 1
   [DeviceA-dhcp-pool-1] network 10.10.1.0 mask 255.255.255.0
   [DeviceA-dhcp-pool-1] gateway-list 10.10.1.1
   ```

```
[DeviceA-dhcp-pool-1] quit
[DeviceA] ip route-static 10.10.1.0 24 10.1.1.2
```

2. Configure Device B:

   # Enable DHCP.

```
<DeviceB> system-view
[DeviceB] dhcp enable
```

   # Specify the IP addresses of GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] ip address 10.1.1.2 24
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ip address 10.10.1.1 24
```

   # Enable DHCP relay agent on GigabitEthernet 1/0/2.

```
[DeviceB-GigabitEthernet1/0/2] dhcp select relay
```

   # Add the DHCP server 10.1.1.1 to DHCP server group 1.

```
[DeviceB-GigabitEthernet1/0/2] dhcp relay server-address 10.1.1.1
```

   # Enable authorized ARP.

```
[DeviceB-GigabitEthernet1/0/2] arp authorized enable
[DeviceB-GigabitEthernet1/0/2] quit
```

   # Enable recording of relay entries on the relay agent.

```
[DeviceB] dhcp relay client-information record
```

3. Configure Device C:

```
<DeviceC> system-view
[DeviceC] ip route-static 10.1.1.0 24 10.10.1.1
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] ip address dhcp-alloc
[DeviceC-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Display authorized ARP information on Device B.

```
[DeviceB] display arp all
  Type: S-Static   D-Dynamic   O-Openflow   R-Rule   M-Multiport  I-Invalid
IP address      MAC address    VLAN/VSI name Interface               Aging Type
10.10.1.2       0012-3f86-e94c --            GE1/0/2                  960   D
```

The output shows that Device A assigned the IP address 10.10.1.2 to Device C.

Device C must use the IP address and MAC address in the authorized ARP entry to communicate with Device B. Otherwise, the communication fails. Thus the user validity is ensured.

Configuring ARP attack detection

About ARP attack detection

ARP attack detection enables access devices to block ARP packets from unauthorized clients to prevent user spoofing and gateway spoofing attacks.

ARP attack detection provides the following features:

- User validity check.

- ARP packet validity check.
- ARP restricted forwarding.
- ARP packet ingress port ignoring during user validity check
- ARP attack detection logging.

If both ARP packet validity check and user validity check are enabled, the former one applies first, and then the latter applies.

Do not configure ARP attack detection together with ARP snooping. Otherwise, ARP snooping entries cannot be generated.

Configuring user validity check

About this task

User validity check does not check ARP packets received on ARP trusted interfaces. This feature compares the sender IP and sender MAC in the ARP packet received on an ARP untrusted interface with the matching criteria in the following order:

1. User validity check rules.
   - If a match is found, the device processes the ARP packet according to the rule.
   - If no match is found or no user validity check rule is configured, proceeds to step 2.
2. Static IPSG bindings, 802.1X security entries, and DHCP snooping entries.
   - If a match is found, the device determines that the ARP packet is valid, and then forwards the ARP packet.
     - If a match is found and the receiving interface is different from the interface in the entry with a matching sender IP address, the device performs Layer 3 forwarding.
     - If a match is found but the receiving interface is the same as the interface in the entry with a matching sender IP address, the device performs Layer 2 forwarding.
     - If no match is found, the device performs Layer 2 forwarding.
   - If no match is found, the device discards the ARP packet.

Static IP source guard bindings are created by using the **`ip source binding`** command. For more information, see "Configuring IP source guard."

DHCP snooping entries are automatically generated by DHCP snooping. For more information, see *Layer 3—IP Services Configuration Guide*.

802.1X security entries record the IP-to-MAC mappings for 802.1X clients. After a client passes 802.1X authentication and uploads its IP address to an ARP attack detection enabled device, the device automatically generates an 802.1X security entry. The 802.1X client must be enabled to upload its IP address to the device. For more information, see "Configuring 802.1X."

Restrictions and guidelines

When you configure user validity check, make sure one or more of the following items are configured:

- User validity check rules.
- Static IP source guard bindings.
- DHCP snooping.
- 802.1X.

If none of the items is configured, no incoming ARP packets on ARP untrusted interfaces can be forwarded normally.

Specify an IP address, a MAC address, and a VLAN where ARP attack detection is enabled for an IP source guard binding. Otherwise, no ARP packets can match the IP source guard binding.

Configuring user validity check rules

1. Enter system view.

   system-view

2. Configure a user validity check rule.

   arp detection rule *rule-id* { **deny** | **permit** } **ip** { *ip-address* [*mask*] | **any** } **mac** { *mac-address* [*mask*] | **any** } [**vlan** *vlan-id* | **vsi** *vsi-name*]

   By default, no user validity check rules are configured.

Enabling ARP attack detection in a VLAN

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Enable ARP attack detection.

   arp detection enable

   By default, ARP attack detection is disabled. The device does not perform ARP user validity check.

4. (Optional.) Configure an interface that does not require ARP user validity check as a trusted interface.

   a. Return to system view.

      quit

   b. Enter interface view.

      interface *interface-type interface-number*

      Supported interface types include Layer 2 Ethernet interface and Layer 2 aggregate interface.

   c. Configure the interface as a trusted interface excluded from ARP attack detection.

      arp detection trust

      By default, an interface is untrusted.

Configuring ARP packet validity check

About this task

ARP packet validity check does not check ARP packets received on ARP trusted interfaces. To check ARP packets received on untrusted interfaces, you can specify the following objects to be checked:

- **src-mac**—Checks whether the sender MAC address in the message body is identical to the source MAC address in the Ethernet header. If they are identical, the packet is forwarded. Otherwise, the packet is discarded.
- **dst-mac**—Checks the target MAC address of ARP replies. If the target MAC address is all-zero, all-one, or inconsistent with the destination MAC address in the Ethernet header, the packet is considered invalid and discarded.
- **ip**—Checks the sender and target IP addresses of ARP replies, and the sender IP address of ARP requests. All-one or multicast IP addresses are considered invalid and the corresponding packets are discarded.

Prerequisites

Before you configure ARP packet validity check, you must first configure user validity check. For more information about user validity check configuration, see "Configuring user validity check."

Enabling ARP packet validity check

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ARP attack detection.
   arp detection enable
   By default, ARP attack detection is disabled. The device does not perform ARP packet validity check.
4. Enable ARP packet validity check.
   a. Return to system view.
      quit
   b. Enable ARP packet validity check and specify the objects to be checked.
      arp detection validate { **dst-mac** | **ip** | **src-mac** } *
      By default, ARP packet validity check is disabled.

Enabling ARP attack detection in a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ARP attack detection.
   arp detection enable
   By default, ARP attack detection is disabled. The device does not perform ARP packet validity check.
4. (Optional.) Configure the interface that does not require ARP packet validity check as a trusted interface.
   a. Return to system view.
      quit
   b. Enter interface view.
      interface *interface-type interface-number*
      Supported interface types include Layer 2 Ethernet interface and Layer 2 aggregate interface.
   c. Configure the interface as a trusted interface excluded from ARP attack detection.
      arp detection trust
      By default, an interface is untrusted.

Configuring ARP restricted forwarding

About this task

ARP restricted forwarding does not take effect on ARP packets received on ARP trusted interfaces and forwards the ARP packets correctly. This feature controls the forwarding of ARP packets that are received on untrusted interfaces and have passed user validity check as follows:

- If the packets are ARP requests, they are forwarded through the trusted interface.
- If the packets are ARP replies, they are forwarded according to their destination MAC address. If no match is found in the MAC address table, they are forwarded through the trusted interface.

Restrictions and guidelines

ARP restricted forwarding does not apply to ARP packets that use multiport destination MAC addresses.

Prerequisites

Configure user validity check before you configure ARP restricted forwarding. For information about user validity check configuration, see "Configuring user validity check."

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ARP restricted forwarding.
   arp restricted-forwarding enable
   By default, ARP restricted forwarding is disabled.

Ignoring ingress ports of ARP packets during user validity check

About this task

ARP attack detection performs user validity check on ARP packets from ARP untrusted interfaces. The sender IP and sender MAC in the received ARP packet are compared with the entries used for user validity check. In addition, user validity check compares the ingress port of the ARP packet with the port in the entries. If no matching port is found, the ARP packet is discarded. For more information about user validity check, see "Configuring user validity check."

You can configure the device to ignore the ingress ports of ARP packets during user validity check. If an ARP packet passes user validity check, the device directly performs Layer 2 forwarding for the packet. It no longer searches for a matching static IPSG binding, 802.1X security entry, or DHCP snooping entry by the target IP address of the ARP packet.

Procedure

1. Enter system view.
   system-view
2. Ignore ingress ports of ARP packets during user validity check.
   arp detection port-match-ignore
   By default, ingress ports of ARP packets are checked during user invalidity.

Configuring ARP attack detection for a VSI

About this task

In VXLAN networks, you can configure a VTEP to perform ARP attack detection in a VSI. ARP attack detection performs user validity check and ARP packet validity check on ARP packets from ARP untrusted ACs. For information about ACs, see *VXLAN Configuration Guide*.

The user validity check and ARP packet validity check mechanisms for a VSI are the same as those for a VLAN. For more information, see "Configuring user validity check" and "Configuring ARP packet validity check."

Configuring user validity check for a VSI

1. Enter system view.
   system-view
2. (Optional.) Configure a user validity check rule.
   arp detection rule *rule-id* { **deny** | **permit** } **ip** { *ip-address* [*mask*] | **any** } **mac** { *mac-address* [*mask*] | **any** } [**vlan** *vlan-id* | **vsi** *vsi-name*]
   By default, no user validity check rule is configured.
3. Enter VSI view.
   vsi *vsi-name*
4. Enable ARP attack detection.
   arp detection enable
   By default, ARP attack detection is disabled.
5. (Optional.) Configure an AC as a trusted AC excluded from ARP attack detection.
   a. Return to system view.
      quit
   b. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
      interface *interface-type interface-number*
   c. Enter Ethernet service instance view.
      service-instance *instance-id*
   d. Configure the AC as a trusted AC excluded from ARP attack detection
      arp detection trust
      By default, an AC is untrusted.

Configuring ARP packet validity check for a VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable ARP attack detection.
   arp detection enable
   By default, ARP attack detection is disabled.
4. Return to system view.
   quit
5. Enable ARP packet validity check and specify the objects to be checked.
   arp detection validate { **dst-mac** | **ip** | **src-mac** } *
   By default, ARP packet validity check is disabled.
6. (Optional.) Configure an AC as a trusted AC excluded from ARP attack detection
   a. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
      interface *interface-type interface-number*
   b. Enter Ethernet service instance view.
      service-instance *instance-id*
   c. Configure the AC as a trusted AC excluded from ARP attack detection.
      arp detection trust
      By default, an AC is untrusted.

Enabling ARP attack detection logging

About this task

The ARP attack detection logging feature enables a device to generate ARP attack detection log messages when illegal ARP packets are detected. An ARP attack detection log message contains the following information:

- Receiving interface of the ARP packets.
- Sender IP address.
- Total number of dropped ARP packets.

Procedure

1. Enter system view.
   system-view
2. Enable ARP attack detection logging.
   arp detection log enable [**interval** *interval* | **number** *number* | **threshold** *threshold-value*]*
   By default, ARP attack detection logging is disabled.

Display and maintenance commands for ARP attack detection

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the VLANs enabled with ARP attack detection.	**display arp detection**
Display ARP attack source statistics.	**display arp detection statistics attack-source slot** *slot-number*
Display statistics for packets dropped by ARP attack detection.	**display arp detection statistics packet-drop** [**interface** *interface-type interface-number*]
Clear ARP attack source statistics.	**reset arp detection statistics attack-source** [**slot** *slot-number*]
Clear statistics for packets dropped by ARP attack detection.	**reset arp detection statistics packet-drop** [**interface** *interface-type interface-number*]

Example: Configuring user validity check

Network configuration

As shown in Figure 5, configure Device B to perform user validity check based on 802.1X security entries for connected hosts.

Figure 5 Network diagram

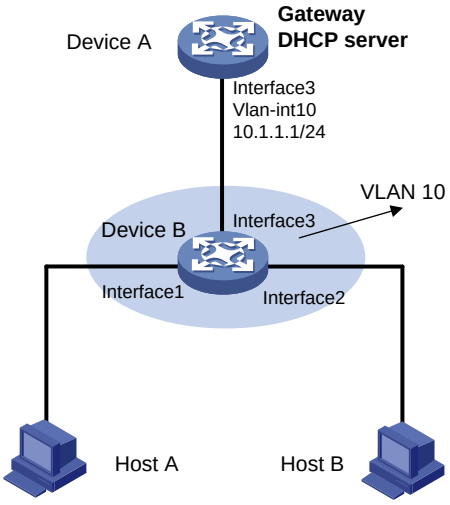


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Add all interfaces on Device B to VLAN 10, and specify the IP address of VLAN-interface 10 on Device A. (Details not shown.)
2. Configure the DHCP server on Device A, and configure DHCP address pool 0.

```
<DeviceA> system-view
[DeviceA] dhcp enable
[DeviceA] dhcp server ip-pool 0
[DeviceA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
```

3. Configure Host A and Host B as 802.1X clients and configure them to upload IP addresses for ARP attack detection. (Details not shown.)
4. Configure Device B:

   # Enable 802.1X.

```
<DeviceB> system-view
[DeviceB] dot1x
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] dot1x
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] dot1x
[DeviceB-GigabitEthernet1/0/2] quit
```

   # Add a local user **test**.

```
[DeviceB] local-user test class network
[DeviceB-luser-network-test] service-type lan-access
[DeviceB-luser-network-test] password simple test
```

```
[DeviceB-luser-network-test] quit
```

Enable ARP attack detection for VLAN 10 to check user validity based on 802.1X entries.

```
[DeviceB] vlan 10
[DeviceB-vlan10] arp detection enable
```

Configure the upstream interface as an ARP trusted interface. By default, an interface is an untrusted interface.

```
[DeviceB-vlan10] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] arp detection trust
[DeviceB-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Verify that ARP packets received on interfaces GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are checked against 802.1X entries.

Example: Configuring user validity check and ARP packet validity check

Network configuration

As shown in Figure 6, configure Device B to perform ARP packet validity check and user validity check based on static IP source guard bindings and DHCP snooping entries for connected hosts.

Figure 6 Network diagram

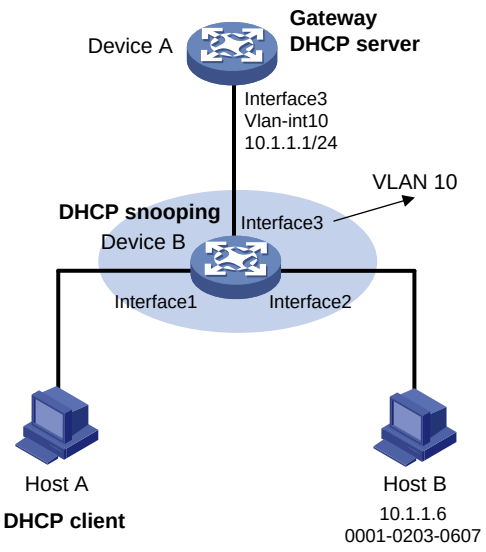


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Add all interfaces on Device B to VLAN 10, and specify the IP address of VLAN-interface 10 on Device A. (Details not shown.)

2. Configure the DHCP server on Device A, and configure DHCP address pool 0.

```
<DeviceA> system-view
[DeviceA] dhcp enable
[DeviceA] dhcp server ip-pool 0
[DeviceA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
```

3. Configure Host A (DHCP client) and Host B. (Details not shown.)
4. Configure Device B:

   # Enable DHCP snooping.

```
<DeviceB> system-view
[DeviceB] dhcp snooping enable
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] dhcp snooping trust
[DeviceB-GigabitEthernet1/0/3] quit
```

   # Enable recording of client information in DHCP snooping entries on GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] dhcp snooping binding record
[DeviceB-GigabitEthernet1/0/1] quit
```

   # Enable ARP attack detection for VLAN 10.

```
[DeviceB] vlan 10
[DeviceB-vlan10] arp detection enable
```

   # Configure the upstream interface as a trusted interface. By default, an interface is an untrusted interface.

```
[DeviceB-vlan10] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] arp detection trust
[DeviceB-GigabitEthernet1/0/3] quit
```

   # Configure a static IP source guard binding entry on interface GigabitEthernet 1/0/2 for user validity check.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ip source binding ip-address 10.1.1.6 mac-address
0001-0203-0607 vlan 10
[DeviceB-GigabitEthernet1/0/2] quit
```

   # Enable ARP packet validity check by checking the MAC addresses and IP addresses of ARP packets.

```
[DeviceB] arp detection validate dst-mac ip src-mac
```

Verifying the configuration

Verify that Device B first checks the validity of ARP packets received on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2. If the ARP packets are confirmed valid, Device B performs user validity check by using the static IP source guard bindings and finally DHCP snooping entries.

Example: Configuring ARP restricted forwarding

Network configuration

As shown in Figure 7, configure ARP restricted forwarding on Device B where ARP attack detection is configured. Port isolation configured on Device B can take effect for broadcast ARP requests.

Figure 7 Network diagram

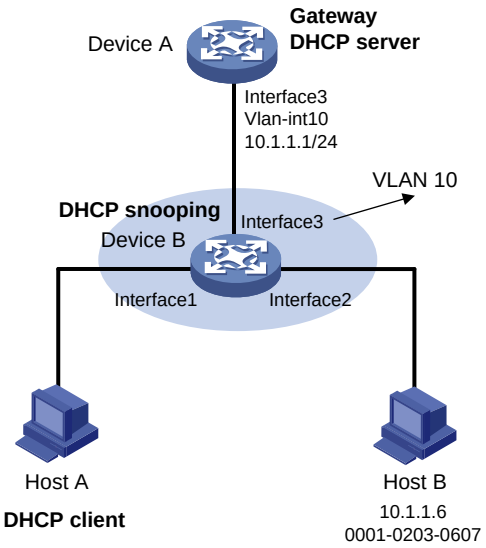


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure VLAN 10, add interfaces to VLAN 10, and specify the IP address of VLAN-interface 10 on Device A. (Details not shown.)
2. Configure the DHCP server on Device A, and configure DHCP address pool 0.

```
<DeviceA> system-view
[DeviceA] dhcp enable
[DeviceA] dhcp server ip-pool 0
[DeviceA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
```

3. Configure Host A (DHCP client) and Host B. (Details not shown.)
4. Configure Device B:

   # Enable DHCP snooping, and configure GigabitEthernet 1/0/3 as a DHCP trusted interface.

```
<DeviceB> system-view
[DeviceB] dhcp snooping enable
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] dhcp snooping trust
[DeviceB-GigabitEthernet1/0/3] quit
```

   # Enable ARP attack detection for user validity check.

```
[DeviceB] vlan 10
[DeviceB-vlan10] arp detection enable
```

   # Configure GigabitEthernet 1/0/3 as an ARP trusted interface.

```
[DeviceB-vlan10] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] arp detection trust
[DeviceB-GigabitEthernet1/0/3] quit
```

Configure a static IP source guard entry on interface GigabitEthernet 1/0/2.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ip source binding ip-address 10.1.1.6 mac-address
0001-0203-0607 vlan 10
[DeviceB-GigabitEthernet1/0/2] quit
```

Enable ARP packet validity check by checking the MAC addresses and IP addresses of ARP packets.

```
[DeviceB] arp detection validate dst-mac ip src-mac
```

Configure port isolation.

```
[DeviceB] port-isolate group 1
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port-isolate enable group 1
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port-isolate enable group 1
[DeviceB-GigabitEthernet1/0/2] quit
```

After the configurations are completed, Device B first checks the validity of ARP packets received on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2. If the ARP packets are confirmed valid, Device B performs user validity check by using the static IP source guard bindings and finally DHCP snooping entries. However, ARP broadcast requests sent from Host A can pass the check on Device B and reach Host B. Port isolation fails.

Enable ARP restricted forwarding.

```
[DeviceB] vlan 10
[DeviceB-vlan10] arp restricted-forwarding enable
[DeviceB-vlan10] quit
```

Verifying the configuration

Verify that Device B forwards ARP broadcast requests from Host A to Device A through the trusted interface GigabitEthernet 1/0/3. Host B cannot receive such packets. Port isolation operates correctly.

Configuring ARP scanning and fixed ARP

About this task

ARP scanning is typically used together with the fixed ARP feature in small-scale and stable networks.

ARP scanning automatically creates ARP entries for devices in an address range. The device performs ARP scanning in the following steps:

1. Sends ARP requests for each IP address in the address range.
2. Obtains their MAC addresses through received ARP replies.
3. Creates dynamic ARP entries.

You can manually trigger ARP scanning or enable automatic ARP scanning. Automatic ARP scanning periodically sends ARP requests for all IP addresses in the range at the specified rate set by using the **arp scan auto send-rate** command.

Fixed ARP converts existing dynamic ARP entries (including those generated through ARP scanning) to static ARP entries. These static ARP entries are of the same attributes as the ARP entries that are manually configured. This feature prevents ARP entries from being modified by attackers.

You can set the ARP packet sending rate if the scanning range has a large number of IP addresses. This setting can avoid high CPU usage and heavy network load caused by a burst of ARP traffic.

Restrictions and guidelines

IP addresses in existing ARP entries are not scanned.

If you trigger ARP scanning and enable automatic ARP scanning, both of them take effect. As a best practice, use automatic ARP scanning only on networks where users come online and go offline frequently.

Due to the limit on the total number of static ARP entries, some dynamic ARP entries might fail the conversion.

The **arp fixup** command is a one-time operation. You can use this command again to convert the dynamic ARP entries learned later to static.

To delete a static ARP entry converted from a dynamic one, use the **undo arp** *ip-address* [*vpn-instance-name*] command. You can also use the **reset arp all** command to delete all ARP entries or the **reset arp static** command to delete all static ARP entries.

Triggering ARP scanning

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Trigger ARP scanning.
   arp scan [*start-ip-address* **to** *end-ip-address*] [**send-rate** { **ppm** *ppm* | *pps* }]

CAUTION:

ARP scanning will take some time and occupy a lot of system and network resources. To stop an ongoing scan, press **Ctrl** + **C**. Dynamic ARP entries are created based on ARP replies received before the scan is terminated.

Configuring automatic ARP scanning

1. Enter system view.
   system-view
2. Set the ARP packet sending rate for automatic ARP scanning.
   arp scan auto send-rate { **ppm** *ppm* | **pps** *pps* }
   By default, the device sends ARP packets at the rate of 48 pps during automatic ARP scanning.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable automatic ARP scanning.
   arp scan auto enable [*start-ip-address* **to** *end-ip-address*]
   By default, automatic ARP scanning is disabled.

Configuring fixed ARP

1. Enter system view.
   system-view
2. Convert existing dynamic ARP entries to static ARP entries.
   arp fixup

Configuring ARP keepalive entry scanning

About ARP keepalive entry scanning

In a large-scale network, it takes a long time for ARP scanning to identify the hosts that go offline abnormally if you specify a large scanning range. After you enable ARP keepalive entry scanning, the system can quickly locate those hosts and monitor the host status within the aging time.

You can set the aging time for ARP keepalive entries in both system view and interface view. The aging time set in interface view takes precedence over the aging time set in system view. In interface view, the default aging time for ARP keepalive entries is the aging time set in system view.

When users come online, the system generates dynamic APR entries and IPSG binding entries. Enabled with ARP keepalive entry scanning, the system also creates online keepalive entries when users come online. If users go offline, the corresponding APR entries are deleted and the status of the keepalive entries is set to offline. The device sends ARP requests at intervals to the IP addresses in the offline keepalive entries until the keepalive entries become online again or are deleted when the aging time expires. The interval varies with the number of ARP requests that have been sent to the IP address in an offline keepalive entry. For more information about IP source guard configuration, see *Security Configuration Guide*.

Configuring ARP keepalive entry scanning

1. Enter system view.
   system-view
2. (Optional.) Set the ARP request sending rate for keepalive entry scanning.
   arp scan keepalive send-rate *pps*
   By default, the device sends ARP requests at a rate of 48 pps during keepalive entry scanning.
3. (Optional.) Set the aging time for ARP keepalive entries.
   arp scan keepalive aging-time *time*
   By default, the aging time for ARP keepalive entries in system view is 60 minutes. The aging time for ARP keepalive entries in interface view is the aging time set in system view.
4. Enter interface view.
   interface *interface-type interface-number*
5. Enable the ARP keepalive entry scanning feature.
   arp scan keepalive enable
   By default, the ARP keepalive entry scanning is disabled on an interface.

Display and maintenance commands for ARP keepalive entry scanning

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ARP keepalive entries.	**display arp scan keepalive entry** [**interface** *interface-type interface-number*] [**count**]
Display statistics about ARP requests sent to the IP addresses in offline keepalive entries for an interface.	**display arp scan keepalive statistics** [**slot** *slot-number*] [**interface** *interface-type interface-number*]
Clear statistics about ARP requests sent to the IP addresses in offline keepalive entries.	**reset arp scan keepalive statistics** [**slot** *slot-number*]

Configuring ARP gateway protection

About ARP gateway protection

Configure this feature on interfaces not connected with a gateway to prevent gateway spoofing attacks.

When such an interface receives an ARP packet, it checks whether the sender IP address in the packet is consistent with that of any protected gateway. If yes, it discards the packet. If not, it handles the packet correctly.

If you configure ARP gateway protection in Ethernet service instance view, this feature filters out ARP packets with IP addresses of protected gateways for the Ethernet service instance. For more information about Ethernet service instances, see EVPN VXLAN configuration in *EVPN Configuration Guide*.

Restrictions and guidelines

You can enable ARP gateway protection for a maximum of eight gateways on an interface or in an Ethernet service instance.

Do not configure both the **arp filter source** and **arp filter binding** commands on an interface. You can configure both the **arp filter binding** command on an interface and the **arp filter source** command in an Ethernet service instance.

If ARP gateway protection works with ARP attack detection, MFF, ARP snooping, and ARP fast-reply, ARP gateway protection applies first.

Procedure

1. Enter system view.
   system-view
2. Enter interface view or Ethernet service instance view.
   - Enter interface view.
     interface *interface-type interface-number*
     Supported interface types include Layer 2 Ethernet interface and Layer 2 aggregate interface.
   - Enter Ethernet service instance view.
     interface *interface-type interface-number*
     service-instance *instance-id*

3. Enable ARP gateway protection for the specified gateway.

   arp filter source *ip-address*

   By default, ARP gateway protection is disabled.

Example: Configuring ARP gateway protection

Network configuration

As shown in Figure 8, Host B launches gateway spoofing attacks to Device B. As a result, traffic that Device B intends to send to Device A is sent to Host B.

Configure Device B to block such attacks.

Figure 8 Network diagram

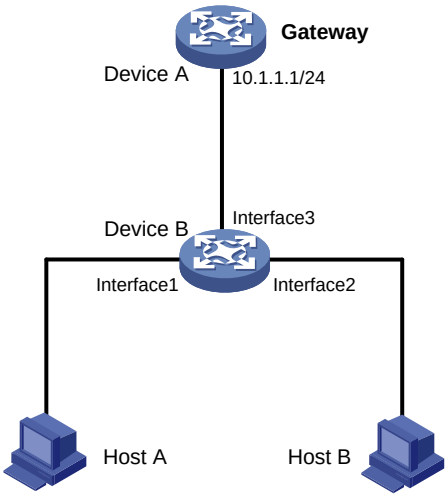


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Configure ARP gateway protection on Device B.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] arp filter source 10.1.1.1
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] arp filter source 10.1.1.1
```

Verifying the configuration

Verify that GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 discard the incoming ARP packets whose sender IP address is the IP address of the gateway.

Configuring ARP filtering

About ARP filtering

The ARP filtering feature can prevent gateway spoofing and user spoofing attacks.

An interface enabled with this feature checks the sender IP and MAC addresses in a received ARP packet against permitted entries. If a match is found, the packet is handled correctly. If not, the packet is discarded.

Restrictions and guidelines

You can configure a maximum of eight permitted entries on an interface.

Do not configure both the **arp filter source** and **arp filter binding** commands on an interface.

If ARP filtering works with ARP attack detection, MFF, ARP snooping, and ARP fast-reply, ARP filtering applies first.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
   Supported interface types include Ethernet interface and Layer 2 aggregate interface.
3. Enable ARP filtering and configure a permitted entry.
   arp filter binding *ip-address mac-address*
   By default, ARP filtering is disabled.

Example: Configuring ARP filtering

Network configuration

As shown in Figure 9, the IP and MAC addresses of Host A are 10.1.1.2 and 000f-e349-1233, respectively. The IP and MAC addresses of Host B are 10.1.1.3 and 000f-e349-1234, respectively.

Configure ARP filtering on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 of Device B to permit ARP packets from only Host A and Host B.

Figure 9 Network diagram

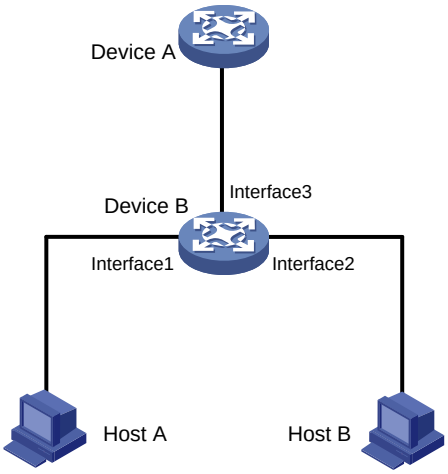


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Configure ARP filtering on Device B.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] arp filter binding 10.1.1.2 000f-e349-1233
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] arp filter binding 10.1.1.3 000f-e349-1234
```

Verifying the configuration

Verify that GigabitEthernet 1/0/1 permits ARP packets from Host A and discards other ARP packets.

Verify that GigabitEthernet 1/0/2 permits ARP packets from Host B and discards other ARP packets.

Configuring ARP sender IP address checking

About ARP sender IP address checking

This feature allows a gateway to check the sender IP address of an ARP packet in a VLAN before ARP learning. If the sender IP address is within the allowed IP address range, the gateway continues ARP learning. If the sender IP address is out of the range, the gateway determines the ARP packet as an attack packet and discards it.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable the ARP sender IP address checking feature and specify the IP address range.
   arp sender-ip-range *start-ip-address end-ip-address*
   By default, the ARP sender IP address checking feature is disabled.

Configuring ARP entry fixing

About this task

Upon receiving the first ARP packet from a user, the device adds an ARP entry to the ARP table. If it receives another ARP packet from the same user, it dynamically refreshes the ARP entry based on the new ARP packet. To prevent attackers from interrupting legitimate user services by sending forged ARP packets that tamper with their ARP entries, enable this feature.

ARP entry fixing supports the following modes:

- **all**—Fixes all information in ARP entries. When the device receives an ARP packet in which the IP address is the same as an ARP entry, it works as follows:
  - If the MAC address, output interface, or VLAN in the ARP packet does not match the ARP entry, the device directly drops the packet and does not update the ARP entry.
  - If the MAC address, output interface, and VLAN in the ARP packet match the ARP entry, the device updates the aging time of the ARP entry.
- **mac**—Fixes only the MAC addresses in ARP entries. When the device receives an ARP packet in which the IP address is the same as an ARP entry, it works as follows:
  - If the MAC address in the ARP packet does not match the ARP entry, the device directly drops the packet and does not update the ARP entry.
  - If the MAC address in the ARP packet matches the ARP entry, the device updates the output interface, VLAN, and aging time of the ARP entry.

Restrictions and guidelines

The global ARP entry fixing feature takes effect on all interfaces.

An interface is enabled with ARP entry fixing once you enable ARP entry fixing in system view or the view of that interface.

To enable ARP entry fixing on a specific interface, make sure global ARP entry fixing is disabled..

If global ARP entry fixing and interface-specific ARP entry fixing are in different modes, the interface-specific ARP entry fixing configuration takes effect.

Enabling global ARP entry fixing

1. Enter system view.
   system-view
2. Enable global ARP entry fixing.
   arp anti-attack entry-fixing { **all** | **mac** } **enable**
   By default, global ARP entry fixing is disabled.

Enabling ARP entry fixing on an interface

1. Enter system view.

system-view

2. (Optional.) Disable global ARP entry fixing.
   undo arp anti-attack entry-fixing { **all** | **mac** } **enable**
   By default, global ARP entry fixing is disabled.
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable ARP entry fixing on the interface.
   arp anti-attack entry-fixing { **all** | **mac** } **enable**
   By default, ARP entry fixing is disabled on an interface.

Enabling SNMP notifications for ARP

About this task

Enable SNMP notifications for ARP as required.

- If you enable ARP active acknowledgement notifications, the device sends a notification to the SNMP module when it does not establish an ARP entry due to active acknowledgement. The notification includes the sender IP address in the received ARP packet and the interface that receives the ARP packet.
- If you enable rate limit notifications for sending ARP Miss messages and ARP packets, the device sends the highest threshold-crossed rate as a notification to the SNMP module when the sending rate exceeds the custom threshold.
- If you enable notifications for packet dropping due to ARP attack detection, the device sends a notification to the SNMP module when a packet is dropped by ARP attack detection. The notification includes the VLAN ID, service ID, VSI, and drop reason for the dropped packet.
- If you enable ARP entry modification notifications, the device sends a notification to the SNMP module when it detects that the ARP entry for a user might be changed to the attacker's ARP entry. The notification includes the sender IP and MAC addresses in the ARP attack packet.
- If you enable ARP entry limit notifications, the device sends the current number of ARP entries as a notification to the SNMP module when the number of global ARP entries exceeds the alarm threshold.
- If you enable ARP gateway protection notifications, the device sends a notification to the SNMP module when it is attacked by gateway spoofing attacks. The notification includes the sender IP and MAC addresses in the ARP attack packet dropped by ARP gateway protection, and index of the interface that receives the ARP attack packet. When an ARP packet with the same sender IP address triggers the SNMP notification again within 3 minutes, the device sends a recovery notification to the SNMP module.
- If you enable endpoint and local device conflict notifications, the device sends a notification to the SNMP module when an endpoint and local device conflict occurs. The notification includes the sender IP address, sender MAC address, target IP address, and target MAC address in the conflicting ARP packet.
- If you enable MAC address inconsistency notifications, the device sends a notification to the SNMP module when the MAC address in the configured static ARP entry and user's actual MAC address are inconsistent. The notification includes the IP address, VPN instance, and MAC address in the ARP entry.
- An invalid ARP packet contains an invalid hardware address type, protocol address type, hardware address length, protocol address length, operation code, or VLAN. For example, when the operation code value is not 1 (ARP request) or 2 (ARP reply), the operation code is invalid and the ARP packet is invalid. If you enable invalid ARP packet check notifications, the device sends a notification to the SNMP module when it receives an invalid ARP packet. The

notification includes the name of the interface that receives the ARP packet, sender MAC and IP addresses, user VLAN ID, and VLAN ID of the service provider.

- If you enable ARP packet queue limit notifications, the device sends the number of packets in the queue as a notification to the SNMP module when the number exceeds 80% of the upper limit.
- If you enable user IP address conflict notifications, the device sends a notification to the SNMP module when a user IP address conflict occurs. The notification includes the sender IP and MAC addresses in the conflicting ARP packet, and MAC address in the corresponding local ARP entry. For more information about enabling recording user IP address conflicts, see ARP configuration in *Layer 3—IP Services Configuration Guide*.
- If you enable user port migration notifications, the device sends a notification to the SNMP module when a user port changes. The notification includes the IP address, MAC address, port before migration, and port after migration of the user. For more information about enabling recording user port migrations, see *ARP configuration in Layer 3—IP Services Configuration Guide*.

For ARP event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for ARP.

   snmp-agent trap enable arp [**active-ack** | **arp-miss** | **detection-drop** | **entry-check** | **entry-limit** | **gateway-check** | **local-conflict** | **mac-mismatch** | **packet-check** | **pkt-threshold** | **rate-limit** | **user-ip-conflict** | **user-move**] *

   By default, SNMP notifications for ARP are disabled.

   If you do not specify any keywords, this command enables all SNMP notifications for ARP.

Contents

Configuring ND attack defense 1

About ND attack defense 1
ND attack defense tasks at a glance 1
Configuring ND packet rate limit 2
Enabling source MAC consistency check for ND messages 2
Configuring ND attack detection 3
About ND attack detection 3
Restrictions and guidelines for ND attack detection configuration 4
Configuring ND attack detection for a VLAN 4
Configuring ND attack detection for a VSI 5
Ignoring ingress ports of ND packets 5
Enabling ND attack detection logging 6
Display and maintenance commands for ND attack detection 6
Example: Configuring ND attack detection 6
Configuring RA guard 8
About RA guard 8
Specifying the role of the attached device 9
Configuring and applying an RA guard policy 9
Enabling the RA guard logging feature 10
Display and maintenance commands for RA guard 10
Example: Configuring RA guard 11
Configuring ND keepalive entry scanning 12
About ND keepalive entry scanning 12
Configuring ND keepalive entry scanning 13
Display and maintenance commands for ND keepalive entry scanning 13
Enabling SNMP notifications for ND 13
Configuring ND blackhole routing 14

Configuring ND attack defense

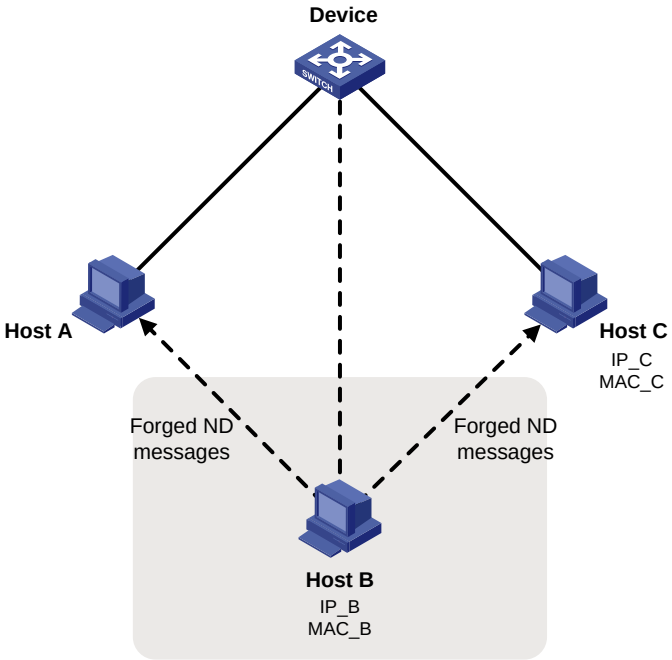
About ND attack defense

IPv6 Neighbor Discovery (ND) attack defense is able to identify forged ND messages to prevent ND attacks.

The IPv6 ND protocol does not provide any security mechanisms and is vulnerable to network attacks. As shown in Figure 1, an attacker can send the following forged ICMPv6 messages to perform ND attacks:

- Forged NS/NA/RS messages with an IPv6 address of a victim host. The gateway and other hosts update the ND entry for the victim with incorrect address information. As a result, all packets intended for the victim are sent to the attacking terminal.
- Forged RA messages with the IPv6 address of a victim gateway. As a result, all hosts attached to the victim gateway maintain incorrect IPv6 configuration parameters and ND entries.

Figure 1 ND attack diagram



ND attack defense tasks at a glance

All ND attack defense tasks are optional.

- Configuring ND packet rate limit
- Enabling source MAC consistency check for ND messages
- Configuring ND attack detection
- Configuring RA guard
- Configuring ND keepalive entry scanning
- Enabling SNMP notifications for ND
- Configuring ND blackhole routing

Configuring ND packet rate limit

About this task

This feature limits the receiving rate of ND packets that are to be delivered to the CPU, preventing the CPU from being overwhelmed by ND packets.

Enabled with ND attack detection, the device redirects the received ND packets to the CPU for security check. If attackers send a large number of forged ND packets to the device, the CPU will be overloaded. As a result, other features might malfunction and the device might crash. To avoid this situation, configure ND packet rate limit on ND packet receiving interfaces. When the receiving rate of ND packets on the interfaces exceeds the rate limit, the packets that exceed the limit will be discarded.

When logging for ND packet rate limit is enabled, the device sends the highest threshold-crossed ND packet rate within the sending interval in a log message to the information center. You can configure the information center module to set the log output rules. For more information about information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

As a best practice, configure this feature if the device is configured with ND attack detection or ND snooping or has detected ND flood attacks.

To prevent the device from sending log messages frequently, set the log message sending interval to a high value.

On a Layer 2 aggregate interface, if you enable ND packet rate limit and the logging for this feature, the device sends log messages upon limit violations on its member ports.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Enable logging for ND packet rate limit.

   ipv6 nd rate-limit log enable

   By default, logging for ND packet rate limit is disabled.

3. (Optional.) Set the log message sending interval.

   ipv6 nd rate-limit log interval *interval*

   By default, the device sends log messages every 60 seconds.

4. Enter interface view.

   interface *interface-type interface-number*

   Supported interfaces include Layer 2 Ethernet interfaces, Layer 2 aggregate interfaces, Layer 3 Ethernet interfaces, and Layer 3 aggregate interfaces.

5. Enable ND packet rate limit.

   ipv6 nd rate-limit [*pps*]

   By default, ND packet rate limit is enabled.

Enabling source MAC consistency check for ND messages

About this task

The source MAC consistency check feature is typically configured on gateways to prevent ND attacks.

This feature checks the source MAC address and the source link-layer address for consistency for each arriving ND message.

- If the source MAC address and the source link-layer address are not the same, the device drops the packet.
- If the addresses are the same, the device continues learning ND entries.

The ND logging feature logs source MAC inconsistency events, and it sends the log messages to the information center. The information center can then output log messages from different source modules to different destinations. For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Enable source MAC consistency check for ND messages.
   `ipv6 nd mac-check enable`
   By default, source MAC consistency check is disabled for ND messages.
3. (Optional.) Enable the ND logging feature.
   `ipv6 nd check log enable`
   By default, the ND logging feature is disabled.
   As a best practice, disable the ND logging feature to avoid excessive ND logs.

Configuring ND attack detection

About ND attack detection

ND attack detection checks incoming ND messages for user validity to prevent spoofing attacks. It is typically configured on access devices.

ND attack detection defines the following types of interfaces on a VLAN network:

- **ND trusted interface**—Performs no user validity check on the received ND messages.
- **ND untrusted interface**—Discards incoming RA and redirect messages, and performs the user validity check on other types of incoming ND messages.

ND attack detection compares the source IPv6 address and the source MAC address in an incoming ND message against security entries from other modules.

- If a match is found, the device verifies the user as legal, and it forwards the ND message.
- If no match is found, the device verifies the user as illegal, and it discards the ND message.

The security entries used by ND attack detection for user validity check vary by the specified keyword when you enable this feature.

- If you do not specify the **`dhcpv6-relay`** or **`nd-ra-prefix`** keyword, ND attack detection uses static IPv6 source guard binding entries, ND snooping entries, and DHCPv6 snooping entries for user validity check.
- If you specify the **`dhcpv6-relay`** keyword, ND attack detection uses static IPv6 source guard binding entries, ND snooping entries, DHCPv6 snooping entries, and DHCPv6 relay agent entries for user validity check.
- If you specify the **`nd-ra-prefix`** keyword, ND attack detection uses static IPv6 source guard binding entries, ND snooping entries, DHCPv6 snooping entries, and IPv6 source guard binding entries with prefixes advertised in RA messages for user validity check.

The entries are generated as follows:

- Static IPv6 source guard binding entries are created by using the **ipv6 source binding** command. For information about IPv6 source guard, see "Configuring IP source guard."
- ND snooping entries are automatically generated by the ND snooping feature. For information about ND snooping, see IPv6 basics configuration in *Layer 3–IP Services Configuration Guide*.
- DHCPv6 snooping entries are automatically generated by the DHCPv6 snooping feature. For more information, see DHCPv6 configuration in *Layer 3–IP Services Configuration Guide*.
- DHCPv6 relay entries are automatically generated by the DHCPv6 relay entry recording feature. For more information, see DHCPv6 configuration in *Layer 3–IP Services Configuration Guide*.
- IPv6 source guard binding entries with prefixes advertised in RA messages are created by using the **ipv6 nd ra prefix source binding** command. For more information, see IPv6 basics configuration in *Layer 3–IP Services Configuration Guide*.

Restrictions and guidelines for ND attack detection configuration

When you configure ND attack detection, follow these restrictions and guidelines:

- To prevent ND untrusted interfaces from dropping all received ND messages, make sure one or more of the these features are configured: IPv6 source guard static bindings, DHCPv6 snooping, and ND snooping.
- To make the IPv6 source guard static bindings effective for ND attack detection, you must perform the following operations:
  - Specify the **vlan** *vlan-id* option in the **ipv6 source binding** command.
  - Enable ND attack detection for the same VLAN.

Configuring ND attack detection for a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable ND attack detection.
   ipv6 nd detection enable [**dhcpv6-relay** | **nd-ra-prefix**]
   By default, ND attack detection is disabled.
4. (Optional.) Configure the interface as ND trusted interface:
   a. Return to system view.
      quit
   b. Enter Layer 2 Ethernet or aggregate interface view.
      interface *interface-type interface-number*
   c. Configure the interface as ND trusted interface.
      ipv6 nd detection trust
      By default, all interfaces are ND untrusted interfaces.

Configuring ND attack detection for a VSI

Restrictions and guidelines

Do not configure ND snooping in a VSI where ND attack detection is configured. If they are both configured in one VSI, ND snooping cannot learn snooping entries.

Procedure

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enable ND attack detection.
   ipv6 nd detection enable
   By default, ND attack detection is disabled.
4. (Optional.) Configure the ND trusted AC.
   a. Return to system view.
      quit
   b. Enter interface view.
      interface *interface-type interface-number*
   c. Enter Ethernet service instance view.
      service-instance *instance-id*
   d. (Optional.) Configure the AC as ND trusted AC.
      ipv6 nd detection trust
      By default, all ACs are ND untrusted ACs.

Ignoring ingress ports of ND packets

About this task

With ND attack detection enabled, the device can perform security check on received packets based on the local and remote IPSG bindings. Remote IPSG bindings do not contain port information. The device drops ND packets that match remote IPSG bindings because it does not find matching ingress ports for these packets. To prevent the device from dropping these packets, you can configure the device to ignore ingress ports of ND packets. This feature does not examine the ingress ports of ND packets, so that ND packets that match remote IPSG bindings will not be dropped.

Procedure

1. Enter system view.
   system-view
2. Ignore ingress ports of ND packets in ND attack detection.
   ipv6 nd detection port-match-ignore
   By default, ingress ports of ND packets are examined in ND attack detection.

Enabling ND attack detection logging

About this task

This feature allows a device to generate logs when it detects invalid ND packets. The log information helps administrators locate and solve problems. Each log records the following information:

- Victim port numbers in a VLAN.
- Source IP address of the invalid ND packets.
- Source MAC address of the invalid ND packets.
- VLAN ID of the invalid ND packets.
- Total number of dropped ND packets.

Procedure

1. Enter system view.
   `system-view`
2. Enable ND attack detection logging.
   `ipv6 nd detection log enable`
   By default, ND attack detection logging is disabled.

Display and maintenance commands for ND attack detection

Execute **`display`** commands in any view and **`reset`** commands in user view.

Task	Command
Display statistics for ND messages dropped by ND attack detection.	**`display ipv6 nd detection statistics`** [**`interface`** *`interface-type interface-number`*]
Clear ND attack detection statistics.	**`reset ipv6 nd detection statistics`** [**`interface`** *`interface-type interface-number`*]

Example: Configuring ND attack detection

Network configuration

As shown in Figure 2, configure ND attack detection on Device B to check user validity for ND messages from Host A and Host B.

Figure 2 Network diagram

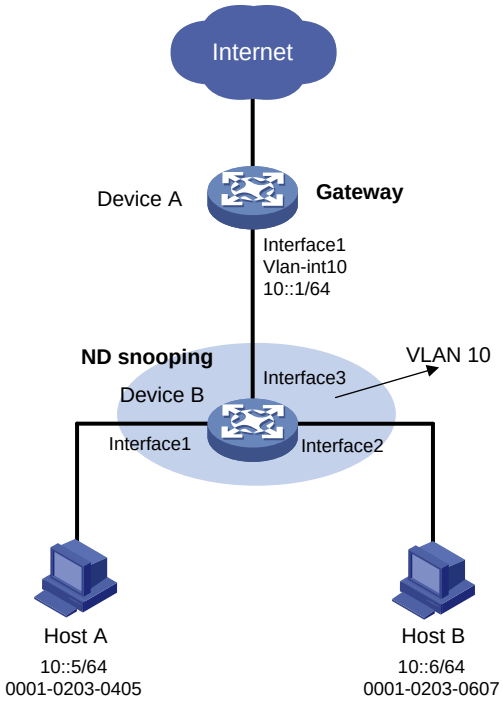


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Device A:

   # Create VLAN 10.

   ```
   <DeviceA> system-view
   [DeviceA] vlan 10
   [DeviceA-vlan10] quit
   ```

   # Configure GigabitEthernet 1/0/1 to trunk VLAN 10.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] port link-type trunk
   [DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 10
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

   # Assign IPv6 address 10::1/64 to VLAN-interface 10.

   ```
   [DeviceA] interface vlan-interface 10
   [DeviceA-Vlan-interface10] ipv6 address 10::1/64
   [DeviceA-Vlan-interface10] quit
   ```

2. Configure Device B:

   # Create VLAN 10.

   ```
   <DeviceB> system-view
   [DeviceB] vlan 10
   ```

```
[DeviceB-vlan10] quit
```

Configure GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 to trunk VLAN 10.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type access
[DeviceB-GigabitEthernet1/0/1] port access vlan 10
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type access
[DeviceB-GigabitEthernet1/0/2] port access vlan 10
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan 10
[DeviceB-GigabitEthernet1/0/3] quit
```

Enable ND attack detection for VLAN 10.

```
[DeviceB] vlan 10
[DeviceB-vlan10] ipv6 nd detection enable
```

Enable ND snooping for IPv6 global unicast addresses and ND snooping for IPv6 link-local addresses in VLAN 10.

```
[DeviceB-vlan10] ipv6 nd snooping enable global
[DeviceB-vlan10] ipv6 nd snooping enable link-local
[DeviceB-vlan10] quit
```

Configure GigabitEthernet 1/0/3 as ND trusted interface.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] ipv6 nd detection trust
```

Verifying the configuration

Verify that Device B inspects all ND messages received by GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 based on the ND snooping entries. (Details not shown.)

Configuring RA guard

About RA guard

RA guard allows Layer 2 access devices to analyze and block unwanted and forged RA messages.

Upon receiving an RA message, the device makes the forwarding or dropping decision based on the role of the attached device or the RA guard policy.

1. If the role of the device attached to the receiving interface is **router**, the device forwards the RA message. If the role is **host**, the device drops the RA message.
2. If no attached device role is set, the device uses the RA guard policy applied to the VLAN of the receiving interface to match the RA message.
   - If the policy does not contain match criteria, the policy will not take effect and the device forwards the RA message.
   - If the RA message content matches every criterion in the policy, the device forwards the message. Otherwise, the device drops the message.

Specifying the role of the attached device

Restrictions and guidelines

Make sure your setting is consistent with the type of the attached device. If you are not aware of the device type, do not specify a role for the device.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Specify the role of the device attached to the interface.
   ipv6 nd raguard role { **host** | **router** }
   By default, the role of the device attached to the interface is not specified.

Configuring and applying an RA guard policy

About this task

Configure an RA guard policy if you do not specify a role for the attached device or if you want to filter the RA messages sent by a router.

Procedure

1. Enter system view.
   system-view
2. Create an RA guard policy and enter its view.
   ipv6 nd raguard policy *policy-name*
3. Configure the RA guard policy.
   Choose the following tasks as needed:
   - Specify an ACL match criterion.
     if-match acl { *ipv6-acl-number* | **name** *ipv6-acl-name* }
   - Specify a prefix match criterion.
     if-match prefix acl { *ipv6-acl-number* | **name** *ipv6-acl-name* }
   - Specify a router preference match criterion.
     if-match router-preference maximum { **high** | **low** | **medium** }
   - Specify an M flag match criterion.
     if-match autoconfig managed-address-flag { **off** | **on** }
   - Specify an O flag match criterion.
     if-match autoconfig other-flag { **off** | **on** }
   - Specify a maximum or minimum hop limit match criterion.
     if-match hop-limit { **maximum** | **minimum** } *limit*

   By default, the RA guard policy is not configured.
4. Quit RA guard policy view.

quit

5. Enter VLAN view.
   vlan *vlan-number*
6. Apply an RA guard policy to the VLAN.
   ipv6 nd raguard apply policy [*policy-name*]
   By default, no RA guard policy is applied to the VLAN.

Enabling the RA guard logging feature

About this task

This feature allows a device to generate logs when it detects forged RA messages. The log information helps administrators locate and solve problems. Each log records the following information:

- Name of the interface that received the forged RA message.
- Source IP address of the forged RA message.
- Number of RA messages dropped on the interface.

The RA guard logging feature sends the log messages to the information center. The information center can then output log messages from different source modules to different destinations. For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable the RA guard logging feature.
   ipv6 nd raguard log enable
   By default, the RA guard logging feature is disabled.

Display and maintenance commands for RA guard

Execute **display** commands in any view.

Task	Command
Display the RA guard policy configuration.	**display ipv6 nd raguard policy** [*policy-name*]
Display RA guard statistics.	**display ipv6 nd raguard statistics** [**interface** *interface-type interface-number*]
Clear RA guard statistics.	**reset ipv6 nd raguard statistics** [**interface** *interface-type interface-number*]

Example: Configuring RA guard

Network configuration

As shown in Figure 3, GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 of Device B are in VLAN 10.

Configure RA guard on Device B to filter forged and unwanted RA messages.

- Configure an RA policy in VLAN 10 for GigabitEthernet 1/0/2 to filter all RA messages received from the unknown device.
- Specify **host** as the role of the host. All RA messages received on GigabitEthernet 1/0/1 are dropped.
- Specify **router** as the role of the Device A. All RA messages received on GigabitEthernet 1/0/3 are forwarded.

Figure 3 Network diagram

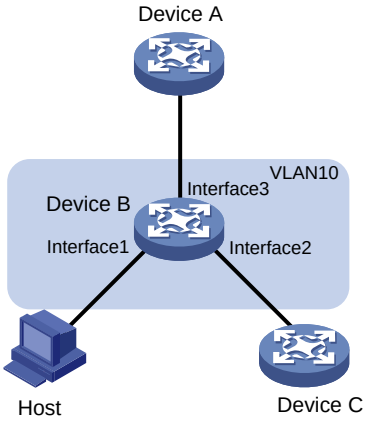


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Create an RA guard policy named **policy1**.

```
<DeviceB> system-view
[DeviceB] ipv6 nd raguard policy policy1
```

Set the maximum router preference to **high** for the RA guard policy.

```
[DeviceB-raguard-policy-policy1] if-match router-preference maximum high
```

Specify **on** as the M flag match criterion for the RA guard policy.

```
[DeviceB-raguard-policy-policy1] if-match autoconfig managed-address-flag on
```

Specify **on** as the O flag match criterion for the RA guard policy.

```
[DeviceB-raguard-policy-policy1] if-match autoconfig other-flag on
```

Set the maximum advertised hop limit to 120 for the RA guard policy.

```
[DeviceB-raguard-policy-policy1] if-match hop-limit maximum 120
```

Set the minimum advertised hop limit to 100 for the RA guard policy.

```
[DeviceB-raguard-policy-policy1] if-match hop-limit minimum 100
[DeviceB-raguard-policy-policy1] quit
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to VLAN 10.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type access
[DeviceB-GigabitEthernet1/0/1] port access vlan 10
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type access
[DeviceB-GigabitEthernet1/0/2] port access vlan 10
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 to trunk VLAN 10.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan 10
[DeviceB-GigabitEthernet1/0/3] quit
```

Apply the RA guard policy **policy1** to VLAN 10.

```
[DeviceB] vlan 10
[DeviceB-vlan10] ipv6 nd raguard apply policy policy1
[DeviceB-vlan10] quit
```

Specify **host** as the role of the device attached to GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] ipv6 nd raguard role host
[DeviceB-GigabitEthernet1/0/1] quit
```

Specify **router** as the role of the device attached to GigabitEthernet 1/0/3.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] ipv6 nd raguard role router
[DeviceB-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Verify that the device forwards or drops RA messages received on GigabitEthernet 1/0/2 based on the RA guard policy. (Details not shown.)

Verify that the device drops RA messages received on GigabitEthernet 1/0/1. (Details not shown.)

Verify that the device forwards RA messages received on GigabitEthernet 1/0/3 to other interfaces in VLAN 10. (Details not shown.)

Configuring ND keepalive entry scanning

About ND keepalive entry scanning

In a large-scale network, it takes a long time for ND scanning to identify the hosts that go offline abnormally if you specify a large scanning range. After you enable ND keepalive entry scanning, the system can quickly locate those hosts and monitor the host status within the aging time.

You can set the aging time for ND keepalive entries in both system view and interface view. The aging time set in interface view takes precedence over the aging time set in system view. In interface view, the default aging time for ND keepalive entries is the aging time set in system view.

When users come online, the system generates dynamic ND entries and IPSG binding entries. Enabled with ND keepalive entry scanning, the system also creates online keepalive entries when users come online. If users go offline, the corresponding ND entries are deleted and the status of the keepalive entries is set to offline. The device sends NS packets at intervals to the IPv6 addresses in the offline keepalive entries until the keepalive entries become online again or are deleted when the aging time expires. The interval varies with the number of NS packets that have been sent to the IPv6 address in an offline keepalive entry. For more information about IP source guard configuration, see *Security Configuration Guide*.

Configuring ND keepalive entry scanning

1. Enter system view.
   system-view
2. (Optional.) Set the NS packet sending rate for keepalive entry scanning.
   ipv6 nd scan keepalive send-rate *pps*
   By default, the device sends NS packets at a rate of 48 pps during keepalive entry scanning.
3. (Optional.) Set the aging time for ND keepalive entries.
   ipv6 nd scan keepalive aging-time *time*
   By default, the aging time for ND keepalive entries in system view is 60 minutes. The aging time for ND keepalive entries in interface view is the aging time set in system view.
4. Enter interface view.
   interface *interface-type interface-number*
5. Enable the ND keepalive entry scanning feature.
   ipv6 nd scan keepalive enable
   By default, the ND keepalive entry scanning is disabled on an interface.

Display and maintenance commands for ND keepalive entry scanning

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ND keepalive entries.	**display ipv6 nd scan keepalive entry** [**interface** *interface-type interface-number*] [**count**]
Display statistics about NS packets sent to the IPv6 addresses in offline keepalive entries.	**display ipv6 nd scan keepalive statistics** [**slot** *slot-number*] [**interface** *interface-type interface-number*]
Clear statistics about NS packets sent to the IPv6 addresses in offline keepalive entries.	**reset ipv6 nd scan keepalive statistics** [**slot** *slot-number*]

Enabling SNMP notifications for ND

About this task

Enable SNMP notifications for ND as required.

- If you enable ND entry limit notifications, the device sends the current ND entry information as a notification to the SNMP module when the number of ND entries exceeds the alarm threshold.
- If you enable endpoint and local device conflict notifications, the device sends a notification to the SNMP module when an endpoint and local device conflict occurs. The notification includes the source IPv6 address, source MAC address, destination IPv6 address, and destination MAC address in the conflicting ND packet.
- If you enable rate limit notifications for sending ND Miss messages and ND packets, the device sends the highest threshold-crossed rate as a notification to the SNMP module. When the device receives an IP packet in which the destination IP address is unresolvable, it sends a ND Miss message to the CPU.
- If you enable ND packet queue limit notifications, the device sends the number of packets in the queue as a notification to the SNMP module when the number exceeds 80% of the upper limit.
- If you enable user IPv6 address conflict notifications, the device sends a notification to the SNMP module when a user IPv6 address conflict occurs. The notification includes the source IPv6 and MAC addresses in the conflicting ND packet, and MAC address in the corresponding local ND entry.

For ND event notifications to be sent correctly, you must also configure SNMP on the device. For more information, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for ND.

   snmp-agent trap enable nd [**entry-limit** | **local-conflict** | **nd-miss** | **pkt-threshold** | **user-ip-conflict**] *

   By default, SNMP notifications for ND are disabled.

   If you do not specify any keywords, this command enables all SNMP notifications for ND.

Configuring ND blackhole routing

About this task

If the device receives a large number of unresolvable IPv6 packets from a host, it frequently sends NS messages and keeps trying to resolve the destination IPv6 addresses, overloading its CPU.

Enabled with ND blackhole routing, the device creates a blackhole route for an unresolved IPv6 address and sends NS messages to the IPv6 address. If the IPv6 address is resolved before the blackhole route ages out, the device immediately deletes the blackhole route and starts forwarding packets destined for the IPv6 address. If the IPv6 address remains unresolvable before the blackhole route is deleted, the device directly drops the packets destined for the IPv6 address.

Procedure

1. Enter system view.

   system-view

2. Enable ND blackhole routing.

   ipv6 nd resolving-route enable

   By default, ND blackhole routing is enabled.

Contents

Configuring uRPF 2
About uRPF 2
uRPF application scenario 2
uRPF check modes 2
uRPF extended functions 3
Network application 3
Restrictions and guidelines: uRPF configuration 3
Enabling uRPF on an interface 4
Display and maintenance commands for uRPF 4
Configuring IPv6 uRPF 4
About IPv6 uRPF 4
IPv6 uRPF application scenario 4
IPv6 uRPF check modes 5
IPv6 uRPF extended functions 5
Network application 5
Restrictions and guidelines: IPv6 uRPF configuration 6
Enabling IPv6 uRPF on an interface 6
Display and maintenance commands for IPv6 uRPF 7

Configuring uRPF

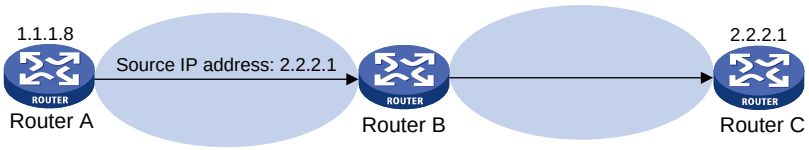
About uRPF

Unicast Reverse Path Forwarding (uRPF) protects a network against source address spoofing attacks, such as DoS and DDoS attacks.

uRPF application scenario

Attackers send packets with a forged source address to access a system that uses IPv4-based authentication, in the name of authorized users or even the administrator. Even if the attackers or other hosts cannot receive any response packets, the attacks are still disruptive to the attacked target.

Figure 1 Source address spoofing attack



As shown in Figure 1, an attacker on Router A sends the server (Router B) requests with a forged source IP address 2.2.2.1 at a high rate. Router B sends response packets to IP address 2.2.2.1 (Router C). Consequently, both Router B and Router C are attacked. If the administrator disconnects Router C by mistake, the network service is interrupted.

Attackers can also send packets with different forged source addresses or attack multiple servers simultaneously to block connections or even break down the network.

uRPF can prevent these source address spoofing attacks. It checks whether an interface that receives a packet is the output interface of the FIB entry that matches the source address of the packet. If not, uRPF considers it a spoofing attack and discards the packet.

uRPF check modes

uRPF supports strict and loose modes.

Strict uRPF check

To pass strict uRPF check, the source address of a packet and the receiving interface must match the destination address and output interface of a FIB entry. In some scenarios (for example, asymmetrical routing), strict uRPF might discard valid packets.

Strict uRPF is often deployed between a PE and a CE.

Loose uRPF check

To pass loose uRPF check, the source address of a packet must match the destination address of a FIB entry. Loose uRPF can avoid discarding valid packets, but might let go attack packets.

Loose uRPF is often deployed between ISPs, especially in asymmetrical routing.

uRPF extended functions

Using the default route in uRPF check

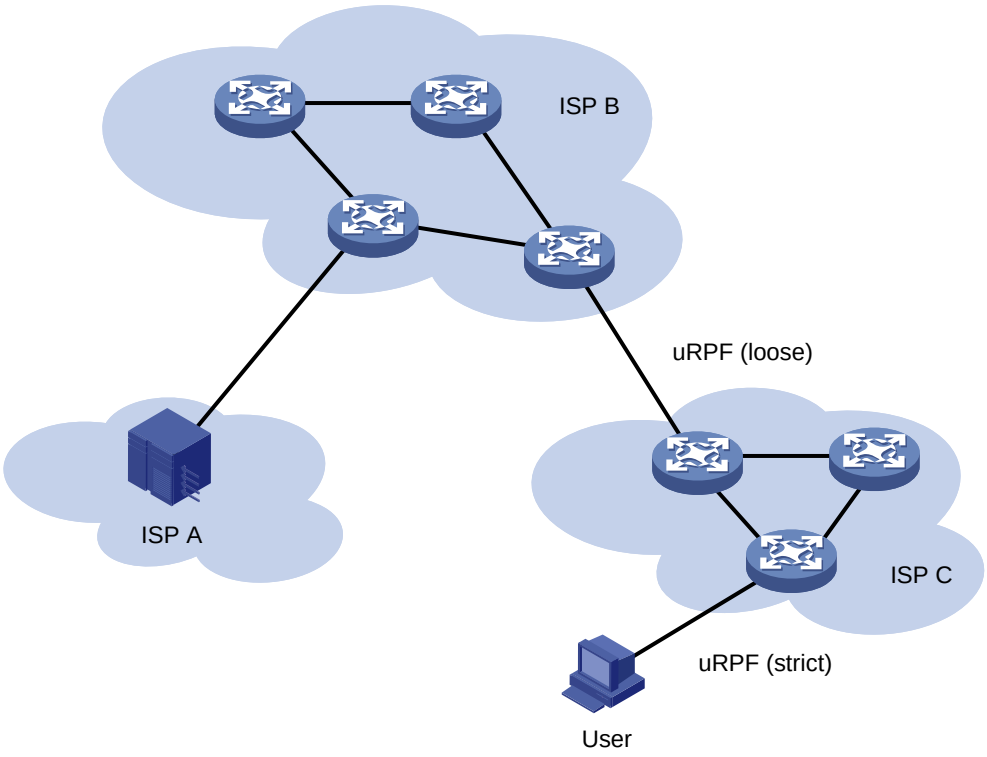
When a default route exists, all packets that fail to match a specific FIB entry match the default route during uRPF check and thus are permitted to pass. To avoid this situation, you can disable uRPF from using any default route to discard such packets. If you allow using the default route (set by using **`allow-default-route`**), uRPF permits packets that only match the default route.

By default, uRPF discards packets that can only match a default route. Typically, you do not need to configure the **`allow-default-route`** keyword on a PE device because it has no default route pointing to the CE. If you enable uRPF on a CE interface and the CE interface has a default route pointing to the PE, specify the **`allow-default-route`** keyword.

Network application

As shown in Figure 2, strict uRPF check is configured between an ISP network and a customer network. Loose uRPF check is configured between ISPs.

Figure 2 Network diagram



Restrictions and guidelines: uRPF configuration

If you enable uRPF on an interface, you can use the **`display ip interface`** command to display statistics about packets discarded by uRPF (displayed as "Drops" and "Suppressed drops").

Do not configure the **`allow-default-route`** keyword for loose uRPF check. Otherwise, uRPF might fail to work.

Do not configure both strict uRPF check and loose IPv6 uRPF check, or loose uRPF check and strict IPv6 uRPF check on an interface.

Do not use strict uRPF if ECMP routing is available in the network. Service packets that travel along ECMP routes cannot pass the strict uRPF check and will be dropped.

Enabling uRPF on an interface

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Enable uRPF.
   ip urpf { **loose** [**allow-default-route**] | **strict** [**allow-default-route**] }
   By default, uRPF is disabled.

Display and maintenance commands for uRPF

Execute **display** commands in any view.

Task	Command
Display uRPF configuration.	**display ip urpf interface** *interface-type interface-number* [**slot** *slot-number*]

Configuring IPv6 uRPF

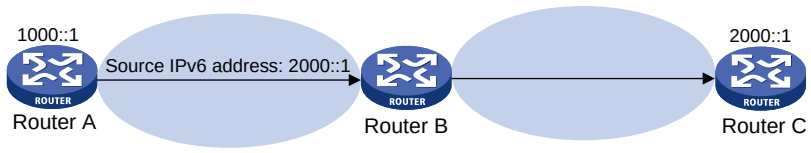
About IPv6 uRPF

IPv6 Unicast Reverse Path Forwarding (uRPF) protects a network against source address spoofing attacks, such as DoS and DDoS attacks.

IPv6 uRPF application scenario

Attackers send packets with a forged source address to access a system that uses IPv6-based authentication, in the name of authorized users or even the administrator. Even if the attackers or other hosts cannot receive any response packets, the attacks are still disruptive to the attacked target.

Figure 3 Source address spoofing attack



As shown in Figure 3, an attacker on Router A sends the server (Router B) requests with a forged source IPv6 address 2000::1 at a high rate. Router B sends response packets to IPv6 address 2000::1 (Router C). Consequently, both Router B and Router C are attacked. If the administrator disconnects Router C by mistake, the network service is interrupted.

Attackers can also send packets with different forged source addresses or attack multiple servers simultaneously to block connections or even break down the network.

IPv6 uRPF can prevent these source address spoofing attacks. It checks whether an interface that receives a packet is the output interface of the FIB entry that matches the source address of the packet. If not, IPv6 uRPF considers it a spoofing attack and discards the packet.

IPv6 uRPF check modes

IPv6 uRPF supports strict and loose check modes.

Strict IPv6 uRPF check

To pass strict IPv6 uRPF check, the source address of a packet and the receiving interface must match the destination address and output interface of an IPv6 FIB entry. In some scenarios (for example, asymmetrical routing), strict IPv6 uRPF might discard valid packets.

Strict IPv6 uRPF is often deployed between a PE and a CE.

Loose IPv6 uRPF check

To pass loose IPv6 uRPF check, the source address of a packet must match the destination address of an IPv6 FIB entry. Loose IPv6 uRPF can avoid discarding valid packets, but might let go attack packets.

Loose IPv6 uRPF is often deployed between ISPs, especially in asymmetrical routing.

IPv6 uRPF extended functions

Using the default route in IPv6 uRPF check

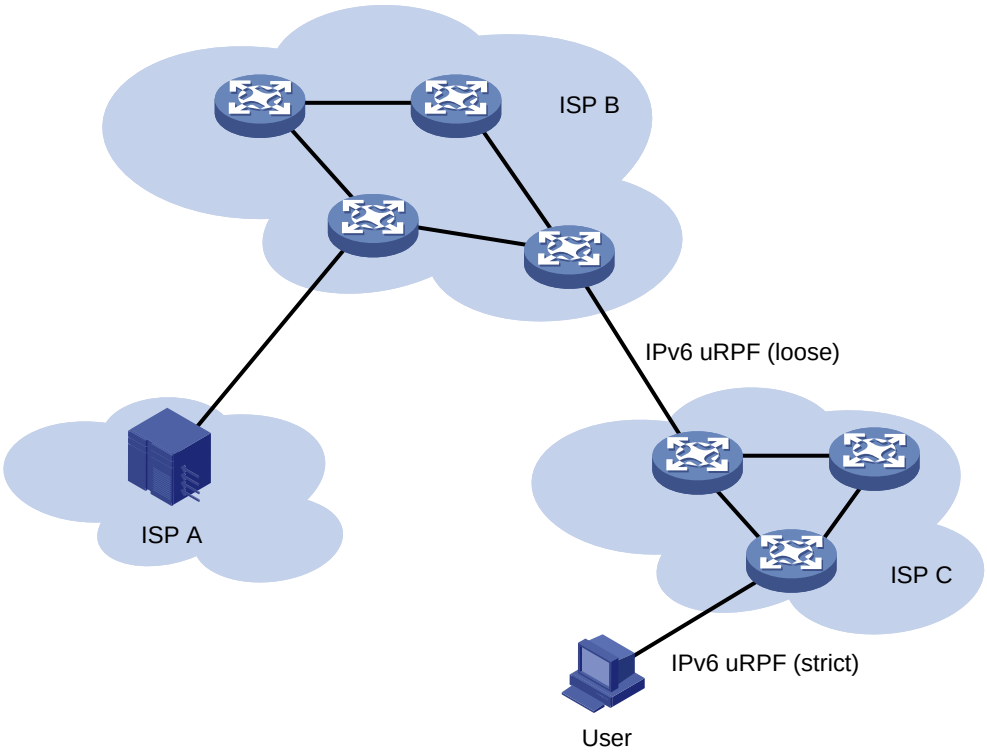
When a default route exists, all packets that fail to match a specific IPv6 FIB entry match the default route during IPv6 uRPF check and thus are permitted to pass. If you allow using the default route (by using **`allow-default-route`**), IPv6 uRPF permits packets that only match the default route.

By default, IPv6 uRPF discards packets that can only match a default route. Typically, you do not need to configure the **`allow-default-route`** keyword on a PE device because it has no default route pointing to the CE device. If you enable IPv6 uRPF on a CE interface and the CE interface has a default route pointing to the PE, specify the **`allow-default-route`** keyword.

Network application

As shown in Figure 4, strict IPv6 uRPF check is configured between an ISP network and a customer network. Loose IPv6 uRPF check is configured between ISPs.

Figure 4 Network diagram



Restrictions and guidelines: IPv6 uRPF configuration

If you enable IPv6 uRPF on an interface, you can use the **display ipv6 interface** command to view statistics about packets discarded by IPv6 uRPF (displayed as "Drops" and "Suppressed drops").

Do not configure the **allow-default-route** keyword for loose IPv6 uRPF check. Otherwise, IPv6 uRPF might fail to work.

Do not configure both strict IPv6 uRPF check and loose uRPF check, or loose IPv6 uRPF check and strict uRPF check on an interface.

Do not use strict IPv6 uRPF if ECMP routing is available in the network. Service packets that travel along ECMP routes cannot pass the strict uRPF check and will be dropped.

Enabling IPv6 uRPF on an interface

1. Enter system view.
   system-view
2. Enter Layer 3 interface view.
   interface *interface-type interface-number*
3. Enable IPv6 uRPF.
   ipv6 urpf { **loose** | **strict** } [**allow-default-route**]
   By default, IPv6 uRPF is disabled.

Display and maintenance commands for IPv6 uRPF

Execute **display** commands in any view.

Task	Command
Display IPv6 uRPF configuration.	**display ipv6 urpf interface** *interface-type interface-number***interface** *interface-type interface-number* [**slot** *slot-number*]

Contents

Configuring MFF 1
About MFF 1
MFF network model 1
Port roles 1
Processing of ARP packets in MFF 2
MFF default gateway 2
Protocols and standards 2
MFF tasks at a glance 2
Enabling MFF 3
Configuring a network port 3
Enabling periodic gateway probe 4
Specifying the IP addresses of servers 4
Display and maintenance commands for MFF 4
MFF configuration examples 5
Example: Configuring MFF in a tree network 5
Example: Configuring MFF in a ring network 6

Configuring MFF

About MFF

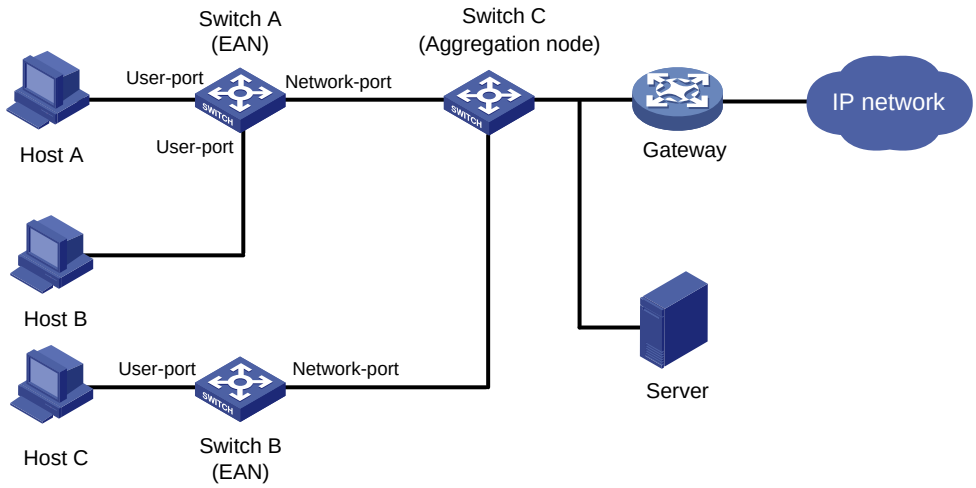
MAC-forced forwarding (MFF) implements Layer 2 isolation and Layer 3 communication between hosts in the same broadcast domain.

An MFF-enabled device intercepts ARP requests and returns the MAC address of a gateway (or server) to the senders. In this way, the senders are forced to send packets to the gateway for traffic monitoring and attack prevention.

MFF network model

As shown in Figure 1, hosts are connected to Switch C through Switch A and Switch B, which are called Ethernet access nodes (EANs). The MFF-enabled EANs forward packets from hosts to the gateway for further forwarding. The hosts are isolated at Layer 2, but they can communicate at Layer 3.

Figure 1 Network diagram for MFF



MFF works with any of the following features to implement traffic filtering and Layer 2 isolation on the EANs:

- ARP snooping (see *Layer 3—IP Services Configuration Guide*).
- IP source guard (see "Configuring IP source guard").
- ARP detection (see "Configuring ARP attack protection").
- VLAN mapping (see *Layer 2—LAN Switching Configuration Guide*).

Port roles

Two types of ports, user port and network port, exist in an MFF-enabled VLAN.

User port

An MFF user port is directly connected to a host and processes the following packets differently:

- Allows multicast packets to pass.
- Delivers ARP packets to the CPU.

- Processes unicast packets as follows:
  - If gateways' MAC addresses have been learned, the user port allows only the unicast packets with the gateways' MAC addresses as the destination MAC addresses to pass.
  - If no gateways' MAC addresses have been learned, the user port discards all received unicast packets.

Network port

An MFF network port is connected to any of the following networking devices:

- An access switch.
- A distribution switch.
- A gateway.
- A server.

A network port processes the following packets differently:

- Allows multicast packets to pass.
- Delivers ARP packets to the CPU.
- Denies broadcast packets other than DHCP and ARP packets.

Processing of ARP packets in MFF

An MFF-enabled device implements Layer 3 communication between hosts by intercepting ARP requests from the hosts and replies with the MAC address of a gateway. This mechanism helps reduce the number of broadcast messages.

The MFF device processes ARP packets as follows:

- After receiving an ARP request from a host, the MFF device sends the MAC address of the corresponding gateway to the host. In this way, hosts in the network have to communicate at Layer 3 through a gateway.
- After receiving an ARP request from a gateway, the MFF device sends the requested host's MAC address to the gateway if the corresponding entry is available. If the entry is not available, the MFF device forwards the ARP request.
- The MFF device forwards ARP replies between hosts and gateways.
- If the source MAC addresses of ARP requests from gateways are different from those recorded, the MFF device updates and broadcasts the IP and MAC addresses of the gateways.

MFF default gateway

MFF applies to only networks where the hosts' IP addresses are manually configured. Because the hosts cannot obtain the gateway information through DHCP, the default gateway must be specified by the **`mac-forced-forwarding default-gateway`** command. MFF maintains only one default gateway for each VLAN. MFF updates the MAC address of the default gateway upon receiving an ARP packet with a different sender MAC address from the default gateway.

Protocols and standards

RFC 4562, *MAC-Forced Forwarding*

MFF tasks at a glance

To configure MFF, perform the following tasks:

1. Enabling MFF
2. Configuring a network port
3. (Optional.) Enabling periodic gateway probe
4. Specifying the IP addresses of servers

   If servers exist in the network, you must perform this task to ensure communication between the servers and hosts.

Enabling MFF

Restrictions and guidelines

- An MFF-enabled device and a host cannot ping each other.
- When MFF works with static IP source guard bindings, you must configure VLAN IDs in the static bindings. Otherwise, IP packets allowed by IP source guard are permitted even if their destination MAC addresses are not the MAC address of the gateway.
- MFF is not supported in a network where VRRP load balancing mode is configured.

Prerequisites

For MFF to take effect, make sure ARP snooping is enabled on the VLAN where MFF is enabled.

Procedure

1. Enter system view.

   system-view
2. Enter VLAN view.

   vlan *vlan-id*
3. Enable MFF.

   mac-forced-forwarding default-gateway *gateway-ip*

   By default, MFF is disabled.

Configuring a network port

Restrictions and guidelines

In a VLAN with MFF enabled, you need to configure the following ports as network ports:

- Upstream ports connected to the gateway.
- Ports connected to other MFF devices.

Link aggregation is supported by network ports in an MFF-enabled VLAN, but it is not supported by user ports in the VLAN. You can add the network ports to link aggregation groups, but cannot add the user ports to link aggregation groups. For more information about link aggregation, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.
   - **interface** *interface-type interface-number*
3. Configure the port as a network port.

   mac-forced-forwarding network-port

   By default, the port is a user port.

Enabling periodic gateway probe

About this task

You can configure the MFF device to detect gateways every 30 seconds for the change of MAC addresses by sending forged ARP packets. The ARP packets use 0.0.0.0 as the sender IP address and bridge MAC address as the sender MAC address.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable periodic gateway probe.
   mac-forced-forwarding gateway probe
   By default, this feature is disabled.

Specifying the IP addresses of servers

About this task

Server IP addresses can be those of the interfaces on a router in a VRRP group and those of the servers collaborating with MFF, such as a RADIUS server.

When the MFF device receives an ARP request from a server, the device searches IP-to-MAC address entries it has stored. Then the device replies with the requested MAC address to the server.

As a result, packets from a host to a server are forwarded by the gateway. However, packets from a server to a host are not forwarded by the gateway.

MFF does not check whether the IP address of a server is on the same network segment as that of a gateway. Instead, it checks whether the IP address of a server is all-zero or all-one. An all-zero or all-one server IP address is invalid.

Restrictions and guidelines

If the server's interface connecting to the MFF device uses secondary IP addresses to send ARP packets, include all these IP addresses in the server IP address list.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Specify the IP addresses of servers.
   mac-forced-forwarding server *server-ip*&<1-10>
   By default, no server IP address is specified.

Display and maintenance commands for MFF

Execute **display** commands in any view.

Task	Command
Display MFF port configuration.	**display mac-forced-forwarding interface**
Display the MFF configuration for a VLAN.	**display mac-forced-forwarding vlan** *vlan-id*

MFF configuration examples

Example: Configuring MFF in a tree network

Network configuration

As shown in Figure 2, all the devices are in VLAN 100. Hosts A, B, and C are assigned IP addresses manually.

Configure MFF to isolate the hosts at Layer 2 and allow them to communicate with each other through the gateway at Layer 3.

Figure 2 Network diagram

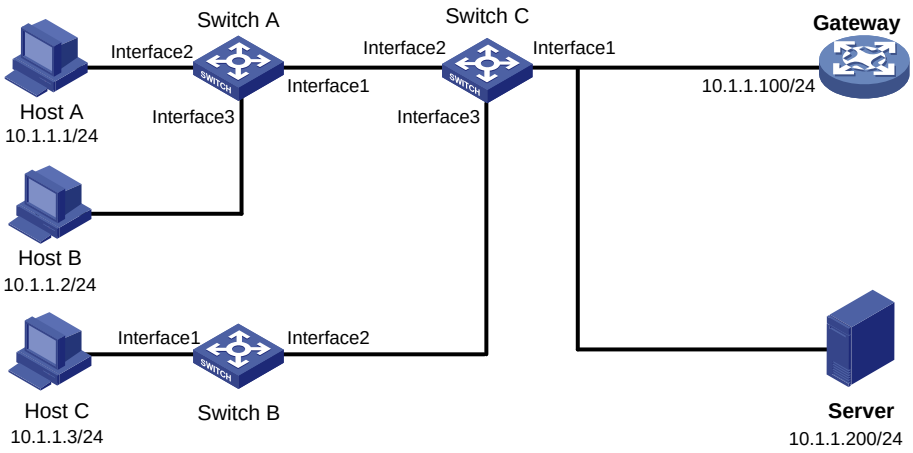


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure the IP addresses of the hosts and the gateway, as shown in Figure 2.
2. Configure Switch A:

   # Configure MFF on VLAN 100.

   ```
   [SwitchA] vlan 100
   [SwitchA-vlan100] mac-forced-forwarding default-gateway 10.1.1.100
   ```

   # Specify the IP address of the server.

```
[SwitchA-vlan100] mac-forced-forwarding server 10.1.1.200
```

Enable ARP snooping on VLAN 100.

```
[SwitchA-vlan100] arp snooping enable
[SwitchA-vlan100] quit
```

Configure GigabitEthernet 1/0/1 as a network port.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] mac-forced-forwarding network-port
```

3. Configure Switch B:

Configure MFF on VLAN 100.

```
[SwitchB] vlan 100
[SwitchB-vlan100] mac-forced-forwarding default-gateway 10.1.1.100
```

Specify the IP address of the server.

```
[SwitchB-vlan100] mac-forced-forwarding server 10.1.1.200
```

Enable ARP snooping on VLAN 100.

```
[SwitchB-vlan100] arp snooping enable
[SwitchB-vlan100] quit
```

Configure GigabitEthernet 1/0/2 as a network port.

```
[SwitchB] interface gigabitethernet 1/0/2 1/0/6
[SwitchB-GigabitEthernet1/0/2] mac-forced-forwarding network-port
```

Example: Configuring MFF in a ring network

Network configuration

As shown in Figure 3, all the devices are in VLAN 100, and the switches form a ring. Hosts A, B, and C are assigned IP addresses manually.

Configure MFF to isolate the hosts at Layer 2 and allow them to communicate with each other through the gateway at Layer 3.

Figure 3 Network diagram

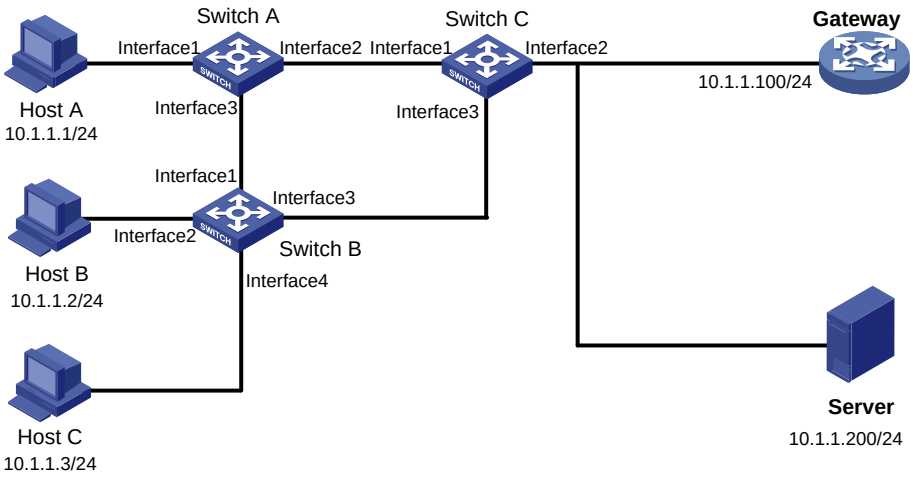


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure the IP addresses of the hosts and the gateway, as in shown in Figure 3.
2. Configure Switch A:

   # Enable STP globally to make sure STP is enabled on interfaces.

   ```
   [SwitchA] stp global enable
   ```

   # Configure MFF on VLAN 100.

   ```
   [SwitchA] vlan 100
   [SwitchA-vlan100] mac-forced-forwarding default-gateway 10.1.1.100
   ```

   # Specify the IP address of the server.

   ```
   [SwitchA-vlan100] mac-forced-forwarding server 10.1.1.200
   ```

   # Enable ARP snooping on VLAN 100.

   ```
   [SwitchA-vlan100] arp snooping enable
   [SwitchA-vlan100] quit
   ```

   # Configure GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 as network ports.

   ```
   [SwitchA] interface gigabitethernet 1/0/2
   [SwitchA-GigabitEthernet1/0/2] mac-forced-forwarding network-port
   [SwitchA-GigabitEthernet1/0/2] quit
   [SwitchA] interface gigabitethernet 1/0/3
   [SwitchA-GigabitEthernet1/0/3] mac-forced-forwarding network-port
   ```

3. Configure Switch B:

   # Enable STP globally to make sure STP is enabled on interfaces.

   ```
   [SwitchB] stp global enable
   ```

   # Configure MFF on VLAN 100.

   ```
   [SwitchB] vlan 100
   [SwitchB-vlan100] mac-forced-forwarding default-gateway 10.1.1.100
   ```

   # Specify the IP address of the server.

   ```
   [SwitchB-vlan100] mac-forced-forwarding server 10.1.1.200
   ```

   # Enable ARP snooping on VLAN 100.

   ```
   [SwitchB-vlan100] arp snooping enable
   [SwitchB-vlan100] quit
   ```

   # Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/3 as network ports.

   ```
   [SwitchB] interface gigabitethernet 1/0/1
   [SwitchB-GigabitEthernet1/0/1] mac-forced-forwarding network-port
   [SwitchB-GigabitEthernet1/0/1] quit
   [SwitchB] interface gigabitethernet 1/0/3
   [SwitchB-GigabitEthernet1/0/3] mac-forced-forwarding network-port
   ```

4. Enable STP on Switch C globally to make sure STP is enabled on interfaces.

   ```
   <SwitchC> system-view
   [SwitchC] stp global enable
   ```

Contents

Configuring FIPS 1

About FIPS 1
FIPS security levels 1
FIPS functionality 1
FIPS self-tests 1
Restrictions and guidelines: FIPS 3
Entering FIPS mode 4
About entering FIPS mode 4
Restrictions and guidelines 4
Using the automatic reboot method to enter FIPS mode 5
a Using the manual reboot method to enter FIPS mode 6
Manually triggering self-tests 7
1. Exiting FIPS mode 7
Display and maintenance commands for FIPS 9
FIPS configuration examples 9
Example: Entering FIPS mode through automatic reboot 9
Example: Entering FIPS mode through manual reboot 11
Example: Exiting FIPS mode through automatic reboot 12
Example: Exiting FIPS mode through manual reboot 13

Configuring FIPS

About FIPS

Federal Information Processing Standards (FIPS) was developed by the National Institute of Standards and Technology (NIST) of the United States. FIPS specifies the requirements for cryptographic modules.

FIPS security levels

FIPS 140-2 defines four levels of security, named Level 1 to Level 4, from low to high. The device supports Level 2.

Unless otherwise noted, the term "FIPS" refers to Level-2 FIPS 140-2 in this document.

FIPS functionality

In FIPS mode, the device has strict security requirements. It performs self-tests on cryptography modules to verify that the modules are operating correctly.

A FIPS device also meets the functionality requirements defined in Network Device Protection Profile (NDPP) of Common Criteria (CC).

FIPS self-tests

To ensure correct operation of cryptography modules, FIPS provides self-test mechanisms, including power-up self-tests and conditional self-tests.

If a power-up self-test fails, the device where the self-test process exists reboots. If a conditional self-test fails, the system outputs a self-test failure message.

NOTE:
If a self-test fails, contact H3C Support.

Power-up self-tests

The power-up self-test examines the availability of FIPS-allowed cryptographic algorithms.

The device supports the following types of power-up self-tests:

- Known-answer test (KAT)

  A cryptographic algorithm is run on data for which the correct output is already known. The calculated output is compared with the known answer. If they are not identical, the KAT test fails.

- Pairwise conditional test (PWCT)
  - **Signature and authentication test**—The test is run when a DSA, RSA, or ECDSA asymmetrical key pair is generated. The system uses the private key to sign the specific data, and then uses the public key to authenticate the signed data. If the authentication is successful, the test succeeds.
  - **Encryption and decryption test**—The test is run when an RSA asymmetrical key pair is generated. The system uses the public key to encrypt a plain text string, and then uses the private key to decrypt the encrypted text. If the decryption result is the same as the original plain text string, the test succeeds.

The power-up self-test examines the cryptographic algorithms listed in Table 1.

Table 1 Power-up self-tests list

Type	Operations
KAT	Tests the following algorithms: • SHA1, SHA224, SHA256, SHA384, and SHA512. • HMAC-SHA1, HMAC-SHA224, HMAC-SHA256, HMAC-SHA384, and HMAC-SHA512. • AES. • RSA (signature and authentication). • ECDH. • DRBG. • GCM. • GMAC.
PWCT	Tests the following algorithms: • RSA (signature and authentication). • RSA (encryption and decryption). • DSA (signature and authentication). • ECDSA (signature and authentication).

Conditional self-tests

A conditional self-test runs when an asymmetrical cryptographic module or a random number generator module is invoked. Conditional self-tests include the following types:

- **PWCT signature and authentication**—This test is run when a DSA or RSA asymmetrical key pair is generated. The system uses the private key to sign the specific data, and then uses the public key to authenticate the signed data. If the authentication is successful, the test succeeds.
- **Continuous random number generator test**—Runs when a random number is generated. The system compares the generated random number with the previously generated random number. If the two numbers are the same, the test fails. This test also runs when a DSA or RSA asymmetrical key pair is generated.

Restrictions and guidelines: FIPS

Requirements for key pairs and passwords

Before you reboot the device to enter FIPS mode, the system automatically removes all key pairs configured in non-FIPS mode and all FIPS-incompliant digital certificates. FIPS-incompliant digital certificates are MD5-based certificates with a key modulus length less than 2048 bits. You cannot log in to the device through SSH after the device enters FIPS mode. To log in to the device in FIPS mode through SSH, log in to the device through a console port and create a key pair for the SSH server.

The password for entering the device in FIPS mode must comply with the password control policies, such as password length, complexity, and aging policy. When the aging timer for a password expires, the system prompts you to change the password. If you adjust the system time after the device enters FIPS mode, the login password might expire before the next login, because the original system time is typically much earlier than the actual time.

Configuration rollback guidelines

Configuration rollback is supported in FIPS mode and also during a switch between FIPS mode and non-FIPS mode. After a configuration rollback between FIPS mode and non-FIPS mode, perform the following tasks:

1. Delete the local user and configure a new local user. Local user attributes include password, user role, and service type.
2. Save the current configuration file.
3. Specify the current configuration file as the startup configuration file.
4. Reboot the device. The new configuration takes effect after the reboot. During this process, do not exit the system or perform other operations.

If a device enters FIPS or non-FIPS mode through automatic reboot, configuration rollback fails. To support configuration rollback, you must execute the **save** command after the device enters FIPS or non-FIPS mode.

IRF compatibility

All devices in an IRF fabric must be operating in the same mode, whether in FIPS mode or non-FIPS mode.

To enable FIPS mode for an IRF fabric, you must reboot the entire IRF fabric.

Feature changes in FIPS mode

After the system enters FIPS mode, the following feature changes occur:

- The user login authentication mode can only be scheme.
- The FTP/TFTP server and client are disabled.
- The Telnet server and client are disabled.
- The HTTP server is disabled.

- SNMPv1 and SNMPv2c are disabled. Only SNMPv3 is available.
- The SSL server supports only TLS1.0, TLS1.1, and TLS1.2.
- The SSH server does not support SSHv1 clients or DSA key pairs.
- The generated RSA and DSA key pairs must have a modulus length of 2048 bits.

  When the device acts as a server to authenticate a client through the public key, the key pair for the client must also have a modulus length of 2048 bits.
- The generated ECDSA key pairs must have a modulus length of more than 256 bits.

  When the device acts as a server to authenticate a client through the public key, the key pair for the client must also have a modulus length of more than 256 bits.
- SSH, SNMPv3, IPsec, and SSL do not support DES, 3DES, RC4, or MD5.
- The password control feature cannot be disabled globally. The **`undo password-control enable`** command does not take effect.
- An AAA shared key, IKE pre-shared key, or SNMPv3 authentication key must have at least 15 characters and must contain uppercase and lowercase letters, digits, and special characters.
- The password for a device management local user and password for switching user roles must comply with the password control policies. By default, the password must have at least 15 characters and must contain uppercase and lowercase letters, digits, and special characters.

Entering FIPS mode

About entering FIPS mode

For the device to enter FIPS mode, you can use one of the following methods:

- **Automatic reboot**—The system automatically performs the following operations:
  a. Prompts you to specify the username and password for the next login.
  b. Creates a default FIPS configuration file named **fips-startup.cfg**.
  c. Specifies the default FIPS configuration file as the startup configuration file.
  d. Reboots and loads the default FIPS configuration file to enter the FIPS mode.
- **Manual reboot**—You must complete the required configuration tasks and reboot the device manually.

Restrictions and guidelines

After you execute the **`fips mode enable`** command, the system prompts you to choose a reboot method.

- If you do not make a choice within 30 seconds or press **Ctrl**+**C**, the system enables FIPS mode and waits for you to manually complete the FIPS mode configuration tasks. You must complete

the tasks or execute the **undo fips mode enable** command before saving the running configuration and rebooting the device.

- If you select the automatic reboot method, you can press **Ctrl**+**C** to abort both the interactive FIPS mode configuration process and the **fips mode enable** command.

Using the automatic reboot method to enter FIPS mode

Prerequisites

To ensure login password effectiveness under the password control policies, set the correct system time.

Procedure

1. **Enter system view.**

   system-view

   - Enable FIPS mode.

     fips mode enable

     By default, the FIPS mode is disabled.

   - After the reboot method choice prompt appears, enter **Y** within 30 minutes.

     The system starts the interactive FIPS mode configuration process.

CAUTION:
System reboot might interrupt ongoing services. Please perform the previous operations with caution.

   - Enter the login username and password as prompted.

     The password must have a minimum of 15 characters and must contain uppercase and lowercase letters, digits, and special characters. After you enter the username and password, the device performs the following operations:

1. Creates a device management local user that uses the entered username and password.
2. Assigns the user the terminal service and the network-admin user role.
3. Saves the running configuration and specifies the configuration file as the startup configuration file.
4. Reboots, loads the startup configuration file, and enters FIPS mode.

To log in to the device, you must enter the configured username and password. After login, you are identified as the FIPS mode crypto officer.

a Using the manual reboot method to enter FIPS mode

Prerequisites

1. **To ensure login password effectiveness under the password control policies, set the correct system time.**

Configure the password control feature.

Enable the password control feature globally.

Configure password control policies.

2. Set the number of character types a password must contain to 4.
3. Set the minimum number of characters for each type to one character.
4. Set the minimum length for a user password to 15 characters.

For more information about the password control feature, see password control in *Security Configuration Guide*.

Configure a local user.

5. Create a device management local user.
6. Specify a password that complies with the password control policies.
7. Assign the terminal service to the user.
8. Assign the network-admin user role to the user.

Procedure

1. **Enter system view.**

```
system-view
```

Enable FIPS mode.

```
fips mode enable
```

By default, the FIPS mode is disabled.

After the reboot method choice prompt appears, enter **N**.

The system enables FIPS mode and waits for you to complete the FIPS mode configuration tasks. Before rebooting the device to enter FIPS mode, do not execute any commands except for **save** and commands used to prepare for entering FIPS mode. If you execute any other commands, the commands might not take effect.

CAUTION:

1. When you enable or disable FIPS mode, the system prompts you to select the reboot method before it reboots. System reboot might interrupt ongoing services. Perform this operation with caution.
2. If you select the manual reboot method to enable FIPS mode, you must manually complete the

configurations for entering FIPS mode. Otherwise, you cannot log in to the system correctly.

Save the running configuration and specify the configuration file as the startup configuration file.

Delete the .mdb startup configuration file.

When loading a .mdb configuration file, the device loads all settings in the file. The settings that are not supported in FIPS mode might affect device operation.

Reboot the device.

The device reboots, loads the startup configuration file, and enters FIPS mode. To log in to the device, you must enter the configured username and password. After login, you are identified as the FIPS mode crypto officer.

Manually triggering self-tests

About this task

You can manually trigger FIPS self-tests to verify operation of cryptography modules anytime as required. The triggered self-tests are the same as the power-up self-tests.

Procedure

1. **Enter system view.**

```
system-view
```

Table 2 Trigger self-tests.

```
fips self-test
```

CAUTION:
A successful self-test requires that all cryptographic algorithms pass the self-test. If the self-test fails, the device where the self-test process exists reboots.

1. Exiting FIPS mode

1. About this task

After you disable FIPS mode and reboot the device, the device operates in non-FIPS mode.

For the device to exit FIPS mode, you can use one of the following reboot methods:

2. **Automatic reboot**—The system automatically creates a default non-FIPS configuration file named **non-fips-startup.cfg**, specifies the file as the startup configuration file, and reboots to enter non-FIPS mode. You can log in to the device without providing username or password.

3. **Manual reboot**—You must manually complete the configuration tasks for entering non-FIPS mode, and then reboot the device. To log in to the device after the reboot, you must enter user information as required by the authentication mode settings.

   The following are the default authentication mode settings:

   1. **VTY line**—Password authentication.
      - **AUX line**—Authentication is disabled.

   You can modify the authentication settings as needed.

   1. Using the automatic reboot method to exit FIPS mode

   1. Enter system view.

   ```
   system-view
   ```

   2. Disable FIPS mode.

      ```
      undo fips mode enable
      ```

      By default, the FIPS mode is disabled.

   3. Select the automatic reboot method.

   CAUTION:
   System reboot might interrupt ongoing services. Please perform the previous operations with caution.

   2. Using the manual reboot method to exit FIPS mode

   1. Enter system view.

   ```
   system-view
   ```

   2. Disable FIPS mode.

      ```
      undo fips mode enable
      ```

      By default, the FIPS mode is disabled.

   3. Select the manual reboot method.

   CAUTION:
   System reboot might interrupt ongoing services. Please perform the previous operations with caution.

   4. Configure login authentication settings.

2. If you logged in to the device through SSH, perform the following tasks without disconnecting the current user line:

   1. Set the authentication mode to **scheme** for VTY lines.

2. Specify the username and password. If you do not specify the username or password, the device uses the current username and password.

3. If you logged in to the device through a console port, configure login authentication settings for the current type of user lines as described in the following table:

Current login method	Login authentication requirements
Scheme	Set the authentication to scheme and specify the username and password. If you do not specify the username or password, the device uses the current username and password.
Password	Set the authentication to password and specify the password. If you do not specify the password, the device uses the current password.
None	Set the authentication to none.

1. Save the running configuration and specify the file as the startup configuration file.
2. Delete the .mdb startup configuration file.
3. Reboot the device.

Display and maintenance commands for FIPS

Execute **display** commands in any view.

Task	Command
Display the version number of the device algorithm base.	**display crypto version**
Display the FIPS mode state.	**display fips status**

FIPS configuration examples

Example: Entering FIPS mode through automatic reboot

Network configuration

Use the automatic reboot method to enter FIPS mode, and use a console port to log in to the device in FIPS mode.

Procedure

If you want to save the current configuration, execute the **save** command before you enable FIPS mode.

Enable FIPS mode and choose the automatic reboot method to enter FIPS mode. Set the username to **root** and the password to **12345zxcvb!@#$%ZXCVB**.

```
<Sysname> system-view
[Sysname] fips mode enable
FIPS mode change requires a device reboot. Continue? [Y/N]:y
Reboot the device automatically? [Y/N]:y
The system will create a new startup configuration file for FIPS mode. After you set the login username and password for FIPS mode, the device will reboot automatically.
Enter username(1-55 characters):root
Enter password(15-63 characters):
Confirm password:
Waiting for reboot... After reboot, the device will enter FIPS mode.
```

Verifying the configuration

After the device reboots, enter a username of **root** and a password of **12345zxcvb!@#$%ZXCVB**. The system prompts you to configure a new password. After you configure the new password, the device enters FIPS mode. The new password must be different from the previous password. It must include at least 15 characters, and contain uppercase and lowercase letters, digits, and special characters. For more information about the requirements for the password, see the system output.

```
Press ENTER to get started.
login: root
Password:
First login or password reset. For security reason, you need to change your password. Please enter your password.
old password:
new password:
confirm:
Updating user information. Please wait ... ...
…
<Sysname>
```

Display the FIPS mode state.

```
<Sysname> display fips status
FIPS mode is enabled.
```

Display the default configuration file.

```
<Sysname> more fips-startup.cfg
#
 password-control enable
#
local-user root class manage
 service-type terminal
 authorization-attribute user-role network-admin
#
 fips mode enable
#
return

<Sysname>
```

Example: Entering FIPS mode through manual reboot

Network configuration

Use the manual reboot method to enter FIPS mode, and use a console port to log in to the device in FIPS mode.

Procedure

Enable the password control feature globally.

```
<Sysname> system-view
[Sysname] password-control enable
```

Set the number of character types a password must contain to 4, and set the minimum number of characters for each type to one character.

```
[Sysname] password-control composition type-number 4 type-length 1
```

Set the minimum length of user passwords to 15 characters.

```
[Sysname] password-control length 15
```

Add a local user account for device management, including a username of **test**, a password of **12345zxcvb!@#$%ZXCVB**, a user role of **network-admin**, and a service type of **terminal**.

```
[Sysname] local-user test class manage
[Sysname-luser-manage-test] password simple 12345zxcvb!@#$%ZXCVB
[Sysname-luser-manage-test] authorization-attribute user-role network-admin
[Sysname-luser-manage-test] service-type terminal
[Sysname-luser-manage-test] quit
```

Enable FIPS mode, and choose the manual reboot method to enter FIPS mode.

```
[Sysname] fips mode enable
FIPS mode change requires a device reboot. Continue? [Y/N]:y
Reboot the device automatically? [Y/N]:n
Change the configuration to meet FIPS mode requirements, save the configuration to the
next-startup configuration file, and then reboot to enter FIPS mode.
```

Save the current configuration to the root directory of the storage medium, and specify it as the startup configuration file.

```
[Sysname] save
The current configuration will be written to the device. Are you sure? [Y/N]:y
Please input the file name(*.cfg)[flash:/startup.cfg]
(To leave the existing filename unchanged, press the enter key):
flash:/startup.cfg exists, overwrite? [Y/N]:y
Validating file. Please wait...
Saved the current configuration to mainboard device successfully.
[Sysname] quit
```

Delete the startup configuration file in binary format.

```
<Sysname> delete flash:/startup.mdb
Delete flash:/startup.mdb?[Y/N]:y
Deleting file flash:/startup.mdb...Done.
```

Reboot the device.

```
<Sysname> reboot
```

Verifying the configuration

After the device reboots, enter a username of **test** and a password of **12345zxcvb!@#$%ZXCVB**. The system prompts you to configure a new password. After you configure the new password, the device enters FIPS mode. The new password must be different from the previous password. It must include at least 15 characters, and contain uppercase and lowercase letters, digits, and special characters. For more information about the requirements for the password, see the system output.

```
Press ENTER to get started.
login: test
Password:
First login or password reset. For security reason, you need to change your pass
word. Please enter your password.
old password:
new password:
confirm:
Updating user information. Please wait ... ...
…
<Sysname>
```

Display the FIPS mode state.

```
<Sysname> display fips status
FIPS mode is enabled.
```

Example: Exiting FIPS mode through automatic reboot

Network configuration

After logging in to the device in FIPS mode through a console port, use the automatic reboot method to exit FIPS mode.

Procedure

Disable FIPS mode.

```
[Sysname] undo fips mode enable
FIPS mode change requires a device reboot. Continue? [Y/N]:y
The system will create a new startup configuration file for non-FIPS mode and then reboot
automatically. Continue? [Y/N]:y
Waiting for reboot... After reboot, the device will enter non-FIPS mode.
```

Verifying the configuration

After the device reboots, you can enter the system.

```
<Sysname>
```

Display the FIPS mode state.

```
<Sysname> display fips status
FIPS mode is disabled.
```

Example: Exiting FIPS mode through manual reboot

Network configuration

After logging in to the device in FIPS mode through the console port with username **test** and password **12345zxcvb!@#$%ZXCVB**, use the manual reboot method to exit FIPS mode.

Procedure

Disable FIPS mode.

```
[Sysname] undo fips mode enable
FIPS mode change requires a device reboot. Continue? [Y/N]:y
The system will create a new startup configuration file for non-FIPS mode, and then reboot automatically. Continue? [Y/N]:n
Change the configuration to meet non-FIPS mode requirements, save the configuration to the next-startup configuration file, and then reboot to enter non-FIPS mode.
```

Save the current configuration to the root directory of the storage medium, and specify it as the startup configuration file.

```
[Sysname] save
The current configuration will be written to the device. Are you sure? [Y/N]:y
Please input the file name(*.cfg)[flash:/startup.cfg]
(To leave the existing filename unchanged, press the enter key):
flash:/startup.cfg exists, overwrite? [Y/N]:y
Validating file. Please wait...
Saved the current configuration to mainboard device successfully.
[Sysname] quit
```

Delete the startup configuration file in binary format.

```
<Sysname> delete flash:/startup.mdb
Delete flash:/startup.mdb?[Y/N]:y
Deleting file flash:/startup.mdb...Done.
```

Reboot the device.

```
<Sysname> reboot
```

Verifying the configuration

After the device reboots, authentication is disabled for console login by default. You can press **Enter** to enter non-FIPS mode without authentication.

```
Press ENTER to get started.
login: test
Password:
Last successfully login time:…
…
<Sysname>
```

Display the FIPS mode state.

```
<Sysname> display fips status
FIPS mode is disabled.
```

Contents

Configuring MACsec 1

About MACsec 1
Basic concepts 1
Security mechanisms 2
Packet structure 2
MACsec application modes 3
MACsec operating mechanism 4
Protocols and standards 6
Restrictions and guidelines: MACsec configuration 6
MACsec tasks at a glance 7
Enabling MKA 7
Enabling MACsec desire 7
Specifying the cipher suite for MACsec encryption 8
Configuring a preshared key 8
Configuring the MKA key server priority 9
Setting the MKA life time 10
Configuring MACsec protection parameters 10
About MACsec protection parameters 10
Restrictions and guidelines for MACsec protection parameter configuration 10
Configuring MACsec protection parameters in interface view 10
Configuring MACsec protection parameters by MKA policy 11
Enabling inclusion of the SCI in the SecTAG field of MACsec frames on an interface 13
Enabling MACsec maintenance mode on an interface 13
Enabling MKA session logging 14
Display and maintenance commands for MACsec 14
MACsec configuration examples 15
Example: Configuring client-oriented MACsec (host as client) 15
Example: Configuring client-oriented MACsec (device as client) 18
Example: Configuring device-oriented MACsec 22
Troubleshooting MACsec 25
Cannot establish MKA sessions between MACsec devices 25

Configuring MACsec

About MACsec

Media Access Control Security (MACsec) secures data communication on IEEE 802 LANs. MACsec provides services such as data encryption, frame integrity check, and data origin validation for frames on the MAC sublayer of the Data Link Layer.

Basic concepts

MKA

MACsec Key Agreement (MKA) protocol is used for encryption key negotiation.

CA

Connectivity association (CA) is a group of participants that use the same key and key algorithm.

CAK

The key used by the CA participants is called a connectivity association key (CAK). The CAK is not directly used for data packet encryption. It and other parameters generate the encryption key. A CAK can be a key generated during 802.1X authentication or a user-configured preshared key. The user-configured preshared key takes precedence over the 802.1X-generated key.

The following types of CAKs are available:

- **Pairwise CAK**—Used by CAs that have two participants.
- **Group CAK**—Used by CAs that have more than two participants.

The pairwise CAK is used most often because MACsec is typically applied to point-to-point networks.

CKN

Connectivity Association Key Name (CKN) is the name of a CAK.

SC

Security Channel (SC) transmits secure data between CA participants.

SCI

Security Channel Identifier (SCI) includes the MAC address and ID of a port, which uniquely identifies an SC.

SA

Secure association (SA) is an agreement negotiated by CA participants. The agreement includes a cipher suite and keys for integrity check.

SAK

Security Association Key (SAK) is generated from the CAK based on an algorithm, which encrypts data transmitted along the SC. An SC can contain multiple SAs, and each SA uses a unique SAK.

MKA limits the number of packets that can be encrypted by an SAK. When the limit is exceeded, the SAK will be refreshed. For example, when packets with the minimum size are sent on a 10 Gbps link, an SAK rekey occurs about every 300 seconds.

ICV

The packet sender calculates an Integrity Check Value (ICV) from packet data units by using an algorithm and includes the ICV in the frame trailer. The packet receiver recalculates an ICV and compares the calculated ICV with the ICV in the frame trailer to check packet data integrity.

Security mechanisms

Data encryption

MACsec enables a port to encrypt outbound frames and decrypt MACsec-encrypted inbound frames. The keys for encryption and decryption are negotiated by MKA.

Integrity check

MACsec performs integrity check when the device receives a MACsec-encrypted frame. The integrity check uses the following process:

- Uses a key negotiated by MKA to calculate an ICV for the frame.
- Compares the calculated ICV with the ICV in the frame trailer.
  - If the ICVs are the same, the device verifies the frame as legal.
  - If the ICVs are different, the device determines whether to drop the frame based on the validation mode. The device supports the following validation modes:
    - **check**—Performs validation only, and does not drop illegal frames.
    - **disabled**—Does not perform validation.
    - **strict**—Performs validation, and drops illegal frames.

Replay protection

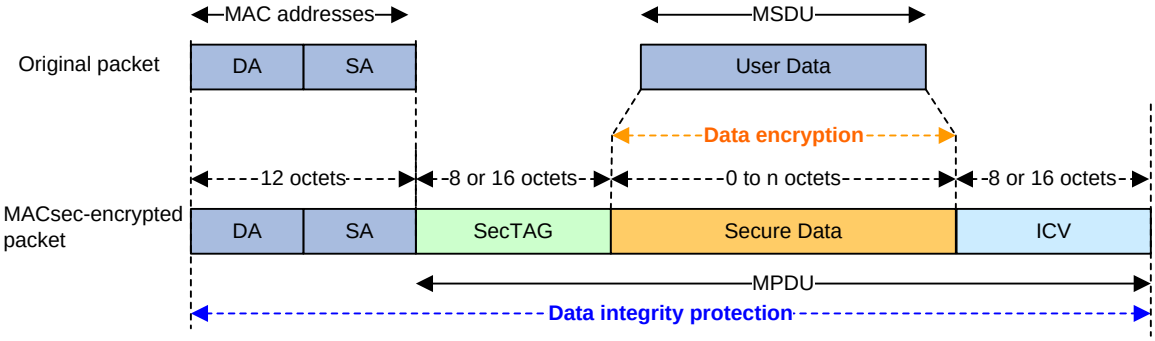
When MACsec frames are transmitted over the network, frame disorder might occur. MACsec replay protection allows the device to accept the out-of-order packets within the replay protection window size and drop other out-of-order packets. Suppose the replay protection window size is **a** on a port. After the port receives a packet with PN **x**, it can accept only packets whose PN is greater than or equal to **x-a**.

Packet structure

After SAK negotiation and installation between CA participants are completed, the negotiated key will be used for data packet forwarding.

- The data packet sender performs the following tasks:
  a. Uses an SAK and an encryption algorithm to encrypt the MAC Service Data Unit (MSDU) of the original packet. The encrypted part is called secure data.
  b. Uses the SAK and a validity algorithm to calculate an ICV from the source MAC address, destination MAC address, SecTAG, and secure data, and then includes the ICV in the packet trailer.
  c. Sends the complete MACsec-encrypted packet to the peer end.
- The packet receiver performs the following tasks:
  d. Uses the SAK and a decryption algorithm to obtain the plaintext from the encrypted packet.
  e. Uses a validity algorithm to calculate an ICV, and compares the calculated ICV with the ICV in the packet trailer. If the ICVs are the same, the packet has not been modified.

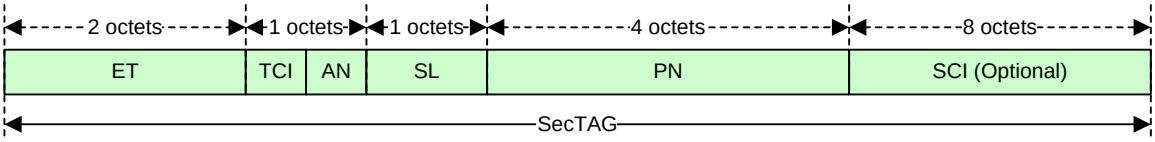
Figure 1 MACsec-encrypted packet



A MACsec-encrypted packet has the following key fields:

- **DA**—Destination MAC address in the packet. The length is 6 bytes.
- **SecTAG**—Security tag. The length is 8 or 16 bytes.
- **Secure Data**—Packet payload after MACsec encryption. MACsec can encrypt only the bytes after the MACsec confidentiality offset in a frame. The offset value can be 0, 30, or 50.
- **ICV**—Integrity check value. The length is 8 to 16 bytes. If a packet has been tampered with, the ICV value will change. This protects the packet from malicious modification.

Figure 2 SecTAG field



- **Ethertype (ET)**—Ethernet type of the MACsec protocol. The length is 2 bytes and the value is 0x88E5.
- **TAG control information (TCI)**—Provides information including the MACsec version number, whether the packet has been encrypted, whether integrity calculation has been performed, and whether the packet trailer carries an ICV. The length is 6 bits.
- **Association Number (AN)**—Number of an SA in the SC that transmits the packet. The length is 2 bits. An SC can include four SAs.
- **Short Length (SL)**—The length is 1 byte. If the length of the **Secure Data** field is less than 48 bytes, the value for this field is the length of **Secure Data** field. if the length of the **Secure Data** field is greater than 48 bytes, the value for this field is 0.
- **Packet Number (PN)**—Number of the packet. The length is 4 bytes. Every time the packet sender sends a packet, the value for this field increases by 1 to prevent packet replay attacks.
- **Secure Channel Identifier (SCI)**—8-byte identifier for the SC, including a 6-byte MAC address and a 2-byte port number.

MACsec application modes

MACsec supports client-oriented and device-oriented modes.

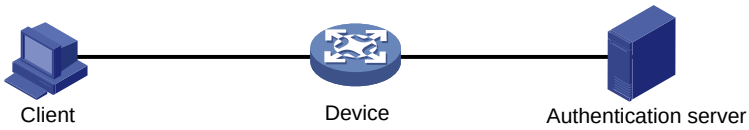
Client-oriented mode

As shown in Figure 3, MACsec secures data transmission between the client and the access device. In this mode, MACsec must operate with 802.1X authentication.

Client-oriented MACsec includes the following entities:

- **Client (supplicant)**—A user terminal seeking access to the LAN or a device that supports the 802.1X client feature. The terminal must have 802.1X software to authenticate to the access device. The terminal also performs CAK negotiation and packet encryption.
- **Access device (authenticator)**—Authenticates the client to control access to the LAN and performs CAK negotiation and packet encryption.
- **Authentication server**—Provides AAA services for the access device. The authentication server is typically a RADIUS server. After the client passes authentication, the authentication server generates and distributes the CAK to the client and the access device.

Figure 3 Client-oriented mode



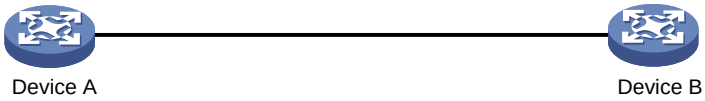
NOTE:

In client-oriented mode, an MKA-enabled port on the access device must perform port-based 802.1X access control. The authentication method must be EAP relay.

Device-oriented mode

As shown in Figure 4, MACsec secures data transmission between devices. In this mode, the same preshared key must be configured on the MACsec ports that connect the devices. The devices use the configured preshared key as the CAK.

Figure 4 Device-oriented mode

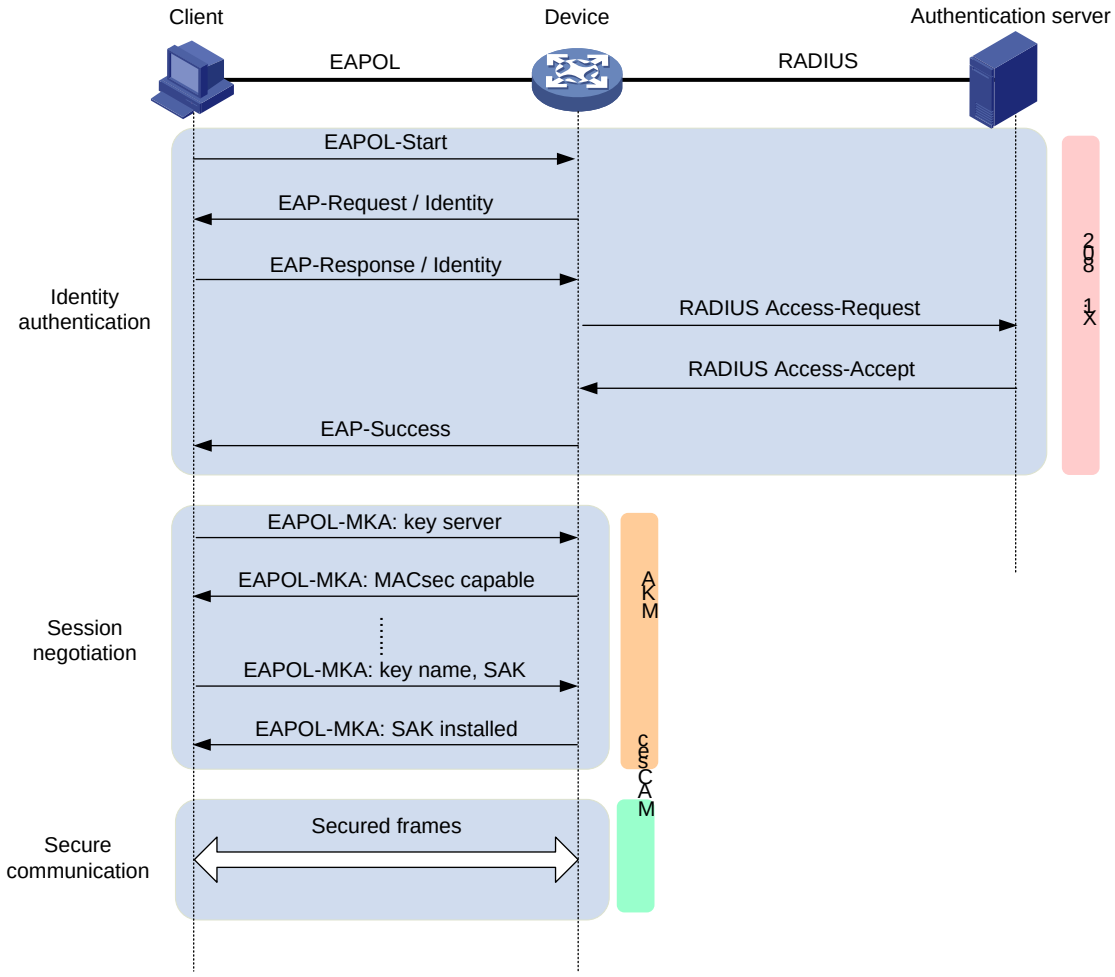


MACsec operating mechanism

Operating mechanism for client-oriented mode

Figure 5 illustrates how MACsec operates in client-oriented mode.

Figure 5 MACsec interactive process in client-oriented mode



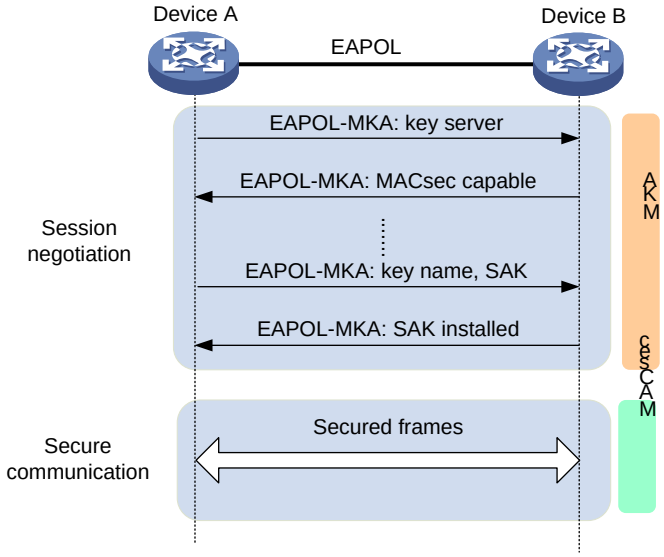
The following shows the MACsec process:

2. After the client passes 802.1X authentication, the RADIUS server distributes the generated CAK to the client and the access device.
3. After receiving the CAK, the client and the access device exchange EAPOL-MKA packets.
4. The client and the access device exchange the MACsec capability and required parameters for session establishment. The parameters include MKA key server priority and MACsec desire.
5. During the negotiation process, the access device with a higher priority automatically becomes the key server.
6. The key server generates an SAK from the CAK for packet encryption, and it distributes the SAK to the client.
7. The client and the access device use the SAK to encrypt packets, and they send and receive the encrypted packets in secure channels.
8. When the access device receives a logoff request from the client, it immediately removes the associated secure session from the port. The remove operation prevents an unauthorized client from using the secure session established by the previous authorized client to access the network.

Operating mechanism for device-oriented mode

As shown in Figure 6, the devices use the configured preshared key to start session negotiation.

Figure 6 MACsec interactive process in device-oriented mode



The following shows the MACsec process:

2. The devices use the configured preshared key as the CAK to exchange EAPOL-MKA packets.
3. The devices exchange the MACsec capability and required parameters for session establishment. The parameters include MKA key server priority and MACsec desire.
4. During the negotiation process, the port with higher MKA key server priority becomes the key server. The key server generates and distributes an SAK.
5. The devices use the SAK to encrypt packets, and they send and receive the encrypted packets in secure channels.
6. When a device receives a logoff request from the peer, it immediately deletes the associated secure session.

Protocols and standards

IEEE 802.1X-2010, *Port-Based Network Access Control*

IEEE 802.1AE-2006, *Media Access Control (MAC) Security*

Restrictions and guidelines: MACsec configuration

To avoid packet dropping, make sure MACsec traffic speed is below 100 Gbps.

You can enable MACsec on a maximum of 32 ports.

If you enable MKA on multiple ports in a serpentine topology, one port will occupy two encryption/decryption hardware resources, and one of these resources will be deleted by the platform later. The current number of encryption/decryption hardware resources is 32. Therefore, in a serpentine network, you can enable MKA on a maximum of 16 ports simultaneously. If this number is exceeded, channels cannot be created until other ports release encryption/decryption hardware resources.

Before the MACsec encryption/decryption channel is stable, traffic might not be forwarded normally.

If ports belong to the same VLAN, you can enable MACsec for a maximum of 5 ports in bulk when establishing channels between each two ports.

MACsec tasks at a glance

To configure MACsec, perform the following tasks:

1. Configuring basic MACsec parameters
   - Enabling MKA
   - Enabling MACsec desire
   - Specifying the cipher suite for MACsec encryption

     This task is applicable only to device-oriented mode.
   - Configuring a preshared key

     This task is required in device-oriented mode. Do not perform this task in client-oriented mode.
2. (Optional.) Configuring extended MACsec features
   - Configuring the MKA key server priority
   - Setting the MKA life time

     This task is applicable only in device-oriented mode.
   - Configuring MACsec protection parameters

     Choose one of the following tasks:
     - Configuring MACsec protection parameters in interface view
     - Configuring MACsec protection parameters by MKA policy
   - Enabling inclusion of the SCI in the SecTAG field of MACsec frames on an interface
   - Enabling MACsec maintenance mode on an interface
   - Enabling MKA session logging

Enabling MKA

About this task

MKA establishes and manages MACsec secure channels on a port. It also negotiates keys used by MACsec.

Procedure

1. Enter system view.

   `system-view`
2. Enter interface view.

   `interface` *`interface-type interface-number`*
3. Enable MKA.

   `mka enable`

   By default, MKA is disabled on the port.

Enabling MACsec desire

About this task

The MACsec desire feature expects MACsec protection for outbound frames. The key server determines whether MACsec protects the outbound frames.

MACsec protects the outbound frames of a port when the following requirements are met:

- The key server is MACsec capable.
- Both the local participant and its peer are MACsec capable.
- A minimum of one participant is enabled with MACsec desire.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable MACsec desire.
   macsec desire
   By default, the port does not expect MACsec protection for outbound frames.

Specifying the cipher suite for MACsec encryption

About this task

MACsec uses the GCM-AES-128, GCM-AES-256, GCM-AES-XPN-128, or GCM-AES-XPN-256 cipher suite to encrypt, validate, and decrypt protected data frames.

H3C's implementations of GCM-AES-256 and GCM-AES-XPN-256 cipher suites differ from their standard open source counterparts. To successfully establish MKA sessions with a peer that uses the standard open source GCM-AES-256 or GCM-AES-XPN-256 cipher suite, you must specify the standard cipher suite for MACsec encryption.

Restrictions and guidelines

This feature is supported only in device-oriented mode. Do not configure this feature on an 802.1X-enabled port.

Make sure the connected MACsec ports are configured with the same cipher suite. If the ports are configured with different cipher suites, they cannot successfully establish MKA sessions.

Specifying the cipher suite for MACsec encryption (interface view)

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the cipher suite for MACsec encryption.
   macsec cipher-suite { **gcm-aes-128** | **gcm-aes-256** [**standard**] | **gcm-aes-xpn-128** | **gcm-aes-xpn-256** [**standard**] }
   By default, MACsec uses the GCM-AES-128 cipher suite for encryption.

Configuring a preshared key

Restrictions and guidelines

In device-oriented mode, configure a preshared key as the CAK. To successfully establish an MKA session between two devices, make sure the following requirements are met:

- The connected MACsec ports are configured with the same CAK name (CKN) and CAK.
- Only the ports are configured with the same CKN in the network.

A user-configured preshared key has higher priority than the 802.1X-generated CAK. To ensure a successful MKA session establishment, do not configure a preshared key in client-oriented mode.

Different cipher suites for MACsec encryption have different requirements for the CKN and CAK configuration.

- The GCM-AES-128 or GCM-AES-XPN-128 cipher suite requires that the CKN and CAK each must be 32 characters long. If the configured CKN or CAK is not 32 characters long, the system performs the following operations when it runs the cipher suite:
  - Automatically increases the length of the CKN or CAK by zero padding if the CKN or CAK contains less than 32 characters.
  - Uses only the first 32 characters if the CKN or CAK contains more than 32 characters.
- The GCM-AES-256 or GCM-AES-XPN-256 cipher suite requires that the CKN and CAK each must be 64 characters long. If the configured CKN or CAK contains less than 64 characters, the system automatically increases the length of the CKN or CAK by zero padding when it runs the cipher suite.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set a preshared key.

   mka psk ckn *name* **cak** { **cipher** | **simple** } *string*

   By default, no MKA preshared key exists.

Configuring the MKA key server priority

Restrictions and guidelines

When you configure the MKA key server priority, follow these restrictions and guidelines:

- In client-oriented mode, the access device port automatically becomes the key server. You do not have to configure the MKA key server priority.
- In device-oriented mode, the port that has higher priority becomes the key server. The lower the priority value, the higher the priority. If a port and its peers have the same priority, MACsec compares the secure channel identifier (SCI) values on the ports. The port with the lowest SCI value (a combination of MAC address and port ID) becomes the key server.
- A port with priority 255 cannot become the key server. For a successful key server selection, make sure a minimum of one participant's key server priority is not 255.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Set the MKA key server priority.

   mka priority *priority-value*

   The default setting is 0.

Setting the MKA life time

Restrictions and guidelines

This task is applicable only in device-oriented mode.

Make sure the participants at each end of a secure session have the same MKA life time.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the MKA life time.
   mka timer mka-life *seconds*
   By default, the MKA life time is 6 seconds.

Configuring MACsec protection parameters

About MACsec protection parameters

You can set the following MACsec protection parameters:

- **MACsec confidentiality offset**—Specifies the number of bytes starting from the frame header. MACsec encrypts only the bytes after the offset in a frame. The offset value can be 0, 30, or 50.
- **MACsec replay protection**—Allows a MACsec port to accept a number of out-of-order or repeated inbound frames.
- **MACsec validation**—Allows a port to perform integrity check based on the validation modes.

Restrictions and guidelines for MACsec protection parameter configuration

You can configure MACsec protection parameters either in interface view or through MKA policies.

The use of MKA policies provides a centralized method to configure MACsec protection parameters. When you need to configure the same settings for MACsec protection parameters on multiple ports, you can configure them in an MKA policy and apply the policy to the ports.

If you configure a protection parameter in interface view after applying an MKA policy, the configuration in interface view overwrites the configuration of that parameter in the MKA policy. The other protection parameters in the MKA policy still take effect.

If you apply an MKA policy to a port after configuring protection parameters on the port, the settings in the policy overwrite all protection parameter settings in interface view. The protection parameters not configured in the policy are restored to the default.

The configured MACsec confidentiality offset does not take effect on the GCM-AES-XPN-128 or GCM-AES-XPN-256 cipher suite. The confidentiality offset is fixed at 0 for these two suites.

Configuring MACsec protection parameters in interface view

1. Enter system view.
   system-view

2. Enter interface view.
   interface *interface-type interface-number*
3. Set the MACsec confidentiality offset.
   macsec confidentiality-offset *offset-value*
   The default setting is 0, and the entire frame needs to be encrypted.
   MACsec uses the confidentiality offset propagated by the key server.
4. Configure MACsec replay protection:
   a. Enable MACsec replay protection.
      macsec replay-protection enable
      By default, MACsec replay protection is enabled on the port.
   b. Set the MACsec replay protection window size.
      macsec replay-protection window-size *size-value*
      The default setting is 0. The device accepts only frames that arrive in the correct order. Out-of-order or duplicated frames will be dropped.
      The configured replay protection window size takes effect only when MACsec replay protection is enabled.
5. Set a MACsec validation mode.
   macsec validation mode { **check** | **disabled** | **strict** }
   The default setting is **check**.
   To avoid data loss, use the **display macsec** command to verify that MKA negotiation with the peer device has succeeded before you change the mode to **strict**.

Parameter	Description
check	Verifies incoming frames but does not drop illegal frames.
disabled	Disables MACsec validation for incoming frames.
strict	Verifies incoming frames and drops illegal frames.

Configuring MACsec protection parameters by MKA policy

Restrictions and guidelines

An MKA policy can be applied to a port or multiple ports. When you apply an MKA policy to a port, follow these restrictions and guidelines:

- The settings in the MKA policy overwrite all protection parameter settings configured in interface view. The protection parameters not configured in the policy are restored to the default.
- Any modifications to the MKA policy take effect immediately.
- When you remove an MKA policy application from the port, the MACsec parameter settings on the port restore to the default.
- When you apply a nonexistent MKA policy to the port, the port automatically uses the system-defined MKA policy named **default-policy**. If you create the policy, the policy will be automatically applied to the port.
- The configured MACsec confidentiality offset does not take effect on the GCM-AES-XPN-128 or GCM-AES-XPN-256 cipher suite. The confidentiality offset is fixed at 0 for these two suites.

Procedure

1. Enter system view.

system-view

2. Create an MKA policy and enter its view.

   mka policy *policy-name*

   By default, a system-defined MKA policy exists. The policy name is **default-policy**.

   The settings for parameters in the system-defined policy are the same as the default settings for the parameters on a port.

   You cannot delete or modify the system-defined MKA policy.

   You can create multiple MKA policies.

3. Set the MACsec confidentiality offset.

   confidentiality-offset *offset-value*

   The default setting is 0, and the entire frame needs to be encrypted.

   MACsec uses the confidentiality offset propagated by the key server.

4. Configure MACsec replay protection:

   a. Enable MACsec replay protection.

      replay-protection enable

      By default, MACsec replay protection is enabled.

   b. Set the replay protection window size.

      replay-protection window-size *size-value*

      The default replay protection window size is 0. The device accepts only frames that arrive in the correct order. Out-of-order or duplicated frames will be dropped.

5. Set a MACsec validation mode.

   validation mode { **check** | **disabled** | **strict** }

   The default setting is **check**.

Parameter	Description
check	Verifies incoming frames but does not drop illegal frames.
disabled	Disables MACsec validation for incoming frames.
strict	Verifies incoming frames and drops illegal frames.

6. Apply an MKA policy:

   a. Return to system view.

      quit

   b. Enter interface view.

      interface *interface-type interface-number*

   c. Apply the MKA policy to the port.

      mka apply policy *policy-name*

      By default, no MKA policy is applied to a port.

Enabling inclusion of the SCI in the SecTAG field of MACsec frames on an interface

About this task

The secure channel identifier (SCI) identifies the source of a MACsec frame. It contains the MAC address and ID of the port that sends the frame.

To have an interface send encrypted MACsec frames with the SCI included in the SecTAG field, perform this task.

Restrictions and guidelines

To forward traffic between a pair of peer MACsec interfaces, you must make sure inclusion of the SCI in the SecTAG field is enabled or disabled on both of them.

Procedure

1. Enter system view.

   `system-view`
2. Enter interface view.

   `interface` *`interface-type interface-number`*
3. Enable inclusion of the SCI in the SecTAG field of MACsec frames sent out of the interface.

   `macsec include-sci`

   By default, a MACsec port sends MACsec frames with the SCI included in the SecTAG field.

Enabling MACsec maintenance mode on an interface

About this task

To protect data security, a MACsec port is blocked by default until it establishes an MKA session with its peer port. In blocked state, the MACsec port can send and receive MKA protocol packets, but it cannot send or receive data packets. On a network that uses a controller to deploy configuration, this might cause a configuration deployment failure if the configuration delivered to a device must traverse a MACsec port.

To make sure a device can always receive its configuration from a controller over the link between a pair of MACsec ports, enable MACsec maintenance mode on the MACsec ports. In this mode, a MACsec port is unblocked regardless of whether it has established an MKA session with its peer port.

For security purposes, a MACsec port in MACsec maintenance mode will still be blocked when it receives frames that contain its own SCI (identified by its MAC address and port ID).

A MACsec port in MACsec maintenance mode issues its SCI information only after it receives MKA protocol packets from its peer port.

Restrictions and guideline

△ **CAUTION:**

If you enable MACsec maintenance mode on an interface before it establishes an MKA session with its peer end, data frames are exposed to security risks. Before you perform this operation, make sure you fully understand this impact on security.

Prerequisites

Before you enable or disable MACsec maintenance mode on a MACsec port:

- Disable MKA on that port.
- Use the **link-delay down** command to increase the delay to report a link down event to CPU. Transient link state flapping might occur after you enable MACsec maintenance mode. This step prevents a link down event from causing MKA packet exchange failure during this period.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable MACsec maintenance mode.
   macsec maintenance-mode enable
   By default, MACsec maintenance mode is disabled.

CAUTION:
Enabling this feature before MKA session establishment reduces security of data packets. Please be cautious.

Enabling MKA session logging

About this task

This feature enables the device to generate logs for MKA session changes, such as peer aging and SAK updates. The logs are sent to the information center of the device. For the logs to be output correctly, you must also configure the information center on the device. For more information about information center configuration, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

As a best practice, disable this feature to prevent excessive MKA session log output.

Procedure

1. Enter system view.
   system-view
2. Enable MKA session logging.
   macsec mka-session log enable
   By default, MKA session logging is disabled.

Display and maintenance commands for MACsec

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display MACsec information on ports.	**display macsec** [**interface** *interface-type interface-number*] [**verbose**]

Task	Command
Display MKA policy information.	**display mka** { **default-policy** | **policy** [**name** *policy-name*] }
Display MKA session information.	**display mka session** [**interface** *interface-type interface-number* | **local-sci** *sci-id*] [**verbose**]
Display MKA statistics on ports.	**display mka statistics** [**interface** *interface-type interface-number*]
Reset MKA sessions on ports.	**reset mka session** [**interface** *interface-type interface-number*]
Clear MKA statistics on ports.	**reset mka statistics** [**interface** *interface-type interface-number*]

MACsec configuration examples

Example: Configuring client-oriented MACsec (host as client)

Network configuration

As shown in Figure 7, the host accesses the network through GigabitEthernet 1/0/1. The device performs RADIUS-based 802.1X authentication for the host to control user access to the Internet.

To ensure secure communication between the host and device, perform the following tasks on the device:

- Enable MACsec desire, and configure MKA to negotiate SAKs for packet encryption.
- Set the MACsec confidentiality offset to 30 bytes.
- Enable MACsec replay protection, and set the replay protection window size to 100.
- Set the MACsec validation mode to **strict**.

Figure 7 Network diagram

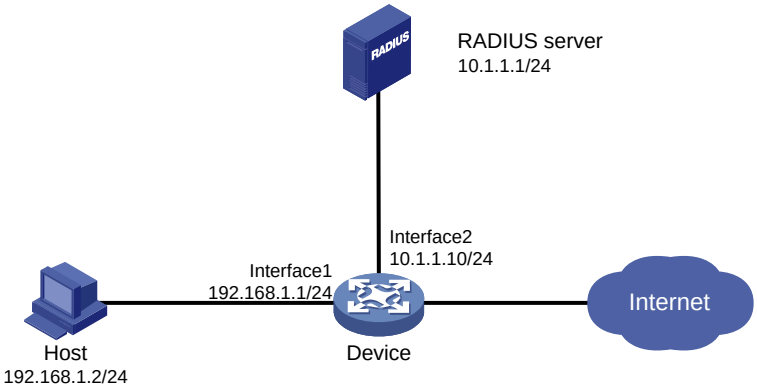


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure IP addresses for the Ethernet ports. (Details not shown.)
2. Configure the RADIUS server to provide authentication, authorization, and accounting services. Add a user account for the host. (Details not shown.)
3. Configure AAA:

   # Enter system view.

   ```
   <Device> system-view
   ```

   # Configure RADIUS scheme **radius1**.

   ```
   [Device] radius scheme radius1
   [Device-radius-radius1] primary authentication 10.1.1.1
   [Device-radius-radius1] primary accounting 10.1.1.1
   [Device-radius-radius1] key authentication simple name
   [Device-radius-radius1] key accounting simple name
   [Device-radius-radius1] user-name-format without-domain
   [Device-radius-radius1] quit
   ```

   # Configure authentication domain **bbb** for 802.1X users.

   ```
   [Device] domain bbb
   [Device-isp-bbb] authentication lan-access radius-scheme radius1
   [Device-isp-bbb] authorization lan-access radius-scheme radius1
   [Device-isp-bbb] accounting lan-access radius-scheme radius1
   [Device-isp-bbb] quit
   ```

4. Configure 802.1X:

   # Enable 802.1X on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] dot1x
   ```

   # Implement port-based access control on GigabitEthernet 1/0/1.

   ```
   [Device-GigabitEthernet1/0/1] dot1x port-method portbased
   ```

   # Specify **bbb** as the mandatory authentication domain for 802.1X users on GigabitEthernet 1/0/1.

   ```
   [Device-GigabitEthernet1/0/1] dot1x mandatory-domain bbb
   [Device-GigabitEthernet1/0/1] quit
   ```

   # Enable 802.1X globally, and sets the device to relay EAP packets.

   ```
   [Device] dot1x
   [Device] dot1x authentication-method eap
   ```

5. Configure MACsec:

   # Create an MKA policy named **pls**.

   ```
   [Device] mka policy pls
   ```

   # Set the MACsec confidentiality offset to 30 bytes.

   ```
   [Device-mka-policy-pls] confidentiality-offset 30
   ```

   # Enable MACsec replay protection.

   ```
   [Device-mka-policy-pls] replay-protection enable
   ```

   # Set the MACsec replay protection window size to 100.

   ```
   [Device-mka-policy-pls] replay-protection window-size 100
   ```

   # Set the MACsec validation mode to **strict**.

   ```
   [Device-mka-policy-pls] validation mode strict
   [Device-mka-policy-pls] quit
   ```

Apply the MKA policy to GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] mka apply policy pls
```

Configure MACsec desire and enable MKA on GigabitEthernet 1/0/1.

```
[Device-GigabitEthernet1/0/1] macsec desire
[Device-GigabitEthernet1/0/1] mka enable
[Device-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Display MACsec information on GigabitEthernet 1/0/1.

```
[Device] display macsec interface gigabitethernet 1/0/1 verbose
Interface GigabitEthernet1/0/1
  Protect frames               : Yes
  Active MKA policy            : pls
  Replay protection            : Enabled
  Config reply window size     : 100 frames
  Active replay window size    : 100 frames
  Config confidentiality offset : 30 bytes
  Active confidentiality offset : 30 bytes
  Config validation mode       : Strict
  Active validation mode       : Strict
  Included SCI                 : No
  SCI conflict                 : No
  Cipher suite                 : GCM-AES-128
  MKA life time                : 6 seconds
  Transmit secure channel:
    SCI           : 00E00100000A0006
      Elapsed time: 00h:02m:07s
      Current SA  : AN 0        PN 1
  Receive secure channels:
    SCI           : 00E0020000000106
      Elapsed time: 00h:02m:03s
      Current SA  : AN 0        LPN 1
      Previous SA : AN N/A      LPN N/A
```

Display MKA session information on GigabitEthernet 1/0/1 after a user logs in.

```
[Device] display mka session interface gigabitethernet 1/0/1 verbose
Interface GigabitEthernet1/0/1
Tx-SCI    : 00E00100000A0006
Priority  : 0
Capability: 3
  CKN for participant: 1234
    Key server            : Yes
    MI (MN)               : A1E0D2897596817209CD2307 (2509)
    Live peers            : 1
    Potential peers       : 0
    Principal actor       : Yes
    MKA session status    : Secured
    Confidentiality offset: 30 bytes
   Tx-SSCI               : N/A
```

```
Current SAK status    : Rx & Tx
Current SAK AN        : 0
Current SAK KI (KN)   : A1E0D2897596817209CD230700000002 (2)
Previous SAK status   : N/A
Previous SAK AN       : N/A
Previous SAK KI (KN)  : N/A
Live peer list:
MI                       MN         Prio  Cap  Rx-SCI            Rx-SSCI
B2CAF896C9BFE2ABFB135E63 2512       0     3    00E0020000000106 N/A
```

Example: Configuring client-oriented MACsec (device as client)

Network configuration

As shown in Figure 8:

- The switch connects to the device through trunk ports GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3. GigabitEthernet 1/0/2 allows packets only from VLAN 1 and VLAN 2, and GigabitEthernet 1/0/3 allows packets only from VLAN 1 and VLAN 3.
- The device acts as a network access device. You cannot configure a preshared key on the device for MKA negotiation and packet encryption.
- The RADIUS server acts as an 802.1X authentication server.

To secure data between the device and the switch, configure MACsec on the switch as follows:

- Enable MKA and MACsec desire on GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 to protect traffic forwarding on a per-port basis.
- Configure MACsec to use 802.1X-generated CAKs for MKA negotiation and packet encryption. For this purpose, configure the 802.1X client feature on the switch.

Figure 8 Network diagram

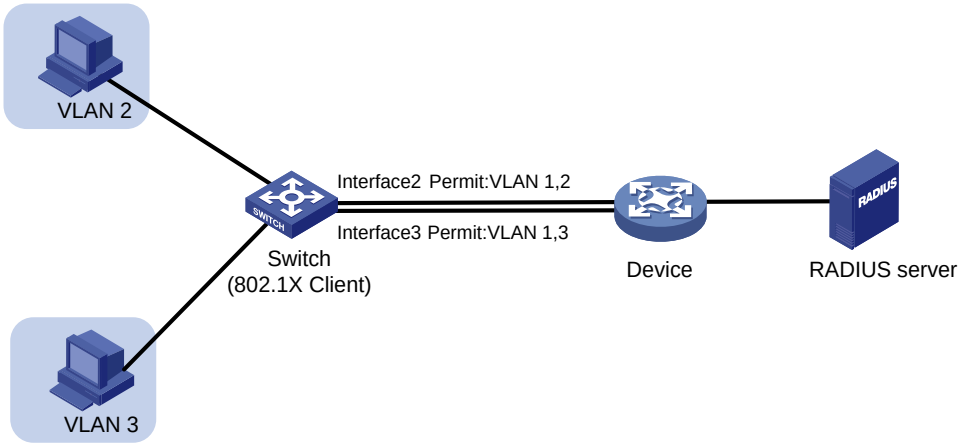


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure IP addresses for the Ethernet ports. Make sure the switch, the device, and the RADIUS server can reach one another. (Details not shown.)
2. Configure the access device. (Details not shown.)

   Configuration on the access device varies by manufacturer. For information about device configuration, see the corresponding product manual. This part covers only the switch configuration, and for information about 802.1X client commands, see *Security Command Reference*.
3. Configure the RADIUS server to provide authentication, authorization, and accounting services. Add user accounts. (Details not shown.)
4. Configure the switch:

   # Create VLAN 2.

   ```
   <Switch> system-view
   [Switch] vlan 2
   [Switch-vlan2] quit
   ```

   # Configure GigabitEthernet 1/0/2 as a trunk port, and assign the port to VLAN 2.

   ```
   [Switch] interface gigabitethernet 1/0/2
   [Switch-GigabitEthernet1/0/2] port link-type trunk
   [Switch-GigabitEthernet1/0/2] port trunk permit vlan 2
   ```

   # Configure the 802.1X client username as **aaaa**, and set the password to **123456** in plaintext form on GigabitEthernet 1/0/2.

   ```
   [Switch-GigabitEthernet1/0/2] dot1x supplicant username aaaa
   [Switch-GigabitEthernet1/0/2] dot1x supplicant password simple 123456
   ```

   # Specify TTLS-GTC as the 802.1X client EAP authentication method on GigabitEthernet 1/0/2.

   ```
   [Switch-GigabitEthernet1/0/2] dot1x supplicant eap-method ttls-gtc
   ```

   # Specify MAC address 1-1-1 for 802.1X client authentication on GigabitEthernet 1/0/2.

   ```
   [Switch-GigabitEthernet1/0/2] dot1x supplicant mac-address 1-1-1
   ```

   # Enable the 802.1X client feature on GigabitEthernet 1/0/2.

   ```
   [Switch-GigabitEthernet1/0/2] dot1x supplicant enable
   ```

   # Configure MACsec desire and enable MKA on GigabitEthernet 1/0/2.

   ```
   [Switch-GigabitEthernet1/0/2] macsec desire
   [Switch-GigabitEthernet1/0/2] mka enable
   [Switch-GigabitEthernet1/0/2] quit
   ```

   # Create VLAN 3.

   ```
   [Switch] vlan 3
   [Switch-vlan3] quit
   ```

   # Configure GigabitEthernet 1/0/3 as a trunk port, and assign the port to VLAN 3.

   ```
   [Switch] interface gigabitethernet 1/0/3
   [Switch-GigabitEthernet1/0/3] port link-type trunk
   [Switch-GigabitEthernet1/0/3] port trunk permit vlan 3
   ```

   # Configure the 802.1X client username as **bbbb**, and set the password to **654321** in plaintext form on GigabitEthernet 1/0/3.

   ```
   [Switch-GigabitEthernet1/0/3] dot1x supplicant username bbbb
   [Switch-GigabitEthernet1/0/3] dot1x supplicant password simple 654321
   ```

   # Specify TTLS-GTC as the 802.1X client EAP authentication method on GigabitEthernet 1/0/3.

   ```
   [Switch-GigabitEthernet1/0/3] dot1x supplicant eap-method ttls-gtc
   ```

   # Specify MAC address 1-1-2 for 802.1X client authentication on GigabitEthernet 1/0/3.

   ```
   [Switch-GigabitEthernet1/0/3] dot1x supplicant mac-address 1-1-2
   ```

Enable the 802.1X client feature on GigabitEthernet 1/0/3.

```
[Switch-GigabitEthernet1/0/3] dot1x supplicant enable
```

Configure MACsec desire and enable MKA on GigabitEthernet 1/0/3.

```
[Switch-GigabitEthernet1/0/3] macsec desire
[Switch-GigabitEthernet1/0/3] mka enable
[Switch-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Display MACsec information on GigabitEthernet 1/0/2.

```
[Switch] display macsec interface gigabitethernet 1/0/2 verbose
Interface GigabitEthernet1/0/2
  Protect frames                : Yes
  Active MKA policy             : pls
  Replay protection             : Enabled
  Config replay window size     : 100 frames
  Active replay window size     : 100 frames
  Config confidentiality offset : 30 bytes
  Active confidentiality offset : 30 bytes
  Config validation mode        : Strict
  Active validation mode        : Strict
  Included SCI                  : No
  SCI conflict                  : No
  Cipher suite                  : GCM-AES-128
  MKA life time                 : 6 seconds
  Transmit secure channel:
    SCI           : 00E00100000A0006
      Elapsed time: 00h:02m:07s
      Current SA  : AN 0        PN 1
  Receive secure channels:
    SCI           : 00E0020000000106
      Elapsed time: 00h:02m:03s
      Current SA  : AN 0        LPN 1
      Previous SA : AN N/A      LPN N/A
```

Display MACsec information on GigabitEthernet 1/0/3.

```
[Switch] display macsec interface gigabitethernet 1/0/3 verbose
Interface GigabitEthernet1/0/3
  Protect frames                : Yes
  Replay protection             : Enabled
  Config replay window size     : 100 frames
  Active replay window size     : 100 frames
  Config confidentiality offset : 30 bytes
  Active confidentiality offset : 30 bytes
  Config validation mode        : Check
  Active validation mode        : Check
  Included SCI                  : No
  SCI conflict                  : No
  Cipher suite                  : GCM-AES-128
  MKA life time                 : 6 seconds
  Transmit secure channel:
```

```
    SCI           : A087100801000103
      Elapsed time: 00h:00m:55s
      Current SA  : AN 0        PN 1
  Receive secure channels:
    SCI           : A0872B3602000003
      Elapsed time: 00h:00m:52s
      Current SA  : AN 0        LPN 1
      Previous SA : AN N/A      LPN N/A
```

Display MKA session information on GigabitEthernet 1/0/2 after 802.1X client user **aaaa** comes online.

```
[Switch] display mka session interface gigabitethernet 1/0/2 verbose
Interface GigabitEthernet1/0/2
Tx-SCI    : 00E00100000A0006
Priority  : 0
Capability: 3
 CKN for participant: 1234
   Key server            : Yes
   MI (MN)               : A1E0D2897596817209CD2307 (2509)
   Live peers            : 1
   Potential peers       : 0
   Principal actor       : Yes
   MKA session status    : Secured
   Confidentiality offset: 30 bytes
   Tx-SSCI               : N/A
   Current SAK status    : Rx & Tx
   Current SAK AN        : 0
   Current SAK KI (KN)   : A1E0D2897596817209CD230700000002 (2)
   Previous SAK status   : N/A
   Previous SAK AN       : N/A
   Previous SAK KI (KN)  : N/A
   Live peer list:
   MI                        MN         Prio  Cap  Rx-SCI            Rx-SSCI
   B2CAF896C9BFE2ABFB135E63  2512       0     3    00E0020000000106  N/A
```

Display MKA session information on GigabitEthernet 1/0/3 after 802.1X client user **bbbb** comes online.

```
[Switch] display mka session interface gigabitethernet 1/0/3 verbose
Interface GigabitEthernet1/0/3
Tx-SCI    : A087100801000103
Priority  : 0
Capability: 3
  CKN for participant: 7B8784F16F85ED8F9D0130AA9B93D0F0
    Key server            : No
    MI (MN)               : D3F6D374598C8FD1F1819D6C (78)
    Live peers            : 1
    Potential peers       : 0
    Principal actor       : Yes
    MKA session status    : Secured
    Confidentiality offset: 0 bytes
```

```
Tx-SSCI               : N/A
Current SAK status    : Rx & Tx
Current SAK AN        : 0
Current SAK KI (KN)   : FCA71854FCAE51398EC2DA7900000001 (1)
Previous SAK status   : N/A
Previous SAK AN       : N/A
Previous SAK KI (KN)  : N/A
Live peer list:
MI                        MN         Prio  Cap  Rx-SCI            Rx-SSCI
FCA71854FCAE51398EC2DA79  71         0     3    A0872B3602000003  N/A
```

Example: Configuring device-oriented MACsec

Network configuration

As shown in Figure 9, Device A is the MACsec key server.

To secure data transmission between the two devices by MACsec, perform the following tasks on Device A and Device B, respectively:

- Set the MACsec confidentiality offset to 30 bytes.
- Enable MACsec replay protection, and set the replay protection window size to 100.
- Set the MACsec validation mode to **strict**.
- Configure the CAK name (CKN) and the CAK as **E9AC** and **09DB3EF1**, respectively.

Figure 9 Network diagram

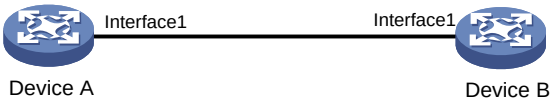


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure Device A:

   # Enter system view.

   ```
   <DeviceA> system-view
   ```

   # Enter GigabitEthernet 1/0/1 interface view.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   ```

   # Enable MACsec desire on GigabitEthernet 1/0/1.

   ```
   [DeviceA-GigabitEthernet1/0/1] macsec desire
   ```

   # Set the MKA key server priority to 5.

   ```
   [DeviceA-GigabitEthernet1/0/1] mka priority 5
   ```

   # Configure the CKN as **E9AC** and the CAK as **09DB3EF1** in plain text.

   ```
   [DeviceA-GigabitEthernet1/0/1] mka psk ckn E9AC cak simple 09DB3EF1
   ```

   # Set the MACsec confidentiality offset to 30 bytes.

   ```
   [DeviceA-GigabitEthernet1/0/1] macsec confidentiality-offset 30
   ```

   # Enable MACsec replay protection.

```
[DeviceA-GigabitEthernet1/0/1] macsec replay-protection enable
```

Set the MACsec replay protection window size to 100.

```
[DeviceA-GigabitEthernet1/0/1] macsec replay-protection window-size 100
```

Set the MACsec validation mode to **strict**.

```
[DeviceA-GigabitEthernet1/0/1] macsec validation mode strict
```

Enable MKA on GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] mka enable
[DeviceA-GigabitEthernet1/0/1] quit
```

2. Configure Device B:

Enter system view.

```
<DeviceB> system-view
```

Enter GigabitEthernet 1/0/1 interface view.

```
[DeviceB] interface gigabitethernet 1/0/1
```

Enable MACsec desire on GigabitEthernet 1/0/1.

```
[DeviceB-GigabitEthernet1/0/1] macsec desire
```

Set the MKA key server priority to 10.

```
[DeviceB-GigabitEthernet1/0/1] mka priority 10
```

Configure the CKN as **E9AC** and the CAK as **09DB3EF1** in plain text.

```
[DeviceB-GigabitEthernet1/0/1] mka psk ckn E9AC cak simple 09DB3EF1
```

Set the MACsec confidentiality offset to 30 bytes.

```
[DeviceB-GigabitEthernet1/0/1] macsec confidentiality-offset 30
```

Enable MACsec replay protection.

```
[DeviceB-GigabitEthernet1/0/1] macsec replay-protection enable
```

Set the MACsec replay protection window size to 100.

```
[DeviceB-GigabitEthernet1/0/1] macsec replay-protection window-size 100
```

Set the MACsec validation mode to **strict**.

```
[DeviceB-GigabitEthernet1/0/1] macsec validation mode strict
```

Enable MKA on GigabitEthernet 1/0/1.

```
[DeviceB-GigabitEthernet1/0/1] mka enable
[DeviceB-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Display MACsec information on GigabitEthernet 1/0/1 of Device A.

```
[DeviceA] display macsec interface gigabitethernet 1/0/1 verbose
Interface GigabitEthernet1/0/1
  Protect frames                : Yes
  Replay protection             : Enabled
  Config replay window size     : 100 frames
  Active replay window size     : 100 frames
  Config confidentiality offset : 30 bytes
  Active confidentiality offset : 30 bytes
  Config validation mode        : Strict
  Active validation mode        : Strict
  Included SCI                  : No
  SCI conflict                  : No
  Cipher suite                  : GCM-AES-128
  MKA life time                 : 6 seconds
```

```
  Transmit secure channel:
    SCI           : 00E00100000A0006
      Elapsed time: 00h:05m:00s
      Current SA  : AN 0        PN 1
  Receive secure channels:
    SCI           : 00E0020000000106
      Elapsed time: 00h:03m:18s
      Current SA  : AN 0        LPN 1
      Previous SA : AN N/A      LPN N/A
```

Display MKA session information on GigabitEthernet 1/0/1 of Device A.

```
[DeviceA] display mka session interface gigabitethernet 1/0/1 verbose
Interface GigabitEthernet1/0/1
Tx-SCI    : 00E00100000A0006
Priority  : 5
Capability: 3
  CKN for participant: E9AC
    Key server            : Yes
    MI (MN)               : 85E004AF49934720AC5131D3 (182)
    Live peers            : 1
    Potential peers       : 0
    Principal actor       : Yes
    MKA session status    : Secured
    Confidentiality offset: 30 bytes
    Tx-SSCI               : N/A
    Current SAK status    : Rx & Tx
    Current SAK AN        : 0
    Current SAK KI (KN)   : 85E004AF49934720AC5131D300000003 (3)
    Previous SAK status   : N/A
    Previous SAK AN       : N/A
    Previous SAK KI (KN)  : N/A
    Live peer list:
    MI                        MN         Prio  Cap  Rx-SCI            Rx-SSCI
    12A1677D59DD211AE86A0128  182        10    3    00E0020000000106  N/A
```

Display MACsec information on GigabitEthernet 1/0/1 of Device B.

```
[DeviceB] display macsec interface gigabitethernet 1/0/1 verbose
Interface GigabitEthernet1/0/1
  Protect frames               : Yes
  Replay protection            : Enabled
  Config replay window size    : 100 frames
  Active replay window size    : 100 frames
  Config confidentiality offset : 30 bytes
  Active confidentiality offset : 30 bytes
  Config validation mode       : Strict
  Active validation mode       : Strict
  Included SCI                 : No
  SCI conflict                 : No
  Cipher suite                 : GCM-AES-128
  MKA life time                : 6 seconds
```

```
  Transmit secure channel:
    SCI           : 00E0020000000106
      Elapsed time: 00h:05m:36s
      Current SA  : AN 0        PN 1
  Receive secure channels:
    SCI           : 00E00100000A0006
      Elapsed time: 00h:03m:21s
      Current SA  : AN 0        LPN 1
      Previous SA : AN N/A      LPN N/A
```

Display MKA session information on GigabitEthernet 1/0/1 of Device B.

```
[DeviceB] display mka session interface gigabitethernet 1/0/1 verbose
Interface GigabitEthernet1/0/1
Tx-SCI    : 00E0020000000106
Priority  : 10
Capability: 3
  CKN for participant: E9AC
    Key server            : No
    MI (MN)               : 12A1677D59DD211AE86A0128 (1219)
    Live peers            : 1
    Potential peers       : 0
    Principal actor       : Yes
    MKA session status    : Secured
    Confidentiality offset: 30 bytes
    Tx-SSCI               : N/A
    Current SAK status    : Rx & Tx
    Current SAK AN        : 0
    Current SAK KI (KN)   : 85E004AF49934720AC5131D300000003 (3)
    Previous SAK status   : N/A
    Previous SAK AN       : N/A
    Previous SAK KI (KN)  : N/A
    Live peer list:
    MI                        MN         Prio  Cap  Rx-SCI            Rx-SSCI
    85E004AF49934720AC5131D3  1216       5     3    00E00100000A0006  N/A
```

Troubleshooting MACsec

Cannot establish MKA sessions between MACsec devices

Symptom

The devices cannot establish MKA sessions when the following conditions exist:

- The link connecting the devices is up.
- The ports at the ends of the link are MACsec capable.

Analysis

The symptom might occur for the following reasons:

- The ports at the link are not enabled with MKA.

- A port at the link is not configured with a preshared key or configured with a preshared key different from the peer.

Solution

To resolve the issue:

1. Enter interface view.
2. Use the **display this** command to check the MACsec configuration:
   - If MKA is not enabled on the port, execute the **mka enable** command.
   - If a preshared key is not configured or the preshared key is different from the peer, use the **mka psk** command to configure a preshared key. Make sure the preshared key is the same as the preshared key on the peer.
3. If the issue persists, contact H3C Support.

Contents

Configuring an 802.1X client 1

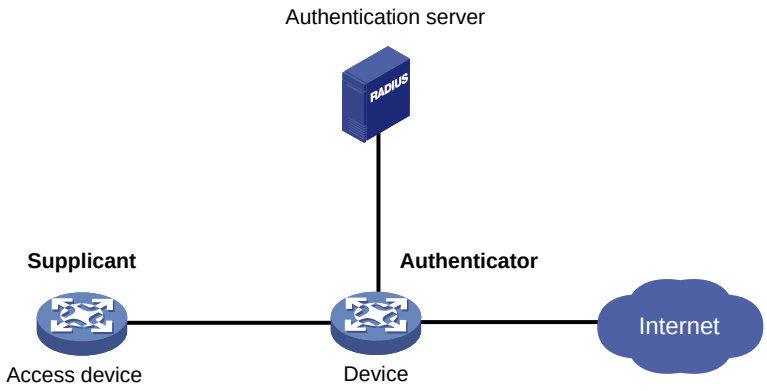
About 802.1X clients 1
802.1X client tasks at a glance 1
Enabling the 802.1X client feature 1
Enabling the 802.1X client feature in Ethernet interface view 1
Configuring an 802.1X client username and password 2
Specifying an 802.1X client EAP authentication method 2
Configuring an 802.1X client MAC address 3
Specifying an 802.1X client mode for sending EAP-Response and EAPOL-Logoff packets 4
Configuring an 802.1X client anonymous identifier 4
Specifying an SSL client policy 5
Display and maintenance commands for 802.1X client 5

Configuring an 802.1X client

About 802.1X clients

As shown in Figure 1, the 802.1X client feature allows the access device to act as the supplicant in the 802.1X architecture. For information about the 802.1X architecture, see "802.1X overview."

Figure 1 802.1X client network diagram



802.1X client tasks at a glance

To configure an 802.1X client, perform the following tasks:

1. Enabling the 802.1X client feature
2. Configuring an 802.1X client username and password
3. Specifying an 802.1X client EAP authentication method
4. (Optional.) Configuring an 802.1X client MAC address
5. (Optional.) Specifying an 802.1X client mode for sending EAP-Response and EAPOL-Logoff packets
6. (Optional.) Configuring an 802.1X client anonymous identifier
7. Specifying an SSL client policy

   This task is required when you specify PEAP-MSCHAPv2, PEAP-GTC, TTLS-MSCHAPv2, or TTLS-GTC authentication as the 802.1X client EAP authentication method.

Enabling the 802.1X client feature

Enabling the 802.1X client feature in Ethernet interface view

1. Enter system view.

   `system-view`
2. Enter Ethernet interface view.

   `interface` *`interface-type interface-number`*
3. Enable the 802.1X client feature.

   `dot1x supplicant enable`

   By default, the 802.1X client feature is disabled.

Configuring an 802.1X client username and password

Restrictions and guidelines

To ensure successful authentication, make sure the username and password configured on the 802.1X client is consistent with the username and password configured on the authentication server.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Configure an 802.1X client username.
   dot1x supplicant username *username*
   By default, no 802.1X client username is configured.
4. Set an 802.1X client password.
   dot1x supplicant password { **cipher** | **simple** } *string*
   By default, no 802.1X client password is configured.

Specifying an 802.1X client EAP authentication method

About this task

The following EAP authentication methods are available for the 802.1X client feature:

- MD5-Challenge.
- PEAP-MSCHAPv2.
- PEAP-GTC.
- TTLS-MSCHAPv2.
- TTLS-GTC.

Restrictions and guidelines

The following matrix shows the restrictions for the selection of authentication methods on the 802.1X client and the authenticator:

Authentication method specified on the 802.1X client	Packet exchange method specified on the authenticator
MD5-Challenge	• CHAP • EAP
• PEAP-MSCHAPv2 • PEAP-GTC • TTLS-MSCHAPv2	EAP

Authentication method specified on the 802.1X client	Packet exchange method specified on the authenticator
• TTLS-GTC	

For information about 802.1X packet exchange methods, see "Configuring 802.1X."

Make sure the specified 802.1X client EAP authentication method is supported by the authentication server.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Specify an 802.1X client EAP authentication method.
   dot1x supplicant eap-method { **md5** | **peap-gtc** | **peap-mschapv2** | **ttls-gtc** | **ttls-mschapv2** }
   By default, the EAP authentication method is MD5-Challenge.

Configuring an 802.1X client MAC address

About this task

The authenticator adds the MAC address of an authenticated 802.1X client to the MAC address table and then assigns access rights to the client.

If the device has multiple Ethernet interfaces that act as 802.1X clients to seek MACsec protection, configure a unique MAC address for each interface to ensure successful 802.1X client authentication. For information about MACsec, see "Configuring MACsec."

You can use either of the following methods to configure a unique MAC address for each interface:

- Execute the **mac-address** command in Ethernet interface view. For information about this command, see *Layer 2—LAN Switching Command Reference*.
- Configure an 802.1X client MAC address.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Configure an 802.1X client MAC address.
   dot1x supplicant mac-address *mac-address*
   By default, the 802.1X client on an Ethernet interface uses the MAC address of the interface for 802.1X authentication. If the interface's MAC address is unavailable, the client uses the device's MAC address for 802.1X authentication.

Specifying an 802.1X client mode for sending EAP-Response and EAPOL-Logoff packets

About this task

802.1X client authentication supports unicast and multicast modes to send EAP-Response and EAPOL-Logoff packets. As a best practice, use multicast mode to avoid 802.1X client authentication failures if the NAS device in the network does not support receiving unicast EAP-Response or EAPOL-Logoff packets.

Procedure

1. Enter system view.

   system-view

2. Enter Ethernet interface view.

   interface *interface-type interface-number*

3. Specify a mode for 802.1X client authentication to send EAP-Response and EAPOL-Logoff packets.

   dot1x supplicant transmit-mode { **multicast** | **unicast** }

   By default, 802.1X client authentication uses unicast mode to send EAP-Response and EAPOL-Logoff packets.

Configuring an 802.1X client anonymous identifier

About this task

At phase 1, packets sent to the authenticator are not encrypted. The use of an 802.1X client anonymous identifier prevents the 802.1X client username from being disclosed at phase 1. The 802.1X client sends the anonymous identifier to the authenticator instead of the 802.1X client username. The 802.1X client username will be sent to the authenticator in encrypted packets at phase 2.

If no 802.1X client anonymous identifier is configured, the 802.1X client sends the 802.1X client username at phase 1.

The configured 802.1X client anonymous identifier takes effect only if one of the following EAP authentication methods is used:

- PEAP-MSCHAPv2.
- PEAP-GTC.
- TTLS-MSCHAPv2.
- TTLS-GTC.

If the MD5-Challenge EAP authentication is used, the configured 802.1X client anonymous identifier does not take effect. The 802.1X client uses the 802.1X client username at phase 1.

Restrictions and guidelines

Do not configure the 802.1X client anonymous identifier if the vendor-specific authentication server cannot identify anonymous identifiers.

Procedure

1. Enter system view.

   system-view

2. Enter Ethernet interface view.

interface *interface-type interface-number*

3. Configure an 802.1X client anonymous identifier.
   dot1x supplicant anonymous identify *identifier*
   By default, no 802.1X client anonymous identifier is configured.

Specifying an SSL client policy

About this task

If the PEAP-MSCHAPv2, PEAP-GTC, TTLS-MSCHAPv2, or TTLS-GTC authentication is used, the 802.1X client authentication process is as follows:

- **Phase 1**—The 802.1X client acts as an SSL client to negotiate with the SSL server.
  The SSL client uses the SSL parameters defined in the specified SSL client policy to establish a connection with the SSL server for negotiation. The SSL parameters include a PKI domain, supported cipher suites, and the SSL version. For information about SSL client policy configuration, see "Configuring SSL."
- **Phase 2**—The 802.1X client uses the negotiated result to encrypt and transmit the interchanged authentication packets.

If the MD5-Challenge authentication is used, the 802.1X client does not use an SSL client policy during the authentication process.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Specify an SSL client policy.
   dot1x supplicant ssl-client-policy *policy-name*
   By default, the default SSL client policy is used.

Display and maintenance commands for 802.1X client

Execute **display** commands in any view.

Task	Command
Display 802.1X client information.	**display dot1x supplicant** [**interface** *interface-type interface-number*]

Contents

Configuring microsegmentation 1

About microsegmentation 1
Basic concepts 1
Components of microsegmentation 1
How microsegmentation works 2
Restrictions and guidelines: Microsegmentation configuration 3
Microsegmentation tasks at a glance 3
Prerequisites for microsegmentation configuration 4
Configuring a microsegment 4
Configuring an aggregate microsegment 5
Configuring an ACL 5
Configuring a GBP 6
Configuring PBR 6
Configuring a QoS policy 7
Configuring packet filtering 8
Configuring the microsegment extended community attribute 8
Configuring the network address match method for microsegments 9
Enabling the device to identify the microsegments of both IPv4 and IPv6 packets 9
Enabling SNMP notifications for microsegmentation 10
Display and maintenance commands for microsegmentation 10
Microsegmentation configuration examples 10
Example: Configuring microsegmentation in an IP network 10
Example: Configuring microsegmentation in an EVPN VXLAN network 24

Configuring microsegmentation

About microsegmentation

The microsegmentation feature, also called group-based security segregation, controls traffic based on groups the traffic assigned to. For example, you can group servers in data centers based on specific criteria and apply traffic control policies to different groups.

Basic concepts

Microsegment

A microsegment groups endpoints (such as servers) based on specific criteria. Each microsegment has a globally unique ID.

Group-based policy

A group-based policy (GBP) is a microsegment-based traffic control policy can be implemented by using the following functions:

- **PBR**—A policy node corresponds to a GBP and controls communication between microsegments through the **`apply next-hop`** or **`apply output-interface null0`** action. For more information about PBR, see *Layer 3—IP Routing Configuration Guide*.
- **QoS policy**—A class-behavior association corresponds to a GBP and controls communication between microsegments through the **`filter deny`** or **`filter permit`** action in the traffic behavior. For more information about QoS policy, see QoS configuration *ACL and QoS Configuration Guide*.
- **Packet filtering**—A packet filter corresponds to a GBP and controls communication between microsegments through the permit or deny rule in an ACL. For more information about packet filtering, see ACL configuration *ACL and QoS Configuration Guide*.

Components of microsegmentation

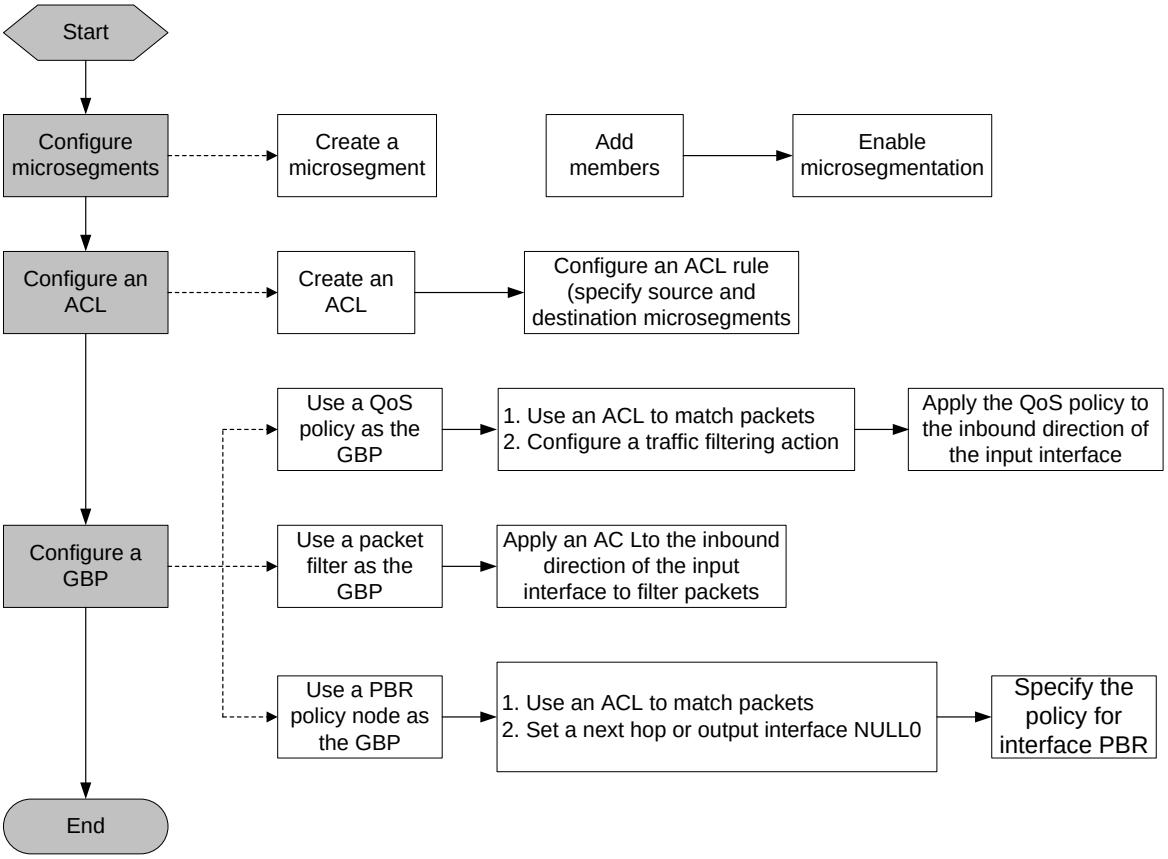
As shown in Figure 1, the microsegmentation feature contains the microsegment, ACL, and GBP settings. A GBP can be a QoS policy, a packet filter, or a PBR policy node.

This feature controls whether members in different microsegments can communicate. The GBP takes effect on the local end of a link. To control bidirectional traffic, configure this feature on both ends. Intermediate nodes do not require the configuration of this feature.

This feature can be used in IP, VXLAN, and EVPN networks. The configurations are slightly different in different networks.

- In an IP network, all settings must be configured on the Layer 3 gateway devices. In a VXLAN or EVPN network, all settings must be configured on the VTEPs.
- In an EVPN network, if the microsegment settings are automatically synchronized to the remote end through the BGP extended community attribute in the MAC/IP advertisement route, you do not need to configure microsegment settings on the remote end.

Figure 1 Microsegmentation configuration workflow



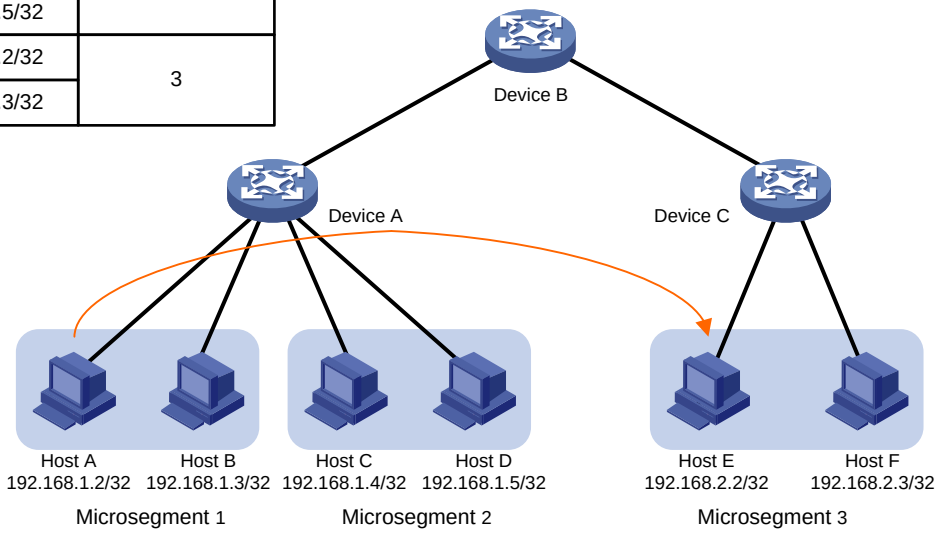
How microsegmentation works

The microsegmentation feature works in the same way in IP, VXLAN, and EVPN networks. As shown in Figure 2, this section takes unidirectional traffic in an IP network as an example to illustrate how this feature works. This example uses a QoS policy as the GBP.

1. After receiving a packet sent from Host A to Host D, Device A obtains its source IP address (192.168.1.2) and destination IP address (192.168.1.5).
2. Device A searches the FIB table for the source IP address according to the longest match rule and determines that Host A belongs to microsegment 1.
3. Device A searches the FIB table for the destination IP address according to the longest match rule and determines that Host B belongs to microsegment 2.
4. Device A matches microsegment 1 and microsegment 2 against ACLs and executes one of the following actions in the QoS policy on matching packets:
   - Forwards matching packets if the action is **`filter permit`**.
   - Drops matching packets if the action is **`filter deny`**.

Figure 2 Forwarding of Layer 3 packets in an IP network

IP address	Microsegment ID
192.168.1.2/32	1
192.168.1.3/32	
192.168.1.4/32	2
192.168.1.5/32	
192.168.2.2/32	3
192.168.2.3/32	



The microsegmentation feature works in the same way for cross-device packet forwarding.

Restrictions and guidelines: Microsegmentation configuration

The packet filtering feature and the microsegmentation feature are mutually exclusive on a VLAN interface. Only the feature configured first takes effect.

When you use this feature to control unidirectional inter-VPN traffic, follow these restrictions and guidelines:

- On the source PE device, if the route guiding traffic forwarding is a network route, you must add the destination address of the route to a microsegment as a member.
- On the destination PE device, if the route is a host route, you must add the destination address of the route to a microsegment as a member.

Microsegmentation tasks at a glance

To configure microsegmentation, perform the following tasks:

1. Configuring a microsegment
2. (Optional.) Configuring an aggregate microsegment
3. Configuring an ACLConfiguring an ACL
4. Configuring a GBP

   Choose one option as needed:

   - Configuring PBR
   - Configuring a QoS policy

- Configuring packet filtering

5. (Optional.) Configuring the microsegment extended community attribute

   Perform this task in an EVPN network to avoid extended community attribute conflicts.
6. (Optional.) Configuring the network address match method for microsegments
7. (Optional.) Enabling the device to identify the microsegments of both IPv4 and IPv6 packets
8. (Optional.) Enabling SNMP notifications for microsegmentation

Prerequisites for microsegmentation configuration

This feature can be used in IP, VXLAN, and EVPN networks. For information about configuring these features, see the relevant configuration guides.

Configuring a microsegment

Restrictions and guidelines

To control bidirectional traffic, follow these rules:

- In an IP or VXLAN network, configure the same microsegment settings on the two ends.
- In an EVPN network, configure a microsegment only on the local end. When distributing a MAC/IP advertisement route, the device carries the microsegment ID in the BGP extended community attribute. If the members (IP addresses) are in the distributed route, the microsegment settings are automatically synchronized to the remote end. The synchronized microsegment settings directly take effect on the remote end and are not subject to the **microsegment enable** command. If you also configure a microsegment on the remote end,, the synchronized microsegment settings overwrite the configured ones in the case of any setting differences.

  For more information about MAC/IP advertisement routes, see EVPN overview in *EVPN Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Create a microsegment and enter microsegment view.

   microsegment *microsegment-id* [**name** *microsegment-name*]
3. Add a member to the microsegment.

   member ipv4 *ipv4-address* { *mask* | *mask-length* } [**vpn-instance** *vpn-instance-name*]

   member ipv6 *ipv6-address prefix-length* [**vpn-instance** *vpn-instance-name*]

   By default, a microsegment does not contain members.
4. (Optional.) Add users connected through the corresponding AC by mapping the microsegment interface to a VSI.

   xconnect vsi

   By default, a microsegment is not bound to any VSI.

   The information about added users is not displayed in the **display microsegment** command output.

   For more information about the **xconnect vsi** command, see *VXLAN Command Reference*.
5. Return to system view.

quit

6. Enable microsegmentation.

   microsegment enable

   By default, microsegmentation is disabled.

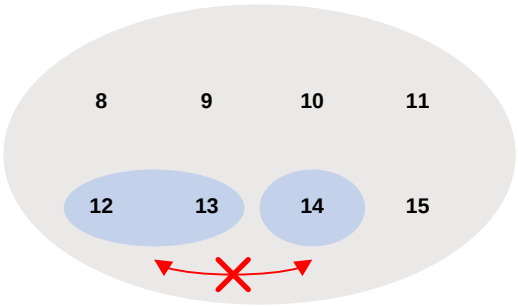
Configuring an aggregate microsegment

About this task

An aggregate microsegment is a group of microsegments with contiguous IDs. The ID of the aggregate microsegment is the start microsegment ID. You can use a mask to specify the member microsegments for an aggregate microsegment.

As shown in Figure 3, microsegments 8 through 15 can communicate with one another. To prevent communication between microsegments 12 and 14 and between microsegments 13 and 14, you can use an aggregate microsegment instead of reconfiguring microsegment settings. In this example, you can combine microsegments 12 and 13 to create an aggregate microsegment with ID 12, and use a GBP to prevent communication between aggregate microsegment 12 and microsegment 14.

Figure 3 Microsegment aggregation



Procedure

1. Enter system view.

   system-view

2. Create an aggregate microsegment and enter its view.

   microsegment aggregation *aggregation-id* **mask-length** *mask-length* [**name** *aggregation-name*]

Configuring an ACL

Restrictions and guidelines

To control bidirectional traffic, you must configure an ACL on both ends and configure an ACL rule with swapped source and destination microsegments on the two ends.

If you use a PBR policy node or a QoS policy as the GBP, the ACL rules must be permit rules. The apply action or QoS action is taken on matching packets.

If you use a packet filter as the GBP, the ACL rules can be permit or deny rules. Matching packets are permitted or denied.

If an ACL does not contain the **vpn-instance** *vpn-instance-name* option, it matches only non-VPN packets.

Procedure

1. Enter system view.
   system-view
2. Create an IPv4 or IPv6 advanced ACL and enter its view. Choose one option as needed:
   - **acl** [**ipv6**] **number** *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]
   - **acl** [**ipv6**] { **advanced** | **basic** } { *acl-number* | **name** *acl-name* } [**match-order** { **auto** | **config** }]
3. Configure a rule.
   For more information, see the **rule** command in *ACL and QoS Command Reference*.
   In the **rule** command, the **destination microsegment** *microsegment-id* and **source microsegment** *microsegment-id* options must be specified, and other parameters can be configured as needed.

Configuring a GBP

Configuring PBR

About this task

You can control communication between microsegments by referencing an ACL and specifying a next hop (permitting traffic) or the output interface NULL0 (dropping traffic) in a PBR policy.

For more information about PBR, see *Layer 3—IP Routing Configuration Guide*.

Restrictions and guidelines

To control bidirectional traffic, you must configure PBR on both ends.

Procedure

1. Enter system view.
   system-view
2. Create a node for a policy, and enter its view.
   policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*
3. Set an ACL match criterion for the node.
   if-match acl { *acl-number* | **name** *acl-name* }
   By default, no ACL match criterion is set.
4. Configure an action for the node. Choose one option as needed:
   - Set a next hop.
     apply next-hop *ip-address*
   - Set NULL0 as the output interface.
     apply output-interface null0

   By default, no action is configured.
5. Return to system view.
   quit
6. Enter interface view.
   interface *interface-type interface-number*
7. Specify the policy for interface PBR.

ip policy-based-route *policy-name*

By default, no interface policy is applied to an interface.

Configuring a QoS policy

About this task

You can use the traffic filtering action in a QoS policy to control communication between microsegments.

Procedure

1. Enter system view.
   system-view
2. Define a traffic class.
   a. Create a traffic class and enter traffic class view.
      traffic classifier *classifier-name* [**operator** { **and** | **or** }]
   b. Configure a match criterion.
      if-match acl [**ipv6**] { *acl-number* | **name** *acl-name* }
      By default, no match criterion is configured.
      Only IPv4 and IPv6 advanced ACLs can be used to match packets.
   c. Return to system view.
      quit
3. Define a traffic behavior.
   a. Create a traffic behavior and enter traffic behavior view.
      traffic behavior *behavior-name*
   b. Configure a traffic filtering action.
      filter { **deny** | **permit** }
      By default, no traffic filtering action is configured.
   c. Return to system view.
      quit
4. Define a QoS policy.
   a. Create a QoS policy and enter QoS policy view.
      qos policy *policy-name*
   b. Associate the traffic class with the traffic behavior in the QoS policy.
      classifier *classifier-name* **behavior** *behavior-name*
      By default, a traffic class is not associated with a traffic behavior.
   c. Return to system view.
      quit
5. Apply the QoS policy to an interface.
   a. Enter interface view.
      interface *interface-type interface-number*
   b. Apply the QoS policy to the inbound direction of the interface.
      qos apply policy *policy-name* **inbound** [**share-mode**]
      By default, no QoS policy is applied to an interface.

Configuring packet filtering

About this task

You can apply an ACL to the inbound direction of an interface to control communication between microsegments.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Apply an ACL to the inbound direction of the interface.
   packet-filter [**ipv6**] { *acl-number* | **name** *acl-name* } **inbound**
   By default, no ACL is applied to an interface.

Configuring the microsegment extended community attribute

About this task

A MAC/IP advertisement route carries microsegment IDs in a BGP extended community attribute and advertises microsegment settings to a peer through the extended community attribute. For more information about MAC/IP advertisement routes, see EVPN overview in *EVPN Configuration Guide*.

To avoid attribute conflicts, you can perform this task to modify the microsegment extended community attribute value.

Procedure

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Set the microsegment extended community attribute value.
   extcommunity-type microsegment-id *microsegment-type-value*
   The default setting is 83ff (hexadecimal).
4. (Optional.) Display the microsegment extended community attribute value in BGP EVPN routes and the microsegment ID.
   display bgp l2vpn evpn
   For more information about this command, see EVPN commands in *EVPN Command Reference*.
5. (Optional.) Display the microsegment IDs carried in routes.
   display ip routing-table
   For more information about this command, see IP basics commands in *Layer 3—IP Routing Command Reference*.

Configuring the network address match method for microsegments

About this task

The device determines the segment membership of packets by matching the source and destination IP addresses of packets. The following match methods are available:

- **Exact match**—The mask lengths of the source and destination IP addresses must be equal to those of members in microsegments. For example, a packet sourced from 10.10.10.1/24 matches member 10.10.10.0/24 instead of 10.10.10.0/23.
- **Longest match**—The mask lengths of the source and destination IP addresses can be greater than or equal to those of members in microsegments. For example, a packet sourced from 10.10.10.1/24 matches member 10.10.10.0/16.

The device uses different match methods for different member types of microsegments:

- Host addresses (IPv4 addresses with a 32-bit mask and IPv6 addresses with a 128-bit prefix) use the longest match method, which cannot be modified.
- The default route (0.0.0.0/0 or 0::0/0) uses the exact match method, which cannot be modified.
- Network addresses (IPv4 addresses with a 1-bit to 31-bit mask and IPv6 addresses with a 1-bit to 127-bit prefix) use the exact match method by default. If you execute the **`microsegment subnet-match longest`** command, the longest match method is used. If you execute the **`microsegment subnet-match disable`** command, no microsegment member can be matched.

The longest match method helps you simplify configuration when you need to add a large number of network addresses to a microsegment. For example, to match network addresses 10.10.10.0/24, 10.10.20.0/24, and 10.10.30.0/24 to microsegment 1, you need to execute only the **`member ipv4 10.10.10.0/16`** command if you use longest match.

Procedure

1. Enter system view.

   `system-view`
2. Configure the network address match method for microsegments.

   `microsegment subnet-match` { **`disable`** | **`longest`** }

   By default, exact match is used for network addresses.

Enabling the device to identify the microsegments of both IPv4 and IPv6 packets

About this task

When receiving a packet, the device uses ACL resources to match it and determines the microsegment that it belongs to according to its source and destination IP addresses. Then, the device implements traffic control according to the GBP.

By default, the device can only match IPv4 packets and determine the microsegments to which they belong. The device cannot identify the microsegments to which IPv6 packets belong. Even if a microsegment contains an IPv6 member, the device cannot implement traffic control for IPv6 packets according to the GBP. Perform this task to achieve traffic control for both IPv4 and IPv6 packets based on microsegment IDs.

To save ACL resources, do not perform this task if traffic control for IPv6 packets is not required.

Procedure

1. Enter system view.
   system-view
2. Enable the device to identify the microsegments of both IPv4 and IPv6 packets.
   microsegment dual-stack-mode
   By default, the device can only identify the microsegment to which IPv4 packets belong.

Enabling SNMP notifications for microsegmentation

About this task

To report critical microsegmentation events to an SNMP NMS, enable SNMP notifications for microsegmentation.

For microsegmentation SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about configuring SNMP, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for microsegmentation.
   snmp-agent trap enable microsegment
   By default, SNMP notifications are enabled for microsegmentation.

Display and maintenance commands for microsegmentation

Execute **display** commands in any view.

Task	Command
Display aggregate microsegment configuration.	**display microsegment aggregation** [*aggregation-id* | **name** *aggregation-name*]
Display microsegment configuration.	**display microsegment** [*microsegment-id* | **name** *microsegment-name*]

Microsegmentation configuration examples

Example: Configuring microsegmentation in an IP network

Network configuration

As shown in Figure 4, configure microsegmentation to meet the following requirements:

- Host A, Host B, and the general server can access the file server.

- Host A and the general server cannot access each other. Host B and the general server cannot access each other.

Figure 4 Network diagram

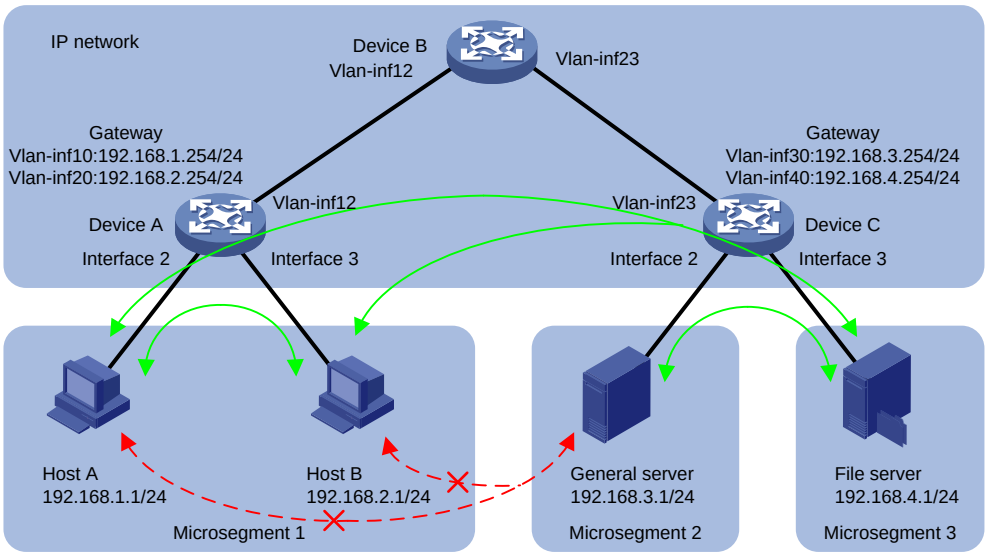


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Analysis

1. Add Host A and Host B to microsegment 1.
2. Add the general server to microsegment 2, and add the file server to microsegment 3.
3. Configure a QoS policy to allow microsegment 1 and microsegment 3 to communicate, allow microsegment 2 and microsegment 3 to communicate, and prevent microsegment 1 and microsegment 2 from communicating.

Restrictions and guidelines

For microsegmentation to operate correctly, execute the **system-working-mode expert** command in system view to set the system operating mode to expert.

For a microsegment in an ACL rule to match packets successfully, make sure the mask length of the microsegment member is the same as the mask length for the source or destination IP address of the packets.

- If the mask length of the microsegment member is shorter than the mask length for the source or destination IP address of the packets, you must specify the longest match method to match the packets successfully.
- If the mask length of the microsegment member is longer than the mask length for the source or destination IP address of the packets, the microsegment member cannot match them.

Procedure

1. Configure Device A:

   # Create VLANs and configure IP address for VLAN interfaces.

   ```
   <Sysname> system-view
   [Sysname] sysname DeviceA
   ```

```
[DeviceA] vlan 10
[DeviceA-vlan10] port gigabitethernet 1/0/2
[DeviceA-vlan10] quit
[DeviceA] interface vlan-interface 10
[DeviceA-Vlan-interface10] ip address 192.168.1.254 24
[DeviceA-Vlan-interface10] undo shutdown
[DeviceA-Vlan-interface10] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] undo shutdown
[DeviceA-GigabitEthernet1/0/2] quit
[DeviceA] vlan 20
[DeviceA-vlan20] port gigabitethernet 1/0/3
[DeviceA-vlan20] quit
[DeviceA] interface vlan-interface 20
[DeviceA-Vlan-interface20] ip address 192.168.2.254 24
[DeviceA-Vlan-interface20] undo shutdown
[DeviceA-Vlan-interface20] quit
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] undo shutdown
[DeviceA-GigabitEthernet1/0/3] quit
[DeviceA] vlan 12
[DeviceA-vlan12] quit
[DeviceA] interface vlan-interface 12
[DeviceA-Vlan-interface12] ip address 12.0.0.2 30
[DeviceA-Vlan-interface12] undo shutdown
[DeviceA-Vlan-interface12] quit
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo shutdown
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 12
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure IS-IS to implement Layer 3 interconnectivity.

```
[DeviceA] isis 1
[DeviceA-isis-1] cost-style wide
[DeviceA-isis-1] network-entity 00.0000.0000.0001.00
[DeviceA-isis-1] is-level level-2
[DeviceA-isis-1] address-family ipv4 unicast
[DeviceA-isis-1-ipv4] quit
[DeviceA-isis-1] quit
[DeviceA] interface vlan-interface 10
[DeviceA-Vlan-interface20] isis enable 1
[DeviceA-Vlan-interface20] quit
[DeviceA] interface vlan-interface 20
[DeviceA-Vlan-interface30] isis enable 1
[DeviceA-Vlan-interface30] quit
[DeviceA] interface vlan-interface 12
[DeviceA-Vlan-interface12] isis enable 1
[DeviceA-Vlan-interface12] quit
```

Create microsegment 1, and add the IP addresses of Host A and Host B as its members.

```
[DeviceA] microsegment 1 name EPG1
[DeviceA-microsegment-1] member ipv4 192.168.1.1 24
[DeviceA-microsegment-1] member ipv4 192.168.2.1 24
[DeviceA-microsegment-1] quit
```

Create microsegment 2, and add the IP address of the general server as its member.

```
[DeviceA] microsegment 2 name EPG2
[DeviceA-microsegment-2] member ipv4 192.168.3.1 24
[DeviceA-microsegment-2] quit
```

Create microsegment 3, and add the IP address of the file server as its member.

```
[DeviceA] microsegment 3 name EPG3
[DeviceA-microsegment-3] member ipv4 192.168.4.1 24
[DeviceA-microsegment-3] quit
```

Enable microsegmentation.

```
[DeviceA] microsegment enable
```

Create an IPv4 advanced ACL named **EPG1-EPG3**, configure a rule to match the IP packets from microsegment 1 to microsegment 3.

```
[DeviceA] acl advanced name EPG1-EPG3
[DeviceA-acl-ipv4-adv-EPG1-EPG3] rule 0 permit ip source microsegment 1 destination
microsegment 3
[Device-acl-ipv4-adv-EPG1-EPG3] quit
```

Create an IPv4 advanced ACL named **EPG1-EPG2**, configure a rule to match the IP packets from microsegment 1 to microsegment 2.

```
[DeviceA] acl advanced name EPG1-EPG2
[DeviceA-acl-ipv4-adv-EPG1-EPG2] rule 0 permit ip source microsegment 1 destination
microsegment 2
[DeviceA-acl-ipv4-adv-EPG1-EPG2] quit
```

Create a traffic class named **CLASSIFIER-GBP13**, and use ACL **EPG1-EPG3** as the match criterion.

```
[DeviceA] traffic classifier CLASSIFIER-GBP13
[DeviceA-classifier-CLASSIFIER-GBP13] if-match acl name EPG1-EPG3
[DeviceA-classifier-CLASSIFIER-GBP13] quit
```

Create a traffic behavior named **BEHAVIOR-GBP13**, configure a permit action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceA] traffic behavior BEHAVIOR-GBP13
[DeviceA-behavior-BEHAVIOR-GBP13] filter permit
[DeviceA-behavior-BEHAVIOR-GBP13] accounting packet
[DeviceA-behavior-BEHAVIOR-GBP13] quit
```

Create a traffic class named **CLASSIFIER-GBP12**, and use ACL **EPG1-EPG2** as the match criterion.

```
[DeviceA] traffic classifier CLASSIFIER-GBP12
[DeviceA-classifier-CLASSIFIER-GBP12] if-match acl name EPG1-EPG2
[DeviceA-classifier-CLASSIFIER-GBP12] quit
```

Create a traffic behavior named **BEHAVIOR-GBP12**, configure a deny action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceA] traffic behavior BEHAVIOR-GBP12
[DeviceA-behavior-BEHAVIOR-GBP12] filter deny
[DeviceA-behavior-BEHAVIOR-GBP12] accounting packet
[DeviceA-behavior-BEHAVIOR-GBP12] quit
```

Create a QoS policy named **GBP1**, and associate the configured traffic classes and traffic behaviors in the QoS policy.

```
[DeviceA] qos policy GBP1
[DeviceA-qospolicy-GBP1] classifier CLASSIFIER-GBP13 behavior BEHAVIOR-GBP13
[DeviceA-qospolicy-GBP1] classifier CLASSIFIER-GBP12 behavior BEHAVIOR-GBP12
[DeviceA-qospolicy-GBP1] quit
```

Apply QoS policy **GBP1** to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] qos apply policy GBP1 inbound
[DeviceA-GigabitEthernet1/0/2] quit
```

Apply QoS policy **GBP1** to the inbound direction of GigabitEthernet 1/0/3.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] qos apply policy GBP1 inbound
[DeviceA-GigabitEthernet1/0/3] quit
```

2. Configure Device B:

Create VLANs and configure IP address for VLAN interfaces.

```
<Sysname> system-view
[Sysname] sysname DeviceB
[DeviceB] vlan 12
[DeviceB-vlan12] quit
[DeviceB] interface vlan-interface 12
[DeviceB-Vlan-interface12] ip address 12.0.0.1 30
[DeviceB-Vlan-interface12] undo shutdown
[DeviceB-Vlan-interface12] quit
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] undo shutdown
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 12
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] vlan 23
[DeviceB-vlan23] quit
[DeviceB] interface vlan-interface 23
[DeviceB-Vlan-interface23] ip address 23.0.0.2 30
[DeviceB-Vlan-interface23] undo shutdown
[DeviceB-Vlan-interface23] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] undo shutdown
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 23
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure IS-IS to implement Layer 3 interconnectivity.

```
[DeviceB] isis 1
[DeviceB-isis-1] cost-style wide
[DeviceB-isis-1] network-entity 00.0000.0000.0002.00
[DeviceB-isis-1] is-level level-2
[DeviceB-isis-1] address-family ipv4 unicast
[DeviceB-isis-1-ipv4] quit
[DeviceB-isis-1] quit
```

```
[DeviceB] interface vlan-interface 12
[DeviceB-Vlan-interface12] isis enable 1
[DeviceB-Vlan-interface12] quit
[DeviceB] interface vlan-interface 23
[DeviceB-Vlan-interface23] isis enable 1
[DeviceB-Vlan-interface23] quit
```

3. Configure Device C:

Create VLANs and configure IP address for VLAN interfaces.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] vlan 30
[DeviceC-vlan30] port gigabitethernet 1/0/2
[DeviceC-vlan30] quit
[DeviceC] interface vlan-interface 30
[DeviceC-Vlan-interface30] ip address 192.168.3.254 24
[DeviceC-Vlan-interface30] undo shutdown
[DeviceC-Vlan-interface30] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo shutdown
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] vlan 40
[DeviceC-vlan40] port gigabitethernet 1/0/3
[DeviceC-vlan40] quit
[DeviceC] interface vlan-interface 40
[DeviceC-Vlan-interface40] ip address 192.168.2.254 24
[DeviceC-Vlan-interface40] undo shutdown
[DeviceC-Vlan-interface40] quit
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] undo shutdown
[DeviceC-GigabitEthernet1/0/3] quit
[DeviceC] vlan 23
[DeviceC-vlan23] quit
[DeviceC] interface vlan-interface 23
[DeviceC-Vlan-interface23] ip address 23.0.0.1 30
[DeviceC-Vlan-interface23] undo shutdown
[DeviceC-Vlan-interface23] quit
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo shutdown
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 23
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure IS-IS to implement Layer 3 interconnectivity.

```
[DeviceC] isis 1
[DeviceC-isis-1] cost-style wide
[DeviceC-isis-1] network-entity 00.0000.0000.0003.00
[DeviceC-isis-1] is-level level-2
[DeviceC-isis-1] address-family ipv4 unicast
[DeviceC-isis-1-ipv4] quit
```

```
[DeviceC-isis-1] quit
[DeviceC] interface vlan-interface 30
[DeviceC-Vlan-interface30] isis enable 1
[DeviceC-Vlan-interface30] quit
[DeviceC] interface vlan-interface 40
[DeviceC-Vlan-interface40] isis enable 1
[DeviceC-Vlan-interface40] quit
[DeviceC] interface vlan-interface 23
[DeviceC-Vlan-interface23] isis enable 1
[DeviceC-Vlan-interface23] quit
```

Create microsegment 1, and add the IP addresses of Host A and Host B as its members.

```
[DeviceC] microsegment 1 name EPG1
[DeviceC-microsegment-1] member ipv4 192.168.1.1 24
[DeviceC-microsegment-1] member ipv4 192.168.2.1 24
[DeviceC-microsegment-1] quit
```

Create microsegment 2, and add the IP address of the general server as its member.

```
[DeviceC] microsegment 2 name EPG2
[DeviceC-microsegment-2] member ipv4 192.168.3.1 24
[DeviceC-microsegment-2] quit
```

Create microsegment 3, and add the IP address of the file server as its member.

```
[DeviceC] microsegment 3 name EPG3
[DeviceC-microsegment-3] member ipv4 192.168.4.1 24
[DeviceC-microsegment-3] quit
```

Enable microsegmentation.

```
[DeviceC] microsegment enable
```

Create an IPv4 advanced ACL named **EPG3-EPG1**, configure a rule to match the IP packets from microsegment 3 to microsegment 1.

```
<DeviceC> system-view
[DeviceC] acl advanced name EPG3-EPG1
[DeviceC-acl-ipv4-adv-EPG3-EPG1] rule 0 permit ip source microsegment 3 destination
microsegment 1
[DeviceC-acl-ipv4-adv-EPG3-EPG1] quit
```

Create an IPv4 advanced ACL named **EPG2-EPG1**, configure a rule to match the IP packets from microsegment 2 to microsegment 1.

```
[DeviceC] acl advanced name EPG2-EPG1
[DeviceC-acl-ipv4-adv-EPG2-EPG1] rule 0 permit ip source microsegment 2 destination
microsegment 1
[DeviceC-acl-ipv4-adv-EPG2-EPG1] quit
```

Create an IPv4 advanced ACL named **EPG2-EPG3**, configure a rule to match the IP packets from microsegment 2 to microsegment 3.

```
[DeviceC] acl advanced name EPG2-EPG3
[DeviceC-acl-ipv4-adv-EPG2-EPG3] rule 0 permit ip source microsegment 2 destination
microsegment 3
[DeviceC-acl-ipv4-adv-EPG2-EPG3] quit
```

Create an IPv4 advanced ACL named **EPG3-EPG2**, configure a rule to match the IP packets from microsegment 3 to microsegment 2.

```
[DeviceC] acl advanced name EPG3-EPG2
[DeviceC-acl-ipv4-adv-EPG3-EPG2] rule 0 permit ip source microsegment 3 destination
microsegment 2
```

```
[DeviceC-acl-ipv4-adv-EPG3-EPG2] quit
```

Create a traffic class named **CLASSIFIER-GBP23**, and use ACL **EPG2-EPG3** as the match criterion.

```
[DeviceC] traffic classifier CLASSIFIER-GBP23
[DeviceC-classifier-CLASSIFIER-GBP23] if-match acl name EPG2-EPG3
[DeviceC-classifier-CLASSIFIER-GBP23] quit
```

Create a traffic behavior named **BEHAVIOR-GBP23**, configure a permit action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceC] traffic behavior BEHAVIOR-GBP23
[DeviceC-behavior-BEHAVIOR-GBP23] filter permit
[DeviceC-behavior-BEHAVIOR-GBP23] accounting packet
[DeviceC-behavior-BEHAVIOR-GBP23] quit
```

Create a traffic class named **CLASSIFIER-GBP21**, and use ACL **EPG2-EPG1** as the match criterion.

```
[DeviceC] traffic classifier CLASSIFIER-GBP21
[DeviceC-classifier-CLASSIFIER-GBP21] if-match acl name EPG2-EPG1
[DeviceC-classifier-CLASSIFIER-GBP21] quit
```

Create a traffic behavior named **BEHAVIOR-GBP21**, configure a deny action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceC] traffic behavior BEHAVIOR-GBP21
[DeviceC-behavior-BEHAVIOR-GBP21] filter deny
[DeviceC-behavior-BEHAVIOR-GBP21] accounting packet
[DeviceC-behavior-BEHAVIOR-GBP21] quit
```

Create a QoS policy named **GBP2**, and associate the configured traffic classes and traffic behaviors in the QoS policy.

```
[DeviceC] qos policy GBP2
[DeviceC-qospolicy-GBP2] classifier CLASSIFIER-GBP23 behavior BEHAVIOR-GBP23
[DeviceC-qospolicy-GBP2] classifier CLASSIFIER-GBP21 behavior BEHAVIOR-GBP21
[DeviceC-qospolicy-GBP2] quit
```

Create a traffic class named **CLASSIFIER-GBP31**, and use ACL **EPG3-EPG1** as the match criterion.

```
[DeviceC] traffic classifier CLASSIFIER-GBP31
[DeviceC-classifier-CLASSIFIER-GBP31] if-match acl name EPG3-EPG1
[DeviceC-classifier-CLASSIFIER-GBP31] quit
```

Create a traffic behavior named **BEHAVIOR-GBP31**, configure a permit action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceC] traffic behavior BEHAVIOR-GBP31
[DeviceC-behavior-BEHAVIOR-GBP31] filter permit
[DeviceC-behavior-BEHAVIOR-GBP31] accounting packet
[DeviceC-behavior-BEHAVIOR-GBP31] quit
```

Create a traffic class named **CLASSIFIER-GBP32**, and use ACL **EPG3-EPG2** as the match criterion.

```
[DeviceC] traffic classifier CLASSIFIER-GBP32
[DeviceC-classifier-CLASSIFIER-GBP32] if-match acl name EPG3-EPG2
[DeviceC-classifier-CLASSIFIER-GBP32] quit
```

Create a traffic behavior named **BEHAVIOR-GBP32**, configure a permit action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceC] traffic behavior BEHAVIOR-GBP32
[DeviceC-behavior-BEHAVIOR-GBP32] filter permit
```

```
[DeviceC-behavior-BEHAVIOR-GBP32] accounting packet
[DeviceC-behavior-BEHAVIOR-GBP32] quit
```

Create a QoS policy named **GBP3**, and associate the configured traffic classes and traffic behaviors in the QoS policy.

```
[DeviceC] qos policy GBP3
[DeviceC-qospolicy-GBP3] classifier CLASSIFIER-GBP31 behavior BEHAVIOR-GBP31
[DeviceC-qospolicy-GBP3] classifier CLASSIFIER-GBP32 behavior BEHAVIOR-GBP32
[DeviceC-qospolicy-GBP3] quit
```

Apply QoS policy **GBP2** to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] qos apply policy GBP2 inbound
[DeviceC-GigabitEthernet1/0/2] quit
```

Apply QoS policy **GBP3** to the inbound direction of GigabitEthernet 1/0/3.

```
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] qos apply policy GBP3 inbound
[DeviceC-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Verify that Host A and Host B cannot successfully ping the general server.

```
C:\> ping 192.168.3.1

Pinging 192.168.3.1 with 32 bytes of data:

Request timed out
Request timed out
Request timed out
Request timed out

Ping statistics for 192.168.3.1:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),
```

The output shows that Host A and Host B cannot successfully ping the general server.

Verify that Host A and Host B can successfully ping the file server.

```
C:\> ping 192.168.4.1

Pinging 192.168.40.100 with 32 bytes of data:

Reply from 192.168.10.100: bytes=32 time=1ms TTL=255
Reply from 192.168.10.100: bytes=32 time<1ms TTL=255
Reply from 192.168.10.100: bytes=32 time<1ms TTL=255
Reply from 192.168.10.100: bytes=32 time<1ms TTL=255

Ping statistics for 192.168.10.100:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 1ms, Average = 0ms
```

The output shows that Host A and Host B can successfully ping the file server.

Display the configuration of each microsegment on Device A.

```
[DeviceA] display microsegment 1
```

```
Microsegment ID    : 1
Microsegment name  : EPG1
  IPv4 member:
    192.168.1.1/32
    192.168.2.1/32
[DeviceA] display microsegment 2
Microsegment ID    : 2
Microsegment name  : EPG2
  IPv4 member:
    192.168.3.1/32
[DeviceA] display microsegment 3
Microsegment ID    : 3
Microsegment name  : EPG3
  IPv4 member:
    192.168.4.1/32
```

Display summary information about microsegments on Device A.

```
[DeviceA] display microsegment
Microsegment status: Enabled
Total microsegments: 3
Microsegment list  :
  Microsegment ID  Members  Microsegment name
  1                2        EPG1
  2                1        EPG2
  3                1        EPG3
[DeviceA] display qos policy interface
Interface: GigabitEthernet1/0/2
  Direction: Inbound
  Policy: GBP1
   Classifier: CLASSIFIER-GBP13
     Operator: AND
     Rule(s) :
      If-match acl name EPG1-EPG3
     Behavior: BEHAVIOR-GBP13
      Accounting enable:
        5 (Packets)
        0 (pps)
      Filter enable: Permit
   Classifier: CLASSIFIER-GBP12
     Operator: AND
     Rule(s) :
      If-match acl name EPG1-EPG2
     Behavior: BEHAVIOR-GBP12
      Accounting enable:
        5 (Packets)
        0 (pps)
      Filter enable: Deny

Interface: GigabitEthernet1/0/3
```

```
Direction: Inbound
Policy: GBP1
 Classifier: CLASSIFIER-GBP13
   Operator: AND
   Rule(s) :
    If-match acl name EPG1-EPG3
   Behavior: BEHAVIOR-GBP13
    Accounting enable:
      5 (Packets)
      0 (pps)
    Filter enable: Permit
 Classifier: CLASSIFIER-GBP12
   Operator: AND
   Rule(s) :
    If-match acl name EPG1-EPG2
   Behavior: BEHAVIOR-GBP12
    Accounting enable:
      5 (Packets)
      0 (pps)
    Filter enable: Deny
```

Display the configuration of microsegments, and ACLs, and QoS policies on Device D. (Details not shown.)

Configuration files

- Device A:

```
#
 sysname DeviceA
#
isis 1
 is-level level-2
 cost-style wide
 network-entity 00.0000.0000.0001.00
 #
 address-family ipv4 unicast
#
vlan 10
#
vlan 12
#
vlan 20
#
 microsegment enable
#
microsegment 1 name EPG1
 member ipv4 192.168.1.0 255.255.255.0
 member ipv4 192.168.2.0 255.255.255.0
#
microsegment 2 name EPG2
 member ipv4 192.168.3.0 255.255.255.0
```

```
#
microsegment 3 name EPG3
 member ipv4 192.168.4.0 255.255.255.0
#
traffic classifier CLASSIFIER-GBP12 operator and
 if-match acl name EPG1-EPG2
#
traffic classifier CLASSIFIER-GBP13 operator and
 if-match acl name EPG1-EPG3
#
traffic behavior BEHAVIOR-GBP12
 accounting packet
 filter deny
#
traffic behavior BEHAVIOR-GBP13
  accounting packet
  filter permit
#
qos policy GBP1
 classifier CLASSIFIER-GBP13 behavior BEHAVIOR-GBP13
 classifier CLASSIFIER-GBP12 behavior BEHAVIOR-GBP12
#
interface Vlan-interface10
 ip address 192.168.1.254 255.255.255.0
 isis enable 1
#
interface Vlan-interface12
 ip address 12.0.0.2 255.255.255.252
 isis enable 1
#
interface Vlan-interface20
 ip address 192.168.2.254 255.255.255.0
 isis enable 1
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 12
#
interface GigabitEthernet1/0/2
 port access vlan 10
 qos apply policy GBP1 inbound
#
interface GigabitEthernet1/0/3
 port access vlan 20
 qos apply policy GBP1 inbound
#
acl advanced name EPG1-EPG2
 rule 0 permit ip source microsegment 1 destination microsegment 2
```

```
#
acl advanced name EPG1-EPG3
 rule 0 permit ip source microsegment 1 destination microsegment 3
#
```

- Device B:

```
#
 sysname DeviceB
#
isis 1
 is-level level-2
 cost-style wide
 network-entity 00.0000.0000.0002.00
 #
 address-family ipv4 unicast
#
vlan 12
#
vlan 23
#
interface Vlan-interface12
 ip address 12.0.0.1 255.255.255.252
 isis enable 1
#
interface Vlan-interface23
 ip address 23.0.0.2 255.255.255.252
 isis enable 1
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 12
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 23
#
```

- Device C:

```
#
 sysname DeviceC
#
isis 1
 is-level level-2
 cost-style wide
 network-entity 00.0000.0000.0003.00
 #
 address-family ipv4 unicast
#
vlan 23
```

```
#
vlan 30
#
vlan 40
#
 microsegment enable
#
microsegment 1 name EPG1
 member ipv4 192.168.1.0 255.255.255.0
 member ipv4 192.168.2.0 255.255.255.0
#
microsegment 2 name EPG2
 member ipv4 192.168.3.0 255.255.255.0
#
microsegment 3 name EPG3
 member ipv4 192.168.4.0 255.255.255.0
#
traffic classifier CLASSIFIER-GBP21 operator and
 if-match acl name EPG2-EPG1
#
traffic classifier CLASSIFIER-GBP23 operator and
 if-match acl name EPG2-EPG3
#
traffic classifier CLASSIFIER-GBP31 operator and
 if-match acl name EPG3-EPG1
#
traffic classifier CLASSIFIER-GBP32 operator and
 if-match acl name EPG3-EPG2
#
traffic behavior BEHAVIOR-GBP21
 accounting packet
 filter deny
#
traffic behavior BEHAVIOR-GBP23
 accounting packet
 filter permit
#
traffic behavior BEHAVIOR-GBP31
 accounting packet
 filter permit
#
traffic behavior BEHAVIOR-GBP32
 accounting packet
 filter permit
#
qos policy GBP2
 classifier CLASSIFIER-GBP23 behavior BEHAVIOR-GBP23
 classifier CLASSIFIER-GBP21 behavior BEHAVIOR-GBP21
```

```
#
qos policy GBP3
 classifier CLASSIFIER-GBP31 behavior BEHAVIOR-GBP31
 classifier CLASSIFIER-GBP32 behavior BEHAVIOR-GBP32
#
interface Vlan-interface23
 ip address 23.0.0.1 255.255.255.252
 isis enable 1
#
interface Vlan-interface30
 ip address 192.168.3.254 255.255.255.0
 isis enable 1
#
interface Vlan-interface40
 ip address 192.168.2.254 255.255.255.0
 isis enable 1
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 23
#
interface GigabitEthernet1/0/2
 port access vlan 30
 qos apply policy GBP2 inbound
#
interface GigabitEthernet1/0/3
 port access vlan 40
 qos apply policy GBP3 inbound
#
acl advanced name EPG2-EPG1
 rule 0 permit ip source microsegment 2 destination microsegment 1
#
acl advanced name EPG2-EPG3
 rule 0 permit ip source microsegment 2 destination microsegment 3
#
acl advanced name EPG3-EPG1
 rule 0 permit ip source microsegment 3 destination microsegment 1
#
acl advanced name EPG3-EPG2
 rule 0 permit ip source microsegment 3 destination microsegment 2
#
```

Example: Configuring microsegmentation in an EVPN VXLAN network

Network configuration

Host A and Host B access the EVPN network via Device A. The general server accesses the EVPN network via Device C. Configure microsegmentation to meet the following requirements:

- Host A and Host B can access each other.
- Host A cannot access the general server. Host B can access the general server.

Figure 5 Network diagram

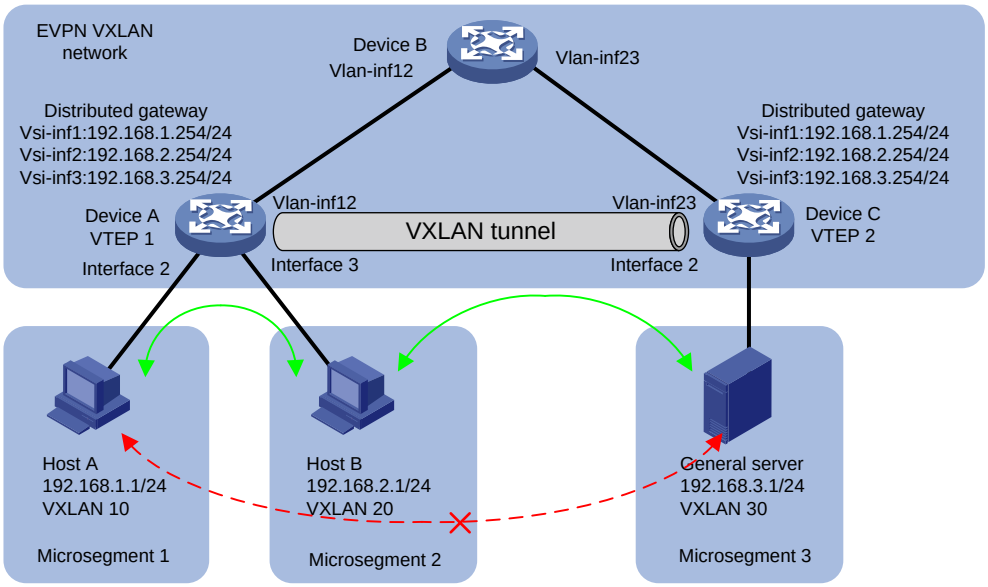


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Analysis

1. Add Host A to microsegment 1, Host B to microsegment 2, and the general server to microsegment 3.
2. Configure PBR to allow microsegment 1 and microsegment 2 to communicate, allow microsegment 2 and microsegment 3 to communicate, and prevent microsegment 1 and microsegment 3 from communicating.
3. Device A and Device C are distributed EVPN gateways. Host A belongs to VXLAN 10. Host B belongs to VXLAN 20. The general server belongs to VXLAN 30. Implement Layer 2 communication within the same VXLAN, and implement communication between different VXLANs through the distributed EVPN gateways.
4. Configure microsegment 1 and microsegment 2 on Device A, and configure microsegment 3 on Device C. Microsegment member information will be synchronized to the peer devices through BGP EVPN routes.

Restrictions and guidelines

For microsegmentation to operate correctly, execute the **`system-working-mode expert`** command in system view to set the system operating mode to expert.

For a microsegment in an ACL rule to match packets successfully, make sure the mask length of the microsegment member is the same as the mask length for the source or destination IP address of the packets.

Procedure

1. Configure Device A:

Create VLANs and configure IP address for VLAN interfaces.

```
<Sysname> system-view
[Sysname] sysname DeviceA
[DeviceA] vlan 10
[DeviceA-vlan10] port gigabitethernet 1/0/2
[DeviceA-vlan10] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] undo shutdown
[DeviceA-GigabitEthernet1/0/2] quit
[DeviceA] vlan 20
[DeviceA-vlan20] port gigabitethernet 1/0/3
[DeviceA-vlan20] quit
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] undo shutdown
[DeviceA-GigabitEthernet1/0/3] quit
[DeviceA] vlan 12
[DeviceA-vlan12] quit
[DeviceA] interface vlan-interface 12
[DeviceA-Vlan-interface12] ip address 12.0.0.2 30
[DeviceA-Vlan-interface12] undo shutdown
[DeviceA-Vlan-interface12] quit
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo shutdown
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 12
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure IS-IS to implement Layer 3 interconnectivity.

```
[DeviceA] isis 1
[DeviceA-isis-1] cost-style wide
[DeviceA-isis-1] network-entity 00.0000.0000.0001.00
[DeviceA-isis-1] is-level level-2
[DeviceA-isis-1] address-family ipv4 unicast
[DeviceA-isis-1-ipv4] quit
[DeviceA-isis-1] quit
[DeviceA] interface vlan-interface 12
[DeviceA-Vlan-interface12] isis enable 1
[DeviceA-Vlan-interface12] quit
[DeviceA] interface LoopBack0
[DeviceA-LoopBack0] ip address 1.1.1.1 32
[DeviceA-LoopBack0] isis enable 1
[DeviceA-LoopBack0] quit
```

Enable L2VPN.

```
[DeviceA] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[DeviceA] vxlan tunnel mac-learning disable
[DeviceA] vxlan tunnel arp-learning disable
```

Create an EVPN instance in VSI **vpna**, and configure the device to automatically generate an RD and a route target for the EVPN instance.

```
[DeviceA] vsi vpna
[DeviceA-vsi-vpna] evpn encapsulation vxlan
[DeviceA-vsi-vpna-evpn-vxlan] route-distinguisher auto
[DeviceA-vsi-vpna-evpn-vxlan] vpn-target auto
[DeviceA-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[DeviceA-vsi-vpna] vxlan 10
[DeviceA-vsi-vpna-vxlan-10] quit
[DeviceA-vsi-vpna] quit
```

Create an EVPN instance in VSI **vpnb**, and configure the device to automatically generate an RD and a route target for the EVPN instance.

```
[DeviceA] vsi vpnb
[DeviceA-vsi-vpnb] evpn encapsulation vxlan
[DeviceA-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[DeviceA-vsi-vpnb-evpn-vxlan] vpn-target auto
[DeviceA-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[DeviceA-vsi-vpnb] vxlan 20
[DeviceA-vsi-vpnb-vxlan-20] quit
[DeviceA-vsi-vpnb] quit
```

Create an EVPN instance in VSI **vpnc**, and configure the device to automatically generate an RD and a route target for the EVPN instance.

```
[DeviceA] vsi vpnc
[DeviceA-vsi-vpnc] evpn encapsulation vxlan
[DeviceA-vsi-vpnc-evpn-vxlan] route-distinguisher auto
[DeviceA-vsi-vpnc-evpn-vxlan] vpn-target auto
[DeviceA-vsi-vpnc-evpn-vxlan] quit
```

Create VXLAN 30.

```
[DeviceA-vsi-vpnc] vxlan 30
[DeviceA-vsi-vpnc-vxlan-30] quit
[DeviceA-vsi-vpnc] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[DeviceA] bgp 100
[DeviceA-bgp-default] peer 3.3.3.3 as-number 100
[DeviceA-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[DeviceA-bgp-default] address-family l2vpn evpn
[DeviceA-bgp-default-evpn] peer 3.3.3.3 enable
[DeviceA-bgp-default-evpn] quit
[DeviceA-bgp-default] quit
```

On GigabitEthernet1/0/2 connected to Host A, create Ethernet service instance 1000 and configure it to match data frames from VLAN 10.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] service-instance 1000
[DeviceA-GigabitEthernet1/0/2-srv1000] encapsulation s-vid 10
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[DeviceA-GigabitEthernet1/0/2-srv1000] xconnect vsi vpna
[DeviceA-GigabitEthernet1/0/2-srv1000] quit
```

On GigabitEthernet1/0/3 connected to Host B, create Ethernet service instance 2000 and configure it to match data frames from VLAN 20.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] service-instance 2000
[DeviceA-GigabitEthernet1/0/3-srv2000] encapsulation s-vid 20
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[DeviceA-GigabitEthernet1/0/3-srv2000] xconnect vsi vpnb
[DeviceA-GigabitEthernet1/0/3-srv2000] quit
[DeviceA-GigabitEthernet1/0/3] quit
```

Configure the RD and RT settings for VPN instance **l3vpna**.

```
[DeviceA] ip vpn-instance l3vpna
[DeviceA-vpn-instance-l3vpna] route-distinguisher 1:1
[DeviceA-vpn-instance-l3vpna] address-family ipv4
[DeviceA-vpn-ipv4-l3vpna] vpn-target 2:2
[DeviceA-vpn-ipv4-l3vpna] quit
[DeviceA-vpn-instance-l3vpna] address-family evpn
[DeviceA-vpn-evpn-l3vpna] vpn-target 1:1
[DeviceA-vpn-evpn-l3vpna] quit
[DeviceA-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[DeviceA] interface vsi-interface 1
[DeviceA-Vsi-interface1] ip binding vpn-instance l3vpna
[DeviceA-Vsi-interface1] ip address 192.168.1.254 24
[DeviceA-Vsi-interface1] mac-address 1-1-1
[DeviceA-Vsi-interface1] distributed-gateway local
[DeviceA-Vsi-interface1] local-proxy-arp enable
[DeviceA-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[DeviceA] interface vsi-interface 2
[DeviceA-Vsi-interface2] ip binding vpn-instance l3vpna
[DeviceA-Vsi-interface2] ip address 192.168.2.254 24
[DeviceA-Vsi-interface2] mac-address 2-2-2
[DeviceA-Vsi-interface2] distributed-gateway local
[DeviceA-Vsi-interface2] local-proxy-arp enable
[DeviceA-Vsi-interface2] quit
```

Configure VSI-interface 3.

```
[DeviceA] interface vsi-interface 3
[DeviceA-Vsi-interface1] ip binding vpn-instance l3vpna
[DeviceA-Vsi-interface1] ip address 192.168.3.254 24
[DeviceA-Vsi-interface1] mac-address 3-3-3
[DeviceA-Vsi-interface1] distributed-gateway local
[DeviceA-Vsi-interface1] local-proxy-arp enable
[DeviceA-Vsi-interface1] quit
```

Create VSI-interface 4, associate it with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[DeviceA] interface vsi-interface 4
[DeviceA-Vsi-interface4] ip binding vpn-instance l3vpna
[DeviceA-Vsi-interface4] l3-vni 1000
```

```
[DeviceA-Vsi-interface4] quit
```

Specify VSI-interface 1 as the gateway interface for the VSI of VXLAN 10.

```
[DeviceA] vsi vpna
[DeviceA-vsi-vpna] gateway vsi-interface 1
[DeviceA-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for the VSI of VXLAN 20.

```
[DeviceA] vsi vpnb
[DeviceA-vsi-vpnb] gateway vsi-interface 2
[DeviceA-vsi-vpnb] quit
```

Specify VSI-interface 3 as the gateway interface for the VSI of VXLAN 30.

```
[DeviceA] vsi vpnc
[DeviceA-vsi-vpnc] gateway vsi-interface 3
[DeviceA-vsi-vpnc] quit
```

Create microsegment 1, and add the IP address of Host A as its member.

```
[DeviceA] microsegment 1 name EPG1
[DeviceA-microsegment-1] member ipv4 192.168.1.1 24 vpn-instance l3vpna
[DeviceA-microsegment-1] quit
```

Create microsegment 2, and add the IP address of Host B as its member.

```
[DeviceA] microsegment 2 name EPG2
[DeviceA-microsegment-2] member ipv4 192.168.2.1 24 vpn-instance l3vpna
[DeviceA-microsegment-2] quit
```

Enable microsegmentation.

```
[DeviceA] microsegment enable
```

Create an IPv4 advanced ACL named **EPG1-EPG3**, configure a rule to match the IP packets of VPN instance **l3vpna** from microsegment 1 to microsegment 3.

```
[DeviceA] acl advanced name EPG1-EPG3
[DeviceA-acl-ipv4-adv-EPG1-EPG3] rule 0 permit ip vpn-instance l3vpna source
microsegment 1 destination microsegment 3
[Device-acl-ipv4-adv-EPG1-EPG3] quit
```

Create an IPv4 advanced ACL named **EPG1-EPG2**, configure a rule to match the IP packets of VPN instance **l3vpna** from microsegment 1 to microsegment 2.

```
[DeviceA] acl advanced name EPG1-EPG2
[DeviceA-acl-ipv4-adv-EPG1-EPG2] rule 0 permit ip vpn-instance l3vpna source
microsegment 1 destination microsegment 2
[DeviceA-acl-ipv4-adv-EPG1-EPG2] quit
```

Create an IPv4 advanced ACL named **EPG2-EPG3**, configure a rule to match the IP packets of VPN instance **l3vpna** from microsegment 2 to microsegment 3.

```
[DeviceA] acl advanced name EPG2-EPG3
[DeviceA-acl-ipv4-adv-EPG2-EPG3] rule 0 permit ip vpn-instance l3vpna source
microsegment 2 destination microsegment 3
[DeviceA-acl-ipv4-adv-EPG2-EPG3] quit
```

Create a traffic class named **CLASSIFIER-GBP13**, and use ACL **EPG1-EPG3** as the match criterion.

```
[DeviceA] traffic classifier CLASSIFIER-GBP13
[DeviceA-classifier-CLASSIFIER-GBP13] if-match acl name EPG1-EPG3
[DeviceA-classifier-CLASSIFIER-GBP13] quit
```

Create a traffic behavior named **BEHAVIOR-GBP13**, configure a deny action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceA] traffic behavior BEHAVIOR-GBP13
```

```
[DeviceA-behavior-BEHAVIOR-GBP13] filter deny
[DeviceA-behavior-BEHAVIOR-GBP13] accounting packet
[DeviceA-behavior-BEHAVIOR-GBP13] quit
```

Create a traffic class named **CLASSIFIER-GBP12**, and use ACL **EPG1-EPG2** as the match criterion.

```
[DeviceA] traffic classifier CLASSIFIER-GBP12
[DeviceA-classifier-CLASSIFIER-GBP12] if-match acl name EPG1-EPG2
[DeviceA-classifier-CLASSIFIER-GBP12] quit
```

Create a traffic behavior named **BEHAVIOR-GBP12**, configure a permit action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceA] traffic behavior BEHAVIOR-GBP12
[DeviceA-behavior-BEHAVIOR-GBP12] filter permit
[DeviceA-behavior-BEHAVIOR-GBP12] accounting packet
[DeviceA-behavior-BEHAVIOR-GBP12] quit
```

Create a traffic class named **CLASSIFIER-GBP23**, and use ACL **EPG2-EPG3** as the match criterion.

```
[DeviceA] traffic classifier CLASSIFIER-GBP23
[DeviceA-classifier-CLASSIFIER-GBP23] if-match acl name EPG2-EPG3
[DeviceA-classifier-CLASSIFIER-GBP23] quit
```

Create a traffic behavior named **BEHAVIOR-GBP23**, configure a permit action, and configure a traffic accounting action to count traffic in packets.

```
[DeviceA] traffic behavior BEHAVIOR-GBP23
[DeviceA-behavior-BEHAVIOR-GBP23] filter permit
[DeviceA-behavior-BEHAVIOR-GBP23] accounting packet
[DeviceA-behavior-BEHAVIOR-GBP23] quit
```

Create a QoS policy named **GBP1**, and associate traffic class **CLASSIFIER-GBP13** with traffic behavior **BEHAVIOR-GBP13** and **CLASSIFIER-GBP12** with **BEHAVIOR-GBP12** in the QoS policy.

```
[DeviceA] qos policy GBP1
[DeviceA-qospolicy-GBP1] classifier CLASSIFIER-GBP13 behavior BEHAVIOR-GBP13
[DeviceA-qospolicy-GBP1] classifier CLASSIFIER-GBP12 behavior BEHAVIOR-GBP12
[DeviceA-qospolicy-GBP1] quit
```

Create a QoS policy **GBP2**, and associate traffic class **CLASSIFIER-GBP23** with traffic behavior **BEHAVIOR-GBP23** in the QoS policy.

```
[DeviceA] qos policy GBP2
[DeviceA-qospolicy-GBP2] classifier CLASSIFIER-GBP23 behavior BEHAVIOR-GBP23
[DeviceA-qospolicy-GBP2] quit
```

Apply QoS policy **GBP1** to the inbound direction of GigabitEthernet1/0/2.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] qos apply policy GBP1 inbound
[DeviceA-GigabitEthernet1/0/2] quit
```

Apply QoS policy **GBP2** to the inbound direction of GigabitEthernet1/0/3.

```
[DeviceA] interface gigabitethernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] qos apply policy GBP2 inbound
[DeviceA-GigabitEthernet1/0/3] quit
```

2. Configure Device B:

Create VLANs and configure IP address for VLAN interfaces.

```
<Sysname> system-view
[Sysname] sysname DeviceB
```

```
[DeviceB] vlan 12
[DeviceB-vlan12] quit
[DeviceB] interface vlan-interface 12
[DeviceB-Vlan-interface12] ip address 12.0.0.1 30
[DeviceB-Vlan-interface12] undo shutdown
[DeviceB-Vlan-interface12] quit
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] undo shutdown
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 12
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] vlan 23
[DeviceB-vlan23] quit
[DeviceB] interface vlan-interface 23
[DeviceB-Vlan-interface23] ip address 23.0.0.2 30
[DeviceB-Vlan-interface23] undo shutdown
[DeviceB-Vlan-interface23] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] undo shutdown
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 23
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure IS-IS to implement Layer 3 interconnectivity.

```
[DeviceB] isis 1
[DeviceB-isis-1] cost-style wide
[DeviceB-isis-1] network-entity 00.0000.0000.0002.00
[DeviceB-isis-1] is-level level-2
[DeviceB-isis-1] address-family ipv4 unicast
[DeviceB-isis-1-ipv4] quit
[DeviceB-isis-1] quit
[DeviceB] interface vlan-interface 12
[DeviceB-Vlan-interface12] isis enable 1
[DeviceB-Vlan-interface12] quit
[DeviceB] interface vlan-interface 23
[DeviceB-Vlan-interface23] isis enable 1
[DeviceB-Vlan-interface23] quit
[DeviceB] interface LoopBack0
[DeviceB-LoopBack0] ip address 2.2.2.2 32
[DeviceB-LoopBack0] isis enable 1
[DeviceB-LoopBack0] quit
```

3. Configure Device C:

Create VLANs and configure IP address for VLAN interfaces.

```
<Sysname> system-view
[Sysname] sysname DeviceC
[DeviceC] vlan 30
[DeviceC-vlan30] port gigabitethernet 1/0/2
[DeviceC-vlan30] quit
[DeviceC] interface gigabitethernet 1/0/2
```

```
[DeviceC-GigabitEthernet1/0/2] undo shutdown
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] vlan 23
[DeviceC-vlan23] quit
[DeviceC] interface vlan-interface 23
[DeviceC-Vlan-interface23] ip address 23.0.0.1 30
[DeviceC-Vlan-interface23] undo shutdown
[DeviceC-Vlan-interface23] quit
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo shutdown
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 23
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure IS-IS to implement Layer 3 interconnectivity.

```
[DeviceC] isis 1
[DeviceC-isis-1] cost-style wide
[DeviceC-isis-1] network-entity 00.0000.0000.0003.00
[DeviceC-isis-1] is-level level-2
[DeviceC-isis-1] address-family ipv4 unicast
[DeviceC-isis-1-ipv4] quit
[DeviceC-isis-1] quit
[DeviceC] interface vlan-interface 23
[DeviceC-Vlan-interface23] isis enable 1
[DeviceC-Vlan-interface23] quit
[DeviceC] interface LoopBack0
[DeviceC-LoopBack0] ip address 3.3.3.3 32
[DeviceC-LoopBack0] isis enable 1
[DeviceC-LoopBack0] quit
```

Enable L2VPN.

```
[DeviceC] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[DeviceC] vxlan tunnel mac-learning disable
[DeviceC] vxlan tunnel arp-learning disable
```

Create an EVPN instance in VSI **vpna**, and configure the switch to automatically generate a route distinguisher (RD) and a route target (RT) for the EVPN instance.

```
[DeviceC] vsi vpna
[DeviceC-vsi-vpna] evpn encapsulation vxlan
[DeviceC-vsi-vpna-evpn-vxlan] route-distinguisher auto
[DeviceC-vsi-vpna-evpn-vxlan] vpn-target auto
[DeviceC-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[DeviceC-vsi-vpna] vxlan 10
[DeviceC-vsi-vpna-vxlan-10] quit
[DeviceC-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the EVPN instance to automatically generate an RD and RT.

```
[DeviceC] vsi vpnb
[DeviceC-vsi-vpnb] evpn encapsulation vxlan
```

```
[DeviceC-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[DeviceC-vsi-vpnb-evpn-vxlan] vpn-target auto
[DeviceC-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[DeviceC-vsi-vpnb] vxlan 20
[DeviceC-vsi-vpnb-vxlan-20] quit
[DeviceC-vsi-vpnb] quit
```

Create an EVPN instance in VSI **vpnc**, and configure the device to automatically generate an RD and a route target for the EVPN instance.

```
[DeviceC] vsi vpnc
[DeviceC-vsi-vpnc] evpn encapsulation vxlan
[DeviceC-vsi-vpnc-evpn-vxlan] route-distinguisher auto
[DeviceC-vsi-vpnc-evpn-vxlan] vpn-target auto
[DeviceC-vsi-vpnc-evpn-vxlan] quit
```

Create VXLAN 30.

```
[DeviceC-vsi-vpnc] vxlan 30
[DeviceC-vsi-vpnc-vxlan-30] quit
[DeviceC-vsi-vpnc] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[DeviceC] bgp 100
[DeviceC-bgp-default] peer 1.1.1.1 as-number 100
[DeviceC-bgp-default] peer 1.1.1.1 connect-interface loopback 0
[DeviceC-bgp-default] address-family l2vpn evpn
[DeviceC-bgp-default-evpn] peer 1.1.1.1 enable
[DeviceC-bgp-default-evpn] quit
[DeviceC-bgp-default] quit
```

On GigabitEthernet1/0/2 connected to the general server, create Ethernet service instance 1000 and configure it to match data frames from VLAN 30.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] service-instance 1000
[DeviceC-GigabitEthernet1/0/2-srv1000] encapsulation s-vid 30
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[DeviceC-GigabitEthernet1/0/2-srv1000] xconnect vsi vpna
[DeviceC-GigabitEthernet1/0/2-srv1000] quit
```

Configure the RD and RT settings for VPN instance **l3vpna**.

```
[DeviceC] ip vpn-instance l3vpna
[DeviceC-vpn-instance-l3vpna] route-distinguisher 1:1
[DeviceC-vpn-instance-l3vpna] address-family ipv4
[DeviceC-vpn-ipv4-l3vpna] vpn-target 2:2
[DeviceC-vpn-ipv4-l3vpna] quit
[DeviceC-vpn-instance-l3vpna] address-family evpn
[DeviceC-vpn-evpn-l3vpna] vpn-target 1:1
[DeviceC-vpn-evpn-l3vpna] quit
[DeviceC-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[DeviceC] interface vsi-interface 1
[DeviceC-Vsi-interface1] ip binding vpn-instance l3vpna
[DeviceC-Vsi-interface1] ip address 192.168.1.254 24
```

```
[DeviceC-Vsi-interface1] mac-address 1-1-1
[DeviceC-Vsi-interface1] distributed-gateway local
[DeviceC-Vsi-interface1] local-proxy-arp enable
[DeviceC-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[DeviceC] interface vsi-interface 2
[DeviceC-Vsi-interface2] ip binding vpn-instance l3vpna
[DeviceC-Vsi-interface2] ip address 192.168.2.254 24
[DeviceC-Vsi-interface2] mac-address 2-2-2
[DeviceC-Vsi-interface2] distributed-gateway local
[DeviceC-Vsi-interface2] local-proxy-arp enable
[DeviceC-Vsi-interface2] quit
```

Configure VSI-interface 3.

```
[DeviceC] interface vsi-interface 3
[DeviceC-Vsi-interface1] ip binding vpn-instance l3vpna
[DeviceC-Vsi-interface1] ip address 192.168.3.254 24
[DeviceC-Vsi-interface1] mac-address 3-3-3
[DeviceC-Vsi-interface1] distributed-gateway local
[DeviceC-Vsi-interface1] local-proxy-arp enable
[DeviceC-Vsi-interface1] quit
```

Create VSI-interface 4, associate it with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[DeviceC] interface vsi-interface 4
[DeviceC-Vsi-interface4] ip binding vpn-instance l3vpna
[DeviceC-Vsi-interface4] l3-vni 1000
[DeviceC-Vsi-interface4] quit
```

Specify VSI-interface 1 as the gateway interface for the VSI of VXLAN 10.

```
[DeviceC] vsi vpna
[DeviceC-vsi-vpna] gateway vsi-interface 1
[DeviceC-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for the VSI of VXLAN 20.

```
[DeviceC] vsi vpnb
[DeviceC-vsi-vpnb] gateway vsi-interface 2
[DeviceC-vsi-vpnb] quit
```

Specify VSI-interface 3 as the gateway interface for the VSI of VXLAN 30.

```
[DeviceC] vsi vpnc
[DeviceC-vsi-vpnc] gateway vsi-interface 3
[DeviceC-vsi-vpnc] quit
```

Create microsegment 3, and add the IP address of the general server as its member.

```
[DeviceC] microsegment 3 name EPG3
[DeviceC-microsegment-3] member ipv4 192.168.3.1 24 vpn-instance l3vpna
[DeviceC-microsegment-3] quit
```

Verifying the configuration

Execute the **display bgp l2vpn evpn** command on Device C to view detailed EVPN route information. You can see that the MAC/IP advertisement route 192.168.1.1 carries microsegment information in the extended community attribute (Ext-Community), which indicates that the EVPN route is learned correctly.

```
[DeviceC] display bgp l2vpn evpn [2][0][48][6e7f-a8c5-0407][32][192.168.1.1]/136
```

```
 BGP local router ID: 3.3.3.3
 Local AS number: 100


 Route distinguisher: 1:2(l3vpna)
 Total number of routes: 1
 Paths:   1 available, 1 best

 BGP routing table information of [2][0][48][6e7f-a8c5-0407][32][192.168.1.1]/136:
 From            : 1.1.1.1  (1.1.1.1)
 Rely nexthop    : 23.0.0.2
 Original nexthop: 1.1.1.1
 Route age       : 00h02m32s
 OutLabel        : NULL
 Ext-Community   : <RT: 1:1>, <RT: 100:10>, <Encapsulation Type: VXLAN>, <Route
                   r's MAC: 6e7f-9995-0100>, <MicroSegment-id: Type 0x83ff, ID
                   1>
 RxPathID        : 0x0
 TxPathID        : 0x0
 AS-path         : (null)
 Origin          : igp
 Attribute value : MED 0, localpref 100, pref-val 0
 State           : valid, internal, best, remoteredist
 IP precedence   : N/A
 QoS local ID    : N/A
 Traffic index   : N/A
 EVPN route type : MAC/IP advertisement route
 ESI             : 0000.0000.0000.0000.0000
 Ethernet tag ID : 0
 MAC address     : 6e7f-a8c5-0407
 IP address      : 192.168.1.1/32
 MPLS label1     : 10
 MPLS label2     : 1000
 Tunnel policy   : NULL
 Rely tunnel IDs : N/A
…
```

Ping the general server from Host A and Host B. The output shows that Host A cannot access the general server.

```
C:\> ping 192.168.3.1

Pinging 192.168.3.1 with 32 bytes of data:

Request timed out
Request timed out
Request timed out
Request timed out
```

```
Ping statistics for 192.168.3.1:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),
```

Host B can ping the general server.

Display the configuration of each microsegment on Device A.

```
[DeviceA] display microsegment 1
Microsegment ID    : 1
Microsegment name  : EPG1
  IPv4 member:
    192.168.1.1/24
[DeviceA] display microsegment 2
Microsegment ID    : 2
Microsegment name  : EPG2
  IPv4 member:
    192.168.2.1/24
[DeviceA] display microsegment 3
Microsegment ID    : 3
Microsegment name  : EPG3
  IPv4 member:
    192.168.2.1/24
```

Display summary and status information about microsegments on Device A.

```
[DeviceA] display microsegment
Microsegment status: Enabled
Total microsegments: 3
Microsegment list  :
  Microsegment ID  Members  Microsegment name
  1                1        EPG1
  2                1        EPG2
  3                1        EPG3
```

Display the configuration and operating status of QoS policies on the interface of Device A.

```
[DeviceA] display qos policy interface GigabitEthernet1/0/2
Interface: GigabitEthernet1/0/2
  Direction: Inbound
  Policy: GBP1
   Classifier: CLASSIFIER-GBP13
     Rule(s) :
      If-match acl name EPG1-EPG3
     Behavior: BEHAVIOR-GBP13
      Accounting enable:
        5 (Packets)
        0 (pps)
      Filter enable: Deny
   Classifier: CLASSIFIER-GBP12
     Rule(s) :
      If-match acl name EPG1-EPG2
     Behavior: BEHAVIOR-GBP12
      Accounting enable:
```

```
    5 (Packets)
    0 (pps)
  Filter enable: Permit
```

Configuration files

- Device A:

```
#
 sysname DeviceA
#
ip vpn-instance l3vpna
 route-distinguisher 1:1
 #
 address-family ipv4
  vpn-target 2:2 import-extcommunity
  vpn-target 2:2 export-extcommunity
 #
 address-family evpn
  vpn-target 1:1 import-extcommunity
  vpn-target 1:1 export-extcommunity
#
 vxlan tunnel mac-learning disable
#
isis 1
 is-level level-2
 cost-style wide
 network-entity 00.0000.0000.0001.00
 #
 address-family ipv4 unicast
#
vlan 10
#
vlan 12
#
vlan 20
#
 microsegment enable
#
microsegment 1 name EPG1
 member ipv4 192.168.1.0 255.255.255.0 vpn-instance l3vpna
#
microsegment 2 name EPG2
 member ipv4 192.168.2.0 255.255.255.0 vpn-instance l3vpna
#
traffic classifier CLASSIFIER-GBP12 operator and
 if-match acl name EPG1-EPG2
#
traffic classifier CLASSIFIER-GBP13 operator and
 if-match acl name EPG1-EPG3
```

```
#
traffic classifier CLASSIFIER-GBP23 operator and
 if-match acl name EPG2-EPG3
#
traffic behavior BEHAVIOR-GBP12
 accounting packet
 filter permit
#
traffic behavior BEHAVIOR-GBP13
 accounting packet
 filter deny
#
traffic behavior BEHAVIOR-GBP23
 accounting packet
 filter permit
#
qos policy GBP1
 classifier CLASSIFIER-GBP13 behavior BEHAVIOR-GBP13
 classifier CLASSIFIER-GBP12 behavior BEHAVIOR-GBP12
#
qos policy GBP2
 classifier CLASSIFIER-GBP23 behavior BEHAVIOR-GBP23
#
 l2vpn enable
 vxlan tunnel arp-learning disable
#
vsi vpna
 gateway Vsi-interface 1
 vxlan 10
 evpn encapsulation vxlan
  route-distinguisher auto
  vpn-target auto export-extcommunity
  vpn-target auto import-extcommunity
#
vsi vpnb
 gateway Vsi-interface 2
 vxlan 20
 evpn encapsulation vxlan
  route-distinguisher auto
  vpn-target auto export-extcommunity
  vpn-target auto import-extcommunity
#
vsi vpnc
 gateway Vsi-interface 3
 vxlan 30
 evpn encapsulation vxlan
  route-distinguisher auto
  vpn-target auto export-extcommunity
```

```
  vpn-target auto import-extcommunity
#
interface LoopBack0
 ip address 1.1.1.1 255.255.255.255
 isis enable 1
#
interface Vlan-interface12
 ip address 12.0.0.2 255.255.255.252
 isis enable 1
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 12
#
interface GigabitEthernet1/0/2
 port access vlan 10
 qos apply policy GBP1 inbound
 #
 service-instance 1000
  encapsulation s-vid 10
  xconnect vsi vpna
#
interface GigabitEthernet1/0/3
 port access vlan 20
 qos apply policy GBP2 inbound
 #
 service-instance 2000
  encapsulation s-vid 20
  xconnect vsi vpnb
#
#
interface Vsi-interface1
 ip binding vpn-instance l3vpna
 ip address 192.168.1.254 255.255.255.0
 mac-address 0001-0001-0001
 local-proxy-arp enable
 distributed-gateway local
#
interface Vsi-interface2
 ip binding vpn-instance l3vpna
 ip address 192.168.2.254 255.255.255.0
 mac-address 0002-0002-0002
 local-proxy-arp enable
 distributed-gateway local
#
interface Vsi-interface3
 ip binding vpn-instance l3vpna
 ip address 192.168.3.254 255.255.255.0
```

```
 mac-address 0003-0003-0003
 local-proxy-arp enable
 distributed-gateway local
#
interface Vsi-interface4
 ip binding vpn-instance l3vpna
 l3-vni 1000
#
bgp 100
 router-id 1.1.1.1
 peer 3.3.3.3 as-number 100
 peer 3.3.3.3 connect-interface LoopBack0
 #
 address-family l2vpn evpn
  peer 3.3.3.3 enable
#
acl advanced name EPG1-EPG2
 rule 0 permit ip vpn-instance l3vpna source microsegment 1 destination microsegment
2
#
acl advanced name EPG1-EPG3
 rule 0 permit ip vpn-instance l3vpna source microsegment 1 destination microsegment
3
#
acl advanced name EPG2-EPG3
 rule 0 permit ip vpn-instance l3vpna source microsegment 2 destination microsegment
3
#
```

- Device B:

```
#
 sysname DeviceB
#
isis 1
 is-level level-2
 cost-style wide
 network-entity 00.0000.0000.0002.00
 #
 address-family ipv4 unicast
#
vlan 12
#
vlan 23
#
interface LoopBack0
 ip address 2.2.2.2 255.255.255.255
 isis enable 1
#
interface Vlan-interface12
 ip address 12.0.0.1 255.255.255.252
```

```
 isis enable 1
#
interface Vlan-interface23
 ip address 23.0.0.2 255.255.255.252
 isis enable 1
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 12
#
interface GigabitEthernet1/0/2
 port link-type trunk
 port trunk permit vlan 1 23
#
```

- Device C:

```
#
 sysname DeviceC
#
ip vpn-instance l3vpna
 route-distinguisher 1:2
 #
 address-family ipv4
  vpn-target 2:2 import-extcommunity
  vpn-target 2:2 export-extcommunity
 #
 address-family evpn
  vpn-target 1:1 import-extcommunity
  vpn-target 1:1 export-extcommunity
#
 vxlan tunnel mac-learning disable
#
isis 1
 is-level level-2
 cost-style wide
 network-entity 00.0000.0000.0003.00
 #
 address-family ipv4 unicast
vlan 23
#
vlan 30
#
vlan 40
#
 microsegment enable
#
microsegment 3 name EPG3
 member ipv4 192.168.3.0 255.255.255.0 vpn-instance l3vpna
#
```

```
 l2vpn enable
 vxlan tunnel arp-learning disable
#
vsi vpna
 gateway Vsi-interface 1
 vxlan 10
 evpn encapsulation vxlan
  route-distinguisher auto
  vpn-target auto export-extcommunity
  vpn-target auto import-extcommunity
#
vsi vpnb
 gateway Vsi-interface 2
 vxlan 20
 evpn encapsulation vxlan
  route-distinguisher auto
  vpn-target auto export-extcommunity
  vpn-target auto import-extcommunity
#
vsi vpnc
 gateway Vsi-interface 3
 vxlan 30
 evpn encapsulation vxlan
  route-distinguisher auto
  vpn-target auto export-extcommunity
  vpn-target auto import-extcommunity
#
interface LoopBack0
 ip address 3.3.3.3 255.255.255.255
 isis enable 1
#
interface Vlan-interface23
 ip address 23.0.0.1 255.255.255.252
 isis enable 1
#
interface GigabitEthernet1/0/1
 port link-type trunk
 port trunk permit vlan 1 23
#
interface GigabitEthernet1/0/2
 port access vlan 30
 #
 service-instance 1000
  encapsulation s-vid 30
  xconnect vsi vpnc
#
interface Vsi-interface1
 ip binding vpn-instance l3vpna
```

```
 ip address 192.168.1.254 255.255.255.0
 mac-address 0001-0001-0001
 local-proxy-arp enable
 distributed-gateway local
#
interface Vsi-interface2
 ip binding vpn-instance l3vpna
 ip address 192.168.2.254 255.255.255.0
 mac-address 0002-0002-0002
 local-proxy-arp enable
 distributed-gateway local
#
interface Vsi-interface3
 ip binding vpn-instance l3vpna
 ip address 192.168.3.254 255.255.255.0
 mac-address 0003-0003-0003
 local-proxy-arp enable
 distributed-gateway local
#
interface Vsi-interface4
 ip binding vpn-instance l3vpna
 l3-vni 1000
#
bgp 100
 router-id 3.3.3.3
 peer 1.1.1.1 as-number 100
 peer 1.1.1.1 connect-interface LoopBack0
 #
 address-family l2vpn evpn
  peer 1.1.1.1 enable
#
```

Contents

Configuring SAVI 1

About SAVI 1
SAVI application scenarios 1
SAVI tasks at a glance 1
Enabling SAVI 1
Configuring IPv6 source guard 2
Configuring DHCPv6 snooping 2
Configuring ND parameters 2
Setting the entry deletion delay 2
Enabling packet spoofing logging and filtering entry logging 3
SAVI configuration examples 3
Example: Configuring DHCPv6-only SAVI 3
Example: Configuring SLAAC-only SAVI 5
Example: Configuring DHCPv6+SLAAC SAVI 6

Configuring SAVI

About SAVI

Source Address Validation Improvement (SAVI) checks the validity of the source addresses of global unicast IPv6 packets. It implements the validity check by using the ND snooping, DHCPv6 snooping, ND attack detection, and IP source guard features. SAVI checks only global unicast addresses and forwards the packets that pass the validity check. Packets sourced from an invalid address are dropped.

SAVI application scenarios

DHCPv6-only

The hosts connected to the SAVI-enabled device obtain addresses only through DHCPv6. DHCPv6 messages, ND messages (RA and RR messages excluded), and IPv6 data packets are checked based on DHCPv6 snooping entries and static IPv6 source guard binding entries.

SLAAC-only

The hosts connected to the SAVI-enabled device obtain addresses only through Stateless Address Autoconfiguration (SLAAC). In this scenario, SAVI drops all DHCPv6 messages. Only ND messages and IPv6 data packets are checked based on DHCPv6 snooping entries and static IPv6 source guard binding entries.

DHCPv6+SLAAC

The hosts connected to the SAVI-enabled device obtain addresses through DHCPv6 and SLAAC. In this scenario, SAVI checks all DHCPv6 messages, ND messages, and IPv6 data packets based on DHCPv6 snooping entries, ND snooping entries, and static IPv6 source guard binding entries.

SAVI tasks at a glance

To configure SAVI, perform the following tasks:

1. Enabling SAVI
2. Configuring IPv6 source guard
3. Configuring DHCPv6 snooping
4. Configuring ND parameters
5. (Optional.) Setting the entry deletion delay
6. (Optional.) Enabling packet spoofing logging and filtering entry logging

Enabling SAVI

1. Enter system view.
   `system-view`
2. Enable SAVI.
   `ipv6 savi strict`
   By default, SAVI is disabled.

Configuring IPv6 source guard

1. Enable IPv6 source guard on an interface.
2. (Optional.) Configure static IPv6SG bindings.

For more information about IPv6 source guard configuration, see "Configuring IP source guard."

Configuring DHCPv6 snooping

Restrictions and guidelines

Enable only DHCPv6 snooping for the SLAAC-only scenario.

Procedure

1. Enable DHCPv6 snooping.
2. Specify DHCPv6 snooping trusted ports.
3. Enable recording client information in DHCPv6 snooping entries.

For more information about DHCPv6 snooping configuration, see *Layer 3—IP Services Configuration Guide*.

Configuring ND parameters

Restrictions and guidelines

Enable only ND attack detection for the DHCPv6-only scenario.

Procedure

1. Enable ND snooping for global unicast addresses.

   For more information about ND snooping, see IPv6 basics in *Layer 3—IP Services Configuration Guide*.
2. Enable ND attack detection.

   For more information about ND attack detection, see "Configuring ND attack defense."
3. Specify ND trusted ports.

   For more information about ND trusted ports, see "Configuring ND attack defense."

Setting the entry deletion delay

About this task

The entry deletion delay is the period of time that the device waits before deleting the DHCPv6 snooping entries and ND snooping entries for a down port.

Procedure

1. Enter system view.

   system-view
2. Set the entry deletion delay.

   ipv6 savi down-delay *delay-time*

   By default, the entry deletion delay is 30 seconds.

Enabling packet spoofing logging and filtering entry logging

About this task

Packet spoofing logging enables the device to generate log messages for the spoofed packets detected by SAVI.

Filtering entries are effective bindings used for filtering IPv6 packets by the source IPv6 address. Filtering entry logging enables the device to generate log messages for filtering entries. A log message contains the IPv6 address, MAC address, VLAN, and interface of a filtering entry.

The device sends packet spoofing and filtering entry log messages to the information center. With the information center, you can set log message filtering and output rules, including output destinations. For more information about using the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable packet spoofing logging.

   ipv6 savi log enable spoofing-packet [**interval** *interval* | **total-number** *number*] *

   By default, packet spoofing logging is disabled.

3. Enable filtering entry logging.

   ipv6 savi log enable filter-entry

   By default, filtering entry logging is disabled.

SAVI configuration examples

Example: Configuring DHCPv6-only SAVI

Network configuration

As shown in Figure 1, configure SAVI on the switch to meet the following requirements:

- Clients obtain IPv6 addresses only through DHCPv6.
- SAVI checks the source addresses of DHCPv6 messages, ND messages (RA and RR messages excluded), and IPv6 data packets on GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3.

Figure 1 Network diagram

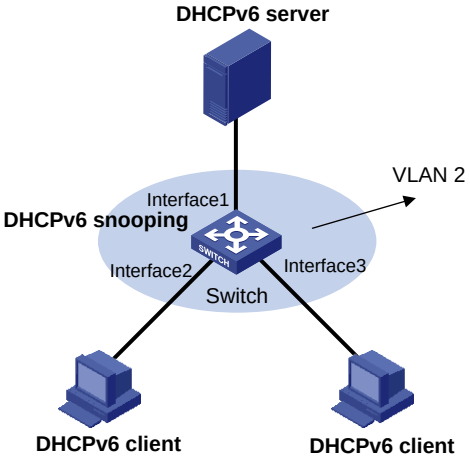


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Enable SAVI.

```
<Switch> system-view
[Switch] ipv6 savi strict
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to VLAN 2.

```
[Switch] vlan 2
[Switch-vlan2] port gigabitethernet 1/0/1 gigabitethernet 1/0/2 gigabitethernet 1/0/3
[Switch-vlan2] quit
```

Enable DHCPv6 snooping.

```
[Switch] ipv6 dhcp snooping enable
```

Configure GigabitEthernet 1/0/1 as a DHCPv6 snooping trusted port.

```
[Switch] interface gigabitethernet 1/0/1
[Switch-GigabitEthernet1/0/1] ipv6 dhcp snooping trust
[Switch-GigabitEthernet1/0/1] quit
```

Enable recording DHCPv6 snooping entries on GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3.

```
[Switch] interface gigabitethernet 1/0/2
[Switch-GigabitEthernet1/0/2] ipv6 dhcp snooping binding record
[Switch-GigabitEthernet1/0/2] quit
[Switch] interface gigabitethernet 1/0/3
[Switch-GigabitEthernet1/0/3] ipv6 dhcp snooping binding record
[Switch-GigabitEthernet1/0/3] quit
```

Enable ND attack detection.

```
[Switch] vlan 2
[Switch-vlan2] ipv6 nd detection enable
```

```
[Switch-vlan2] quit
```

Enable IPv6 source guard on GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3.

```
[Switch] interface gigabitethernet 1/0/2
[Switch-GigabitEthernet1/0/2] ipv6 verify source ip-address mac-address
[Switch-GigabitEthernet1/0/2] quit
[Switch] interface gigabitethernet 1/0/3
[Switch-GigabitEthernet1/0/3] ipv6 verify source ip-address mac-address
[Switch-GigabitEthernet1/0/3] quit
```

Example: Configuring SLAAC-only SAVI

Network configuration

As shown in Figure 2, configure SAVI on Switch B to meet the following requirements:

- Hosts obtain IPv6 addresses only through SLAAC.
- DHCPv6 messages are dropped on GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 in VLAN 2.
- SAVI checks the source addresses of ND messages and IPv6 data packets on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

Figure 2 Network diagram

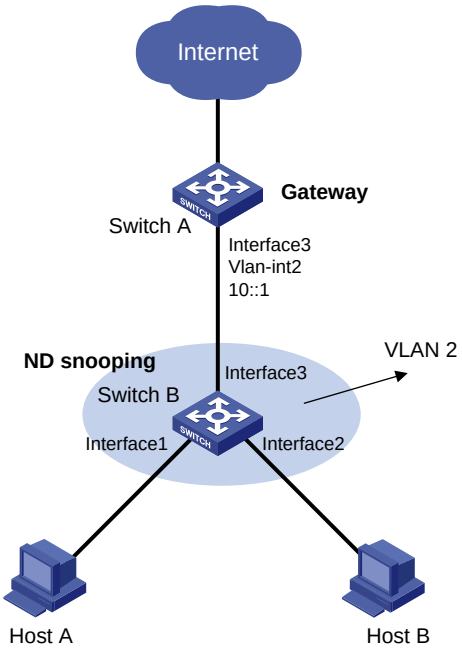


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

Enable SAVI.

```
<SwitchB> system-view
[SwitchB] ipv6 savi strict
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to VLAN 2.

```
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1 gigabitethernet 1/0/2 gigabitethernet 1/0/3
[SwitchB-vlan2] quit
```

Enable ND snooping for global unicast addresses in VLAN 2.

```
[SwitchB] vlan 2
[SwitchB-vlan2] ipv6 nd snooping enable global
```

Enable ND attack detection for VLAN 2.

```
[SwitchB-vlan2] ipv6 nd detection enable
[SwitchB-vlan2] quit
```

Enable DHCPv6 snooping.

```
[SwitchB] ipv6 dhcp snooping enable
```

Configure GigabitEthernet 1/0/3 as an ND trusted port.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] ipv6 nd detection trust
[SwitchB-GigabitEthernet1/0/3] quit
```

Enable IPv6 source guard on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] ipv6 verify source ip-address mac-address
[SwitchB-GigabitEthernet1/0/1] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] ipv6 verify source ip-address mac-address
[SwitchB-GigabitEthernet1/0/2] quit
```

Example: Configuring DHCPv6+SLAAC SAVI

Network configuration

As shown in Figure 3, configure SAVI on Switch B to meet the following requirements:

- Hosts obtain IP addresses through DHCPv6 or SLAAC.
- SAVI checks the source addresses of DHCPv6 messages, ND messages, and IPv6 data packets on GigabitEthernet 1/0/3 through GigabitEthernet 1/0/5.

Figure 3 Network diagram

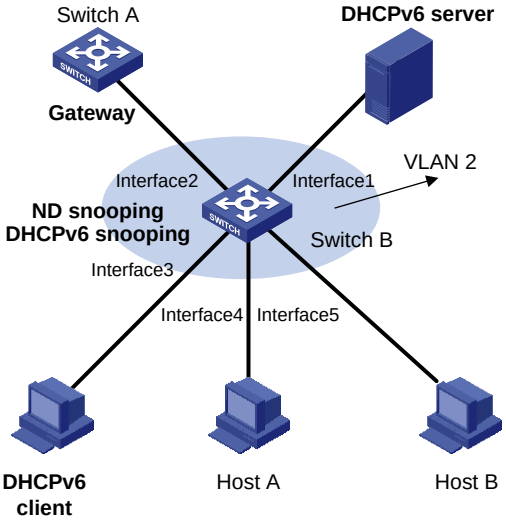


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

Enable SAVI.

```
<SwitchB> system-view
[SwitchB] ipv6 savi strict
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/5 to VLAN 2.

```
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1 gigabitethernet 1/0/2 gigabitethernet 1/0/3
gigabitethernet 1/0/4 gigabitethernet 1/0/5
```

Enable DHCPv6 snooping.

```
[SwitchB] ipv6 dhcp snooping enable
```

Enable recording DHCPv6 snooping entries on GigabitEthernet 1/0/3 through GigabitEthernet 1/0/5.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] ipv6 dhcp snooping binding record
[SwitchB-GigabitEthernet1/0/3] quit
[SwitchB] interface gigabitethernet 1/0/4
[SwitchB-GigabitEthernet1/0/4] ipv6 dhcp snooping binding record
[SwitchB-GigabitEthernet1/0/4] quit
[SwitchB] interface gigabitethernet 1/0/5
[SwitchB-GigabitEthernet1/0/5] ipv6 dhcp snooping binding record
[SwitchB-GigabitEthernet1/0/5] quit
```

Configure GigabitEthernet 1/0/1 as a DHCPv6 snooping trusted port.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] ipv6 dhcp snooping trust
[SwitchB-GigabitEthernet1/0/1] quit
```

Enable ND snooping for global unicast addresses in VLAN 2.

```
[SwitchB] vlan 2
[SwitchB-vlan2] ipv6 nd snooping enable global
```

Enable ND attack detection for VLAN 2.

```
[SwitchB-vlan2] ipv6 nd detection enable
[SwitchB-vlan2] quit
```

Configure GigabitEthernet 1/0/2 as an ND trusted port.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] ipv6 nd detection trust
[SwitchB-GigabitEthernet1/0/2] quit
```

Enable IPv6 source guard on GigabitEthernet 1/0/3 through GigabitEthernet 1/0/5.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] ipv6 verify source ip-address mac-address
[SwitchB-GigabitEthernet1/0/3] quit
[SwitchB] interface gigabitethernet 1/0/4
[SwitchB-GigabitEthernet1/0/4] ipv6 verify source ip-address mac-address
[SwitchB-GigabitEthernet1/0/4] quit
[SwitchB] interface gigabitethernet 1/0/5
[SwitchB-GigabitEthernet1/0/5] ipv6 verify source ip-address mac-address
```

Contents

Configuring object groups 1
About object groups 1
Configuring an IPv4 address object group 1
Configuring an IPv6 address object group 1
Configuring a port object group 2
Configuring aging of DNS-resolved IP addresses from host names 2
Display and maintenance commands for object groups 2

Configuring object groups

About object groups

An object group is a group of objects that can be used by an ACL or object group to identify packets. Object groups are divided into the following types:

- **IPv4 address object group**—A group of IPv4 address objects used to match the IPv4 address in a packet.
- **IPv6 address object group**—A group of IPv6 address objects used to match the IPv6 address in a packet.
- **Port object group**—A group of port objects used to match the protocol port number in a packet.

Configuring an IPv4 address object group

1. Enter system view.

   system-view
2. Configure an IPv4 address object group and enter its view.

   object-group ip address *object-group-name*

   The system has one default IPv4 address object group named **any**.
3. (Optional.) Configure a description for the IPv4 address object group.

   description *text*

   By default, an object group does not have a description.
4. Configure an IPv4 address object.

   [*object-id*] **network** { **host** { **address** *ip-address* | **name** *host-name* [**vpn-instance** *vpn-instance-name*] } | **subnet** *ip-address* { *mask-length* | *mask* }| **group-object** *object-group-name* | **range** *ip-address1 ip-address2* }

Configuring an IPv6 address object group

1. Enter system view.

   system-view
2. Configure an IPv6 address object group and enter its view.

   object-group ipv6 address *object-group-name*

   The system has one default IPv6 address object group named **any**.
3. (Optional.) Configure a description for the IPv6 address object group.

   description *text*

   By default, an object group does not have a description.
4. Configure an IPv6 address object.

   [*object-id*] **network** { **host** { **address** *ipv6-address* | **name** *host-name* [**vpn-instance** *vpn-instance-name*] } | **subnet** *ipv6-address prefix-length* | **group-object** *object-group-name* | **range** *ip-address1 ip-address2* }

Configuring a port object group

1. Enter system view.
   system-view
2. Configure a port object group and enter its view.
   object-group port *object-group-name*
   The system has one default port object group named **any**.
3. (Optional.) Configure a description for the port object group.
   description *text*
   By default, an object group does not have a description.
4. Configure a port object.
   [*object-id*] **port** { { **eq** | **lt** | **gt** } *port* | **range** *port1 port2* | **group-object** *object-group-name* }

Configuring aging of DNS-resolved IP addresses from host names

About this task

In load balancing scenarios where one host name maps to several IP addresses, DNS-resolved IP address for a host name changes between these mapping addresses. Upon every change, the object group module notifies relevant policies (such as security policy) of the change, which causes policies to submit changes frequently and consumes memory. To resolve this issue, you can enable aging of DNS-resolved IP addresses from host names.

With this feature enabled, the system maintains an IP address group for each host name. If a resolved IP address is not in the group, the system adds the address to the group and notifies relevant policies of the change. If a resolved IP address is in the group, the system does not notify relevant policies.

Restrictions and guidelines

As a best practice, set the aging time to be longer than the TTL of resolution records on the DNS server.

Procedure

1. Enter system view.
   system-view
2. Enable aging of DNS-resolved IP addresses from host names.
   object-group dns-aging [**time** *aging-time*]
   By default, aging of DNS-resolved IP addresses from host names is disabled.

Display and maintenance commands for object groups

Execute **display** commands in any view.

Task	Command
Display information about object groups.	**display object-group** [{ { **ip** | **ipv6** } **address** | **port** }[**default**] [**name** *object-group-name*] | **name** *object-group-name*]
Display IPv4 or IPv6 addresses for host names.	**display object-group** { **ip** | **ipv6** } **host** { **object-group-name** *object-group-name* | **name** *host-name* [**vpn-instance** *vpn-instance-name*] } * [**slot** *slot-number*
Display information about the IP address corresponding to the kernel host name.	**display object-group kernel** { **ip** | **ipv6** } **host** { **object-group-name** *object-group-name* | **name** *host-name* [**vpn-instance** *vpn-instance-name*] } * [**slot** *slot-number*]

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

High Availability Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the high availability fundamentals and configuration procedures. The high availability technologies include fault detection and fault failover. Failure detection technologies focus on fault detection and isolation. Failover technologies focus on network recovery.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** >

	Folder.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring Ethernet OAM 1

About Ethernet OAM 1
Major functions of Ethernet OAM 1
Ethernet OAMPDUs 1
How Ethernet OAM works 1
Protocols and standards 3
Restrictions and guidelines: Ethernet OAM configuration 3
Ethernet OAM tasks at a glance 3
Configuring basic Ethernet OAM functions 4
Configuring the Ethernet OAM connection detection timers 4
About the Ethernet OAM connection detection timers 4
Restrictions and guidelines for configuring Ethernet OAM connection detection timers 4
Configuring the Ethernet OAM connection detection timers globally 5
Configuring the Ethernet OAM connection detection timers on a port 5
Configuring errored symbol event detection 5
Restrictions and guidelines for configuring errored symbol event detection 5
Configuring errored symbol event detection globally 5
Configuring errored symbol event detection on a port 6
Configuring errored frame event detection 6
Restrictions and guidelines for configuring errored frame event detection 6
Configuring errored frame event detection globally 6
Configuring errored frame event detection on a port 6
Configuring errored frame period event detection 7
Restrictions and guidelines for configuring errored frame period event detection 7
Configuring errored frame period event detection globally 7
Configuring errored frame period event detection on a port 7
Configuring errored frame seconds event detection 8
Restrictions and guidelines for configuring errored frame seconds event detection 8
Configuring errored frame seconds event detection globally 8
Configuring errored frame seconds event detection on a port 8
Configuring the action a port takes after it receives an Ethernet OAM event from the remote end 8
Display and maintenance commands for Ethernet OAM 9
Ethernet OAM configuration examples 9
Example: Configuring Ethernet OAM 9

Configuring Ethernet OAM

About Ethernet OAM

Ethernet Operation, Administration, and Maintenance (OAM) is a tool that monitors Layer 2 link status and addresses common link-related issues on the "last mile." Ethernet OAM improves Ethernet management and maintainability. You can use it to monitor the status of the point-to-point link between two directly connected devices.

Major functions of Ethernet OAM

Ethernet OAM provides the following functions:

- **Link performance monitoring**—Monitors the performance indices of a link, including packet loss, delay, and jitter, and collects traffic statistics of various types.
- **Fault detection and alarm**—Checks the connectivity of a link by sending OAM protocol data units (OAMPDUs) and reports to the network administrators when a link error occurs.
- **Remote loopback**—Checks link quality and locates link errors by looping back OAMPDUs.

Ethernet OAMPDUs

Ethernet OAM operates on the data link layer. Ethernet OAM reports the link status by periodically exchanging OAMPDUs between devices, so that the administrator can effectively manage the network.

Ethernet OAMPDUs include the following types shown in Table 1.

Table 1 Functions of different types of OAMPDUs

OAMPDU type	Function
Information OAMPDU	Used for transmitting state information of an Ethernet OAM entity, including the information about the local device and remote devices, and customized information, to the remote Ethernet OAM entity, and maintaining OAM connections.
Event Notification OAMPDU	Used by link monitoring to notify the remote OAM entity when it detects problems on the link in between.
Loopback Control OAMPDU	Used for remote loopback control. By inserting the information used to enable/disable loopback to a loopback control OAMPDU, you can enable/disable loopback on a remote OAM entity.

NOTE:

Throughout this document, an Ethernet OAM-enabled port is called an Ethernet OAM entity or an OAM entity.

How Ethernet OAM works

This section describes the working procedures of Ethernet OAM.

Ethernet OAM connection establishment

OAM connection establishment is also known as the Discovery phase, where an Ethernet OAM entity discovers the remote OAM entity to establish a session.

In this phase, two connected OAM entities exchange Information OAMPDUs to advertise their OAM configuration and capabilities to each other for a comparison. If their Loopback, link detection, and link event settings match, the OAM entities establish an OAM connection.

An OAM entity operates in active mode or passive mode. OAM entities in active mode initiate OAM connections, and OAM entities in passive mode wait and respond to the OAM connection requests. To set up an OAM connection between two OAM entities, you must set at least one entity to operate in active mode.

Table 2 shows the actions that a device can perform in different modes.

Table 2 Active Ethernet OAM mode and passive Ethernet OAM mode

Item	Active Ethernet OAM mode	Passive Ethernet OAM mode
Initiating OAM Discovery	Available	Unavailable
Responding to OAM Discovery	Available	Available
Transmitting Information OAMPDUs	Available	Available
Transmitting Event Notification OAMPDUs	Available	Available
Transmitting Information OAMPDUs without any TLV	Available	Available
Transmitting Loopback Control OAMPDUs	Available	Unavailable
Responding to Loopback Control OAMPDUs	Available	Available

After an Ethernet OAM connection is established, the Ethernet OAM entities exchange Information OAMPDUs at the handshake packet transmission interval to detect the availability of the Ethernet OAM connection. If an Ethernet OAM entity receives no Information OAMPDU within the Ethernet OAM connection timeout time, the Ethernet OAM connection is considered disconnected.

Link monitoring

Error detection in an Ethernet is difficult, especially when the physical connection in the network is not disconnected, but network performance is degrading gradually.

Link monitoring detects link faults in various environments. Ethernet OAM entities monitor link status by exchanging Event Notification OAMPDUs. When detecting one of the link error events listed in Table 3, an OAM entity sends an Event Notification OAMPDU to its peer OAM entity. The network administrator can keep track of network status changes by retrieving the log.

Table 3 Ethernet OAM link error events

Ethernet OAM link events	Description
Errored symbol event	An errored symbol event occurs when the number of detected symbol errors in the detection window (specified number of received symbols) exceeds the predefined threshold.
Errored frame event	An errored frame event occurs when the number of detected error frames in the detection window (specified detection interval) exceeds the predefined threshold.
Errored frame period event	An errored frame period event occurs when the number of frame errors in the detection window (specified number of received frames) exceeds the predefined threshold.
Errored frame seconds event	An errored frame seconds event occurs when the number of errored frame seconds (the second in which an errored frame appears is called an

Ethernet OAM link events	Description
	errored frame second) detected on a port in the detection window (specified detection interval) reaches the predefined threshold.

Remote fault detection

Information OAMPDUs are exchanged periodically among Ethernet OAM entities across established OAM connections. When traffic is interrupted due to device failure or unavailability, the Ethernet OAM entity at the faulty end sends error information to its peer. The Ethernet OAM entity uses the flag field in Information OAMPDUs to indicate the error information (any critical link event type as shown in Table 4). You can use the log information to track ongoing link status and troubleshoot problems promptly.

Table 4 Critical link events

Type	Description	OAMPDU transmission frequencies
Link Fault	Peer link signal is lost.	Once per second.
Dying Gasp	An unexpected fault, such as power failure, occurred.	Non-stop.
Critical Event	An undetermined critical event happened.	Non-stop.

Protocols and standards

IEEE 802.3ah, *Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*

Restrictions and guidelines: Ethernet OAM configuration

The device's support for sending and receiving Information OAMPDUs carrying critical link events is as follows:

- Can receive Information OAMPDUs carrying the critical link events listed in Table 4.
- Can send Information OAMPDUs carrying Link Fault events.
- Can send Information OAMPDUs carrying Dying Gasp events when the device is rebooted or relevant ports are manually shut down. Physical IRF ports, however, are unable to send this type of OAMPDUs.
- Cannot send Information OAMPDUs carrying Critical Events.

Ethernet OAM tasks at a glance

To configure Ethernet OAM, perform the following tasks:

1. Configuring basic Ethernet OAM functions
2. (Optional.) Configuring the Ethernet OAM connection detection timers
3. (Optional.) Configuring link event detection
   - Configuring errored symbol event detection
   - Configuring errored frame event detection
   - Configuring errored frame period event detection

- Configuring errored frame seconds event detection

4. (Optional.) Configuring the action a port takes after it receives an Ethernet OAM event from the remote end

Configuring basic Ethernet OAM functions

About this task

To set up an Ethernet OAM connection between two Ethernet OAM entities, you must set at least one entity to operate in active mode. An Ethernet OAM entity can initiate OAM connection only in active mode.

Restrictions and guidelines

To change the Ethernet OAM mode on an Ethernet OAM-enabled port, first disable Ethernet OAM on the port.

Procedure

1. Enter system view.
   System-view
2. Enter Layer 2/Layer 3 Ethernet port view.
   interface *interface-type interface-number*
3. Set the Ethernet OAM mode.
   oam mode { **active** | **passive** }
   The default is active Ethernet OAM mode.
4. Enable Ethernet OAM.
   oam enable
   Ethernet OAM is disabled by default.

Configuring the Ethernet OAM connection detection timers

About the Ethernet OAM connection detection timers

After an Ethernet OAM connection is established, the Ethernet OAM entities exchange Information OAMPDUs at the handshake packet transmission interval to detect the availability of the Ethernet OAM connection. If an Ethernet OAM entity receives no Information OAMPDU within the Ethernet OAM connection timeout time, the Ethernet OAM connection is considered disconnected.

By adjusting the handshake packet transmission interval and the connection timeout timer, you can change the detection time resolution for Ethernet OAM connections.

Restrictions and guidelines for configuring Ethernet OAM connection detection timers

When you configure Ethernet OAM, follow these restrictions and guidelines:

- You can configure this command in system view or port view. The configuration in system view takes effect on all ports, and the configuration in port view takes effect on the specified port. For a port, the configuration in port view takes precedence.

- After the timeout timer of an Ethernet OAM connection expires, the local OAM entity ages out and terminates its connection with the peer OAM entity. To keep the Ethernet OAM connections stable, set the connection timeout timer to be at least five times the handshake packet transmission interval.

Configuring the Ethernet OAM connection detection timers globally

1. Enter system view.
   System-view
2. Configure the Ethernet OAM handshake packet transmission interval.
   oam global timer hello *interval*
   The default is 1000 milliseconds.
3. Configure the Ethernet OAM connection timeout timer.
   oam global timer keepalive *interval*
   The default is 5000 milliseconds.

Configuring the Ethernet OAM connection detection timers on a port

1. Enter system view.
   System-view
2. Enter Layer 2/Layer 3 Ethernet port view.
   interface *interface-type interface-number*
3. Configure the Ethernet OAM handshake packet transmission interval.
   oam timer hello *interval*
   By default, an interface uses the value configured globally.
4. Configure the Ethernet OAM connection timeout timer.
   oam timer keepalive *interval*
   By default, an interface uses the value configured globally.

Configuring errored symbol event detection

Restrictions and guidelines for configuring errored symbol event detection

You can configure this function in system view or port view. The configuration in system view takes effect on all ports, and the configuration in port view takes effect on the specified port. For a port, the configuration in port view takes precedence.

Configuring errored symbol event detection globally

1. Enter system view.
   system-view
2. Configure the errored symbol event detection window.

oam global errored-symbol-period window *window-value*

By default, the errored symbol event detection window is 100000000.

3. Configure the errored symbol event triggering threshold.

oam global errored-symbol-period threshold *threshold-value*

By default, the errored symbol event triggering threshold is 1.

Configuring errored symbol event detection on a port

1. Enter system view.

system-view

2. Enter Layer 2/Layer 3 Ethernet port view.

interface *interface-type interface-number*

3. Configure the errored symbol event detection window.

oam errored-symbol-period window *window-value*

By default, an interface uses the value configured globally.

4. Configure the errored symbol event triggering threshold.

oam errored-symbol-period threshold *threshold-value*

By default, an interface uses the value configured globally.

Configuring errored frame event detection

Restrictions and guidelines for configuring errored frame event detection

You can configure this function in system view or port view. The configuration in system view takes effect on all ports, and the configuration in port view takes effect on the specified port. For a port, the configuration in port view takes precedence.

Configuring errored frame event detection globally

1. Enter system view.

system-view

2. Configure the errored frame event detection window.

oam global errored-frame window *window-value*

By default, the errored frame event detection window is 1000 milliseconds.

3. Configure the errored frame event triggering threshold.

oam global errored-frame threshold *threshold-value*

By default, the errored frame event triggering threshold is 1.

Configuring errored frame event detection on a port

1. Enter system view.

system-view

2. Enter Layer 2/Layer 3 Ethernet port view.

interface *interface-type interface-number*

3. Configure the errored frame event detection window.
   oam errored-frame window *window-value*
   By default, an interface uses the value configured globally.
4. Configure the errored frame event triggering threshold.
   oam errored-frame threshold *threshold-value*
   By default, an interface uses the value configured globally.

Configuring errored frame period event detection

Restrictions and guidelines for configuring errored frame period event detection

You can configure this function in system view or port view. The configuration in system view takes effect on all ports, and the configuration in port view takes effect on the specified port. For a port, the configuration in port view takes precedence.

Configuring errored frame period event detection globally

1. Enter system view.
   system-view
2. Configure the errored frame period event detection window.
   oam global errored-frame-period window *window-value*
   By default, the errored frame period event detection window is 10000000.
3. Configure the errored frame period event triggering threshold.
   oam global errored-frame-period threshold *threshold-value*
   By default, the errored frame period event triggering threshold is 1.

Configuring errored frame period event detection on a port

1. Enter system view.
   system-view
2. Enter Layer 2/Layer 3 Ethernet port view.
   interface *interface-type interface-number*
3. Configure the errored frame period event detection window.
   oam errored-frame-period window *window-value*
   By default, an interface uses the value configured globally.
4. Configure the errored frame period event triggering threshold.
   oam errored-frame-period threshold *threshold-value*
   By default, an interface uses the value configured globally.

Configuring errored frame seconds event detection

Restrictions and guidelines for configuring errored frame seconds event detection

- You can configure this function in system view or port view. The configuration in system view takes effect on all ports, and the configuration in port view takes effect on the specified port. For a port, the configuration in port view takes precedence.
- Make sure the errored frame seconds triggering threshold is less than the errored frame seconds detection window. Otherwise, no errored frame seconds event can be generated.

Configuring errored frame seconds event detection globally

1. Enter system view.
   system-view
2. Configure the errored frame seconds event detection window.
   oam global errored-frame-seconds window *window-value*
   By default, the errored frame seconds event detection window is 60000 milliseconds.
3. Configure the errored frame seconds event triggering threshold.
   oam global errored-frame-seconds threshold *threshold-value*
   By default, the errored frame seconds event triggering threshold is 1.

Configuring errored frame seconds event detection on a port

1. Enter system view.
   system-view
2. Enter Layer 2/Layer 3 Ethernet port view.
   interface *interface-type interface-number*
3. Configure the errored frame seconds event detection window.
   oam errored-frame-seconds window *window-value*
   By default, an interface uses the value configured globally.
4. Configure the errored frame seconds event triggering threshold.
   oam errored-frame-seconds threshold *threshold-value*
   By default, an interface uses the value configured globally.

Configuring the action a port takes after it receives an Ethernet OAM event from the remote end

About this task

This feature enables a port to log events and automatically terminate the OAM connection and set the link state to down.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2/Layer 3 Ethernet port view.
   interface *interface-type interface-number*
3. Configure the action the port takes after it receives an Ethernet OAM event from the remote end.
   oam remote-failure { **connection-expired** | **critical-event** | **dying-gasp** | **link-fault** } **action error-link-down**
   By default, the port only logs the Ethernet OAM event it receives from the remote end.

Display and maintenance commands for Ethernet OAM

Execute **display** commands in any view and **reset** commands in user view:

Task	Command
Display information about an Ethernet OAM connection.	**display oam** { **local** | **remote** } [**interface** *interface-type interface-number*]
Display Ethernet OAM configuration.	**display oam configuration** [**interface** *interface-type interface-number*]
Display the statistics on critical events after an Ethernet OAM connection is established.	**display oam critical-event** [**interface** *interface-type interface-number*]
Display the statistics on Ethernet OAM link error events after an Ethernet OAM connection is established.	**display oam link-event** { **local** | **remote** } [**interface** *interface-type interface-number*]
Clear statistics on Ethernet OAM packets and Ethernet OAM link error events.	**reset oam** [**interface** *interface-type interface-number*]

Ethernet OAM configuration examples

Example: Configuring Ethernet OAM

Network configuration

On the network shown in Figure 1, perform the following operations:

- Enable Ethernet OAM on Device A and Device B to auto-detect link errors between the two devices
- Determine the performance of the link between Device A and Device B by collecting statistics about the error frames received by Device A

Figure 1 Network diagram

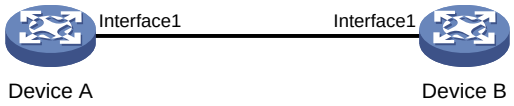


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure Device A:

   # Configure GigabitEthernet 1/0/1 to operate in active Ethernet OAM mode, and enable Ethernet OAM for it.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] oam mode active
[DeviceA-GigabitEthernet1/0/1] oam enable
```

   # Set the errored frame event detection window to 20000 milliseconds, and set the errored frame event triggering threshold to 10.

```
[DeviceA-GigabitEthernet1/0/1] oam errored-frame window 200
[DeviceA-GigabitEthernet1/0/1] oam errored-frame threshold 10
[DeviceA-GigabitEthernet1/0/1] quit
```

2. Configure Device B:

   # Configure GigabitEthernet 1/0/1 to operate in passive Ethernet OAM mode (the default), and enable Ethernet OAM for it.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] oam mode passive
[DeviceB-GigabitEthernet1/0/1] oam enable
[DeviceB-GigabitEthernet1/0/1] quit
```

Verifying the configuration

Use the **display oam critical-event** command to display the statistics of Ethernet OAM critical link events. For example:

Display the statistics of Ethernet OAM critical link events on all the ports of Device A.

```
[DeviceA] display oam critical-event
-----------[GigabitEthernet1/0/1] -----------
 Local link status   : UP
 Event statistics
   Link fault        : Not occurred
   Dying gasp        : Not occurred
   Critical event    : Not occurred
```

The output shows that no critical link event occurred on the link between Device A and Device B.

Use the **display oam link-event** command to display the statistics of Ethernet OAM link events. For example:

Display Ethernet OAM link event statistics of the local end of Device A.

```
[DeviceA] display oam link-event local
------------ [GigabitEthernet1/0/1] -----------
 Link status: UP
 OAM local errored frame event
   Event time stamp        : 5789 x 100 milliseconds
   Errored frame window    : 200 x 100 milliseconds
   Errored frame threshold : 10 error frames
   Errored frame           : 13 error frames
   Error running total     : 350 error frames
   Event running total     : 17 events
```

The output shows the following:

- 350 errors occurred after Ethernet OAM is enabled on Device A.
- 17 errors were caused by error frames.
- The link is unstable.

Contents

Configuring DLDP 1

About DLDP 1
Application scenario 1
Basic concepts 2
How DLDP works 3
Restrictions and guidelines: DLDP configuration 5
DLDP tasks at a glance 6
Enabling DLDP 6
Setting the interval to send advertisement packets 6
Setting the DelayDown timer 7
Setting the port shutdown mode 7
Configuring DLDP authentication 8
Display and maintenance commands for DLDP 8
DLDP configuration examples 9
Example: Configuring the auto port shutdown mode 9
Example: Configuring the manual port shutdown mode 12
Example: Configuring the hybrid port shutdown mode 16

Configuring DLDP

About DLDP

Physical layer detection mechanisms, such as auto-negotiation, can detect physical signals and faults. However, they cannot detect communication failures for unidirectional links where the physical layer is in connected state.

As a data link layer protocol, the H3C-proprietary Device Link Detection Protocol (DLDP) can do the following:

- Monitor status of the fiber link or twisted-pair link at the link layer.
- Cooperate with physical layer protocols to detect whether the link is correctly connected and whether the two ends of the link can exchange packets correctly.

When DLDP detects unidirectional links, it can automatically shut down the faulty port to avoid network problems. Alternatively, a user can manually shut down the faulty port.

Application scenario

A link becomes unidirectional when only one end of the link can receive packets from the other end.

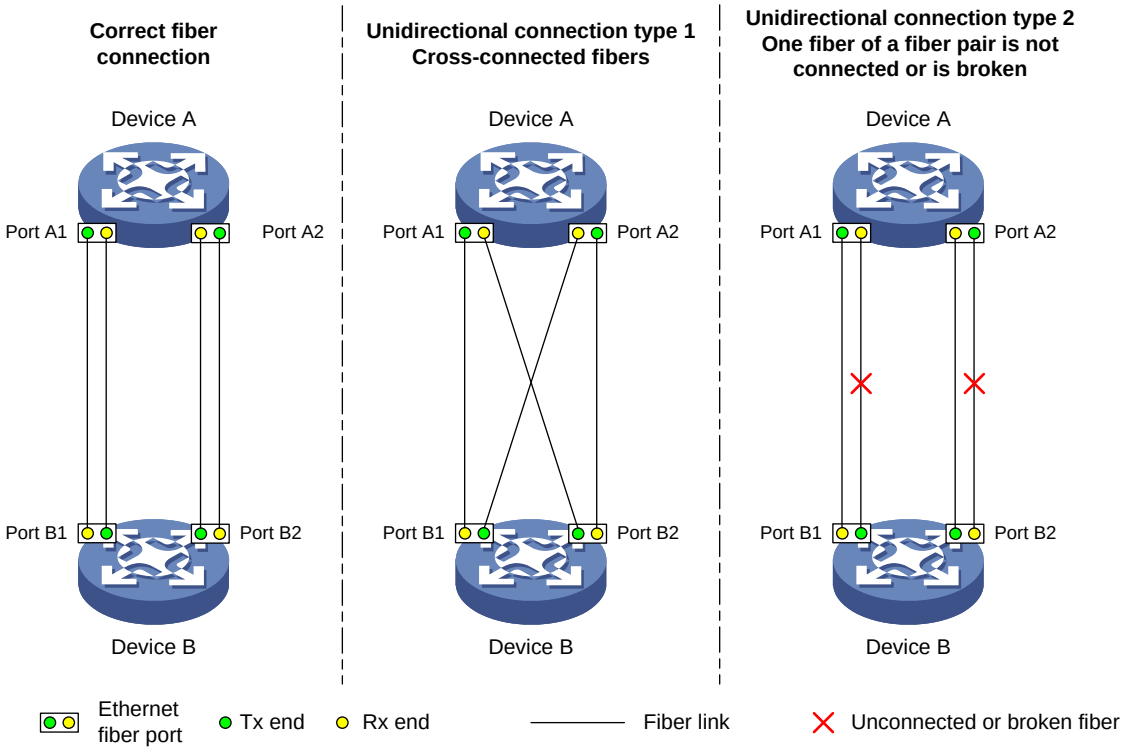
Unidirectional fiber links occur in the following cases:

- Fibers are cross-connected.
- A fiber is not connected at one end or one fiber of a fiber pair is broken.

Figure 1 shows a correct fiber connection and two types of unidirectional fiber connections.

When DLDP detects unidirectional links, it can automatically shut down the faulty port to avoid network problems. Alternatively, a user can manually shut down the faulty port.

Figure 1 Correct and incorrect fiber connections



Basic concepts

DLDP neighbor states

If port A can receive link-layer packets from port B on the same link, port B is a DLDP neighbor of port A. Two ports that can exchange packets are neighbors.

Table 1 DLDP neighbor states

DLDP timer	Description
Confirmed	The link to a DLDP neighbor is bidirectional.
Unconfirmed	The state of the link to a newly discovered neighbor is not determined.

DLDP port states

A DLDP-enabled port is called a DLDP port. A DLDP port can have multiple neighbors, and its state varies by the DLDP neighbor state.

Table 2 DLDP port states

State	Description
Initial	DLDP is enabled on the port, but is disabled globally.
Inactive	DLDP is enabled on the port and globally, and the link is physically down.
Bidirectional	DLDP is enabled on the port and globally, and at least one neighbor in Confirmed state exists.
Unidirectional	DLDP is enabled on the port and globally, and no neighbor in Confirmed state exists. In this state, a port does not send or receive packets other than DLDP packets any more.

DLDP timers

Table 3 DLDP timers

DLDP timer	Description
Advertisement timer	Advertisement packet sending interval (the default is 5 seconds and is configurable).
Probe timer	Probe packet sending interval. This timer is set to 1 second.
Echo timer	The Echo timer is triggered when a probe is launched for a new neighbor. The neighbor information is deleted when the timer expires. This timer is set to 10 seconds.
Entry timer	When a new neighbor joins, a neighbor entry is created and the corresponding Entry timer is triggered if the neighbor is in Confirmed state. When an Advertisement is received, the device updates the corresponding neighbor entry and the Entry timer. When the Entry timer expires, the Enhanced timer and Echo timer are triggered for the neighbor. The value of an Entry timer is three times that of the Advertisement timer.
Enhanced timer	Probe packet sending interval. This timer is set to 1 second. When the Entry timer expires, the Enhanced timer is triggered and probe packets are sent. At the same time, the Echo timer is triggered.
DelayDown timer	If a port is physically down, the device triggers the DelayDown timer, rather than removing the corresponding neighbor entry. The default DelayDown timer is 1 second and is configurable. When the DelayDown timer expires, the device removes the corresponding DLDP neighbor information if the port is down, and does not perform any operation if the port is up.
RecoverProbe timer	This timer is set to 2 seconds. A port in unidirectional state regularly sends RecoverProbe packets to detect whether a unidirectional link has been restored to bidirectional.

DLDP authentication mode

You can use DLDP authentication to prevent network attacks and illegal detecting.

Table 4 DLDP authentication mode

Authentication mode	Processing at the DLDP packet sending side	Processing at the DLDP packet receiving side
Non-authentication	The sending side sets the Authentication field of DLDP packets to 0.	The receiving side examines the authentication information of received DLDP packets and drops packets where the authentication information conflicts with the local configuration.
Plaintext authentication	The sending side sets the Authentication field to the password configured in plain text.	
MD5 authentication	The sending side encrypts the user configured password by using MD5 algorithm, and assigns the digest to the Authentication field.	

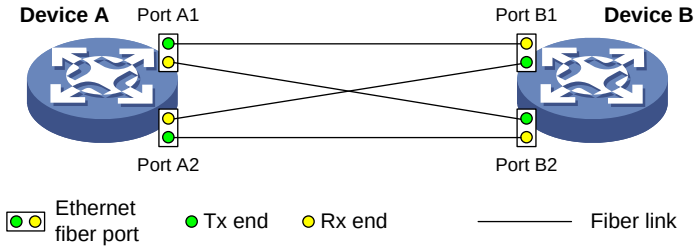
How DLDP works

Detecting one neighbor

When two devices are connected through an optical fiber or a network cable, enable DLDP to detect unidirectional links to the neighbor. The following illustrates the unidirectional link detection process in two cases:

- Unidirectional links occur before you enable DLDP.

  Figure 2 Cross-connected fibers

  As shown in Figure 2, before you enable DLDP, the optical fibers between Device A and Device B are cross-connected. After you enable DLDP, the four ports are all up and in unidirectional state, and they send RecoverProbe packets. Take Port A1 as an example to illustrate the unidirectional link detection process.

  a. Port A1 receives the RecoverProbe packet from Port B2, and returns a RecoverEcho packet.

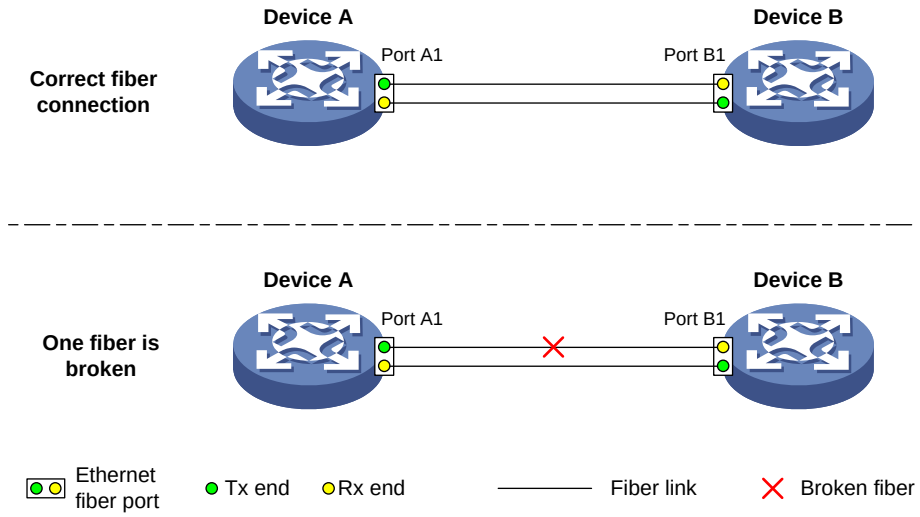
  b. Port B2 cannot receive any RecoverEcho packet from Port A1, so Port B2 cannot become the neighbor of Port A1.

  c. Port B1 can receive the RecoverEcho packet from Port A1, but Port B1 is not the intended destination, so Port B1 cannot become the neighbor of Port A1.

  The same process occurs on the other three ports. The four ports are all in unidirectional state.

- Unidirectional links occur after you enable DLDP.

  Figure 3 Broken fiber

  As shown in Figure 3, Device A and Device B are connected through an optical fiber. After you enable DLDP, Port A1 and Port B1 establish the bidirectional neighborship in the following way:

  a. Port A1 that is physically up enters the unidirectional state and sends a RecoverProbe packet.

  b. After receiving the RecoverProbe packet, Port B1 returns a RecoverEcho packet.

  c. After Port A1 receives the RecoverEcho packet, it examines the neighbor information in the packet. If the neighbor information matches the local information, Port A1 establishes the neighborship with Port B1 and transits to bidirectional state. Port A1 then starts the Entry timer and periodically sends Advertisement packets.

d. After Port B1 receives the Advertisement packet, it establishes the Unconfirmed neighborship with Port A1. Port B1 then starts the Echo timer and Probe timer, and periodically sends Probe packets.

e. After receiving the Probe packet, Port A1 returns an Echo packet.

f. After Port B1 receives the Echo packet, it examines the neighbor information in the packet. If the neighbor information matches the local information, the neighbor state of Port A1 becomes Confirmed. Port B1 then transits to bidirectional state, starts the Entry timer, and periodically sends Advertisement packets.

The bidirectional neighborship between Port A1 and Port B1 is now established.

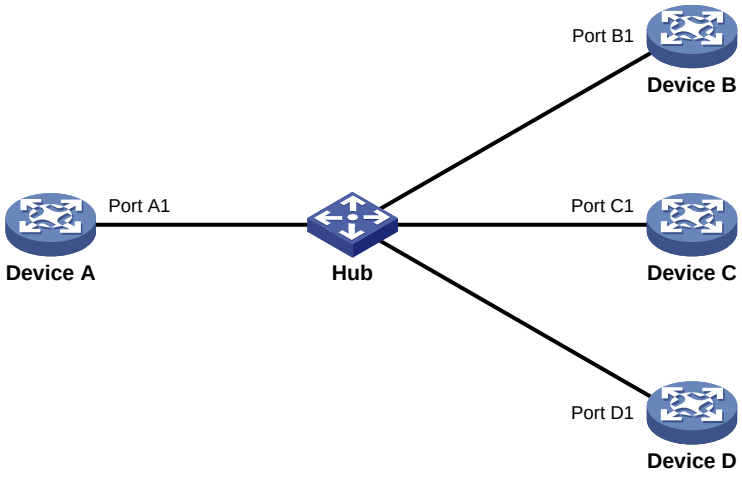
After that, when Port B1's Rx end fails to receive signals, Port B1 is physically down and enters the Inactive state. Because Port B1's Tx end can still send signals to Port A1, Port A1 stays up. After the Entry timer for Port B1 expires, Port A1 starts the Enhanced timer and Echo timer, and sends a probe packet to Port B1. Because Port A1's Tx line is broken, Port A1 cannot receive the Echo packet from Port B1 after the Echo timer expires. Port A1 then enters the unidirectional state, and sends a Disable packet to Port B1. At the same time, Port A1 deletes the neighborship with Port B1, and starts the RecoverProbe timer. Port B1 stays in Inactive state during this process.

When an interface is physically down, but the Tx end of the interface is still operating, DLDP sends a LinkDown packet to inform the peer to delete the relevant neighbor entry.

Detecting multiple neighbors

When multiple devices are connected through a hub, enable DLDP on all interfaces connected to the hub to detect unidirectional links among the neighbors. When no Confirmed neighbor exists, an interface enters the unidirectional state.

Figure 4 Network diagram



As shown in Figure 4, Device A through Device D are connected through a hub, and enabled with DLDP. When Port A1, Port B1, and Port C1 detect that the link to Port D1 fails, they delete the neighborship with Port D1, but stay in bidirectional state.

Restrictions and guidelines: DLDP configuration

When you configure DLDP, follow these configuration restrictions and guidelines:

- For DLDP to operate correctly, enable DLDP on both sides and make sure the following settings are consistent:
  - Interval to send Advertisement packets.
  - DLDP authentication mode.

- Password.

- For DLDP to operate correctly, configure the full duplex mode for the ports at the two ends of the link, and configure the same speed for the two ports.

DLDP tasks at a glance

To configure DLDP, perform the following tasks:

1. Enabling DLDP
2. (Optional.) Setting the interval to send advertisement packets
3. (Optional.) Setting the DelayDown timer
4. (Optional.) Setting the port shutdown mode
5. (Optional.) Configuring DLDP authentication

Enabling DLDP

About this task

After a port is enabled with DLDP, the port enters Initial state and then transitions to Unidirectional state. By default, DLDP blocks a port immediately when the port transitions to Unidirectional state. This behavior causes a traffic disruption that lasts until the port enters Bidirectional state when establishing a Confirmed neighbor.

You can set the delay time for DLDP to block a port upon an Initial-to-Unidirectional state transition. DLDP does not block the port until the delay time expires.

Procedure

1. Enter system view.

   system-view

2. Enable DLDP globally.

   dldp global enable

   By default, DLDP is globally disabled.

3. Enter Ethernet interface view.

   interface *interface-type interface-number*

4. Enable DLDP and set the delay time for DLDP to block the port upon an Initial-to-Unidirectional state transition.

   dldp enable [**initial-unidirectional-delay** *time*]

   By default, DLDP is disabled on a port, and when DLDP is enabled, a port is blocked immediately upon an Initial-to-Unidirectional state transition.

Setting the interval to send advertisement packets

About this task

To make sure DLDP can detect unidirectional links before network performance deteriorates, set the advertisement interval appropriate for your network environment. As a best practice, use the default interval.

Procedure

1. Enter system view.

   system-view

2. Set the interval to send Advertisement packets.

   dldp interval *interval*

   By default, the interval is 5 seconds.

Setting the DelayDown timer

About this task

When the Tx line fails, some ports might go down and then come up again, causing optical signal jitters on the Rx line. To prevent the device from removing neighbor entries in such cases, set the DelayDown timer for the device. The device starts the DelayDown timer when a port goes down due to a Tx line failure. If the port remains down when the timer expires, the device removes the DLDP neighbor information. If the port comes up, the device takes no action.

Procedure

1. Enter system view.

   system-view

2. Set the DelayDown timer.

   dldp delaydown-timer *time*

   The default is 1 second.

   The DelayDown timer setting applies to all DLDP-enabled ports.

Setting the port shutdown mode

About this task

On detecting a unidirectional link, DLDP shuts down the ports in one of the following modes:

- **Auto mode**—When DLDP detects a unidirectional link, it shuts down the unidirectional port. When the link becomes bidirectional, DLDP brings up the port that was shut down.
- **Manual mode**—When DLDP detects a unidirectional link, it does not shut down the port. You need to manually shut it down. To verify the link status, use the **undo shutdown** command to bring up the port. If the link becomes bidirectional, the port becomes bidirectional. Use this mode to prevent normal links from being shut down because of false unidirectional link reports in the following cases:
  - The network performance is low.
  - The device is busy.
  - The CPU usage is high.
- **Hybrid mode**—When DLDP detects a unidirectional link, it shuts down the unidirectional port and stops link detection. To verify the link status, use the **undo shutdown** command to bring up the port. If the link becomes bidirectional, the port becomes bidirectional.

Restrictions and guidelines

You can set the port shutdown mode for all interfaces in system view or for a single interface in interface view. The setting in interface view takes precedence over the setting in system view.

Setting the global port shutdown mode

1. Enter system view.

   system-view

2. Set the global port shutdown mode.

   dldp unidirectional-shutdown { **auto** | **hybrid** | **manual** }

The default mode is **auto**.

Setting the port shutdown mode for an interface

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Set the the port shutdown mode for the interface.
   dldp port unidirectional-shutdown { **auto** | **hybrid** | **manual** }
   By default, the global setting is used.

Configuring DLDP authentication

About this task

You can guard your network against attacks and malicious probes by configuring an appropriate DLDP authentication mode, which can be plain text authentication or MD5 authentication. If your network is safe, you can choose not to authenticate.

Procedure

1. Enter system view.
   system-view
2. Configure a DLDP authentication mode.
   dldp authentication-mode { **md5** | **none** | **simple** }
   The default authentication mode is **none**.
3. Configure the password for DLDP authentication.
   dldp authentication-password { **cipher** | **simple** } *string*
   By default, no password is configured for DLDP authentication.
   If you do not configure the authentication password after you configure the authentication mode, the authentication mode is **none** no matter which authentication mode you configure.

Display and maintenance commands for DLDP

Execute **display** commands in any view and the **reset** command in user view.

Task	Command
Display the DLDP configuration globally and of a port.	**display dldp** [**interface** *interface-type interface-number*]
Display the statistics on DLDP packets passing through a port.	**display dldp statistics** [**interface** *interface-type interface-number*]
Clear the statistics on DLDP packets passing through a port.	**reset dldp statistics** [**interface** *interface-type interface-number*]

DLDP configuration examples

Example: Configuring the auto port shutdown mode

Network configuration

As shown in Figure 5, Device A and Device B are connected through two fiber pairs.

Configure DLDP to automatically shut down the faulty port upon detecting a unidirectional link, and automatically bring up the port after you clear the fault.

Figure 5 Network diagram

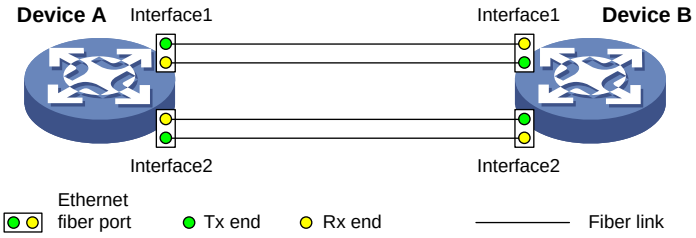


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

   # Enable DLDP globally.

   ```
   <DeviceA> system-view
   [DeviceA] dldp global enable
   ```

   # Configure GigabitEthernet 1/0/1 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on the port.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] duplex full
   [DeviceA-GigabitEthernet1/0/1] speed 1000
   [DeviceA-GigabitEthernet1/0/1] dldp enable
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

   # Configure GigabitEthernet 1/0/2 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on the port.

   ```
   [DeviceA] interface gigabitethernet 1/0/2
   [DeviceA-GigabitEthernet1/0/2] duplex full
   [DeviceA-GigabitEthernet1/0/2] speed 1000
   [DeviceA-GigabitEthernet1/0/2] dldp enable
   [DeviceA-GigabitEthernet1/0/2] quit
   ```

   # Set the global port shutdown mode to auto.

   ```
   [DeviceA] dldp unidirectional-shutdown auto
   ```

2. Configure Device B:

   # Enable DLDP globally.

```
<DeviceB> system-view
[DeviceB] dldp global enable
```

Configure GigabitEthernet 1/0/1 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on it.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] duplex full
[DeviceB-GigabitEthernet1/0/1] speed 1000
[DeviceB-GigabitEthernet1/0/1] dldp enable
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on it.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] duplex full
[DeviceB-GigabitEthernet1/0/2] speed 1000
[DeviceB-GigabitEthernet1/0/2] dldp enable
[DeviceB-GigabitEthernet1/0/2] quit
```

Set the global port shutdown mode to auto.

```
[DeviceB] dldp unidirectional-shutdown auto
```

Verifying the configuration

Display the DLDP configuration globally and on all the DLDP-enabled ports of Device A.

```
[DeviceA] display dldp
 DLDP global status: Enabled
 DLDP advertisement interval: 5s
 DLDP authentication-mode: None
 DLDP unidirectional-shutdown mode: Auto
 DLDP delaydown-timer value: 1s
 Number of enabled ports: 2

Interface GigabitEthernet1/0/1
 DLDP port state: Bidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 1
  Neighbor MAC address: 0023-8956-3600
  Neighbor port index: 1
  Neighbor state: Confirmed
  Neighbor aged time: 11s
  Neighbor echo time: -

Interface GigabitEthernet1/0/2
 DLDP port state: Bidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 1
  Neighbor MAC address: 0023-8956-3600
  Neighbor port index: 2
  Neighbor state: Confirmed
  Neighbor aged time: 12s
```

```
  Neighbor echo time: -
```

The output shows that both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are bidirectional.

Enable the monitoring of logs on the current terminal on Device A. Set the lowest level of the logs that can be output to the current terminal to 6.

```
[DeviceA] quit
<DeviceA> terminal monitor
The current terminal is enabled to display logs.
<DeviceA> terminal logging level 6
```

The following log information is displayed on Device A:

```
<DeviceA>%Jul 11 17:40:31:089 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface Physical state on the interface GigabitEthernet1/0/1 changed to down.
%Jul 11 17:40:31:091 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface GigabitEthernet1/0/1 changed to down.
%Jul 11 17:40:31:677 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface Physical state on the interface GigabitEthernet1/0/2 changed to down.
%Jul 11 17:40:31:678 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface GigabitEthernet1/0/2 changed to down.
%Jul 11 17:40:38:544 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface Physical state on the interface GigabitEthernet1/0/1 l changed to up.
%Jul 11 17:40:38:836 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface Physical state on the interface GigabitEthernet1/0/2 changed to up.
```

The output shows the following:

- The port status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 is down and then up.
- The link status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 is always down.

Display the DLDP configuration globally and of all the DLDP-enabled ports.

```
<DeviceA> display dldp
 DLDP global status: Enabled
 DLDP advertisement interval: 5s
 DLDP authentication-mode: None
 DLDP unidirectional-shutdown mode: Auto
 DLDP delaydown-timer value: 1s
 Number of enabled ports: 2

Interface GigabitEthernet1/0/1
 DLDP port state: Unidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 0 (Maximum number ever detected: 1)

Interface GigabitEthernet1/0/2
 DLDP port state: Unidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 0 (Maximum number ever detected: 1)
```

The output shows that the DLDP port status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 is unidirectional. DLDP detects unidirectional links on them and automatically shuts down the two ports.

The unidirectional links are caused by cross-connected fibers. Correct the fiber connections. As a result, the ports shut down by DLDP automatically recover, and Device A displays the following log information:

```
<DeviceA>%Jul 11 17:42:57:709 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the
interface GigabitEthernet1/0/1 changed to down.
%Jul 11 17:42:58:603 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/2 changed to down.
%Jul 11 17:43:02:342 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/1 changed to up.
%Jul 11 17:43:02:343 2012 DeviceA DLDP/6/DLDP_NEIGHBOR_CONFIRMED: A neighbor was confirmed
on interface GigabitEthernet1/0/1. The neighbor's system MAC is 0023-8956-3600, and the
port index is 1.
%Jul 11 17:43:02:344 2012 DeviceA DLDP/6/DLDP_LINK_BIDIRECTIONAL: DLDP detected a
bidirectional link on interface GigabitEthernet1/0/1.
%Jul 11 17:43:02:353 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/1 changed to up.
%Jul 11 17:43:02:357 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/2 changed to up.
%Jul 11 17:43:02:362 2012 DeviceA DLDP/6/DLDP_NEIGHBOR_CONFIRMED: A neighbor was confirmed
on interface GigabitEthernet1/0/2. The neighbor's system MAC is 0023-8956-3600, and the
port index is 2.
%Jul 11 17:43:02:362 2012 DeviceA DLDP/6/DLDP_LINK_BIDIRECTIONAL: DLDP detected a
bidirectional link on interface GigabitEthernet1/0/2.
%Jul 11 17:43:02:368 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/2 changed to up.
```

The output shows that the port status and link status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are now up and their DLDP neighbors are determined.

Example: Configuring the manual port shutdown mode

Network configuration

As shown in Figure 6, Device A and Device B are connected through two fiber pairs.

Configure DLDP to detect unidirectional links. When a unidirectional link is detected, the administrator must manually shut down the port.

Figure 6 Network diagram

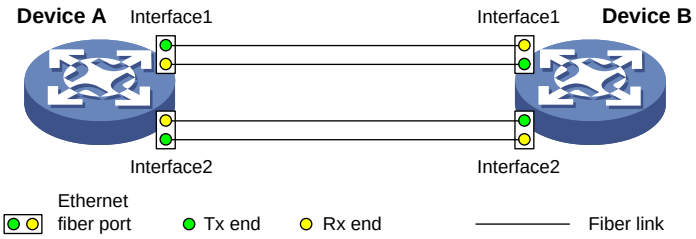


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

   # Enable DLDP globally.

   ```
   <DeviceA> system-view
   [DeviceA] dldp enable
   ```

   # Configure GigabitEthernet 1/0/1 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on the port.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] duplex full
   [DeviceA-GigabitEthernet1/0/1] speed 1000
   [DeviceA-GigabitEthernet1/0/1] dldp enable
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

   # Configure GigabitEthernet 1/0/2 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on the port.

   ```
   [DeviceA] interface gigabitethernet 1/0/2
   [DeviceA-GigabitEthernet1/0/2] duplex full
   [DeviceA-GigabitEthernet1/0/2] speed 1000
   [DeviceA-GigabitEthernet1/0/2] dldp enable
   [DeviceA-GigabitEthernet1/0/2] quit
   ```

   # Set the global port shutdown mode to manual.

   ```
   [DeviceA] dldp unidirectional-shutdown manual
   ```

2. Configure Device B:

   # Enable DLDP globally.

   ```
   <DeviceB> system-view
   [DeviceB] dldp global enable
   ```

   # Configure GigabitEthernet 1/0/1 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on it.

   ```
   [DeviceB] interface gigabitethernet 1/0/1
   [DeviceB-GigabitEthernet1/0/1] duplex full
   [DeviceB-GigabitEthernet1/0/1] speed 1000
   [DeviceB-GigabitEthernet1/0/1] dldp enable
   [DeviceB-GigabitEthernet1/0/1] quit
   ```

   # Configure GigabitEthernet 1/0/2 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on it.

   ```
   [DeviceB] interface gigabitethernet 1/0/2
   [DeviceB-GigabitEthernet1/0/2] duplex full
   [DeviceB-GigabitEthernet1/0/2] speed 1000
   [DeviceB-GigabitEthernet1/0/2] dldp enable
   [DeviceB-GigabitEthernet1/0/2] quit
   ```

   # Set the global port shutdown mode to manual.

   ```
   [DeviceB] dldp unidirectional-shutdown manual
   ```

Verifying the configuration

Display the DLDP configuration globally and on all the DLDP-enabled ports of Device A.

```
[DeviceA] display dldp
 DLDP global status: Enabled
 DLDP advertisement interval: 5s
 DLDP authentication-mode: None
```

```
 DLDP unidirectional-shutdown mode: Manual
 DLDP delaydown-timer value: 1s
 Number of enabled ports: 2

Interface GigabitEthernet1/0/1
 DLDP port state: Bidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 1
  Neighbor MAC address: 0023-8956-3600
  Neighbor port index: 1
  Neighbor state: Confirmed
  Neighbor aged time: 11s
  Neighbor echo time: -

Interface GigabitEthernet1/0/2
 DLDP port state: Bidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 1
  Neighbor MAC address: 0023-8956-3600
  Neighbor port index: 2
  Neighbor state: Confirmed
  Neighbor aged time: 12s
  Neighbor echo time: -
```

The output shows that both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are in bidirectional state, which means both links are bidirectional.

Enable the monitoring of logs on the current terminal on Device A. Set the lowest level of the logs that can be output to the current terminal to 6.

```
[DeviceA] quit
<DeviceA> terminal monitor
The current terminal is enabled to display logs.
<DeviceA> terminal logging level 6
```

The following log information is displayed on Device A:

```
<DeviceA>%Jul 12 08:29:17:786 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the
interface GigabitEthernet1/0/1 changed to down.
%Jul 12 08:29:17:787 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/1 changed to down.
%Jul 12 08:29:17:800 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/2 changed to down.
%Jul 12 08:29:17:800 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/2 changed to down.
%Jul 12 08:29:25:004 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/1 changed to up.
%Jul 12 08:29:25:005 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/1 changed to up.
%Jul 12 08:29:25:893 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/2 changed to up.
```

```
%Jul 12 08:29:25:894 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/2 changed to up.
```

The output shows that the port status and link status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are down and then up.

Display the DLDP configuration globally and of all the DLDP-enabled ports.

```
<DeviceA> display dldp
 DLDP global status: Enabled
 DLDP advertisement interval: 5s
 DLDP authentication-mode: None
 DLDP unidirectional-shutdown mode: Manual
 DLDP delaydown-timer value: 1s
 Number of enabled ports: 2

Interface GigabitEthernet1/0/1
 DLDP port state: Unidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 0 (Maximum number ever detected: 1)

Interface GigabitEthernet1/0/2
 DLDP port state: Unidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 0 (Maximum number ever detected: 1)
```

The output shows that the DLDP port status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 is unidirectional. DLDP detects unidirectional links on the two ports but does not shut them down.

The unidirectional links are caused by cross-connected fibers. Manually shut down the two ports:

Shut down GigabitEthernet 1/0/1.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] shutdown
```

The following log information is displayed on Device A:

```
[DeviceA-GigabitEthernet1/0/1]%Jul 12 08:34:23:717 2012 DeviceA IFNET/3/PHY_UPDOWN:
Physical state on the interface GigabitEthernet1/0/1 changed to down.
%Jul 12 08:34:23:718 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/1 changed to down.
%Jul 12 08:34:23:778 2012 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/2 changed to down.
%Jul 12 08:34:23:779 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/2 changed to down.
```

The output shows that the port status and link status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are now down.

Shut down GigabitEthernet 1/0/2.

```
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] shutdown
```

Correct the fiber connections and bring up the two ports:

Bring up GigabitEthernet 1/0/2.

```
[DeviceA-GigabitEthernet1/0/2] undo shutdown
```

The following log information is displayed on Device A:

```
[DeviceA-GigabitEthernet1/0/2]%Jul 12 08:46:17:677 2012 DeviceA IFNET/3/PHY_UPDOWN:
Physical state on the interface GigabitEthernet1/0/2 changed to up.

%Jul 12 08:46:17:678 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/2 changed to up.

%Jul 12 08:46:17:959 2012 DeviceA DLDP/6/DLDP_NEIGHBOR_CONFIRMED: A neighbor was confirmed
on interface GigabitEthernet1/0/2. The neighbor's system MAC is 0023-8956-3600, and the
port index is 2.

%Jul 12 08:46:17:959 2012 DeviceA DLDP/6/DLDP_LINK_BIDIRECTIONAL: DLDP detected a
bidirectional link on interface GigabitEthernet1/0/2.
```

The output shows that the port status and link status of GigabitEthernet 1/0/2 are now up and its DLDP neighbors are determined.

Bring up GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/2] quit

[DeviceA] interface gigabitethernet 1/0/1

[DeviceA-GigabitEthernet1/0/1] undo shutdown
```

The following log information is displayed on Device A:

```
[DeviceA-GigabitEthernet1/0/1]%Jul 12 08:48:25:952 2012 DeviceA IFNET/3/PHY_UPDOWN:
Physical state on the interface GigabitEthernet1/0/1 changed to up.

%Jul 12 08:48:25:952 2012 DeviceA DLDP/6/DLDP_NEIGHBOR_CONFIRMED: A neighbor was confirmed
on interface GigabitEthernet1/0/1. The neighbor's system MAC is 0023-8956-3600, and the
port index is 1.

%Jul 12 08:48:25:953 2012 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/1 changed to up.

%Jul 12 08:48:25:953 2012 DeviceA DLDP/6/DLDP_LINK_BIDIRECTIONAL: DLDP detected a
bidirectional link on interface GigabitEthernet1/0/1.
```

The output shows that the port status and link status of GigabitEthernet 1/0/1 are now up and its DLDP neighbors are determined.

Example: Configuring the hybrid port shutdown mode

Network configuration

As shown in Figure 7Figure 6, Device A and Device B are connected through two fiber pairs.

Configure DLDP to detect unidirectional links. When a unidirectional link is detected, DLDP automatically shuts down the unidirectional port. The administrator needs to bring up the port after clearing the fault.

Figure 7 Network diagram

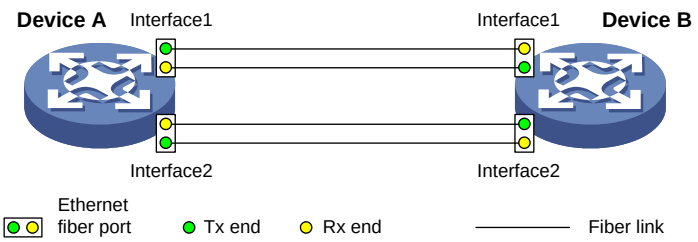


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

   # Enable DLDP globally.

   ```
   <DeviceA> system-view
   [DeviceA] dldp enable
   ```

   # Configure GigabitEthernet 1/0/1 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on the port.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] duplex full
   [DeviceA-GigabitEthernet1/0/1] speed 1000
   [DeviceA-GigabitEthernet1/0/1] dldp enable
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

   # Configure GigabitEthernet 1/0/2 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on the port.

   ```
   [DeviceA] interface gigabitethernet 1/0/2
   [DeviceA-GigabitEthernet1/0/2] duplex full
   [DeviceA-GigabitEthernet1/0/2] speed 1000
   [DeviceA-GigabitEthernet1/0/2] dldp enable
   [DeviceA-GigabitEthernet1/0/2] quit
   ```

   # Set the global port shutdown mode to hybrid.

   ```
   [DeviceA] dldp unidirectional-shutdown hybrid
   ```

2. Configure Device B:

   # Enable DLDP globally.

   ```
   <DeviceB> system-view
   [DeviceB] dldp global enable
   ```

   # Configure GigabitEthernet 1/0/1 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on it.

   ```
   [DeviceB] interface gigabitethernet 1/0/1
   [DeviceB-GigabitEthernet1/0/1] duplex full
   [DeviceB-GigabitEthernet1/0/1] speed 1000
   [DeviceB-GigabitEthernet1/0/1] dldp enable
   [DeviceB-GigabitEthernet1/0/1] quit
   ```

   # Configure GigabitEthernet 1/0/2 to operate in full duplex mode and at 1000 Mbps, and enable DLDP on it.

   ```
   [DeviceB] interface gigabitethernet 1/0/2
   [DeviceB-GigabitEthernet1/0/2] duplex full
   [DeviceB-GigabitEthernet1/0/2] speed 1000
   [DeviceB-GigabitEthernet1/0/2] dldp enable
   [DeviceB-GigabitEthernet1/0/2] quit
   ```

   # Set the global port shutdown mode to hybrid.

   ```
   [DeviceB] dldp unidirectional-shutdown hybrid
   ```

Verifying the configuration

Display the DLDP configuration globally and on all the DLDP-enabled ports of Device A.

```
[DeviceA] display dldp
 DLDP global status: Enabled
 DLDP advertisement interval: 5s
 DLDP authentication-mode: None
 DLDP unidirectional-shutdown mode: Hybrid
 DLDP delaydown-timer value: 1s
 Number of enabled ports: 2

Interface GigabitEthernet1/0/1
 DLDP port state: Bidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 1
  Neighbor MAC address: 0023-8956-3600
  Neighbor port index: 1
  Neighbor state: Confirmed
  Neighbor aged time: 11s
  Neighbor echo time: -

Interface GigabitEthernet1/0/2
 DLDP port state: Bidirectional
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port’s neighbors: 1
  Neighbor MAC address: 0023-8956-3600
  Neighbor port index: 2
  Neighbor state: Confirmed
  Neighbor aged time: 12s
  Neighbor echo time: -
```

The output shows that both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are in bidirectional state, which means both links are bidirectional.

Enable the monitoring of logs on the current terminal on Device A. Set the lowest level of the logs that can be output to the current terminal to 6.

```
[DeviceA] quit
<DeviceA> terminal monitor
The current terminal is enabled to display logs.
<DeviceA> terminal logging level 6
```

The following log information is displayed on Device A:

```
<DeviceA>%Jan  4 07:16:06:556 2011 DeviceA DLDP/5/DLDP_NEIGHBOR_AGED: A neighbor on
interface
GigabitEthernet1/0/1 was deleted because the neighbor was aged. The neighbor's system MAC
is 0023-8956-3600, and the port index is 162.
%Jan  4 07:16:06:560 2011 DeviceA DLDP/5/DLDP_NEIGHBOR_AGED: A neighbor on interface
GigabitEthernet1/0/2 was deleted because the neighbor was aged. The neighbor's system MAC
is 0023-8956-3600, and the port index is 165.
```

```
%Jan  4 07:16:06:724 2011 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/1 changed to down.
%Jan  4 07:16:06:730 2011 DeviceA IFNET/3/PHY_UPDOWN: Physical state on the interface
GigabitEthernet1/0/2 changed to down.
%Jan 4 07:16:06:736 2011 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/1 changed to down.
%Jan 4 07:16:06:738 2011 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/2 changed to down.
%Jan  4 07:16:07:152 2011 DeviceA DLDP/3/DLDP_LINK_UNIDIRECTIONAL: DLDP detected a
unidirectional link on interface GigabitEthernet1/0/1. DLDP automatically shut down the
interface. Please manually bring up the interface.
%Jan  4 07:16:07:156 2011 DeviceA DLDP/3/DLDP_LINK_UNIDIRECTIONAL: DLDP detected a
unidirectional link on interface GigabitEthernet1/0/2. DLDP automatically shut down the
interface. Please manually bring up the interface.
```

The output shows that the port status and link status of both GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are down.

Display the DLDP configuration globally and of all the DLDP-enabled ports.

```
<DeviceA> display dldp
 DLDP global status: Enabled
 DLDP advertisement interval: 5s
 DLDP authentication-mode: None
 DLDP unidirectional-shutdown mode: Hybrid
 DLDP delaydown-timer value: 1s
 Number of enabled ports: 2

Interface GigabitEthernet1/0/1
 DLDP port state: Inactive
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port's neighbors: 0 (Maximum number ever detected: 1)
Interface GigabitEthernet1/0/2
 DLDP port state: Inactive
 DLDP port unidirectional-shutdown mode: None
 DLDP initial-unidirectional-delay: 0s
 Number of the port's neighbors: 0 (Maximum number ever detected: 1)
```

The output shows that DLDP detects a unidirectional link and shuts down GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

The unidirectional links are caused by cross-connected fibers. Bring up the two ports after correcting the fiber connection:

Bring up GigabitEthernet 1/0/1.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo shutdown
```

The following log information is displayed on Device A:

```
[DeviceA-GigabitEthernet1/0/1]%Jan  4 07:33:26:574 2011 DeviceA IFNET/3/PHY_UPDOWN:
Physical state on the interface GigabitEthernet1/0/1 changed to up.
%Jan 4 07:33:57:562 2011 DeviceA DLDP/6/DLDP_NEIGHBOR_CONFIRMED: A neighbor was confirmed
on interface GigabitEthernet1/0/1. The neighbor's system MAC is 0023-8956-3600, and the
port index is 162.
```

```
%Jan  4 07:33:57:563 2011 DeviceA DLDP/6/DLDP_LINK_BIDIRECTIONAL: DLDP detected a
bidirectional link on interface GigabitEthernet1/0/1.
%Jan 4 07:33:57:590 2011 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/1 changed to up.
%Jan  4 07:33:57:609 2011 DeviceA STP/6/STP_DETECTED_TC: Instance 0's port
GigabitEthernet1/0/1 detected a topology change.
```

The output shows that the port status and link status of GigabitEthernet 1/0/1 are now up and its DLDP neighbors are determined.

Bring up GigabitEthernet 1/0/2.

```
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] undo shutdown
```

The following log information is displayed on Device A:

```
[DeviceA-GigabitEthernet1/0/2]%Jan  4 07:35:26:574 2011 DeviceA IFNET/3/PHY_UPDOWN:
Physical state on the interface GigabitEthernet1/0/2 changed to up.
%Jan 4 07:35:57:562 2011 DeviceA DLDP/6/DLDP_NEIGHBOR_CONFIRMED: A neighbor was confirmed
on interface GigabitEthernet1/0/2. The neighbor's system MAC is 0023-8956-3600, and the
port index is 162.
%Jan  4 07:35:57:563 2011 DeviceA DLDP/6/DLDP_LINK_BIDIRECTIONAL: DLDP detected a
bidirectional link on interface GigabitEthernet1/0/2.
%Jan 4 07:35:57:590 2011 DeviceA IFNET/5/LINK_UPDOWN: Line protocol state on the interface
GigabitEthernet1/0/2 changed to up.
%Jan  4 07:35:57:609 2011 DeviceA STP/6/STP_DETECTED_TC: Instance 0's port
GigabitEthernet1/0/2 detected a topology change.
```

The output shows that the port status and link status of GigabitEthernet 1/0/2 are now up and its DLDP neighbors are determined.

Contents

Configuring RRPP 1

About RRPP 1
RRPP networking 1
RRPPDUs 3
RRPP timers 4
How RRPP works 5
Typical RRPP networking 6
Protocols and standards 10
Restrictions and guidelines: RRPP configuration 10
RRPP tasks at a glance 10
Prerequisites for RRPP 10
Creating an RRPP domain 11
Configuring control VLANs 11
Configuring protected VLANs 12
Configuring RRPP rings 12
Prerequisites 12
Configuring RRPP ports 12
Configuring RRPP nodes 13
Activating an RRPP domain 14
Configuring RRPP timers 15
Restrictions and guidelines for RRPP timer configuration 15
Configuring the Hello timer and Fail timer 15
Configuring the link-up delay timer 15
Configuring an RRPP ring group 16
Enabling SNMP notifications for RRPP 16
Display and maintenance commands for RRPP 17
RRPP configuration examples 17
Example: Configuring a single ring 17
Example: Configuring intersecting rings 22
Example: Configuring load-balanced intersecting rings 28
Troubleshooting RRPP 38
The primary node cannot receive Hello packets when the link state is normal 38

Configuring RRPP

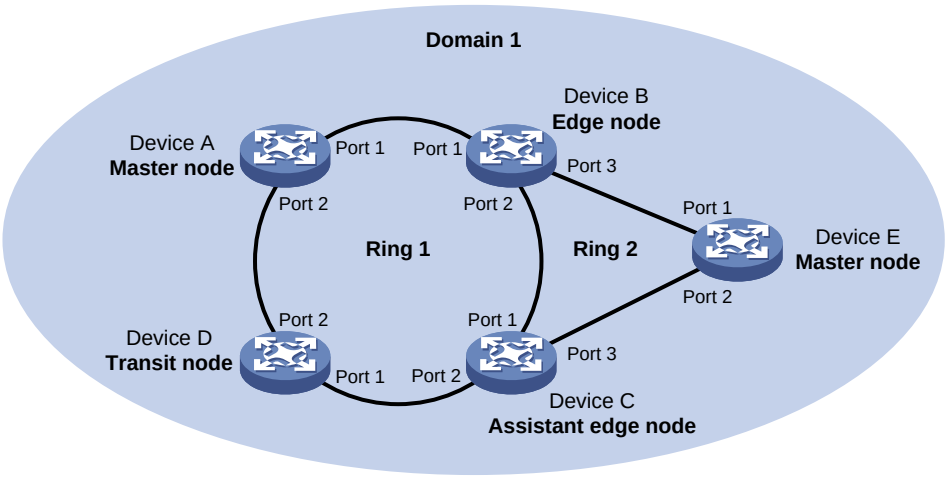
About RRPP

The Rapid Ring Protection Protocol (RRPP) is an H3C-proprietary link layer protocol designed for Ethernet rings. RRPP can prevent broadcast storms caused by data loops when an Ethernet ring is healthy. RRPP can also rapidly restore the communication paths between the nodes when a link is disconnected on the ring. Compared with the spanning tree protocol, the convergence time of RRPP is fast and independent of the number of nodes in the Ethernet ring. RRPP is applicable to large-diameter networks.

RRPP networking

Figure 1 shows a typical RRPP network with two Ethernet rings and multiple nodes. RRPP detects ring status and sends topology change information by exchanging Rapid Ring Protection Protocol Data Units (RRPPDUs) among the nodes.

Figure 1 `RRPP networking diagram`



RRPP domain

An RRPP domain is uniquely identified by a domain ID. The interconnected devices with the same domain ID and control VLANs constitute an RRPP domain. An RRPP domain contains the following elements:

- Primary ring and subring.
- Control VLAN.
- Master node, transit node, edge node, and assistant edge node.
- Primary port, secondary port, common port, and edge port.

As shown in Figure 1, Domain 1 is an RRPP domain, containing two RRPP rings: Ring 1 and Ring 2. All the nodes on the two RRPP rings belong to the RRPP domain.

RRPP ring

A ring-shaped Ethernet topology is called an RRPP ring. RRPP rings include primary rings and subrings. You can configure a ring as either the primary ring or a subring by specifying its ring level. The primary ring is of level 0, and a subring is of level 1. An RRPP domain contains one or multiple RRPP rings, one serving as the primary ring and the others serving as subrings. A ring can be in one of the following states:

- **`Health state`**—All physical links on the Ethernet ring are connected.
- **`Disconnect state`**—Some physical links on the Ethernet ring are not connected.

As shown in Figure 1, Domain 1 contains two RRPP rings: Ring 1 and Ring 2. The level is set to 0 for Ring 1 and 1 for Ring 2. Ring 1 is configured as the primary ring, and Ring 2 is configured as a subring.

Control VLAN and protected VLAN

1. Control VLAN

   In an RRPP domain, a control VLAN is dedicated to transferring RRPPDUs. On a device, the ports accessing an RRPP ring belong to the control VLANs of the ring, and only these ports can join the control VLANs.

   An RRPP domain is configured with the following control VLANs:

   - One primary control VLAN, which is the control VLAN for the primary ring.
   - One secondary control VLAN, which is the control VLAN for subrings.

   After you specify a VLAN as the primary control VLAN, the system automatically configures the secondary control VLAN. The VLAN ID is the primary control VLAN ID plus one. All subrings in the same RRPP domain share the same secondary control VLAN. IP address configuration is prohibited on the control VLAN interfaces.

2. Protected VLAN

   A protected VLAN is dedicated to transferring data packets. Both RRPP ports and non-RRPP ports can be assigned to a protected VLAN.

Node role

Each device on an RRPP ring is a node. The role of a node is configurable. RRPP has the following node roles:

- **Master node**—Each ring has only one master node. The master node initiates the polling mechanism and determines the operations to be performed after a topology change.
- **Transit node**—On the primary ring, transit nodes refer to all nodes except the master node. On the subring, transit nodes refer to all nodes except the master node and the nodes where the primary ring intersects with the subring. A transit node monitors the state of its directly connected RRPP links and notifies the master node of the link state changes, if any. Based on the link state changes, the master node determines the operations to be performed.
- **Edge node**—A special node residing on both the primary ring and a subring at the same time. An edge node acts as a master node or transit node on the primary ring and as an edge node on the subring.
- **Assistant edge node**—A special node residing on both the primary ring and a subring at the same time. An assistant edge node acts as a master node or transit node on the primary ring and as an assistant edge node on the subring. This node works in conjunction with the edge node to detect the integrity of the primary ring and to perform loop guard.

As shown in Figure 1, Ring 1 is the primary ring and Ring 2 is a subring. Device A is the master node of Ring 1. Device B, Device C, and Device D are the transit nodes of Ring 1. Device E is the master node of Ring 2, Device B is the edge node of Ring 2, and Device C is the assistant edge node of Ring 2.

Port role

1. Primary port and secondary port

   Each master node or transit node has two ports connected to an RRPP ring, a primary port and a secondary port. You can determine the role of a port.

   In terms of functionality, the primary port and the secondary port of a master node have the following differences:

   - The primary port and the secondary port are designed to play the role of sending and receiving Hello packets, respectively.

- When an RRPP ring is in Health state, the secondary port logically denies protected VLANs and permits only the packets from the control VLANs.
- When an RRPP ring is in Disconnect state, the secondary port forwards packets from protected VLANs.

In terms of functionality, the primary port and the secondary port of a transit node are the same. Both are designed for transferring protocol packets and data packets over an RRPP ring.

As shown in Figure 1, Device A is the master node of Ring 1. Port 1 and Port 2 are the primary port and the secondary port of the master node on Ring 1, respectively. Device B, Device C, and Device D are the transit nodes of Ring 1. Their Port 1 and Port 2 are the primary port and the secondary port on Ring 1, respectively.

2. Common port and edge port

   The ports connecting the edge node and assistant edge node to the primary ring are common ports. The ports connecting the edge node and assistant edge node only to the subrings are edge ports. You can determine the role of a port.

   As shown in Figure 1, Device B and Device C reside on Ring 1 and Ring 2. Device B's Port 1 and Port 2 and Device C's Port 1 and Port 2 access the primary ring, so they are common ports. Device B's Port 3 and Device C's Port 3 access only the subring, so they are edge ports.

RRPP ring group

To reduce Edge-Hello traffic, you can configure a group of subrings on the edge node or assistant edge node. You must configure a device as the edge node of these subrings, and another device as the assistant edge node of these subrings. Additionally, the subrings of the edge node and assistant edge node must connect to the same subring packet tunnels in major ring (SRPTs). Edge-Hello packets of the edge node of these subrings travel to the assistant edge node of these subrings over the same link.

An RRPP ring group configured on the edge node is an edge node RRPP ring group. An RRPP ring group configured on an assistant edge node is an assistant edge node RRPP ring group. Only one subring in an edge node RRPP ring group is allowed to send Edge-Hello packets.

RRPPDUs

RRPPDUs of subrings are transmitted as data packets in the primary ring, and RRPPDUs of the primary ring can only be transmitted within the primary ring. In the primary ring, Common-Flush-FDB packets and Complete-Flush-FDB packets of subrings are sent to the CPU on each node for processing.

Table 1 RRPPDU types and their functions

Type	Description
Hello	The master node sends Hello packets (also known as Health packets) to detect the integrity of a ring in a network.
Link-Down	When a port on the transit node, edge node, or assistant edge node fails, the node initiates Link-Down packets to notify the master node of the disconnection of the ring.
Common-Flush-FDB	When an RRPP ring transits to Disconnect state, the master node initiates Common-Flush-FDB (FDB stands for Forwarding Database) packets. It uses the packets to instruct the transit nodes, edge nodes, and assistant edge nodes to update their own MAC address entries and ARP/ND entries.
Complete-Flush-FDB	When an RRPP ring transits to Health state, the master node sends Complete-Flush-FDB packets for the following purposes: • Instruct the transit nodes, edge nodes, and assistant edge nodes to update their MAC address entries and ARP/ND entries. • Instruct the transit nodes to unblock temporarily blocked ports.

Edge-Hello	The edge node sends Edge-Hello packets to examine the SRPTs between the edge node and the assistant edge node.
Major-Fault	The assistant edge node sends Major-Fault packets to notify the edge node of SRPT failure when an SRPT between assistant edge node and edge node is disconnected.

RRPP timers

When RRPP determines the link state of an Ethernet ring, it uses the following timers:

Hello timer

The Hello timer specifies the interval at which the master node sends Hello packets out of the primary port.

Fail timer

The Fail timer specifies the maximum delay of Hello packets sent from the primary port to the secondary port of the master node. If the secondary port receives the Hello packets sent by the local master node before the Fail timer expires, the ring is in Health state. If the secondary port does not receive the Hello packets before the Fail timer expires, the ring transits to Disconnect state.

In an RRPP domain, a transit node learns the Fail timer value on the master node through the received Hello packets. This ensures that all nodes in the ring network have consistent Fail timer settings.

Link-up delay timer

This feature prevents frequent switchover of RRPP traffic forwarding paths caused by unstable RRPP port states. This feature behaves differently depending on whether you specify the **`distribute`** keyword in the **`linkup-delay-timer`** command.

- If you do not specify the **`distribute`** keyword, the master node starts the link-up delay timer when a faulty port comes up and the master node receives Hello packets from the secondary port.
  - If the master node can still receive Hello packets from the secondary port after the link-up delay timer expires, the master node performs the following operations:
    - Changes the RRPP ring state from Disconnect to Health.
    - Switches the traffic from the secondary port to the primary port.
  - If the master node cannot receive Hello packets from the secondary port after the Fail timer expires and before the link-up delay timer expires, the master node performs the following operations:
    - Stops the link-up delay timer.
    - Keeps the RRPP ring in Disconnect state.
- If you specify the **`distribute`** keyword, all nodes in the RRPP domain can learn the value of the link-up delay timer through Hello packets. When the faulty port comes up, the master node performs the following operations:
  - The hosting RRPP node blocks the faulty port (the faulty port cannot send or receive any packets).
  - Starts the link-up delay timer.

  If the port does not become faulty after the link-up delay timer expires, the hosting RRPP node sets the port state to up. The master node can receive Hello packets from its secondary port again. Then, the master node changes the RRPP ring state from Disconnect to Health and switches the traffic from the secondary port to the primary port.

  If the port becomes faulty again before the link-up delay timer expires, the hosting RRPP node blocks the port and stops the link-up delay timer.

How RRPP works

Polling mechanism

The polling mechanism is used by the master node of an RRPP ring to examine the Health state of the ring network.

The master node sends Hello packets out of its primary port at the Hello interval. These Hello packets travel through each transit node on the ring in turn.

- If the ring is complete, the secondary port of the master node receives Hello packets before the Fail timer expires. The master node keeps the secondary port blocked.
- If the ring is disconnected, the secondary port of the master node fails to receive Hello packets before the Fail timer expires. The master node releases the secondary port from blocking protected VLANs. It sends Common-Flush-FDB packets to instruct all transit nodes to update their own MAC address entries and ARP/ND entries.

Load balancing

In a ring network, traffic from multiple VLANs might be transmitted at the same time. RRPP can implement load balancing by transmitting traffic from different VLANs along different paths.

You can configure multiple RRPP domains for a ring network. Each RRPP domain transmits the traffic from the specified VLANs (protected VLANs). Traffic from different VLANs can be transmitted according to different topologies in the ring network for load balancing.

As shown in the schematic diagram for a single-ring load balancing network, Ring 1 is configured as the primary ring of Domain 1 and Domain 2, which are configured with different protected VLANs. Device A is the master node of Ring 1 in Domain 1. Device B is the master node of Ring 1 in Domain 2. With such configurations, traffic from different VLANs can be transmitted on different links for load balancing in the single-ring network.

Link down alarm mechanism

In an RRPP domain, when the transit node, edge node, or assistant edge node finds that any of its ports is down, it immediately sends Link-Down packets to the master node. When the master node receives a Link-Down packet, it takes the following actions:

- Releases the secondary port from blocking protected VLANs.
- Sends Common-Flush-FDB packets to instruct all the transit nodes, edge nodes, and assistant edge nodes to update their MAC address entries and ARP/ND entries.

After each node updates its own entries, traffic is switched to the normal link.

Ring recovery

When the ports in an RRPP domain on the transit nodes, edge nodes, or assistant edge nodes come up again, the ring is recovered. However, the master node might detect the ring recovery after a period of time. A temporary loop might arise in the protected VLAN during this period. As a result, a broadcast storm occurs.

To prevent such cases, non-master nodes block the ports immediately when they find the ports accessing the ring are brought up again. The nodes block only the packets from the protected VLAN, and they permit only the packets from the control VLAN to pass through. The blocked ports are activated only when the nodes determine that no loop will be generated by these ports.

Broadcast storm suppression mechanism in case of SRPT failure in a multi-homed subring

As shown in the schematic diagram for a dual-homed-ring network, Ring 1 is the primary ring, and Ring 2 and Ring 3 are subrings. When the two SRPTs between the edge node and the assistant edge node are down, the master nodes of Ring 2 and Ring 3 will open their secondary ports. A loop is generated among Device B, Device C, Device E, and Device F, causing a broadcast storm.

To avoid generating a loop, the edge node will temporarily block the edge port. The blocked edge port is activated only when the edge node determines that no loop will be generated when the edge port is activated.

RRPP ring group

In an edge node RRPP ring group, only the activated subring with the smallest domain ID and ring ID can send Edge-Hello packets. In an assistant edge node RRPP ring group, any activated subring that has received Edge-Hello packets will forward these packets to the other activated subrings. When an edge node RRPP ring group and an assistant edge node RRPP ring group are configured, the CPU workload is reduced because of the following reasons:

- Only one subring sends Edge-Hello packets on the edge node.
- Only one subring receives Edge-Hello packets on the assistant edge node.

As shown in the schematic diagram for a dual-homed-ring network, Device B is the edge node of Ring 2 and Ring 3. Device C is the assistant edge node of Ring 2 and Ring 3. Device B and Device C need to send or receive Edge-Hello packets frequently. If more subrings are configured or more domains are configured for load balancing, Device B and Device C will send or receive a large number of Edge-Hello packets.

To reduce Edge-Hello traffic, perform the following tasks:

- Assign Ring 2 and Ring 3 to an RRPP ring group configured on the edge node Device B.
- Assign Ring 2 and Ring 3 to an RRPP ring group configured on the assistant edge node Device C.

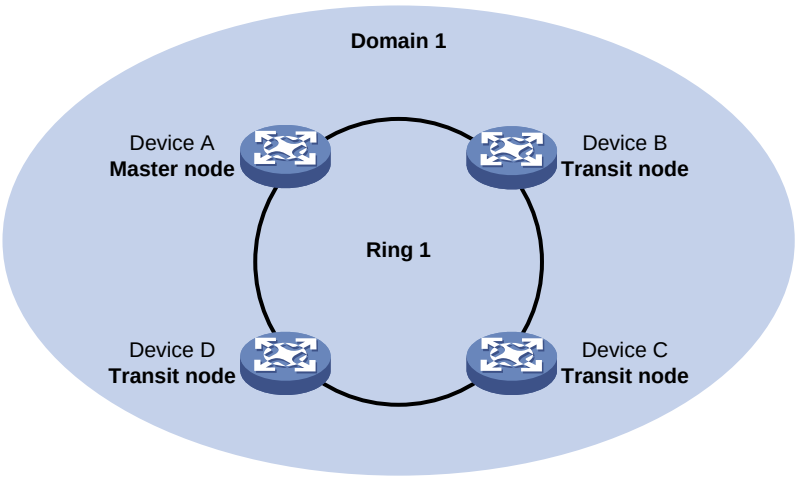
If all rings are activated, only Ring 2 on Device B sends Edge-Hello packets.

Typical RRPP networking

Single ring

As shown in Figure 2, only a single ring exists in the network topology. You need only define an RRPP domain.

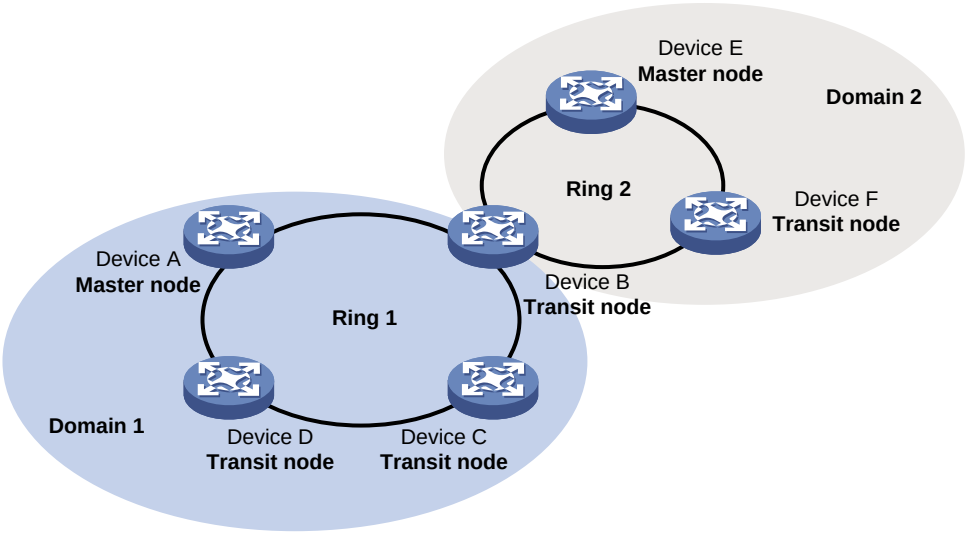
Figure 2 Schematic diagram for a single-ring network



Tangent rings

As shown in Figure 3, two or more rings exist in the network topology and only one common node exists between rings. You must define an RRPP domain for each ring.

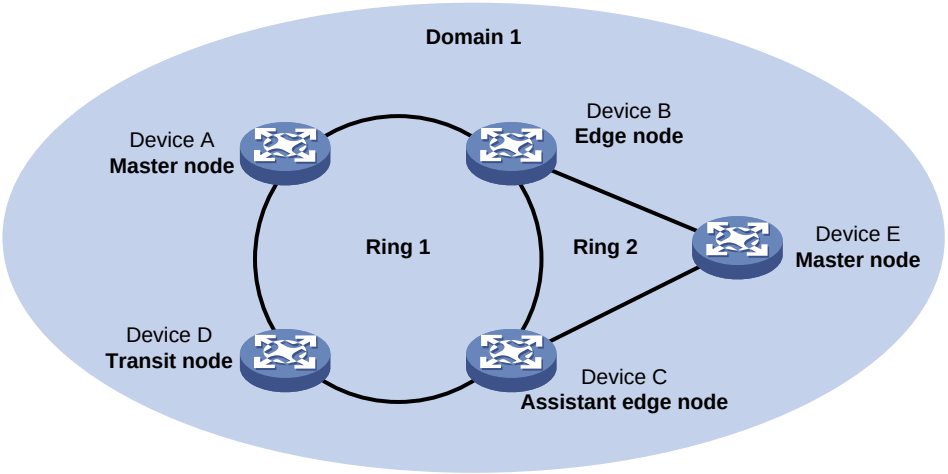
Figure 3 Schematic diagram for a tangent-ring network



Intersecting rings

As shown in Figure 4, two or more rings exist in the network topology and two common nodes exist between rings. You need only define an RRPP domain and configure one ring as the primary ring and the other rings as subrings.

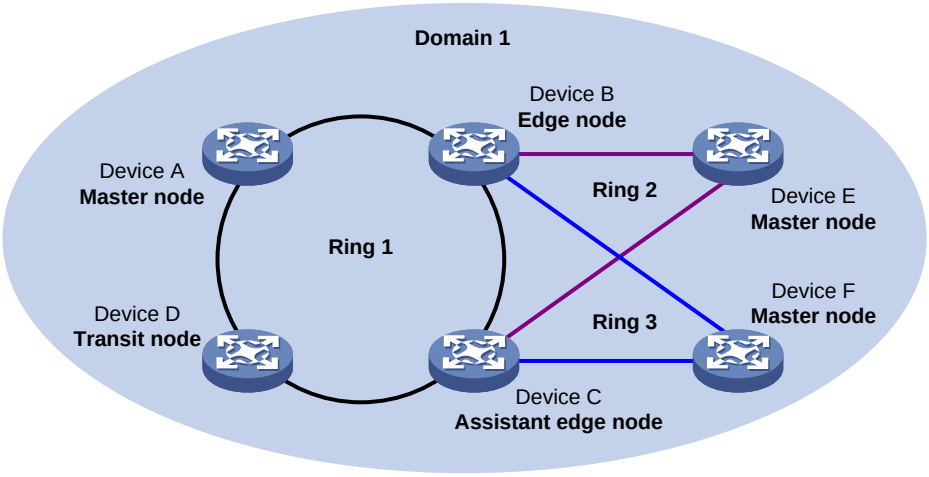
Figure 4 Schematic diagram for an intersecting-ring network



Dual-homed rings

As shown in Figure 5, two or more rings exist in the network topology and two similar common nodes exist between rings. You need only define an RRPP domain and configure one ring as the primary ring and the other rings as subrings.

Figure 5 Schematic diagram for a dual-homed-ring network



Single-ring load balancing

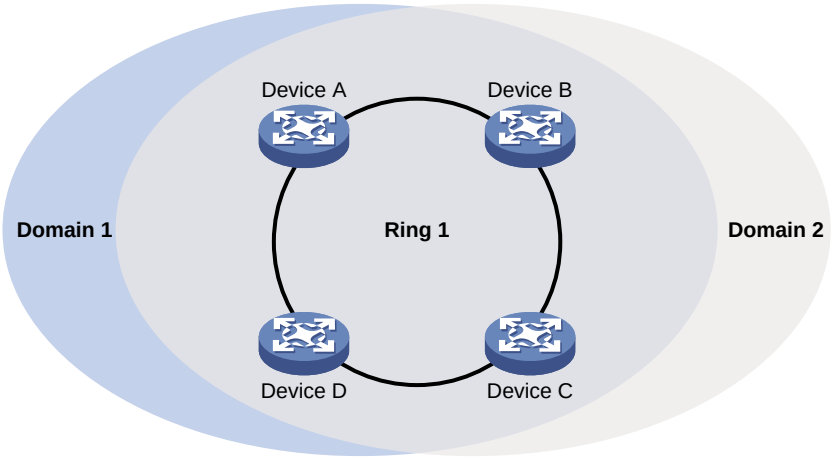
In a single-ring network, you can achieve load balancing by configuring multiple domains.

As shown in Figure 6:

- Ring 1 is configured as the primary ring of both Domain 1 and Domain 2.
- Domain 1 and Domain 2 are configured with different protected VLANs.
- In Domain 1, Device A is configured as the master node of Ring 1.
- In Domain 2, Device B is configured as the master node of Ring 1.

Such configurations enable the ring to block different links based on VLANs and achieve single-ring load balancing.

Figure 6 Schematic diagram for a single-ring load balancing network



Intersecting-ring load balancing

In an intersecting-ring network, you can also achieve load balancing by configuring multiple domains.

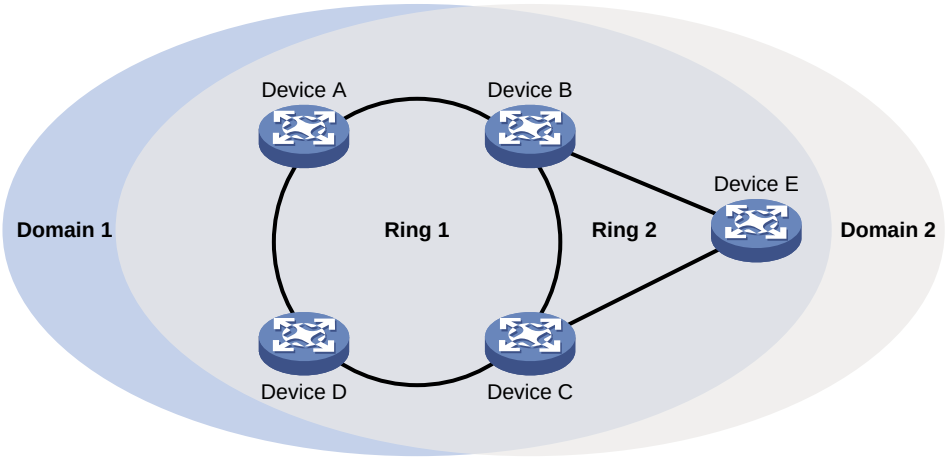
As shown in Figure 7:

- Ring 1 is the primary ring and Ring 2 is the subring in both Domain 1 and Domain 2.
- Domain 1 and Domain 2 are configured with different protected VLANs.
- Device A is configured as the master node of Ring 1 in Domain 1.

- Device D is configured as the master node of Ring 1 in Domain 2.
- Device E is configured as the master node of Ring 2 in both Domain 1 and Domain 2. However, different ports on Device E are blocked in Domain 1 and Domain 2.

Traffic from different VLANs can travel over different paths in the subring and primary ring.

Figure 7 Schematic diagram for an intersecting-ring load balancing network

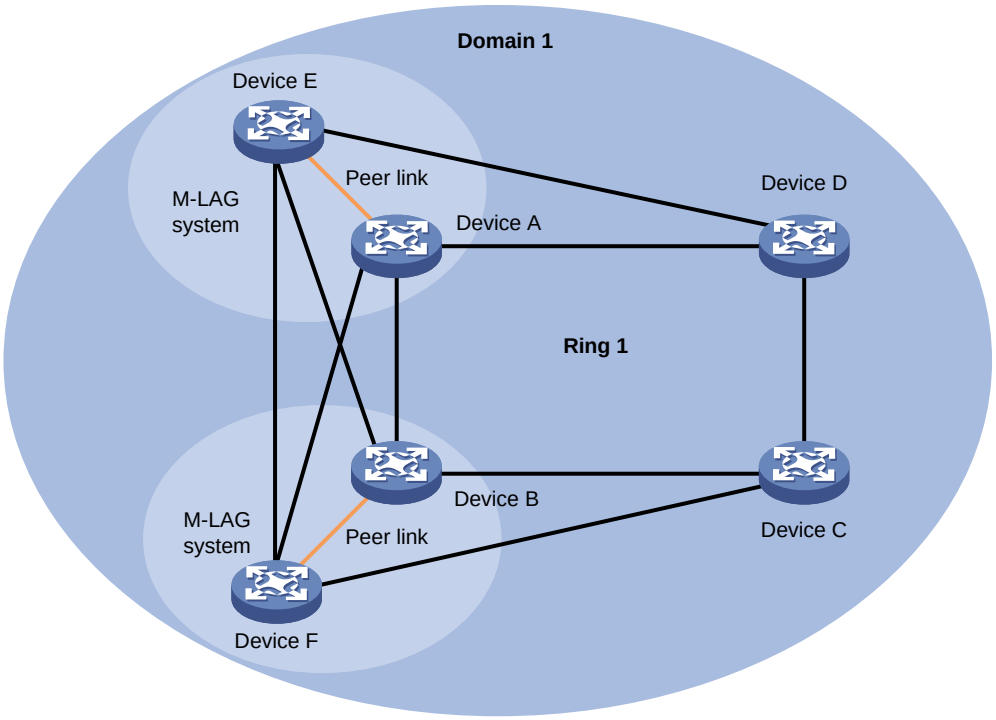


RRPP with M-LAG

M-LAG allows you to virtualize the devices in an RRPP ring into one system through multichassis link, thereby achieving device-level redundancy and load balancing.

As shown in Figure 8, the M-LAG system composed of Device A and Device E is the master node of Ring 1, and the M-LAG system composed of Device B and Device F is the transit node of Ring 1. For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*.

Figure 8 RRPP with M-LAG



Protocols and standards

RFC 3619, *Extreme Networks' Ethernet Automatic Protection Switching (EAPS) Version 1*

Restrictions and guidelines: RRPP configuration

- RRPP does not have an auto election mechanism. You must configure each node in the ring network correctly for RRPP to monitor and protect the ring network.
- You can configure RRPP in the following order:
  - Create RRPP domains based on service planning.
  - Specify control VLANs and protected VLANs for each RRPP domain.
  - Determine the ring roles and node roles based on the traffic paths in each RRPP domain.
- To ensure correct operation of the RRPP feature, do not configure user authentication settings (such as AAA, 802.1X, MAC address authentication, and portal) on RRPP ports.

RRPP tasks at a glance

To configure RRPP, perform the following tasks:

1. Creating an RRPP domain

   Perform this task on devices you want to configure as nodes in the RRPP domain.
2. Configuring control VLANs

   Perform this task on all nodes in the RRPP domain.
3. Configuring protected VLANs

   Perform this task on all nodes in the RRPP domain.
4. Configuring RRPP rings
   a. Configuring RRPP ports

      Perform this task on each node's ports intended for accessing RRPP rings.
   b. Configuring RRPP nodes

      Perform this task on all nodes in the RRPP domain.
5. Activating an RRPP domain

   Perform this task on all nodes in the RRPP domain.
6. (Optional.) Configuring RRPP timers

   Perform this task on the master node in the RRPP domain.
   - Configuring the Hello timer and Fail timer
   - Configuring the link-up delay timer
7. (Optional.) Configuring an RRPP ring group

   Perform this task on the edge node and assistant edge node in the RRPP domain.
8. (Optional.) Enabling SNMP notifications for RRPP

Prerequisites for RRPP

Before you configure RRPP, you must physically construct a ring-shaped Ethernet topology.

Creating an RRPP domain

About this task

When you create an RRPP domain, specify a domain ID to uniquely identify the RRPP domain. All devices in the same RRPP domain must be configured with the same domain ID.

Restrictions and guidelines

Perform this task on devices you want to configure as nodes in the RRPP domain.

Procedure

1. Enter system view.
   `system-view`
2. Create an RRPP domain and enter RRPP domain view.
   `rrpp domain` *`domain-id`*

Configuring control VLANs

Restrictions and guidelines

- Perform this task on all nodes in the RRPP domain.
- Before you configure RRPP rings in an RRPP domain, configure the same control VLANs for all nodes in the RRPP domain first. You need only configure the primary control VLAN for an RRPP domain. The system automatically configures the secondary control VLAN. It uses the primary control VLAN ID plus 1 as the secondary control VLAN ID. For the control VLAN configuration to succeed, make sure the IDs of the two control VLANs are consecutive and have not been previously assigned.
- Do not configure the default VLAN of a port accessing an RRPP ring as the control VLAN.
- A VLAN with VLAN interface resource reservation cannot be configured as the control VLAN. For more information about VLAN interface resource reservation, see *Layer 2—LAN Switching Configuration Guide*.
- The primary and secondary control VLAN IDs must be different from the Layer 3 Ethernet subinterface IDs of the master ring and subrings.
- For RRPPDUs to be correctly forwarded, do not enable QinQ or VLAN mapping on control VLANs.
- After you configure RRPP rings for an RRPP domain, you cannot delete or modify the primary control VLAN of the domain. You can only use the **`undo control-vlan`** command to delete a primary control VLAN.
- To transparently transmit RRPPDUs on a device not configured with RRPP, make sure only the two ports accessing the RRPP ring permit packets from the control VLANs. Otherwise, the packets from other VLANs might enter the control VLANs in transparent transmission mode and strike the RRPP ring.
- When configuring RRPP with M-LAG, make sure the peer-link interface is a trunk port. The peer-link interface will automatically join or leave a control VLAN as the control VLAN is created or deleted.

Procedure

1. Enter system view.
   `system-view`
2. Enter RRPP domain view.
   `rrpp domain` *`domain-id`*

3. Configure the primary control VLAN for the RRPP domain.
   control-vlan *vlan-id*

Configuring protected VLANs

Restrictions and guidelines

- Perform this task on all nodes in the RRPP domain.
- Before you configure RRPP rings in an RRPP domain, configure the same protected VLANs for all nodes in the RRPP domain. All VLANs to which the RRPP ports are assigned must be protected by the RRPP domains.
- When you configure load balancing, you must configure different protected VLANs for different RRPP domains.

Prerequisites

Before you configure protected VLANs, you must configure an MST region and the VLAN-to-instance mapping table. For more information about MST regions, see spanning tree configuration in *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter RRPP domain view.
   rrpp domain *domain-id*
3. Configure protected VLANs for the RRPP domain.
   protected-vlan reference-instance *instance-id-list*

Configuring RRPP rings

When you configure an RRPP ring, you must configure the ports connecting each node to the RRPP ring before configuring the nodes.

Prerequisites

Before you configure an RRPP ring, you must configure control VLANs and protected VLANs.

Configuring RRPP ports

Restrictions and guidelines

- Perform this task on each node's ports intended for accessing RRPP rings.
- To accelerate topology convergence, use the **link-delay** command to enable link status rapid report function on an RRPP port. Use this command to set the physical state change suppression interval to 0 seconds. For more information about the **link-delay** command, see *Layer 2—LAN Switching Command Reference*.
- Do not assign a port to both an aggregation group and an RRPP ring. If you do so, the port does not take effect on the RRPP ring.
- Do not assign a peer-link interface to an RRPP ring. A peer-link interface does not work for an RRPP ring.

- Do not assign both an M-LAG interface and a non-M-LAG interface to an RRPP ring. If you do so, the M-LAG interface does not take effect on the RRPP ring.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   interface *interface-type interface-number*
3. Configure the link type of the interface as trunk.
   port link-type trunk
   By default, the link type of an interface is access.
   For more information about the command, see *Layer 2—LAN Switching Command Reference*.
4. Assign the trunk port to the protected VLANs of the RRPP domain.
   port trunk permit vlan { *vlan-id-list* | **all** }
   By default, a trunk port allows only packets from VLAN 1 to pass through.
   RRPP ports always allow packets from the control VLANs to pass through.
   For more information about the command, see *Layer 2—LAN Switching Command Reference*.
5. Disable the spanning tree feature.
   undo stp enable
   By default, the spanning tree feature is enabled.
   For more information about the command, see *Layer 2—LAN Switching Command Reference*.

Configuring RRPP nodes

Restrictions and guidelines

- Perform this task on all nodes in the RRPP domain.
- If a device carries multiple RRPP rings in an RRPP domain, it can only be an edge node or an assistant edge node on a subring.
- When you configure an edge node or an assistant edge node, you must configure the primary ring before configuring the subrings.

Specifying a master node

1. Enter system view.
   system-view
2. Enter RRPP domain view.
   rrpp domain *domain-id*
3. Specify the current device as the master node of the ring, and specify the primary port and the secondary port.
   ring *ring-id* **node-mode master** [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*

Specifying a transit node

1. Enter system view.
   system-view
2. Enter RRPP domain view.
   rrpp domain *domain-id*

3. Specify the current device as a transit node of the ring, and specify the primary port and the secondary port.

   ring *ring-id* **node-mode transit** [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*

Specifying an edge node

1. Enter system view.

   system-view

2. Enter RRPP domain view.

   rrpp domain *domain-id*

3. Specify the current device as a master node or transit node of the primary ring, and specify the primary port and the secondary port.

   ring *ring-id* **node-mode** { **master** | **transit** } [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*

4. Specify the current device as the edge node of a subring, and specify the edge port.

   ring *ring-id* **node-mode edge** [**edge-port** *interface-type interface-number*]

Specifying an assistant edge node

1. Enter system view.

   system-view

2. Enter RRPP domain view.

   rrpp domain *domain-id*

3. Specify the current device as a master node or transit node of the primary ring, and specify the primary port and the secondary port.

   ring *ring-id* **node-mode** { **master** | **transit** } [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*

   By default, a device is not a node of the RRPP ring.

4. Specify the current device as the assistant edge node of the subring, and specify an edge port.

   ring *ring-id* **node-mode assistant-edge** [**edge-port** *interface-type interface-number*]

Activating an RRPP domain

Restrictions and guidelines

- Perform this task on all nodes in the RRPP domain.
- Before you activate an RRPP domain on the current device, enable the RRPP protocol and RRPP rings for the RRPP domain on the current device.
- Before you enable subrings on a device, you must enable the primary ring. Before you disable the primary ring on the device, you must disable all subrings. Otherwise, the system displays error prompts.
- To prevent Hello packets of subrings from being looped on the primary ring, enable the primary ring on its master node first. Then enable the subrings on their respective master nodes.

Procedure

1. Enter system view.

system-view

2. Enable RRPP.

   rrpp enable

   By default, RRPP is disabled.

3. Enter RRPP domain view.

   rrpp domain *domain-id*

4. Enable the specified RRPP ring.

   ring *ring-id* **enable**

   By default, an RRPP ring is disabled.

Configuring RRPP timers

Restrictions and guidelines for RRPP timer configuration

Perform this task on the master node of an RRPP domain.

Configuring the Hello timer and Fail timer

Restrictions and guidelines

- The Fail timer must be greater than or equal to three times the Hello timer.
- In a dual-homed-ring network, to avoid temporary loops when the primary ring fails, make sure the value of *A* is greater than the value of *B*, where:
  - *A* is the difference between the Fail timer values on the master node of the subring and the master node of the primary ring.
  - *B* is twice the Hello timer value on the master node of the subring.

Procedure

1. Enter system view.

   system-view

2. Enter RRPP domain view.

   rrpp domain *domain-id*

3. Set the Hello timer and Fail timer for the RRPP domain.

   timer hello-timer *hello-value* **fail-timer** *fail-value*

   By default, the Hello timer value is 1 second and the Fail timer value is 3 seconds.

Configuring the link-up delay timer

Restrictions and guidelines

If the **distribute** keyword is not specified, the link-up delay timer value cannot be greater than the Fail timer value minus twice the Hello timer value.

If you specify the **distribute** keyword in an RRPP network implementing load balancing, you must configure the link-up delay timer for each RRPP domain for the timer to take effect. If you set different timer values for different RRPP domains, the smallest timer value takes effect.

Procedure

1. Enter system view.

system-view

2. Enter RRPP domain view.

   rrpp domain *domain-id*

3. Set the link-up delay timer for the RRPP domain.

   linkup-delay-timer *delay-time* [**distribute**]

   By default, the link-up delay timer value is 0 seconds, and the **distribute** keyword is not specified.

Configuring an RRPP ring group

About this task

To reduce Edge-Hello traffic, assign subrings with the same edge node and assistant edge node to an RRPP ring group. An RRPP ring group must be configured on both the edge node and the assistant edge node. It can only be configured on these two types of nodes.

Restrictions and guidelines

- Perform this task on the edge node and assistant edge node in the RRPP domain.
- You can assign a subring to only one RRPP ring group. The RRPP ring groups configured on the edge node and the assistant edge node must contain the same subrings. Otherwise, the RRPP ring group cannot operate correctly.
- The subrings in an RRPP ring group must share the same edge node and assistant edge node. The edge node and the assistant edge node must have the same SRPTs.
- Make sure a device is either the edge node or the assistant edge node on the subrings in an RRPP ring group.
- Make sure the RRPP ring groups on the edge node and the assistant edge node have the same configurations and activation status.
- Make sure all subrings in an RRPP ring group have the same SRPTs. If the SRPTs of these subrings are different, the RRPP ring group cannot operate correctly.

Procedure

1. Enter system view.

   system-view

2. Create an RRPP ring group and enter RRPP ring group view.

   rrpp ring-group *ring-group-id*

   By default, no RRPP ring groups exist.

3. Assign the specified subrings to the RRPP ring group.

   domain *domain-id* **ring** *ring-id-list*

   By default, no subrings are assigned to an RRPP ring group.

Enabling SNMP notifications for RRPP

About this task

To report critical RRPP events to an NMS, enable SNMP notifications for RRPP. For RRPP event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for RRPP.
   snmp-agent trap enable rrpp [**major-fault** | **multi-master** | **ring-fail** | **ring-recover**] *
   By default, SNMP notifications for RRPP are disabled.

Display and maintenance commands for RRPP

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display brief RRPP information.	**display rrpp brief**
Display RRPP group configuration information.	**display rrpp ring-group** [*ring-group-id*]
Display RRPPDU statistics.	**display rrpp statistics domain** *domain-id* [**ring** *ring-id*]
Display detailed RRPP information.	**display rrpp verbose domain** *domain-id* [**ring** *ring-id*]
Clear RRPPDU statistics.	**reset rrpp statistics domain** *domain-id* [**ring** *ring-id*]

RRPP configuration examples

Example: Configuring a single ring

Network configuration

As shown in Figure 9:

- Device A, Device B, Device C, and Device D form RRPP domain 1. Specify the primary control VLAN of RRPP domain 1 as VLAN 4092. Specify the protected VLANs of RRPP domain 1 as VLANs 1 through 30.
- Device A, Device B, Device C, and Device D form primary ring 1.
- Specify Device A as the master node of primary ring 1, GigabitEthernet 1/0/1 as the primary port, and GigabitEthernet 1/0/2 as the secondary port.
- Specify Device B, Device C, and Device D as the transit nodes of primary ring 1. Specify GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port on Device B, Device C, and Device D.

Figure 9 Network diagram

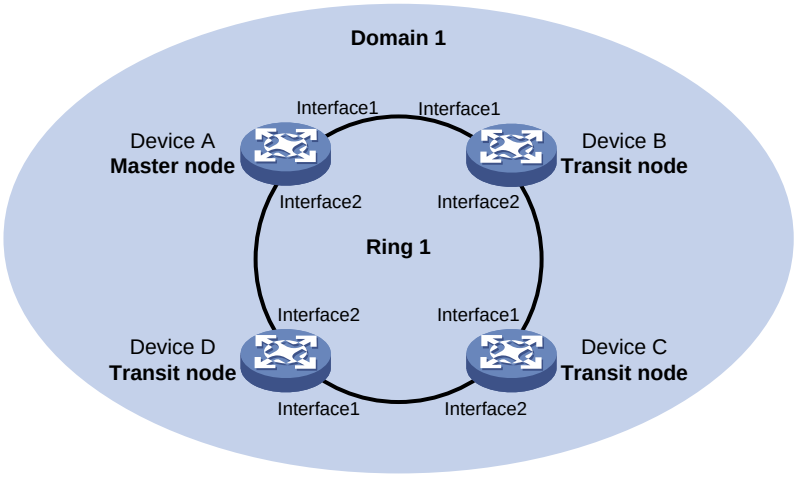


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

Create VLANs 1 through 30.

```
<DeviceA> system-view
[DeviceA] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceA] stp region-configuration
[DeviceA-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceA-mst-region] active region-configuration
[DeviceA-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] link-delay up 0
[DeviceA-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceA-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] link-delay up 0
[DeviceA-GigabitEthernet1/0/2] link-delay down 0
```

```
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceA-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceA] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceA-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceA-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceA-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceA-rrpp-domain1] ring 1 enable
[DeviceA-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceA] rrpp enable
```

2. Configure Device B:

Create VLANs 1 through 30.

```
<DeviceB> system-view
[DeviceB] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceB] stp region-configuration
[DeviceB-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceB-mst-region] active region-configuration
[DeviceB-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] link-delay up 0
[DeviceB-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] link-delay up 0
[DeviceB-GigabitEthernet1/0/2] link-delay down 0
[DeviceB-GigabitEthernet1/0/2] undo stp enable
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceB-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceB] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceB-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceB-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device B as the transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceB-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain1] ring 1 enable
[DeviceB-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceB] rrpp enable
```

3. Configure Device C:

Create VLANs 1 through 30.

```
<DeviceC> system-view
[DeviceC] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceC] stp region-configuration
[DeviceC-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] link-delay up 0
[DeviceC-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceC-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] link-delay up 0
[DeviceC-GigabitEthernet1/0/2] link-delay down 0
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceC-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceC] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceC-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceC-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device C as the transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceC-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceC-rrpp-domain1] ring 1 enable
[DeviceC-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceC] rrpp enable
```

4. Configure Device D:

Create VLANs 1 through 30.

```
<DeviceD> system-view
[DeviceD] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceD] stp region-configuration
[DeviceD-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] link-delay up 0
[DeviceD-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceD-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceD-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] link-delay up 0
[DeviceD-GigabitEthernet1/0/2] link-delay down 0
[DeviceD-GigabitEthernet1/0/2] undo stp enable
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceD-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceD] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceD-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceD-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device D as the transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceD-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceD-rrpp-domain1] ring 1 enable
[DeviceD-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceD] rrpp enable
```

Verifying the configuration

Use the **`display rrpp verbose`** command to display detailed RRPP information on the device, such as ring state and port state.

```
[DeviceA] display rrpp verbose domain 1
 Domain ID     : 1
 Control VLAN  : Primary 4092, Secondary 4093
 Protected VLAN: Reference instance 1
 Hello timer   : 1 sec, Fail timer: 3 sec
 Linkup-Delay timer: 0 sec

 Ring ID       : 1
 Ring level    : 0
 Node mode     : Master
 Ring state    : Complete
 Enable status : Yes, Active status: Yes
 Primary port  : GE1/0/1                   Port status: UP
 Secondary port: GE1/0/2                   Port status: BLOCKED
```

Example: Configuring intersecting rings

Network configuration

As shown in Figure 10:

- Device A, Device B, Device C, Device D, and Device E form RRPP domain 1. VLAN 4092 is the primary control VLAN of RRPP domain 1. RRPP domain 1 protects VLANs 1 through 30.
- Device A, Device B, Device C, and Device D form primary ring 1. Device B, Device C, and Device E form subring 2.
- Device A is the master node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 the secondary port.
- Device E is the master node of subring 2, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 the secondary port.
- Device B is the transit node of primary ring 1 and the edge node of subring 2, with GigabitEthernet 1/0/3 as the edge port.
- Device C is the transit node of primary ring 1 and the assistant edge node of subring 1, with GigabitEthernet 1/0/3 as the edge port.
- Device D is the transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 the secondary port.

Figure 10 Network diagram

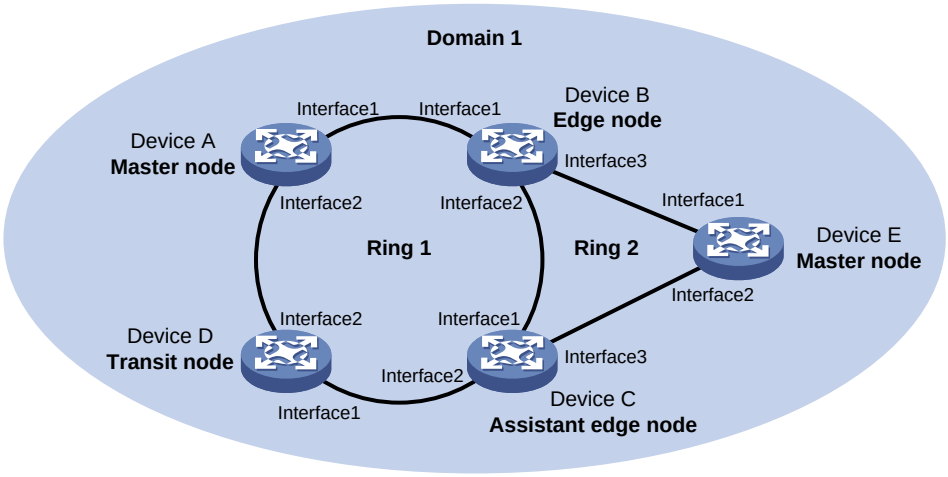


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Device A:

Create VLANs 1 through 30.

```
<DeviceA> system-view
[DeviceA] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceA] stp region-configuration
[DeviceA-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceA-mst-region] active region-configuration
[DeviceA-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] link-delay up 0
[DeviceA-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceA-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] link-delay up 0
```

```
[DeviceA-GigabitEthernet1/0/2] link-delay down 0
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceA-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceA] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceA-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceA-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceA-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceA-rrpp-domain1] ring 1 enable
[DeviceA-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceA] rrpp enable
```

2. Configure Device B:

Create VLANs 1 through 30.

```
<DeviceB> system-view
[DeviceB] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceB] stp region-configuration
[DeviceB-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceB-mst-region] active region-configuration
[DeviceB-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] link-delay up 0
[DeviceB-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] link-delay up 0
[DeviceB-GigabitEthernet1/0/2] link-delay down 0
[DeviceB-GigabitEthernet1/0/2] undo stp enable
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
```

```
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] link-delay up 0
[DeviceB-GigabitEthernet1/0/3] undo stp enable
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan 1 to 30
[DeviceB-GigabitEthernet1/0/3] quit
```

Create RRPP domain 1.

```
[DeviceB] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceB-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceB-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device B as a transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceB-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain1] ring 1 enable
```

Configure Device B as the edge node of subring 2, with GigabitEthernet 1/0/3 as the edge port. Enable ring 2.

```
[DeviceB-rrpp-domain1] ring 2 node-mode edge edge-port gigabitethernet 1/0/3
[DeviceB-rrpp-domain1] ring 2 enable
[DeviceB-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceB] rrpp enable
```

3. Configure Device C:

Create VLANs 1 through 30.

```
<DeviceC> system-view
[DeviceC] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceC] stp region-configuration
[DeviceC-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] link-delay up 0
[DeviceC-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceC-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] link-delay up 0
[DeviceC-GigabitEthernet1/0/2] link-delay down 0
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceC-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] link-delay up 0
[DeviceC-GigabitEthernet1/0/3] link-delay down 0
[DeviceC-GigabitEthernet1/0/3] undo stp enable
[DeviceC-GigabitEthernet1/0/3] port link-type trunk
[DeviceC-GigabitEthernet1/0/3] port trunk permit vlan 1 to 30
[DeviceC-GigabitEthernet1/0/3] quit
```

Create RRPP domain 1.

```
[DeviceC] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceC-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceC-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device C as a transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceC-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceC-rrpp-domain1] ring 1 enable
```

Configure Device C as the assistant edge node of subring 2, with GigabitEthernet 1/0/3 as the edge port. Enable ring 2.

```
[DeviceC-rrpp-domain1] ring 2 node-mode assistant-edge edge-port gigabitethernet
1/0/3
[DeviceC-rrpp-domain1] ring 2 enable
[DeviceC-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceC] rrpp enable
```

4. Configure Device D:

Create VLANs 1 through 30.

```
<DeviceD> system-view
[DeviceD] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceD] stp region-configuration
[DeviceD-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] link-delay up 0
```

```
[DeviceD-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceD-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceD-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] link-delay up 0
[DeviceD-GigabitEthernet1/0/2] link-delay down 0
[DeviceD-GigabitEthernet1/0/2] undo stp enable
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceD-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceD] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceD-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceD-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device D as the transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceD-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceD-rrpp-domain1] ring 1 enable
[DeviceD-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceD] rrpp enable
```

5. Configure Device E:

Create VLANs 1 through 30.

```
<DeviceE> system-view
[DeviceE] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceE] stp region-configuration
[DeviceE-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceE-mst-region] active region-configuration
[DeviceE-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] link-delay up 0
[DeviceE-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceE-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceE-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceE-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceE-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] link-delay up 0
[DeviceE-GigabitEthernet1/0/2] link-delay down 0
[DeviceE-GigabitEthernet1/0/2] undo stp enable
[DeviceE-GigabitEthernet1/0/2] port link-type trunk
[DeviceE-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceE-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceE] rrpp domain 1
```

Configure VLAN 4092 as the primary control VLAN of RRPP domain 1.

```
[DeviceE-rrpp-domain1] control-vlan 4092
```

Configure the VLANs mapped to MSTI 1 as the protected VLANs of RRPP domain 1.

```
[DeviceE-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device E as the master node of subring 2, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 2.

```
[DeviceE-rrpp-domain1] ring 2 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 1
[DeviceE-rrpp-domain1] ring 2 enable
[DeviceE-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceE] rrpp enable
```

Verifying the configuration

Use the **`display rrpp verbose`** command to display detailed RRPP information on the device, such as ring state and port state.

Example: Configuring load-balanced intersecting rings

Network configuration

As shown in Figure 11:

- Device A, Device B, Device C, Device D, and Device F form RRPP domain 1. VLAN 100 is the primary control VLAN of the RRPP domain. Device A is the master node of the primary ring, Ring 1. Device D is the transit node of Ring 1. Device F is the master node of the subring Ring 3. Device C is the edge node of the subring Ring 3. Device B is the assistant edge node of the subring Ring 3.
- Device A, Device B, Device C, Device D, and Device E form RRPP domain 2. VLAN 105 is the primary control VLAN of the RRPP domain. Device A is the master node of the primary ring, Ring 1. Device D is the transit node of Ring 1. Device E is the master node of the subring Ring 2. Device C is the edge node of the subring Ring 2. Device B is the assistant edge node of the subring Ring 2.
- Specify VLAN 11 as the protected VLAN of domain 1, and VLAN 12 the protected VLAN of domain 2. You can implement VLAN-based load balancing on Ring 1.
- Ring 2 and Ring 3 have the same edge node and assistant edge node, and the two subrings have the same SRPTs. You can add Ring 2 and Ring 3 to an RRPP ring group to reduce Edge-Hello traffic.

Figure 11 Network diagram

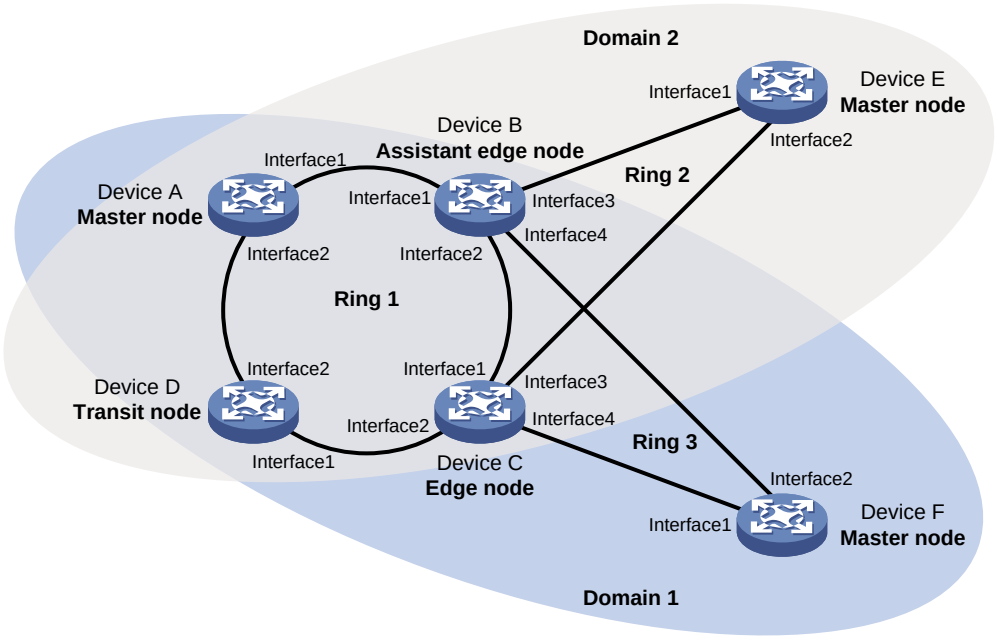


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure Device A:

   # Create VLANs 11 and 12.

   ```
   <DeviceA> system-view
   [DeviceA] vlan 11 to 12
   ```

   # Map VLAN 11 to MSTI 1 and VLAN 12 to MSTI 2.

   ```
   [DeviceA] stp region-configuration
   [DeviceA-mst-region] instance 1 vlan 11
   [DeviceA-mst-region] instance 2 vlan 12
   ```

   # Activate the MST region configuration.

   ```
   [DeviceA-mst-region] active region-configuration
   [DeviceA-mst-region] quit
   ```

   # Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] link-delay up 0
   [DeviceA-GigabitEthernet1/0/1] link-delay down 0
   ```

   # Disable the spanning tree feature on the port.

   ```
   [DeviceA-GigabitEthernet1/0/1] undo stp enable
   ```

   # Configure the port as a trunk port.

```
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLANs 11 and 12.

```
[DeviceA-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 11 12
```

Configure VLAN 11 as the default VLAN.

```
[DeviceA-GigabitEthernet1/0/1] port trunk pvid vlan 11
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] link-delay up 0
[DeviceA-GigabitEthernet1/0/2] link-delay down 0
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 11 12
[DeviceA-GigabitEthernet1/0/2] port trunk pvid vlan 11
[DeviceA-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceA] rrpp domain 1
```

Configure VLAN 100 as the primary control VLAN of RRPP domain 1.

```
[DeviceA-rrpp-domain1] control-vlan 100
```

Configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceA-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceA-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceA-rrpp-domain1] ring 1 enable
[DeviceA-rrpp-domain1] quit
```

Create RRPP domain 2.

```
[DeviceA] rrpp domain 2
```

Configure VLAN 105 as the primary control VLAN of RRPP domain 2.

```
[DeviceA-rrpp-domain2] control-vlan 105
```

Configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceA-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet 1/0/2 as the master port and GigabitEthernet 1/0/1 as the secondary port. Enable ring 1.

```
[DeviceA-rrpp-domain2] ring 1 node-mode master primary-port gigabitethernet 1/0/2
secondary-port gigabitethernet 1/0/1 level 0
[DeviceA-rrpp-domain2] ring 1 enable
[DeviceA-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceA] rrpp enable
```

2. Configure Device B:

Create VLANs 11 and 12.

```
<DeviceB> system-view
[DeviceB] vlan 11 to 12
```

Map VLAN 11 to MSTI 1 and VLAN 12 to MSTI 2.

```
[DeviceB] stp region-configuration
[DeviceB-mst-region] instance 1 vlan 11
[DeviceB-mst-region] instance 2 vlan 12
```

Activate the MST region configuration.

```
[DeviceB-mst-region] active region-configuration
[DeviceB-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] link-delay up 0
[DeviceB-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLANs 11 and 12.

```
[DeviceB-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 11 12
```

Configure VLAN 11 as the default VLAN.

```
[DeviceB-GigabitEthernet1/0/1] port trunk pvid vlan 11
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] link-delay up 0
[DeviceB-GigabitEthernet1/0/2] link-delay down 0
[DeviceB-GigabitEthernet1/0/2] undo stp enable
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 11 12
[DeviceB-GigabitEthernet1/0/2] port trunk pvid vlan 11
[DeviceB-GigabitEthernet1/0/2] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/3.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] link-delay up 0
[DeviceB-GigabitEthernet1/0/3] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/3] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLAN 12.

```
[DeviceB-GigabitEthernet1/0/3] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan 12
```

Configure VLAN 12 as the default VLAN.

```
[DeviceB-GigabitEthernet1/0/3] port trunk pvid vlan 12
[DeviceB-GigabitEthernet1/0/3] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/4.

```
[DeviceB] interface gigabitethernet 1/0/4
```

```
[DeviceB-GigabitEthernet1/0/4] link-delay up 0
[DeviceB-GigabitEthernet1/0/4] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/4] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceB-GigabitEthernet1/0/4] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLAN 11.

```
[DeviceB-GigabitEthernet1/0/4] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/4] port trunk permit vlan 11
```

Configure VLAN 11 as the default VLAN.

```
[DeviceB-GigabitEthernet1/0/4] port trunk pvid vlan 11
[DeviceB-GigabitEthernet1/0/4] quit
```

Create RRPP domain 1.

```
[DeviceB] rrpp domain 1
```

Configure VLAN 100 as the primary control VLAN of RRPP domain 1.

```
[DeviceB-rrpp-domain1] control-vlan 100
```

Configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceB-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device B as a transit node of primary ring 1 in RRPP domain 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceB-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain1] ring 1 enable
```

Configure Device B as the assistant edge node of subring 3 in RRPP domain 1, with GigabitEthernet 1/0/4 as the edge port. Enable subring 3.

```
[DeviceB-rrpp-domain1] ring 3 node-mode assistant-edge edge-port gigabitethernet
1/0/4
[DeviceB-rrpp-domain1] ring 3 enable
[DeviceB-rrpp-domain1] quit
```

Create RRPP domain 2.

```
[DeviceB] rrpp domain 2
```

Configure VLAN 105 as the primary control VLAN of RRPP domain 2.

```
[DeviceB-rrpp-domain2] control-vlan 105
```

Configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceB-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device B as the transit node of primary ring 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceB-rrpp-domain2] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain2] ring 1 enable
```

Configure Device B as the assistant edge node of subring 2 in RRPP domain 2, with GigabitEthernet 1/0/3 as the edge port. Enable subring 2.

```
[DeviceB-rrpp-domain2] ring 2 node-mode assistant-edge edge-port gigabitethernet
1/0/3
[DeviceB-rrpp-domain2] ring 2 enable
[DeviceC-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceB] rrpp enable
```

3. Configure Device C:

Create VLANs 11 and 12.

```
<DeviceC> system-view
[DeviceC] vlan 11 to 12
```

Map VLAN 11 to MSTI 1 and VLAN 12 to MSTI 2.

```
[DeviceC] stp region-configuration
[DeviceC-mst-region] instance 1 vlan 11
[DeviceC-mst-region] instance 2 vlan 12
```

Activate the MST region configuration.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] link-delay up 0
[DeviceC-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceC-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLANs 11 and 12.

```
[DeviceC-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 11 12
```

Configure VLAN 11 as the default VLAN.

```
[DeviceC-GigabitEthernet1/0/1] port trunk pvid vlan 11
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] link-delay up 0
[DeviceC-GigabitEthernet1/0/2] link-delay down 0
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 11 12
[DeviceC-GigabitEthernet1/0/2] port trunk pvid vlan 11
[DeviceC-GigabitEthernet1/0/2] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/3.

```
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] link-delay up 0
[DeviceC-GigabitEthernet1/0/3] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceC-GigabitEthernet1/0/3] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceC-GigabitEthernet1/0/3] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLAN 12.

```
[DeviceC-GigabitEthernet1/0/3] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/3] port trunk permit vlan 12
```

Configure VLAN 12 as the default VLAN.

```
[DeviceC-GigabitEthernet1/0/3] port trunk pvid vlan 12
[DeviceC-GigabitEthernet1/0/3] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/4.

```
[DeviceC] interface gigabitethernet 1/0/4
[DeviceC-GigabitEthernet1/0/4] link-delay up 0
[DeviceC-GigabitEthernet1/0/4] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceC-GigabitEthernet1/0/4] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceC-GigabitEthernet1/0/4] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLAN 11.

```
[DeviceC-GigabitEthernet1/0/4] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/4] port trunk permit vlan 11
```

Configure VLAN 11 as the default VLAN.

```
[DeviceC-GigabitEthernet1/0/4] port trunk pvid vlan 11
[DeviceC-GigabitEthernet1/0/4] quit
```

Create RRPP domain 1.

```
[DeviceC] rrpp domain 1
```

Configure VLAN 100 as the primary control VLAN of RRPP domain 1.

```
[DeviceC-rrpp-domain1] control-vlan 100
```

Configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceC-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device C as the transit node of primary ring 1 in RRPP domain 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceC-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceC-rrpp-domain1] ring 1 enable
```

Configure Device C as the edge node of subring 3 in RRPP domain 1, with GigabitEthernet 1/0/4 as the edge port. Enable subring 3.

```
[DeviceC-rrpp-domain1] ring 3 node-mode edge edge-port gigabitethernet 1/0/4
[DeviceC-rrpp-domain1] ring 3 enable
[DeviceC-rrpp-domain1] quit
```

Create RRPP domain 2.

```
[DeviceC] rrpp domain 2
```

Configure VLAN 105 as the primary control VLAN of RRPP domain 2.

```
[DeviceC-rrpp-domain2] control-vlan 105
```

Configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceC-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device C as the transit node of primary ring 1 in RRPP domain 2, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceC-rrpp-domain2] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceC-rrpp-domain2] ring 1 enable
```

Configure Device C as the edge node of subring 2 in RRPP domain 2, with GigabitEthernet 1/0/3 as the edge port. Enable subring 2.

```
[DeviceC-rrpp-domain2] ring 2 node-mode edge edge-port gigabitethernet 1/0/3
[DeviceC-rrpp-domain2] ring 2 enable
[DeviceC-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceC] rrpp enable
```

4. Configure Device D:

Create VLANs 11 and 12.

```
<DeviceD> system-view
[DeviceD] vlan 11 to 12
```

Map VLAN 11 to MSTI 1 and VLAN 12 to MSTI 2.

```
[DeviceD] stp region-configuration
[DeviceD-mst-region] instance 1 vlan 11
[DeviceD-mst-region] instance 2 vlan 12
```

Activate the MST region configuration.

```
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] link-delay up 0
[DeviceD-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceD-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLANs 11 and 12.

```
[DeviceD-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 11 12
```

Configure VLAN 11 as the default VLAN.

```
[DeviceD-GigabitEthernet1/0/1] port trunk pvid vlan 11
[DeviceD-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] link-delay up 0
[DeviceD-GigabitEthernet1/0/2] link-delay down 0
[DeviceD-GigabitEthernet1/0/2] undo stp enable
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 11 12
[DeviceD-GigabitEthernet1/0/2] port trunk pvid vlan 11
[DeviceD-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceD] rrpp domain 1
```

Configure VLAN 100 as the primary control VLAN of RRPP domain 1.

```
[DeviceD-rrpp-domain1] control-vlan 100
```

Configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceD-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device D as the transit node of primary ring 1 in RRPP domain 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceD-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceD-rrpp-domain1] ring 1 enable
[DeviceD-rrpp-domain1] quit
```

Create RRPP domain 2.

```
[DeviceD] rrpp domain 2
```

Configure VLAN 105 as the primary control VLAN of RRPP domain 2.

```
[DeviceD-rrpp-domain2] control-vlan 105
```

Configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceD-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device D as the transit node of primary ring 1 in RRPP domain 2, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable ring 1.

```
[DeviceD-rrpp-domain2] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceD-rrpp-domain2] ring 1 enable
[DeviceD-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceD] rrpp enable
```

5. Configure Device E:

Create VLAN 12.

```
<DeviceE> system-view
[DeviceE] vlan 12
```

Map VLAN 12 to MSTI 2.

```
[DeviceE-vlan12] quit
[DeviceE] stp region-configuration
[DeviceE-mst-region] instance 2 vlan 12
```

Activate the MST region configuration.

```
[DeviceE-mst-region] active region-configuration
[DeviceE-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] link-delay up 0
[DeviceE-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceE-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceE-GigabitEthernet1/0/1] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLAN 12.

```
[DeviceE-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceE-GigabitEthernet1/0/1] port trunk permit vlan 12
```

Configure VLAN 12 as the default VLAN.

```
[DeviceE-GigabitEthernet1/0/1] port trunk pvid vlan 12
[DeviceE-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] link-delay up 0
[DeviceE-GigabitEthernet1/0/2] link-delay down 0
[DeviceE-GigabitEthernet1/0/2] undo stp enable
[DeviceE-GigabitEthernet1/0/2] port link-type trunk
[DeviceE-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceE-GigabitEthernet1/0/2] port trunk permit vlan 12
[DeviceE-GigabitEthernet1/0/2] port trunk pvid vlan 12
[DeviceE-GigabitEthernet1/0/2] quit
```

Create RRPP domain 2.

```
[DeviceE] rrpp domain 2
```

Configure VLAN 105 as the primary control VLAN of RRPP domain 2.

```
[DeviceE-rrpp-domain2] control-vlan 105
```

Configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceE-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device E as the master mode of subring 2 in RRPP domain 2, with GigabitEthernet 1/0/2 as the primary port and GigabitEthernet 1/0/1 as the secondary port. Enable ring 2.

```
[DeviceE-rrpp-domain2] ring 2 node-mode master primary-port gigabitethernet 1/0/2
secondary-port gigabitethernet 1/0/1 level 1
[DeviceE-rrpp-domain2] ring 2 enable
[DeviceE-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceE] rrpp enable
```

6. Configure Device F:

Create VLAN 11.

```
<DeviceF> system-view
[DeviceF] vlan 11
[DeviceF-vlan11] quit
```

Map VLAN 11 to MSTI 1.

```
[DeviceF] stp region-configuration
[DeviceF-mst-region] instance 1 vlan 11
```

Activate the MST region configuration.

```
[DeviceF-mst-region] active region-configuration
[DeviceF-mst-region] quit
```

Set the physical state change suppression interval to 0 seconds on GigabitEthernet 1/0/1.

```
[DeviceF] interface gigabitethernet 1/0/1
[DeviceF-GigabitEthernet1/0/1] link-delay up 0
[DeviceF-GigabitEthernet1/0/1] link-delay down 0
```

Disable the spanning tree feature on the port.

```
[DeviceF-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceF-GigabitEthernet1/0/1] port link-type trunk
```

Remove the port from VLAN 1, and assign it to VLAN 11.

```
[DeviceF-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceF-GigabitEthernet1/0/1] port trunk permit vlan 11
```

Configure VLAN 11 as the default VLAN.

```
[DeviceF-GigabitEthernet1/0/1] port trunk pvid vlan 11
```

```
[DeviceF-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceF] interface gigabitethernet 1/0/2
[DeviceF-GigabitEthernet1/0/2] link-delay up 0
[DeviceF-GigabitEthernet1/0/2] link-delay down 0
[DeviceF-GigabitEthernet1/0/2] undo stp enable
[DeviceF-GigabitEthernet1/0/2] port link-type trunk
[DeviceF-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceF-GigabitEthernet1/0/2] port trunk permit vlan 11
[DeviceF-GigabitEthernet1/0/2] port trunk pvid vlan 11
[DeviceF-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1.

```
[DeviceF] rrpp domain 1
```

Configure VLAN 100 as the primary control VLAN of RRPP domain 1.

```
[DeviceF-rrpp-domain1] control-vlan 100
```

Configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceF-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device F as the master node of subring 3 in RRPP domain 1, with GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port. Enable subring 3.

```
[DeviceF-rrpp-domain1] ring 3 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 1
[DeviceF-rrpp-domain1] ring 3 enable
[DeviceF-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceF] rrpp enable
```

7. Configure RRPP ring group settings on Device B and Device C:

Create RRPP ring group 1 on Device B, and add subrings 2 and 3 to the RRPP ring group.

```
[DeviceB] rrpp ring-group 1
[DeviceB-rrpp-ring-group1] domain 2 ring 2
[DeviceB-rrpp-ring-group1] domain 1 ring 3
```

Create RRPP ring group 1 on Device C, and add subrings 2 and 3 to the RRPP ring group.

```
[DeviceC] rrpp ring-group 1
[DeviceC-rrpp-ring-group1] domain 2 ring 2
[DeviceC-rrpp-ring-group1] domain 1 ring 3
```

Verifying the configuration

Use the **`display rrpp verbose`** command to display detailed RRPP information on the device, such as ring state and port state.

Troubleshooting RRPP

The primary node cannot receive Hello packets when the link state is normal

Symptom

When the link state is normal, the master node cannot receive Hello packets, and it unblocks the secondary port.

Analysis

The following are possible causes:

- One or more nodes on the RRPP ring are not enabled with RRPP.
- The domain IDs or control VLANs are not the same for all nodes on the RRPP ring.
- The port on the RRPP ring does not work correctly.

Solution

To resolve the problem:

1. Use the **display rrpp brief** command to determine whether RRPP is enabled for all nodes. If it is not, use the **rrpp enable** command and the **ring enable** command to enable RRPP and RRPP rings for all nodes.
2. Use the **display rrpp brief** command to determine whether the domain ID and primary control VLAN ID are the same for all nodes. If they are not, set the same domain ID and primary control VLAN ID for the nodes.
3. Use the **display rrpp verbose** command to examine the link state of each port in each ring.
4. Use the **debugging rrpp** command on each node to determine whether a port receives or transmits Hello packets. If it does not, Hello packets are lost.
5. If the problem persists, contact H3C Support.

Contents

Configuring ERPS 1

About ERPS 1
ERPS structure 1
Instances 2
ERPS protocol packets 2
ERPS node states 3
ERPS timers 3
ERPS operation mechanism 4
ERPS network diagrams 5
Protocols and standards 9
Restrictions and guidelines: ERPS configuration 9
ERPS tasks at a glance 9
Prerequisites 10
Enabling ERPS globally 10
Configuring an ERPS ring 10
Creating an ERPS ring 10
Configuring ERPS ring member ports 11
Configuring control VLANs 12
Configuring protected VLANs 12
Configuring the node role 13
Enabling ERPS for an instance 13
Enabling R-APS packets to carry the ring ID in the destination MAC address 14
Configuring R-APS packet levels 14
Setting ERPS timers 14
Setting the non-revertive mode 15
Setting a switchover mode 15
Associating a ring with a subring 16
Enabling flush packet transparent transmission 16
Removing the MS mode and FS mode settings for an ERPS ring 17
Displaying and maintaining ERPS 17
Troubleshooting ERPS 17
The owner node cannot receive SF packets from a faulty node when the link state is normal 17

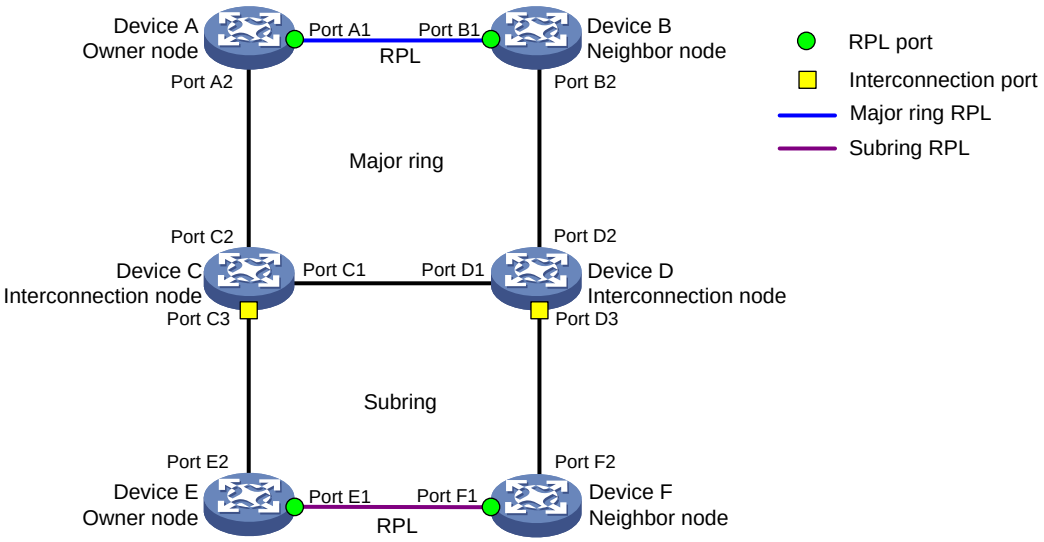
Configuring ERPS

About ERPS

Ethernet Ring Protection Switching (ERPS) is a robust link layer protocol that ensures a loop-free topology and implements quick link recovery.

ERPS structure

Figure 1 ERPS ring structure



Rings

ERPS rings can be divided into major rings and subrings. An ERPS network consists of one major ring or multiple major rings, and multiple subrings. By default, a ring is a major ring. You can configure a ring as a subring manually.

As shown in the ERPS ring structure diagram, a major ring is a closed ring formed by Device A, Device B, Device C, and Device D. A subring is an open ring formed by the link Device C<—>Device E<—>Device F<—>Device D.

RPL

An ERPS ring is composed of many nodes. Some nodes use ring protection links (RPLs) to prevent loops on the ERPS ring. As shown in the ERPS ring structure diagram, the link between Device A and Device B or the link between Device E and Device F is an RPL.

Nodes

ERPS nodes include owner nodes, neighbor nodes, interconnection nodes, and normal nodes.

- The owner node and neighbor node block and unblock ports on the RPL to prevent loops and switch traffic. An RPL connects an owner node and a neighbor node.
- Interconnection nodes connect different rings. Interconnection nodes reside on subrings and forward service packets but not protocol packets.
- Normal nodes forward both service packets and protocol packets.

As shown in the ERPS ring structure diagram, on the major ring, Device A is the owner node and Device B is the neighbor node. On the subring, Device E is the owner node and Device F is the neighbor node. Devices C and D are interconnection nodes.

Ports

Each node consists of two ERPS ring member ports: Port 0 and port 1. ERPS ring member ports have the following types:

- **RPL port**—Port on an RPL link.
- **Interconnection port**—Port that connects a subring to a major ring.
- **Normal port**—Default type of a port that forwards both service packets and protocol packets.

As shown in the ERPS ring structure diagram, ports A1, B1, E1, and F1 are RPL ports. Ports C3 and D3 are interconnection ports. Other ports are normal ports.

Instances

An ERPS ring supports multiple ERPS instances. An ERPS instance is a logical ring to process service and protocol packets. Each ERPS instance has its own owner node and maintains its own state and data. An ERPS instance is uniquely identified by the ring ID and VLAN ID of ERPS packets. The ring ID indicates the ring of ERPS packets. It can be represented by the last byte in the destination MAC address of the packets. The VLAN ID indicates the ERPS instance of the packets.

ERPS protocol packets

ERPS protocol packets are Ring Automatic Protection Switching (R-APS) packets. You can configure the R-APS packet level. A node does not process R-APS packets whose levels are greater than the level of the packets sent by the node. On a ring, the levels of R-APS packets must be the same for all nodes in an ERPS instance.

Table 1 R-APS packet types and functions

Packet type	Function
No request, RPL block (NR-RB)	When the link is stable, an owner node in idle state periodically sends NR-RB packets to inform other nodes that the RPL ports are blocked. The nodes that receive the NR-RB packets unblock available ports and update MAC address entries.
No request (NR)	After the link fault is cleared, the node that detects the recovery periodically sends NR packets. When the owner node receives the NR packets, it starts the WTR timer. The node stops sending NR packets after receiving NR-RB packets from the owner node.
Signal fail (SF)	When a link fails to send or receive signals, the node that detects the fault periodically sends SF packets. When the owner node and neighbor node receive the FS packets, they unblock the RPL ports. The node stops sending SF packets after the fault is cleared.
Manual switch (MS)	A port configured with the MS mode is blocked and periodically sends MS packets. When other nodes receive the MS packets, they unblock available ports and update MAC address entries.
Forced switch (FS)	A port configured with the FS mode is blocked and periodically sends FS packets. When other nodes receive the FS packets, they unblock all ports and update MAC address entries.
Flush	If the topology of a subring changes, the interconnection ports on the subring broadcasts flush packets. All nodes that receive the flush packets update MAC address entries.

NOTE:

- Typically R-APS packets are transmitted within a ring. The flush packets sourced from the subring can be forwarded to the major ring.
- Service packets can be transmitted between different rings.

ERPS node states

Table 2 ERPS states

State	Description
Init	State for a non-interconnection node that has less than two ERPS ring member ports or for an interconnection node that does not have ERPS ring member ports.
Idle	Stable state when all non-RPL links are available. In this state, the owner node blocks the RPL port and periodically sends NR-RB packets. The neighbor node blocks the RPL port. All nodes enter the idle state after the owner node enters the idle state.
Protection	State when a non-RPL link is faulty. In this state, the RPL link is unblocked to forward traffic. All nodes enter the protection state after a node enters the protection state.
MS	State when traffic paths are manually switched. All nodes enter the MS state after a node is configured with the MS mode.
FS	State when traffic paths are forcibly switched. All nodes enter the FS state after a node is configured with the FS mode.
Pending	Transient state between the previous states.

ERPS timers

Hold-off timer

The hold-off timer starts when the port detects a link fault. The port reports the link fault if the fault persists when the timer expires.

This timer delays the fault report time and affects the link switching performance.

Guard timer

The guard timer starts when the port detects a link recovery. The port does not process R-APS packets before the timer expires.

This timer prevents R-APS packets from impacting the network and affects the link switching performance when multiple points of failures exist.

WTR timer

In revertive mode, the WTR timer starts when the owner node in protection state receives NR packets. The RPL is unblocked and the recovered node is blocked before the timer expires. The owner node blocks the RPL and sends NR-RB packets when the timer expires. If the port receives SF packets before the timer expires, the timer stops and the RPL remains unblocked.

This timer prevents intermittent link failures from impacting the network.

WTB timer

In revertive mode, the WTB timer starts when the owner node in MS or FS state receives NR packets. The RPL is unblocked and the recovered node sends NR packets before the timer expires. The owner node blocks the RPL and sends NR-RB packets when the timer expires. If the port receives SF packets before the timer expires, the timer stops and the RPL remains unblocked.

This timer prevents the RPL ports from being blocked and unblocked frequently.

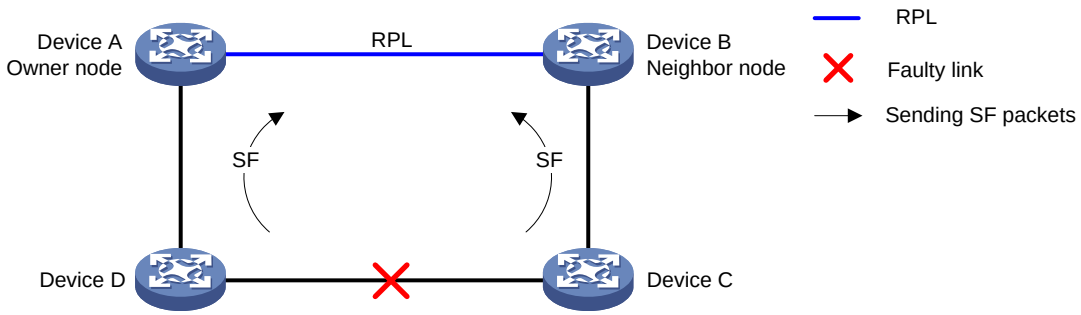
ERPS operation mechanism

ERPS uses the detection mechanism defined in ITU-T G.8032/Y.1344 to locate the point of failure and identify unidirectional or bidirectional faults.

ERPS uses the SF packets to report signal failures on a link and the NR packets to report link recovery. When a node detects a link status change, the node sends three packets first and then sends subsequent packets every five seconds.

Link-down report mechanism

Figure 2 Link-down report mechanism



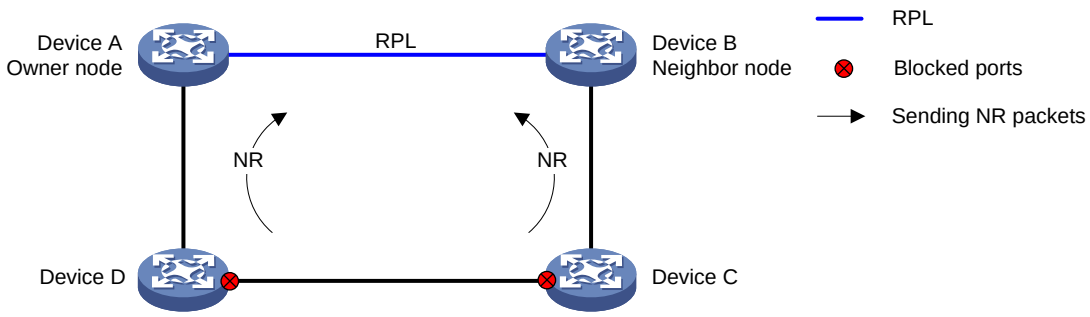
As shown in Figure 2, the link-down report mechanism uses the following process:

1. Device C and Device D detect the link failure and perform the following operations:
   a. Block the ports on both side of the faulty link.
   b. Periodically send SF packets to other nodes.
2. Device A and Device B receive the SF packets and perform the following operations:
   a. Unblock RPL ports.
   b. Update the MAC address entries.

Service packets are switched to the RPL link.

Link recovery mechanism

Figure 3 Link recovery mechanism



As shown in Figure 3, the link recovery mechanism uses the following process:

1. Device C and Device D detect the link recovery and perform the following operations:
   a. Block the recovered ports.
   b. Start the guard timer.
   c. Send NR packets.
2. When Device A (owner node) receives the NR packets, it does not perform any operations if it is in non-revertive mode. If Device A is in revertive mode, it performs the following operations:

a. Starts the WTR timer.

b. Blocks the RPL port and periodically sends NR-RB packets when the WTR timer expires.

3. When other nodes receive the NR-RB packets, they perform the following operations:

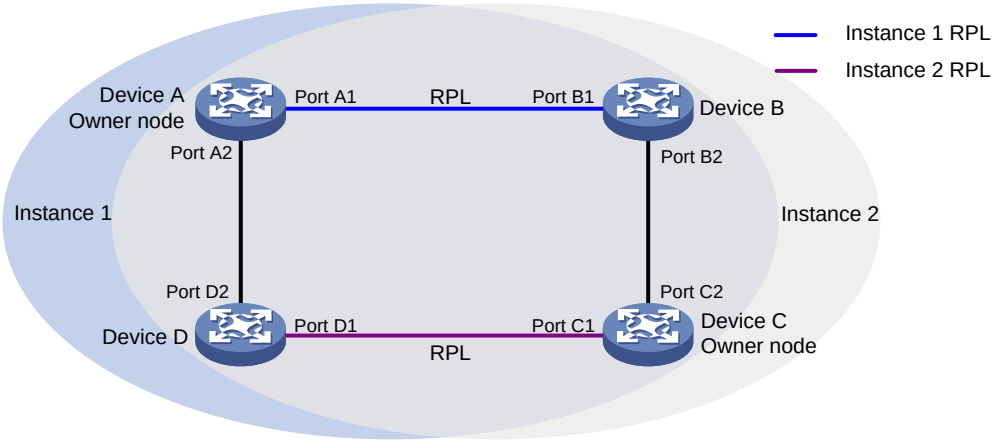
a. Device B (neighbor port) blocks the RPL port.

b. Device C and Device D unblock the recovered ports.

Service packets are switched to the recovered link.

Multi-instance load balancing mechanism

Figure 4 Multi-instance load balancing mechanism



An ERPS ring topology might carry traffic from multiple VLANs. Traffic from different VLANs can be load balanced among different ERPS instances.

ERPS uses the following types of VLANs:

- **Control VLAN**—Carries ERPS protocol packets. Each ERPS instance has its own control VLAN.
- **Protected VLAN**—Carries data packets. Each ERPS instance has its own protected VLAN. Protected VLANs are configured by using the mappings between VLANs and MSTIs.

As shown in Figure 4, the ERPS ring is configured with instance 1 and instance 2. For instance 1, the owner node is Device A, and the RPL is the link between Device A and Device B. For instance 2, the owner node is Device C, and the RPL is the link between Device C and Device D. Traffic from different VLANs can be load balanced among different links.

Manual configuration mechanism

ERPS supports the following manual configuration modes:

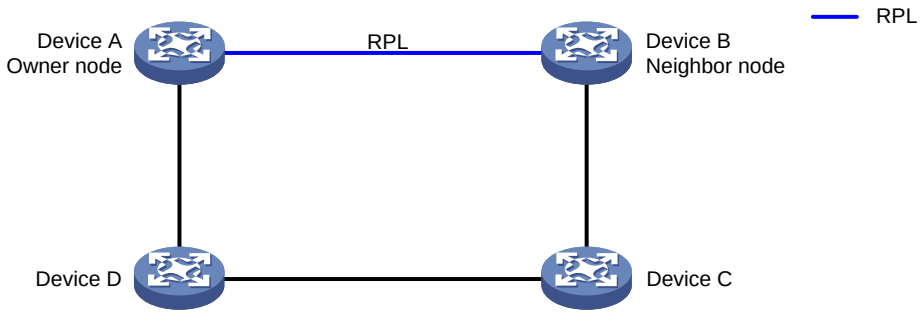
- **MS**—Use the **`erps switch manual`** command to block an ERPS ring member port. A port in MS mode is blocked and sends MS packets. The nodes that receive the MS packets unblock available ports. If the nodes in MS mode receive an SF packet, they unblock the blocked ports.
- **FS**—Use the **`erps switch force ring`** command to block an ERPS ring member port. A port in FS mode is blocked and sends FS packets. The nodes that receive the FS packets unblock available ports. If the nodes in FS mode receive an SF packet, they do not unblock the blocked ports.

ERPS network diagrams

One major ring

The network has one major ring.

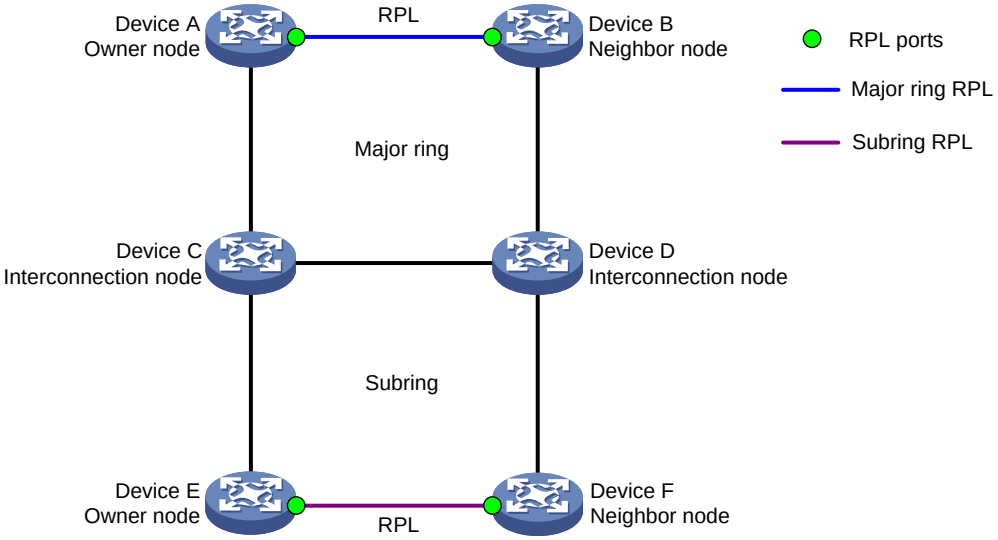
Figure 5 Network diagram



One major ring connecting one subring

The network has one major ring and one subring.

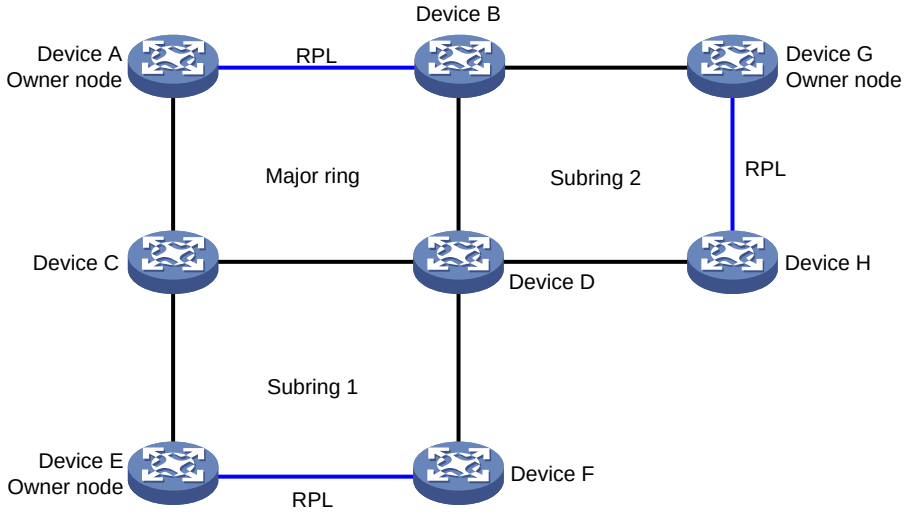
Figure 6 Network diagram



One major ring connecting multiple subrings

The network has three or more rings. Each subring is connected to the major ring by two interconnection nodes.

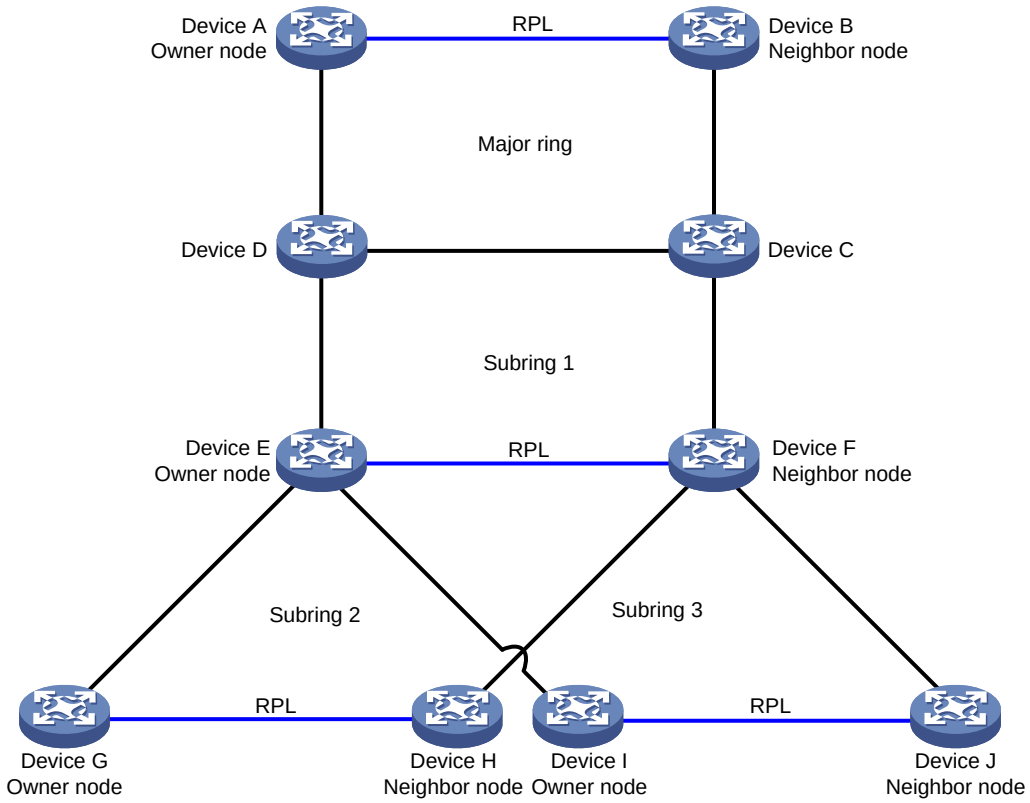
Figure 7 Network diagram



One subring connecting multiple subrings

The network has three or more rings. As shown in Figure 8, subring 1 is connected to the major ring. Other subrings are connected to subring 1 by two interconnection nodes.

Figure 8 Network diagram



One subring connecting multiple rings

The network has three or more rings. A minimum of one subring is connected to two rings. As shown in Figure 9, one interconnection node on subring 2 is connected to the major ring; and another interconnection node is connected to subring 1. As shown in Figure 10, subring 3 is connected to subring 1 and subring 2.

Figure 9 Network diagram 1

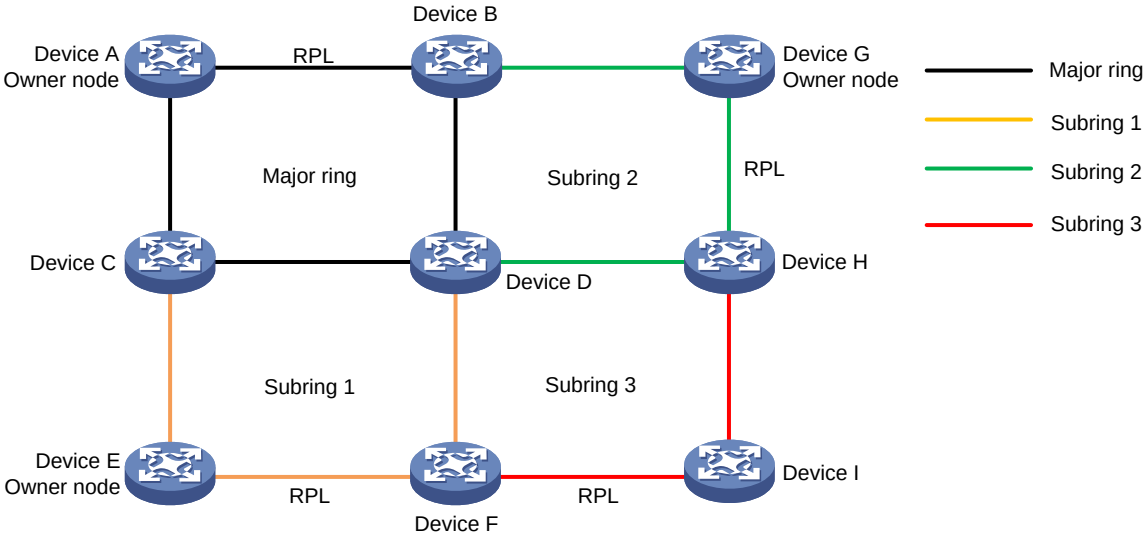
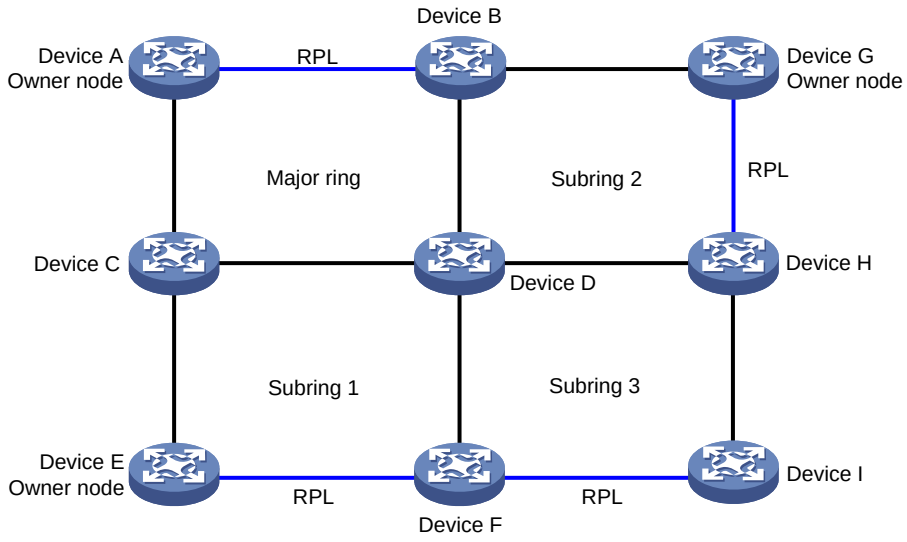


Figure 10 Network diagram 2

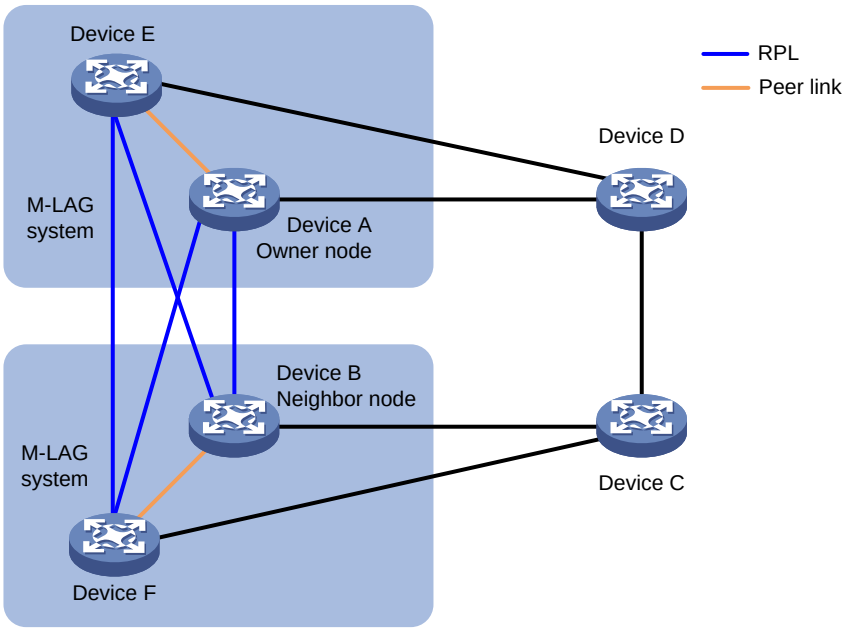


ERPS with M-LAG

M-LAG allows you to virtualize the devices in an ERPS ring into one system through multichassis link, thereby achieving device-level redundancy and load balancing.

You can use the **`display erps detail ring`** command to determine whether ERPS works correctly in an M-LAG network. For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*.

Figure 11 ERPS with M-LAG



Protocols and standards

- ITU-T G.8032, *Recommendation ITU-T G.8032/Y.1344, Ethernet ring protection switching*
- IEEE 802.1D, *IEEE Std 802.1D™-2004, IEEE Standard for Local and Metropolitan Area Networks—Media Access Control (MAC) Bridges*
- IEEE 802.3, *IEEE Std 802.3-2008, IEEE Standard for Information technology*

Restrictions and guidelines: ERPS configuration

ERPS does not provide an election mechanism. To implement ring detection and protection, configure all nodes correctly.

ERPS tasks at a glance

To configure ERPS, perform the following tasks:

1. Enabling ERPS globally

   Perform this task on devices you want to configure as ERPS nodes.
2. Configuring an ERPS ring

   Perform this task on all nodes on an ERPS ring.

   a. Creating an ERPS ring
   b. Configuring ERPS ring member ports
   c. Configuring control VLANs
   d. Configuring protected VLANs
   e. Configuring the node role
3. Enabling ERPS for an instance

   Perform this task on all nodes on an ERPS ring.

4. (Optional.) Enabling R-APS packets to carry the ring ID in the destination MAC address
   Perform this task on all nodes on an ERPS ring.
5. (Optional.) Configuring R-APS packet levels
6. (Optional.) Setting ERPS timers
   Perform this task on the owner node on an ERPS ring.
7. (Optional.) Setting the non-revertive mode
   Perform this task on the owner node on an ERPS ring.
8. (Optional.) Setting a switchover mode
   Perform this task on the nodes that you want to block their ports.
9. (Optional.) Removing the MS mode and FS mode settings for an ERPS ring

Prerequisites

Before you configure ERPS, complete the following tasks:

- Establish the Ethernet ring topology.
- Determine the ERPS rings, ERPS instances, control VLANs, protected VLANs, and node roles.

Enabling ERPS globally

Restrictions and guidelines

- Perform this task on devices you want to configure as ERPS nodes.
- For ERPS to take effect for an instance, enable it globally first.

Procedure

1. Enter system view.
   `system-view`
2. Enable ERPS globally.
   `erps enable`
   By default, ERPS is disabled globally.

Configuring an ERPS ring

Creating an ERPS ring

Restrictions and guidelines

- Perform this task on all nodes on an ERPS ring.
- A ring ID uniquely identifies an ERPS ring. All nodes on an ERPS ring must be configured with the same ring ID.

Procedure

1. Enter system view.
   `system-view`
2. Create an ERPS ring.
   `erps ring` *`ring-id`*
3. (Optional.) Configure the ring type.

ring-type sub-ring

By default, an ERPS ring is a major ring.

Configuring ERPS ring member ports

Restrictions and guidelines

- Perform this task on each node's ports intended for accessing ERPS rings.
- ERPS ring member ports automatically allow packets from the control VLAN to pass through.
- For faster topology convergence, use the **link-delay** command on ERPS ring member ports to set the physical state change suppression interval to 0 seconds. For more information about the **link-delay** command, see *Layer 2—LAN Switching Command Reference*.
- You must configure ERPS ring member ports as trunk ports.
- Do not assign an interface to both an aggregation group and an ERPS ring. If you do so, the interface does not take effect on the ERPS ring and cannot be displayed by using the **display erps detail** command.
- Do not assign a peer-link interface to an ERPS ring. A peer-link interface does not work for an ERPS ring and cannot be displayed by using the **display erps detail** command.
- Do not assign both an M-LAG interface and a non-M-LAG interface on a non-interconnection node to an ERPS ring. If you do so, the M-LAG interface does not take effect on the ERPS ring.

Configuring ERPS ring member port attributes

1. Enter system view.

   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.

   interface *interface-type interface-number*
3. Configure the port as a trunk port.

   port link-type trunk

   By default, a port is an access port.

   For more information about this command, see *Layer 2—LAN Switching Command Reference*.
4. Assign the trunk port to protected VLANs.

   port trunk permit vlan { *vlan-id-list* | **all** }

   By default, a trunk port is assigned only to VLAN 1.

   For more information about this command, see *Layer 2—LAN Switching Command Reference*.
5. Disable the spanning tree feature.

   undo stp enable

   By default, the spanning tree feature is enabled.

   For more information about this command, see *Layer 2—LAN Switching Command Reference*.

Configuring an ERPS ring member port

1. Enter system view.

   system-view
2. Enter ERPS ring view.

   erps ring *ring-id*
3. Configure an ERPS ring member port.

   { **port0** | **port1** } **interface** *interface-type interface-number*

   By default, an ERPS ring does not have ERPS ring member ports.

Configuring control VLANs

Restrictions and guidelines

- Perform this task on all nodes on an ERPS ring.
- The control VLAN must be a VLAN that has not been created on the device.
- Configure the same control VLAN for all nodes in an ERPS instance.
- Do not configure the default VLAN of an ERPS ring member port as the control VLAN.
- Do not enable QinQ or VLAN mapping on control VLANs. If you do, ERPS packets cannot be correctly forwarded and received.
- Make sure the ERPS instance has been configured. After the ERPS instance is enabled, the control VLAN cannot be changed.
- For a device not configured with ERPS to transparently transmit ERPS packets, make sure only the two ports accessing the ERPS ring permit packets from the control VLAN. If other ports on the device permit packets from the control VLAN, the packets from other VLANs might enter the control VLAN and strike the ERPS ring.
- When configuring ERPS with M-LAG, make sure the peer-link interface is a trunk port. The peer-link interface will automatically join or leave a control VLAN as the control VLAN is created or deleted.

Procedure

1. Enter system view.
   `system-view`
2. Enter ERPS ring view.
   `erps ring` *`ring-id`*
3. Enable ERPS instance view.
   `instance` *`instance-id`*
4. Configure a control VLAN.
   `control-vlan` *`vlan-id`*

Configuring protected VLANs

Restrictions and guidelines

- Perform this task on all nodes on an ERPS ring.
- Configure the same protected VLAN for all nodes of an ERPS instance. To implement load balancing, configure different protected VLANs for different ERPS instances.

Prerequisites

Before you configure protected VLANs, you must configure an MST region and the VLAN-to-instance mapping table. For more information about MST regions, see spanning tree configuration in *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Enter ERPS ring view.
   `erps ring` *`ring-id`*
3. Enable ERPS instance view.
   `instance` *`instance-id`*

4. Configure the protected VLANs.
   protected-vlan reference-instance *instance-id-list*

Configuring the node role

Restrictions and guidelines

- Perform this task on all nodes on an ERPS ring.
- For the owner node to work correctly, you must configure only one owner node for an ERPS ring.
- You can only configure the interconnection node for subrings.

Procedure

1. Enter system view.
   system-view
2. Enter ERPS ring view.
   erps ring *ring-id*
3. Enter ERPS instance view.
   instance *instance-id*
4. Configure the node role.
   node-role { { **owner** | **neighbor** } **rpl** | **interconnection** } { **port0** | **port1** }
   By default, a node is a normal node.

Enabling ERPS for an instance

Restrictions and guidelines

- Perform this task on all nodes on an ERPS ring.
- You can enable ERPS for an instance only when it is configured with a control VLAN and a protected VLAN.

Procedure

1. Enter system view.
   system-view
2. Enter ERPS ring view.
   erps ring *ring-id*
3. Enter ERPS instance view.
   instance *instance-id*
4. Enable ERPS for the instance.
   instance enable
   By default, ERPS is disabled for an instance.

Enabling R-APS packets to carry the ring ID in the destination MAC address

About this task

Perform this task to configure the ring ID as the last byte of the destination MAC address for R-APS packets. The ring of R-APS packets can be identified by their destination MAC addresses.

Restrictions and guidelines

Perform this task on all nodes on an ERPS ring.

Procedure

1. Enter system view.
   system-view
2. Enter ERPS ring view.
   erps ring *ring-id*
3. Enable R-APS packets to carry the ring ID in the destination MAC address.
   r-aps ring-mac
   By default, R-APS packets do not carry ring IDs in their destination MAC addresses. The last byte of the destination MAC address is 1.

Configuring R-APS packet levels

Restrictions and guidelines

Perform this task on all nodes on an ERPS ring.

On a ring, the levels of R-APS packets must be the same for all nodes in an ERPS instance.

A node does not process R-APS packets whose levels are greater than the level of R-APS packets sent by the node.

Procedure

1. Enter system view.
   system-view
2. Enter ERPS ring view.
   erps ring *ring-id*
3. Enter ERPS instance view.
   instance *instance-id*
4. Configure the R-APS packet level.
   r-aps level *level-value*
   By default, the level for R-APS packets is 7.

Setting ERPS timers

Restrictions and guidelines

Perform this task on the owner node on an ERPS ring.

Procedure

1. Enter system view.
   system-view
2. Enter ERPS ring view.
   erps ring *ring-id*
3. Enter ERPS instance view.
   instance *instance-id*
4. Set the guard timer.
   timer guard *guard-value*
   By default, the guard timer is 500 milliseconds.
5. Set the hold-off timer.
   timer hold-off *hold-off-value*
   By default, the hold-off timer is 0 milliseconds.
6. Set the WTR timer.
   timer wtr *wtr-value*
   By default, the WTR timer is 5 minutes.

Setting the non-revertive mode

About this task

Perform this task if you do not want to switch back to the recovered link after the link fault is cleared.

Restrictions and guidelines

Perform this task on the owner node on an ERPS ring.

Procedure

1. Enter system view.
   system-view
2. Enter ERPS ring view.
   erps ring *ring-id*
3. Enter ERPS instance view.
   instance *instance-id*
4. Set the non-revertive mode.
   revertive-operation non-revertive
   By default, revertive mode is used.

Setting a switchover mode

Restrictions and guidelines

Perform this task on the nodes that you want to block their ports.

Procedure

1. Enter system view.
   system-view
2. Set a switchover mode.

erps switch { **force** | **manual** } **ring** *ring-id* **instance** *instance-id* { **port0** | **port1** }

By default, no switchover mode is not set.

Associating a ring with a subring

About this task

On a multi-ring network, perform this task if you want to advertise topology changes in a subring to a ring.

Restrictions and guidelines

Perform this task on the interconnection node on an ERPS ring.

Procedure

1. Enter system view.
   system-view
2. Enter ERPS ring view.
   erps ring *ring-id*
3. Configure the ERPS ring as a subring.
   ring-type sub-ring
   By default, an ERPS ring is a major ring.
4. Enter ERPS instance view.
   instance *instance-id*
5. Associate a ring with the subring.
   sub-ring connect ring *ring-id* **instance** *instance-id*
   By default, a subring is not associated with any rings.

Enabling flush packet transparent transmission

About this task

This feature enables the interconnection nodes to forward flush packets for topology changes in the subring to the ring associated with the subring. The associated ring can flush the MAC address table quickly to speed up convergence.

Restrictions and guidelines

Perform this task on the interconnection node on an ERPS ring.

To use this feature, you must also associate a subring on the interconnection node with the ring.

Procedure

1. Enter system view.
   system-view
2. Enable flush packet transparent transmission.
   erps tcn-propagation
   By default, flush packet transparent transmission is disabled.

Removing the MS mode and FS mode settings for an ERPS ring

About this task

After you configure this task, the owner node can ignore the WTR timer and immediately switch traffic to the recovered link upon link recovery.

This task also switches an ERPS ring in non-revertive mode to revertive mode.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   erps clear ring *ring-id* **instance** *instance-id*

Displaying and maintaining ERPS

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display brief ERPS information.	**display erps**
Display detailed ERPS information.	**display erps detail ring** *ring-id* [**instance** *instance-id*]
Display ERPS packet statistics.	**display erps statistics** [**ring** *ring-id* [**instance** *instance-id*]]
Clear ERPS packet statistics.	**reset erps statistics ring** *ring-id* [**instance** *instance-id*]

Troubleshooting ERPS

The owner node cannot receive SF packets from a faulty node when the link state is normal

Symptom

The link between the owner node and the faulty node is available, but the owner node cannot receive SF packets sent by the faulty node. The RPL port is blocked.

Analysis

Possible reasons include:

- ERPS is not enabled for some nodes on the ERPS ring.
- The ring IDs are different for the nodes on the same ERPS ring.
- The control VLAN IDs are different for the nodes in the same ERPS instance.
- A port on the ERPS ring is faulty.

Solutions

To resolve the problem:

- Use the **display erps** command to examine whether ERPS is enabled for all nodes on the ERPS ring. If ERPS is disabled for some nodes, use the **erps enable** command to enable ERPS for the nodes.
- Set the same ring ID for all nodes on a ERPS ring and configure the same control VLAN for all nodes in an ERPS instance.
- Use the **display erps detail** command to examine the port status for all nodes. Bring up the ports in down state.
- Use the **debugging erps** command on all nodes to view debugging information about packets and node status.

Contents

Configuring Smart Link 1

About Smart Link 1
Application scenario 1
Terminology 2
How Smart Link works 2
Collaboration between Smart Link and Monitor Link for port status detection 3
Restrictions and guidelines: Smart Link configuration 3
Smart Link tasks at a glance 4
Configuring a Smart Link device 4
Prerequisites for Smart Link device configuration 4
Configuring protected VLANs for a smart link group 4
Configuring member ports for a smart link group 4
Configuring a preemption mode for a smart link group 5
Enabling the sending of flush messages 5
Enabling the sending of dummy messages 6
Enabling an associated device to receive flush messages 6
Display and maintenance commands for Smart Link 7
Smart Link configuration examples 7
Example: Configuring a single smart link group 7
Example: Configuring multiple smart link groups load sharing 12

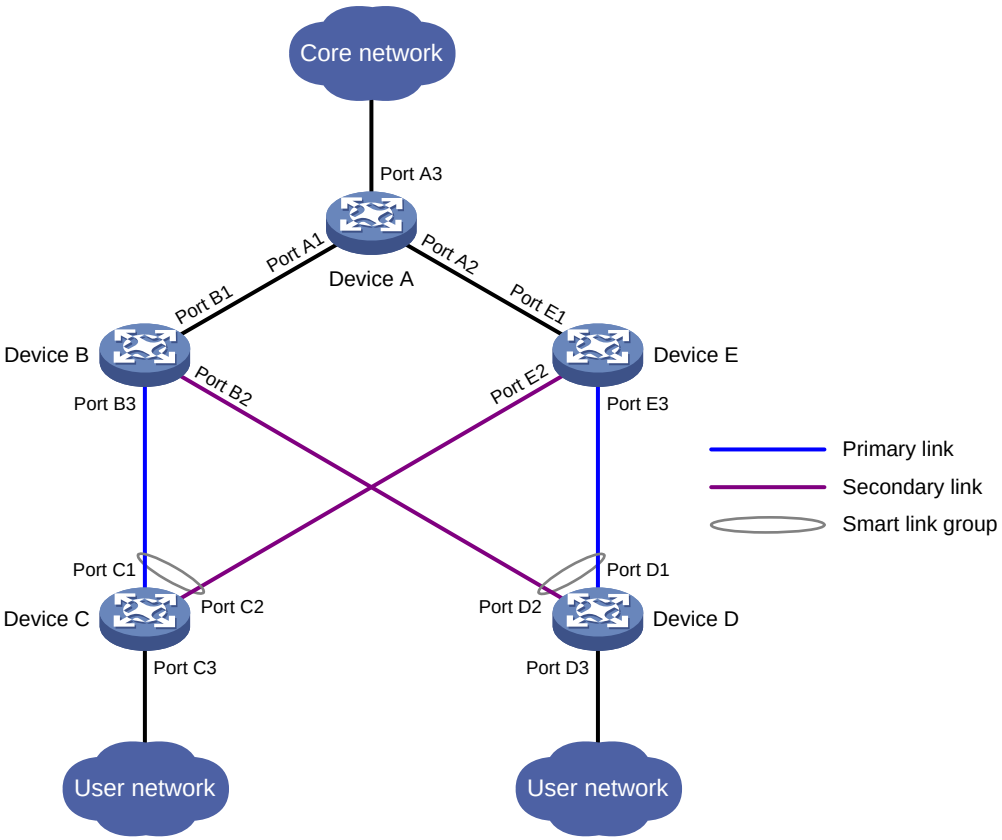
Configuring Smart Link

About Smart Link

Application scenario

Smart Link is an H3C-proprietary protocol that provides link redundancy and subsecond convergence time in a dual uplink network. As shown in Figure 1, Smart Link is configured on Device C and Device D. The secondary link takes over quickly when the primary link fails.

Figure 1 Smart link application scenario diagram



A Smart Link network has the following devices:

- **Smart Link devices**—A Smart Link device has two uplinks. A Smart Link device must be configured with a smart link group and a transmit control VLAN to transmit flush messages. Device C and Device D in Figure 1 are Smart Link devices.
- **Associated devices**—An associated device is an uplink device to which Smart Link devices are connected. An associated device supports Smart Link, and receives flush messages sent from the specified control VLAN. When a primary/secondary link switchover occurs, the associated device updates the MAC address entries and ARP/ND entries according to received flush messages. Device A, Device B, and Device E in Figure 1 are associated devices.

Terminology

Smart link group

A smart link group consists of only two member ports: the primary and the secondary ports. Only one port is active for forwarding at a time, and the other port is blocked and in standby state. When link failure occurs on the active port due to port shutdown or the presence of unidirectional link, the standby port becomes active and takes over. The original active port transits to the blocked state.

As shown in the Smart link application scenario diagram, Port C1 and Port C2 of Device C form a smart link group. Port C1 is active, and Port C2 is standby. Port D1 and Port D2 of Device D form another smart link group. Port D1 is active, and Port D2 is standby.

Primary/secondary port

Primary port and secondary port are two port types in a smart link group. When both ports in a smart link group are up, the primary port preferentially transits to the forwarding state. The secondary port stays in standby state. When the primary port fails, the secondary port takes over to forward traffic.

As shown in the Smart link application scenario diagram, Port C1 of Device C and Port D1 of Device D are primary ports. Port C2 of Device C and Port D2 of Device D are secondary ports.

Primary/secondary link

The link that connects the primary port in a smart link group is the primary link. The link that connects the secondary port is the secondary link.

Flush message

When link switchover occurs, the smart link group uses flush messages to notify other devices to refresh their MAC address entries and ARP/ND entries. Flush messages are common multicast data packets, and will be dropped by a blocked receiving port.

Protected VLAN

A smart link group controls the forwarding state of protected VLANs. Each smart link group on a port controls a different protected VLAN. The state of the port in a protected VLAN is determined by the state of the port in the smart link group.

Transmit control VLAN

The transmit control VLAN is used for transmitting flush messages. When link switchover occurs, the devices (such as Device C and Device D in the Smart link application scenario diagram) send flush messages within the transmit control VLAN.

Receive control VLAN

The receive control VLAN is used for receiving and processing flush messages. When link switchover occurs, the devices (such as Device A, Device B, and Device E in the Smart link application scenario diagram) receive and process flush messages in the receive control VLAN. In addition, they refresh their MAC address entries and ARP/ND entries.

How Smart Link works

Link backup

As shown in the Smart link application scenario diagram, the link on Port C1 of Device C is the primary link. The link on Port C2 of Device C is the secondary link. Port C1 is in forwarding state, and Port C2 is in standby state. When the primary link fails, Port C2 takes over to forward traffic and Port C1 is blocked and placed in standby state.

When a port switches to the forwarding state, the system outputs log information to notify the user of the port state change.

Topology change

Link switchover might outdate the MAC address entries and ARP/ND entries on all devices. A flush update mechanism is provided to ensure correct packet transmission. With this mechanism, a Smart Link-enabled device updates its information by transmitting flush messages over the backup link to its upstream devices. This mechanism requires the upstream devices to be capable of recognizing Smart Link flush messages to update their MAC address forwarding entries and ARP/ND entries.

Preemption mode

As shown in the Smart link application scenario diagram, the link on Port C1 of Device C is the primary link. The link on Port C2 of Device C is the secondary link. When the primary link fails, Port C1 is automatically blocked and placed in standby state, and Port C2 takes over to forward traffic. When the primary link recovers, one of the following actions occurs:

- If the smart link group is not configured with a preemption mode, Port C1 stays blocked to keep traffic forwarding stable. Port C1 does not take over to forward traffic until the next link switchover.
- If the smart link group is configured with a preemption mode and the preemption conditions are met, Port C1 takes over to forward traffic as soon as its link recovers. Port C2 is automatically blocked and placed in standby state.

Load sharing

A ring network might carry traffic of multiple VLANs. Smart Link can forward traffic from different VLANs in different smart link groups for load sharing.

To implement load sharing, you can assign a port to multiple smart link groups. Configure each group with a different protected VLAN. Make sure the state of the port is different in these smart link groups, so traffic from different VLANs can be forwarded along different paths.

You can configure protected VLANs for a smart link group by referencing Multiple Spanning Tree Instances (MSTIs). For more information about MSTIs, see *Layer 2—LAN Switching Configuration Guide*.

Collaboration between Smart Link and Monitor Link for port status detection

Smart Link cannot detect when faults occur on the uplink of the upstream devices or when faults are cleared. You can configure the Monitor Link function to monitor the status of the uplink ports of the upstream devices. Monitor Link adapts the up/down state of downlink ports to uplink ports, and triggers Smart Link to perform link switchover on the downstream device. For more information about Monitor Link, see "Configuring Monitor Link."

Restrictions and guidelines: Smart Link configuration

When you configure Smart Link, follow these restrictions and guidelines:

- If you configure a port as both an aggregation group member and a smart link group member, only the aggregation group configuration takes effect. The port is not shown in the output from the **`display smart-link group`** command. The smart link group configuration takes effect after the port leaves the aggregation group.
- In an M-LAG network:
  - The two M-LAG member devices must have the same Smart Link settings.
  - For a smart link group to work with an M-LAG system, assign only M-LAG interfaces to the smart link group.

- Do not assign a peer-link interface to a smart link group.

Smart Link tasks at a glance

To configure Smart Link, perform the following tasks:

1. Configuring a Smart Link device
   a. Configuring protected VLANs for a smart link group
   b. Configuring member ports for a smart link group
   c. (Optional.) Configuring a preemption mode for a smart link group
   d. (Optional.) Enabling the sending of flush messages
2. Enabling an associated device to receive flush messages

Configuring a Smart Link device

Prerequisites for Smart Link device configuration

Before configuring a Smart Link device, complete the following tasks:

- To prevent loops, shut down a port before configuring it as a smart link group member. You can bring up the port only after completing the smart link group configuration.
- Disable the spanning tree feature, RRPP, and ERPS on the ports you want to add to the smart link group.

Configuring protected VLANs for a smart link group

Prerequisites

Before you configure protected VLANs, you must configure an MST region and the VLAN-to-instance mapping table. For more information about MST regions, see spanning tree configuration in *Layer 2—LAN Switching Configration Guide*.

Procedure

1. Enter system view.
   `system-view`
2. Create a smart link group and enter smart link group view.
   `smart-link group` *`group-id`*
3. Configure protected VLANs for the smart link group.
   `protected-vlan reference-instance` *`instance-id-list`*

Configuring member ports for a smart link group

Restrictions and guidelines

You can configure member ports for a smart link group either in smart link group view or in interface view. The configurations made in these two views have the same effect.

In smart link group view

1. Enter system view.
   `system-view`
2. Create a smart link group and enter smart link group view.

smart-link group *group-id*

3. Configure member ports for a smart link group.
   port *interface-type interface-number* { **primary** | **secondary** }
   By default, no member port is configured for a smart link group.

In interface view

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   interface *interface-type interface-number*
3. Configure member ports for a smart link group.
   port smart-link group *group-id* { **primary** | **secondary** }
   By default, an interface is not a smart link group member.

Configuring a preemption mode for a smart link group

1. Enter system view.
   system-view
2. Enter smart link group view.
   smart-link group *group-id*
3. Configure a preemption mode for the smart link group.
   preemption mode { **role** | **speed** [**threshold** *threshold-value*] }
   By default, preemption is disabled.
4. Configure the preemption delay.
   preemption delay *delay*
   By default, the preemption delay is 1 second.
   The preemption delay configuration takes effect only after a preemption mode is configured.

Enabling the sending of flush messages

Restrictions and guidelines

- The control VLAN configured for a smart link group must be different from the control VLAN configured for any other smart link group.
- Make sure the configured control VLAN already exists, and assign the smart link group member ports to the control VLAN.
- The control VLAN of a smart link group must also be one of its protected VLANs. Do not remove the control VLAN. Otherwise, flush messages cannot be sent correctly.

Procedure

1. Enter system view.
   system-view
2. Enter smart link group view.
   smart-link group *group-id*
3. Enable flush update.
   flush enable [**control-vlan** *vlan-id*]
   By default, flush update is enabled, and VLAN 1 is the control VLAN.

Enabling the sending of dummy messages

Restrictions and guidelines

After you enable this feature, the device can send the MAC addresses learned from downlink ports to the peer through dummy messages. The device performs such actions when the state of ports in the smart link group changes from standby to active.

Restrictions and guidelines

Sending of flush messages and sending of dummy messages are mutually exclusive.

After you enable sending of dummy messages, configure the **undo stp enable** command on the connected device to prevent packet loss.

Procedure

1. Enter system view.
   system-view
2. Enter smart link group view.
   smart-link group *group-id*
3. Enable sending of dummy messages.
   dummy enable [**max-number** *number*] [**times** *times*]
   By default, sending of dummy messages is enabled.

Enabling an associated device to receive flush messages

Restrictions and guidelines

- You do not need to enable all ports on the associated devices to receive flush messages. Enable the feature only on all control VLANs of ports on the primary and secondary links between the Smart Link device and the destination device.
- If no control VLAN is specified for processing flush messages, the device forwards the received flush messages without any processing.
- Make sure the receive control VLAN is the same as the transmit control VLAN configured on the Smart Link device. If they are not the same, the associated device will forward the received flush messages directly without any processing.
- Do not remove the control VLANs. Otherwise, flush messages cannot be sent correctly.
- Make sure the control VLANs are existing VLANs, and assign the ports capable of receiving flush messages to the control VLANs.

Prerequisites

Disable the spanning tree feature on the associated device's ports that connect to the member ports of the smart link group. Otherwise, the ports will discard flush messages when they are not in forwarding state if a topology change occurs.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 2 aggregate interface view.
   interface *interface-type interface-number*
3. Configure the control VLANs for receiving flush messages.

smart-link flush enable [**control-vlan** *vlan-id-list*]

By default, no control VLAN receives flush messages.

Display and maintenance commands for Smart Link

Execute **display** commands in any view and the **reset** command in user view:

Task	Command
Display information about the received flush messages.	**display smart-link flush**
Display smart link group information.	**display smart-link group** { *group-id* | **all** }
Clear the statistics about flush messages.	**reset smart-link statistics**

Smart Link configuration examples

Example: Configuring a single smart link group

Network configuration

As shown in Figure 2:

- Device C and Device D are Smart Link devices. Device A, Device B, and Device E are associated devices. Traffic of VLANs 1 through 30 on Device C and Device D is dually uplinked to Device A.
- Configure Smart Link on Device C and Device D for dual uplink backup.

Figure 2 Network diagram

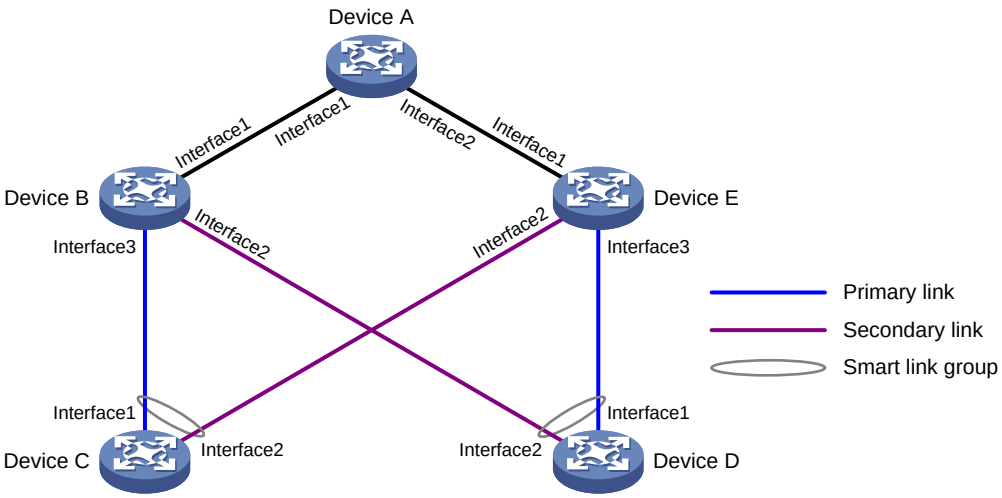


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Device C:

Create VLANs 1 through 30.

```
<DeviceC> system-view
[DeviceC] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceC] stp region-configuration
[DeviceC-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Shut down GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] shutdown
```

Disable the spanning tree feature on the port.

```
[DeviceC-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] shutdown
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceC-GigabitEthernet1/0/2] quit
```

Create smart link group 1, and configure all VLANs mapped to MSTI 1 as the protected VLANs.

```
[DeviceC] smart-link group 1
[DeviceC-smlk-group1] protected-vlan reference-instance 1
```

Configure GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port for smart link group 1.

```
[DeviceC-smlk-group1] port gigabitethernet 1/0/1 primary
[DeviceC-smlk-group1] port gigabitethernet 1/0/2 secondary
```

Enable flush message sending in smart link group 1, and configure VLAN 10 as the transmit control VLAN.

```
[DeviceC-smlk-group1] flush enable control-vlan 10
[DeviceC-smlk-group1] quit
```

Bring up GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 again.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo shutdown
[DeviceC-GigabitEthernet1/0/1] quit
```

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo shutdown
[DeviceC-GigabitEthernet1/0/2] quit
```

2. Configure Device D:

Create VLANs 1 through 30.

```
<DeviceD> system-view
[DeviceD] vlan 1 to 30
```

Map these VLANs to MSTI 1.

```
[DeviceD] stp region-configuration
[DeviceD-mst-region] instance 1 vlan 1 to 30
```

Activate the MST region configuration.

```
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

Shut down GigabitEthernet 1/0/1.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] shutdown
```

Disable the spanning tree feature on the port.

```
[DeviceD-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
[DeviceD-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] shutdown
[DeviceD-GigabitEthernet1/0/2] undo stp enable
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceD-GigabitEthernet1/0/2] quit
```

Create smart link group 1, and configure all VLANs mapped to MSTI 1 as the protected VLANs.

```
[DeviceD] smart-link group 1
[DeviceD-smlk-group1] protected-vlan reference-instance 1
```

Configure GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port for smart link group 1.

```
[DeviceD-smlk-group1] port gigabitethernet 1/0/1 primary
[DeviceD-smlk-group1] port gigabitethernet 1/0/2 secondary
```

Enable flush message sending in smart link group 1, and configure VLAN 20 as the transmit control VLAN.

```
[DeviceD-smlk-group1] flush enable control-vlan 20
[DeviceD-smlk-group1] quit
```

Bring up GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 again.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] undo shutdown
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] undo shutdown
```

```
[DeviceD-GigabitEthernet1/0/2] quit
```

3. Configure Device B:

Create VLANs 1 through 30.

```
<DeviceB> system-view
[DeviceB] vlan 1 to 30
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
```

Enable flush message receiving and configure VLAN 10 and VLAN 20 as the receive control VLANs on the port.

```
[DeviceB-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 20
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/2] undo stp enable
```

Enable flush message receiving and configure VLAN 20 as the receive control VLAN on the port.

```
[DeviceB-GigabitEthernet1/0/2] smart-link flush enable control-vlan 20
[DeviceB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan 1 to 30
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/3] undo stp enable
```

Enable flush message receiving and configure VLAN 10 as the receive control VLAN on the port.

```
[DeviceB-GigabitEthernet1/0/3] smart-link flush enable control-vlan 10
[DeviceB-GigabitEthernet1/0/3] quit
```

4. Configure Device E:

Create VLANs 1 through 30.

```
<DeviceE> system-view
[DeviceE] vlan 1 to 30
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceE-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
```

Enable flush message receiving and configure VLAN 10 and VLAN 20 as the receive control VLANs on the port.

```
[DeviceE-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 20
[DeviceE-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port.

```
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceE-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
```

Disable the spanning tree feature on the port.

```
[DeviceE-GigabitEthernet1/0/2] undo stp enable
```

Enable flush message receiving and configure VLAN 10 as the receive control VLAN on the port.

```
[DeviceE-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10
[DeviceE-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 as a trunk port.

```
[DeviceE] interface gigabitethernet 1/0/3
[DeviceE-GigabitEthernet1/0/3] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceE-GigabitEthernet1/0/3] port trunk permit vlan 1 to 30
```

Disable the spanning tree feature on the port.

```
[DeviceE-GigabitEthernet1/0/3] undo stp enable
```

Enable flush message receiving and configure VLAN 20 as the receive control VLAN on the port.

```
[DeviceE-GigabitEthernet1/0/3] smart-link flush enable control-vlan 20
[DeviceE-GigabitEthernet1/0/3] quit
```

5. Configure Device A:

Create VLANs 1 through 30.

```
<DeviceA> system-view
[DeviceA] vlan 1 to 30
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 30.

```
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
```

Enable flush message receiving and configure VLAN 10 and VLAN 20 as the receive control VLANs on the port.

```
[DeviceA-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 20
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceA-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10 20
[DeviceA-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Display the smart link group configuration on Device C.

```
[DeviceC] display smart-link group 1
Smart link group 1 information:
```

```
  Device ID       : 000f-e23d-5af0
  Preemption mode : None
  Preemption delay: 1(s)
  Control VLAN    : 10
  Protected VLAN  : Reference Instance 1

  Member                  Role      State   Flush-count     Last-flush-time
  ------------------------------------------------------------------------------
  GE1/0/1                 PRIMARY   ACTIVE  5               16:45:20 2012/04/21
  GE1/0/2                 SECONDARY STANDBY 1               16:37:20 2012/04/21
```

Display the flush messages received on Device B.

```
[DeviceB] display smart-link flush
 Received flush packets                              : 5
 Receiving interface of the last flush packet        : GigabitEthernet1/0/3
 Receiving time of the last flush packet             : 16:50:21 2012/04/21
 Device ID of the last flush packet                  : 000f-e23d-5af0
 Control VLAN of the last flush packet               : 10
```

Example: Configuring multiple smart link groups load sharing

Network configuration

As shown in Figure 3:

- Device C is a Smart Link device. Device A, Device B, and Device D are associated devices. Traffic of VLANs 1 through 200 on Device C is dually uplinked to Device A by Device B and Device D.
- Implement dual uplink backup and load sharing on Device C. Traffic of VLANs 1 through 100 is uplinked to Device A by Device B. Traffic of VLANs 101 through 200 is uplinked to Device A by Device D.

Figure 3 Network diagram

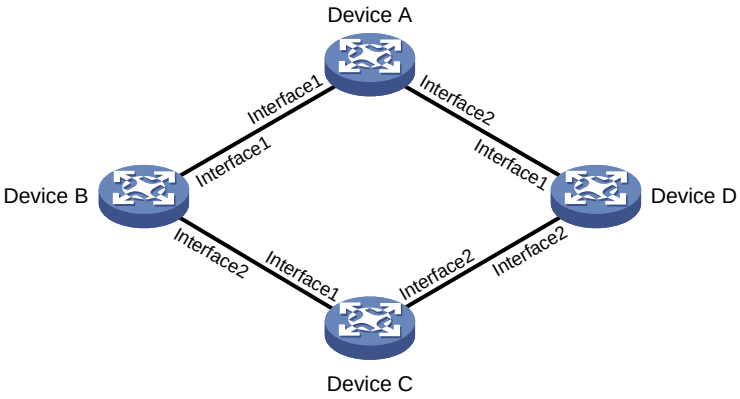


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device C:

Create VLAN 1 through VLAN 200.

```
<DeviceC> system-view
[DeviceC] vlan 1 to 200
```

Map VLANs 1 through 100 to MSTI 1, and VLANs 101 through 200 to MSTI 2.

```
[DeviceC] stp region-configuration
[DeviceC-mst-region] instance 1 vlan 1 to 100
[DeviceC-mst-region] instance 2 vlan 101 to 200
```

Activate the MST region configuration.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Shut down GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] shutdown
```

Disable the spanning tree feature on the port.

```
[DeviceC-GigabitEthernet1/0/1] undo stp enable
```

Configure the port as a trunk port.

```
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLAN 1 through VLAN 200.

```
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
[DeviceC-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] shutdown
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
[DeviceC-GigabitEthernet1/0/2] quit
```

Create smart link group 1, and configure all VLANs mapped to MSTI 1 as the protected VLANs for smart link group 1.

```
[DeviceC] smart-link group 1
[DeviceC-smlk-group1] protected-vlan reference-instance 1
```

Configure GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port for smart link group 1.

```
[DeviceC-smlk-group1] port gigabitethernet 1/0/1 primary
[DeviceC-smlk-group1] port gigabitethernet 1/0/2 secondary
```

Enable role preemption in smart link group 1, enable flush message sending, and configure VLAN 10 as the transmit control VLAN.

```
[DeviceC-smlk-group1] preemption mode role
[DeviceC-smlk-group1] flush enable control-vlan 10
[DeviceC-smlk-group1] quit
```

Create smart link group 2, and configure all VLANs mapped to MSTI 2 as the protected VLANs for smart link group 2.

```
[DeviceC] smart-link group 2
[DeviceC-smlk-group2] protected-vlan reference-instance 2
```

Configure GigabitEthernet 1/0/1 as the secondary port and GigabitEthernet 1/0/2 as the primary port for smart link group 2.

```
[DeviceC-smlk-group2] port gigabitethernet 1/0/2 primary
[DeviceC-smlk-group2] port gigabitethernet 1/0/1 secondary
```

Enable role preemption in smart link group 2, enable flush message sending, and configure VLAN 110 as the transmit control VLAN.

```
[DeviceC-smlk-group2] preemption mode role
[DeviceC-smlk-group2] flush enable control-vlan 110
[DeviceC-smlk-group2] quit
```

Bring up GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo shutdown
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo shutdown
[DeviceC-GigabitEthernet1/0/2] quit
```

2. Configure Device B:

Create VLAN 1 through VLAN 200.

```
<DeviceB> system-view
[DeviceB] vlan 1 to 200
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 200.

```
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
```

Enable flush message receiving and configure VLAN 10 and VLAN 110 as the receive control VLANs on the port.

```
[DeviceB-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 110
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
```

Assign the port to VLANs 1 through 200.

```
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
```

Disable the spanning tree feature on the port.

```
[DeviceB-GigabitEthernet1/0/2] undo stp enable
```

Enable flush message receiving and configure VLAN 10 and VLAN 110 as the receive control VLANs on the port.

```
[DeviceB-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10 110
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device D:

Create VLAN 1 through VLAN 200.

```
<DeviceD> system-view
[DeviceD] vlan 1 to 200
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 200.

```
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
```

Enable flush message receiving and configure VLAN 10 and VLAN 110 as the receive control VLANs on the port.

```
[DeviceD-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 110
[DeviceD-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
```

Assign the port to VLANs 1 through 200.

```
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
```

Disable the spanning tree feature on the port.

```
[DeviceD-GigabitEthernet1/0/2] undo stp enable
```

Enable flush message receiving and configure VLAN 10 and VLAN 110 as the receive control VLANs on the port.

```
[DeviceD-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10 110
[DeviceD-GigabitEthernet1/0/2] quit
```

4. Configure Device A:

Create VLAN 1 through VLAN 200.

```
<DeviceA> system-view
[DeviceA] vlan 1 to 200
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
```

Assign the port to VLANs 1 through 200.

```
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
```

Enable flush message receiving and configure VLAN 10 and VLAN 110 as the receive control VLANs on the port.

```
[DeviceA-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 110
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
[DeviceA-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10 110
[DeviceA-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Display the smart link group configuration on Device C.

```
[DeviceC] display smart-link group all
Smart link group 1 information:
  Device ID       : 000f-e23d-5af0
  Preemption mode : Role
  Preemption delay: 1(s)
  Control VLAN    : 10
  Protected VLAN  : Reference Instance 1

  Member                  Role      State   Flush-count     Last-flush-time
  ------------------------------------------------------------------------------
  GE1/0/1                 PRIMARY   ACTIVE  5               16:45:20 2012/04/21
```

```
  GE1/0/2                 SECONDARY STANDBY 1              16:37:20 2012/04/21

Smart link group 2 information:
  Device ID       : 000f-e23d-5af0
  Preemption mode : Role
  Preemption delay: 1(s)
  Control VLAN    : 110
  Protected VLAN  : Reference Instance 2

  Member                  Role      State   Flush-count     Last-flush-time
  ------------------------------------------------------------------------------
  GE1/0/2                 PRIMARY   ACTIVE  5               16:45:20 2012/04/21
  GE1/0/1                 SECONDARY STANDBY 1               16:37:20 2012/04/21
```

Display the flush messages received on Device B.

```
[DeviceB] display smart-link flush
 Received flush packets                              : 5
 Receiving interface of the last flush packet        : GigabitEthernet1/0/2
 Receiving time of the last flush packet             : 16:25:21 2012/04/21
 Device ID of the last flush packet                  : 000f-e23d-5af0
 Control VLAN of the last flush packet               : 10
```

Contents

Configuring Monitor Link 1

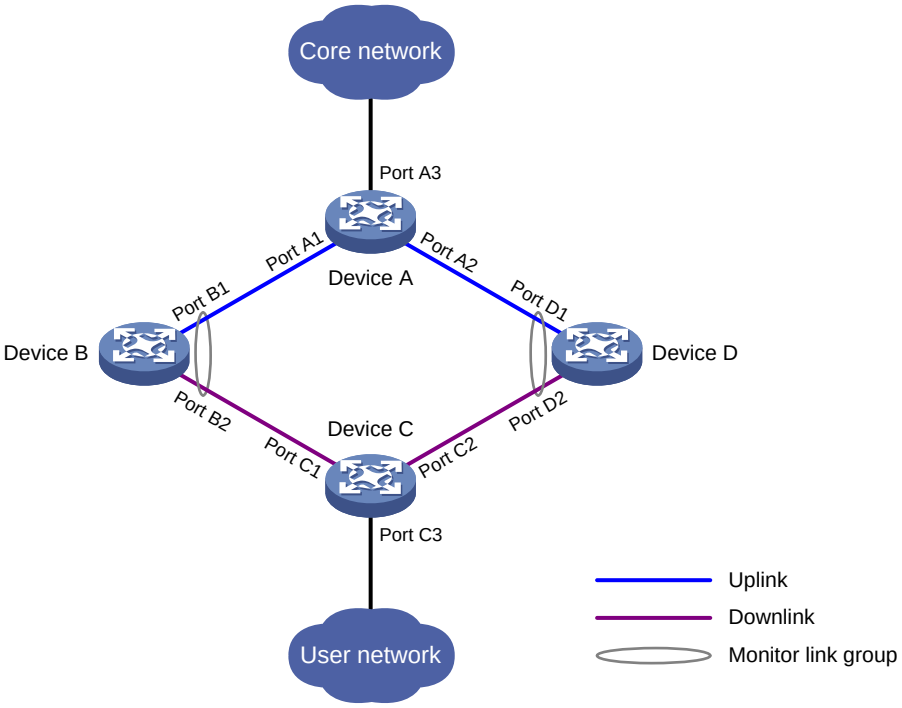
About Monitor Link 1
Restrictions and guidelines: Monitor Link configuration 2
Monitor Link tasks at a glance 2
Enabling Monitor Link globally 2
Creating a monitor link group 2
Configuring monitor link group member interfaces 3
Restrictions and guidelines 3
Configuring monitor link group member interfaces in monitor link group view 3
Configuring monitor link group member interfaces in interface view 3
Enabling Monitor Link to monitor BGP instance process state 4
Configuring the uplink interface threshold for triggering monitor link group state switchover 4
Configuring the switchover delay for the downlink interfaces in a monitor link group 5
Display and maintenance commands for Monitor Link 5
Monitor Link configuration examples 5
Example: Configuring Monitor Link 5

Configuring Monitor Link

About Monitor Link

Monitor Link is an H3C-proprietary protocol that associates the state of downlink interfaces with the state of uplink interfaces in a monitor link group. When Monitor Link shuts down the downlink interfaces because of an uplink failure, the downstream device changes connectivity to another link.

Figure 1 Monitor Link application scenarios



A monitor link group contains uplink and downlink interfaces. An interface can belong to only one monitor link group.

- Uplink interfaces are the monitored interfaces.
- Downlink interfaces are the monitoring interfaces.

As shown in Figure 1:

- Port B1 and Port B2 of Device B form a monitor link group. Port B1 is an uplink interface, and Port B2 is a downlink interface.
- Port D1 and Port D2 of Device D form another monitor link group. Port D1 is an uplink interface, and Port D2 is a downlink interface.

A monitor link group works independently of other monitor link groups. When a monitor link group does not contain any uplink interface or all its uplink interfaces are down, the monitor link group goes down. It forces all downlink interfaces down at the same time. When any uplink interface comes up, the monitor link group comes up and brings up all the downlink interfaces.

Restrictions and guidelines: Monitor Link configuration

Follow these restrictions and guidelines when you configure Monitor Link:

- Do not manually shut down or bring up the downlink interfaces in a monitor link group.
- To avoid frequent state changes of downlink interfaces in the event that uplink interfaces in the monitor link group flap, you can configure a switchover delay. The switchover delay is the time that the downlink interfaces wait before they come up following an uplink interface.

Monitor Link tasks at a glance

To configure Monitor Link, perform the following tasks:

- Enabling Monitor Link globally
- Creating a monitor link group
- Configuring monitor link group member interfaces
- Enabling Monitor Link to monitor BGP instance process state
- (Optional.) Configuring the uplink interface threshold for triggering monitor link group state switchover
- (Optional.) Configuring the switchover delay for the downlink interfaces in a monitor link group

Enabling Monitor Link globally

About this task

All monitor link groups can operate only after you enable Monitor Link globally. When you disable Monitor Link globally, all monitor link groups cannot operate and the downlink interfaces brought down by the monitor link groups resume their original states.

Procedure

1. Enter system view.
   system-view
2. Enable Monitor Link globally.
   undo monitor-link disable
   By default, Monitor Link is enabled globally.

Creating a monitor link group

1. Enter system view.
   system-view
2. Create a monitor link group and enter monitor link group view.
   monitor-link group *group-id*

Configuring monitor link group member interfaces

Restrictions and guidelines

- An interface can be assigned to only one monitor link group.
- To avoid undesired down/up state changes on the downlink interfaces, configure uplink interfaces before you configure downlink interfaces.
- If you have configured an interface as the downlink interface of a monitor link group, do not configure its subinterfaces as the uplink interfaces of any monitor link group. Otherwise, the Monitor Link operation might be interrupted.
- The state of subinterfaces is associated with the state of the interface. Do not add the interface and its subinterfaces to the same monitor link group. Otherwise, the monitor link group performance might be affected.
- If you have configured a Selected port of an aggregation group as the downlink interface of a monitor link group, do not configure an Unselected port of the aggregation group as the uplink interface of the monitor link group.
- Do not add an aggregate interface and its member ports to the same monitor link group.
- You can configure member interfaces for a monitor link group in monitor link group view or interface view. Configurations made in these views have the same effect. The configuration is supported by the following interfaces:
  - Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces.
  - Layer 3 Ethernet interfaces/subinterfaces and Layer 3 aggregate interfaces/subinterfaces.
  - Loopback interfaces.
  - VLAN interfaces.

Configuring monitor link group member interfaces in monitor link group view

1. Enter system view.

   system-view
2. Enter monitor link group view.

   monitor-link group *group-id*
3. Configure member interfaces for the monitor link group.

   port *interface-type* { *interface-number* | *interface-number*.*subnumber* } { **downlink** | **uplink** }

   By default, no member interfaces exist in a monitor link group.

Configuring monitor link group member interfaces in interface view

1. Enter system view.

   system-view
2. Enter interface view or subinterface view.

   interface *interface-type* { *interface-number* | *interface-number.subnumber* }
3. Configure the interface as a member of a monitor link group.

port monitor-link group *group-id* { **downlink** | **uplink** }

By default, the interface is not a monitor link group member.

Enabling Monitor Link to monitor BGP instance process state

About this task

The Monitor Link group can monitor not only uplink interfaces but also BGP process and session state information:

- BGP process state changes are counted as uplink interface state changes. When a BGP process goes down (the BGP process exists), Monitor Link takes it as an uplink interface down event. When the BGP process comes up, Monitor Link takes it as an uplink interface up event.
- BGP session state changes are counted as uplink interface state changes. When a BGP session goes down, Monitor Link takes it as an uplink interface down event. When the BGP session comes up, Monitor Link takes it as an uplink interface up event. For more information about monitoring BGP session state, see advanced BGP configuration in *Layer 3—IP Routing Configuration Guide*.

The monitor link group goes down and forces all downlink interfaces down when the number of monitored uplink interfaces in up state is below the uplink interface threshold. When the number of uplink interfaces in up state is greater than or equal to the uplink interface threshold, the monitor link group comes up and restores all downlink interfaces to up state.

Restrictions and guidelines

When the BGP processes and sessions monitored by the monitor link group belong to the same BGP instance, the following situations exist:

- When BGP sessions in up state exist in the BGP instance, Monitor link counts only the number of BGP sessions in up state when counting the number of uplink interfaces in up state. In this case, Monitor link does not count the number of BGP processes in up state.
- When no BGP sessions in up state exist in the BGP instance, Monitor link only checks whether the BGP instance process is up when counting the number of uplink interfaces in up state.

Procedure

1. Enter system view.

   system-view
2. Enter monitor link group view.

   monitor-link group *group-id*
3. Enable Monitor Link to monitor the process state of the specified BGP instance.

   monitor bgp *instance-name*

Configuring the uplink interface threshold for triggering monitor link group state switchover

1. Enter system view.

   system-view
2. Enter monitor link group view.

   monitor-link group *group-id*
3. Configure the uplink interface threshold for triggering monitor link group state switchover.

uplink up-port-threshold *number-of-port*

By default, the uplink interface threshold for triggering monitor link group state switchover is 1.

Configuring the switchover delay for the downlink interfaces in a monitor link group

1. Enter system view.
   system-view
2. Enter monitor link group view.
   monitor-link group *group-id*
3. Configure the switchover delay for the downlink interfaces in the monitor link group.
   downlink up-delay *delay*
   By default, the switchover delay is 0 seconds. The downlink interfaces come up as soon as an uplink interface comes up.

Display and maintenance commands for Monitor Link

Execute the **display** command in any view:

Task	Command
Display monitor link group information.	**display monitor-link** [**bgp**] **group** { *group-id* | **all** }

Monitor Link configuration examples

Example: Configuring Monitor Link

Network configuration

As shown in Figure 2:

- Device C is a Smart Link device, and Device A, Device B, and Device D are associated devices. Traffic of VLANs 1 through 30 on Device C is dual-uplinked to Device A through a smart link group.
- Implement dual uplink backup on Device C. When the link between Device A and Device B (or Device D) fails, Device C can detect the link fault. It then performs uplink switchover in the smart link group.

For more information about Smart Link, see "Configuring Smart Link."

Figure 2 Network diagram

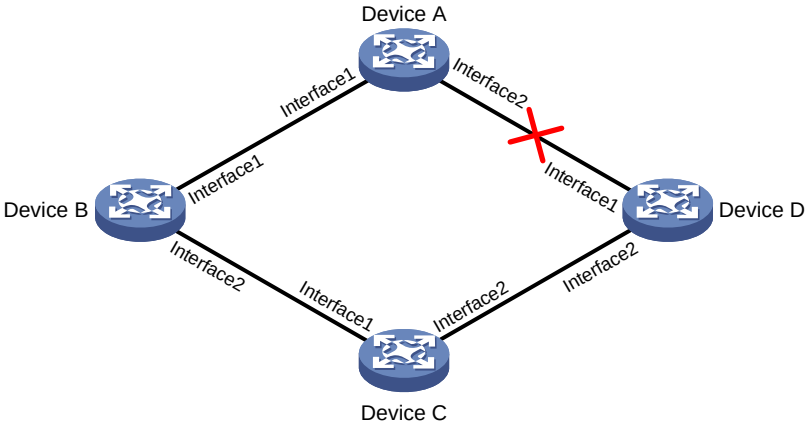


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device C:

   # Create VLANs 1 through 30.

   ```
   <DeviceC> system-view
   [DeviceC] vlan 1 to 30
   ```

   # Map these VLANs to MSTI 1.

   ```
   [DeviceC] stp region-configuration
   [DeviceC-mst-region] instance 1 vlan 1 to 30
   ```

   # Activate MST region configuration.

   ```
   [DeviceC-mst-region] active region-configuration
   [DeviceC-mst-region] quit
   ```

   # Shut down GigabitEthernet 1/0/1.

   ```
   [DeviceC] interface gigabitethernet 1/0/1
   [DeviceC-GigabitEthernet1/0/1] shutdown
   ```

   # Disable the spanning tree feature on the interface.

   ```
   [DeviceC-GigabitEthernet1/0/1] undo stp enable
   ```

   # Configure the interface as a trunk port.

   ```
   [DeviceC-GigabitEthernet1/0/1] port link-type trunk
   ```

   # Assign the interface to VLANs 1 through 30.

   ```
   [DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
   [DeviceC-GigabitEthernet1/0/1] quit
   ```

   # Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

   ```
   [DeviceC] interface gigabitethernet 1/0/2
   [DeviceC-GigabitEthernet1/0/2] shutdown
   [DeviceC-GigabitEthernet1/0/2] undo stp enable
   [DeviceC-GigabitEthernet1/0/2] port link-type trunk
   [DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
   ```

```
[DeviceC-GigabitEthernet1/0/2] quit
```

Create smart link group 1, and configure all the VLANs mapped to MSTI 1 as the protected VLANs for smart link group 1.

```
[DeviceC] smart-link group 1
[DeviceC-smlk-group1] protected-vlan reference-instance 1
```

Configure GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port for smart link group 1.

```
[DeviceC-smlk-group1] port gigabitethernet 1/0/1 primary
[DeviceC-smlk-group1] port gigabitethernet 1/0/2 secondary
```

Enable the smart link group to transmit flush messages.

```
[DeviceC-smlk-group1] flush enable
[DeviceC-smlk-group1] quit
```

Bring up GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo shutdown
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo shutdown
[DeviceC-GigabitEthernet1/0/2] quit
```

2. Configure Device A:

Create VLANs 1 through 30.

```
<DeviceA> system-view
[DeviceA] vlan 1 to 30
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
```

Assign the interface to VLANs 1 through 30.

```
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
```

Enable flush message receiving on the interface.

```
[DeviceA-GigabitEthernet1/0/1] smart-link flush enable
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 in the same way GigabitEthernet 1/0/1 is configured.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
[DeviceA-GigabitEthernet1/0/2] smart-link flush enable
[DeviceA-GigabitEthernet1/0/2] quit
```

3. Configure Device B:

Create VLANs 1 through 30.

```
<DeviceB> system-view
[DeviceB] vlan 1 to 30
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
```

Assign the interface to VLANs 1 through 30.

```
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
```

Enable flush message receiving on the interface.

```
[DeviceB-GigabitEthernet1/0/1] smart-link flush enable
[DeviceB-GigabitEthernet1/0/1] quit
```

Disable the spanning tree feature on GigabitEthernet 1/0/2.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] undo stp enable
```

Configure the interface as a trunk port.

```
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
```

Assign the interface to VLANs 1 through 30.

```
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
```

Enable flush message receiving on the interface.

```
[DeviceB-GigabitEthernet1/0/2] smart-link flush enable
[DeviceB-GigabitEthernet1/0/2] quit
```

Create monitor link group 1.

```
[DeviceB] monitor-link group 1
```

Configure GigabitEthernet 1/0/1 as an uplink interface for monitor link group 1.

```
[DeviceB-mtlk-group1] port gigabitethernet 1/0/1 uplink
```

Configure GigabitEthernet 1/0/2 as a downlink interface for monitor link group 1.

```
[DeviceB-mtlk-group1] port gigabitethernet 1/0/2 downlink
[DeviceB-mtlk-group1] quit
```

4. Configure Device D:

Create VLANs 1 through 30.

```
<DeviceD> system-view
[DeviceD] vlan 1 to 30
```

Configure GigabitEthernet 1/0/1 as a trunk port.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
```

Assign the interface to VLANs 1 through 30.

```
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 1 to 30
```

Enable flush message receiving on the interface.

```
[DeviceD-GigabitEthernet1/0/1] smart-link flush enable
[DeviceD-GigabitEthernet1/0/1] quit
```

Disable the spanning tree feature on GigabitEthernet 1/0/2.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] undo stp enable
```

Configure the interface as a trunk port.

```
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
```

Assign the interface to VLANs 1 through 30.

```
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 1 to 30
```

Enable flush message receiving on the interface.

```
[DeviceD-GigabitEthernet1/0/2] smart-link flush enable
[DeviceD-GigabitEthernet1/0/2] quit
```

Create monitor link group 1.

```
[DeviceD] monitor-link group 1
```

Configure GigabitEthernet 1/0/1 as an uplink interface for monitor link group 1.

```
[DeviceD-mtlk-group1] port gigabitethernet 1/0/1 uplink
```

Configure GigabitEthernet 1/0/2 as a downlink interface for monitor link group 1.

```
[DeviceD-mtlk-group1] port gigabitethernet 1/0/2 downlink
[DeviceD-mtlk-group1] quit
```

Verifying the configuration

When GigabitEthernet 1/0/2 on Device A goes down because of a link fault, verify information about monitor link group 1 on Device B.

```
[DeviceB] display monitor-link group 1
Monitor link group 1 information:
  Group status     : UP
  Downlink up delay: 0(s)
  Last up time     : 16:38:26 2012/4/21
  Last down time   : 16:37:20 2012/4/21
  Up port threshold: 1

  Member                    Role       Status
  ----------------------------------------------
  GE1/0/1                   UPLINK     UP
  GE1/0/2                   DOWNLINK   UP
```

Verify information about monitor link group 1 on Device D.

```
[DeviceD] display monitor-link group 1
Monitor link group 1 information:
  Group status     : DOWN
  Downlink up delay: 0(s)
  Last up time     : 16:37:20 2012/4/21
  Last down time   : 16:38:26 2012/4/21
  Up port threshold: 1

  Member                    Role       Status
  ----------------------------------------------
  GE1/0/1                   UPLINK     DOWN
  GE1/0/2                   DOWNLINK   DOWN (Monitor Link)
```

Contents

Configuring error code detection 1

About error code detection 1
Error code detection fundamentals 1
Interface error code detection 1
Link quality detection 1
Background traffic 2
Error code detection deployment 2
Restrictions and guidelines: Error code detection configuration 3
Configuring interface error code detection 3
Configuring link quality detection 3
Configuring the bit error ratio calculation factors for an interface 4
Specifying the protocol for transferring error code detection information 5
Display and maintenance commands for error code detection 5

Configuring error code detection

About error code detection

As a high availability mechanism, error code detection triggers the upper-layer service module to automatically take protection actions when error code events occur. This feature reduces the influence of error codes on the service.

Error code detection fundamentals

Interface error code detection

With this feature configured on an interface, the device periodically samples, calculates, and detects the packets received on the interface.

- If the bit error ratio exceeds the upper threshold, an error code failure occurs on the interface. Then, an error code event is reported to trigger the upper-layer service module to take the corresponding protection action.
- If the bit error ratio drops below the lower threshold, the error code failure is cleared on the interface. Then, an error code clearing event is reported to trigger the upper-layer service module to cancel the protection action.

Interface error code detection supports the following protection actions:

- **Trigger LSP**—Error code failures on an interface can trigger LSP switchover. An error code failure occurring on an interface is reported to the upper-layer service module to switch traffic from the primary LSP to the secondary LSP. After the error code failure is cleared, traffic is switched back to the primary LSP.
- **Trigger Section**—Error code failures on an interface can trigger section layer service protection. An error code failure occurring on an interface is reported to the interface management module to set the link layer protocol status of the interface to **DOWN(Bit-error down)**. Then, the upper-layer service module associated with the interface triggers section layer service protection. After the error code failure is cleared, the link layer protocol status of the interface restores to UP and triggers the upper-layer service module to cancel section layer service protection.

Link quality detection

With this feature configured on an interface, the device periodically samples, calculates, and detects the packets received on the interface.

- If the bit error ratio exceeds the upper threshold, an error code failure occurs on the interface. Then, the link quality of the interface becomes LOW, and an error code event is reported. In this case, IGP (for example, OSPF and IS-IS) increases the cost for the link and recalculates the routes, so that IGP does not preferentially select the link and keeps off error code failures.
- If the bit error ratio drops below the lower threshold, the error code failure is cleared on the interface. Then, the link quality of the interface restores to GOOD, and an error code clearing event is reported. In this case, IGP restores the cost value for the interface and recalculates the routes, so that IGP can preferentially select the link again.

Background traffic

To guarantee the error code detection accuracy on an interface, make sure the interface has enough traffic. For error code detection to take effect on an interface with no or little service traffic, sending a type of specific traffic must be automatically started or stopped according to the service traffic size on the interface. In this way, the error code failures can be accurately located on the interface. This type of traffic is called background traffic.

After an error code failure occurs on an interface, the error code failure on the interface cannot be automatically recovered if no service traffic exists on the interface. To solve this problem, the device will automatically enable the background traffic sending function when error code detection is configured on an interface. The background traffic contains 64-byte all-0 packets with the lowest priority. Therefore, the background traffic can provide enough samples for error code detection without affecting services, so that the error code failure can be fast recovered

Background traffic sending is automatically started or stopped according to the service traffic size as follows:

- When the interface speed is greater than or equal to 1000 Mbps and the outgoing traffic rate is smaller than 100 Mbps, background traffic sending automatically starts. The interface will automatically send background traffic at 300 Mbps.
- When the outgoing traffic rate of an interface exceeds 300 Mbps, the interface will automatically stop sending background traffic.

Error code detection deployment

In the current software version, you can deploy error code detection only on Ethernet interfaces and aggregate interfaces.

- With this feature configured on an Ethernet interface, this feature samples, calculates, and detects the bit error ratio of the Ethernet interface, and the bit error ratio of the Ethernet interface can trigger an error code event.
- With this feature configured on an aggregate interface, this feature samples, calculates, and detects the bit error ratios for the member ports of the aggregate interface. Interface error code detection and link quality detection are different in how to trigger error code events.

Interface error code detection deployment on an aggregate interface

An aggregate interface goes down when one of the following conditions exist:

- Error code failures occur on all member ports of the aggregate interface.
- The number of Selected member ports without error code failures is smaller than the minimum number of Selected member ports required for bringing up the aggregate interface.

Then, the upper-layer service module of the aggregate interface is triggered to perform the protection action.

When the number of Selected member ports without error code failures reaches the minimum number of Selected member ports required for bringing up the aggregate interface, the aggregate interface comes up, and the upper-layer service module of the aggregate interface cancels the protection action.

For more information about the minimum number of Selected member ports required for bringing up an aggregate interface, see Ethernet link aggregation configuration in *Layer 2—LAN Switching Configuration Guide*.

Link quality detection deployment on an aggregate interface

An aggregate interface goes down and the upper-layer service module is triggered to perform the protection action when the following conditions exist:

- Error code failures occur on all member ports of the aggregate interface.

- The minimum number of Selected member ports required for bringing up the aggregate interface is not configured.

The status of an aggregate interface is still up but the link quality of the aggregate interface will become LOW when the number of Selected member ports without error code failures is smaller than the minimum number of Selected member ports required for bringing up the aggregate interface. In this case, IGP (for example, OSPF and IS-IS) increases the cost for the link and recalculates the routes, so that IGP does not preferentially select the link and keeps off error code failures.

When the number of Selected member ports without error code failures reaches the minimum number of Selected member ports required for bringing up the aggregate interface, the link quality of the aggregate interface restores to GOOD. In this case, IGP restores the cost value for the interface and recalculates the routes, so that IGP can preferentially select the link again.

Restrictions and guidelines: Error code detection configuration

The interface error code detection feature and the link quality detection feature are mutually exclusive. Select a proper detection method according to the scenario.

Configuring interface error code detection

Restrictions and guidelines

The **trigger-lsp** keyword is not supported on a Layer 2 Ethernet or aggregate interface.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 3 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Configure error code detection parameters for the interface.
   port ifmonitor crc-error bit-error-ratio high-threshold *high-value-coefficient high-value-power* **low-threshold** *low-value-coefficient low-value-power* { **trigger-lsp** | **trigger-section** }

   By default, no error code detection parameters are configured for an interface.

Configuring link quality detection

1. Enter system view.
   system-view

2. Enter interface view.
   - Enter Layer 3 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Configure link quality detection parameters.
   link-quality crc-error bit-error-ratio high-threshold *high-value-coefficient high-value-power* **low-threshold** *low-value-coefficient low-value-power*

   By default, no link quality detection parameters are configured.

Configuring the bit error ratio calculation factors for an interface

About this task

With error code detection enabled, the device periodically samples and checks the packets sent and received by an interface and performs the following operations:

- When the interface is in lower bit error ratio status, the device uses the upper bit error ratio calculation factor to calculate a result. Then, the device compares the calculation result with the upper bit error ratio threshold to determine whether the interface is in upper bit error ratio status.
- When the interface is in upper bit error ratio status, the device uses the lower bit error ratio calculation factor to calculate a result. Then, the device compares the calculation result with the lower bit error ratio threshold to determine whether the interface is in lower bit error ratio status.

You can use this feature to adjust the bit error ratio calculation factors to adjust the timeliness and reliability of the calculation results. A smaller calculation factor means a higher timeliness and lower reliability.

Restrictions and guidelines

This command takes effect only when the **link-quality crc-error bit-error-ratio** or **port ifmonitor crc-error bit-error-ratio** command is executed.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 3 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Configure the bit error ratio calculation factors for the interface.
   crc-error bit-error-ratio algorithm-parameter high-factor *high-factor-value* **low-factor** *low-factor-value*

   By default, the calculation factor is 50 for both the upper bit error ratio and lower bit error ratio.

Specifying the protocol for transferring error code detection information

About this task

When an error code failure occurs on an interface, the error code status of the interface must be sent to the peer device. The device supports using BFD or LLDP for transferring the error code status of interfaces. When the device is connected to a third-party device, you can select a protocol for transferring error code information, so that the peer device can recognize the error code information.

Restrictions and guidelines

To use LLDP for transferring error code detection information, make sure LLDP is enabled and LLDP operates in nearest bridge agent mode. For more information about LLDP, see LLDP configuration in *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface.
     interface *interface-type interface-number*
   - Enter Layer 3 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
   - Enter Layer 3 aggregate interface view.
     interface route-aggregation *interface-number*
3. Specify the protocol for transferring error code detection information.
   crc-error bit-error-ratio transfer-protocol { **bfd** | **lldp** }
   By default, BFD is used for transferring error code detection information.

Display and maintenance commands for error code detection

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display background traffic information.	**display ifmonitor background-traffic interface** [*interface-type* [*interface-number*]]
Display alarm information of an interface.	**display ifmonitor interface** *interface-type interface-number*
Clear the bit error ratio status for interfaces.	**reset crc-error bit-error-ratio status interface** [*interface-type* [*interface-number*]]

Contents

Configuring VRRP 1

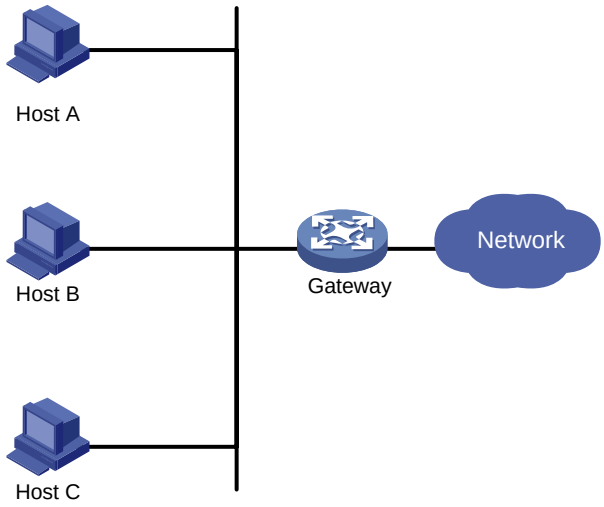
About VRRP 1
VRRP group 1
VRRP standard mode 2
VRRP networking 2
Virtual IP address and IP address owner 2
Router priority in a VRRP group 2
Preemption 2
Authentication method 3
VRRP timers 3
Master election 4
VRRP tracking 4
VRRP application 4
VRRP load balancing mode 6
Virtual MAC address assignment 6
Virtual forwarder 8
Protocols and standards 10
Configuring IPv4 VRRP 10
Restrictions and guidelines: IPv4 VRRP configuration 10
IPv4 VRRP tasks at a glance 10
Specifying an IPv4 VRRP operating mode 10
Specifying the IPv4 VRRP version 11
Configuring an IPv4 VRRP group 11
Configuring the virtual MAC address for an IPv4 VRRP group 13
Configuring IPv4 VRRP packet attributes 14
Setting the packet sending mode for IPv4 VRRPv3 15
Enabling periodic sending of gratuitous ARP packets for IPv4 VRRP 16
Configuring a subordinate IPv4 VRRP group to follow a master IPv4 VRRP group 16
Setting the state transition delay timer for an IPv4 VRRP group 17
Enabling SNMP notifications for VRRP 18
Enabling the isolation mode for IPv4 VRRP 18
Display and maintenance commands for IPv4 VRRP 18
Configuring IPv6 VRRP 19
Restrictions and guidelines: IPv6 VRRP configuration 19
IPv6 VRRP tasks at a glance 19
Specifying an IPv6 VRRP operating mode 19
Configuring an IPv6 VRRP group 20
Configuring the virtual MAC address for an IPv6 VRRP group 22
Configuring IPv6 VRRP packet attributes 22
Enabling periodic sending of ND packets for IPv6 VRRP 23
Configuring a subordinate IPv6 VRRP group to follow a master IPv6 VRRP group 23
Setting the state transition delay timer for an IPv6 VRRP group 24
Enabling the isolation mode for IPv6 VRRP 25
Display and maintenance commands for IPv6 VRRP 25
IPv4 VRRP configuration examples 26
Example: Configuring a single VRRP group 26
Example: Configuring multiple VRRP groups 29
IPv6 VRRP configuration examples 32
Example: Configuring a single VRRP group 32
Example: Configuring multiple VRRP groups 35
Troubleshooting VRRP 38
An error prompt is displayed 38
Multiple masters appear in a VRRP group 39
Fast VRRP state flapping 39

Configuring VRRP

About VRRP

Typically, you can configure a default gateway for every host on a LAN. All packets destined for other networks are sent through the default gateway. As shown in Figure 1, when the default gateway fails, no hosts can communicate with external networks.

Figure 1 LAN networking



Using a default gateway facilitates your configuration but requires high availability. Using more egress gateways improves link availability but introduces the problem of routing among the egresses.

Virtual Router Redundancy Protocol (VRRP) is designed to address this issue. VRRP adds a group of network gateways to a VRRP group called a virtual router. The VRRP group has one master and multiple backups, and provides a virtual IP address. The hosts on the subnet use the virtual IP address as their default network gateway to communicate with external networks.

VRRP avoids single points of failure and simplifies the configuration on hosts. When the master in the VRRP group on a multicast or broadcast LAN (for example, an Ethernet network) fails, another router in the VRRP group takes over. The switchover is complete without causing dynamic route recalculation, route re-discovery, gateway reconfiguration on the hosts, or traffic interruption.

VRRP has two versions: VRRPv2 and VRRPv3. VRRPv2 supports IPv4 VRRP. VRRPv3 supports IPv4 VRRP and IPv6 VRRP.

VRRP group

VRRP adds a group of network gateways to a VRRP group called a virtual router. The VRRP group has one master and multiple backups, and provides a virtual IP address. The hosts on the subnet use the virtual IP address as their default network gateway to communicate with external networks.

The administrator can add a router to a VRRP group by creating the VRRP group on a Layer 3 interface on the router.

NOTE:

On a device, the interfaces added to VRRP groups having the same VRRP group number belong to

different VRRP groups.

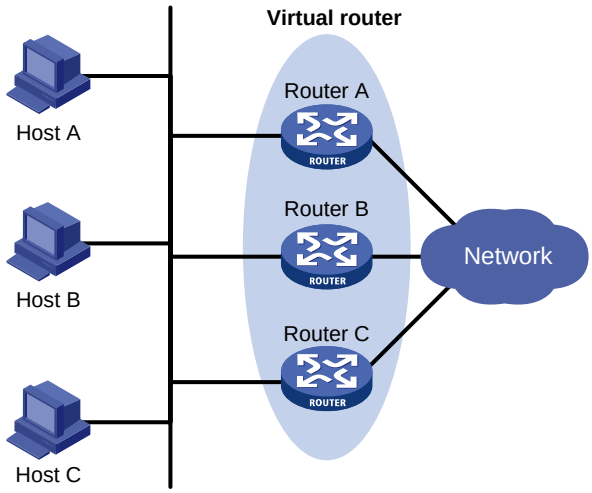
VRRP standard mode

VRRP networking

As shown in Figure 2, Router A, Router B, and Router C form a virtual router, which has a manually configured virtual IP address. Hosts on the subnet use the virtual router as the default gateway.

The router with the highest priority among the three routers is elected as the master, and the other two are backups. Only the master in the VRRP group can provide gateway service. When the master fails, the backup routers elect a new master to take over for nonstop gateway service.

Figure 2 VRRP networking



Virtual IP address and IP address owner

The virtual IP address of the virtual router can be either of the following IP addresses:

- Unused IP address on the subnet where the VRRP group resides.
- IP address of an interface on a router in the VRRP group.

In the latter case, the router is called the IP address owner. A VRRP group can have only one IP address owner.

Router priority in a VRRP group

VRRP determines the role (master or backup) of each router in a VRRP group by priority. A router with higher priority is more likely to become the master.

A VRRP priority can be in the range of 0 to 255, and a greater number represents a higher priority. Priorities 1 to 254 are configurable. Priority 0 is reserved for special uses, and priority 255 is for the IP address owner. The IP address owner in a VRRP group always has a running priority of 255 and acts as the master as long as it operates correctly.

Preemption

A router in a VRRP group operates in either non-preemptive mode or preemptive mode.

- **Non-preemptive mode**—The master router acts as the master as long as it operates correctly, even if a backup router is later assigned a higher priority. Non-preemptive mode helps avoid frequent switchover between the master and backup routers.
- **Preemptive mode**—A backup starts a new master election and takes over as master when it detects that it has a higher priority than the current master. Preemptive mode ensures that the router with the highest priority in a VRRP group always acts as the master.

Authentication method

To avoid attacks from unauthorized users, VRRP member routers add authentication keys in VRRP packets to authenticate one another. VRRP provides the following authentication methods:

- Simple authentication

  The sender fills an authentication key into the VRRP packet, and the receiver compares the received authentication key with its local authentication key. If the two authentication keys match, the received VRRP packet is legitimate. Otherwise, the received packet is illegitimate and gets discarded.
- MD5 authentication

  The sender computes a digest for the VRRP packet by using the authentication key and MD5 algorithm, and saves the result to the packet. The receiver performs the same operation with the authentication key and MD5 algorithm, and compares the result with the content in the authentication header. If the results match, the received VRRP packet is legitimate. Otherwise, the received packet is illegitimate and gets discarded.

On a secure network, you can choose to not authenticate VRRP packets.

NOTE:

IPv4 VRRPv3 and IPv6 VRRPv3 do not support VRRP packet authentication.

VRRP timers

Skew_Time

Skew_Time helps avoid the situation that multiple backups in a VRRP group become the master when the master in the VRRP group fails.

Skew_Time is not configurable; its value depends on the VRRP version.

- In VRRPv2 (described in RFC 3768), Skew_Time is (256 – Router priority)/256.
- In VRRPv3 (described in RFC 5798), Skew_Time is ((256 – Router priority) × VRRP advertisement interval)/256.

VRRP advertisement interval

The master in a VRRP group periodically sends VRRP advertisements to declare its presence.

You can configure the interval at which the master sends VRRP advertisements. If a backup does not receive any VRRP advertisement when the timer (3 × VRRP advertisement interval + Skew_Time) expires, it takes over as the master.

VRRP preemption delay timer

You can configure the VRRP preemption delay timer for the following purposes:

- Avoid frequent state changes among members in a VRRP group.
- Provide the backups with enough time to collect information (such as routing information).

In preempt mode, a backup does not immediately become the master after it receives an advertisement with lower priority than the local priority. Instead, it waits for a period of time (preemption delay time + Skew_Time) before taking over as the master.

Master election

Routers in a VRRP group determine their roles by priority. When a router joins a VRRP group, it has a backup role. The router role changes according to the following situations:

- If the backup does not receive any VRRP advertisement when the timer (3 × advertisement interval + Skew_Time) expires, it becomes the master.
- If the backup receives a VRRP advertisement with the same or greater priority within the timer (3 × advertisement interval + Skew_Time), it remains a backup.
- If the backup receives a VRRP advertisement with a smaller priority within the timer (3 × advertisement interval + Skew_Time), the following results apply:
  - It remains a backup when operating in non-preemptive mode.
  - It becomes the master when operating in preemptive mode.

The elected master starts a VRRP advertisement interval to periodically send VRRP advertisements to notify the backups that it is operating correctly. Each of the backups starts a timer to wait for advertisements from the master.

When multiple routers in a VRRP group declare that they are the master because of network problems, the one with the highest priority becomes the master. If two routers have the same priority, the one with the highest IP address becomes the master.

VRRP tracking

The VRRP tracking function uses network quality analyzer (NQA) or bidirectional forwarding detection (BFD) to monitor the state of the master or the upstream link. The collaboration between VRRP and NQA or BFD through a track entry implements the following functions:

- Monitors the upstream link and changes the priority of the router according to the state of the link. If the upstream link fails, the hosts on the subnet cannot access external networks through the router and the state of the track entry becomes Negative. The priority of the master decreases by a specified value, and a router with a higher priority in the VRRP group becomes the master. The switchover ensures uninterrupted communication between the hosts on the subnet and external networks.
- Monitors the state of the master on the backups. When the master fails, a backup immediately takes over to ensure uninterrupted communication.

When the track entry changes from Negative to Positive or Notready, the router automatically restores its priority. For more information about track entries, see "Configuring Track."

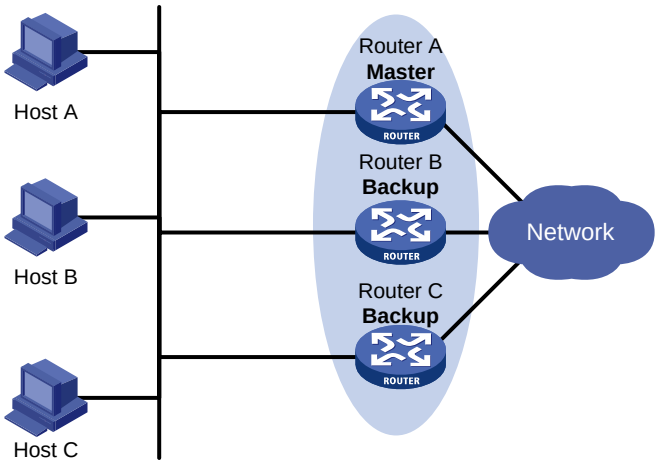
To enable VRRP tracking, configure the routers in the VRRP group to operate in preemptive mode first. This configuration ensures that only the router with the highest priority operates as the master.

VRRP application

Master/backup

In master/backup mode, only the master forwards packets, as shown in Figure 3. When the master fails, a new master is elected from among the backups. This mode requires only one VRRP group, and each router in the group has a different priority. The one with the highest priority becomes the master.

Figure 3 VRRP in master/backup mode



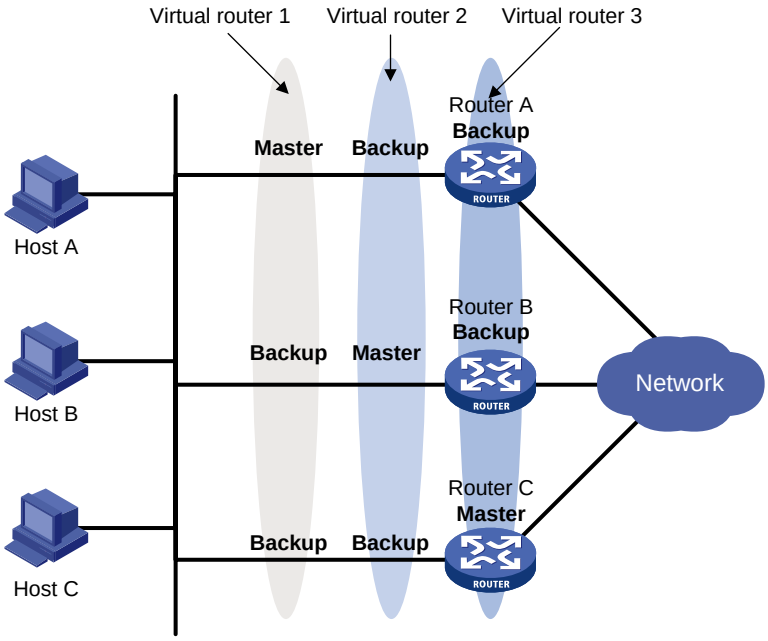
Assume that Router A is acting as the master to forward packets to external networks, and Router B and Router C are backups in listening state. When Router A fails, Router B and Router C elect a new master to forward packets for hosts on the subnet.

Load sharing

A router can join multiple VRRP groups. With different priorities in different VRRP groups, the router can act as the master in one VRRP group and a backup in another.

In load sharing mode, multiple VRRP groups provide gateway services. This mode requires a minimum of two VRRP groups, and each group has one master and multiple backups. The master roles in the VRRP groups are assumed by different routers, as shown in Figure 4.

Figure 4 Load sharing of VRRP



A router can be in multiple VRRP groups and have a different priority in each group.

As shown in Figure 4, the following VRRP groups exist:

- **VRRP group 1**—Router A is the master. Router B and Router C are the backups.
- **VRRP group 2**—Router B is the master. Router A and Router C are the backups.

- **VRRP group 3**—Router C is the master. Router A and Router B are the backups.

To implement load sharing among Router A, Router B, and Router C, perform the following tasks:

- Configure the virtual IP addresses of VRRP group 1, 2, and 3 as default gateway IP addresses for hosts on the subnet.
- Assign the highest priority to Router A, B, and C in VRRP group 1, 2, and 3, respectively.

VRRP load balancing mode

In a standard-mode VRRP group, only the master can forward packets and backups are in listening state. You can create multiple VRRP groups to share traffic, but you must configure different gateways for hosts on the subnet.

In load balancing mode, a VRRP group maps its virtual IP address to multiple virtual MAC addresses, assigning one virtual MAC address to each member router. Every router in this VRRP group can forward traffic and respond to IPv4 ARP requests or IPv6 ND requests from hosts. Because their virtual MAC addresses are different, traffic from hosts is distributed across the VRRP group members. Load balancing mode simplifies configuration and improves forwarding efficiency.

VRRP load balancing mode uses the same master election, preemption, and tracking mechanisms as the standard mode. New mechanisms have been introduced to VRRP load balancing mode, as described in the following sections.

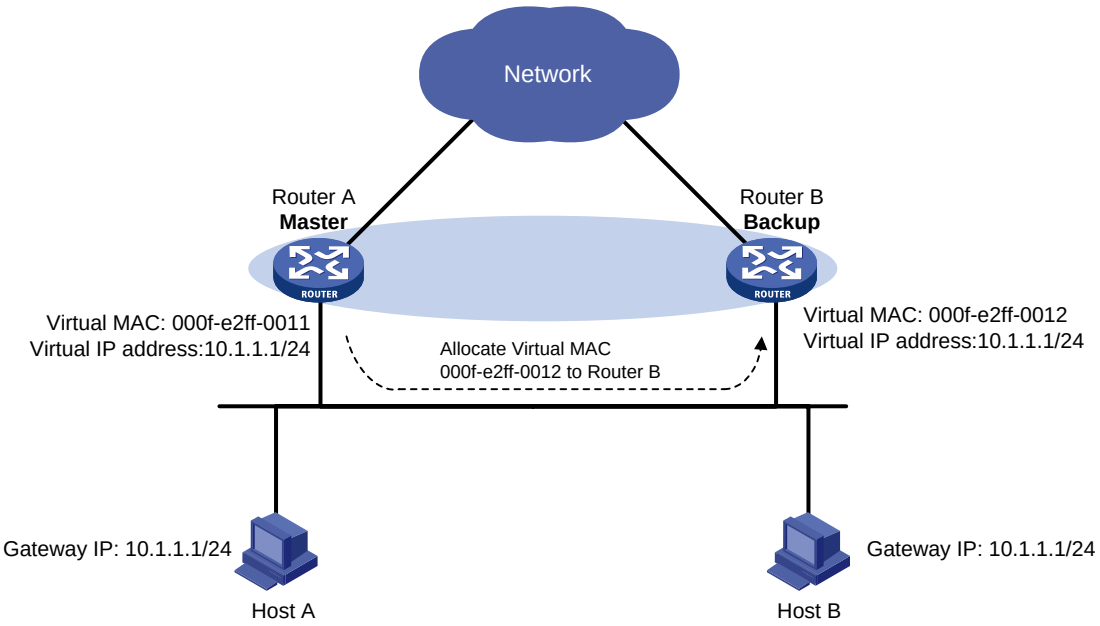
Virtual MAC address assignment

In load balancing mode, the master assigns virtual MAC addresses to routers in the VRRP group. The master uses different MAC addresses to respond to ARP requests or ND requests from different hosts. The backup routers, however, do not answer ARP requests or ND requests from hosts.

In an IPv4 network, a load balanced VRRP group works as follows:

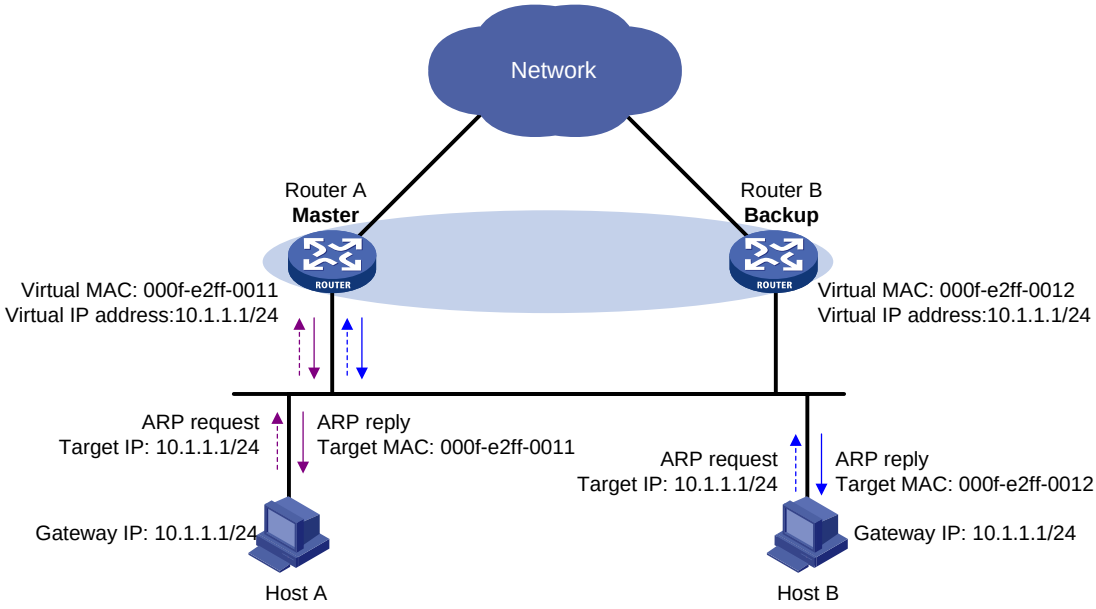
1. The master assigns virtual MAC addresses to all member routers, including itself. This example assumes that the virtual IP address of the VRRP group is 10.1.1.1/24, Router A is the master, and Router B is the backup. Router A assigns 000f-e2ff-0011 for itself and 000f-e2ff-0012 for Router B. See Figure 5.

Figure 5 Virtual MAC address assignment



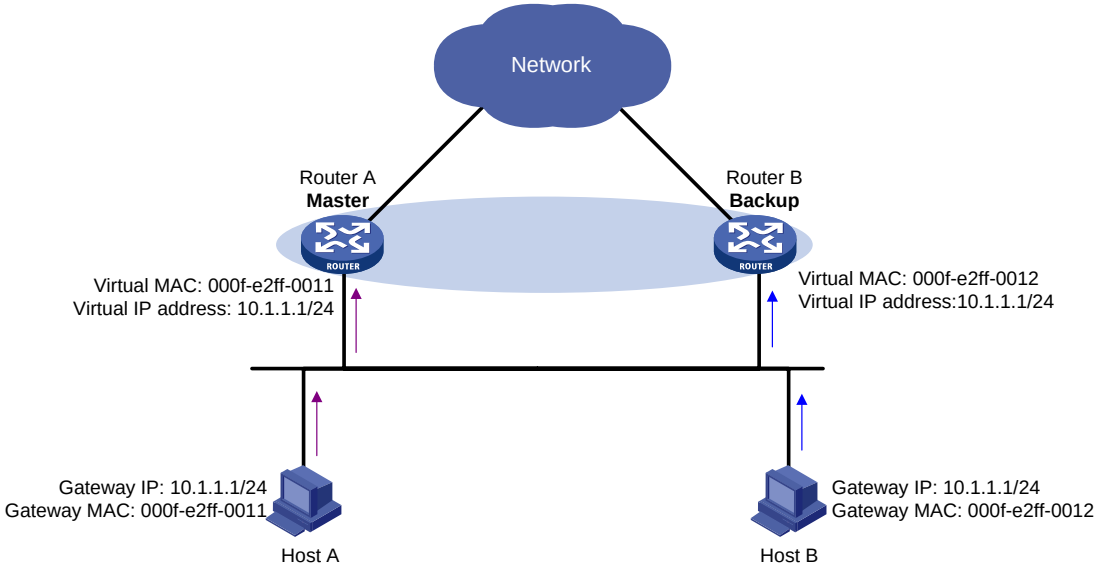
2. When an ARP request arrives, the master (Router A) selects a virtual MAC address based on the load balancing algorithm to answer the ARP request. In this example, Router A returns the virtual MAC address of itself in response to the ARP request from Host A. Router A returns the virtual MAC address of Router B in response to the ARP request from Host B. See Figure 6.

Figure 6 Answering ARP requests



3. Each host sends packets to the returned MAC address. As shown in Figure 7, Host A sends packets to Router A and Host B sends packets to Router B.

Figure 7 Sending packets to different routers for forwarding



In the ARP reply sent by the master, the source MAC address in the Ethernet header is different from the sender MAC address in the message body. For the Layer 2 device to forward the ARP packet, follow these configuration guidelines on the Layer 2 device:

- Do not enable ARP packet source MAC address consistency check.
- Do not specify the **src-mac** keyword when you enable ARP packet validity check for ARP detection.

For more information about ARP packet source MAC address consistency check and ARP detection, see *Security Configuration Guide*.

Virtual forwarder

Virtual forwarder creation

Virtual MAC addresses enable traffic distribution across routers in a VRRP group. To enable routers in the VRRP group to forward packets, VFs must be created on them. Each VF is associated with a virtual MAC address in the VRRP group and forwards packets that are sent to this virtual MAC address.

VFs are created on routers in a VRRP group, as follows:

1. The master assigns virtual MAC addresses to all routers in the VRRP group. Each member router creates a VF for this MAC address and becomes the owner of this VF.
2. Each VF owner advertises its VF information to the other member routers.
3. After receiving the VF advertisement, each of the other routers creates the advertised VF.

Eventually, every member router maintains one VF for each virtual MAC address in the VRRP group.

VF weight and priority

The weight of a VF indicates the forwarding capability of a VF. A higher weight means higher forwarding capability. When the weight is lower than the lower limit of failure, the VF cannot forward packets.

The priority of a VF determines the VF state. Among the VFs created on different member routers for the same virtual MAC address, the VF with the highest priority is in active state. This VF, known as the active virtual forwarder (AVF), forwards packets. All other VFs listen to the state of the AVF and are known as the listening virtual forwarders (LVFs). VF priority is in the range of 0 to 255, where 255 is reserved for the VF owner. When the weight of a VF owner is higher than or equal to the lower limit of failure, the priority of the VF owner is 255.

The priority of a VF is calculated based on its weight.

- If the VF weight is higher than or equal to the lower limit of failure, the following VF priorities apply:
  - On a VF owner, the VF priority is 255.
  - On a non-VF owner, the VF priority is calculated as weight/(number of local AVFs + 1).
- If the VF weight is lower than the lower limit of failure, the VF priority is 0.

VF backup

The VFs corresponding to a virtual MAC address on different routers in the VRRP group back up one another.

Figure 8 VF information

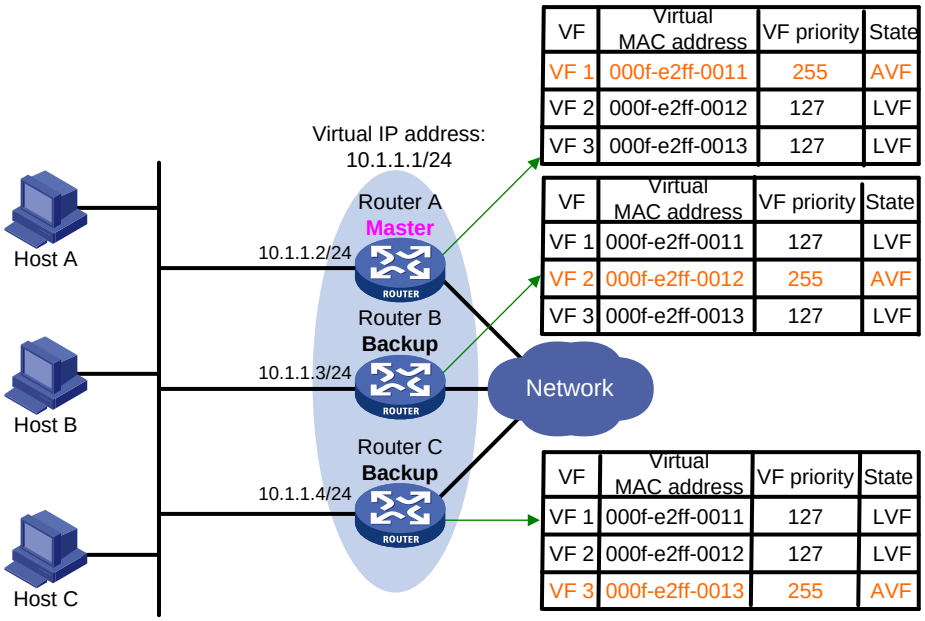


Figure 8 shows the VF table on each router in the VRRP group and how the VFs back up one another. The master, Router A, assigns virtual MAC addresses 000f-e2ff-0011, 000f-e2ff-0012, and 000f-e2ff-0013 to itself, Router B, and Router C, respectively. Each router creates VF 1, VF 2, and VF 3 for virtual MAC addresses 000f-e2ff-0011, 000f-e2ff-0012, and 000f-e2ff-0013, respectively. The VFs for the same virtual MAC address on different routers back up one another. For example, the VF 1 instances on Router A, Router B, and Router C back up one another.

- The VF 1 instance on Router A (the VF 1 owner) has priority 255. It acts as the AVF to forward packets sent to virtual MAC address 000f-e2ff-0011.
- The VF 1 instances on Router B and Router C have a priority of 255/(1 + 1), or 127. Because their priorities are lower than the priority of the VF 1 instance on Router A, they act as LVFs. These LVFs listen to the state of the VF 1 instance on Router A.
- When the VF 1 instance on Router A fails, the VF 1 instances on Router B and Router C elect the one with higher priority as the new AVF. This AVF forwards packets destined for virtual MAC address 000f-e2ff-0011. If the two LVFs' priorities are the same, the LVF with a greater device MAC address becomes the new AVF.

A VF always operates in preemptive mode. When an LVF finds its priority value higher than the one advertised by the AVF, the LVF declares itself as the AVF.

VF timers

When the AVF on a router fails, the new AVF on another router creates the following timers for the failed AVF:

- **Redirect timer**—Before this timer expires, the master still uses the virtual MAC address corresponding to the failed AVF to respond to ARP/ND requests from hosts. The VF owner can share traffic load if the VF owner resumes normal operation within this time. When this timer expires, the master stops using the virtual MAC address corresponding to the failed AVF to respond to ARP/ND requests from hosts.
- **Timeout timer**—The duration after which the new AVF takes over responsibilities of the failed VF owner. Before this timer expires, all routers in the VRRP group keep the VFs that correspond to the failed AVF. The new AVF forwards packets destined for the virtual MAC address of the failed AVF. When this timer expires, all routers in the VRRP group remove the VFs that correspond to the failed AVF, including the new AVF. Packets destined for the virtual MAC address of the failed AVF are not forwarded any longer.

VF tracking

An AVF forwards packets destined for the MAC address of the AVF. If the AVF's upstream link fails but no LVF takes over, the hosts that use the AVF's MAC address as their gateway MAC address cannot access the external network.

The VF tracking function can solve this problem. You can use NQA or BFD to monitor the upstream link state of the VF owner, and associate the VFs with NQA or BFD through the tracking function. This enables the collaboration between VRRP and NQA or BFD through the Track module. When the upstream link fails, the state of the track entry changes to Negative. The weights of the VFs (including the AVF) on the router decrease by a specific value. The corresponding LVF with a higher priority on another router becomes the AVF and forwards packets.

Protocols and standards

- RFC 3768, *Virtual Router Redundancy Protocol (VRRP)*
- RFC 5798, *Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6*

Configuring IPv4 VRRP

Restrictions and guidelines: IPv4 VRRP configuration

IPv4 VRRP does not take effect on member ports of aggregation groups.

Configuration on the routers in an IPv4 VRRP group must be consistent.

Each IPv4 VRRP group corresponds to a virtual MAC address. The maximum number of IPv4 VRRP groups supported on an interface depends on the maximum number of virtual MAC addresses supported on the interface.

IPv4 VRRP tasks at a glance

To configure IPv4 VRRP, perform the following tasks:

1. Specifying an IPv4 VRRP operating mode
2. (Optional.) Specifying the IPv4 VRRP version
3. Configuring an IPv4 VRRP group
4. (Optional.) Configuring the virtual MAC address for an IPv4 VRRP group
5. (Optional.) Configuring IPv4 VRRP packet attributes
6. (Optional.) Setting the packet sending mode for IPv4 VRRPv3
7. (Optional.) Enabling periodic sending of gratuitous ARP packets for IPv4 VRRP
8. (Optional.) Configuring a subordinate IPv4 VRRP group to follow a master IPv4 VRRP group
9. (Optional.) Setting the state transition delay timer for an IPv4 VRRP group
10. (Optional.) Enabling SNMP notifications for VRRP
11. (Optional.) Enabling the isolation mode for IPv4 VRRP

Specifying an IPv4 VRRP operating mode

Restrictions and guidelines

After an IPv4 VRRP operating mode is configured on a router, all IPv4 VRRP groups on the router operate in the specified operating mode.

When you change the VRRP group operating mode from standard to load balancing, follow these restrictions and guidelines:

- The VRRP virtual MAC address will also change.
- The original VRRP virtual MAC address in the ARP table of the neighbor device is not refreshed until the ARP entry ages out.

Before the VRRP virtual MAC address is refreshed, temporary service traffic interruption might occur. As a best practice, perform such operations during off-peak hours to minimize service impact.

Procedure

1. Enter system view.
   system-view
2. Specify an IPv4 VRRP operating mode.
   - Specify the standard mode.
   - **undo vrrp mode**
   - Specify the load balancing mode.
   - **vrrp mode load-balance** [**version-8**]

   By default, VRRP operates in standard mode.

Specifying the IPv4 VRRP version

About this task

IPv4 VRRP can use VRRPv2 and VRRPv3.

Restrictions and guidelines

For an IPv4 VRRP group to operate correctly, make sure the same VRRP version is used on all routers in the IPv4 VRRP group.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the version of VRRP.
   vrrp version *version-number*

   By default, VRRPv3 is used.

Configuring an IPv4 VRRP group

About this task

A VRRP group can operate correctly after you create it and assign a minimum of one virtual IP address to it. You can configure multiple virtual IP addresses for the VRRP group on an interface that connects to multiple subnets for router backup on different subnets.

If you disable an IPv4 VRRP group, the VRRP group enters **Initialize** state, and the existing configuration on the VRRP group remains unchanged. You can modify the configuration of the VRRP group. The modification takes effect when you enable the VRRP group again.

Restrictions and guidelines

Item	Remarks
Maximum number of VRRP groups and virtual IP addresses	In VRRP load balancing mode, the device supports a maximum of *MaxVRNum/N* VRRP groups. *MaxVRNum* refers to the maximum number of VRRP groups supported by the device in VRRP standard mode. *N* refers to the number of devices in the VRRP group.
Virtual IP address	When VRRP is operating in standard mode, the virtual IP address of a VRRP group can be either of the following addresses: • Unused IP address on the subnet where the VRRP group resides. • IP address of an interface on a router in the VRRP group. An IPv4 VRRP group without virtual IP addresses configured can exist on a device provided that other settings (for example, priority and preemption mode) are available. Such a VRRP group stays in inactive state and does not function. For hosts in the subnet to access external networks, as a best practice, configure the following addresses in the same subnet: • Virtual IP address of an IPv4 VRRP group. • Downlink interface IP addresses of the VRRP group members.
IP address owner	On an IP address owner, as a best practice, do not use the **network** command to enable OSPF on the interface owning the virtual IP address of the VRRP group. For more information about the **network** command, see *Layer 3—IP Routing Command Reference*. After you remove the VRRP group on an IP address owner, a backup in the VRRP group takes over as the master. The actual IP address of the interface on the IP address owner will conflict with the virtual IP address of the new master. To avoid the conflict, change the IP address of the interface on the IP address owner before you remove the VRRP group from the interface. The running priority of an IP address owner is always 255, and you do not need to configure it. An IP address owner always operates in preemptive mode. If you configure the **vrrp vrid track priority reduced** or **vrrp vrid track switchover** command on an IP address owner, the configuration does not take effect until the router becomes a non-IP address owner.
VRRP association with a track entry	When the track entry changes from Negative to Positive or Notready, the router automatically restores its priority or the failed master router becomes the master again.

Creating a VRRP group and assigning a virtual IP address

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Create a VRRP group and assign a virtual IP address.
   vrrp vrid *virtual-router-id* **virtual-ip** *virtual-address*

Configuring an IPv4 VRRP group

1. Enter system view.
   system-view
2. Enter interface view.

interface *interface-type interface-number*

3. Set the priority of the router in the VRRP group.

   vrrp vrid *virtual-router-id* **priority** *priority-value*

   The default setting is 100.

4. Enable the preemptive mode for the router in a VRRP group and set the preemption delay time.

   vrrp vrid *virtual-router-id* **preempt-mode** [**delay** *delay-value*]

   By default, the router in a VRRP group operates in preemptive mode and the preemption delay time is 0 centiseconds, which means an immediate preemption.

5. Associate a VRRP group with a track entry.

   vrrp vrid *virtual-router-id* **track** *track-entry-number* { **forwarder-switchover member-ip** *ip-address* | **priority reduced** [*priority-reduced*] | **switchover** | **weight reduced** [*weight-reduced*] }

   By default, a VRRP group is not associated with any track entries.

Disabling an IPv4 VRRP group

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Disable a VRRP group.

   vrrp vrid *virtual-router-id* **shutdown**

   By default, an IPv4 VRRP group is enabled.

CAUTION:

You can use this command to temporarily disable an IPv4 VRRP group. After you execute this command, packets sent to the IPv4 VRRP group might be discarded.

Configuring the virtual MAC address for an IPv4 VRRP group

About this task

By default, after the network administrator creates an IPv4 VRRP group, the device automatically generates a virtual MAC address for the IPv4 VRRP group based on the VRRP protocol requirements for communication purposes. After you configure this feature, the device will set the virtual MAC address of the IPv4 VRRP group to the specified value. This can meet the requirements in specific scenarios (such as in a static ARP network).

Restrictions and guidelines

This feature takes effect only when the local device acts as the master in the IPv4 VRRP group.

The virtual MAC addresses of the IPv4 VRRP group and IPv6 VRRP group deployed on an interface cannot be the same.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Configure the virtual MAC address for the IPv4 VRRP group.

vrrp vrid *virtual-router-id* **virtual-mac** *virtual-mac*

By default, no virtual MAC address is configured for the IPv4 VRRP group. The device will automatically generate a virtual MAC address for the IPv4 VRRP group.

Configuring IPv4 VRRP packet attributes

Restrictions and guidelines

- You can configure different authentication modes and authentication keys for VRRP groups on an interface. However, members of the same VRRP group must use the same authentication mode and authentication key.
- In VRRPv2, all routers in a VRRP group must have the same VRRP advertisement interval.
- In VRRPv3, authentication mode and authentication key settings do not take effect.
- In VRRPv3, routers in an IPv4 VRRP group can have different intervals for sending VRRP advertisements. The master in the VRRP group sends VRRP advertisements at specified intervals, and carries the interval in the advertisements. After a backup receives the advertisement, it records the interval in the advertisement. If the backup does not receive a VRRP advertisement before the timer (3 x recorded interval + Skew_Time) expires, it regards the master as failed and takes over.
- If you use M-LAG and VRRP together, make sure the keepalive hold timer is shorter than the VRRP advertisement interval. Violation of this restriction might cause a VRRP master/backup switchover to occur before peer link failure is confirmed. To set the keepalive hold timer, use the **m-lag keepalive hold-time** command. For more information about this command, see *Layer 2—LAN Switching Command Reference*.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the authentication mode and authentication key for an IPv4 VRRP group to send and receive VRRP packets.

   vrrp vrid *virtual-router-id* **authentication-mode** { **md5** | **simple** } { **cipher** | **plain** } *string*

   By default, authentication is disabled.
4. Set the interval at which the master in an IPv4 VRRP group sends VRRP advertisements.

   vrrp vrid *virtual-router-id* **timer advertise** *adver-interval*

   The default setting is 100 centiseconds.

   As a best practice to maintain system stability, set the VRRP advertisement interval to be greater than 100 centiseconds.
5. Specify the source interface for receiving and sending VRRP packets.

   vrrp vrid *virtual-router-id* **source-interface** *interface-type interface-number*

   By default, the source interface for receiving and sending VRRP packets is not specified. The interface where the VRRP group resides sends and receives VRRP packets.
6. Enable TTL check for IPv4 VRRP packets.

   vrrp check-ttl enable

   By default, TTL check for IPv4 VRRP packets is enabled.
7. Use the standard algorithm to calculate the checksum of IPv4 VRRP packets.

   vrrp vrid *virtual-router-id* **checksum-standard**

By default, the device uses the private algorithm to calculate the checksum of IPv4 VRRP packets.

Configure this command when the device forms a VRRP group with a device from another vendor. This can avoid checksum check failure and IPv4 VRRP group setup failure when the two ends use different methods to calculate the checksum in IPv4 VRRP packets.

8. Return to system view.

   quit

9. Set a DSCP value for VRRP packets.

   vrrp dscp *dscp-value*

   By default, the DSCP value for VRRP packets is 48.

   The DSCP value identifies the packet priority during transmission.

Setting the packet sending mode for IPv4 VRRPv3

About this task

A router configured with VRRPv3 can process incoming VRRPv2 packets, but a router configured with VRRPv2 cannot process incoming VRRPv3 packets. When the VRRP version of the routers in a VRRP group is changed from VRRPv2 to VRRPv3, multiple masters might be elected in the VRRP group. To resolve the problem, you can set the packet sending mode for IPv4 VRRPv3. This task enables a router configured with VRRPv3 to send VRRPv2 packets and communicate with routers configured with VRRPv2.

Restrictions and guidelines

- The packet sending mode for IPv4 VRRPv3 takes effect only on outgoing VRRP packets. A router configured with VRRPv3 can process incoming VRRPv2 and VRRPv3 packets.
- If you set the packet sending mode for IPv4 VRRPv3 and configure VRRP packet authentication, authentication information will be carried in outgoing VRRPv2 packets but not in outgoing VRRPv3 packets.
- The VRRP advertisement interval is set in centiseconds by using the **vrrp vrid timer advertise** command. The VRRP advertisement interval carried in VRRPv2 packets sent from routers configured with VRRPv3 might be different from the configured value. For information about the VRRP advertisement interval, see the **vrrp vrid timer advertise** command in *High Availability Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Set the packet sending mode for IPv4 VRRPv3.

   vrrp vrid *virtual-router-id* **vrrpv3-send-packet** { **v2-only** | **v2v3-both** }

   By default, a router configured with VRRPv3 sends only VRRPv3 packets.

Enabling periodic sending of gratuitous ARP packets for IPv4 VRRP

About this task

This feature enables the master router in a VRRP group to periodically send gratuitous ARP packets. Then the downstream devices can update the MAC address entry for the virtual MAC address of the VRRP group in a timely manner.

Restrictions and guidelines

- This feature takes effect only in VRRP standard mode.
- If you change the sending interval for gratuitous ARP packets, the configuration takes effect at the next sending interval.
- The master sends the first gratuitous ARP packet at a random time in the second half of the set interval after you execute the **vrrp send-gratuitous-arp** command. This prevents too many gratuitous ARP packets from being sent at the same time.
- The sending interval for gratuitous ARP packets might be much longer than the set interval when the following conditions are met:
  - Multiple VRRP groups exist on the device.
  - A short sending interval is set.

Procedure

1. Enter system view.

   system-view
2. Enable periodic sending of gratuitous ARP packets for IPv4 VRRP.

   vrrp send-gratuitous-arp [**interval** *interval*]

   By default, periodic sending of gratuitous ARP packets is disabled for IPv4 VRRP.

Configuring a subordinate IPv4 VRRP group to follow a master IPv4 VRRP group

About this task

Each VRRP group determines the device role (master or backup) by exchanging VRRP packets among member devices, which might consume excessive bandwidth and CPU resources. To reduce the number of VRRP packets in the network, you can configure a subordinate VRRP group to follow a master VRRP group.

A master VRRP group determines the device role through exchanging VRRP packets among member devices. A VRRP group that follows a master group, called a subordinate VRRP group, does not exchange VRRP packets among its member devices. The state of the subordinate VRRP group follows the state of the master group.

Restrictions and guidelines

- To ensure the master router election, configure the settings such as the router priority, preemptive mode, and tracking function for the master IPv4 VRRP group. The settings are not required for subordinate IPv4 VRRP groups.
- An IPv4 VRRP group cannot be both a master group and a subordinate group.
- An IPv4 VRRP group stays in **Inactive** state if it is configured to follow a nonexistent master group.
- If an IPv4 VRRP group in **Inactive** or **Initialize** state follows a master group that is not in **Inactive** state, the state of the VRRP group does not change.

- A subordinate IPv4 VRRP group does not exchange VRRP packets, which might cause the MAC address entry for its virtual MAC address not to be updated on downstream devices. As a best practice, enable periodic sending of gratuitous ARP packets for IPv4 VRRP by using the **vrrp send-gratuitous-arp** command.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure an IPv4 VRRP group as a master group and assign a name to it.
   vrrp vrid *virtual-router-id* **name** *name*
   By default, an IPv4 VRRP group does not act as a master group.
4. Return to system view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Configure an IPv4 VRRP group to follow a master group.
   vrrp vrid *virtual-router-id* **follow** *name*
   By default, an IPv4 VRRP group does not follow a master VRRP group.

Setting the state transition delay timer for an IPv4 VRRP group

About this task

Typically, an IPv4 VRRP group immediately transitions from **Initialize** to **Master** or **Backup** state when the interface where the IPv4 VRRP group resides comes up. If the device restarts or the interface state changes frequently, the IPv4 VRRP group state might be unstable, and the VRRP group operation might be affected.This task allows you to solve this issue.

Restrictions and guidelines

The delay time for an IPv4 VRRP group to transition from **Initialize** to **Master** or **Backup** state is in the range of 0 to 60 seconds.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the state transition delay timer for the IPv4 VRRP group.
   vrrp state-transition-delay *delay-value*
   By default, the delay time is 0 seconds for the IPv4 VRRP group to transition from **Initialize** to **Master** or **Backup** state.

Enabling SNMP notifications for VRRP

About this task

To report critical VRRP events to an NMS, enable SNMP notifications for VRRP. For VRRP event notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see the network management and monitoring configuration guide for the device.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for VRRP.
   snmp-agent trap enable vrrp [**auth-failure** | **new-master**]
   By default, SNMP notifications for VRRP are enabled.

Enabling the isolation mode for IPv4 VRRP

About this task

To maintain the master in an IPv4 VRRP group, use this feature to perform a master/backup switchover. In isolation mode, the maintenance operation on the device is performed after a master/backup switchover and does not affect services.

Enabling the isolation mode sets the priority of the device to 0 in the IPv4 VRRP group.

Restrictions and guidelines

If you set the priority with the **vrrp vrid priority** command for a device in isolation mode, the setting does not take effect until the isolation mode is disabled.

Procedure

1. Enter system view.
   system-view
2. Enable the isolation mode for IPv4 VRRP.
   vrrp isolate enable
   By default, the isolation mode is disabled for IPv4 VRRP.

Display and maintenance commands for IPv4 VRRP

Execute **display** commands in any view and the **reset** command in user view.

Task	Command
Display states of IPv4 VRRP groups.	**display vrrp** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]] [**verbose**]
Display IPv4 VRRP packet anomaly information.	**display vrrp abnormal-info** [**interface** *interface-type interface-number* [**vrid** *vrid-id*]]
Display master-to-subordinate IPv4 VRRP group bindings.	**display vrrp binding** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*] | **name** *name*]

Task	Command
Display statistics for IPv4 VRRP groups.	**display vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]
Clear statistics for IPv4 VRRP groups.	**reset vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]

Configuring IPv6 VRRP

Restrictions and guidelines: IPv6 VRRP configuration

IPv6 VRRP does not take effect on member ports of aggregation groups.

Configuration on the routers in an IPv6 VRRP group must be consistent.

Each IPv6 VRRP group corresponds to a virtual MAC address. The maximum number of IPv6 VRRP groups supported on an interface depends on the maximum number of virtual MAC addresses supported on the interface.

IPv6 VRRP tasks at a glance

To configure IPv6 VRRP, perform the following tasks:

1. Configuring an IPv6 VRRP group
2. (Optional.) Configuring the virtual MAC address for an IPv6 VRRP group
3. (Optional.) Configuring IPv6 VRRP packet attributes
4. (Optional.) Enabling periodic sending of ND packets for IPv6 VRRP
5. (Optional.) Configuring a subordinate IPv6 VRRP group to follow a master IPv6 VRRP group
6. (Optional.) Setting the state transition delay timer for an IPv6 VRRP group
7. (Optional.) Enabling the isolation mode for IPv6 VRRP

Specifying an IPv6 VRRP operating mode

Restrictions and guidelines

After the IPv6 VRRP operating mode is specified on a router, all IPv6 VRRP groups on the router operate in the specified operating mode.

Procedure

1. Enter system view.

   system-view
2. Specify an IPv6 VRRP operating mode.
   - Specify the standard mode.
   - **undo vrrp ipv6 mode**
   - Specify the load balancing mode.
   - **vrrp ipv6 mode load-balance**

   By default, VRRP operates in standard mode.

Configuring an IPv6 VRRP group

About this task

A VRRP group can work correctly after you create it and assign a minimum of one virtual IPv6 address for it. You can configure multiple virtual IPv6 addresses for the VRRP group on an interface that connects to multiple subnets for router backup.

If you disable an IPv6 VRRP group, the VRRP group enters **Initialize** state, and the existing configuration on the VRRP group remains unchanged. You can modify the configuration of the VRRP group. The modification takes effect when you enable the VRRP group again.

Restrictions and guidelines

Item	Remarks
Maximum number of VRRP groups and virtual IPv6 addresses	In VRRP load balancing mode, the device supports a maximum of *MaxVRNum/N* VRRP groups. *MaxVRNum* refers to the maximum number of VRRP groups supported by the device in VRRP standard mode. *N* refers to the number of devices in the VRRP group.
Virtual IPv6 address	An IPv6 VRRP group without virtual IPv6 addresses configured can exist on a device provided that other settings (for example, priority and preemption mode) are available. Such a VRRP group stays in inactive state and does not function. For hosts in the subnet to access external networks, as a best practice, configure the following addresses in the same subnet: • Virtual IPv6 address of an IPv6 VRRP group. • Downlink interface IPv6 addresses of the VRRP group members.
IP address owner	On an IP address owner, as a best practice, do not use the **ospfv3 area** command to enable OSPF on the interface owning the virtual IPv6 address of the VRRP group. For more information about the **ospfv3 area** command, see *Layer 3—IP Routing Command Reference*. After you remove the IPv6 VRRP group on an IP address owner, a backup in the IPv6 VRRP group takes over as the master. The actual IPv6 address of the interface on the IP address owner will conflict with the virtual IPv6 address of the new master. To avoid the conflict, change the IPv6 address of the interface on the IP address owner before you remove the IPv6 VRRP group from the interface. The running priority of an IP address owner is always 255, and you do not need to configure it. An IP address owner always operates in preemptive mode. If you configure the **vrrp ipv6 vrid track priority reduced** or **vrrp ipv6 vrid track switchover** command on an IP address owner, the configuration does not take effect until the router becomes a non-IP address owner. On an IP address owner, disable Duplicate Address Detection (DAD) on the interface configured with VRRP. To disable DAD, set the *interval* argument to 0 for the **ipv6 nd dad attempts** command. For more information about the command, see IPv6 basics commands in *Layer 3—IP Services Command Reference*.
VRRP association with a track entry	When the track entry changes from Negative to Positive or Notready, the router automatically restores its priority or the failed master router becomes the master again.

Creating a VRRP group and assign a virtual IPv6 address

1. Enter system view.

system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Create a VRRP group and assign a virtual IPv6 address, which is a link-local address.

   vrrp ipv6 vrid *virtual-router-id* **virtual-ip** *virtual-address* **link-local**

   The first virtual IPv6 address that you assign to an IPv6 VRRP group must be a link-local address. It must be the last address you remove. Only one link-local address is allowed in a VRRP group.

Configuring an IPv6 VRRP group

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Assign a virtual IPv6 address, which is a global unicast address.

   vrrp ipv6 vrid *virtual-router-id* **virtual-ip** *virtual-address*

   By default, no global unicast address is assigned to an IPv6 VRRP group.

4. Set the priority of the router in the VRRP group.

   vrrp ipv6 vrid *virtual-router-id* **priority** *priority-value*

   The default setting is 100.

5. Enable the preemptive mode for the router in a VRRP group and set the preemption delay time.

   vrrp ipv6 vrid *virtual-router-id* **preempt-mode** [**delay** *delay-value*]

   By default, the router in a VRRP group operates in preemptive mode and the preemption delay time is 0 centiseconds, which means an immediate preemption.

6. Associate a VRRP group with a track entry.

   vrrp ipv6 vrid *virtual-router-id* **track** *track-entry-number* { **forwarder-switchover member-ip** *ipv6-address* | **priority reduced** [*priority-reduced*] | **switchover** | **weight reduced** [*weight-reduced*] }

   By default, a VRRP group is not associated with any track entries.

Disabling an IPv6 VRRP group

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Disable an IPv6 VRRP group.

   vrrp ipv6 vrid *virtual-router-id* **shutdown**

   By default, an IPv6 VRRP group is enabled.

CAUTION:

You can use this command to temporarily disable an IPv6 VRRP group. After you execute this command, packets sent to the IPv6 VRRP group might be discarded.

Configuring the virtual MAC address for an IPv6 VRRP group

About this task

By default, after the network administrator creates an IPv6 VRRP group, the device automatically generates a virtual MAC address for the IPv6 VRRP group based on the VRRP protocol requirements for communication purposes. After you configure this feature, the device will set the virtual MAC address of the IPv6 VRRP group to the specified value. This can meet the requirements in specific scenarios (such as in a static ARP network).

Restrictions and guidelines

This feature takes effect only when the local device acts as the master in the IPv6 VRRP group.

The virtual MAC addresses of the IPv4 VRRP group and IPv6 VRRP group deployed on an interface cannot be the same.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the virtual MAC address for the IPv6 VRRP group.
   vrrp ipv6 vrid *virtual-router-id* **virtual-mac** *virtual-mac*

   By default, no virtual MAC address is configured for the IPv6 VRRP group. The device will automatically generate a virtual MAC address for the IPv6 VRRP group.

Configuring IPv6 VRRP packet attributes

Restrictions and guidelines

- The routers in an IPv6 VRRP group can have different intervals for sending VRRP advertisements. The master in the VRRP group sends VRRP advertisements at the specified interval and carries the interval attribute in the advertisements. After a backup receives the advertisement, it records the interval in the advertisement. If the backup does not receive a VRRP advertisement before the timer (3 x recorded interval + Skew_Time) expires, it regards the master as failed and takes over.
- A high volume of network traffic might cause a backup to fail to receive VRRP advertisements from the master within the specified time. As a result, an unexpected master switchover occurs. To solve this problem, configure a larger interval.
- If you use M-LAG and VRRP together, make sure the keepalive hold timer is shorter than the IPv6 VRRP advertisement interval. Violation of this restriction might cause a VRRP master/backup switchover to occur before peer link failure is confirmed. To set the keepalive hold timer, use the **m-lag keepalive hold-time** command. For more information about this command, see *Layer 2—LAN Switching Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Set the IPv6 VRRP advertisement interval.
   vrrp ipv6 vrid *virtual-router-id* **timer advertise** *adver-interval*

   The default setting is 100 centiseconds.

As a best practice to maintain system stability, set the VRRP advertisement interval to be greater than 100 centiseconds.

4. Return to system view.

   quit

5. Set a DSCP value for IPv6 VRRP packets.

   vrrp ipv6 dscp *dscp-value*

   By default, the DSCP value for IPv6 VRRP packets is 56.

   The DSCP value identifies the packet priority during transmission.

Enabling periodic sending of ND packets for IPv6 VRRP

About this task

This feature enables the master router in an IPv6 VRRP group to periodically send ND packets. Then the downstream devices can update the MAC address entry for the virtual MAC address of the IPv6 VRRP group in a timely manner.

Restrictions and guidelines

- This feature takes effect only in VRRP standard mode.
- If you change the sending interval for ND packets, the configuration takes effect at the next sending interval.
- The master sends the first ND packet at a random time in the second half of the set interval after you execute the **vrrp ipv6 send-nd** command. This prevents too many ND packets from being sent at the same time.
- The sending interval for ND packets might be much longer than the set interval when the following conditions are met:
  - Multiple IPv6 VRRP groups exist on the device.
  - A short sending interval is set.

Procedure

1. Enter system view.

   system-view

2. Enable periodic sending of ND packets for IPv6 VRRP.

   vrrp ipv6 send-nd [**interval** *interval*]

   By default, periodic sending of ND packets is disabled for IPv6 VRRP.

Configuring a subordinate IPv6 VRRP group to follow a master IPv6 VRRP group

About this task

Each IPv6 VRRP group determines the device role (master or backup) by exchanging VRRP packets among member devices, which might consume excessive bandwidth and CPU resources. To reduce the number of VRRP packets in the network, you can configure a subordinate IPv6 VRRP group to follow a master IPv6 VRRP group.

A master IPv6 VRRP group determines the device role through exchanging VRRP packets among member devices. An IPv6 VRRP group that follows a master group, called a subordinate VRRP group, does not exchange VRRP packets among its member devices. The state of the subordinate VRRP group follows the state of the master group.

Restrictions and guidelines

- To ensure the master router election, configure the settings such as the router priority, preemptive mode, and tracking function for the master IPv6 VRRP group. The settings are not required for subordinate IPv6 VRRP groups.
- An IPv6 VRRP group cannot be both a master group and a subordinate group.
- An IPv6 VRRP group stays in **Inactive** state if it is configured to follow a nonexistent master IPv6 VRRP group.
- If an IPv6 VRRP group in **Inactive** or **Initialize** state follows a master group that is not in **Inactive** state, the state of the VRRP group does not change.
- A subordinate IPv6 VRRP group does not exchange VRRP packets, which might cause the MAC address entry for its virtual MAC address not to be updated on downstream devices. As a best practice, enable periodic sending of ND packets for IPv6 VRRP by using the **vrrp ipv6 send-nd** command.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure an IPv6 VRRP group as a master group and assign a name to it.
   vrrp ipv6 vrid *virtual-router-id* **name** *name*
   By default, an IPv6 VRRP group does not act as a master group.
4. Return to system view.
   quit
5. Enter interface view.
   interface *interface-type interface-number*
6. Configure an IPv6 VRRP group to follow a master group.
   vrrp ipv6 vrid *virtual-router-id* **follow** *name*
   By default, an IPv6 VRRP group does not follow a master VRRP group.

Setting the state transition delay timer for an IPv6 VRRP group

About this task

Typically, an IPv6 VRRP group immediately transitions from **Initialize** to **Master** or **Backup** state when the interface where the IPv6 VRRP group resides comes up. If the device restarts or the interface state changes frequently, the IPv6 VRRP group state might be unstable, and the VRRP group operation might be affected.This task allows you to solve this issue.

Restrictions and guidelines

The delay time for an IPv6 VRRP group to transition from **Initialize** to **Master** or **Backup** state is in the range of 0 to 60 seconds.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*

3. Set the state transition delay timer for the IPv6 VRRP group.

   vrrp ipv6 state-transition-delay *delay-value*

   By default, the delay time is 0 seconds for the IPv6 VRRP group to transition from **Initialize** to **Master** or **Backup** state.

Enabling the isolation mode for IPv6 VRRP

About this task

To maintain the master in an IPv6 VRRP group, use this feature to perform a master/backup switchover. In isolation mode, the maintenance operation on the device is performed after a master/backup switchover and does not affect services.

Enabling the isolation mode sets the priority of the device to 0 in the IPv6 VRRP group.

Restrictions and guidelines

If you set the priority with the **vrrp ipv6 vrid priority** command for a device in isolation mode, the setting does not take effect until the isolation mode is disabled.

Procedure

1. Enter system view.

   system-view

2. Enable the isolation mode for IPv6 VRRP.

   vrrp ipv6 isolate enable

   By default, the isolation mode is disabled for IPv6 VRRP.

Display and maintenance commands for IPv6 VRRP

Execute **display** commands in any view and the **reset** command in user view.

Task	Command
Display the states of IPv6 VRRP groups.	**display vrrp ipv6** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]] [**verbose**]
Display IPv6 VRRP packet anomaly information.	**display vrrp ipv6 abnormal-info** [**interface** *interface-type interface-number* [**vrid** *vrid-id*]]
Display master-to-subordinate IPv6 VRRP group bindings.	**display vrrp ipv6 binding** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*] | **name** *name*]
Display statistics for IPv6 VRRP groups.	**display vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]
Clear statistics for IPv6 VRRP groups.	**reset vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]

IPv4 VRRP configuration examples

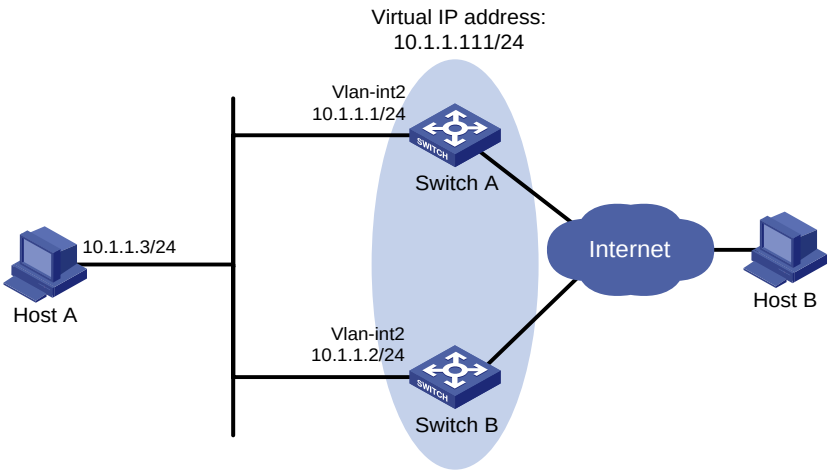
Example: Configuring a single VRRP group

Network configuration

As shown in Figure 9, Switch A and Switch B form a VRRP group. They use the virtual IP address 10.1.1.111/24 to provide gateway service for the subnet where Host A resides.

Switch A operates as the master to forward packets from Host A to Host B. When Switch A fails, Switch B takes over to forward packets for Host A.

Figure 9 Network diagram



Procedure

1. Configure Switch A:

   # Configure VLAN 2.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 2
   [SwitchA-vlan2] port gigabitethernet 1/0/5
   [SwitchA-vlan2] quit
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ip address 10.1.1.1 255.255.255.0
   ```

   # Create VRRP group 1 on VLAN-interface 2, and set its virtual IP address to 10.1.1.111.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.111
   ```

   # Assign Switch A a higher priority than Switch B in VRRP group 1, so Switch A can become the master.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
   ```

   # Configure Switch A to operate in preemptive mode, so it can become the master whenever it operates correctly. Set the preemption delay to 5000 centiseconds to avoid frequent status switchover.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 preempt-mode delay 5000
   ```

2. Configure Switch B:

   # Configure VLAN 2.

   ```
   <SwitchB> system-view
   [SwitchB] vlan 2
   ```

```
[SwitchB-Vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 10.1.1.2 255.255.255.0
```

Create VRRP group 1 on VLAN-interface 2, and set its virtual IP address to 10.1.1.111.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.111
```

Set the priority of Router B to 100 in VRRP group 1.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 priority 100
```

Configure Switch B to operate in preemptive mode, and set the preemption delay to 5000 centiseconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 preempt-mode delay 5000
```

Verifying the configuration

Ping Host B from Host A. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                  Adver Timer  : 100
     Admin Status     : Up                 State        : Master
     Config Pri       : 110                Running Pri  : 110
     Preempt Mode     : Yes                Delay Time   : 5000
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.111
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 10.1.1.1
     First connection : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                  Adver Timer  : 100
     Admin Status     : Up                 State        : Backup
     Config Pri       : 100                Running Pri  : 100
     Preempt Mode     : Yes                Delay Time   : 5000
     Become Master    : 401ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.111
     Master IP        : 10.1.1.1
```

```
    First connection : 2023/03/11 16:29:42:111
    Last status change : 2023/03/11 16:30:42:111
```

The output shows that Switch A is operating as the master in VRRP group 1 to forward packets from Host A to Host B.

Disconnect the link between Host A and Switch A, and verify that Host A can still ping Host B. (Details not shown.)

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                    Adver Timer  : 100
     Admin Status     : Up                   State        : Master
     Config Pri       : 100                  Running Pri  : 100
     Preempt Mode     : Yes                  Delay Time   : 5000
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.111
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 10.1.1.2
     First connection : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111
```

The output shows that when Switch A fails, Switch B takes over to forward packets from Host A to Host B.

Recover the link between Host A and Switch A, and display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                    Adver Timer  : 100
     Admin Status     : Up                   State        : Master
     Config Pri       : 110                  Running Pri  : 110
     Preempt Mode     : Yes                  Delay Time   : 5000
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.111
     Master IP        : 10.1.1.1
     First connection : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111
```

The output shows that after Switch A resumes normal operation, it becomes the master to forward packets from Host A to Host B.

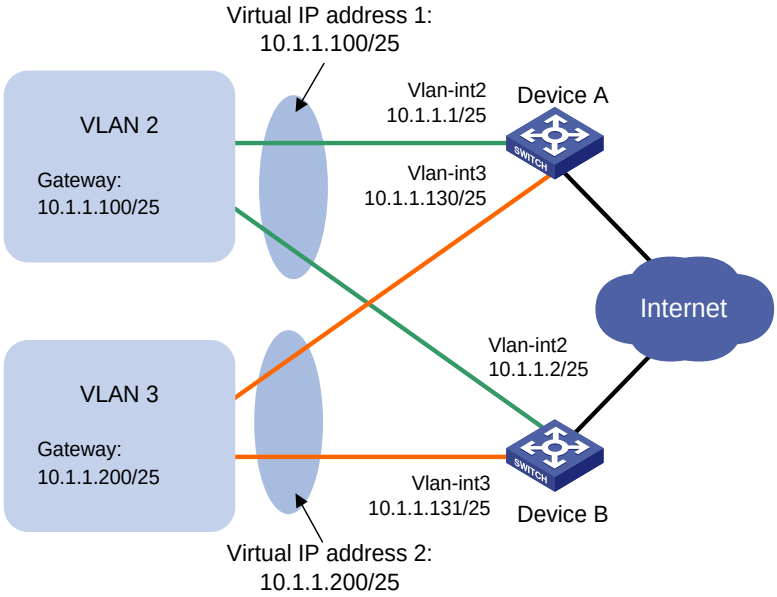
Example: Configuring multiple VRRP groups

Network configuration

As shown in Figure 10, Switch A and Switch B form two VRRP groups. VRRP group 1 uses the virtual IP address 10.1.1.100/25 to provide gateway service for hosts in VLAN 2, and VRRP group 2 uses the virtual IP address 10.1.1.200/25 to provide gateway service for hosts in VLAN 3.

Assign a higher priority to Switch A than Switch B in VRRP group 1, but a lower priority in VRRP group 2. Traffic from VLAN 2 and VLAN 3 can then be distributed between the two switches. When one of the switches fails, the healthy switch provides gateway service for both VLANs.

Figure 10 Network diagram



Procedure

1. Configure Switch A:

   # Configure VLAN 2.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 2
   [SwitchA-vlan2] port gigabitethernet 1/0/5
   [SwitchA-vlan2] quit
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ip address 10.1.1.1 255.255.255.128
   ```

   # Create VRRP group 1, and set its virtual IP address to 10.1.1.100.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.100
   ```

   # Assign Switch A a higher priority than Switch B in VRRP group 1, so Switch A can become the master in the group.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
   [SwitchA-Vlan-interface2] quit
   ```

   # Configure VLAN 3.

   ```
   [SwitchA] vlan 3
   [SwitchA-vlan3] port gigabitethernet 1/0/6
   [SwitchA-vlan3] quit
   [SwitchA] interface vlan-interface 3
   ```

```
[SwitchA-Vlan-interface3] ip address 10.1.1.130 255.255.255.128
```

Create VRRP group 2, and set its virtual IP address to 10.1.1.200.

```
[SwitchA-Vlan-interface3] vrrp vrid 2 virtual-ip 10.1.1.200
```

2. Configure Switch B:

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 10.1.1.2 255.255.255.128
```

Create VRRP group 1, and set its virtual IP address to 10.1.1.100.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.100
[SwitchB-Vlan-interface2] quit
```

Configure VLAN 3.

```
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/6
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ip address 10.1.1.131 255.255.255.128
```

Create VRRP group 2, and set its virtual IP address to 10.1.1.200.

```
[SwitchB-Vlan-interface3] vrrp vrid 2 virtual-ip 10.1.1.200
```

Assign Switch B a higher priority than Switch A in VRRP group 2, so Switch B can become the master in the group.

```
[SwitchB-Vlan-interface3] vrrp vrid 2 priority 110
```

Verifying the configuration

Display detailed information about the VRRP groups on Switch A.

```
[SwitchA-Vlan-interface3] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 2
   Interface Vlan-interface2
     VRID              : 1                 Adver Timer  : 100
     Admin Status      : Up                State        : Master
     Config Pri        : 110               Running Pri  : 110
     Preempt Mode      : Yes               Delay Time   : 0
     Auth Type         : Not supported
     Version           : 3
     Virtual IP        : 10.1.1.100
     Virtual MAC       : 0000-5e00-0101
     Master IP         : 10.1.1.1
     First connection  : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111

   Interface Vlan-interface3
     VRID              : 2                 Adver Timer  : 100
```

```
    Admin Status     : Up                   State        : Backup
    Config Pri       : 100                  Running Pri  : 100
    Preempt Mode     : Yes                  Delay Time   : 0
    Become Master    : 203ms left
    Auth Type        : Not supported
    Version          : 3
    Virtual IP       : 10.1.1.200
    Master IP        : 10.1.1.131
    First connection : 2023/03/11 16:29:42:111
    Last status change : 2023/03/11 16:30:42:111
```

Display detailed information about the VRRP groups on Switch B.

```
[SwitchB-Vlan-interface3] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 2
   Interface Vlan-interface2
     VRID             : 1                    Adver Timer  : 100
     Admin Status     : Up                   State        : Backup
     Config Pri       : 100                  Running Pri  : 100
     Preempt Mode     : Yes                  Delay Time   : 0
     Become Master    : 211ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.100
     Master IP        : 10.1.1.1
     First connection : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111

   Interface Vlan-interface3
     VRID             : 2                    Adver Timer  : 100
     Admin Status     : Up                   State        : Master
     Config Pri       : 110                  Running Pri  : 110
     Preempt Mode     : Yes                  Delay Time   : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.200
     Virtual MAC      : 0000-5e00-0102
     Master IP        : 10.1.1.131
     First connection : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111
```

The output shows the following information:

- Switch A is operating as the master in VRRP group 1 to forward Internet traffic for hosts that use the default gateway 10.1.1.100/25.
- Switch B is operating as the master in VRRP group 2 to forward Internet traffic for hosts that use the default gateway 10.1.1.200/25.

IPv6 VRRP configuration examples

Example: Configuring a single VRRP group

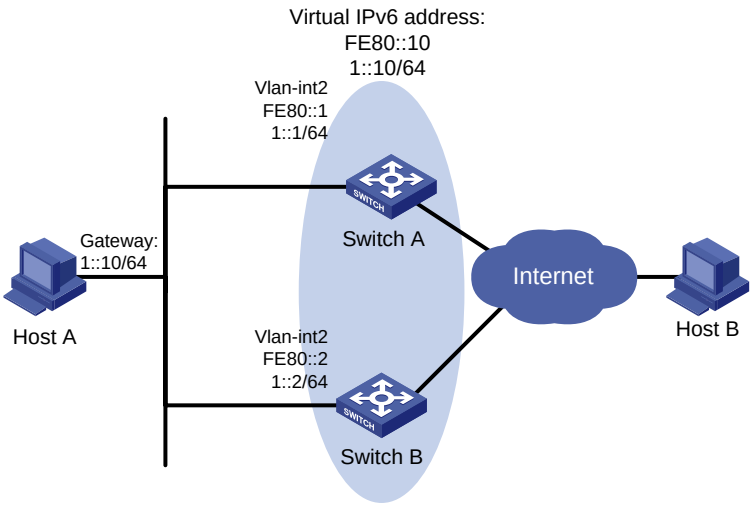
Network configuration

As shown in Figure 11, Switch A and Switch B form a VRRP group. They use the virtual IP addresses 1::10/64 and FE80::10 to provide gateway service for the subnet where Host A resides.

Host A learns 1::10/64 as its default gateway from RA messages sent by the switches.

Switch A operates as the master to forward packets from Host A to Host B. When Switch A fails, Switch B takes over to forward packets for Host A.

Figure 11 Network diagram



Procedure

1. Configure Switch A:

   # Configure VLAN 2.

   ```
   <SwitchA> system-view
   [SwitchA] vlan 2
   [SwitchA-vlan2] port gigabitethernet 1/0/5
   [SwitchA-vlan2] quit
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
   [SwitchA-Vlan-interface2] ipv6 address 1::1 64
   ```

   # Create VRRP group 1, and set its virtual IPv6 addresses to FE80::10 and 1::10.

   ```
   [SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
   [SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
   ```

   # Assign Switch A a higher priority than Switch B in VRRP group 1, so Switch A can become the master.

   ```
   [SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
   ```

   # Configure Switch A to operate in preemptive mode, so it can become the master whenever it operates correctly. Set the preemption delay to 5000 centiseconds to avoid frequent status switchover.

   ```
   [SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode delay 5000
   ```

Enable Switch A to send RA messages, so Host A can learn the default gateway address.

```
[SwitchA-Vlan-interface2] undo ipv6 nd ra halt
```

2. Configure Switch B:

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
[SwitchB-Vlan-interface2] ipv6 address 1::2 64
```

Create VRRP group 1 and set its virtual IPv6 addresses to FE80::10 and 1::10.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Configure Switch B to operate in preemptive mode, and set the preemption delay to 5000 centiseconds.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode delay 5000
```

Enable Switch B to send RA messages, so Host A can learn the default gateway address.

```
[SwitchB-Vlan-interface2] undo ipv6 nd ra halt
```

Verifying the configuration

Ping Host B from Host A. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                    Adver Timer  : 100
     Admin Status     : Up                   State        : Master
     Config Pri       : 110                  Running Pri  : 110
     Preempt Mode     : Yes                  Delay Time   : 5000
     Auth Type        : Not supported
     Virtual IP       : FE80::10
                        1::10
     Virtual MAC      : 0000-5e00-0201
     Master IP        : FE80::1
     First connection : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Virtual Router Information:
 Running Mode        : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                    Adver Timer  : 100
```

```
    Admin Status     : Up                   State        : Backup
    Config Pri       : 100                  Running Pri  : 100
    Preempt Mode     : Yes                  Delay Time   : 5000
    Become Master    : 403ms left
    Auth Type        : Not supported
    Virtual IP       : FE80::10
                       1::10
    Master IP        : FE80::1
    First connection : 2023/03/11 16:29:42:111
    Last status change : 2023/03/11 16:30:42:111
```

The output shows that Switch A is operating as the master in VRRP group 1 to forward packets from Host A to Host B.

Disconnect the link between Host A and Switch A, and verify that Host A can still ping Host B. (Details not shown.)

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                    Adver Timer  : 100
     Admin Status     : Up                   State        : Master
     Config Pri       : 100                  Running Pri  : 100
     Preempt Mode     : Yes                  Delay Time   : 5000
     Auth Type        : Not supported
     Virtual IP       : FE80::10
                        1::10
     Virtual MAC      : 0000-5e00-0201
     Master IP        : FE80::2
     First connection : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111
```

The output shows that when Switch A fails, Switch B takes over to forward packets from Host A to Host B.

Recover the link between Host A and Switch A, and display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                    Adver Timer  : 100
     Admin Status     : Up                   State        : Master
     Config Pri       : 110                  Running Pri  : 110
     Preempt Mode     : Yes                  Delay Time   : 5000
     Auth Type        : Not supported
     Virtual IP       : FE80::10
```

```
                    1::10
Virtual MAC       : 0000-5e00-0201
Master IP         : FE80::1
First connection  : 2023/03/11 16:29:42:111
Last status change : 2023/03/11 16:30:42:111
```

The output shows that after Switch A resumes normal operation, it becomes the master to forward packets from Host A to Host B.

Example: Configuring multiple VRRP groups

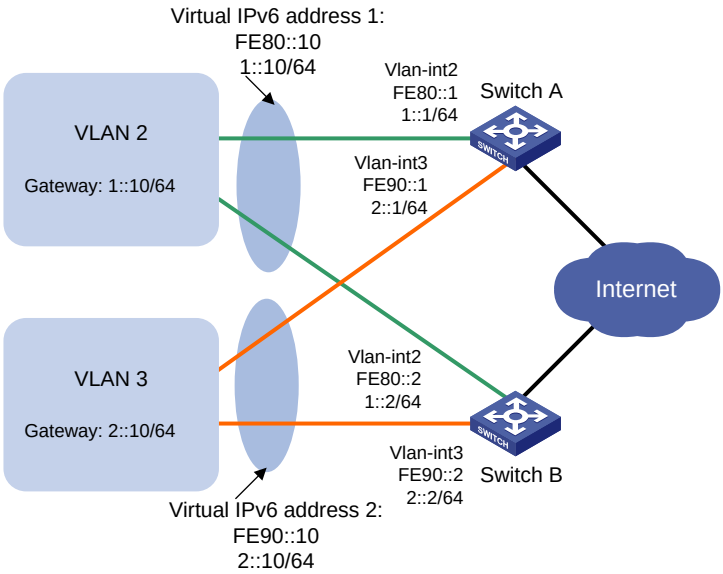
Network configuration

As shown in Figure 12, Switch A and Switch B form two VRRP groups. VRRP group 1 uses the virtual IPv6 addresses 1::10/64 and FE80::10 to provide gateway service for hosts in VLAN 2. VRRP group 2 uses the virtual IPv6 addresses 2::10/64 and FE90::10 to provide gateway service for hosts in VLAN 3.

From RA messages sent by the switches, hosts in VLAN 2 learn 1::10/64 as their default gateway. Hosts in VLAN 3 learn 2::10/64 as their default gateway.

Assign Switch A a higher priority than Switch B in VRRP group 1 but a lower priority in VRRP group 2. Traffic from VLAN 2 and VLAN 3 can then be distributed between the two switches. When one of the switches fails, the healthy switch provides gateway service for both VLANs.

Figure 12 Network diagram



Procedure

1. Configure Switch A:

   # Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
[SwitchA-Vlan-interface2] ipv6 address 1::1 64
```

Create VRRP group 1, and set its virtual IPv6 addresses to FE80::10 to 1::10.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Assign Switch A a higher priority than Switch B in VRRP group 1, so Switch A can become the master in the group.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

Enable Switch A to send RA messages, so hosts in VLAN 2 can learn the default gateway address.

```
[SwitchA-Vlan-interface2] undo ipv6 nd ra halt
[SwitchA-Vlan-interface2] quit
```

Configure VLAN 3.

```
[SwitchA] vlan 3
[SwitchA-vlan3] port gigabitethernet 1/0/6
[SwitchA-vlan3] quit
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ipv6 address fe90::1 link-local
[SwitchA-Vlan-interface3] ipv6 address 2::1 64
```

Create VRRP group 2, and set its virtual IPv6 addresses to FE90::10 and 2::10.

```
[SwitchA-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip fe90::10 link-local
[SwitchA-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip 2::10
```

Enable Switch A to send RA messages, so hosts in VLAN 3 can learn the default gateway address.

```
[SwitchA-Vlan-interface3] undo ipv6 nd ra halt
```

2. Configure Switch B:

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
[SwitchB-Vlan-interface2] ipv6 address 1::2 64
```

Create VRRP group 1, and set its virtual IPv6 addresses to FE80::10 and 1::10.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Enable Switch B to send RA messages, so hosts in VLAN 2 can learn the default gateway address.

```
[SwitchB-Vlan-interface2] undo ipv6 nd ra halt
[SwitchB-Vlan-interface2] quit
```

Configure VLAN 3.

```
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/6
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ipv6 address fe90::2 link-local
[SwitchB-Vlan-interface3] ipv6 address 2::2 64
```

Create VRRP group 2, and set its virtual IPv6 addresses to FE90::10 and 2::10.

```
[SwitchB-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip fe90::10 link-local
[SwitchB-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip 2::10
```

Assign Switch B a higher priority than Switch A in VRRP group 2, so Switch B can become the master in the group.

```
[SwitchB-Vlan-interface3] vrrp ipv6 vrid 2 priority 110
```

Enable Switch B to send RA messages, so hosts in VLAN 3 can learn the default gateway address.

```
[SwitchB-Vlan-interface3] undo ipv6 nd ra halt
```

Verifying the configuration

Display detailed information about the VRRP groups on Switch A.

```
[SwitchA-Vlan-interface3] display vrrp ipv6 verbose
IPv6 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 2
   Interface Vlan-interface2
     VRID              : 1                   Adver Timer  : 100
     Admin Status      : Up                  State        : Master
     Config Pri        : 110                 Running Pri  : 110
     Preempt Mode      : Yes                 Delay Time   : 0
     Auth Type         : Not supported
     Virtual IP        : FE80::10
                         1::10
     Virtual MAC       : 0000-5e00-0201
     Master IP         : FE80::1
     First connection  : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111

   Interface Vlan-interface3
     VRID              : 2                   Adver Timer  : 100
     Admin Status      : Up                  State        : Backup
     Config Pri        : 100                 Running Pri  : 100
     Preempt Mode      : Yes                 Delay Time   : 0
     Become Master     : 402ms left
     Auth Type         : Not supported
     Virtual IP        : FE90::10
                         2::10
     Master IP         : FE90::2
     First connection  : 2023/03/11 16:29:42:111
     Last status change : 2023/03/11 16:30:42:111
```

Display detailed information about the VRRP groups on Switch B.

```
[SwitchB-Vlan-interface3] display vrrp ipv6 verbose
IPv6 Virtual Router Information:
 Running Mode      : Standard
 M-LAG group traffic forwarding for backup router : Enabled
 Total number of virtual routers : 2
   Interface Vlan-interface2
     VRID              : 1                   Adver Timer  : 100
     Admin Status      : Up                  State        : Backup
     Config Pri        : 100                 Running Pri  : 100
```

```
  Preempt Mode     : Yes                 Delay Time   : 0
  Become Master    : 401ms left
  Auth Type        : Not supported
  Virtual IP       : FE80::10
                     1::10
  Master IP        : FE80::1
  First connection : 2023/03/11 16:29:42:111
  Last status change : 2023/03/11 16:30:42:111

Interface Vlan-interface3
  VRID             : 2                   Adver Timer  : 100
  Admin Status     : Up                  State        : Master
  Config Pri       : 110                 Running Pri  : 110
  Preempt Mode     : Yes                 Delay Time   : 0
  Auth Type        : Not supported
  Virtual IP       : FE90::10
                     2::10
  Virtual MAC      : 0000-5e00-0202
  Master IP        : FE90::2
  First connection : 2023/03/11 16:29:42:111
  Last status change : 2023/03/11 16:30:42:111
```

The output shows the following information:

- Switch A is operating as the master in VRRP group 1 to forward Internet traffic for hosts that use the default gateway 1::10/64.
- Switch B is operating as the master in VRRP group 2 to forward Internet traffic for hosts that use the default gateway 2::10/64.

Troubleshooting VRRP

An error prompt is displayed

Symptom

An error prompt "The virtual router detected a VRRP configuration error." is displayed during configuration.

Analysis

This symptom is probably caused by the following reasons:

- The VRRP advertisement interval in the packet is not the same as that for the current VRRP group (in VRRPv2 only).
- The number of virtual IP addresses in the packet is not the same as that for the current VRRP group.
- The virtual IP address list is not the same as that for the current VRRP group.
- A device in the VRRP group receives illegitimate VRRP packets. For example, the IP address owner receives a VRRP packet with the priority 255.

Solution

To resolve the problem:

1. Modify the configuration on routers in VRRP groups to ensure consistent configuration.

2. Take fault location and anti-attack measures to eliminate potential threats.
3. If the problem persists, contact H3C Support.

Multiple masters appear in a VRRP group

Symptom

Multiple masters appear in a VRRP group.

Analysis

It is normal for a VRRP group to have multiple masters for a short time, and this situation requires no manual intervention.

If multiple masters coexist for a longer period, check for the following conditions:

- The masters cannot receive advertisements from each other.
- The received advertisements are illegitimate.

Solution

To resolve the problem:

1. Ping between these masters:
   - If the ping operation fails, examine network connectivity.
   - If the ping operation succeeds, check for configuration inconsistencies in the number of virtual IP addresses, virtual IP addresses, and authentication. For IPv4 VRRP, also make sure the same version of VRRP is configured on all routers in the VRRP group. For VRRPv2, make sure the same VRRP advertisement interval is configured on the routers in the VRRP group.
2. If the problem persists, contact H3C Support.

Fast VRRP state flapping

Symptom

Fast VRRP state flapping occurs.

Analysis

The VRRP advertisement interval is set too short.

Solution

To resolve the problem:

1. Increase the interval for sending VRRP advertisements or introduce a preemption delay.
2. If the problem persists, contact H3C Support.

Contents

Configuring BFD 1

About BFD 1
Mechanism 1
Single-hop detection and multihop detection 2
BFD session states 2
BFD session modes 3
BFD session creation methods 7
BFD session flapping suppression 7
Hardware BFD 7
Supported features 7
Link types supported by BFD 8
Technical benefits of BFD 9
Protocols and standards 9
Restrictions and guidelines: BFD configuration 9
BFD tasks at a glance 10
Configuring BFD sessions in echo packet mode 10
About BFD session creation methods 10
Creating a static BFD session 10
Configuring detection time settings 12
Configuring BFD sessions in control packet mode 13
About BFD session creation methods 13
Restrictions and guidelines 14
Configuring the session establishment mode 14
Configuring the detection mode 14
Creating a static BFD session 15
Configuring detection time settings 16
Configuring the authentication mode 19
Configuring the notification delay timer for session establishment failures 20
Enabling the echo function 20
Creating a static BFD session for detecting the aggregate link of a Layer 3 aggregate interface 21
Configuring the TTL value for BFD packets 24
Configuring BFD session flapping suppression 24
Delaying BFD session negotiation 25
About this task 25
Delaying BFD session negotiation upon a device reboot 25
Delaying BFD session negotiation for down BFD sessions 26
Configuring a BFD template 26
Configuring event logging and SNMP notifications for BFD 27
Enabling BFD event logging 27
Enabling SNMP notifications for BFD 27
Display and maintenance commands for BFD 27

Configuring BFD

About BFD

Bidirectional forwarding detection (BFD) provides a general-purpose, standard, medium- and protocol-independent fast failure detection mechanism. It can detect and monitor the connectivity of forwarding paths to detect communication failures quickly so that measures can be taken to ensure service continuity and enhance network availability.

BFD can uniformly and quickly detect the failures of the bidirectional forwarding paths between two devices for upper-layer protocols such as routing protocols. The hello mechanism used by upper-layer protocols needs seconds to detect a link failure, while BFD can provide detection measured in milliseconds.

Mechanism

BFD establishes a session between two network devices to detect failures on the bidirectional forwarding paths between the devices and provide services for upper-layer protocols. BFD provides no neighbor discovery mechanism. Protocols that BFD services notify BFD of devices to which it needs to establish sessions. After a session is established, if no BFD control packet is received from the peer within the negotiated BFD interval, BFD notifies a failure to the protocol, which then takes appropriate measures.

Figure 1 describes the operation of BFD for OSPF.

1. OSPF discovers neighbors by sending Hello packets and establishes neighbor relationships.
2. After establishing neighbor relationships, OSPF notifies BFD of the neighbor information, including destination and source addresses.
3. BFD uses the information to establish BFD sessions.
4. After session establishment, BFD starts periodic link detection.

Figure 1 BFD session establishment

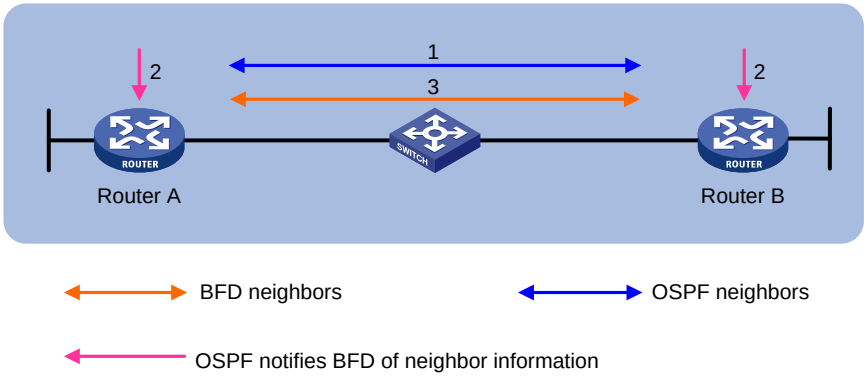
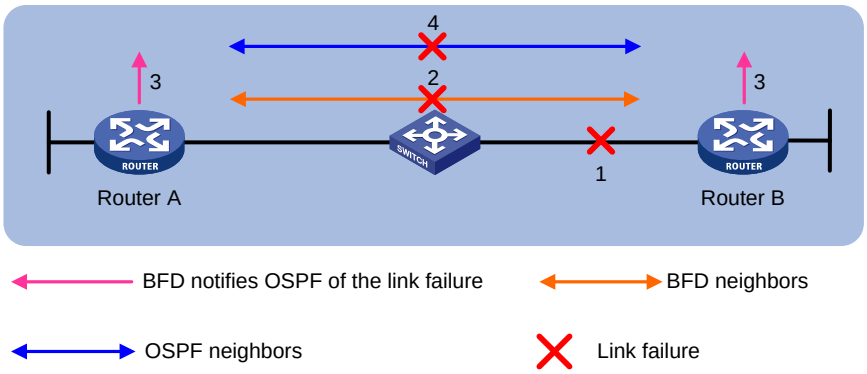


Figure 2 describes the BFD fault detection process.

5. The detected link fails.
6. BFD detects the link failure and tears down the session.
7. BFD notifies the neighbor unreachability to OSPF.
8. OSPF terminates the neighbor relationship on the link.

Figure 2 BFD fault detection



BFD establishes sessions in echo packet mode or control packet mode depending on the BFD packets used.

Single-hop detection and multihop detection

BFD can be used for single-hop and multihop detections.

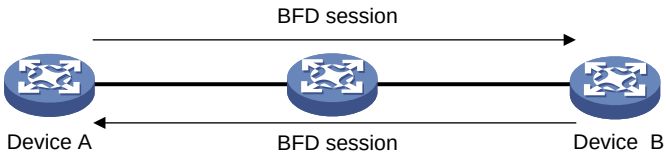
- **Single-hop detection**—Detects the IP connectivity between two directly connected systems, for example, Device A and Device B as shown in Figure 3.

  Figure 3 Single-hop detection

- **Multihop detection**—Detects any of the paths between two systems, for example, Device A and Device B as shown in Figure 4. These paths have multiple hops, and might overlap.

  Figure 4 Multihop detection

BFD session states

A BFD session can be in DOWN, ADMIN DOWN, INIT or UP state. The meanings of the states are as follows:

- **DOWN**—The local BFD session has been shut down or is newly created. This state indicates that the forwarding path is unavailable and that the upper-layer application must take an appropriate action, such as primary/backup path switchover.
- **ADMIN DOWN**—The BFD session of the local system is administratively shut down. This state does not indicate that the forwarding path is unavailable. The upper-layer application is not required to take an action.
- **INIT**—The local end can communicate with the remote end and requires the BFD session to transition to UP state.

- **UP**—The BFD session has been established on the local end. This state indicates that the forwarding path is available.

BFD session modes

BFD sessions use echo packets and control packets.

Echo packet mode

Echo packets are encapsulated into UDP packets. The packet contains destination port 3785, a destination IP address (transmitting interface's IP address), and a source IP address (transmitting interface's IP address).

The local end of the link sends echo packets to establish BFD sessions and monitor link status. The peer end does not establish BFD sessions and only forwards the packets back to the originating end. If the local end does not receive echo packets from the peer end within the detection time, it considers the session down.

BFD sessions in echo packet mode do not require both ends to support BFD. Upon receiving BFD echo packets, the peer end loops back the packets for detection purposes.

In echo packet mode, BFD supports only single-hop detection and the BFD sessions are independent of the operating mode.

Control packet mode

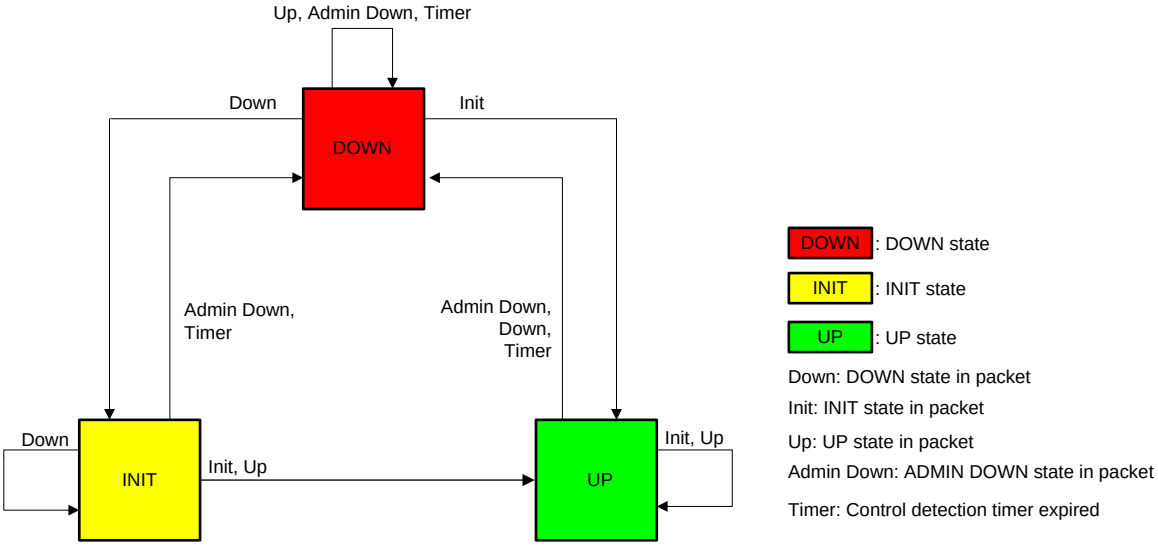
Control packets are encapsulated into UDP packets with port number 3784 for single-hop detection or port number 4784 for multihop detection.

The two ends of the link negotiate the establishment of BFD sessions by using the session parameters carried in control packets. Session parameters include session discriminators, desired minimum packet sending and receiving intervals, and local BFD session state.

State machine transition mechanism for a BFD session in control packet mode

Figure 5 shows the state machine transition mechanism for a BFD session in control packet mode. The ADMIN DOWN state in the state machine is also the DOWN state.

Figure 5 State machine transition mechanism for a BFD session



BFD control packet format

Figure 6 shows the format of a BFD control packet. The BFD control packet contains the required portion and optional authentication portion.

Figure 6 BFD control packet format

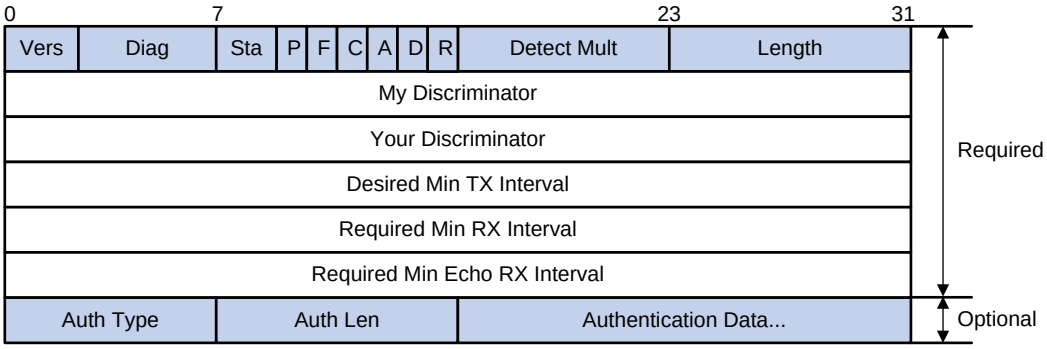


Table 1 shows the description of the fields in a BFD control packet.

Table 1 BFD control packet field description

Field	Length	Description
Vers (Version)	3 bits	BFD version number. The current version number is 1.
Diag (Diagnostic)	5 bits	Diagnostic code, which indicates the reason for the most recent session down of the sender. Options include: • **0**—No diagnostic information. • **1**—A control-mode BFD session goes down because local detection times out. • **2**—An echo-mode BFD session goes down, because local detection times out or the source IP address of echo packets is deleted. • **3**—The remote end notifies the local end of BFD session down upon detection timeout. • **7**—The local system prevents a BFD session from being established. • **31**—A packet in AdminDown state is received from the remote end.
Sta (State)	2 bits	BFD session state of the sender: • **0**—AdminDown. • **1**—Down. • **2**—Init. • **3**—Up.
P (Poll)	1 bit	Upon session parameter changes, the sender sends a BFD packet with the P bit set. The receiver must immediately respond this packet. • **0**—The sender does not request verification. • **1**—The sender requests verification of the connectivity or parameter changes.
F (Final)	1 bit	The F bit determines whether BFD sends a response packet upon receiving a packet with the P bit set. • **0**—The sending system does not respond to packets with the P bit set. • **1**—The sending system responds to received packets with the P bit set.
C (Control Plane Independent)	1 bit	Forwarding- and control-plane separation bit. If this bit is set, the BFD implementation is independent of its control plane. That is, changes in the control plane do not affect BFD detection. For example, when IS-IS on the control

Field	Length	Description
		plane performs GR, BFD can continue to detect link state. • **1**—The BFD implementation on the sending system is independent of the control plane. That is, BFD packets are transmitted on the forwarding plane. BFD can still perform link detection even if the control plane becomes invalid. • **0**—Indicates that BFD packets are transmitted on the control plane.
A (Authentication Present)	1 bit	If this bit is set, the control packet contains the authentication field and the session requires authentication.
D (Demand)	1 bit	If this bit is set, the sender wants to operate in Demand mode. If this bit is not set, the sender does not want to operate in Demand mode or does not support the Demand mode.
M (Multipoint)	1 bit	Reserved bit. It is set to 0 during transmission and ignored during reception.
Detect Mult	8 bits	Detect time multiplier used by the detecting party to calculate the detection interval.
Length	8 bits	BFD control packet length, in bytes.
My Discriminator	32 bits	Local discriminator of the BFD session. It is a unique, non-zero value generated by the sender to identify a BFD session.
Your Discriminator	32 bits	Remote discriminator of the BFD session. If a BFD control packet is received from the session-associated neighbor, the value is My Discriminator in the received packet. If no BFD control packet is received from the neighbor, this value is 0.
Desired Min TX Interval	32 bits	Minimum interval for sending BFD control packets desired by the local system, in microseconds.
Required Min RX Interval	32 bits	Minimum interval for receiving BFD control packets supported by the local system, in milliseconds.
Required Min Echo RX Interval	32 bits	Minimum interval for receiving BFD echo packets supported by the local system, in milliseconds. Value 0 indicates that the echo function is not supported.
Auth Type	8 bits	Authentication type.
Auth Len	8 bits	Authentication field length in bytes, including Auth Type field and Auth Len fields.
Authentication Data	Variable	Authentication data field.

Establishing a BFD session in control packet mode

BFD can use the following operating modes to establish a session:

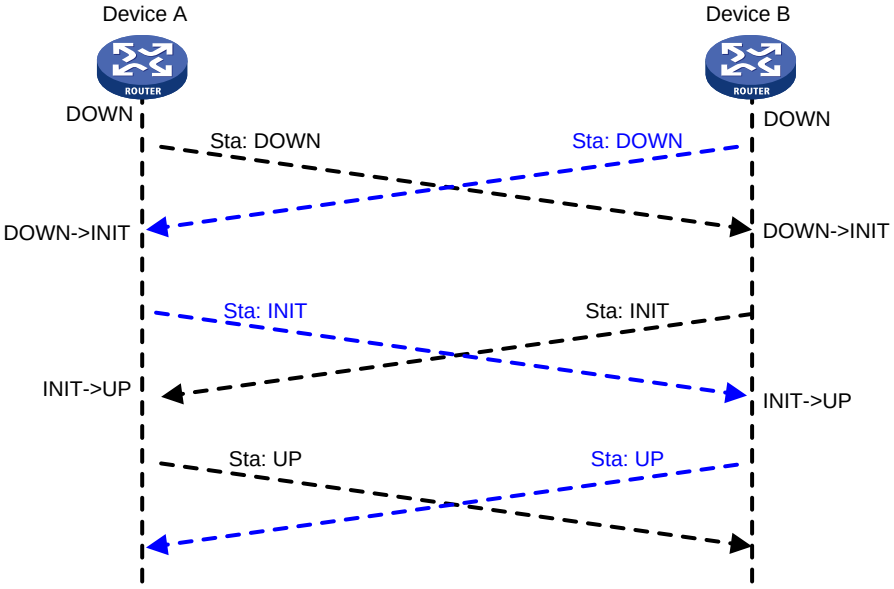
- **Active mode**—BFD actively sends BFD control packets regardless of whether any BFD control packet is received from the peer.
- **Passive mode**—BFD does not send control packets until a BFD control packet is received from the peer.

At least one end must operate in active mode for a BFD session to be established.

BFD uses a three-way handshake mechanism to establish sessions. The sender fills the Sta field with the current session state in the transmitted BFD control packets. The receiver transits the state

machine states based on the Sta field value in the received BFD control packets and local session status to establish a session.

Figure 7 BFD session establishment process



Detection mechanism of the BFD session in control packet mode

After a BFD session in control packet mode is established, the two ends can operate in the following modes:

- **Asynchronous mode**—The device periodically sends BFD control packets. The device considers that the session is down if it does not receive any BFD control packets within the detection interval. By default, a BFD session is in Asynchronous mode.
- **Demand mode**—If a large number of BFD sessions exist in the system, you can specify the Demand mode to prevent periodic sending of BFD control packets from affecting system operations. As a best practice, enable the BFD echo function when the detection mode is Demand mode.
  - If the local end operates in Demand mode, the local device periodically sends BFD control packets with the D bit set to 1. Upon receiving such packets, if the peer end operates in Asynchronous mode (default setting), it stops sending BFD control packets.
  - If the peer end is operating in Demand mode, both ends stop sending BFD control packets. When the connectivity to another system needs to be verified explicitly, a system sends several BFD control packets with the P bit set at the negotiated transmit interval. If no packets with the F bit set to 1 are received from the peer end within the detection interval, the session is considered down. If a packet with the F bit set to 1 is received from the peer end, the connectivity is found to be up, and no more BFD control packets are sent until the next command is issued.

Echo function of a BFD session in control packet mode

When you use Asynchronous mode BFD to detect the connectivity between directly connected devices, you can enable the echo function. This function enables the local system to periodically send echo packets to the remote system and reduces the control packet receiving rate to save bandwidth usage. The remote system loops back the echo packets to the local system without processing them. If the local system does not receive echo packets looped back from the remote system in a consecutive number of times, the local system declares the BFD session down.

The echo function is supported only for single-hop detection.

BFD session creation methods

BFD sessions in control packet mode can be created statically or established dynamically.

BFD sessions are distinguished by the local discriminator and remote discriminator in control packets. The main difference between a statically created session and a dynamically established session is that they obtain the local discriminator and remote discriminator in different ways.

- The local discriminator and remote discriminator of a static BFD session are manually configured.
- The local discriminator of a dynamic BFD session is automatically assigned by the device, and the remote discriminator is obtained during BFD session negotiation. Typically, BFD assigns a fixed range of local discriminators. In an M-LAG network, when M-LAG devices use dynamic BFD sessions to detect downlink connectivity, BFD assigns local discriminator ranges based on the M-LAG system ID, which means that M-LAG devices with different system IDs are assigned different local discriminator ranges.

BFD session flapping suppression

When BFD detects a link failure, it tears down the BFD session and notifies the upper-layer protocol of the failure. When the upper-layer protocol re-establishes a neighbor relationship, the BFD session comes up again. BFD session flaps occur when a link fails and recovers repeatedly, which consumes significant system resources and causes network instability. You can use BFD session flapping suppression to solve the previous problems.

This feature allows you to suppress BFD session flapping by setting different suppression intervals based on the BFD session down times. Within the suppression intervals, BFD sessions cannot be re-established. The more the BFD session down times, the longer the suppression time to avoid BFD flapping.

Hardware BFD

Software BFD uses the CPU to receive and send BFD packets and to maintain the BFD state machine. Software BFD greatly consumes CPU resources and cannot support a large number of BFD sessions.

Hardware BFD uses hardware chips to process CPU-intensive tasks, such as receiving and sending BFD packets and fault detection. Hardware BFD improves CPU utilization and can support a large number of BFD sessions.

The **Hardware mode** field in the output from the **`display bfd session verbose`** command indicates whether BFD packets are processed in hardware:

- **Enabled**—BFD packets are processed in hardware.
- **Disabled**—BFD packets are processed in software.

You can take the following actions to disable hardware BFD:

- Execute commands that are not supported by hardware BFD.
- Remove cards that support hardware BFD.

When you disable hardware BFD, the BFD session might flap. Before performing this operation, make sure you understand its potential impact on the network.

Supported features

Table 2 shows the features supported by BFD.

Table 2 Features supported by BFD

Feature	Reference
Link aggregation	Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*
IPv4 static routing	Static routing in *Layer 3—IP Routing Configuration Guide*
IPv6 static routing	IPv6 static routing in *Layer 3—IP Routing Configuration Guide*
RIP	RIP in *Layer 3—IP Routing Configuration Guide*
OSPF	OSPF in *Layer 3—IP Routing Configuration Guide*
OSPFv3	OSPFv3 in *Layer 3—IP Routing Configuration Guide*
IS-IS	IS-IS in *Layer 3—IP Routing Configuration Guide*
BGP	BGP in *Layer 3—IP Routing Configuration Guide*
PIM	PIM *in IP Multicast Configuration Guide*
IPv6 PIM	IPv6 PIM *in IP Multicast Configuration Guide*
MPLS	MPLS OAM *in IP Multicast Configuration Guide*
Track	"Configuring Track"

Link types supported by BFD

BFD can detect connectivity for various types of links, as shown in Table 3.

Table 3 Link types supported by BFD

Link category	Link types	Description
IP link	• Layer 3 Ethernet interface • Layer 3 Ethernet subinterface (including aggregate subinterface)	You can establish BFD sessions separately for an Ethernet interface and each of its subinterfaces. As a best practice, do not perform that operation.
Ethernet aggregate link	• Layer 2 Ethernet aggregate link • Member links of the Layer 2 Ethernet aggregate link • Layer 3 Ethernet aggregate link • Member links of the Layer 3 Ethernet aggregate link	You can establish BFD sessions separately for trunk member ports and the trunk port. As a best practice, do not perform that operation.
MPLS LSP	Statically established BFD sessions that can detect the following types of LSPs: • LDP LSPs Dynamic established BFD sessions that can detect the following types of LSPs: • LDP LSPs	The BFD session used to detect MPLS LSP connectivity is negotiated in one of the following ways: • **Statically establish the BFD session**—Manually configure the local discriminator and remote discriminator of the session, and use the negotiation mechanism provided by BFD to establish the BFD session. • **Dynamically establish the**

Link category	Link types	Description
		BFD session—The local discriminator is automatically assigned by the local device, and the remote discriminator is obtained during BFD session negotiation. Use the negotiation mechanism provided by BFD to establish the BFD session.
PW	Single segment of PW	Supports using both statically and dynamically established BFD sessions to detect PWs.

Technical benefits of BFD

BFD has the following technical benefits:

- BFD can detect the failures of the bidirectional forwarding paths of any type (including logical links, such as tunnels) between two network devices.
- BFD can provide services for upper-layer protocols and unified detection mechanisms.
- BFD provides millisecond-level detection speed to accelerate network convergence, reduce application interruption time, and enhance network reliability.

Protocols and standards

- RFC 5880, *Bidirectional Forwarding Detection (BFD)*
- RFC 5881, *Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)*
- RFC 5882, *Generic Application of Bidirectional Forwarding Detection (BFD)*
- RFC 5883, *Bidirectional Forwarding Detection (BFD) for Multihop Paths*
- RFC 7130, *Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces*

Restrictions and guidelines: BFD configuration

- By default, the device runs BFD version 1 and is compatible with BFD version 0. You cannot change the BFD version to 0 through commands. When the peer device runs BFD version 0, the local device automatically switches to BFD version 0.
- After a BFD session is established, the two ends negotiate BFD parameters, including minimum sending interval, minimum receiving interval, initialization mode, and packet authentication, by exchanging negotiation packets. They use the negotiated parameters without affecting the session status.
- As a best practice, use control-packet-mode BFD session on a downstream device to detect connectivity with an M-LAG device. Echo-packet-mode BFD sessions might fail to detect link failures.
- BFD session flapping might occur on an aggregate interface with member ports on different IRF member devices. When the master device, which receives and sends BFD packets, is removed or restarted, a subordinate device might not immediately take over. For example, a subordinate device will not take over when the subordinate device has a short detection time or a large number of BFD sessions.

- In a scenario where centralized devices form an IRF fabric and establish a BGP neighbor relationship with another device, using a multihop BFD session to detect link state might lead to BFD session flapping during IRF master/subordinate switchover when BFD packets cannot be sent or received in a timely manner. This can further result in BGP neighbor flapping. Before you use BFD in such a situation, make sure you understand the protection impact.
- You cannot specify a VSI interface as the output interface for a hardware BFD session.

BFD tasks at a glance

All the BFD tasks are optional.

9. Configuring BFD sessions in echo packet mode
10. Configuring BFD sessions in control packet mode
    - Configuring the session establishment mode
    - Configuring the detection mode
    - Creating a static BFD session
    - Configuring detection time settings
    - Configuring the authentication mode
    - Configuring the notification delay timer for session establishment failures
    - Enabling the echo function
    - Creating a static BFD session for detecting the aggregate link of a Layer 3 aggregate interface
    - Configuring the TTL value for BFD packets
11. Configuring BFD session flapping suppression
12. Delaying BFD session negotiation
13. Configuring a BFD template
14. Configuring event logging and SNMP notifications for BFD
    - Enabling BFD event logging
    - Enabling SNMP notifications for BFD

Configuring BFD sessions in echo packet mode

About BFD session creation methods

A BFD session can be created manually by using the **`bfd static`** command or created dynamically through collaboration between an application module and BFD.

Creating a static BFD session

About this task

A static BFD session in echo packet mode can be used to perform single-hop detection or multihop detection.

Restrictions and guidelines

You need to create a static BFD session in echo packet mode on only the local device to perform detection.

When creating a static BFD session, you must specify a peer IPv4 or IPv6 address. The system checks only the format of the IP address but not its correctness. If the peer IPv4 or IPv6 address is incorrect, the static BFD session cannot be established.

Different static BFD sessions cannot have the same local discriminator.

As a best practice, specify the source IP address for echo packets when creating a static BFD session. If you do not specify the source IP address, the device uses the IP address specified in the **bfd echo-source-ip** or **bfd echo-source-ipv6** command as the source IP address of echo packets.

Creating a static BFD session for single-hop detection

15. Enter system view.

system-view

16. Configure the source IP address of echo packets.

- Configure the source IP address of echo packets.

  bfd echo-source-ip *ip-address*

  By default, no source IPv4 address is configured for echo packets.

  As a best practice, do not configure the source IPv4 address to be on the same network segment as any local interfaces. This can avoid network congestion caused a large number of ICMP redirect packets sent from the peer.

- Configure the source IPv6 address of echo packets.

  bfd echo-source-ipv6 *ipv6-address*

  By default, no source IPv6 address is configured for echo packets.

  The source IPv6 address of echo packets can only be a global unicast address.

17. Create a static BFD session and enter static BFD session view.

IPv4:

bfd static *session-name* [**peer-ip** *ipv4-address* **interface** *interface-type interface-number* **destination-ip** *ipv4-address* [**source-ip** *ipv4-address*] **one-arm-echo** [**discriminator** { **auto** | **local** *local-value* }]]

IPv6:

bfd static *session-name* [**peer-ipv6** *ipv6-address* **interface** *interface-type interface-number* **destination-ipv6** *ipv6-address* [**source-ipv6** *ipv6-address*] **one-arm-echo** [**discriminator** { **auto** | **local** *local-value* }]]

18. (Optional.) Specify a local discriminator for the static BFD session.

discriminator local *local-value*

By default, no local discriminator is specified.

You can use this command only if you do not specify a local discriminator when creating a static BFD session.

Creating a static BFD session for multihop detection

19. Enter system view.

system-view

20. (Optional.) Configure the source IP address of echo packets.

- Configure the source IPv4 address of echo packets.

  bfd echo-source-ip *ip-address*

  By default, no source IPv4 address is configured for echo packets.

As a best practice, do not configure the source IPv4 address to be on the same network segment as any local interfaces. If you configure such a source IPv4 address, a large number of ICMP redirect packets might be sent from the peer, resulting in link congestion.

- Configure the source IPv6 address of echo packets.

  bfd echo-source-ipv6 *ipv6-address*

  By default, no source IPv6 address is configured for echo packets.

  The source IPv6 address of echo packets can only be a global unicast address.

21. Create a static BFD session and enter static BFD session view.

IPv4:

bfd static *session-name* [**peer-ip** *ipv4-address* [**vpn-instance** *vpn-instance-name*] **destination-ip** *ipv4-address* [**source-ip** *ipv4-address*] **one-arm-echo**[**discriminator** { **auto** | **local** *local-value* }]]

IPv6:

bfd static *session-name* [**peer-ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*] **destination-ipv6** *ipv6-address* [**source-ipv6** *ipv6-address*] **one-arm-echo** [**discriminator** { **auto** | **local** *local-value* }]]

22. (Optional.) Specify a local discriminator for the static BFD session.

discriminator local *local-value*

By default, no local discriminator is specified.

You can use this command only if you do not specify a local discriminator when creating a static BFD session.

Configuring detection time settings

About this task

When creating a BFD session in echo packet mode, you can configure the minimum interval for receiving BFD echo packets and the detection time multiplier for the device.

Restrictions and guidelines

When the same BFD session (with the same source IP address, destination IP address, interface, VPN, and destination port number) is associated with multiple upper layer applications, BFD selects the detection time settings for the BFD session as follows:

23. Minimum values of the following settings:

- Settings configured in static BFD session view.
- Settings configured in the BFD template specified by the upper layer application.
- Settings configured by the upper layer application.

24. Settings configured in interface view or system view.

25. Default settings.

When a BFD session is associated with only one upper layer application, BFD selects the detection time settings for the BFD session as follows:

26. BFD selects the detection time settings for the BFD session based on whether the upper layer application supports using a BFD template and configuring detection time settings:

- If a BFD template is specified and detection time settings are also configured for the upper layer application, BFD selects the settings configured in the specified BFD template.
- If a BFD template is specified for the upper layer application that does not support configuring detection time settings, BFD selects the settings configured in the specified BFD template.

- If detection time settings are configured for the upper layer application that does not support using a BFD template, BFD selects the configured settings.

27. Settings configured in interface view or system view.
28. Default settings.

Configure detection time settings for single-hop BFD detection

29. Enter system view.
    system-view
30. Enter interface view or static BFD session view.
    - Enter interface view.
      interface *interface-type interface-number*
    - Enter static BFD session view.
      bfd static *session-name*

      The static BFD session must already exist.

      To configure parameters for a static BFD session, you must enter its view.
31. Set the minimum interval for receiving BFD echo packets.
    bfd min-echo-receive-interval *interval*

    The default setting is 400 milliseconds.
32. Set the detection time multiplier.
    bfd detect-multiplier *value*

    The default setting is 5.

Configure detection time settings for multihop BFD detection

33. Enter system view.
    system-view
34. (Optional.) Enter static BFD session view.
    bfd static *session-name*

    The static BFD session must already exist.

    To configure parameters for a static BFD session, you must enter its view.
35. Set the minimum interval for receiving BFD echo packets.
    bfd multi-hop min-echo-receive-interval *interval*

    The default setting is 400 milliseconds.
36. Set the detection time multiplier.
    bfd multi-hop detect-multiplier *value*

    The default setting is 5.

Configuring BFD sessions in control packet mode

About BFD session creation methods

BFD sessions in control packet mode can be created statically or established dynamically.

BFD sessions are distinguished by the local discriminator and remote discriminator in control packets. The main difference between a statically created session and a dynamically established session is that they obtain the local discriminator and remote discriminator in different ways.

- The local discriminator and remote discriminator of a static BFD session are specified manually in the **bfd static** command or in the commands that associate specific applications with BFD.
- The local discriminator of a dynamic BFD session is assigned by the device, and the remote discriminator is obtained during BFD session negotiation. The device can automatically assign the local discriminator to the BFD session in the following conditions:
  - The **auto** keyword is specified for the **bfd static** command.
  - The local and remote discriminators are not specified for the BFD session associated with a specific application.

Restrictions and guidelines

After an upper-layer protocol is configured to support BFD, the device automatically creates BFD sessions in c**ontrol packet mode. You do not need to perform this task.**

Configuring the session establishment mode

About this task

BFD can use the following operating modes to establish a session:

- **Active mode**—BFD actively sends BFD control packets regardless of whether any BFD control packet is received from the peer.
- **Passive mode**—BFD does not send control packets until a BFD control packet is received from the peer.

At least one end must operate in active mode for a BFD session to be established.

Procedure

37. Enter system view.

```
system-view
```

38. Configure the BFD session establishment mode.

```
bfd session init-mode { active | passive }
```

By default, BFD uses the active mode to establish a session.

Configuring the detection mode

About this task

After a BFD session is established, the two ends can operate in the following modes:

- **Asynchronous mode**—The device periodically sends BFD control packets. The device considers that the session is down if it does not receive any BFD control packets within the detection interval.
- **Demand mode**—The Demand mode can be used to reduce the overhead when a large number of BFD sessions exist.
  - If the peer end is operating in Asynchronous mode (default), the peer end stops sending BFD control packets after receiving control packets with the D bit set. In this scenario, BFD detects only the link between the local end and the peer end. The device considers that the session is down if it does not receive any BFD control packets within the detection interval.
  - If the peer end is operating in Demand mode, both ends stop sending BFD control packets. In this scenario, the system uses features other than BFD, such as Hello packets and hardware CC, to verify the connectivity to another system.

Restrictions and guidelines

The **bfd demand enable** command cannot take effect when any of the following conditions exist:

- Echo-packet-mode BFD sessions exist on the device.
- BFD sessions for detecting aggregate links exist on the device.
- After the **bfd detect-interface source-ip** command is configured, the **bfd detect-interface first-fail-timer** and **bfd detect-interface special-processing** commands are configured.

Procedure

39. Enter system view.

system-view

40. Enter interface view.

interface *interface-type interface-number*

41. Specify the Demand mode for the BFD session.

bfd demand enable

By default, a BFD session is in Asynchronous mode.

Creating a static BFD session

About this task

A static BFD session can be used for single-hop detection and multihop detection. By working with Track, a static BFD session can provide fast failure detection. For more information about Track association with BFD, see "Configuring Track."

Restrictions and guidelines for static BFD session configuration

If a static BFD session is created on the remote end, the static BFD session must be created on the local end.

When creating a static BFD session, you must specify a peer IPv4 or IPv6 address. The system checks only the format of the IP address but not its correctness. If the peer IPv4 or IPv6 address is incorrect, the static BFD session cannot be established.

Different static BFD sessions cannot have the same local discriminator.

Creating a static BFD session for single-hop detection

42. Enter system view.

system-view

43. Create a static BFD session and enter static BFD session view.

IPv4:

bfd static *session-name* **peer-ip** *ipv4-address* **interface** *interface-type interface-number* **source-ip** *ipv4-address* [**discriminator** { **auto** | **local** *local-value* **remote** *remote-value* }]

For a static BFD session to be established, specify the IPv4 address of the peer interface where the static BFD session resides for the **peer-ip** *ipv4-address* option. Specify the IPv4 address of the local interface where the static BFD session resides for the **source-ip** *ipv4-address* option.

IPv6:

bfd static *session-name* **peer-ipv6** *ipv6-address* **interface** *interface-type interface-number* **source-ipv6** *ipv6-address* [**discriminator** { **auto** | **local** *local-value* **remote** *remote-value* }]

For a static BFD session to be established, specify the IPv6 address of the peer interface where the static BFD session resides for the **peer-ipv6** *ipv6-address* option. Specify the IPv6 address of the local interface where the static BFD session resides for the **source-ipv6** *ipv6-address* option.

44. (Optional.) Specify the local and remote discriminators for the static BFD session.

discriminator { **local** *local-value* | **remote** *remote-value* }

By default, no local discriminator or remote discriminator is specified for a static BFD session.

Use this command only if you do not specify the local or remote discriminator when creating a static BFD session.

Create a static BFD session for multihop detection

45. Enter system view.

system-view

46. Create a static BFD session and enter static BFD session view.

IPv4:

bfd static *session-name* **peer-ip** *ipv4-address* [**vpn-instance** *vpn-instance-name*] **source-ip** *ipv4-address* [**discriminator** { **auto** | **local** *local-value* **remote** *remote-value* }] [**track-interface** *interface-type interface-number*]

IPv6:

bfd static *session-name* **peer-ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*] **source-ipv6** *ipv6-address* [**discriminator** { **auto** | **local** *local-value* **remote** *remote-value* }] [**track-interface** *interface-type interface-number*]

47. (Optional.) Specify the local and remote discriminators for the static BFD session.

discriminator { **local** *local-value* | **remote** *remote-value* }

By default, no local discriminator or remote discriminator is specified for a static BFD session.

Use this command only if you do not specify the local or remote discriminator when creating a static BFD session.

Configuring detection time settings

About this task

When creating a BFD session in control packet mode, you can configure the minimum interval for receiving and transmitting BFD control packets and the detection time multiplier for the device.

Restrictions and guidelines

When the same BFD session (with the same source IP address, destination IP address, interface, VPN, and destination port number) is associated with multiple upper layer applications, BFD selects the detection time settings for the BFD session as follows:

1. Minimum values of the following settings:
   - Settings configured in static BFD session view.
   - Settings configured in the BFD template specified by the upper layer application.
   - Settings configured by the upper layer application.
2. Settings configured in interface view or system view.
3. Default settings.

When a BFD session is associated with only one upper layer application, BFD selects the detection time settings for the BFD session as follows:

1. BFD selects the detection time settings for the BFD session based on whether the upper layer application supports using a BFD template and configuring detection time settings:
   - If a BFD template is specified and detection time settings are also configured for the upper layer application, BFD selects the settings configured in the specified BFD template.
   - If a BFD template is specified for the upper layer application that does not support configuring detection time settings, BFD selects the settings configured in the specified BFD template.
   - If detection time settings are configured for the upper layer application that does not support using a BFD template, BFD selects the configured settings.
2. Settings configured in interface view or system view.
3. Default settings.

You can use the **bfd min-receive-interval** and **bfd min-control-interval** commands to adjust the minimum interval for receiving single-hop BFD control packets. Follow these guidelines when you configure the commands:

- If you configure both the **bfd min-receive-interval** and **bfd min-control-interval** commands, the time interval configured by the **bfd min-receive-interval** command takes precedence.
- If you do not configure the **bfd min-receive-interval** or **bfd min-control-interval** command, the device uses the default value for the **bfd min-receive-interval** command as the minimum interval for receiving single-hop BFD control packets.
- If you configure either the **bfd min-receive-interval** or **bfd min-control-interval** command, the device uses the value configured for that command as the minimum interval for receiving single-hop BFD control packets.

You can use the **bfd min-transmit-interval** and **bfd min-control-interval** commands to adjust the minimum interval for transmitting single-hop BFD control packets. Follow these guidelines when you configure the commands:

- If you configure both the **bfd min-transmit-interval** and **bfd min-control-interval** commands, the time interval configured by the **bfd min-transmit-interval** command takes precedence.
- If you do not configure the **bfd min-transmit-interval** or **bfd min-control-interval** command, the device uses the default value for the **bfd min-transmit-interval** command as the minimum interval for transmitting single-hop BFD control packets.
- If you configure either the **bfd min-transmit-interval** or **bfd min-control-interval** command, the device uses the value configured for that command as the minimum interval for transmitting single-hop BFD control packets.

You can use the **bfd multi-hop min-receive-interval** and **bfd multi-hop min-control-interval** commands to adjust the minimum interval for receiving multihop BFD control packets. Follow these guidelines when you configure the commands:

- If you configure both the **bfd multi-hop min-receive-interval** and **bfd multi-hop min-control-interval** commands, the time interval configured by the **bfd multi-hop min-receive-interval** command takes precedence.
- If you do not configure the **bfd multi-hop min-receive-interval** or **bfd multi-hop min-control-interval** command, the device uses the default value for the **bfd multi-hop min-receive-interval** command as the minimum interval for receiving multihop BFD control packets.

- If you configure either the **bfd multi-hop min-receive-interval** or **bfd multi-hop min-control-interval** command, the device uses the value configured for that command as the minimum interval for receiving multihop BFD control packets.

You can use the **bfd multi-hop min-transmit-interval** and **bfd multi-hop min-control-interval** commands to adjust the minimum interval for transmitting multihop BFD control packets. Follow these guidelines when you configure the commands:

- If you configure both the **bfd multi-hop min-transmit-interval** and **bfd multi-hop min-control-interval** commands, the time interval configured by the **bfd multi-hop min-transmit-interval** command takes precedence.
- If you do not configure the **bfd multi-hop min-transmit-interval** or **bfd multi-hop min-control-interval** command, the device uses the default value for the **bfd multi-hop min-transmit-interval** command as the minimum interval for transmitting multihop BFD control packets.
- If you configure either the **bfd multi-hop min-transmit-interval** or **bfd multi-hop min-control-interval** command, the device uses the value configured for that command as the minimum interval for transmitting multihop BFD control packets.

Configure detection time settings for single-hop BFD detection

1. Enter system view.
   system-view
2. Enter interface view or static BFD session view.
   - Enter interface view.
     interface *interface-type interface-number*
   - Enter static BFD session view.
     bfd static *session-name*
     To configure parameters for a static BFD session, you must enter its view.
3. Set the minimum interval for receiving echo packets.
   bfd min-echo-receive-interval *interval*
   The default setting is 400 milliseconds.
4. Set the minimum interval for transmitting and receiving single-hop BFD control packets.
   - Set the minimum interval for transmitting single-hop BFD control packets.
     bfd min-transmit-interval *interval*
     The default setting is 400 milliseconds.
   - Set the minimum interval for receiving single-hop BFD control packets.
     bfd min-receive-interval *interval*
     The default setting is 400 milliseconds.
5. Set the minimum interval for transmitting and receiving single-hop BFD control packets.
   bfd min-control-interval *interval*
   The default setting is 400 milliseconds.
6. Set the single-hop detection time multiplier.
   bfd detect-multiplier *value*
   The default setting is 5.

Configure detection time settings for multihop BFD detection

7. Enter system view.
   system-view
8. (Optional.) Enter static BFD session view.

bfd static *session-name*

To configure parameters for a static BFD session, you must enter its view.

9. Set the multihop detection time multiplier.

   bfd multi-hop detect-multiplier *value*

   The default setting is 5.

10. Set the minimum interval for transmitting and receiving multihop BFD control packets.
    - Set the minimum interval for transmitting multihop BFD control packets.

      bfd multi-hop min-transmit-interval *interval*

      The default setting is 400 milliseconds.
    - Set the minimum interval for receiving multihop BFD control packets.

      bfd multi-hop min-receive-interval *interval*

      The default setting is 400 milliseconds.

11. Set the minimum interval for transmitting and receiving multihop BFD control packets.

    bfd multi-hop min-control-interval *interval*

    By default, the minimum interval for transmitting and receiving multihop BFD control packets is 400 milliseconds.

Configuring the authentication mode

About this task

You can configure authentication settings, including algorithm and key, to enhance BFD session security.

Configure the authentication mode for a single-hop BFD session

12. Enter system view.

    system-view

13. Enter interface view or static BFD session view.
    - Enter interface view.

      interface *interface-type interface-number*
    - Enter static BFD session view.

      bfd static *session-name*

      The static BFD session must already exist.

      To configure parameters for a static BFD session, you must enter its view.

14. Configure the authentication mode for single-hop control packets.

    bfd authentication-mode { **hmac-md5** | **hmac-mmd5** | **hmac-msha1** | **hmac-sha1** | **m-md5** | **m-sha1** | **md5** | **sha1** | **simple** } *key-id* { **cipher** *cipher-string* | **plain** *plain-string* }

    By default, no authentication mode is configured for single-hop control packets.

Configure the authentication mode for a multihop BFD session

15. Enter system view.

    system-view

16. (Optional.) Enter static BFD session view.

    bfd static *session-name*

    The static BFD session must already exist.

    To configure parameters for a static BFD session, you must enter its view.

17. Configure the authentication mode for multihop BFD control packets.

    bfd multi-hop authentication-mode { **hmac-md5** | **hmac-mmd5** | **hmac-msha1** | **hmac-sha1** | **m-md5** | **m-sha1** | **md5** | **sha1** | **simple** } *key-id* { **cipher** *cipher-string* | **plain** *plain-string* }

    By default, no authentication mode is configured for multihop BFD control packets.

Configuring the notification delay timer for session establishment failures

About this task

By default, BFD does not notify upper-layer protocols of session establishment failures. The notification is required in some scenarios. For example, upon a session establishment failure in an aggregate link, the aggregate link can place the associated member port in Unselected state based on the failure notification. You can configure the delay timer for BFD to notify upper-layer protocols of session establishment failures.

Restrictions and guidelines

This feature does not apply to BFD sessions in echo packet mode.

Procedure

1. Enter system view.

   system-view

2. Set the delay timer for BFD to notify upper-layer protocols of session establishment failures.

   bfd init-fail-timer *seconds*

   By default, BFD does not notify upper-layer protocols of session establishment failures.

CAUTION:

For session establishment failures caused by configuration mismatches at the two ends, this command can cause the upper-layer protocol to act incorrectly. Therefore, use this command with caution. BFD status mismatch and BFD authentication configuration mismatch are examples of configuration mismatches.

Enabling the echo function

About this task

When you use Asynchronous mode BFD to detect the connectivity between directly connected devices, you can enable the echo function. This function enables the local system to periodically send echo packets to the remote system and reduces the control packet receiving rate to save bandwidth usage. The remote system loops back the echo packets to the local system without processing them. If the local system does not receive echo packets looped back from the remote system in a consecutive number of times, the local system declares the BFD session down.

This function is supported only for single-hop detection.

Restrictions and guidelines

This function does not take effect on BFD sessions associated with interface states.

Procedure

3. Enter system view.

   system-view

4. Enter interface view or static BFD session view.
   - Enter interface view.
     interface *interface-type interface-number*
   - Enter static BFD session view.
     bfd static *session-name*

     The static BFD session must already exist.
5. Enable the echo function.
   bfd echo [**receive** | **send**] **enable**

   By default, the echo function is disabled.

Creating a static BFD session for detecting the aggregate link of a Layer 3 aggregate interface

About this task

To use BFD to detect an aggregate link, choose one as needed:

- Method 1: Execute the **link-aggregation bfd** command to enable BFD for an aggregation group.
- Method 2: Execute the **bfd static** command with the **bind-route-aggregation** keyword specified to create static BFD sessions for detecting an aggregate link and its member links.

If you use Method 1, the Ethernet link aggregation module will mistakenly consider the aggregation link unavailable when the BFD session goes down due to the failure of a single member link. To avoid this issue, use Method 2. To perform multihop detection, use Method 1, because Method 2 only supports single-hop detection.

The operating mechanisms and restrictions and guidelines of the two methods are as shown in Table 4. You can choose only one of the methods.

Table 4 Methods of using BFD to detect an aggregate link

Method	Operating mechanism	Single-hop/multihop detection	Restrictions and guidelines
Method 1	Create a BFD session for detecting the aggregate link. Each member interface of the aggregate link takes turns to send BFD packets.	Both single-hop detection and multihop detection are supported.	• Make sure the source and destination IP addresses are reversed between the two ends of the aggregate link. For example, if you execute **link-aggregation bfd ipv4 source** *1.1.1.1* **destination** *2.2.2.2* at the local end, execute **link-aggregation bfd ipv4 source** *2.2.2.2* **destination** *1.1.1.1* at the peer end. The source and destination IP addresses must be different unicast addresses other than 0.0.0.0. • You can configure only one static BFD session to detect the aggregate link of an aggregate interface. • You cannot execute the **bfd static bind-route-aggregation** command for aggregate sub-interfaces.

Method	Operating mechanism	Single-hop/multihop detection	Restrictions and guidelines
Method 2	1. Create a static BFD session to associate with the aggregated link. 2. Create a static BFD sub-session for detecting each member interface of the aggregated link. 3. The BFD session is in up state as long as a sub-session is in up state, indicating the aggregated link is available. The BFD session goes down only when all sub-sessions are in down state, indicating the aggregated link is unavailable.	Only single-hop detection is supported.	• Make sure the source and destination IP addresses are reversed between the two ends. For example, if you execute **bfd static** *abc* **bind-route-aggregation peer-ipv4** *1.1.1.1* **interface** *route-aggregation 1* **source-ipv4** *1.1.1.2* **discriminator auto** at the local end, execute **bfd static** *abc* **bind-route-aggregation peer-ipv4** *1.1.1.2* **interface** *route-aggregation 1* **source-ipv4** *1.1.1.1* **discriminator auto** at the peer end. The source and destination IP addresses must be different unicast addresses other than 0.0.0.0. • You can configure only one static BFD session to detect the aggregate link of an aggregate interface. • You cannot execute the **bfd static bind-route-aggregation** command for aggregate sub-interfaces.

Procedure

6. Enter system view.

   system-view

7. Create a static BFD session and enter static BFD session view.

   IPv4:

   bfd static *session-name* [**bind-route-aggregation peer-ip** *ipv4-address* **interface** *interface-type interface-number* **source-ip** *ipv4-address* **discriminator auto**]

   Make sure the source and destination IP addresses are reversed between the two ends of the aggregate link. The source and destination IP addresses must be different unicast addresses other than 0.0.0.0.

   IPv6:

   bfd static *session-name* [**bind-route-aggregation peer-ipv6** *ipv6-address* **interface** *interface-type interface-number* **source-ipv6** *ipv6-address* **discriminator auto**]

   Make sure the source and destination IP addresses are reversed between the two ends of the aggregate link. The source and destination IP addresses must be different unicast addresses other than 0::0.

Configuring the TTL value for BFD packets

About this task

When you connect an H3C device to a third-party device, for successful BFD session negotiation, make sure the TTL value settings for BFD packets on both ends of the BFD session are the same. When the device receives a BFD packet in DOWN or INIT state from its peer, it verifies the TTL value of the packet and performs one of the following actions:

- For a single-hop BFD session, if the device receives a packet that carries a TTL value that is different from the TTL value specified with the **bfd ttl** command, the device drops the packet.
- For a multihop BFD session, if the device receives a packet that carries a TTL value that is greater than the TTL value specified with the **bfd ttl** command, the device drops the packet.
- If the BFD session is already up, the device sets the BFD session state to AdminDown, which triggers a BFD session renegotiation.

Restrictions and guidelines

- The **bfd ttl** command does not take effect on BFD sessions in echo packet mode.
- For an IPv4 or IPv6 address with different mask lengths or prefix lengths, the device uses the TTL value specified with the longest mask or prefix length.
- For an IPv4 or IPv6 subnet, the TTL value of single-hop BFD packets must be greater than the TTL value of multihop BFD packets.

Procedure

8. Enter system view.

system-view

9. Specify the TTL value for BFD packets.

bfd { **peer-ip** *ipv4-address mask-length* | **peer-ipv6** *ipv6-address prefix-length* } **ttl** { **single-hop** | **multi-hop** } *ttl-value*

By default, the TTL value of single-hop BFD packets is 255 and the TTL value of multihop BFD packets is 64.

Configuring BFD session flapping suppression

About this task

When BFD detects a link failure, it tears down the BFD session and notifies the upper-layer protocol of the failure. When the upper-layer protocol re-establishes a neighbor relationship, the BFD session comes up again. BFD session flaps occur when a link fails and recovers repeatedly, which consumes significant system resources and causes network instability.

This feature allows you to suppress BFD session flapping by using the *initial-interval*, *secondary-interval*, and *maximum-interval* arguments.

- A BFD session is suppressed within the specified interval. The suppression time does not exceed the *maximum-interval*.
- After a BFD session goes down for the second time, it cannot be re-established within the *initial-interval*.
- After a BFD session goes down for the third time, it cannot be re-established within the *secondary-interval*.
- After a BFD session goes down for the fourth time and at any later time, the following rules apply:

- If *secondary-interval* × 2^{n-3} is smaller than or equal to the *maximum-interval*, the BFD session cannot be re-established within the *secondary-interval* × 2^{n-3}.
- If *secondary-interval* × 2^{n-3} is greater than the *maximum-interval*, the BFD session cannot be re-established within the *maximum-interval*.

The letter n, starting from 4, is the number of times the BFD session flaps.

Procedure

10. Enter system view.
    system-view
11. Configure BFD session flapping suppression.
    bfd dampening [**maximum** *maximum-interval* **initial** *initial-interval* **secondary** *secondary-interval*]
    By default, BFD sessions are not suppressed.
    The values for the *initial-interval* and *secondary-interval* arguments cannot be greater than or equal to the value for the *maximum-interval* argument.

Delaying BFD session negotiation

About this task

If an upper-layer protocol uses BFD to detect the active path, packet loss might occur in the following situations:

- An active/standby path switchover is triggered upon a device reboot. If the BFD session comes up after a device reboot, the traffic on the standby path will be switched over to the active path. In this case, traffic loss will occur if the active path of the upper-layer protocol has not recovered completely.
- An active/standby path switchover is triggered after the active path goes down. If the BFD session comes up before the active path recovers, the traffic on the standby path will be switched over to the active path. In this case, traffic loss will occur.

To solve the traffic loss problem, you can delay BFD session negotiation.

Delaying BFD session negotiation upon a device reboot

About this task

This function starts a delay timer when the device reboots successfully. Before this timer expires, the device does not perform session negotiation on BFD sessions in down state and newly created sessions.

Procedure

12. Enter system view.
    system-view
13. Enable BFD session negotiation delay upon a device reboot and set a delay timer.
    bfd session-negotiation delay-upon-reboot *interval*
    By default, BFD session negotiation delay is disabled for a device reboot.

Delaying BFD session negotiation for down BFD sessions

About this task

This function starts a delay timer for each BFD session from up to down state and each newly created session. Before the timer expires, the device does not perform session negotiation on each BFD session.

Procedure

14. Enter system view.
    system-view
15. Enable session negotiation delay for down BFD sessions and set a delay timer.
    bfd session-negotiation delay-upon-down *interval*
    By default, session negotiation delay is disabled for down BFD sessions.

Configuring a BFD template

About this task

Perform this task to specify BFD parameters in a template for sessions without next hops. You can configure BFD parameters for LSPs and PWs through a BFD template.

This task can also be used to specify BFD parameters for BFD sessions that detect interface states.

Procedure

16. Enter system view.
    system-view
17. Create a BFD template and enter BFD template view.
    bfd template *template-name*
18. (Optional.) Configure the authentication mode for BFD control packets.
    bfd authentication-mode { **hmac-md5** | **hmac-mmd5** | **hmac-msha1** | **hmac-sha1** | **m-md5** | **m-sha1** | **md5** | **sha1** | **simple** } *key-id* { **cipher** *cipher-string* | **plain** *plain-string* }
    By default, no authentication is performed.
19. Set the detection time multiplier.
    bfd detect-multiplier *value*
    The default setting is 5.
20. Set the minimum interval for transmitting and receiving BFD control packets.
    - Set the minimum interval for transmitting BFD control packets.
      bfd min-transmit-interval *interval*
      The default setting is 400 milliseconds.
    - Set the minimum interval for receiving BFD control packets.
      bfd min-receive-interval *interval*
      The default setting is 400 milliseconds.

Configuring event logging and SNMP notifications for BFD

Enabling BFD event logging

About this task

Use this feature to enable the device to record the following types of BFD event logs to help you locate the reason for BFD session state anomalies:

- BFD session down events.
- Driver packet-sending events.

Procedure

21. Enter system view.

system-view

22. Enable BFD event logging.

bfd event-log enable { **down** [*log-size*] | **driver-packet** [**detection-abnormal-autostop**] }

By default, BFD event logging is disabled.

You can repeat this command to modify the maximum number of logs allowed for BFD session down events. After modification, the device clears recorded logs.

The device can record a maximum number of 8192 logs for driver packet-sending events.

Enabling SNMP notifications for BFD

About this task

To report critical BFD events to an NMS, enable SNMP notifications for BFD. For BFD event notifications to be sent correctly, you must also configure SNMP as described in *Network Management and Monitoring Configuration Guide*.

Procedure

23. Enter system view.

system-view

24. Enable SNMP notifications for BFD.

snmp-agent trap enable bfd

By default, SNMP notifications are enabled for BFD.

Display and maintenance commands for BFD

Execute the **display** command in any view and the **reset** command in user view.

Task	Command
Display BFD session information.	**display bfd session** [**discriminator local** *local-value* | **static** *session-name* | **verbose**] **display bfd session** [[**dynamic**] [**control** | **echo**] [**ip**] [**state** { **down** | **admin-down** | **init** | **up** }] [**discriminator remote** *remote-value*] [**peer-ip** *ipv4-address* [**vpn-instance** *vpn-instance-name*]] [**verbose**]] **display bfd session** [[**dynamic**] [**control** | **echo**] [**ipv6**] [**state** { **down** | **admin-down** | **init** | **up** }] [**discriminator remote** *remote-value*] [**peer-ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*]] [**verbose**]] **display bfd session** [[**static**] [**ip**] [**state** { **down** | **admin-down** | **init** | **up** }] [**discriminator remote** *remote-value*] [**peer-ip** *ipv4-address* [**vpn-instance** *vpn-instance-name*]] [**verbose**] **display bfd session** [[**static**] [**ipv6**] [**state** { **down** | **admin-down** | **init** | **up** }] [**discriminator remote** *remote-value*] [**peer-ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*]] [**verbose**]
Display information about static BFD sessions created to detect the aggregate links and member links of Layer 3 aggregate interfaces.	**display bfd session route-aggregation** [**interface** *interface-type interface-number*]
Display the TTL values for BFD packets.	**display bfd ttl**
Clear BFD session statistics.	**reset bfd session statistics**

Contents

Configuring Track 1

About Track 1
Collaboration mechanism 1
Supported detection modules 2
Supported application modules 2
Restrictions and guidelines: Track configuration 2
Collaboration application example 3
Track tasks at a glance 3
Associating the Track module with a detection module 4
Associating Track with NQA 4
Associating Track with BFD 4
Associating Track with interface management 5
Associating Track with route management 6
Associating Track with LLDP 7
Associating Track with M-LAG MAD state 7
Associating Track with PTP 8
Associating Track with a tracked list 8
Associating Track with a Boolean list 8
Associating Track with a percentage threshold list 9
Associating Track with a weight threshold list 10
Associating the Track module with an application module 10
Prerequisites for associating the Track module with an application module 10
Associating Track with VRRP 11
Associating Track with static routing 12
Associating Track with PBR 13
Associating Track with traffic redirection 15
Associating Track with EAA 15
Associating Track with PoE 16
Display and maintenance commands for Track 17
Track configuration examples 17
Example: Configuring VRRP-Track-NQA collaboration 17
Example: Configuring an echo-mode BFD session for a VRRP backup to monitor the master 20
Example: Configuring a control-mode BFD session for a VRRP backup to monitor the master 23
Example: Configuring an echo-mode BFD session for the VRRP master to monitor the uplink 26
Example: Configuring a control-mode BFD session for the VRRP master to monitor the uplink 30
Example: Configuring static routing-Track-NQA collaboration 33
Example: Configuring static routing-Track-BFD (echo mode) collaboration 38
Example: Configuring static routing-Track-BFD (control mode) collaboration 41
Example: Configuring VRRP-Track-interface management collaboration 45
Example: Configuring VRRP-Track-route management collaboration 47
Example: Configuring static routing-Track-LLDP collaboration 50
Example: Configuring NTP-PTP-Track collaboration 54

Configuring Track

About Track

The Track module works between application modules and detection modules. It shields the differences between various detection modules from application modules.

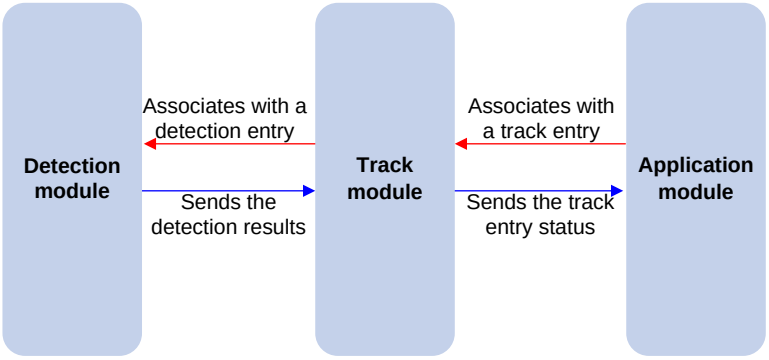
Collaboration mechanism

The Track module collaborates with detection modules and application modules.

As shown in Figure 1, collaboration is enabled when you associate the Track module with a detection module and an application module, and it operates as follows:

1. The detection module probes specific objects such as interface status, link status, network reachability, and network performance, and informs the Track module of detection results.
2. The Track module sends the detection results to the application module.
3. When notified of changes for the tracked object, the application modules can react to avoid communication interruption and network performance degradation.

Figure 1 Collaboration through the Track module



Collaboration between the Track module and a detection module

The detection module sends the detection result of the tracked object to the Track module. The Track module changes the status of the track entry as follows:

- If the tracked object operates correctly, the state of the track entry is Positive. For example, the track entry state is Positive in one of the following conditions:
  - The target interface is up.
  - The target network is reachable.
- If the tracked object does not operate correctly, the state of the track entry is Negative. For example, the track entry state is Negative in one of the following conditions:
  - The target interface is down.
  - The target network is unreachable.
- If the detection result is invalid, the state of the track entry is NotReady. For example, the track entry state is NotReady if its associated NQA operation does not exist.

Collaboration between the Track module and an application module

The track module reports the track entry status changes to the application module. The application module can then take correct actions to avoid communication interruption and network performance degradation.

Supported detection modules

The following detection modules can be associated with the Track module:

- NQA.
- BFD.
- Interface management.
- Route management.
- LLDP.
- M-LAG.
- PTP.

You can also associate a track entry with a list of objects called a tracked list. The state of a tracked list is determined by the states of all objects in the list. The following types of tracked lists are supported:

- **Boolean AND list**—The state of a Boolean AND list is determined by the states of the tracked objects using the Boolean AND operation.
- **Boolean OR list**—The state of a Boolean OR list is determined by the states of the tracked objects using the Boolean OR operation.
- **Percentage threshold list**—The state of a percentage threshold list is determined by comparing the percentage of positive objects in the list with the percentage thresholds configured for the list.
- **Weight threshold list**—The state of a weight threshold list is determined by comparing the weight of positive objects in the list with the weight thresholds configured for the list.

Supported application modules

The following application modules can be associated with the Track module:

- VRRP.
- Static routing.
- PBR.
- Traffic redirection.
- Smart Link.
- Embedded Automation Architecture (EAA).
- Power over Ethernet (PoE).

Restrictions and guidelines: Track configuration

When configuring a track entry for an application module, you can set a notification delay to avoid immediate notification of status changes.

When the delay is not configured and the route convergence is slower than the link state change notification, communication failures occur. For example, when the master in a VRRP group detects an uplink interface failure through Track, Track immediately notifies the master to decrease its priority. A backup with a higher priority then preempts as the new master. When the failed uplink interface recovers, the Track module immediately notifies the original master to restore its priority. If the uplink route has not recovered, forwarding failure will occur.

Collaboration application example

The following is an example of collaboration between NQA, Track, and static routing.

Configure a static route with next hop 192.168.0.88 on the device. If the next hop is reachable, the static route is valid. If the next hop becomes unreachable, the static route is invalid. For this purpose, configure NQA-Track-static routing collaboration as follows:

1. Create an NQA operation to monitor the accessibility of IP address 192.168.0.88.
2. Create a track entry and associate it with the NQA operation.
   - When next hop 192.168.0.88 is reachable, NQA sends the result to the Track module. The Track module sets the track entry to Positive state.
   - When the next hop becomes unreachable, NQA sends the result to the Track module. The Track module sets the track entry to Negative state.
3. Associate the track entry with the static route.
   - When the track entry is in Positive state, the static routing module considers the static route to be valid.
   - When the track entry is in Negative state, the static routing module considers the static route to be invalid.

Track tasks at a glance

To implement the collaboration function, establish associations between the Track module and detection modules, and between the Track module and application modules.

To configure the Track module, perform the following tasks:

1. Associating the Track module with a detection module
   - Associating Track with NQA
   - Associating Track with BFD
   - Associating Track with interface management
   - Associating Track with route management
   - Associating Track with LLDP
   - Associating Track with M-LAG MAD state
   - Associating Track with PTP
2. Associating Track with a tracked list
   - Associating Track with a Boolean list
   - Associating Track with a percentage threshold list
   - Associating Track with a weight threshold list
3. Associating the Track module with an application module
   - Associating Track with VRRP
   - Associating Track with static routing
   - Associating Track with PBR
   - Associating Track with traffic redirection
   - Associating Track with EAA
   - Associating Track with PoE

Associating the Track module with a detection module

Associating Track with NQA

About this task

NQA supports multiple operation types to analyze network performance and service quality. For example, an NQA operation can periodically detect whether a destination is reachable, or whether a TCP connection can be established.

An NQA operation operates as follows when it is associated with a track entry:

- If the consecutive probe failures reach the specified threshold, the NQA module notifies the Track module that the tracked object has malfunctioned. The Track module then sets the track entry to Negative state.
- If the specified threshold is not reached, the NQA module notifies the Track module that the tracked object is operating correctly. The Track module then sets the track entry to Positive state.

For more information about NQA, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

If you associate a track entry with a nonexistent NQA operation or reaction entry, the state of the track entry is NotReady.

Procedure

1. Enter system view.

   system-view

2. Create a track entry, associate it with an NQA reaction entry, and enter its view.

   track *track-entry-number* **nqa entry** *admin-name operation-tag* **reaction** *item-number*

3. Set the delay for notifying the application module of track entry state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the track entry state changes.

Associating Track with BFD

About this task

You can associate a track entry with an echo-mode BFD session, a control-mode BFD session, or a static BFD session. For more information about BFD, see "Configuring BFD."

The associated Track and BFD operate as follows:

- If the BFD detects that the link fails, it informs the Track module of the link failure. The Track module sets the track entry to Negative state.
- If the BFD detects that the link is operating correctly, the Track module sets the track entry to Positive state.

After associating a track entry with an echo-packet-mode BFD session, by default, the device checks the echo response receiving interface. If it is different from the echo packet output interface, the device takes the echo response packet as an invalid packet, and discards it. The BFD session then goes down. When multiple paths are available for the BFD session, as a best practice, enable

checking for the echo response receiving interface. This can avoid BFD session failure when the echo response receiving interface is different from the echo packet output interface.

Restrictions and guidelines

When you associate a track entry with BFD, do not configure the virtual IP address of a VRRP group as the local or remote address of the BFD session.

For correct establishment of a control-mode BFD session, follow these guidelines when you execute the **track bfd ctrl** command:

- Execute the **track bfd ctrl** command on both the local and peer ends.
- Specify the BFD session parameters (outgoing interface, source IP address, and destination IP address of BFD control packets) correctly for the command.

Prerequisites

Before you associate Track with an echo-mode BFD session, configure the source IP address of BFD echo packets. For more information, see "Configuring BFD."

Procedure

1. Enter system view.

   system-view

2. Create a track entry, associate it with a BFD session, and enter its view. Choose the options to configure as needed:
   - Create a track entry, and associate it with an echo-mode BFD session.

     track *track-entry-number* **bfd echo interface** *interface-type interface-number* **remote** { **ip** *remote-ipv4-address* | **ipv6** *remote-ipv6-address* } **local** { **ip** *local-ipv4-address* | **ipv6** *local-ipv6-address* } [**ignore-rx-interface**]
   - Create a track entry, and associate it with a control-mode BFD session.

     track *track-entry-number* **bfd ctrl** [**interface** *interface-type interface-number* | **vpn-instance** *vpn-instance-name*] **remote** { **ip** *remote-ipv4-address* | **ipv6** *remote-ipv6-address* } **local** { **ip** *local-ipv4-address* | **ipv6** *local-ipv6-address* }
   - Create a track entry, and associate it with a static BFD session.

     track *track-entry-number* **bfd static** *bfd-session-name*

3. Set the delay for notifying the application module of track entry state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the track entry state changes.

Associating Track with interface management

About this task

The interface management module monitors the link status or network-layer protocol status of interfaces. The associated Track and interface management operate as follows:

- When the link or network-layer protocol status of the interface changes to up, the interface management module informs the Track module of the change. The Track module sets the track entry to Positive state.
- When the link or network-layer protocol status of the interface changes to down, the interface management module informs the Track module of the change. The Track module sets the track entry to Negative state.

Procedure

1. Enter system view.

   system-view

2. Create a track entry, associate it with interface management, and enter its view.
   - Create a track entry to monitor the link status of an interface.

     track *track-entry-number* **interface** *interface-type interface-number*
   - Create a track entry to monitor the physical status of an interface.

     track *track-entry-number* **interface** *interface-type interface-number* **physical**
   - Create a track entry to monitor the network layer protocol status of an interface.

     track *track-entry-number* **interface** *interface-type interface-number* **protocol** { **ipv4** | **ipv6** }

3. Set the delay for notifying the application module of track entry state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the track entry state changes.

Associating Track with route management

About this task

The route management module monitors route entry changes in the routing table. The associated Track and route management operate as follows:

- Monitor the specified route entry:
  - If the route entry exists in the routing table, the route management module informs the Track module. The Track module then sets the track entry to Positive state.
  - When the route entry is removed from the routing table, the route management module informs the Track module of the change. The Track module then sets the track entry to Negative state.
- Monitor the number of ECMP routes associated with the specified route entry:
  - When the number of ECMP routes is within the specified range, the route management module informs the Track module. The Track module then sets the track entry to Positive state.
  - When the number of ECMP routes is out of the specified range, the route management module informs the Track module. The Track module then sets the track entry to Negative state.

Associating a track entry with an IPv4 route entry

1. Enter system view.

   system-view

2. Create a track entry, associate it with a route entry, and enter its view.

   track *track-entry-number* **ip route** [**vpn-instance** *vpn-instance-name*] *ipv4-address* { *mask-length* | *mask* } { **ecmp-number** { **greater-equal** *min-ecmp-number* | **less-equal** *max-ecmp-number* } * | **reachability** }

3. Set the delay for notifying the application module of track entry state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the track entry state changes.

Associating Track with LLDP

About this task

The LLDP module detects neighbor availability and informs the detection result to the Track module. The associated Track and LLDP operate as follows:

- When the neighbor is available, the Track module sets the track entry to Positive state.
- When the neighbor is unavailable, the Track module sets the track entry to Negative state.

For more information about LLDP, see *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Create a track entry, associate it with an LLDP interface, and enter its view.

   track *track-entry-number* **lldp neighbor interface** *interface-type interface-number*

3. Set the delay for notifying the application module of track entry state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the track entry state changes.

Associating Track with M-LAG MAD state

About this task

M-LAG provides the MAD mechanism to set associated device interfaces to M-LAG MAD DOWN state when the peer link between M-LAG member devices goes down. This can avoid traffic forwarding failure on the interfaces and avoid impacts on services. When the peer link comes up and the delay timer expires, M-LAG brings up the interfaces placed in M-LAG MAD DOWN state.

The associated Track and M-LAG MAD state operate as follows:

- If specific interfaces on the device are set to M-LAG MAD DOWN state, the Track module sets the track entry to Negative state.
- If the interfaces in M-LAG MAD DOWN state come up, the Track module sets the track entry to Positive state.
- If no M-LAG MAD state change events are triggered or received, the Track module sets the track entry to NotReady state.

For more information about M-LAG MAD, see M-LAG configuration in *Layer 2—LAN Switching Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Create a track entry, associate it with M-LAG MAD state, and enter its view.

   track *track-entry-number* **mlag-mad-status**

3. Set the delay for notifying the application module of track entry state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the track entry state changes.

Associating Track with PTP

About this task

If you associate a track entry with PTP, and then associate the track entry with an EAA policy, you can implement automatic monitoring of the PTP clock source and PTP clock state. This enables the device to use PTP to synchronize the system clock when PTP is operating correctly, and use NTP to synchronize the system clock when PTP is not operating correctly.

After you configure this feature, the device creates a track entry associated with PTP. The state of the track entry depends on the state of the associated PTP object. The device supports associating Track with the following types of PTP objects:

- **Clock source**—Determines the state of the track entry based on the clock source state in the specified instance.
  - If the PTP clock source already exists in the specified PTP instance (that is, the device has a subordinate port, and the device is not a GM in the network), the track entry state is Positive.
  - If the PTP clock source is lost in the specified PTP instance (that is, the device is a GM in the network), the track entry state is Negative.
- **PTP time state**—Determines the state of the track entry based on the PTP time state.
  - If the PTP time state is Lock, the track entry state is Positive.
  - If the PTP time state is Unlock, the track entry state is Negative.

You can use the Boolean AND tracked list to monitor both the PTP clock source state and PTP time state. For more information about PTP, see the PTP in Network Management and Monitoring Configuration Guide.

Procedure

1. Enter system view.

   system-view
2. Create a track entry associated with PTP and enter track view.

   track *track-entry-number* **ptp** { **source-state** [**instance** *ptp-instance-id*] | **time-state** }
3. Set the delay for notifying the application module of track entry state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the track entry state changes.

Associating Track with a tracked list

Associating Track with a Boolean list

About this task

A Boolean list is a list of tracked objects based on a Boolean logic. It can be further divided into the following types:

- **Boolean AND list**—A Boolean AND list is set to the Positive state only when all objects are in Positive state. If one or more objects are in Negative state, the tracked list is set to the Negative state.
- **Boolean OR list**—A Boolean OR list is set to the Positive state if any object is in Positive state. If all objects are in Negative state, the tracked list is set to the Negative state.

Procedure

1. Enter system view.

   system-view

2. Create a Boolean tracked list and enter its view.

   track *track-entry-number* **list boolean** { **and** | **or** }

3. (Optional.) Set the delay for notifying the application module of tracked list state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the tracked list state changes.

4. Add the track entry as an object to the tracked list.

   object *track-entry-number* [**not**]

   By default, a tracked list does not contain any objects.

   Repeat this step to add all interested objects to the tracked list.

Associating Track with a percentage threshold list

About this task

The state of a percentage threshold list is determined by comparing the percentage of positive objects in the list with the percentage thresholds configured for the list.

- If the percentage of positive objects is equal to or above the positive state threshold, the list is set to the Positive state.
- If the percentage of positive objects is equal to or below the negative state threshold, the list is set to the Negative state.
- The state of a percentage threshold list remains unchanged if the percentage of positive objects is below the positive state threshold and above the negative state threshold.

Procedure

1. Enter system view.

   system-view

2. Create a percentage threshold list and enter its view.

   track *track-entry-number* **list threshold percentage**

3. (Optional.) Set the delay for notifying the application module of tracked list state changes.

   delay { **negative** *negative-time* | **positive** *positive-time* } *

   By default, the Track module notifies the application module immediately when the tracked list state changes.

4. Add the track entry as an object to the tracked list.

   object *track-entry-number*

   By default, a tracked list does not contain any objects.

   Repeat this step to add all interested objects to the tracked list.

5. Configure the threshold values used to determine the state of the percentage threshold list.

   threshold percentage { **negative** *negative-threshold* | **positive** *positive-threshold* } *

   By default, the negative state threshold is 0% and the positive state threshold is 1%.

Associating Track with a weight threshold list

About this task

The state of a weight threshold list is determined by comparing the weight of positive objects in the list with the weight thresholds configured for the list.

- If the total weight of positive objects is equal to or above the positive state threshold, the list is set to the Positive state.
- If the total weight of positive objects is equal to or below the negative state threshold, the list is set to the Negative state.
- The state of a weight threshold list remains unchanged if the total weight of positive objects is below the positive state threshold and above the negative state threshold.

Procedure

1. Enter system view.
   system-view
2. Create a weight threshold list and enter its view.
   track *track-entry-number* **list threshold weight**
3. (Optional.) Set the delay for notifying the application module of tracked list state changes.
   delay { **negative** *negative-time* | **positive** *positive-time* } *
   By default, the Track module notifies the application module immediately when the tracked list state changes.
4. Add the track entry as an object to the tracked list.
   object *track-entry-number* [**weight** *weight*]
   By default, a tracked list does not contain any objects.
   Repeat this step to add all interested objects to the tracked list.
5. Configure the threshold values used to determine the state of the weight threshold list.
   threshold weight { **negative** *negative-threshold* | **positive** *positive-threshold* } *
   By default, the negative state threshold is 0 and the positive state threshold is 1.

Associating the Track module with an application module

Before you associate the Track module with an application module, make sure the associated track entry has been created.

Prerequisites for associating the Track module with an application module

Create a track entry first before you associate it with an application module.

An application module might obtain incorrect track entry status information if the associated track entry does not exist.

For more information about the Track and application module collaboration commands, see the command reference of the associated application module.

Associating Track with VRRP

About this task

When VRRP is operating in standard mode or load balancing mode, associate the Track module with the VRRP group to implement the following actions:

- Change the priority of a router according to the status of the uplink. If a fault occurs on the uplink of the router, the VRRP group is not aware of the uplink failure. If the router is the master, hosts in the LAN cannot access the external network. To resolve this problem, configure a detection module-Track-VRRP collaboration. The detection module monitors the status of the uplink of the router and notifies the Track module of the detection result.

  When the uplink fails, the detection module notifies the Track module to change the status of the monitored track entry to Negative. The priority of the master decreases by a user-specified value. A router with a higher priority in the VRRP group becomes the master.
- Monitor the master on a backup. If a fault occurs on the master, the backup operating in switchover mode will switch to the master immediately to maintain normal communication.

When VRRP is operating in load balancing mode, associate the Track module with the VRRP VF to implement the following functions:

- Change the priority of the AVF according to its uplink state. When the uplink of the AVF fails, the track entry changes to Negative state. The weight of the AVF decreases by a user-specified value. The VF with a higher priority becomes the new AVF to forward packets.
- Monitor the AVF status from the LVF. When the AVF fails, the LVF that is operating in switchover mode becomes the new AVF to ensure continuous forwarding.

For more information about configuring VRRP, see "Configuring VRRP."

Restrictions and guidelines for Track association with VRRP

- VRRP tracking does not take effect on an IP address owner. The configuration takes effect when the router does not act as the IP address owner.

  An IP address owner is the router with its interface IP address used as the virtual IP address of the VRRP group.
- When the status of the track entry changes from Negative to Positive or NotReady, the associated router or VF restores its priority automatically.

Associating Track with a VRRP group

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Associate a track entry with a VRRP group.

   vrrp [**ipv6**] **vrid** *virtual-router-id* **track** *track-entry-number* { **forwarder-switchover member-ip** *ip-address* | **priority reduced** [*priority-reduced*] **switchover** | **weight reduced** [*weight-reduced*] }

   By default, no track entry is associated with a VRRP group.

   This command is supported when VRRP is operating in both standard mode and load balancing mode.

Associating Track with a VRRP VF

1. Enter system view.

   system-view
2. Enter interface view.

interface *interface-type interface-number*

3. Associate Track with a VRRP VF.

   vrrp [**ipv6**] **vrid** *virtual-router-id* **track** *track-entry-number* { **forwarder-switchover member-ip** *ip-address* | **priority reduced** [*priority-reduced*] **switchover** | **weight reduced** [*weight-reduced*] }

   By default, no track entry is associated with a VRRP VF.

   This command is configurable when VRRP is operating in standard mode or load balancing mode. However, the configuration takes effect only when VRRP is operating in load balancing mode.

Associating Track with static routing

About this task

A static route is a manually configured route to route packets. For more information about static route configuration, see *Layer 3—IP Routing Configuration Guide*.

Static routes cannot adapt to network topology changes. Link failures or network topological changes can make the routes unreachable and cause communication interruption.

To resolve this problem, configure another route to back up the static route. When the static route is reachable, packets are forwarded through the static route. When the static route is unreachable, packets are forwarded through the backup route.

To check the accessibility of a static route in real time, associate the Track module with the static route.

If you specify the next hop but not the output interface when configuring a static route, you can configure the static routing-Track-detection module collaboration. This collaboration enables you to verify the accessibility of the static route based on the track entry state.

- If the track entry is in Positive state, the following conditions exist:
  - The next hop of the static route is reachable.
  - The configured static route is valid.
- If the track entry is in Negative state, the following conditions exist:
  - The next hop of the static route is not reachable.
  - The configured static route is invalid.
- If the track entry is in NotReady state, the following conditions exist:
  - The accessibility of the next hop of the static route is unknown.
  - The static route is valid.

Restrictions and guidelines

In static routing-Track-NQA collaboration, you must configure the same VPN instance name for the NQA operation and the next hop of the static route. Otherwise, the accessibility detection cannot operate correctly.

If a static route needs route recursion, the associated track entry must monitor the next hop of the recursive route. The next hop of the static route cannot be monitored. Otherwise, a valid route might be considered invalid.

Associating Track with an IPv4 static route

1. Enter system view.

   system-view

2. Associate an IPv4 static route with a track entry to check the accessibility of the next hop.

   Public network:

ip route-static { *dest-address* { *mask-length* | *mask* } | **group** *group-name* } { *interface-type interface-number* [*next-hop-address*] **track** *track-entry-number* | *next-hop-address* [**recursive-lookup host-route**] **track** *track-entry-number* | **vpn-instance** *d-vpn-instance-name next-hop-address* [**recursive-lookup host-route**] **track** *track-entry-number* } [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

VPN:

ip route-static vpn-instance *s-vpn-instance-name* { *dest-address* { *mask-length* | *mask* } | **group** *group-name* } { *interface-type interface-number* [*next-hop-address*] **track** *track-entry-number* | *next-hop-address* [**recursive-lookup host-route**] [**public**] **track** *track-entry-number* | **vpn-instance** *d-vpn-instance-name next-hop-address* [**recursive-lookup host-route**] **track** *track-entry-number* } [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

By default, an IPv4 static route is not associated with any track entries.

Associating Track with an IPv6 static route

1. Enter system view.

   system-view

2. Associate an IPv6 static route with a track entry to check the accessibility of the next hop.

   Public network:

   ipv6 route-static *ipv6-address prefix-length* { *interface-type interface-number* [*next-hop-address*] **track** *track-entry-number* | [**vpn-instance** *d-vpn-instance-name*] *next-hop-address* [**recursive-lookup host-route**] **track** *track-entry-number* } [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   VPN:

   ipv6 route-static vpn-instance *s-vpn-instance-name ipv6-address prefix-length* { *interface-type interface-number* [*next-hop-address*] **track** *track-entry-number* | *next-hop-address* [**public**] **track** *track-entry-number* | **vpn-instance** *d-vpn-instance-name next-hop-address* [**recursive-lookup host-route**] **track** *track-entry-number* } [**preference** *preference*] [**tag** *tag-value*] [**description** *text*]

   By default, an IPv6 static route is not associated with any track entries.

Associating Track with PBR

About this task

PBR uses user-defined policies to route packets. You can specify parameters in a PBR policy to guide the forwarding of the packets that match specific criteria. For more information about PBR, see *Layer 3—IP Routing Configuration Guide*.

PBR cannot detect the availability of any action taken on packets. When an action is not available, packets processed by the action might be discarded. For example, if the output interface specified for PBR fails, PBR cannot detect the failure, and continues to forward matching packets out of the interface.

To enable PBR to detect topology changes and improve the flexibility of the PBR application, configure Track-PBR-detection module collaboration.

After you associate a track entry with an apply clause, the detection module associated with the track entry sends Track the detection result of the availability of the tracked object.

- The Positive state of the track entry indicates that the object is available, and the apply clause is valid.
- The Negative state of the track entry indicates that the object is not available, and the apply clause is invalid.
- The NotReady state of the track entry indicates that the apply clause is valid.

Prerequisites for Track association with PBR

Before you associate Track with PBR, create a policy node, and set the match criteria.

Associating Track with PBR

1. Enter system view.

   system-view

2. Create a policy node and enter its view.

   policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*

3. Set match criteria. Choose the options to configure as needed:
   - Set an ACL match criterion.

     if-match acl { *acl-number* | **name** *acl-name* }

     By default, no ACL match criterion is set.

     The ACL match criterion cannot match Layer 2 information.

     When using the ACL to match packets, PBR ignores the time range settings in the ACL. If the specified ACL does not exist, no packets will be matched.

     When using the ACL to match packets, interface PBR ignores the action (**permit** or **deny**) setting in the ACL.

     When using the ACL to match packets, local PBR uses only the **permit** action in the ACL.

4. Set actions and associate the policy node with a track entry. Choose the options to configure as needed:
   - Set the next hop.

     apply next-hop [**vpn-instance** *vpn-instance-name*] { *ip-address* [**direct**] [**track** *track-entry-number*]}&<1-4>

     By default, no next hop is set.

Associating Track with IPv6 PBR

1. Enter system view.

   system-view

2. Create an IPv6 policy node and enter its view.

   ipv6 policy-based-route *policy-name* [**deny** | **permit**] **node** *node-number*

3. Set match criteria. Choose the options to configure as needed:
   - Set an ACL match criterion.

     if-match acl { *ipv6-acl-number* | **name** *ipv6-acl-name* }

     By default, no ACL match criterion is set.

     The ACL match criterion cannot match Layer 2 information.

     When using the ACL to match packets, IPv6 PBR ignores the action (**permit** or **deny**) and time range settings in the ACL.

4. Set actions and associate the policy node with a track entry. Choose the options to configure as needed:

- Set the next hop.

  apply next-hop [**vpn-instance** *vpn-instance-name*] { *ipv6-address* [**direct**] [**track** *track-entry-number*] } &<1-4>

  By default, no next hop is set.

Associating Track with traffic redirection

About this task

Traffic redirection redirects the packets matching specific criteria to a certain location, such as to an interface, a CPU, or a next hop. When the location (for example, next hop) is not reachable, packets redirected to it are lost. To check the accessibility of the next hop in real time, associate Track with traffic redirection. For more information about traffic redirection, see QoS configuration in *ACL and QoS Configuration Guide*.

- If the track entry is in Positive state, the following conditions exist:
  - The next hop is reachable.
  - Traffic redirection redirects packets to the next hop.
- If the track entry is in Negative state, the following conditions exist:
  - The next hop is not reachable.
  - Traffic redirection does not redirect packets.

  The packets are forwarded based on the routing table lookup result.
- If the track entry is in NotReady state, the following conditions exist:
  - The accessibility of the next hop is unknown.
  - Traffic redirection redirects packets to the next hop of the unknown state.

Procedure

1. Enter system view.

   system-view
2. Create a behavior and enter behavior view.

   traffic behavior *behavior-name*
3. Associate Track with traffic redirection.

   redirect next-hop { *ipv4-add* [**track** *track-entry-number*] [*ipv4-add* [**track** *track-entry-number*]] | *ipv6-add* [*interface-type interface-number*] [**track** *track-entry-number*] [*ipv6-add* [*interface-type interface-number*] [**track** *track-entry-number*]] }

   By default, no track entry is associated with traffic redirection.

Associating Track with EAA

About this task

You can configure EAA track event monitor policies to monitor the positive-to-negative or negative-to-positive state changes of track entries.

- If you specify only one track entry for a policy, EAA triggers the policy when it detects the specified state change on the track entry.
- If you specify multiple track entries for a policy, EAA triggers the policy when it detects the specified state change on the last monitored track entry. For example, if you configure a policy to monitor the positive-to-negative state change of multiple track entries, EAA triggers the policy when the last positive track entry monitored by the policy is changed to the Negative state.

You can set a suppression time for a track event monitor policy. The timer starts when the policy is triggered. The system does not process messages that report the monitored track event until the timer times out.

For more information about EAA, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Create a CLI-defined monitor policy and enter its view, or enter the view of an existing CLI-defined monitor policy.
   rtm cli-policy *policy-name*
3. Configure a track event.
   event track *track-entry-number-list* **state** { **negative** | **positive** } [**suppress-time** *suppress-time*]
   By default, a monitor policy does not monitor any track event.

Associating Track with PoE

About this task

The PoE module can collaborate with the Track module to monitor the link status between the device and the PD connected to a PI. For example, if the PD supports the NQA ICMP echo test, you can specify a track entry associated with NQA for the PI to test the reachability of the PD. The NQA ICMP echo test must be configured on a Layer 3 interface. If the PI is a Layer 2 interface, create a VLAN interace and assign the PI to the VLAN interface, or configure the Layer 2 PI to operate in Layer 3 mode.

The Track module notifies the PoE module of the following monitoring results:

- **Positive**—The monitored object is reachable.
- **Negative**—The monitored object is unreachable.
- **NotReady**—The monitoring result is not ready because of reasons such as nonexistence of the NQA group associated with the track entry.

When the Track module detects failure of the link, it changes the track entry state from Positive to Negative, which triggers the PoE module to take the following actions:

- **alarm-only**: Outputs an SNMP notification and log.
- **alarm-reboot-pd**: Outputs an SNMP notification and log and reboots the PD connected to the PI.

For information about SNMP notifications, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

For information about logs, see information center configuration in *Network Management and Monitoring Configuration Guide*.

For information about PoE, see PoE configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Associate a PI with a track entry.
   poe track *track-entry-number* **action** { **alarm** | **alarm-reboot-pd** }
   By default, a PI does not associate with any track entry.

Display and maintenance commands for Track

Execute **display** commands in any view.

Task	Command
Display information about track entries.	**display track** { *track-entry-number* | **all** [**negative** | **positive**] } [**brief**]

The output shows that when the control-mode BFD session detects that Router A fails, the Track module notifies VRRP to change the status of Router B to master. The backup can quickly preempt as the master without waiting for a period three times the advertisement interval plus the Skew_Time.

Track configuration examples

Example: Configuring VRRP-Track-NQA collaboration

Network configuration

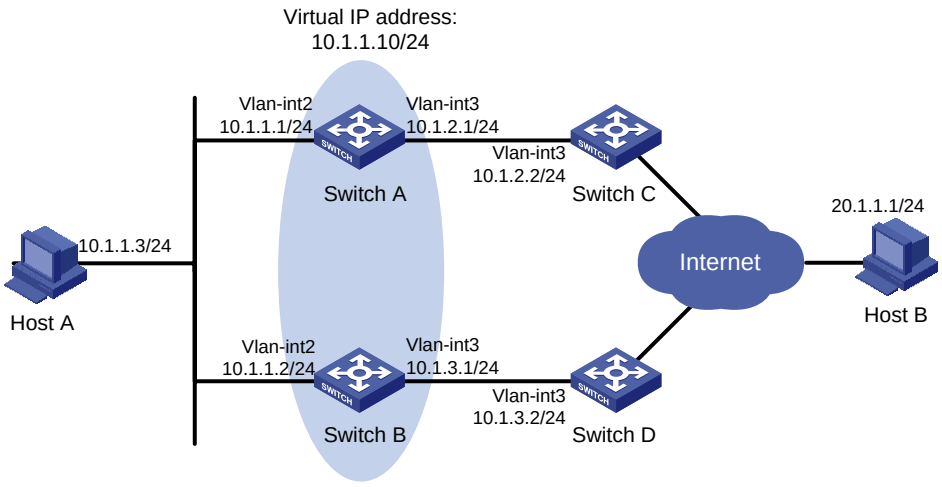
As shown in Figure 2:

- Host A requires access to Host B. The default gateway of Host A is 10.1.1.10/24.
- Switch A and Switch B belong to VRRP group 1. The virtual IP address of VRRP group 1 is 10.1.1.10.

Configure VRRP-Track-NQA collaboration to monitor the uplink on the master and meet the following requirements:

- When Switch A operates correctly, Switch A forwards packets from Host A to Host B.
- When NQA detects a fault on the uplink of Switch A, Switch B forwards packets from Host A to Host B.

Figure 2 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 2. (Details not shown.)
2. Configure an NQA operation on Switch A:

Create an NQA operation with administrator name **admin** and operation tag **test**.

```
<SwitchA> system-view
[SwitchA] nqa entry admin test
```

Specify the **ICMP echo** operation type.

```
[SwitchA-nqa-admin-test] type icmp-echo
```

Specify 10.1.2.2 as the destination address of ICMP echo requests.

```
[SwitchA-nqa-admin-test-icmp-echo] destination ip 10.1.2.2
```

Configure the ICMP echo operation to repeat every 100 milliseconds.

```
[SwitchA-nqa-admin-test-icmp-echo] frequency 100
```

Configure reaction entry 1, specifying that five consecutive probe failures trigger the Track module.

```
[SwitchA-nqa-admin-test-icmp-echo] reaction 1 checked-element probe-fail
threshold-type consecutive 5 action-type trigger-only
[SwitchA-nqa-admin-test-icmp-echo] quit
```

Start the NQA operation.

```
[SwitchA] nqa schedule admin test start-time now lifetime forever
```

3. On Switch A, configure track entry 1, and associate it with reaction entry 1 of the NQA operation.

```
[SwitchA] track 1 nqa entry admin test reaction 1
[SwitchA-track-1] quit
```

4. Configure VRRP on Switch A:

Specify VRRPv2 to run on VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] vrrp version 2
```

Create VRRP group 1, and configure virtual IP address 10.1.1.10 for the group.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
```

Set the priority of Switch A to 110 in VRRP group 1.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Set the authentication mode of VRRP group 1 to **simple**, and the authentication key to **hello**.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Configure the master to send VRRP packets every 500 centiseconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 timer advertise 500
```

Configure Switch A to operate in preemptive mode and set the preemption delay to 5000 centiseconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5000
```

Associate VRRP group 1 with track entry 1 and decrease the router priority by 30 when the state of track entry 1 changes to Negative.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 track 1 priority reduced 30
```

5. Configure VRRP on Switch B:

Specify VRRPv2 to run on VLAN-interface 2.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] vrrp version 2
```

Create VRRP group 1, and configure virtual IP address 10.1.1.10 for the group.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
```

Set the authentication mode of VRRP group 1 to **simple**, and the authentication key to **hello**.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Configure the master to send VRRP packets every 500 centiseconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 timer advertise 500
```

Configure Switch B to operate in preemptive mode and set the preemption delay to 5000 centiseconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5000
```

Verifying the configuration

Ping Host B from Host A to verify that Host B is reachable. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 500
     Admin Status     : Up              State        : Master
     Config Pri       : 110             Running Pri  : 110
     Preempt Mode     : Yes             Delay Time   : 5000
     Auth Type        : Simple          Key          : ******
     Version          : 2
     Virtual IP       : 10.1.1.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 10.1.1.1
   VRRP Track Information:
     Track Object   : 1              State : Positive          Pri Reduced : 30
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 500
     Admin Status     : Up              State        : Backup
     Config Pri       : 100             Running Pri  : 100
     Preempt Mode     : Yes             Delay Time   : 5000
     Become Master    : 2200ms left
     Auth Type        : Simple          Key          : ******
     Version          : 2
     Virtual IP       : 10.1.1.10
     Master IP        : 10.1.1.1
```

The output shows that in VRRP group 1, Switch A is the master, and Switch B is a backup. Switch A forwards packets from Host A to Host B.

Disconnect the link between Switch A and Switch C, and verify that Host A can still ping Host B. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
```

```
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID              : 1                Adver Timer   : 500
     Admin Status      : Up               State         : Backup
     Config Pri        : 110              Running Pri   : 80
     Preempt Mode      : Yes              Delay Time    : 5000
     Become Master     : 2200ms left
     Auth Type         : Simple           Key           : ******
     Version           : 2
     Virtual IP        : 10.1.1.10
     Master IP         : 10.1.1.2
   VRRP Track Information:
     Track Object    : 1              State : Negative          Pri Reduced : 30
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID              : 1                Adver Timer   : 500
     Admin Status      : Up               State         : Master
     Config Pri        : 100              Running Pri   : 100
     Preempt Mode      : Yes              Delay Time    : 5000
     Auth Type         : Simple           Key           : ******
     Version           : 2
     Virtual IP        : 10.1.1.10
     Virtual MAC       : 0000-5e00-0101
     Master IP         : 10.1.1.2
```

The output shows that Switch A becomes the backup, and Switch B becomes the master. Switch B forwards packets from Host A to Host B.

Example: Configuring an echo-mode BFD session for a VRRP backup to monitor the master

Network configuration

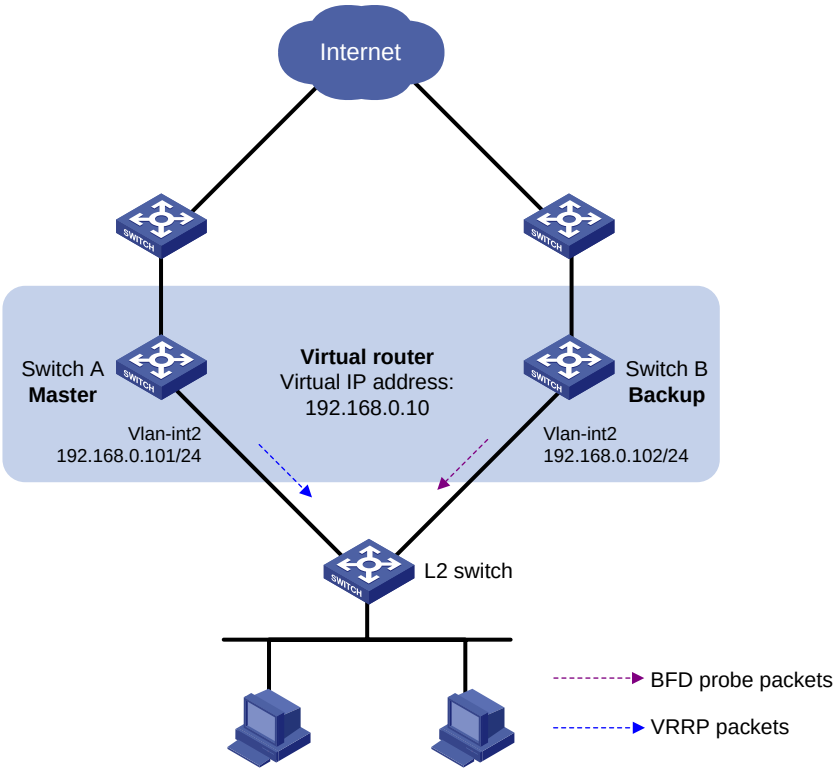
As shown in Figure 3:

- Switch A and Switch B belong to VRRP group 1. The virtual IP address of VRRP group 1 is 192.168.0.10.
- The default gateway of the hosts in the LAN is 192.168.0.10.

Configure VRRP-Track-BFD (echo mode) collaboration to monitor the master on the backup and meet the following requirements:

- When Switch A operates correctly, the hosts in the LAN access the Internet through Switch A.
- When Switch A fails, the backup (Switch B) can detect the state change of the master through the echo-mode BFD session and become the new master. The hosts in the LAN access the Internet through Switch B.

Figure 3 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 3. (Details not shown.)
2. Configure Switch A:

   # Create VRRP group 1, and configure virtual IP address 192.168.0.10 for the group.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
   ```

   # Set the priority of Switch A to 110 in VRRP group 1.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
   [SwitchA-Vlan-interface2] return
   ```

3. Configure Switch B:

   # Specify 10.10.10.10 as the source address of BFD echo packets.

   ```
   <SwitchB> system-view
   [SwitchB] bfd echo-source-ip 10.10.10.10
   ```

   # Create track entry 1, and associate it with the echo-mode BFD session to verify the reachability of Switch A.

   ```
   [SwitchB] track 1 bfd echo interface vlan-interface 2 remote ip 192.168.0.101 local
   ip 192.168.0.102
   [SwitchB-track-1] quit
   ```

   # Create VRRP group 1, and configure virtual IP address 192.168.0.10 for the group.

   ```
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
   ```

   # Configure VRRP group 1 to monitor the status of track entry 1.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 track 1 switchover
[SwitchB-Vlan-interface2] return
```

Verifying the configuration

Display detailed information about VRRP group 1 on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
     Admin Status     : Up              State        : Master
     Config Pri       : 110             Running Pri  : 110
     Preempt Mode     : Yes             Delay Time   : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 192.168.0.101
```

Display detailed information about VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
     Admin Status     : Up              State        : Backup
     Config Pri       : 100             Running Pri  : 100
     Preempt Mode     : Yes             Delay Time   : 0
     Become Master    : 2200ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Master IP        : 192.168.0.101
   VRRP Track Information:
     Track Object   : 1              State : Positive          Switchover
```

Display information about track entry 1 on Switch B.

```
<SwitchB> display track 1
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD echo
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Echo
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 192.168.0.101
```

```
    Local IP: 192.168.0.102
```

The output shows that when the status of the track entry becomes Positive, Switch A is the master and Switch B is the backup.

Enable VRRP state debugging and BFD event notification debugging on Switch B.

```
<SwitchB> terminal debugging
<SwitchB> terminal monitor
<SwitchB> debugging vrrp fsm
<SwitchB> debugging bfd ntfy
```

When Switch A fails, the following output is displayed on Switch B.

```
*Dec 17 14:44:34:142 2019 SwitchB BFD/7/DEBUG: Notify application:TRACK State:DOWN
*Dec 17 14:44:34:144 2019 SwitchB VRRP4/7/FSM:
 IPv4 Vlan-interface2 | Virtual Router 1 : Backup --> Master   reason: The status of the
tracked object changed
```

Display detailed information about the VRRP group on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                Adver Timer   : 100
     Admin Status     : Up               State         : Master
     Config Pri       : 100              Running Pri   : 100
     Preempt Mode     : Yes              Delay Time    : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 192.168.0.102
   VRRP Track Information:
     Track Object   : 1              State : Negative          Switchover
```

The output shows that when the echo-mode BFD session detects that Switch A fails, the Track module notifies VRRP to change the status of Switch B to master. The backup can quickly preempt as the master without waiting for a period three times the advertisement interval plus the Skew_Time.

Example: Configuring a control-mode BFD session for a VRRP backup to monitor the master

Network configuration

As shown in Figure 4:

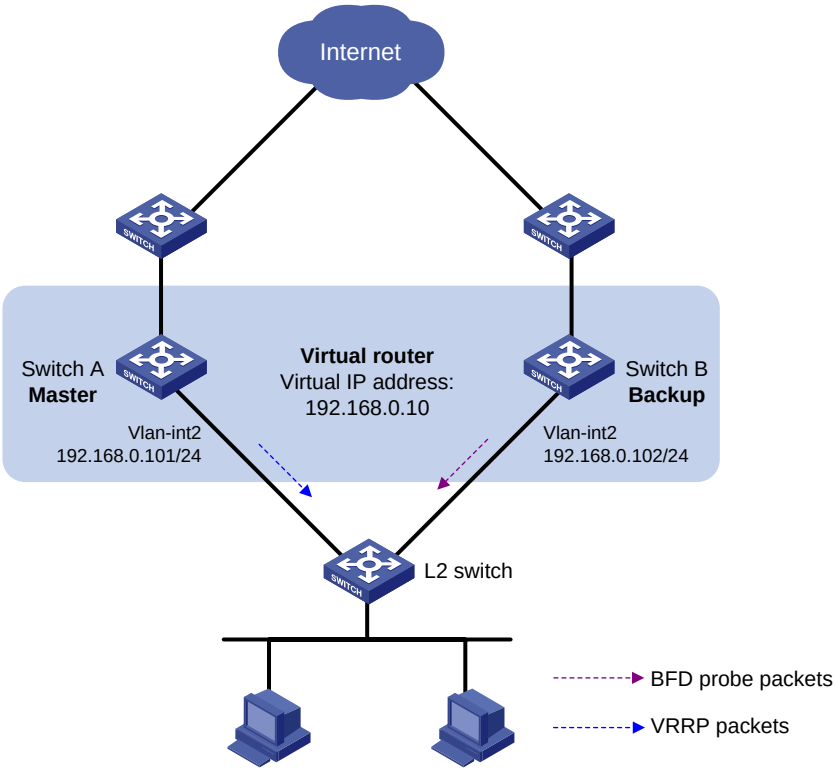
- Switch A and Switch B belong to VRRP group 1. The virtual IP address of VRRP group 1 is 192.168.0.10.
- The default gateway of the hosts in the LAN is 192.168.0.10.

Configure VRRP-Track-BFD (control mode) collaboration to monitor the master on the backup and meet the following requirements:

- When Switch A operates correctly, the hosts in the LAN access the Internet through Switch A.

- When Switch A fails, the backup (Switch B) can detect the state change of the master through the control-mode BFD session and become the new master. The hosts in the LAN access the Internet through Switch B.

Figure 4 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 4. (Details not shown.)
2. Configure Switch A:

   # Create VRRP group 1, and configure virtual IP address 192.168.0.10 for the group.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
   ```

   # Set the priority of Switch A to 110 in VRRP group 1.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
   [SwitchA-Vlan-interface2] return
   ```

   # Create track entry 1, and associate it with the control-mode BFD session to verify the reachability of Switch B.

   ```
   [SwitchA] track 1 bfd ctrl interface vlan-interface 2 remote ip 192.168.0.102 local
   ip 192.168.0.101
   [SwitchA-track-1] quit
   ```
3. Configure Switch B:

   # Create VRRP group 1, and configure virtual IP address 192.168.0.10 for the group.

   ```
   <SwitchB> system-view
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
   ```

Configure VRRP group 1 to monitor track entry 1 so Switch B can take over as the master once the track entry changes to the Negative state.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 track 1 switchover
[SwitchB-Vlan-interface2] return
```

Create track entry 1, and associate it with the control-mode BFD session to verify the reachability of Switch A.

```
[SwitchB] track 1 bfd ctrl interface vlan-interface 2 remote ip 192.168.0.101 local
ip 192.168.0.102
[SwitchB-track-1] quit
```

Verifying the configuration

Display detailed information about VRRP group 1 on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1              Adver Timer  : 100
     Admin Status     : Up             State        : Master
     Config Pri       : 110            Running Pri  : 110
     Preempt Mode     : Yes            Delay Time   : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 192.168.0.101
```

Display detailed information about VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1              Adver Timer  : 100
     Admin Status     : Up             State        : Backup
     Config Pri       : 100            Running Pri  : 100
     Preempt Mode     : Yes            Delay Time   : 0
     Become Master    : 2200ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Master IP        : 192.168.0.101
   VRRP Track Information:
     Track Object   : 1              State : Positive          Switchover
```

Display information about track entry 1 on Switch B.

```
<SwitchB> display track 1
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD ctrl
```

```
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Ctrl
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 192.168.0.101
    Local IP: 192.168.0.102
```

The output shows that when the status of the track entry becomes Positive, Switch A is the master and Switch B is the backup.

Enable VRRP state debugging and BFD event notification debugging on Switch B.

```
<SwitchB> terminal debugging
<SwitchB> terminal monitor
<SwitchB> debugging vrrp fsm
<SwitchB> debugging bfd ntfy
```

When Switch A fails, the following output is displayed on Switch B.

```
*Dec 17 14:44:34:142 2019 SwitchB BFD/7/DEBUG: Notify application:TRACK State:DOWN
*Dec 17 14:44:34:144 2019 SwitchB VRRP4/7/FSM:
 IPv4 Vlan-interface2 | Virtual Router 1 : Backup --> Master   reason: The status of the
tracked object changed
```

Display detailed information about the VRRP group on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1                Adver Timer   : 100
     Admin Status     : Up               State         : Master
     Config Pri       : 100              Running Pri   : 100
     Preempt Mode     : Yes              Delay Time    : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 192.168.0.102
   VRRP Track Information:
     Track Object   : 1              State : Negative          Switchover
```

The output shows that when the control-mode BFD session detects that Switch A fails, the Track module notifies VRRP to change the status of Switch B to master. The backup can quickly preempt as the master without waiting for a period three times the advertisement interval plus the Skew_Time.

Example: Configuring an echo-mode BFD session for the VRRP master to monitor the uplink

Network configuration

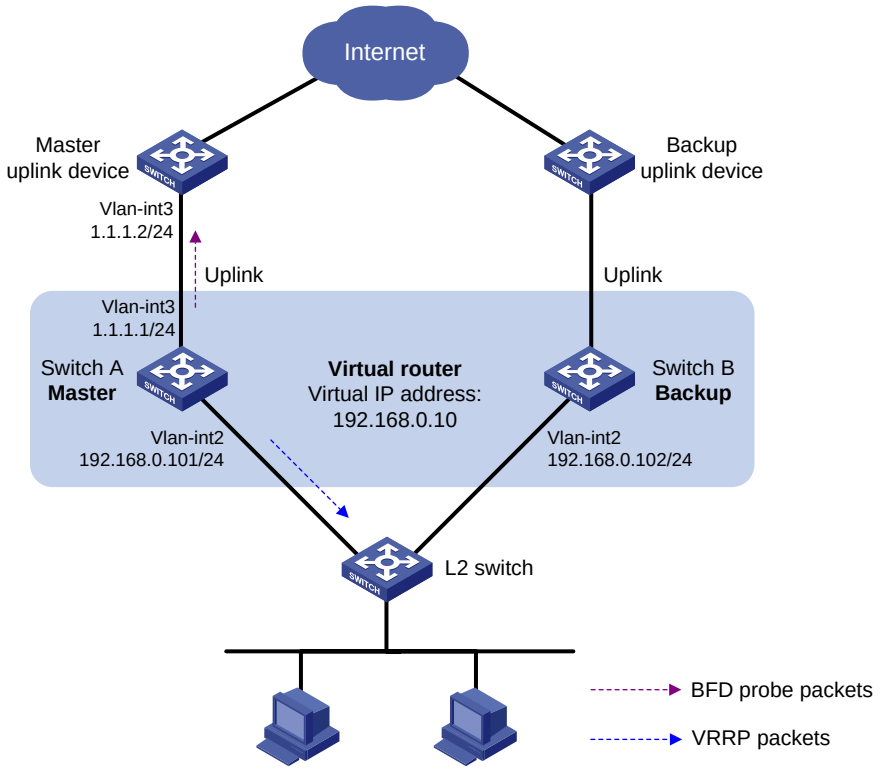
As shown in Figure 5:

- Switch A and Switch B belong to VRRP group 1. The virtual IP address of VRRP group 1 is 192.168.0.10.
- The default gateway of the hosts in the LAN is 192.168.0.10.

Configure VRRP-Track-BFD (echo mode) collaboration to monitor the uplink on the master and meet the following requirements:

- When Switch A operates correctly, the hosts in the LAN access the Internet through Switch A.
- When Switch A detects that the uplink is down through the echo-mode BFD session, Switch B can preempt as the master. The hosts in the LAN can access the Internet through Switch B.

Figure 5 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 5. (Details not shown.)
2. Configure Switch A:

   # Specify 10.10.10.10 as the source address of BFD echo packets.

   ```
   <SwitchA> system-view
   [SwitchA] bfd echo-source-ip 10.10.10.10
   ```

   # Create track entry 1 for the echo-mode BFD session to verify the reachability of the uplink device (1.1.1.2).

   ```
   [SwitchA] track 1 bfd echo interface vlan-interface 3 remote ip 1.1.1.2 local ip
   1.1.1.1
   [SwitchA-track-1] quit
   ```

   # Create VRRP group 1, and specify 192.168.0.10 as the virtual IP address of the group.

   ```
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
   ```

   # Set the priority of Switch A to 110 in VRRP group 1.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
   ```

Associate VRRP group 1 with track entry 1 and decrease the router priority by 20 when the state of track entry 1 changes to Negative.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 track 1 priority reduced 20
[SwitchA-Vlan-interface2] return
```

3. On Switch B, create VRRP group 1, and specify 192.168.0.10 as the virtual IP address of the group.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
[SwitchB-Vlan-interface2] return
```

Verifying the configuration

Display detailed information about the VRRP group on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
     Admin Status     : Up              State        : Master
     Config Pri       : 110             Running Pri  : 110
     Preempt Mode     : Yes             Delay Time   : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 192.168.0.101
   VRRP Track Information:
     Track Object   : 1              State : Positive          Pri Reduced : 20
```

Display information about track entry 1 on Switch A.

```
<SwitchA> display track 1
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD echo
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Echo
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 1.1.1.2
    Local IP: 1.1.1.1
```

Display detailed information about the VRRP group on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
```

```
    Admin Status     : Up              State        : Backup
    Config Pri       : 100             Running Pri  : 100
    Preempt Mode     : Yes             Delay Time   : 0
    Become Master    : 2200ms left
    Auth Type        : Not supported
    Version          : 3
    Virtual IP       : 192.168.0.10
    Master IP        : 192.168.0.101
```

The output shows that when the status of track entry 1 becomes Positive, Switch A is the master and Switch B is the backup.

Display information about track entry 1 when the uplink of Switch A goes down.

```
<SwitchA> display track 1
Track ID: 1
  State: Negative
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD echo
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Echo
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 1.1.1.2
    Local IP: 1.1.1.1
```

Display detailed information about VRRP group 1 on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
     Admin Status     : Up              State        : Backup
     Config Pri       : 110             Running Pri  : 90
     Preempt Mode     : Yes             Delay Time   : 0
     Become Master    : 2200ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Master IP        : 192.168.0.102
   VRRP Track Information:
     Track Object   : 1              State : Negative          Pri Reduced : 20
```

Display detailed information about VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
     Admin Status     : Up              State        : Master
```

```
Config Pri       : 100              Running Pri  : 100
Preempt Mode     : Yes              Delay Time   : 0
Auth Type        : Not supported
Version          : 3
Virtual IP       : 192.168.0.10
Virtual MAC      : 0000-5e00-0101
Master IP        : 192.168.0.102
```

The output shows that when Switch A detects that the uplink fails through the echo-mode BFD session, it decreases its priority by 20. Switch B then preempts as the master.

Example: Configuring a control-mode BFD session for the VRRP master to monitor the uplink

Network configuration

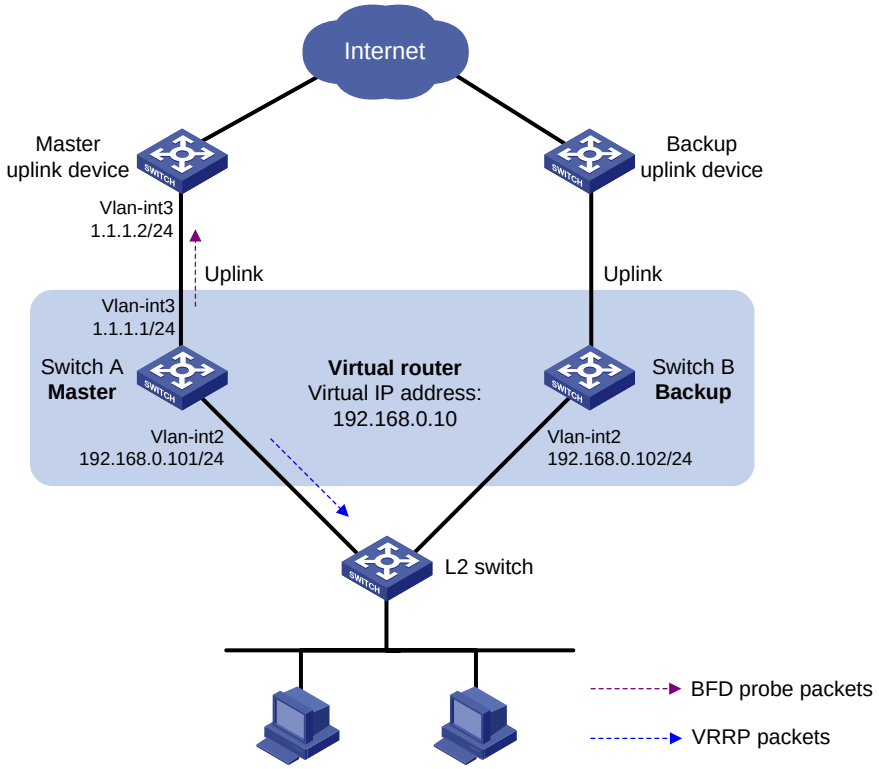
As shown in Figure 6:

- Switch A and Switch B belong to VRRP group 1. The virtual IP address of VRRP group 1 is 192.168.0.10.
- The default gateway of the hosts in the LAN is 192.168.0.10.

Configure VRRP-Track-BFD (control mode) collaboration to monitor the uplink on the master and meet the following requirements:

- When Switch A operates correctly, the hosts in the LAN access the Internet through Switch A.
- When Switch A detects that the uplink is down through the control-mode BFD session, Switch B can preempt as the master. The hosts in the LAN can access the Internet through Switch B.

Figure 6 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 6. (Details not shown.)
2. Configure Switch A:

   # Create VRRP group 1, and specify 192.168.0.10 as the virtual IP address of the group.

   ```
   <SwitchA> system-view
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
   ```

   # Set the priority of Switch A to 110 in VRRP group 1.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
   ```

   # Associate VRRP group 1 with track entry 1 and decrease the router priority by 20 when the state of track entry 1 changes to Negative.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 track 1 priority reduced 20
   [SwitchA-Vlan-interface2] quit
   ```

   # Create track entry 1 for the control-mode BFD session to monitor the reachability of the uplink device (1.1.1.2).

   ```
   [SwitchA] track 1 bfd ctrl interface vlan-interface 3 remote ip 1.1.1.2 local ip
   1.1.1.1
   [SwitchA-track-1] quit
   ```

3. On the uplink device of Switch A, create track entry 1 and associate it with the control-mode BFD session to verify the reachability of the Switch A.

   ```
   <Master> system-view
   [Master] track 1 bfd ctrl interface vlan-interface 3 remote ip 1.1.1.1 local ip 1.1.1.2
   ```

4. On Switch B, create VRRP group 1, and specify 192.168.0.10 as the virtual IP address of the group.

   ```
   <SwitchB> system-view
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
   [SwitchB-Vlan-interface2] return
   ```

Verifying the configuration

Display detailed information about the VRRP group on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1              Adver Timer  : 100
     Admin Status     : Up             State        : Master
     Config Pri       : 110            Running Pri  : 110
     Preempt Mode     : Yes            Delay Time   : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 192.168.0.101
   VRRP Track Information:
     Track Object   : 1             State : Positive         Pri Reduced : 20
```

Display information about track entry 1 on Switch A.

```
<SwitchA> display track 1
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD ctrl
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Ctrl
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 1.1.1.2
    Local IP: 1.1.1.1
```

Display detailed information about the VRRP group on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1              Adver Timer  : 100
     Admin Status     : Up             State        : Backup
     Config Pri       : 100            Running Pri  : 100
     Preempt Mode     : Yes            Delay Time   : 0
     Become Master    : 2200ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 192.168.0.10
     Master IP        : 192.168.0.101
```

The output shows that when the status of track entry 1 becomes Positive, Switch A is the master and Switch B is the backup.

Display information about track entry 1 when the uplink of Switch A goes down.

```
<SwitchA> display track 1
Track ID: 1
  State: Negative
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD ctrl
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Ctrl
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 1.1.1.2
    Local IP: 1.1.1.1
```

Display detailed information about VRRP group 1 on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
```

```
  Interface Vlan-interface2
    VRID              : 1                Adver Timer   : 100
    Admin Status      : Up               State         : Backup
    Config Pri        : 110              Running Pri   : 90
    Preempt Mode      : Yes              Delay Time    : 0
    Become Master     : 2200ms left
    Auth Type         : Not supported
    Version           : 3
    Virtual IP        : 192.168.0.10
    Master IP         : 192.168.0.102
  VRRP Track Information:
    Track Object   : 1              State : Negative          Pri Reduced : 20
```

Display detailed information about VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode       : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID              : 1                Adver Timer   : 100
     Admin Status      : Up               State         : Master
     Config Pri        : 100              Running Pri   : 100
     Preempt Mode      : Yes              Delay Time    : 0
     Auth Type         : Not supported
     Version           : 3
     Virtual IP        : 192.168.0.10
     Virtual MAC       : 0000-5e00-0101
     Master IP         : 192.168.0.102
```

The output shows that when Switch A detects that the uplink fails through the control-mode BFD session, it decreases its priority by 20. Switch B then preempts as the master.

Example: Configuring static routing-Track-NQA collaboration

Network configuration

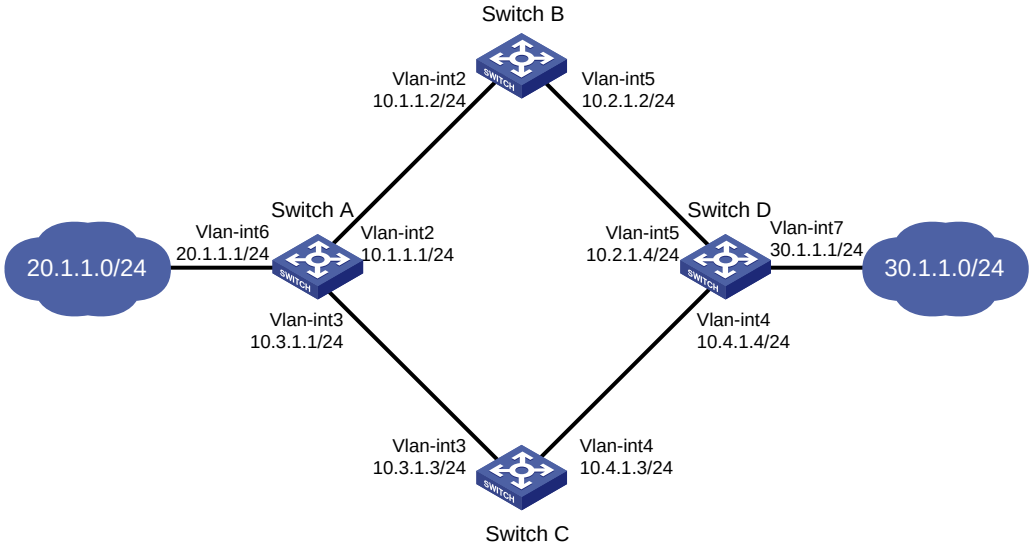
As shown in Figure 7:

- Switch A is the default gateway of the hosts in network 20.1.1.0/24.
- Switch D is the default gateway of the hosts in network 30.1.1.0/24.
- Hosts in the two networks communicate with each other through static routes.

To ensure network availability, configure route backup and static routing-Track-NQA collaboration on Switch A and Switch D as follows:

- On Switch A, assign a higher priority to the static route to 30.1.1.0/24 with next hop Switch B. This route is the master route. The static route to 30.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the backup route takes effect. Switch A forwards packets to 30.1.1.0/24 through Switch C.
- On Switch D, assign a higher priority to the static route to 20.1.1.0/24 with next hop Switch B. This route is the master route. The static route to 20.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the backup route takes effect. Switch D forwards packets to 20.1.1.0/24 through Switch C.

Figure 7 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 7. (Details not shown.)
2. Configure Switch A:

   # Configure a static route to 30.1.1.0/24 with next hop 10.1.1.2 and the default priority (60). Associate this static route with track entry 1.

   ```
   <SwitchA> system-view
   [SwitchA] ip route-static 30.1.1.0 24 10.1.1.2 track 1
   ```

   # Configure a static route to 30.1.1.0/24 with next hop 10.3.1.3 and priority 80.

   ```
   [SwitchA] ip route-static 30.1.1.0 24 10.3.1.3 preference 80
   ```

   # Configure a static route to 10.2.1.4 with next hop 10.1.1.2.

   ```
   [SwitchA] ip route-static 10.2.1.4 24 10.1.1.2
   ```

   # Create an NQA operation with administrator name **admin** and operation tag **test**.

   ```
   [SwitchA] nqa entry admin test
   ```

   # Specify the **ICMP echo** operation type.

   ```
   [SwitchA-nqa-admin-test] type icmp-echo
   ```

   # Specify 10.2.1.4 as the destination address of the operation.

   ```
   [SwitchA-nqa-admin-test-icmp-echo] destination ip 10.2.1.4
   ```

   # Specify 10.1.1.2 as the next hop of the operation.

   ```
   [SwitchA-nqa-admin-test-icmp-echo] next-hop ip 10.1.1.2
   ```

   # Configure the ICMP echo operation to repeat every 100 milliseconds.

   ```
   [SwitchA-nqa-admin-test-icmp-echo] frequency 100
   ```

   # Configure reaction entry 1, specifying that five consecutive probe failures trigger the Track module.

   ```
   [SwitchA-nqa-admin-test-icmp-echo] reaction 1 checked-element probe-fail
   threshold-type consecutive 5 action-type trigger-only
   [SwitchA-nqa-admin-test-icmp-echo] quit
   ```

   # Start the NQA operation.

   ```
   [SwitchA] nqa schedule admin test start-time now lifetime forever
   ```

   # Configure track entry 1, and associate it with reaction entry 1 of the NQA operation.

```
[SwitchA] track 1 nqa entry admin test reaction 1
[SwitchA-track-1] quit
```

3. Configure Switch B:

Configure a static route to 30.1.1.0/24 with next hop 10.2.1.4.

```
<SwitchB> system-view
[SwitchB] ip route-static 30.1.1.0 24 10.2.1.4
```

Configure a static route to 20.1.1.0/24 with next hop 10.1.1.1.

```
[SwitchB] ip route-static 20.1.1.0 24 10.1.1.1
```

4. Configure Switch C:

Configure a static route to 30.1.1.0/24 with next hop 10.4.1.4.

```
<SwitchC> system-view
[SwitchC] ip route-static 30.1.1.0 24 10.4.1.4
```

Configure a static route to 20.1.1.0/24 with next hop 10.3.1.1.

```
[SwitchC] ip route-static 20.1.1.0 24 10.3.1.1
```

5. Configure Switch D:

Configure a static route to 20.1.1.0/24 with next hop 10.2.1.2 and the default priority (60). Associate this static route with track entry 1.

```
<SwitchD> system-view
[SwitchD] ip route-static 20.1.1.0 24 10.2.1.2 track 1
```

Configure a static route to 20.1.1.0/24 with next hop 10.4.1.3 and priority 80.

```
[SwitchD] ip route-static 20.1.1.0 24 10.4.1.3 preference 80
```

Configure a static route to 10.1.1.1 with next hop 10.2.1.2.

```
[SwitchD] ip route-static 10.1.1.1 24 10.2.1.2
```

Create an NQA operation with administrator name **admin** and operation tag **test**.

```
[SwitchD] nqa entry admin test
```

Specify the **ICMP echo** operation type.

```
[SwitchD-nqa-admin-test] type icmp-echo
```

Specify 10.1.1.1 as the destination address of the operation.

```
[SwitchD-nqa-admin-test-icmp-echo] destination ip 10.1.1.1
```

Specify 10.2.1.2 as the next hop of the operation.

```
[SwitchD-nqa-admin-test-icmp-echo] next-hop ip 10.2.1.2
```

Configure the ICMP echo operation to repeat every 100 milliseconds.

```
[SwitchD-nqa-admin-test-icmp-echo] frequency 100
```

Configure reaction entry 1, specifying that five consecutive probe failures trigger the Track module.

```
[SwitchD-nqa-admin-test-icmp-echo] reaction 1 checked-element probe-fail
threshold-type consecutive 5 action-type trigger-only
[SwitchD-nqa-admin-test-icmp-echo] quit
```

Start the NQA operation.

```
[SwitchD] nqa schedule admin test start-time now lifetime forever
```

Configure track entry 1, and associate it with reaction entry 1 of the NQA operation.

```
[SwitchD] track 1 nqa entry admin test reaction 1
[SwitchD-track-1] quit
```

Verifying the configuration

Display information about the track entry on Switch A.

```
[SwitchA] display track all
```

```
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: NQA
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    NQA entry: admin test
    Reaction: 1
    Remote IP/URL: 10.2.1.4
    Local IP:--
    Interface:--
```

The output shows that the status of the track entry is Positive, indicating that the NQA operation has succeeded and the master route is available.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 10       Routes : 10

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.1.1.0/24         Direct 0    0            10.1.1.1        Vlan2
10.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Static 60   0            10.1.1.2        Vlan2
10.3.1.0/24         Direct 0    0            10.3.1.1        Vlan3
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        Vlan6
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.0/24         Static 60   0            10.1.1.2        Vlan2
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets to 30.1.1.0/24 through Switch B.

Remove the IP address of interface VLAN-interface 2 on Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] undo ip address
```

Display information about the track entry on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Negative
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: NQA
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    NQA entry: admin test
    Reaction: 1
    Remote IP/URL: 10.2.1.4
    Local IP:--
    Interface:--
```

The output shows that the status of the track entry is Negative, indicating that the NQA operation has failed and the master route is unavailable.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 10       Routes : 10

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.1.1.0/24         Direct 0    0            10.1.1.1        Vlan2
10.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Static 60   0            10.1.1.2        Vlan2
10.3.1.0/24         Direct 0    0            10.3.1.1        Vlan3
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        Vlan6
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.0/24         Static 80   0            10.3.1.3        Vlan3
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets to 30.1.1.0/24 through Switch C. The backup static route has taken effect.

Verify that hosts in 20.1.1.0/24 can communicate with the hosts in 30.1.1.0/24 when the master route fails.

```
[SwitchA] ping -a 20.1.1.1 30.1.1.1
Ping 30.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 30.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
Reply from 30.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 30.1.1.1: bytes=56 Sequence=4 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 30.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Verify that the hosts in 30.1.1.0/24 can communicate with the hosts in 20.1.1.0/24 when the master route fails.

```
[SwitchB] ping -a 30.1.1.1 20.1.1.1
Ping 20.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 20.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 20.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=4 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 20.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Example: Configuring static routing-Track-BFD (echo mode) collaboration

Network configuration

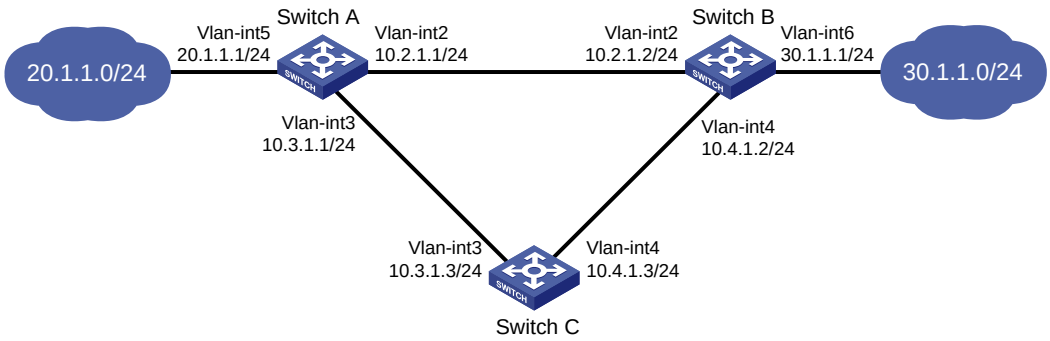
As shown in Figure 8:

- Switch A is the default gateway of the hosts in network 20.1.1.0/24.
- Switch B is the default gateway of the hosts in network 30.1.1.0/24.
- Hosts in the two networks communicate with each other through static routes.

To ensure network availability, configure route backup and static routing-Track-BFD (echo mode) collaboration on Switch A and Switch B as follows:

- On Switch A, assign a higher priority to the static route to 30.1.1.0/24 with next hop Switch B. This route is the master route. The static route to 30.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the echo-mode BFD session can quickly detect the route failure to make the backup route take effect.
- On Switch B, assign a higher priority to the static route to 20.1.1.0/24 with next hop Switch A. This route is the master route. The static route to 20.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the echo-mode BFD session can quickly detect the route failure to make the backup route take effect.

Figure 8 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 8. (Details not shown.)
2. Configure Switch A:

   # Configure a static route to 30.1.1.0/24 with next hop 10.2.1.2 and the default priority (60). Associate this static route with track entry 1.

   ```
   <SwitchA> system-view
   [SwitchA] ip route-static 30.1.1.0 24 10.2.1.2 track 1
   ```

   # Configure a static route to 30.1.1.0/24 with next hop 10.3.1.3 and priority 80.

   ```
   [SwitchA] ip route-static 30.1.1.0 24 10.3.1.3 preference 80
   ```

   # Specify 10.10.10.10 as the source address of BFD echo packets.

   ```
   [SwitchA] bfd echo-source-ip 10.10.10.10
   ```

   # Configure track entry 1, and associate it with the echo-mode BFD session to verify the connectivity between Switch A and Switch B.

   ```
   [SwitchA] track 1 bfd echo interface vlan-interface 2 remote ip 10.2.1.2 local ip
   10.2.1.1
   [SwitchA-track-1] quit
   ```

3. Configure Switch B:

   # Configure a static route to 20.1.1.0/24 with next hop 10.2.1.1 and the default priority (60). Associate this static route with track entry 1.

```
<SwitchB> system-view
[SwitchB] ip route-static 20.1.1.0 24 10.2.1.1 track 1
```

   # Configure a static route to 20.1.1.0/24 with next hop 10.4.1.3 and priority 80.

```
[SwitchB] ip route-static 20.1.1.0 24 10.4.1.3 preference 80
```

   # Specify 1.1.1.1 as the source address of BFD echo packets.

```
[SwitchB] bfd echo-source-ip 1.1.1.1
```

   # Configure track entry 1, and associate it with the echo-mode BFD session to verify the connectivity between Switch B and Switch A.

```
[SwitchB] track 1 bfd echo interface vlan-interface 2 remote ip 10.2.1.1 local ip
10.2.1.2
[SwitchB-track-1] quit
```

4. Configure Switch C:

   # Configure a static route to 30.1.1.0/24 with next hop 10.4.1.2.

```
<SwitchC> system-view
[SwitchC] ip route-static 30.1.1.0 24 10.4.1.2
```

   # Configure a static route to 20.1.1.0/24 with next hop 10.3.1.1.

```
[SwitchC] ip route-static 20.1.1.0 24 10.3.1.1
```

Verifying the configuration

Display information about the track entry on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD echo
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Echo
    Outgoing interface: Vlan-interface2
    Remote IP: 10.2.1.2
    Local IP: 10.2.1.1
```

The output shows that the status of the track entry is Positive, indicating that next hop 10.2.1.2 is reachable.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 9        Routes : 9

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan2
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         Direct 0    0            10.3.1.1        Vlan3
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        Vlan5
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
```

```
30.1.1.0/24         Static 60   0            10.2.1.2        Vlan2
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets to 30.1.1.0/24 through Switch B. The master static route has taken effect.

Remove the IP address of VLAN-interface 2 on Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] undo ip address
```

Display information about the track entry on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Negative
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD echo
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Echo
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 10.2.1.2
    Local IP: 10.2.1.1
```

The output shows that the status of the track entry is Negative, indicating that next hop 10.2.1.2 is unreachable.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 9        Routes : 9

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan2
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         Direct 0    0            10.3.1.1        Vlan3
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        Vlan5
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.0/24         Static 80   0            10.3.1.3        Vlan3
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets to 30.1.1.0/24 through Switch C. The backup static route has taken effect.

Verify that the hosts in 20.1.1.0/24 can communicate with the hosts in 30.1.1.0/24 when the master route fails.

```
[SwitchA] ping -a 20.1.1.1 30.1.1.1
Ping 30.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 30.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
```

```
Reply from 30.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 30.1.1.1: bytes=56 Sequence=4 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 30.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Verify that the hosts in 30.1.1.0/24 can still communicate with the hosts in 20.1.1.0/24 when the master route fails.

```
[SwitchB] ping -a 30.1.1.1 20.1.1.1
Ping 20.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 20.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 20.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=4 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 20.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Example: Configuring static routing-Track-BFD (control mode) collaboration

Network configuration

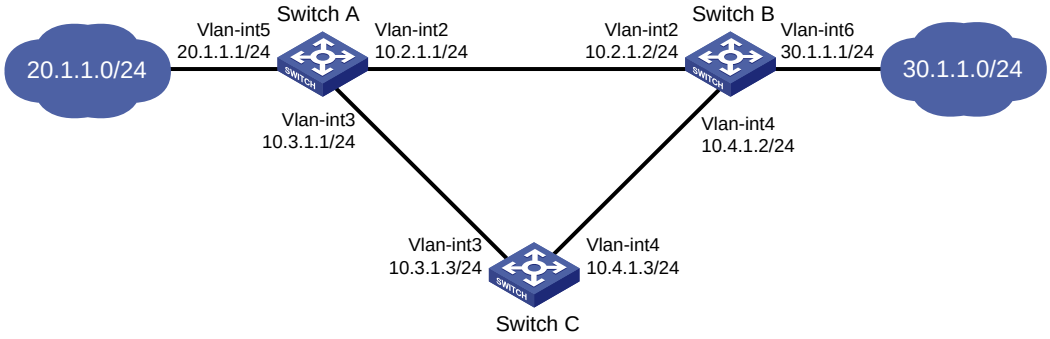
As shown in Figure 9:

- Switch A is the default gateway of the hosts in network 20.1.1.0/24.
- Switch B is the default gateway of the hosts in network 30.1.1.0/24.
- Hosts in the two networks communicate with each other through static routes.

To ensure network availability, configure route backup and static routing-Track-BFD (control mode) collaboration on Switch A and Switch B as follows:

- On Switch A, assign a higher priority to the static route to 30.1.1.0/24 with next hop Switch B. This route is the master route. The static route to 30.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the control-mode BFD session can quickly detect the route failure to make the backup route take effect.
- On Switch B, assign a higher priority to the static route to 20.1.1.0/24 with next hop Switch A. This route is the master route. The static route to 20.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the control-mode BFD session can quickly detect the route failure to make the backup route take effect.

Figure 9 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 9. (Details not shown.)
2. Configure Switch A:

   # Configure a static route to 30.1.1.0/24 with next hop 10.2.1.2 and the default priority (60). Associate this static route with track entry 1.

   ```
   <SwitchA> system-view
   [SwitchA] ip route-static 30.1.1.0 24 10.2.1.2 track 1
   ```

   # Configure a static route to 30.1.1.0/24 with next hop 10.3.1.3 and priority 80.

   ```
   [SwitchA] ip route-static 30.1.1.0 24 10.3.1.3 preference 80
   ```

   # Configure track entry 1, and associate it with the control-mode BFD session to verify the connectivity between Switch A and Switch B.

   ```
   [SwitchA] track 1 bfd ctrl interface vlan-interface 2 remote ip 10.2.1.2 local ip 10.2.1.1
   [SwitchA-track-1] quit
   ```

3. Configure Switch B:

   # Configure a static route to 20.1.1.0/24 with next hop 10.2.1.1 and the default priority (60). Associate this static route with track entry 1.

   ```
   <SwitchB> system-view
   [SwitchB] ip route-static 20.1.1.0 24 10.2.1.1 track 1
   ```

   # Configure a static route to 20.1.1.0/24 with next hop 10.4.1.3 and priority 80.

   ```
   [SwitchB] ip route-static 20.1.1.0 24 10.4.1.3 preference 80
   ```

   # Configure track entry 1, and associate it with the control-mode BFD session to verify the connectivity between Switch B and Switch A.

   ```
   [SwitchB] track 1 bfd ctrl interface vlan-interface 2 remote ip 10.2.1.1 local ip 10.2.1.2
   [SwitchB-track-1] quit
   ```

4. Configure Switch C:

   # Configure a static route to 30.1.1.0/24 with next hop 10.4.1.2.

   ```
   <SwitchC> system-view
   [SwitchC] ip route-static 30.1.1.0 24 10.4.1.2
   ```

   # Configure a static route to 20.1.1.0/24 with next hop 10.3.1.1.

   ```
   [SwitchC] ip route-static 20.1.1.0 24 10.3.1.1
   ```

Verifying the configuration

Display information about the track entry on Switch A.

```
[SwitchA] display track all
```

```
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD ctrl
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Ctrl
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 10.2.1.2
    Local IP: 10.2.1.1
```

The output shows that the status of the track entry is Positive, indicating that next hop 10.2.1.2 is reachable.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 9        Routes : 9

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan2
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         Direct 0    0            10.3.1.1        Vlan3
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        Vlan5
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.0/24         Static 60   0            10.2.1.2        Vlan2
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets to 30.1.1.0/24 through Switch B. The master static route has taken effect.

Remove the IP address of VLAN-interface 2 on Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] undo ip address
```

Display information about the track entry on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Negative
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: BFD ctrl
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    BFD session mode: Ctrl
    Outgoing interface: Vlan-interface2
    VPN instance name: --
    Remote IP: 10.2.1.2
    Local IP: 10.2.1.1
```

The output shows that the status of the track entry is Negative, indicating that next hop 10.2.1.2 is unreachable.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 9        Routes : 9

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan2
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         Direct 0    0            10.3.1.1        Vlan3
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        Vlan5
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.0/24         Static 80   0            10.3.1.3        Vlan3
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets to 30.1.1.0/24 through Switch C. The backup static route has taken effect.

Verify that the hosts in 20.1.1.0/24 can communicate with the hosts in 30.1.1.0/24 when the master route fails.

```
[SwitchA] ping -a 20.1.1.1 30.1.1.1
Ping 30.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 30.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
Reply from 30.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 30.1.1.1: bytes=56 Sequence=4 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 30.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Verify that the hosts in 30.1.1.0/24 can still communicate with the hosts in 20.1.1.0/24 when the master route fails.

```
[SwitchB] ping -a 30.1.1.1 20.1.1.1
Ping 20.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 20.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 20.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=4 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 20.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Example: Configuring VRRP-Track-interface management collaboration

Network configuration

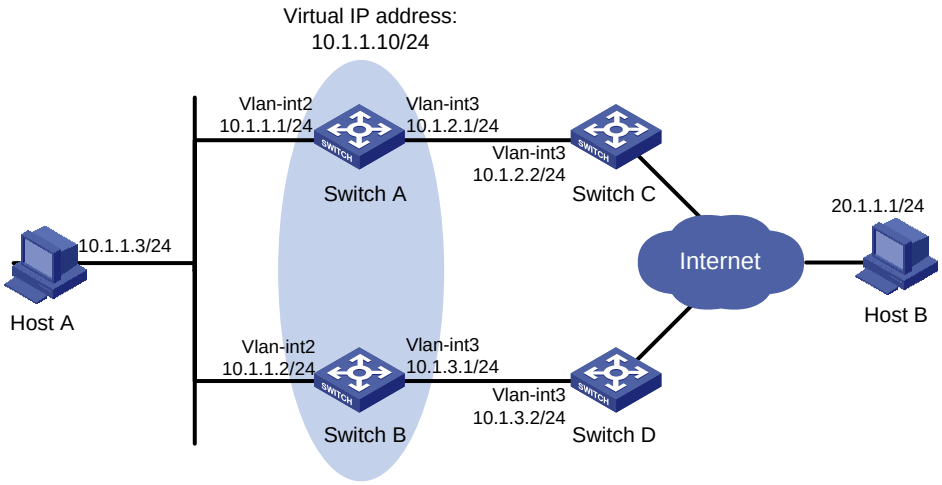
As shown in Figure 10:

- Host A requires access to Host B. The default gateway of Host A is 10.1.1.10/24.
- Switch A and Switch B belong to VRRP group 1. The virtual IP address of VRRP group 1 is 10.1.1.10.

Configure VRRP-Track-interface management collaboration to monitor the uplink interface on the master and meet the following requirements:

- When Switch A operates correctly, Switch A forwards packets from Host A to Host B.
- When VRRP detects a fault on the uplink interface of Switch A through the interface management module, Switch B forwards packets from Host A to Host B.

Figure 10 Network diagram



Procedure

1. Create VLANs and assign ports to them. Configure the IP address of each VLAN interface, as shown in Figure 10. (Details not shown.)
2. Configure Switch A:

   # Configure track entry 1 and associate it with the link status of the uplink interface VLAN-interface 3.

   ```
   [SwitchA] track 1 interface vlan-interface 3
   [SwitchA-track-1] quit
   ```

   # Create VRRP group 1 and configure virtual IP address 10.1.1.10 for the group.

   ```
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
   ```

   # Set the priority of Switch A to 110 in VRRP group 1.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
   ```

   # Associate VRRP group 1 with track entry 1 and decrease the router priority by 30 when the state of track entry 1 changes to Negative.

   ```
   [SwitchA-Vlan-interface2] vrrp vrid 1 track 1 priority reduced 30
   ```

3. On Switch B, create VRRP group 1, and configure virtual IP address 10.1.1.10 for the group.

   ```
   <SwitchB> system-view
   ```

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
```

Verifying the configuration

Ping Host B from Host A to verify that Host B is reachable. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
     Admin Status     : Up              State        : Master
     Config Pri       : 110             Running Pri  : 110
     Preempt Mode     : Yes             Delay Time   : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 10.1.1.1
   VRRP Track Information:
     Track Object   : 1              State : Positive          Pri Reduced : 30
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode       : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1               Adver Timer  : 100
     Admin Status     : Up              State        : Backup
     Config Pri       : 100             Running Pri  : 100
     Preempt Mode     : Yes             Delay Time   : 0
     Become Master    : 2200ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.10
     Master IP        : 10.1.1.1
```

The output shows that in VRRP group 1, Switch A is the master, and Switch B is a backup. Switch A forwards packets from Host A to Host B.

Shut down the uplink interface VLAN-interface 3 on Switch A.

```
[SwitchA-Vlan-interface2] interface vlan-interface 3
[SwitchA-Vlan-interface3] shutdown
```

Ping Host B from Host A to verify that Host B is reachable. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface3] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
```

```
  Interface Vlan-interface2
    VRID              : 1                Adver Timer   : 100
    Admin Status      : Up               State         : Backup
    Config Pri        : 110              Running Pri   : 80
    Preempt Mode      : Yes              Delay Time    : 0
    Become Master     : 2200ms left
    Auth Type         : Not supported
    Version           : 3
    Virtual IP        : 10.1.1.10
    Master IP         : 10.1.1.2
  VRRP Track Information:
    Track Object   : 1              State : Negative          Pri Reduced : 30
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID              : 1                Adver Timer   : 100
     Admin Status      : Up               State         : Master
     Config Pri        : 100              Running Pri   : 100
     Preempt Mode      : Yes              Delay Time    : 0
     Auth Type         : Not supported
     Version           : 3
     Virtual IP        : 10.1.1.10
     Virtual MAC       : 0000-5e00-0101
     Master IP         : 10.1.1.2
```

The output shows that Switch A becomes the backup, and Switch B becomes the master. Switch B forwards packets from Host A to Host B.

Example: Configuring VRRP-Track-route management collaboration

Network configuration

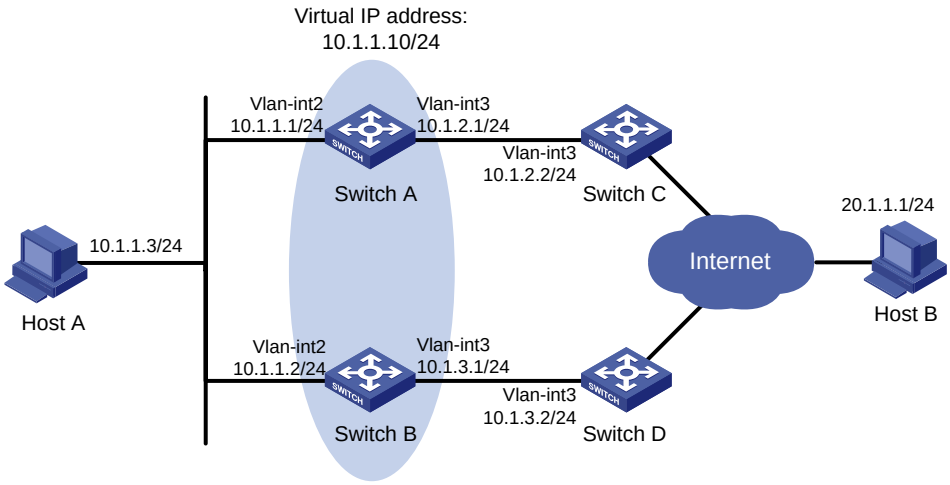
As shown in Figure 11:

- Host A requires access to Host B. The default gateway of Host A is 10.1.1.10/24.
- Switch A and Switch B belong to VRRP group 1. The virtual IP address of VRRP group 1 is 10.1.1.10.
- BGP peer relationships are established between Switch A and Switch C and between Switch B and Switch D. Switch C and Switch D advertise the default route 0.0.0.0/0 to Switch A and Switch B.

Configure VRRP-Track-route management collaboration to meet the following requirements:

- When Switch A operates correctly, Switch A forwards packets from Host A to Host B.
- When VRRP detects the removal of the default route from the routing table of Switch A through route management, Switch B forwards packets from Host A to Host B.

Figure 11 Network diagram



Procedure

1. Configure the IP address of each interface, as shown in Figure 11. (Details not shown.)
2. Establish an IBGP peer relationship between Switch A and Switch C, and configure Switch C to advertise the default route 0.0.0.0/0 to Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp-default] peer 10.1.2.2 as-number 100
[SwitchA-bgp-default] address-family ipv4
[SwitchA-bgp-default-ipv4] peer 10.1.2.2 enable
<SwitchC> system-view
[SwitchC] bgp 100
[SwitchC-bgp-default] peer 10.1.2.1 as-number 100
[SwitchC-bgp-default] address-family ipv4
[SwitchC-bgp-default-ipv4] peer 10.1.2.1 enable
[SwitchC-bgp-default-ipv4] peer 10.1.2.1 default-route-advertise
[SwitchC-bgp-default-ipv4] quit
```

3. Configure Switch B and Switch D in the same way Switch A and Switch C are configured. (Details not shown.)
4. Configure Track and VRRP on Switch A:

   # Configure track entry 1, and associate it with the default route 0.0.0.0/0.

```
[SwitchA] track 1 ip route 0.0.0.0 0.0.0.0 reachability
[SwitchA-track-1] quit
```

   # Create VRRP group 1, and configure virtual IP address 10.1.1.10 for the group.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
```

   # Set the priority of Switch A to 110 in VRRP group 1.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

   # Associate VRRP group 1 with track entry 1 and decrease the Switch priority by 30 when the state of track entry 1 changes to Negative.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 track 1 priority reduced 30
[SwitchA-Vlan-interface2] quit
```

5. On Switch B, create VRRP group 1, and configure virtual IP address 10.1.1.10 for the group.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
[SwitchB-Vlan-interface2] quit
```

Verifying the configuration

Ping Host B from Host A to verify that Host B is reachable. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode       : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1              Adver Timer  : 100
     Admin Status     : Up             State        : Master
     Config Pri       : 110            Running Pri  : 110
     Preempt Mode     : Yes            Delay Time   : 0
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.10
     Virtual MAC      : 0000-5e00-0101
     Master IP        : 10.1.1.1
   VRRP Track Information:
     Track Object   : 1              State : Positive          Pri Reduced : 30
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode       : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID             : 1              Adver Timer  : 100
     Admin Status     : Up             State        : Backup
     Config Pri       : 100            Running Pri  : 100
     Preempt Mode     : Yes            Delay Time   : 0
     Become Master    : 2200ms left
     Auth Type        : Not supported
     Version          : 3
     Virtual IP       : 10.1.1.10
     Master IP        : 10.1.1.1
```

The output shows that in VRRP group 1, Switch A is the master and Switch B is a backup. Switch A forwards packets from Host A to Host B.

Disable Switch C from exchanging routing information with Switch A so that the default route 0.0.0.0/0 is removed from the routing table of Switch A.

```
[SwitchC-bgp-default-ipv4] undo peer 10.1.2.1 enable
```

Ping Host B from Host A to verify that Host B is reachable. (Details not shown.)

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA] display vrrp verbose
IPv4 Virtual Router Information:
```

```
Running Mode      : Standard
Total number of virtual routers : 1
  Interface Vlan-interface2
    VRID              : 1               Adver Timer  : 100
    Admin Status      : Up              State        : Backup
    Config Pri        : 110             Running Pri  : 80
    Preempt Mode      : Yes             Delay Time   : 0
    Become Master     : 2200ms left
    Auth Type         : Not supported
    Version           : 3
    Virtual IP        : 10.1.1.10
    Master IP         : 10.1.1.2
  VRRP Track Information:
    Track Object   : 1              State : Negative          Pri Reduced : 30
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB] display vrrp verbose
IPv4 Virtual Router Information:
 Running Mode      : Standard
 Total number of virtual routers : 1
   Interface Vlan-interface2
     VRID              : 1               Adver Timer  : 100
     Admin Status      : Up              State        : Master
     Config Pri        : 100             Running Pri  : 100
     Preempt Mode      : Yes             Delay Time   : 0
     Auth Type         : Not supported
     Version           : 3
     Virtual IP        : 10.1.1.10
     Virtual MAC       : 0000-5e00-0101
     Master IP         : 10.1.1.2
```

The output shows that Switch A becomes the backup, and Switch B becomes the master. Switch B forwards packets from Host A to Host B.

Example: Configuring static routing-Track-LLDP collaboration

Network requirements

As shown in Figure 12:

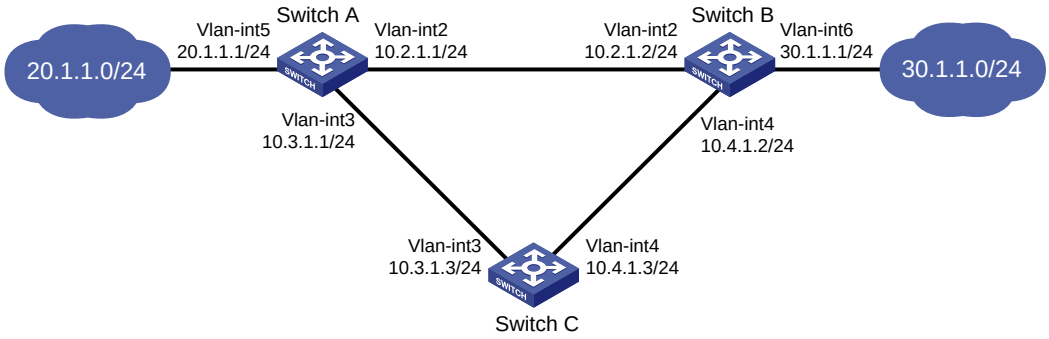
- Switch A is the default gateway of the hosts in network 20.1.1.0/24.
- Switch B is the default gateway of the hosts in network 30.1.1.0/24.
- Hosts in the two networks communicate with each other through static routes.

To ensure network availability, configure route backup and static routing-Track-LLDP collaboration on Switch A and Switch B as follows:

- On Switch A, assign a higher priority to the static route to 30.1.1.0/24 with next hop Switch B. This route is the master route. The static route to 30.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the backup route takes effect. Switch A forwards packets destined for 30.1.1.0/24 to Switch C.

- On Switch B, assign a higher priority to the static route to 20.1.1.0/24 with next hop Switch A. This route is the master route. The static route to 20.1.1.0/24 with next hop Switch C acts as the backup route. When the master route is unavailable, the backup route takes effect. Switch B forwards packets destined for 20.1.1.0/24 to Switch C.

Figure 12 Network diagram



Procedure

1. Configure the IP address of each interface, as shown in Figure 12. (Details not shown.)
2. Configure Switch A:

   # Configure a static route to 30.1.1.0/24 with next hop 10.1.1.2 and the default priority (60). Associate this static route with track entry 1.

   ```
   <SwitchA> system-view
   [SwitchA] ip route-static 30.1.1.0 24 10.2.1.2 track 1
   ```

   # Configure a static route to 30.1.1.0/24 with next hop 10.3.1.3 and priority 80.

   ```
   [SwitchA] ip route-static 30.1.1.0 24 10.3.1.3 preference 80
   ```

   # Enable LLDP globally.

   ```
   [SwitchA] lldp global enable
   ```

   # Enable LLDP on GigabitEthernet 1/0/1. (This step is optional because LLDP is enabled on the port by default.)

   ```
   [SwitchA] interface gigabitethernet 1/0/1
   [SwitchA-GigabitEthernet1/0/1] lldp enable
   ```

   # Configure track entry 1 and associate it with the availability of the neighbor for LLDP interface Gigabitethernet 1/0/1.

   ```
   [SwitchA] track 1 lldp neighbor interface gigabitethernet 1/0/1
   [SwitchA-track-1] quit
   ```

3. Configure Switch B:

   # Configure a static route to 20.1.1.0/24 with next hop 10.2.1.1 and the default priority (60). Associate this static route with track entry 1.

   ```
   <SwitchB> system-view
   [SwitchB] ip route-static 20.1.1.0 24 10.2.1.1 track 1
   ```

   # Configure a static route to 20.1.1.0/24 with next hop 10.4.1.3 and priority 80.

   ```
   [SwitchB] ip route-static 20.1.1.0 24 10.4.1.3 preference 80
   ```

   # Enable LLDP globally.

   ```
   [SwitchB] lldp global enable
   ```

   # Enable LLDP on GigabitEthernet 1/0/1. (This step is optional because LLDP is enabled on the port by default.)

   ```
   [SwitchB] interface gigabitethernet 1/0/1
   [SwitchB-GigabitEthernet1/0/1] lldp enable
   ```

Configure track entry 1 and associate it with the availability of the neighbor for LLDP interface GigabitEthernet 1/0/1.

```
[SwitchB] track 1 lldp neighbor interface gigabitethernet 1/0/1
[SwitchB-track-1] quit
```

4. Configure Switch C:

Configure a static route to 30.1.1.0/24 with next hop 10.4.1.2.

```
<SwitchC> system-view
[SwitchC] ip route-static 30.1.1.0 24 10.4.1.2
```

Configure a static route to 20.1.1.0/24 with next hop 10.3.1.1.

```
[SwitchC] ip route-static 20.1.1.0 24 10.3.1.1
```

Verifying the configuration

Display track entry information on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 32 seconds
  Tracked object type: LLDP
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    LLDP interface: GigabitEthernet1/0/1
```

The output shows that the status of track entry 1 is Positive, indicating that the neighbor of LLDP interface GigabitEthernet 1/0/1 is available. The master route takes effect.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 9        Routes : 9

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         Direct 0    0            10.2.1.1        GE1/0/1
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         Direct 0    0            10.3.1.1        GE1/0/2
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        GE1/0/3
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.0/24         Static 60   0            10.2.1.2        GE1/0/1
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets to 30.1.1.0/24 through Switch B.

On Switch B, disable LLDP on GigabitEthernet 1/0/1.

```
<SwitchB> system-view
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] undo lldp enable
```

Display track entry information on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Negative
  Duration: 0 days 0 hours 0 minutes 32 seconds
```

```
  Tracked object type: LLDP
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    LLDP interface: GigabitEthernet1/0/1
```

The output shows that the status of track entry 1 is Negative, indicating that the neighbor of LLDP interface GigabitEthernet 1/0/1 is unavailable. The master route fails.

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table

Destinations : 9        Routes : 9

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         Direct 0    0            10.2.1.1        GE1/0/1
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         Direct 0    0            10.3.1.1        GE1/0/2
10.3.1.1/32         Direct 0    0            127.0.0.1       InLoop0
20.1.1.0/24         Direct 0    0            20.1.1.1        GE1/0/3
20.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
30.1.1.0/24         Static 80   0            10.3.1.3        GE1/0/2
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output shows that Switch A forwards packets destined for 30.1.1.0/24 to Switch C. The backup static route has taken effect.

Verify that hosts in 20.1.1.0/24 can communicate with the hosts in 30.1.1.0/24 when the master route fails.

```
[SwitchA] ping -a 20.1.1.1 30.1.1.1
Ping 30.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 30.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
Reply from 30.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 30.1.1.1: bytes=56 Sequence=4 ttl=254 time=2 ms
Reply from 30.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 30.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Verify that the hosts in 30.1.1.0/24 can communicate with the hosts in 20.1.1.0/24 when the master route fails.

```
[SwitchB] ping -a 30.1.1.1 20.1.1.1
Ping 20.1.1.1: 56  data bytes, press CTRL_C to break
Reply from 20.1.1.1: bytes=56 Sequence=1 ttl=254 time=2 ms
Reply from 20.1.1.1: bytes=56 Sequence=2 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=3 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=4 ttl=254 time=1 ms
Reply from 20.1.1.1: bytes=56 Sequence=5 ttl=254 time=1 ms

--- Ping statistics for 20.1.1.1 ---
5 packet(s) transmitted, 5 packet(s) received, 0.00% packet loss
```

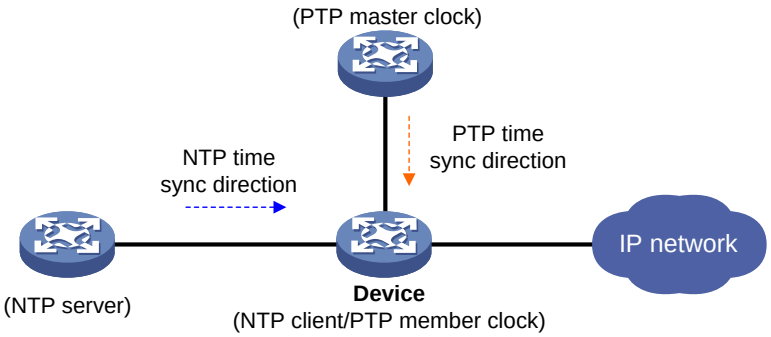
```
round-trip min/avg/max/std-dev = 1/1/2/1 ms
```

Example: Configuring NTP-PTP-Track collaboration

Network configuration

To ensure correct clock information on the devices in the system, deploy both PTP and NTP clocks in the network. The NTP clock is used as the backup solution for the PTP clock. The device preferentially uses the PTP clock. If the PTP clock cannot operate correctly, the device automatically switched to the NTP clock.

Figure 13 Network diagram



Procedure

1. Complete PTP and NTP settings. (Details not shown.)
2. Configure a Boolean AND tracked list to monitor both the PTP clock source state and PTP clock state.

   # Configure track entry 11 to monitor the PTP clock source state.

   ```
   <Device> system-view
   [Device] track 11 ptp source-state
   [Device-track-11] quit
   ```

   # Configure track entry 12 to monitor the PTP clock source state.

   ```
   [Device] track 12 ptp time-state
   [Device-track-12] quit
   ```

   # Configure Boolean AND tracked list 1 to monitor both the PTP clock source state and PTP clock state 1.

   ```
   [Device] track 1 list boolean and
   [Device-track-1] object 11
   [Device-track-1] object 12
   [Device-track-1] quit
   ```

3. Associate EAA monitoring policy ptp with the Boolean AND tracked list. When PTP operates correctly, the PTP time is used as the system time.

   ```
   [Device] rtm cli-policy ptp
   [Device-rtm-ptp] event track 1 state positive
   [Device-rtm-ptp] action 0 cli system-view
   [Device-rtm-ptp] action 1 cli clock protocol ptp
   [Device-rtm-ptp] user-role network-admin
   [Device-rtm-ptp] commit
   [Device-rtm-ptp] quit
   ```

4. Associate EAA monitoring policy ntp with the Boolean AND tracked list. When PTP cannot operate correctly, the NTP time is used as the system time.

```
[Device] rtm cli-policy ntp
[Device-rtm-ntp] event track 1 state negative
[Device-rtm-ntp] action 0 cli system-view
[Device-rtm-ntp] action 1 cli clock protocol ntp
[Device-rtm-ntp] user-role network-admin
[Device-rtm-ntp] commit
[Device-rtm-ntp] quit
```

Verifying the configuration

If the link between the device and PTP master clock fails, you can see that the state of track entry 1 is Negative in the output from the display track all command. The device executes EAA policy ntp, and outputs the log of successful execution of EAA policy ntp. The system uses the NTP time as the system time.

Contents

Configuring process placement 1

About process placement 1
Process 1
Node 1
Process redundancy 1
Process placement 1
Default process placement policy 1
Process placement affinities 2
Process placement optimization 2
Restrictions and guidelines: process placement configuration 2
Process placement tasks at a glance 3
Configuring process placement policy 3
Configuring a location affinity 3
Configuring a location type affinity 3
Configuring a process affinity 4
Configuring a self affinity 4
Optimizing process placement 5
Display and maintenance commands for process placement 5

Configuring process placement

About process placement

Process

A process contains a set of codes and provides specific functionality. For example, an AAA process provides AAA functions.

Each process runs in a protected memory space to prevent problems with one process from impacting the entire system.

Node

The term "node" in this document refers to CPUs that run the Comware system. If the IRF fabric has multiple member devices, each member device has a node.

Process redundancy

The system backs up each active process running on one node to all the other nodes. When an active process fails, one of its standby processes promptly takes over without impacting any other service.

Process redundancy provides the following benefits:

- Improves service availability.
- Enables the system to quickly regain reliability after device status changes in such conditions as removal of an IRF member.

Process placement

The execution of process placement policies varies by the location of active processes.

- An active process running only on the master device does not support placement optimization. If you configure a process placement policy for the process, the system displays a configuration failure message. When such an active process fails, the system automatically restarts the process. The standby processes are used for active/standby switchover and ISSU.
- Some active processes can run on either the master or subordinate device. When such an active process fails, the system uses a placement policy to select a new active process among standby processes.

Default process placement policy

The system provides a default process placement policy that takes effect for all processes. By default, the default process placement policy defines the following rules:

- The active process runs on the CPU of the master, and the standby processes run on the CPU of the subordinate device.
- A process runs at the location where it ran the last time and does not move to any other location during startup or operation.

- The addition of a new node does not impact current active processes. A new active process selects one node with sufficient CPU and memory resources. (You can use the **display cpu-usage** and **display memory** commands to view CPU and memory usage information.)

You can modify the default placement policy in the view you enter by using the **placement program default** command. You can also configure a placement policy for a specific process in the view you enter by using the **placement program** *program-name* [**instance** *instance-name*] command. A placement policy for a process takes precedence over the default process placement policy.

Process placement affinities

You can configure the following settings for a process placement policy to optimize process placement:

- **affinity location-set**—Location affinity, the preference for the process to run on a specific node.
- **affinity location-type**—Location type affinity, the preference for the process to run on a particular type of node. For more information about node types, see "Configuring a location type affinity."
- **affinity program**—Process affinity, the preference for the process to run on the same node as a particular process.
- **affinity self**—Self affinity, the preference for one instance of the process to run on the same node as any other instance of the process.

Affinities include positive affinities (**attract**) and negative affinities (**repulse**), all represented by integers in the range of 1 to 100000.

- The higher the attract value, the stronger the preference.
- The higher the repulse value, the weaker the preference.

Process placement optimization

After you apply new placement policies, the system makes placement decisions based on the new policies, node resources, and topology status. If the new location for an active process is different from the current node, the system changes the state of the process to standby, and uses the standby process on the preferred location as the new active process.

Restrictions and guidelines: process placement configuration

- Configuring process placement on an IRF fabric with only one member device does not change the location of processes. All processes run on the CPU of the IRF member device.
- Configuring process placement on an IRF fabric with multiple member devices places specific active processes to specific CPUs. In case of multiple CPUs, the system performs process redundancy. The number of standby processes and their CPU locations vary by function module. The system by default automatically determines the location for each active process, and process placement optimization is not required. If optimization is required, contact H3C Support to avoid service interruption.
- For an instance of a process, the settings in placement policy view of the instance, in placement policy view of the process, and in default placement policy view are in descending order of priority. For a process, the settings in placement policy view of the process take precedence over the settings in default placement policy view.

- To view the current location of an active process and its predicted new location after optimization, use the **display placement reoptimize** command.

Process placement tasks at a glance

To configure process placement, perform the following tasks:

1. Configuring process placement policy

   Choose the following tasks as needed:

   - Configuring a location affinity
   - Configuring a location type affinity
   - Configuring a process affinity
   - Configuring a self affinity
2. Optimizing process placement

Configuring process placement policy

Configuring a location affinity

1. Enter system view.

   system-view
2. Enter placement process view.
   - Enter default placement process view.
   - **placement program default**
   - Enter specified placement process view.
   - **placement program** { *program-name* [**instance** *instance-name*]

   Settings in default placement process view take effect for all processes. Settings in placement process view take effect only for the specified process.
3. Set the location affinity.

   affinity location-set { **slot** *slot-number* }&<1-5> { **attract** *strength* | **default** | **none** | **repulse** *strength* }

   By default, no location affinity is set.

Configuring a location type affinity

About this task

The following location types are available:

- **current**—Current location of the active process, which can be displayed with the **display placement program** command.
- **paired**—Locations of standby processes.
- **primary**—Master device.

Procedure

1. Enter system view.

   system-view
2. Enter placement process view.

- Enter default placement process view.
- **placement program default**
- Enter specified placement process view.
- **placement program** { *program-name* [**instance** *instance-name*]

Settings in default placement process view take effect for all processes. Settings in placement process view take effect only for the specified process.

3. Set the location type affinity.

   affinity location-type { **current** | **paired** | **primary** } { **attract** *strength* | **repulse** *strength* | **default** | **none** }

   By default, no location type affinity is set.

Configuring a process affinity

1. Enter system view.

   system-view

2. Enter placement process view.
   - Enter default placement process view.
   - **placement program default**
   - Enter specified placement process view.
   - **placement program** { *program-name* [**instance** *instance-name*]

   Settings in default placement process view take effect for all processes. Settings in placement process view take effect only for the specified process.

3. Configure the affinity for the process to run on the same location as another process.

   affinity program *program-name* { **attract** *strength* | **default** | **none** | **repulse** *strength* }

   By default, no process affinity is set.

Configuring a self affinity

About this task

A self affinity determines the preference for one instance of a process to run on the same node as any other instance of the process. The self affinity setting does not take effect for a process that has only one instance.

Procedure

1. Enter system view.

   system-view

2. Enter placement process view.
   - Enter default placement process view.
   - **placement program default**
   - Enter specified placement process view.
   - **placement program** { *program-name* [**instance** *instance-name*]

   Settings in default placement process view take effect for all processes. Settings in placement process view take effect only for the specified process.

3. Configure a self affinity.

   affinity self { **attract** *strength* | **repulse** *strength* | **default** | **none** }

   By default, no self affinity is set.

Optimizing process placement

Restrictions and guidelines

CAUTION:
To avoid neighbor flapping of related protocols, make sure HA features such as NSR or GR are configured for the processes and are stable before optimizing process placement.

To keep the system stable, do not perform any tasks that require process restart when you optimize process placement.

Procedure

1. Enter system view.
   system-view
2. Optimize process placement.
   placement reoptimize

Display and maintenance commands for process placement

Execute **display** commands in any view.

Task	Command
Display service group information.	**display ha service-group** { *program-name* [**instance** *instance-name*] | **all** }
Display the running processes on a specific location.	**display placement location** { **all** | **slot** *slot-number* }
Display process placement policy information.	**display placement policy program** { program-name | **all** | **default** }
Display the location of a process.	**display placement program** { *program-name* | **all** }
Display the predicted location of a process after process placement optimization.	**display placement reoptimize program** { *program-name* [**instance** *instance-name*] | **all** }

Contents

Configuring interface collaboration 1

About interface collaboration 1
Typical application 1
How interface collaboration works 1
Restrictions and guidelines: Interface collaboration configuration 2
Interface collaboration tasks at a glance 2
Prerequisites for interface collaboration 2
Creating a collaboration group 2
Configuring collaboration group member interfaces 2
Restrictions and guidelines 2
Configuring a collaboration group member interface in collaboration group view 3
Configuring a collaboration group member interface in interface view 3
Configuring the delay for the member interfaces in a collaboration group to come up 3
Removing ineffective member interfaces from all collaboration groups 4
Display and maintenance commands for interface collaboration 4
Interface collaboration configuration examples 4
Example: Configuring interface collaboration for Layer 3 interfaces 4
Example: Configuring interface collaboration for Layer 2 interfaces 8

Configuring interface collaboration

About interface collaboration

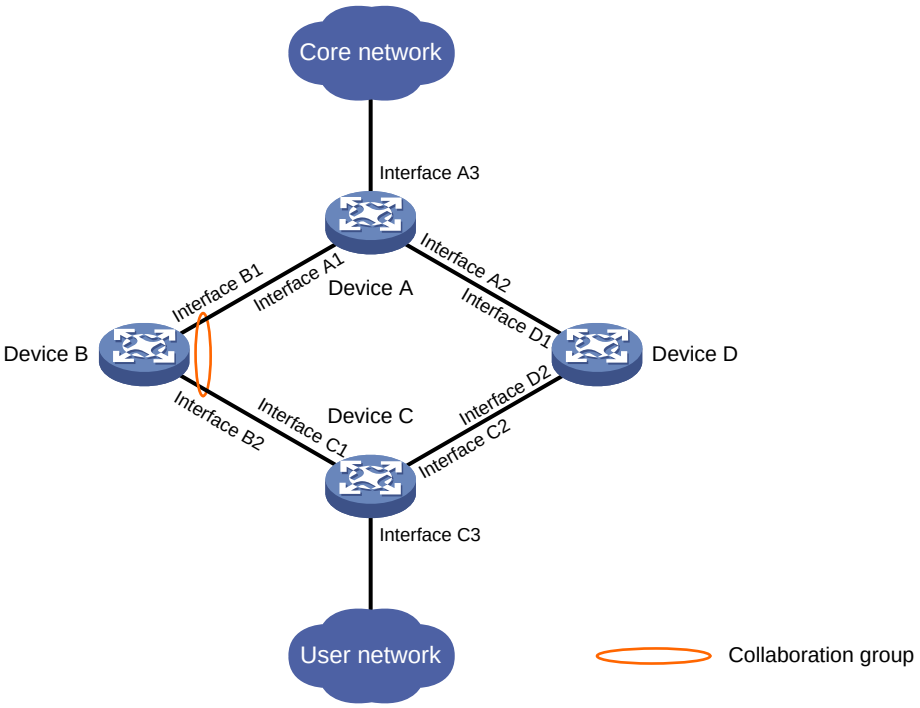
The interface collaboration feature assigns different interfaces on a device to a collaboration group and associates the states of these interfaces. All member interfaces in a collaboration group can or cannot transmit packets at the same time. This feature is typically used to associate the states of downlink and uplink interfaces.

Typical application

As shown in Figure 1, Device C accesses the core network through Device B. When Interface B1 goes down, the switchover of traffic from Device C to Device D is slow because Interface B2 is still up.

If Interface B1 and Interface B2 belong to one collaboration group, Device B brings down Interface B2 when Interface B1 goes down to achieve fast traffic switchover. Similarly, the device brings down Interface B1 when Interface B2 goes down.

Figure 1 Interface collaboration application scenario



How interface collaboration works

The interface collaboration feature works as follows:

- When any member interface in a collaboration group goes down, the device sets all other member interfaces in the collaboration group to the **Collaboration-down** state. The state of the collaboration group is down, and no member interfaces in the collaboration group can transmit packets.

- When any member interface in **DOWN** or **Collaboration-down** state comes up, the device attempts to bring up all other member interfaces in the collaboration group.
  - If all other member interfaces come up in 10 seconds, the collaboration group comes up. All member interfaces in the collaboration group can transmit packets.
  - If any member interface cannot come up in 10 seconds, the device sets that member interface to **DOWN** state and sets all other member interfaces to the **Collaboration-down** state. The collaboration group goes down, and no member interfaces in the collaboration group can transmit packets.

Restrictions and guidelines: Interface collaboration configuration

This feature cannot be used together with the external security service bypass feature. For more information about external security service bypass, see Layer 2 forwarding configuration in *Layer 2—LAN Switching Configuration Guide*.

Interface collaboration tasks at a glance

To configure interface collaboration, perform the following tasks:

1. Creating a collaboration group
2. Configuring collaboration group member interfaces
3. (Optional.) Configuring the delay for the member interfaces in a collaboration group
4. (Optional.) Removing ineffective member interfaces from all collaboration groups

Prerequisites for interface collaboration

Before configuring interface collaboration, you must enable Monitor Link globally (see "Configuring Monitor Link").

Creating a collaboration group

1. Enter system view.
   `system-view`
2. Create a collaboration group and enter collaboration group view.
   `collaboration-group` *`group-id`*
   By default, no collaboration groups exist.

Configuring collaboration group member interfaces

Restrictions and guidelines

An interface can belong to only one collaboration group.

For a collaboration group to work correctly, do not assign its member interfaces to a redundancy group.

Do not add both an aggregate interface and its member ports to the same collaboration group.

Do not add both Selected ports and Unselected ports in a dynamic aggregation group to the same collaboration group.

If you have configured a member interface as the uplink/downlink interface of a monitor link group, do not configure any other member interface in the same collaboration group as the downlink/uplink interface of any monitor link group.

You can assign only one interface of a link to a collaboration group.

You can configure member interfaces for a collaboration group in collaboration group view or interface view. Configurations made in these views have the same effect.

Configuring a collaboration group member interface in collaboration group view

1. Enter system view.
   system-view
2. Enter collaboration group view.
   collaboration-group *group-id*
3. Configure an interface as a member of the collaboration group.
   port *interface-type interface-number*
   By default, no member interfaces exist in a collaboration group.

Configuring a collaboration group member interface in interface view

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the interface as a member of a collaboration group.
   port collaboration-group *group-id*
   By default, an interface is not a collaboration group member.

Configuring the delay for the member interfaces in a collaboration group to come up

About this task

By default, member interfaces in a collaboration group come up and forward service traffic immediately after the device restarts. Traffic loss occurs if the service modules have not returned to normal state.

Perform this task to enable the member interfaces to come up after a configurable delay time upon a device restart.

Procedure

1. Enter system view.
   system-view

2. Enter collaboration group view.
   collaboration-group *group-id*
3. Configure the delay for the member interfaces in a collaboration group to come up.
   up-delay *delay*
   By default, the delay time is 0 seconds. The member interfaces come up as soon as the device restarts.

Removing ineffective member interfaces from all collaboration groups

About this task

A member interface in a collaboration group becomes ineffective when the card that hosts the interface is removed or changed to another slot or the ID of the IRF member device that hosts the interface changes. This task prevents an ineffective interface from causing all other member interfaces in the same collaboration group to go down.

An ineffective interface cannot be automatically assigned to the original collaboration group when its hosting card is reinstalled or changed back to the original slot or the IRF member ID is changed back to the original ID. You must assign it to the original collaboration group manually.

Procedure

1. Enter system view.
   system-view
2. Enter collaboration group view.
   collaboration-group clean

Display and maintenance commands for interface collaboration

Execute the **display** command in any view:

Task	Command
Display collaboration group information.	**display collaboration-group** { *group-id* | **all** } [**verbose**]

Interface collaboration configuration examples

Example: Configuring interface collaboration for Layer 3 interfaces

Network configuration

As shown in Figure 2, configure interface collaboration on Device B. When GigabitEthernet 1/0/1 on Device B goes down, Device B brings down GigabitEthernet 1/0/2. When GigabitEthernet 1/0/2 on Device B goes down, Device B brings down GigabitEthernet 1/0/1.

Figure 2 Network diagram

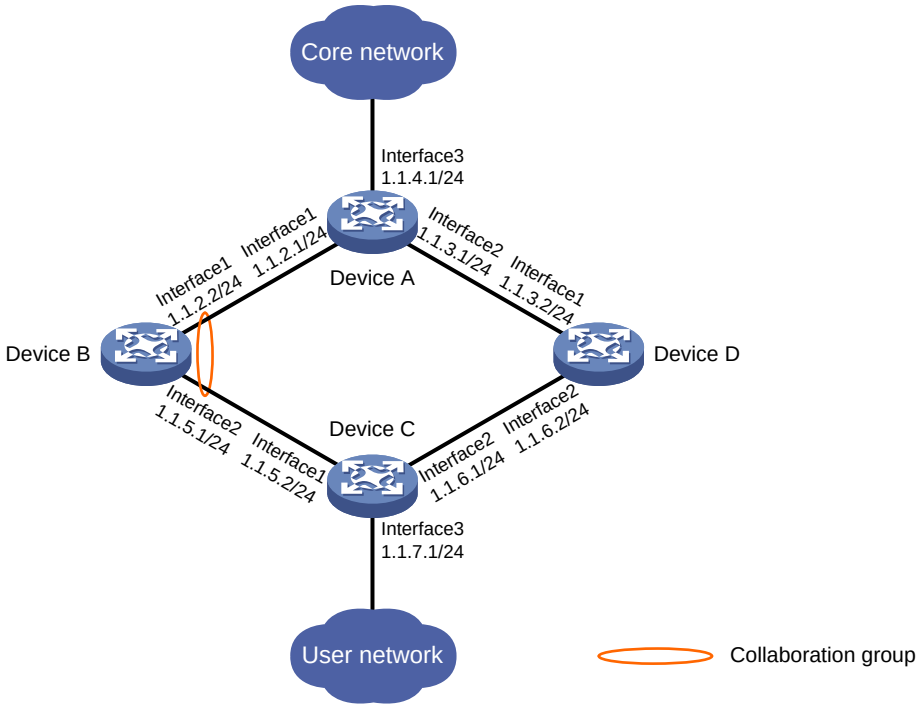


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure routes to enable Device A and Device C to prefer the route through Device B for forwarding traffic to each other. (This example uses static routes):

   # Configure IP addresses for interfaces. (Details not shown.)

   # Configure four static routes on Device A.

   ```
   <DeviceA> system-view
   [DeviceA] ip route-static 1.1.5.0 255.255.255.0 1.1.2.2
   [DeviceA] ip route-static 1.1.7.0 255.255.255.0 1.1.2.2 preference 50
   [DeviceA] ip route-static 1.1.6.0 255.255.255.0 1.1.3.2
   [DeviceA] ip route-static 1.1.7.0 255.255.255.0 1.1.3.2 preference 70
   ```

   # Configure two static routes on Device B.

   ```
   <DeviceB> system-view
   [DeviceB] ip route-static 1.1.4.0 255.255.255.0 1.1.2.1
   [DeviceB] ip route-static 1.1.7.0 255.255.255.0 1.1.5.2
   ```

   # Configure four static routes on Device C.

   ```
   <DeviceC> system-view
   [DeviceC] ip route-static 1.1.2.0 255.255.255.0 1.1.5.1
   [DeviceC] ip route-static 1.1.4.0 255.255.255.0 1.1.5.1 preference 50
   [DeviceC] ip route-static 1.1.3.0 255.255.255.0 1.1.6.2
   ```

```
[DeviceC] ip route-static 1.1.4.0 255.255.255.0 1.1.6.2 preference 70
```

Configure two static routes on Device D.

```
<DeviceD> system-view
[DeviceD] ip route-static 1.1.4.0 255.255.255.0 1.1.3.1
[DeviceD] ip route-static 1.1.7.0 255.255.255.0 1.1.6.1
```

2. Configure interface collaboration on Device B:

Enter system view.

```
<DeviceB> system-view
```

Enable Monitor Link globally. By default, Monitor Link is enabled globally.

```
[DeviceB] undo monitor-link disable
```

Create collaboration group 1.

```
[DeviceB] collaboration-group 1
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to collaboration group 1.

```
[DeviceB-collaboration-group1] port gigabitethernet 1/0/1
[DeviceB-collaboration-group1] port gigabitethernet 1/0/2
```

Verifying the configuration

1. Verify that the state of the uplink interface affects the state of the downlink interface.

Display the static routing information on Device C.

```
[DeviceC] display ip routing-table protocol static

Summary count : 3

Static Routing table status : <Active>
Summary count : 3

Destination/Mask   Proto   Pre Cost        NextHop         Interface
1.1.2.0/24         Static  60  0           1.1.5.1         GE1/0/1
1.1.3.0/24         Static  60  0           1.1.6.2         GE1/0/2
1.1.4.0/24         Static  50  0           1.1.5.1         GE1/0/1

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that the effective route from Device C to Device A goes through Device B.

Shut down GigabitEthernet 1/0/1 on Device B, and display information about collaboration group 1.

```
[DeviceB] display collaboration-group 1 verbose
Collaboration group protocol status: Enabled
Collaboration group 1 information:
  Group status     : DOWN
  Member up delay  : 0 seconds
  Last up time     : 16:55:34 2017/04/05
  Last down time   : 16:57:22 2017/04/05
  Member                    Status
  GE1/0/1                   DOWN
  GE1/0/2                   Collaboration-down
```

The output shows that collaboration group 1 is in DOWN state, GigabitEthernet 1/0/1 is in DOWN state, and GigabitEthernet 1/0/2 is in Collaboration-down state.

Display the static routing information on Device C.

```
[DeviceC] display ip routing-table protocol static

Summary count : 2

Static Routing table status : <Active>
Summary count : 2

Destination/Mask   Proto   Pre Cost        NextHop         Interface
1.1.3.0/24         Static  60  0           1.1.6.2         GE1/0/2
1.1.4.0/24         Static  70  0           1.1.6.2         GE1/0/2

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that the effective route from Device C to Device A goes through Device D.

2. Verify that the state of the downlink interface affects the state of the uplink interface.

Display the static routing information on Device A.

```
[DeviceA] display ip routing-table protocol static

Summary count : 3

Static Routing table status : <Active>
Summary count : 3

Destination/Mask   Proto   Pre Cost        NextHop         Interface
1.1.5.0/24         Static  60  0           1.1.2.2         GE1/0/1
1.1.6.0/24         Static  60  0           1.1.3.2         GE1/0/2
1.1.7.0/24         Static  50  0           1.1.2.2         GE1/0/1

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that the effective route from Device A to Device C goes through Device B.

Shut down GigabitEthernet 1/0/2 on Device B, and display information about collaboration group 1.

```
[DeviceB] display collaboration-group 1 verbose
Collaboration group protocol status: Enabled
Collaboration group 1 information:
  Group status     : DOWN
  Member up delay  : 0 seconds
  Last up time     : 17:11:47 2017/04/05
  Last down time   : 17:14:37 2017/04/05
  Member                    Status
  GE1/0/1                   Collaboration-down
  GE1/0/2                   DOWN
```

The output shows that collaboration group 1 is in DOWN state, GigabitEthernet 1/0/1 is in Collaboration-down state, and GigabitEthernet 1/0/2 is in DOWN state.

Display the static routing information on Device A.

```
[DeviceA] display ip routing-table protocol static
```

```
Summary count : 2

Static Routing table status : <Active>
Summary count : 2

Destination/Mask    Proto   Pre Cost        NextHop         Interface
1.1.6.0/24          Static  60  0           1.1.3.2         GE1/0/2
1.1.7.0/24          Static  70  0           1.1.3.2         GE1/0/2

Static Routing table status : <Inactive>
Summary count : 0
```

The output shows that the effective route from Device A to Device C goes through Device D.

Example: Configuring interface collaboration for Layer 2 interfaces

Network configuration

As shown in Figure 3, Device C is a Smart Link device, and Device A cannot receive flush messages.

Configure interface collaboration on Device B. When GigabitEthernet 1/0/1 on Device B goes down, Device B brings down GigabitEthernet 1/0/2. When GigabitEthernet 1/0/2 on Device B goes down, Device B brings down GigabitEthernet 1/0/1.

Figure 3 Network diagram

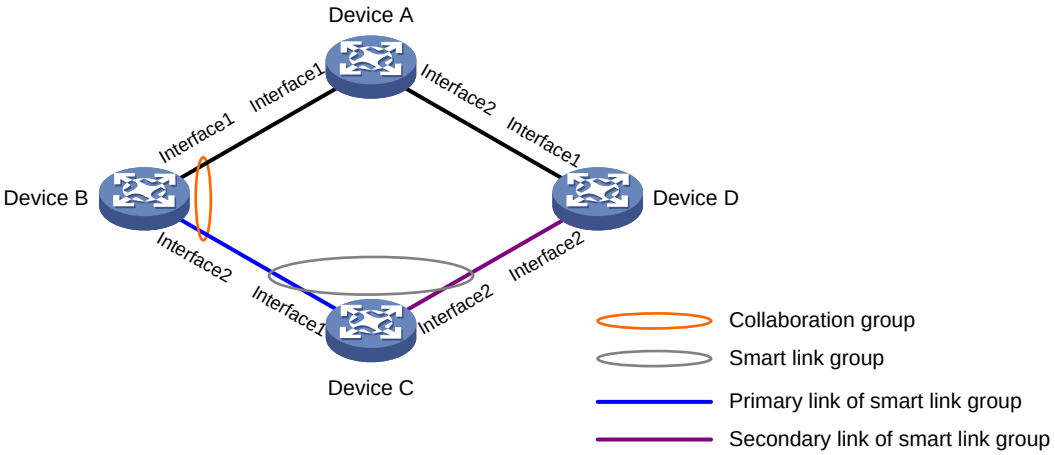


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure Smart Link on Device C:

Create a smart link group, and configure GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port for the smart link group. (Details not shown.)

For more information about configuring Smart Link, see "Configuring Smart Link."

2. Configure Device B:

Enable Monitor Link globally. By default, Monitor Link is enabled globally.

```
<DeviceB> system-view
[DeviceB] undo monitor-link disable
```

Create collaboration group 1.

```
[DeviceB] collaboration-group 1
```

Assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to collaboration group 1.

```
[DeviceB-collaboration-group1] port gigabitethernet 1/0/1
[DeviceB-collaboration-group1] port gigabitethernet 1/0/2
```

Verifying the configuration

1. Verify that the state of the uplink interface affects the state of the downlink interface.

Display information about smart link group 1 on Device C.

```
[DeviceC] display smart-link group 1
Smart link group 1 information:
  Device ID       : 2007-6f03-0700
  Preemption mode : None
  Preemption delay: 1(s)
  Control VLAN    : 1
  Protected VLAN  : Reference Instance 1

  Member                  Role      State   Flush-count     Last-flush-time
  ------------------------------------------------------------------------------
  GE1/0/1                 PRIMARY   ACTIVE  2               20:19:28 2017/04/05
  GE1/0/2                 SECONDARY STANDBY 0               NA
```

The output shows that the primary port of smart link group 1 is GigabitEthernet 1/0/1, and the traffic is sent out of GigabitEthernet 1/0/1.

Shut down GigabitEthernet 1/0/1 on Device B, and display information about collaboration group 1.

```
[DeviceB] display collaboration-group 1 verbose
Collaboration group protocol status: Enabled
Collaboration group 1 information:
  Group status     : DOWN
  Member up delay  : 0 seconds
  Last up time     : 20:19:58 2017/04/05
  Last down time   : 20:27:02 2017/04/05
  Member                    Status
  GE1/0/1                   DOWN
  GE1/0/2                   Collaboration-down
```

The output shows that collaboration group 1 is in DOWN state, GigabitEthernet 1/0/1 is in DOWN state, and GigabitEthernet 1/0/2 is in Collaboration-down state.

Display information about smart link group 1 on Device C.

```
[DeviceC] display smart-link group 1
Smart link group 1 information:
  Device ID       : 2007-6f03-0700
  Preemption mode : None
```

```
  Preemption delay: 1(s)
  Control VLAN    : 1
  Protected VLAN  : Reference Instance 1

  Member                  Role      State   Flush-count     Last-flush-time
  ------------------------------------------------------------------------------
  GE1/0/1                 PRIMARY   DOWN    2               20:19:28 2017/04/05
  GE1/0/2                 SECONDARY ACTIVE  1               20:26:35 2017/04/05
```

The output shows that GigabitEthernet 1/0/2 becomes the primary port of smart link group 1, and the traffic is sent out of GigabitEthernet 1/0/2.

2. Verify that the state of the downlink interface affects the state of the uplink interface.

Display information about GigabitEthernet 1/0/1 on Device A.

```
[DeviceA] display interface gigabitethernet 1/0/1 brief
Brief information on interfaces in bridge mode:
Link: ADM - administratively down; Stby - standby
Speed: (a) - auto
Duplex: (a)/A - auto; H - half; F - full
Type: A - access; T - trunk; H - hybrid
Interface            Link Speed     Duplex Type PVID Description
GE1/0/1              UP   1G(a)     F(a)   A    1
```

The output shows that GigabitEthernet 1/0/1 is in UP state.

Shut down GigabitEthernet 1/0/2 on Device B, and display information about collaboration group 1.

```
[DeviceB] display collaboration-group 1 verbose
Collaboration group protocol status: Enabled
Collaboration group 1 information:
  Group status     : DOWN
  Member up delay  : 0 seconds
  Last up time     : 20:19:58 2017/04/05
  Last down time   : 20:27:02 2017/04/05
  Member                    Status
  GE1/0/1                   Collaboration-down
  GE1/0/2                   DOWN
```

The output shows that collaboration group 1 is in DOWN state, GigabitEthernet 1/0/1 is in Collaboration-down state, and GigabitEthernet 1/0/2 is in DOWN state.

Display information about GigabitEthernet 1/0/1 on Device A.

```
[DeviceA] display interface gigabitethernet 1/0/1 brief
Brief information on interfaces in bridge mode:
Link: ADM - administratively down; Stby - standby
Speed: (a) - auto
Duplex: (a)/A - auto; H - half; F - full
Type: A - access; T - trunk; H - hybrid
Interface            Link Speed     Duplex Type PVID Description
GE1/0/1              DOWN auto      A      A    1
```

The output shows that GigabitEthernet 1/0/1 is in DOWN state.

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Network Management and Monitoring Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes network management and monitoring fundamentals and configuration procedures. It describes how to view system information, collect traffic statistics, assess the network performance, synchronize time for all devices with clocks in your network, and use the ping, tracert, and debug commands to check and debug the current network connectivity.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.

>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Using ping, tracert, and system debugging 1

Ping 1
About ping 1
Using a ping command to test network connectivity 1
Example: Using the ping utility 2
Configuring proxy probe 3
Tracert 3
About tracert 3
Prerequisites 4
Using a tracert command to identify failed or all nodes in a path 5
Example: Using the tracert utility 5
System debugging 6
About system debugging 6
Configuring debugging 7
Display and maintenance commands for debugging 7

Using ping, tracert, and system debugging

This chapter covers ping, tracert, and information about debugging the system.

Ping

About ping

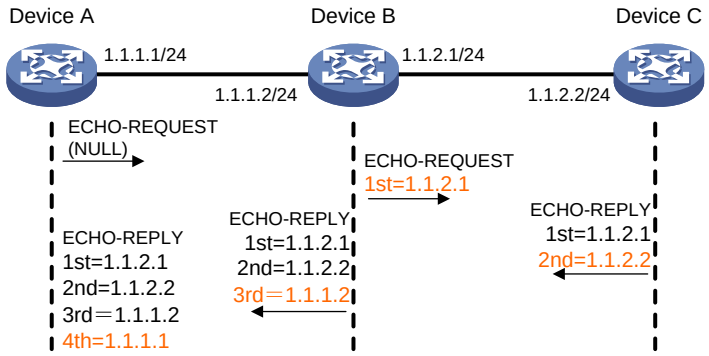
Use the ping utility to determine if an address is reachable.

Ping sends ICMP echo requests (ECHO-REQUEST) to the destination device. Upon receiving the requests, the destination device responds with ICMP echo replies (ECHO-REPLY) to the source device. The source device outputs statistics about the ping operation, including the number of packets sent, number of echo replies received, and the round-trip time. You can measure the network performance by analyzing these statistics.

You can use the **`ping -r`** command to display the routers through which ICMP echo requests have passed. The test procedure of **`ping -r`** is as shown in Figure 1:

1. The source device (Device A) sends an ICMP echo request to the destination device (Device C) with the RR option empty.
2. The intermediate device (Device B) adds the IP address of its outbound interface (1.1.2.1) to the RR option of the ICMP echo request, and forwards the packet.
3. Upon receiving the request, the destination device copies the RR option in the request and adds the IP address of its outbound interface (1.1.2.2) to the RR option. Then the destination device sends an ICMP echo reply.
4. The intermediate device adds the IP address of its outbound interface (1.1.1.2) to the RR option in the ICMP echo reply, and then forwards the reply.
5. Upon receiving the reply, the source device adds the IP address of its inbound interface (1.1.1.1) to the RR option. The detailed information of routes from Device A to Device C is formatted as: 1.1.1.1 <-> {1.1.1.2; 1.1.2.1} <-> 1.1.2.2.

Figure 1 Ping operation



Using a ping command to test network connectivity

Perform the following tasks in any view:

- Determine if an IPv4 address is reachable.

ping [**ip**] [**-a** *source-ip* | **-c** *count* | **-f** | **-h** *ttl* | **-i** *interface-type interface-number* | **-m** *interval* | **-n** | **-no-log** | **-p** *pad* | **-q** | **-r** | **-s** *packet-size* | **-t** *timeout* | **-tos** *tos* | **-v** | **-vpn-instance** *vpn-instance-name*] * *host*

Increase the timeout time (indicated by the **-t** keyword) on a low-speed network.

- Determine if an IPv6 address is reachable.

  ping ipv6 [**-a** *source-ipv6* | **-c** *count* | **-i** *interface-type interface-number* | **-m** *interval* | **-no-log** | **-q** | **-s** *packet-size* | **-t** *timeout* | **-tc** *traffic-class* | **-v** | **-vpn-instance** *vpn-instance-name*] * *host*

  Increase the timeout time (indicated by the **-t** keyword) on a low-speed network.

- Determine if a node in an MPLS network is reachable.

  ping mpls ipv4

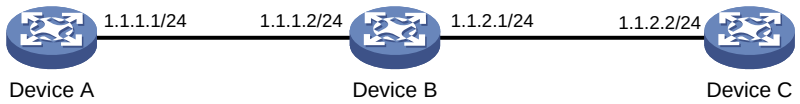
  For more information about this command, see *MPLS Command Reference*.

Example: Using the ping utility

Network configuration

As shown in Figure 2, determine if Device A and Device C can reach each other.

Figure 2 Network diagram



Procedure

Test the connectivity between Device A and Device C.

```
<DeviceA> ping 1.1.2.2
Ping 1.1.2.2 (1.1.2.2): 56 data bytes, press CTRL+C to break
56 bytes from 1.1.2.2: icmp_seq=0 ttl=254 time=2.137 ms
56 bytes from 1.1.2.2: icmp_seq=1 ttl=254 time=2.051 ms
56 bytes from 1.1.2.2: icmp_seq=2 ttl=254 time=1.996 ms
56 bytes from 1.1.2.2: icmp_seq=3 ttl=254 time=1.963 ms
56 bytes from 1.1.2.2: icmp_seq=4 ttl=254 time=1.991 ms

--- Ping statistics for 1.1.2.2 ---
5 packet(s) transmitted, 5 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.963/2.028/2.137/0.062 ms
```

The output shows the following information:

- Device A sends five ICMP packets to Device C and Device A receives five ICMP packets.
- No ICMP packet is lost.
- The route is reachable.

Configuring proxy probe

About this task

Proxy probe is a network reachability diagnostic tool. It test reachability from the local device to a directly connected peer interface through a proxy node. It is applicable to scenarios where the local device cannot communicate directly with the target interface, for example, in cross-protocol scenarios such as probing from an IPv4 device to an IPv6 device. The ping command can test connectivity between two nodes and requires the source to have a route to the destination for the probe to succeed.

As per RFC 8335, this feature allows indirect probe by using ICMP extension messages and a local protocol table as follows:

1. With the **probe** command configured, the probe node sends an ICMP Extended Echo Request to a proxy node.
2. Upon receiving the probe packet, the proxy node does not forward it directly to the probed node. Instead, it checks local ARP or ND entries to determine reachability of the probed node and then replies with an ICMP Extended Echo Reply.
3. Upon receiving the reply from the proxy node, the probe node obtains the status of the probed interface.

Restrictions and guidelines

- If you specify a proxy node by its host name, you must first configure domain name system (DNS) on the device.
- For the proxy node to retrieve target node status through local ARP or ND entries, make sure the probe device, proxy device, and target device have Layer 3 connectivity.
- You can press Ctrl+C to terminate the probe. Use this key combination to terminate long-running or large-scale probing tasks.

Prerequisites

- Verify that the the proxy node and the target node has Layer 3 connectivity.
- If you specify a proxy node by its host name, configure the DNS server address in advance.

Procedure

Perform the following tasks in any view.

- Use an IPv4 proxy node to probe an IPv4 or IPv6 node.

  probe *proxy-host* **by-remote-address** [**ipv6**] *remote-ip* [**count** *count*]
- Use an IPv6 proxy node to probe an IPv4 or IPv6 node.

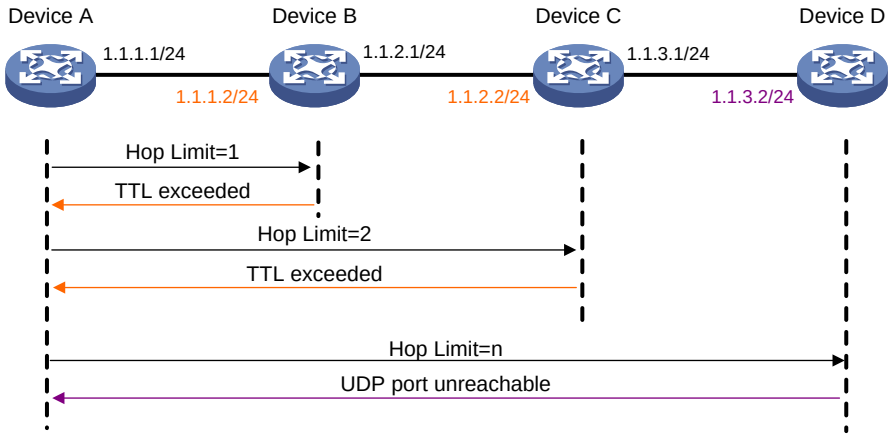
  probe ipv6 *proxy-host* **by-remote-address** [**ipv6**] *remote-ip* [**count** *count*]

Tracert

About tracert

Tracert (also called Traceroute) enables retrieval of the IP addresses of Layer 3 devices in the path to a destination. In the event of network failure, use tracert to test network connectivity and identify failed nodes.

Figure 3 Tracert operation



Tracert uses received ICMP error messages to get the IP addresses of devices. Tracert works as shown in Figure 3:

1. The source device sends a UDP packet with a TTL value of 1 to the destination device. The destination UDP port is not used by any application on the destination device.
2. The first hop (Device B, the first Layer 3 device that receives the packet) responds by sending a TTL-expired ICMP error message to the source, with its IP address (1.1.1.2) encapsulated. This way, the source device can get the address of the first Layer 3 device (1.1.1.2).
3. The source device sends a packet with a TTL value of 2 to the destination device.
4. The second hop (Device C) responds with a TTL-expired ICMP error message, which gives the source device the address of the second Layer 3 device (1.1.2.2).
5. This process continues until a packet sent by the source device reaches the ultimate destination device. Because no application uses the destination port specified in the packet, the destination device responds with a port-unreachable ICMP message to the source device, with its IP address encapsulated. This way, the source device gets the IP address of the destination device (1.1.3.2).
6. The source device determines that:
   - The packet has reached the destination device after receiving the port-unreachable ICMP message.
   - The path to the destination device is 1.1.1.2 to 1.1.2.2 to 1.1.3.2.

Prerequisites

Before you use a tracert command, perform the tasks in this section.

For an IPv4 network:

- Enable sending of ICMP timeout packets on the intermediate devices (devices between the source and destination devices). If the intermediate devices are H3C devices, execute the **ip ttl-expires enable** command on the devices. For more information about this command, see *Layer 3—IP Services Command Reference*.
- Enable sending of ICMP destination unreachable packets on the destination device. If the destination device is an H3C device, execute the **ip unreachables enable** command. For more information about this command, see *Layer 3—IP Services Command Reference*.

For an IPv6 network:

- Enable sending of ICMPv6 timeout packets on the intermediate devices (devices between the source and destination devices). If the intermediate devices are H3C devices, execute the

ipv6 hoplimit-expires enable command on the devices. For more information about this command, see *Layer 3—IP Services Command Reference*.

- Enable sending of ICMPv6 destination unreachable packets on the destination device. If the destination device is an H3C device, execute the **ipv6 unreachables enable** command. For more information about this command, see *Layer 3—IP Services Command Reference*.

Using a tracert command to identify failed or all nodes in a path

Perform the following tasks in any view:

- Trace the route to an IPv4 destination.

  tracert [**-a** *source-ip* | **-e** | **-f** *first-ttl* | **-m** *max-ttl* | **-p** *port* | **-q** *packet-number* | **-t** *tos* | **-vpn-instance** *vpn-instance-name* [**-resolve-as** { **global** | **none** | **vpn** }] | **-w** *timeout*] * *host*

- Trace the route to an IPv6 destination.

  tracert ipv6 [**-a** *source-ipv6* | **-e** | **-f** *first-hop* | **-m** *max-hops* | **-p** *port* | **-q** *packet-number* | **-t** *traffic-class* | **-vpn-instance** *vpn-instance-name* [**-resolve-as** { **global** | **none** | **vpn** }] | **-w** *timeout*] * *host*

- Trace the route to a destination in an MPLS network.

  tracert mpls ipv4

  For more information about this command, see *MPLS Command Reference*.

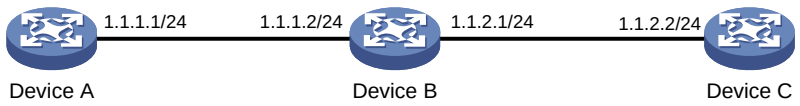
Example: Using the tracert utility

Network configuration

As shown in Figure 4, Device A failed to Telnet to Device C.

Test the network connectivity between Device A and Device C. If they cannot reach each other, locate the failed nodes in the network.

Figure 4 Network diagram



Procedure

1. Configure IP addresses for the devices as shown in Figure 4.
2. Configure a static route on Device A.

   ```
   <DeviceA> system-view
   [DeviceA] ip route-static 0.0.0.0 0.0.0.0 1.1.1.2
   ```

3. Test connectivity between Device A and Device C.

   ```
   [DeviceA] ping 1.1.2.2
   Ping 1.1.2.2(1.1.2.2): 56 -data bytes, press CTRL+C to break
   Request time out
   Request time out
   Request time out
   Request time out
   Request time out
   ```

```
--- Ping statistics for 1.1.2.2 ---
5 packet(s) transmitted,0 packet(s) received,100.0% packet loss
```

The output shows that Device A and Device C cannot reach each other.

4. Identify failed nodes:

Enable sending of ICMP timeout packets on Device B.

```
<DeviceB> system-view
[DeviceB] ip ttl-expires enable
```

Enable sending of ICMP destination unreachable packets on Device C.

```
<DeviceC> system-view
[DeviceC] ip unreachables enable
```

Identify failed nodes.

```
[DeviceA] tracert 1.1.2.2
traceroute to 1.1.2.2 (1.1.2.2) 30 hops at most,40 bytes each packet, press CTRL+C
to break
 1  1.1.1.2 (1.1.1.2) 1 ms 2 ms 1 ms
 2  * * *
 3  * * *
 4  * * *
 5
[DeviceA]
```

The output shows that Device A can reach Device B but cannot reach Device C. An error has occurred on the connection between Device B and Device C.

5. To identify the cause of the issue, execute the following commands on Device A and Device C:
   - Execute the **`debugging ip icmp`** command and verify that Device A and Device C can send and receive the correct ICMP packets.
   - Execute the **`display ip routing-table`** command to verify that Device A and Device C have a route to each other.

System debugging

About system debugging

The device supports debugging for the majority of protocols and features, and provides debugging information to help users diagnose errors.

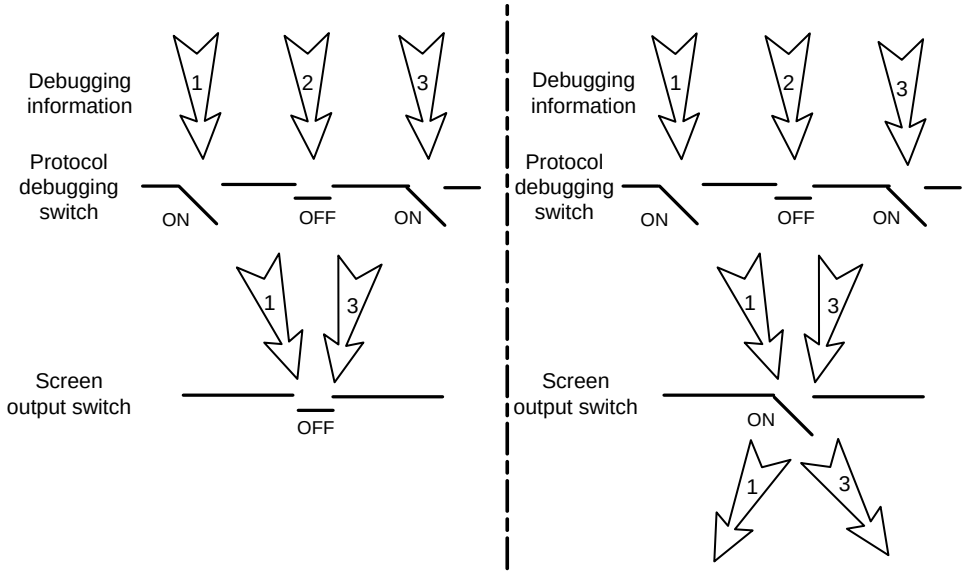
The following switches control the display of debugging information:

- **Module debugging switch**—Controls whether to generate the module-specific debugging information.
- **Screen output switch**—Controls whether to display the debugging information on a certain screen. Use **`terminal monitor`** and **`terminal logging level`** commands to turn on the screen output switch. For more information about these two commands, see *Network Management and Monitoring Command Reference*.

As shown in Figure 5, the device can provide debugging for the three modules 1, 2, and 3. The debugging information can be output on a terminal only when both the module debugging switch and the screen output switch are turned on.

Debugging information is typically displayed on a console. You can also send debugging information to other destinations. For more information, see "Configuring the information center."

Figure 5 Relationship between the module and screen output switch



Configuring debugging

Restrictions and guidelines

CAUTION:

Output of excessive debugging messages increases the CPU usage and downgrades the system performance. To guarantee system performance, enable debugging only for modules that are in an exceptional condition.

Enable debugging for modules for troubleshooting purposes. When debugging is complete, use the **undo debugging all** command to disable all the debugging functions.

Procedure

1. Enable debugging for a module.

   debugging *module-name* [*option*]

   By default, debugging is disabled for all modules.

   This command is available in user view.

Display and maintenance commands for debugging

Execute **display** commands in any view.

Task	Command
Display the enabled debugging features.	**display debugging** [*module-name*]

Contents

Configuring NQA 1

About NQA 1
NQA architecture 1
NQA operating mechanism 1
Collaboration with Track 2
Threshold monitoring 2
NQA templates 3
Restrictions and guidelines: NQA configuration 3
NQA tasks at a glance 4
Configuring the NQA server 4
Enabling the NQA client 5
Configuring NQA operations on the NQA client 5
NQA operations tasks at a glance 5
Configuring the ICMP echo operation 5
Configuring the ICMP jitter operation 7
Configuring the DHCP operation 8
Configuring the DNS operation 9
Configuring the FTP operation 10
Configuring the HTTP operation 11
Configuring the UDP jitter operation 13
Configuring the SNMP operation 16
Configuring the TCP operation 18
Configuring the UDP echo operation 19
Configuring the UDP tracert operation 20
Configuring the voice operation 23
Configuring the DLSw operation 25
Configuring the path jitter operation 25
Configuring optional parameters for the NQA operation 27
Configuring the collaboration feature 28
Configuring threshold monitoring 29
Configuring the NQA statistics collection feature 31
Configuring the saving of NQA history records 32
Scheduling the NQA operation on the NQA client 32
Configuring NQA templates on the NQA client 33
Restrictions and guidelines 33
NQA template tasks at a glance 33
Configuring the ICMP template 33
Configuring the DNS template 35
Configuring the TCP template 36
Configuring the TCP half open template 37
Configuring the UDP template 38
Configuring the HTTP template 40
Configuring the HTTPS template 41
Configuring the FTP template 43
Configuring the RADIUS template 44
Configuring the SSL template 46
Configuring optional parameters for the NQA template 46
Display and maintenance commands for NQA 47
NQA configuration examples 48
Example: Configuring the ICMP echo operation 48
Example: Configuring the ICMP jitter operation 50
Example: Configuring the DHCP operation 52
Example: Configuring the DNS operation 53
Example: Configuring the FTP operation 54
Example: Configuring the HTTP operation 55
Example: Configuring the UDP jitter operation 57
Example: Configuring the SNMP operation 59

Example: Configuring the TCP operation 60
Example: Configuring the UDP echo operation 62
Example: Configuring the UDP tracert operation 63
Example: Configuring the voice operation 64
Example: Configuring the DLSw operation 67
Example: Configuring the path jitter operation 68
Example: Configuring NQA collaboration 70
Example: Configuring the ICMP template 72
Example: Configuring the DNS template 73
Example: Configuring the TCP template 74
Example: Configuring the TCP half open template 75
Example: Configuring the UDP template 76
Example: Configuring the HTTP template 76
Example: Configuring the HTTPS template 77
Example: Configuring the FTP template 78
Example: Configuring the RADIUS template 79
Example: Configuring the SSL template 80

Configuring NQA

About NQA

Network quality analyzer (NQA) allows you to measure network performance, verify the service levels for IP services and applications, and troubleshoot network problems.

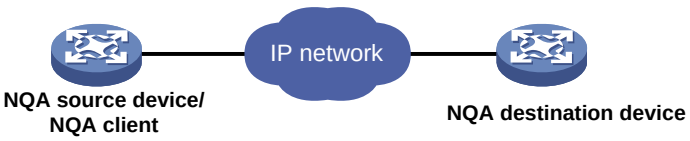
NQA architecture

An NQA operation contains a set of parameters such as the operation type, destination IP address, and port number to define how the operation is performed. Each NQA operation is identified by the combination of the administrator name and the operation tag. You can configure the NQA client to run the operations at scheduled time periods.

As shown in Figure 1, the NQA source device (NQA client) sends data to the NQA destination device by simulating IP services and applications to measure network performance.

All types of NQA operations require the NQA client, but only the TCP, UDP echo, UDP jitter, and voice operations require the NQA server. The NQA operations for services that are already provided by the destination device such as FTP do not need the NQA server. You can configure the NQA server to listen and respond to specific IP addresses and ports to meet various test needs.

Figure 1 Network diagram



NQA operating mechanism

The NQA client and NQA server interact as follows:

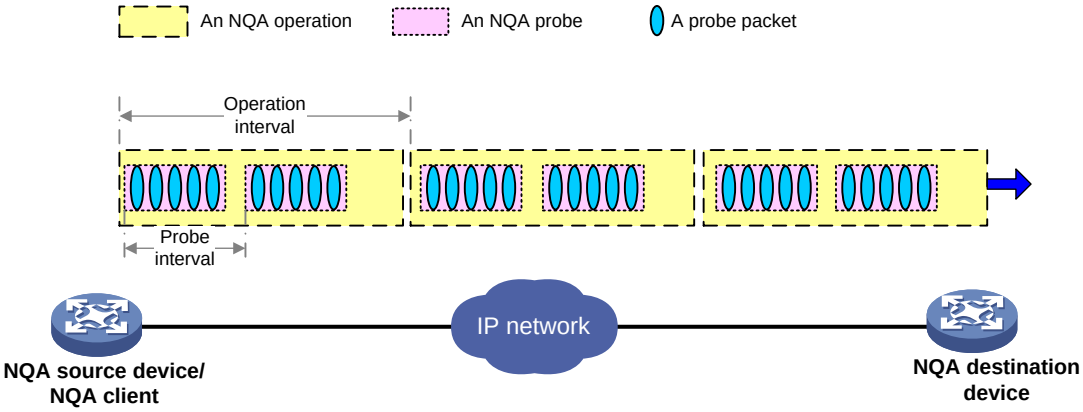
1. The NQA client constructs probe packets of the operation type, and sends them to the NQA server.
2. The NQA server processes the probe packets and sends back reply packets.
3. Upon receiving the replies, the NQA client calculates the packet loss ratio and round-trip time to determine the network performance and service quality.

After an NQA operation starts, the NQA client repeats the operation at the specified interval.

An NQA operation can contain multiple probes. You can set the number of probes the NQA client performs in an operation.

For the voice and path jitter operations, the NQA client performs only one probe per operation.

Figure 2 NQA operation diagram



Collaboration with Track

NQA can collaborate with the Track module to notify application modules of state or performance changes so that the application modules can take predefined actions.

- VRRP
- Static routing.
- Policy-based routing.
- Traffic redirecting.

The following describes how a static route destined for 192.168.0.100 is monitored through collaboration:

1. NQA monitors the reachability to 192.168.0.100.
2. When 192.168.0.100 becomes unreachable, NQA notifies the Track module of the change.
3. The Track module notifies the static routing module of the state change.
4. The static routing module sets the static route to invalid according to a predefined action.

For more information about collaboration, see *High Availability Configuration Guide*.

Threshold monitoring

Threshold monitoring enables the NQA client to take a predefined action when the NQA operation performance metrics violate the specified thresholds.

Table 1 describes the relationships between performance metrics and NQA operation types.

Table 1 Performance metrics and NQA operation types

Performance metric	NQA operation types that can gather the metric
Probe duration	All NQA operation types except UDP jitter, UDP tracert, path jitter, and voice
Number of probe failures	All NQA operation types except UDP jitter, UDP tracert, path jitter, and voice
Round-trip time	ICMP jitter, UDP jitter, and voice
Number of discarded packets	ICMP jitter, UDP jitter, and voice
One-way jitter (source-to-destination or	ICMP jitter, UDP jitter, and voice

Performance metric	NQA operation types that can gather the metric
destination-to-source)	
One-way delay (source-to-destination or destination-to-source)	ICMP jitter, UDP jitter, and voice
Calculated Planning Impairment Factor (ICPIF) (see "Configuring the voice operation")	Voice
Mean Opinion Scores (MOS) (see "Configuring the voice operation")	Voice

NQA templates

An NQA template is a set of parameters (such as destination address and port number) that defines how an NQA operation is performed. Features can use the NQA template to collect statistics.

You can create multiple NQA templates on the NQA client. Each template must be identified by a unique template name.

Restrictions and guidelines: NQA configuration

An NQA operation depends on a series of conditions such as routing table, FIB table, ARP table, and probe packet output interface state. During the initial startup of the device, the NQA operation might fail when it does not meet the previous conditions. If the NQA operation is associated with a track entry, the track entry state might also change, resulting in collaboration with the service function associated with the track entry.

To avoid probe failures, follow these restrictions and guidelines when configuring the listening ports on the NQA client and NQA server:

- Do not specify a well-known port.
- Make sure the specified port number is not used by any services on the device.
  - To obtain the IPv4 addresses and the port numbers in use on this device, see the **Local Addr:port** field in the output from the **`display tcp`** and **`display udp`** commands.
  - To obtain the IPv6 addresses and the port numbers in use on this device, see the **LAddr->port** field in the output from the **`display ipv6 tcp`** and **`display ipv6 udp`** commands.

If the source IP address of NQA probe packets has been NATed through the Easy IP feature, the local device might fail to receive NQA response packets, resulting in NQA operation failure. As a best practice, use the **`source interface`**, **`source ip`**, or **`source ipv6`** command to specify the source IP address of NQA probe packets to prevent the source IP address from being translated by NAT. In the output from the **`display nat all`** command, if the **NAT address group information** field contains the output interface IP address of NQA probe packets, or the **NAT outbound information** field contains the output interface of NQA probe packets, the source IP address of NQA probe packets will be NATed by the Easy IP feature. Avoid specifying the IP address and interface when configuring the source IP address. For more information about Easy IP, see NAT configuration in *Layer 3—IP Services Configuration Guide*.

The destination port configured for the operation (with the **`destination port`** command) on the NQA client must be the same as the listening port configured on the server.

An NQA operation depends on a series of conditions such as routing table, FIB table, ARP table, and probe packet output interface state. During the initial startup of the device, the NQA operation might fail when it does not meet the previous conditions.

For latency operation results to be accurate, synchronize time between the source and the destination with an accuracy higher than or equal to the operation result calculation accuracy.

NQA tasks at a glance

To configure NQA, perform the following tasks:

1. Configuring the NQA server

   Perform this task on the destination device before you configure the TCP, UDP echo, UDP jitter, and voice operations.
2. Enabling the NQA client
3. Configuring NQA operations or NQA templates

   Choose the following tasks as needed:

   - Configuring NQA operations on the NQA client
   - Configuring NQA templates on the NQA client

   After you configure an NQA operation, you can schedule the NQA client to run the NQA operation.

   An NQA template does not run immediately after it is configured. The template creates and run the NQA operation only when it is required by the feature (such as load balancing) to which the template is applied.

Configuring the NQA server

Restrictions and guidelines

To perform TCP, UDP echo, UDP jitter, and voice operations, you must configure the NQA server on the destination device. The NQA server listens and responds to requests on the specified IP addresses and ports.

You can configure multiple TCP or UDP listening services on an NQA server, where each corresponds to a specific IP address and port number.

The IP address, port number, and VPN instance for a listening service must be unique on the NQA server and match the configuration on the NQA client.

Procedure

1. Enter system view.

   system-view
2. Enable the NQA server.

   nqa server enable

   By default, the NQA server is disabled.
3. Configure a TCP listening service.

   nqa server tcp-connect *ip-address port-number* [**vpn-instance** *vpn-instance-name*] [**tos** *tos*]

   This task is required for only TCP and DLSw operations. For the DLSw operation, the port number for the TCP listening service must be 2065.
4. Configure a UDP listening service.

   nqa server udp-echo *ip-address port-number* [**vpn-instance** *vpn-instance-name*] [**tos** *tos*]

   This task is required for only UDP echo, UDP jitter, and voice operations.

Enabling the NQA client

1. Enter system view.
   `system-view`
2. Enable the NQA client.
   `nqa agent enable`
   By default, the NQA client is enabled.
   The NQA client configuration takes effect after you enable the NQA client.

Configuring NQA operations on the NQA client

NQA operations tasks at a glance

To configure NQA operations, perform the following tasks:

1. Configuring an NQA operation
   - Configuring the ICMP echo operation
   - Configuring the ICMP jitter operation
   - Configuring the DHCP operation
   - Configuring the DNS operation
   - Configuring the FTP operation
   - Configuring the HTTP operation
   - Configuring the UDP jitter operation
   - Configuring the SNMP operation
   - Configuring the TCP operation
   - Configuring the UDP echo operation
   - Configuring the UDP tracert operation
   - Configuring the voice operation
   - Configuring the DLSw operation
   - Configuring the path jitter operation
2. (Optional.) Configuring optional parameters for the NQA operation
3. (Optional.) Configuring the collaboration feature
4. (Optional.) Configuring threshold monitoring
5. (Optional.) Configuring the NQA statistics collection feature
6. (Optional.) Configuring the saving of NQA history records
7. Scheduling the NQA operation on the NQA client

Configuring the ICMP echo operation

About this task

ICMP echo is the most basic feature of NQA, and it is implemented based on RFC 2925. An ICMP echo operation is used to test the reachability of the destination, and calculate the response time and packet loss rate by sending ICMP packets.

An ICMP echo operation is similar to ping in terms of functionality. They have the following differences:

- The TTL value of ICMP packets sent in an ICMP echo operation is 20 by default. In a complex network with multiple hops, you can change the TTL value based on the actual network requirements.
- An ICMP echo operation supports various operation parameters. For example, you can specify a next hop address for the operation. When multiple paths exist between the source and destination devices, you can configure a next hop address to specify a path to be tested.

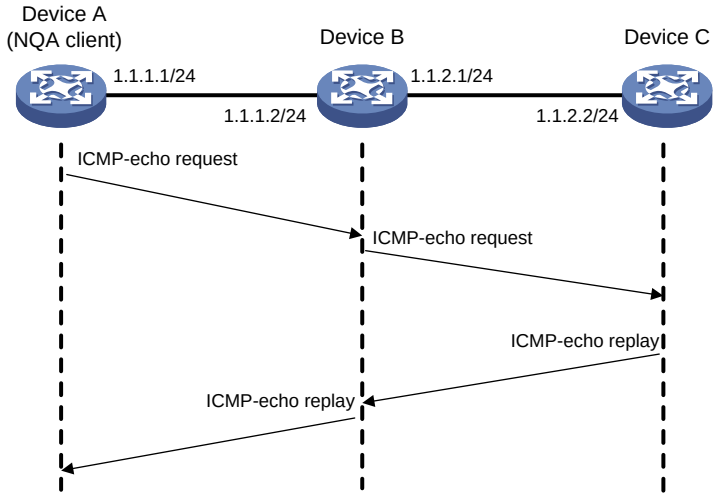
For an ICMP echo operation, a single probe operation refers to sending one probe packet to the destination.

As shown in Figure 3, an ICMP echo operation is performed as follows:

1. An NQA client sends an ICMP echo request to the destination IP address according to the configured probe time and frequency.
2. Upon receiving the ICMP echo request, the destination IP address sends an ICMP echo reply.
3. The NQA client calculates the response time of the destination IP address and the packet loss rate based on the receiving time packet count of the ICMP-echo reply, which reflects the network performance and status.

An ICMP echo operation can succeed only if the destination device can correctly respond to ICMP echo requests.

Figure 3 ICMP echo operation diagram



Procedure

1. Enter system view.
   system-view
2. Create an NQA operation and enter NQA operation view.
   nqa entry *admin-name operation-tag*
3. Specify the ICMP echo type and enter its view.
   type icmp-echo
4. Specify the destination IP address for ICMP echo requests.
   IPv4:
   destination ip *ip-address*
   IPv6:
   destination ipv6 *ipv6-address*
   By default, no destination IP address is specified.
5. Specify the source IP address for ICMP echo requests. Choose one option as needed:

- Use the IP address of the specified interface as the source IP address.

  source interface *interface-type interface-number*

  By default, the source IP address of ICMP echo requests is the primary IP address of their output interface.

  The specified source interface must be up.
- Specify the source IPv4 address.

  source ip *ip-address*

  By default, the source IPv4 address of ICMP echo requests is the primary IPv4 address of their output interface.

  The specified source IPv4 address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.
- Specify the source IPv6 address.

  source ipv6 *ipv6-address*

  By default, the source IPv6 address of ICMP echo requests is the primary IPv6 address of their output interface.

  The specified source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

6. Specify the output interface or the next hop IP address for ICMP echo requests. Choose one option as needed: (If you configure both the output interface and next hop IP address, only the next hop IP address takes effect. To avoid probe failure, make sure the IP address belongs to a directly connected next hop.)
   - Specify the output interface for ICMP echo requests.

     out interface *interface-type interface-number*

     By default, the output interface for ICMP echo requests is not specified. The NQA client determines the output interface based on the routing table lookup.
   - Specify the next hop IPv4 address.

     next-hop ip *ip-address*

     By default, no next hop IPv4 address is specified.
   - Specify the next hop IPv6 address.

     next-hop ipv6 *ipv6-address*

     By default, no next hop IPv6 address is specified.
7. (Optional.) Set the payload size for each ICMP echo request.

   data-size *size*

   The default payload size is 100 bytes.
8. (Optional.) Specify the payload fill string for ICMP echo requests.

   data-fill *string*

   The default payload fill string is 00010203040506070809.

Configuring the ICMP jitter operation

About this task

The ICMP jitter operation measures unidirectional and bidirectional jitters. The operation result helps you to determine whether the network can carry jitter-sensitive services such as real-time voice and video services.

The ICMP jitter operation works as follows:

1. The NQA client sends ICMP packets to the destination device.

2. The destination device time stamps each packet it receives, and then sends the packet back to the NQA client.
3. Upon receiving the responses, the NQA client calculates the jitter according to the timestamps.

The ICMP jitter operation sends a number of ICMP packets to the destination device per probe. The number of packets to send is determined by using the **probe packet-number** command.

Restrictions and guidelines

The **display nqa history** command does not display the results or statistics of the ICMP jitter operation. To view the results or statistics of the operation, use the **display nqa result** or **display nqa statistics** command.

Before starting the operation, make sure the network devices are time synchronized by using NTP. For more information about NTP, see "Configuring NTP."

Procedure

1. Enter system view.
   system-view
2. Create an NQA operation and enter NQA operation view.
   nqa entry *admin-name operation-tag*
3. Specify the ICMP jitter type and enter its view.
   type icmp-jitter
4. Specify the destination IP address for ICMP packets.
   destination ip *ip-address*
   By default, no destination IP address is specified.
5. Set the number of ICMP packets sent per probe.
   probe packet-number *number*
   The default setting is 10.
6. Set the interval for sending ICMP packets.
   probe packet-interval *interval*
   The default setting is 20 milliseconds.
7. Specify how long the NQA client waits for a response from the server before it regards the response times out.
   probe packet-timeout *timeout*
   The default setting is 3000 milliseconds.
8. Specify the source IP address for ICMP packets.
   source ip *ip-address*
   By default, the source IP address of ICMP packets is the primary IP address of their output interface.
   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no ICMP packets can be sent out.

Configuring the DHCP operation

About this task

The DHCP operation measures whether or not the DHCP server can respond to client requests. DHCP also measures the amount of time it takes the NQA client to obtain an IP address from a DHCP server.

The NQA client simulates the DHCP relay agent to forward DHCP requests for IP address acquisition from the DHCP server. The interface that performs the DHCP operation does not change its IP address. When the DHCP operation completes, the NQA client sends a packet to release the obtained IP address.

The DHCP operation acquires an IP address from the DHCP server per probe.

Procedure

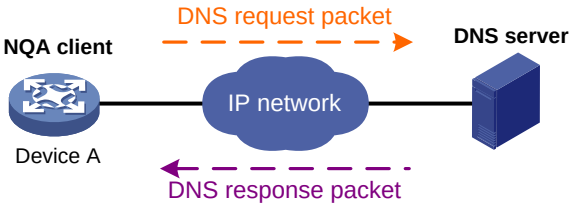
1. Enter system view.
   system-view
2. Create an NQA operation and enter NQA operation view.
   nqa entry *admin-name operation-tag*
3. Specify the DHCP type and enter its view.
   type dhcp
4. Specify the IP address of the DHCP server as the destination IP address of DHCP packets.
   destination ip *ip-address*
   By default, no destination IP address is specified.
5. Specify the output interface for DHCP request packets.
   out interface *interface-type interface-number*
   By default, the NQA client determines the output interface based on the routing table lookup.
6. Specify the source IP address of DHCP request packets.
   source ip *ip-address*
   By default, the source IP address of DHCP request packets is the primary IP address of their output interface.
   The specified source IP address must be the IP address of a local interface, and the local interface must be up. Otherwise, no probe packets can be sent out.

Configuring the DNS operation

About this task

The DNS operation measures the time for the NQA client to resolve a domain name into an IP address through a DNS server.

Figure 4 DNS operation diagram



The DNS operation simulates domain name resolution, and the obtained DNS entry is not saved.

The DNS operation resolves a domain name into an IP address per probe.

Procedure

1. Enter system view.
   system-view
2. Create an NQA operation and enter NQA operation view.
   nqa entry *admin-name operation-tag*

3. Specify the DNS type and enter its view.
   type dns
4. Specify the IP address of the DNS server as the destination IP address of DNS packets.
   destination ip *ip-address*
   By default, no destination IP address is specified.
5. Specify the domain name to be translated.
   resolve-target *domain-name*
   By default, no domain name is specified.

Configuring the FTP operation

About this task

The FTP operation measures the time for the NQA client to transfer a file to or download a file from an FTP server.

The FTP operation uploads or downloads a file from an FTP server per probe.

An FTP operation supports the GET and PUT operations:

- A GET operation temporarily downloads a file to the local file system, calculates the time required to download the file, and immediately deletes the file after obtaining the data to automatically releasing the occupied memory. If the name of the downloaded file is the same as an existing file on the NQA client, the original file on the NQA client will be overwritten.
- A PUT operation uploads a file with a fixed size and content to the FTP server. You can configure the name of the uploaded file, and the file content is fixed data specified internally by the system. If the configured file name is the same as the name of an existing file on the FTP server, the original file on the FTP server will be overwritten.

Restrictions and guidelines

To upload (**put**) a file to the FTP server, use the **filename** command to specify the name of the file you want to upload. The file must exist on the NQA client.

To download (**get**) a file from the FTP server, include the name of the file you want to download in the **url** command. The file must exist on the FTP server. The NQA client does not save the file obtained from the FTP server.

Use a small file for the FTP operation. A big file might result in transfer failure because of timeout, or might affect other services because of the amount of network bandwidth it occupies.

Procedure

1. Enter system view.
   system-view
2. Create an NQA operation and enter NQA operation view.
   nqa entry *admin-name operation-tag*
3. Specify the FTP type and enter its view.
   type ftp
4. Specify an FTP login username.
   username *username*
   By default, no FTP login username is specified.
5. Specify an FTP login password.
   password { **cipher** | **simple** } *string*
   By default, no FTP login password is specified.

6. Specify the source IP address for FTP request packets.

   source ip *ip-address*

   By default, the source IP address of FTP request packets is the primary IP address of their output interface.

   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no FTP requests can be sent out.

7. Set the data transmission mode.

   mode { **active** | **passive** }

   The default mode is **active**.

8. Specify the FTP operation type.

   operation { **get** | **put** }

   The default FTP operation type is **get**.

9. Specify the destination URL for the FTP operation.

   url *url*

   By default, no destination URL is specified for an FTP operation.

   Enter the URL in one of the following formats:

   - ftp://*host*/*filename*.
   - ftp://*host:port*/*filename*.

   If the value for the *host* parameter is an IPv6 address, you must use a pair of square brackets ([]) to include the IPv6 address, for example, ftp://[2001::1]/test. The *filename* argument is required only for the **get** operation.

10. Specify the name of the file to be uploaded.

    filename *file-name*

    By default, no file is specified.

    This task is required only for the **put** operation.

    The configuration does not take effect for the **get** operation.

Configuring the HTTP operation

About this task

The HTTP operation measures the time for the NQA client to obtain responses from an HTTP server.

The HTTP operation supports the following operation types:

- **Get**—Retrieves data such as a Web page from the HTTP server.
- **Post**—Sends data to the HTTP server for processing.
- **Raw**—Sends a user-defined HTTP request to the HTTP server. You must manually configure the content of the HTTP request to be sent.

In an HTTP operation, the NQA client sends a GET or POST request to the HTTP server at the specified address, and then calculates the operation time after receiving a reply. This process is to establish a connection with the HTTP server. If the connection is successfully established, the operation succeeds.

The HTTP operation completes the operation of the specified type per probe.

Procedure

1. Enter system view.

   system-view

2. Create an NQA operation and enter NQA operation view.

nqa entry *admin-name operation-tag*

3. Specify the HTTP type and enter its view.

   type http

4. Specify the destination URL for the HTTP operation.

   url *url*

   By default, no destination URL is specified for an HTTP operation.

   Enter the URL in one of the following formats:

   - http://*host*/*resource*
   - http://*host:port/resource*

   If the value for the *host* parameter is an IPv6 address, you must use a pair of square brackets ([]) to include the IPv6 address, for example, http://[2001::1]/test.

5. Specify an HTTP login username.

   username *username*

   By default, no HTTP login username is specified.

6. Specify an HTTP login password.

   password { **cipher** | **simple** } *string*

   By default, no HTTP login password is specified.

7. Specify the HTTP version.

   version { **v1.0** | **v1.1** }

   By default, HTTP 1.0 is used.

8. Specify the HTTP operation type.

   operation { **get** | **post** | **raw** }

   The default HTTP operation type is **get**.

   If you set the operation type to **raw**, the client pads the content configured in raw request view to the HTTP request to send to the HTTP server.

9. Configure the HTTP raw request.

   a. Enter raw request view.

      raw-request

      Every time you enter raw request view, the previously configured raw request content is cleared.

   b. Enter or paste the request content.

      By default, no request content is configured.

      To ensure successful operations, make sure the request content does not contain command aliases configured by using the **alias** command. For more information about the **alias** command, see CLI commands in *Fundamentals Command Reference*.

   c. Save the input and return to HTTP operation view:

      quit

   This step is required only when the operation type is set to **raw**.

10. Specify the source IP address for the HTTP packets.

    source ip *ip-address*

    By default, the source IP address of HTTP packets is the primary IP address of their output interface.

    The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no request packets can be sent out.

Configuring the UDP jitter operation

About this task

UDP jitter overview

UDP jitter is an important tool to detect network status and monitor real-time service quality. Real-time services (such as voice and video services) have high delay and jitter requirements. You can use a UDP jitter operation to determine whether a network can guarantee service quality for real-time services.

UDP jitter probe packets are proprietary packets, requiring the peer device to be an H3C device with the NQA server feature enabled and associated NQA server parameters configured.

In a UDP jitter operation, a group of packets are sent in one probe, and this group of packets corresponds to one history record only. As a best practice to obtain the UDP jitter operation results, view the probe results, rather than the history records.

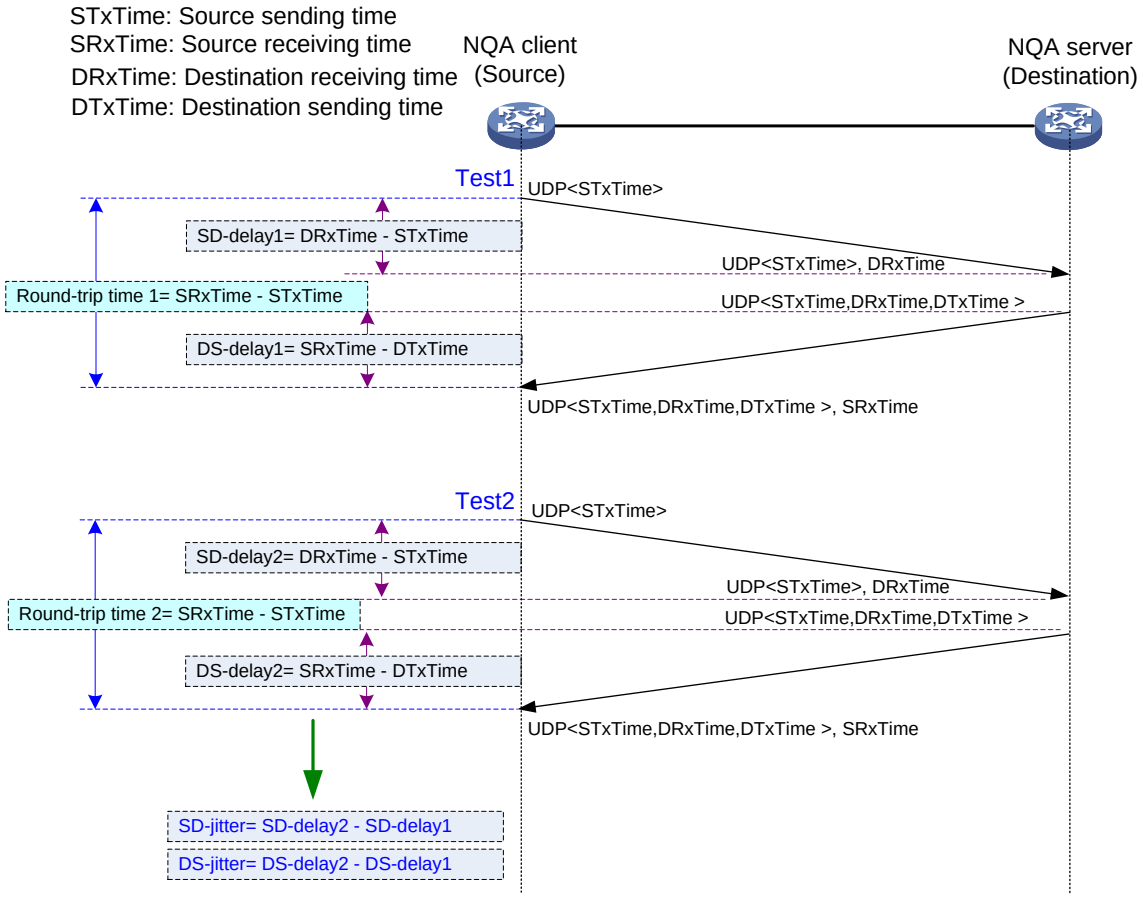
The UDP jitter operation results include one-way delay information. If this information is critical, make sure the time on both devices is synchronized to the millisecond level by using a time synchronization feature (such as PTP). If the information is not critical, time synchronization is not required.

UDP jitter operations

As shown in Figure 5, a UDP jitter operation is performed as follows:

1. The NQA client sends a UDP jitter packet to the NQA server, and records the packet leaving time STxTime.
2. When the UDP jitter packet arrives at the NQA server, the NQA server adds the packet receiving time DRxTime to the packet.
3. When the UDP jitter packet leaves the NQA server, the NQA server adds the packet leaving time DTxTime.
4. When NQA client receives the reply, it records the packet receiving time SRxTime.
5. The NQA client sends multiple probe packets at a fixed interval, and the above process repeats. By recording the timestamps, the device can calculate the delay jitter from the source to the destination, as well as the delay jitter from the destination back to the source.

Figure 5 UDP jitter operation diagram



UDP-jitter one-way packet loss statistics

As shown in Figure 6, the UDP jitter client and server can collaborate to collect the one-way packet loss statistics.

Each packet sent by the NQA client includes a packet ID, which increments by 1 each time a packet is sent. The NQA server updates the greatest packet ID and the number of received packets every time it receives a packet, and sends them back to the NQA client in the reply. The NQA client then records the number of replies, and obtains the information of the NQA sever from these replies.

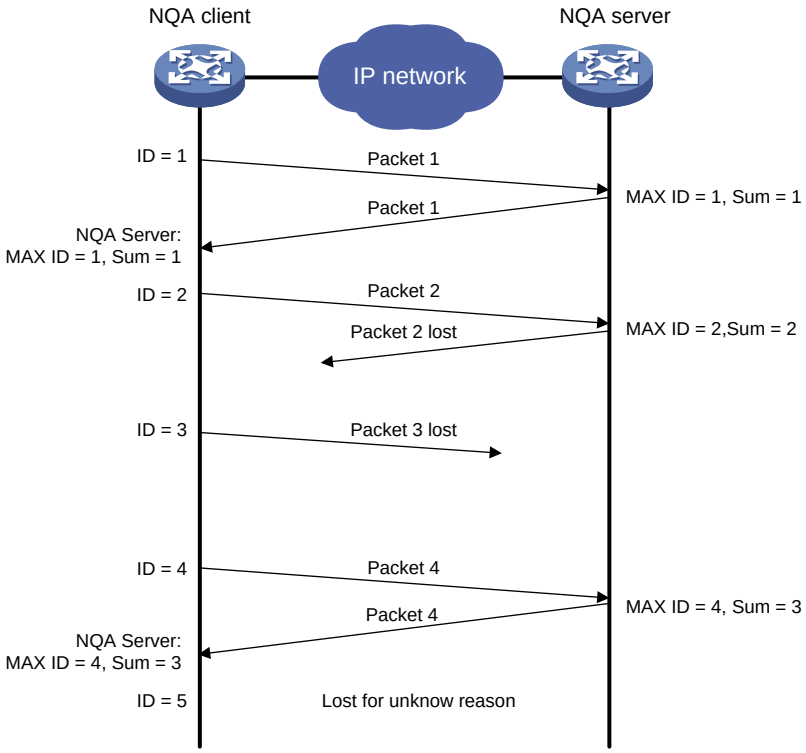
The NQA client can obtain the following information:

- Number of packets sent by the NQA client.
- The greatest packet ID and the number of packets received by the NQA server.
- Number of packets received by the NQA client.

It can calculate the following based on the obtained information:

- Number of packets lost from source to destination = Greatest packet ID received by the NQA server – Number of packets received by the NQA server
- Number of packets lost from destination to source = Number of packets received by the NQA server – Number of packets received by the NQA client
- Number of packets lost on the unknown directions = Number of packets sent by the NQA client – Number of packets received by the NQA client – Number of packets lost from source to destination – Number of packets lost from destination to source

Figure 6 UDP-jitter one-way packet loss statistics diagram



Restrictions and guidelines

The UDP jitter operation requires both the NQA server and the NQA client. Before you perform the UDP jitter operation, configure the UDP listening service on the NQA server. For more information about UDP listening service configuration, see "Configuring the NQA server."

To ensure successful UDP jitter operations and avoid affecting existing services, do not perform the operations on well-known ports from 1 to 1023.

The **display nqa history** command does not display the results or statistics of the UDP jitter operation. To view the results or statistics of the UDP jitter operation, use the **display nqa result** or **display nqa statistics** command.

Before starting the operation, make sure the network devices are time synchronized by using NTP. For more information about NTP, see "Configuring NTP."

Procedure

1. Enter system view.

   system-view
2. Create an NQA operation and enter NQA operation view.

   nqa entry *admin-name operation-tag*
3. Specify the UDP jitter type and enter its view.

   type udp-jitter
4. Specify the destination IP address for UDP packets.

   destination ip *ip-address*

   By default, no destination IP address is specified.

   The destination IP address must be the same as the IP address of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.

5. Specify the destination port number for UDP packets.

   destination port *port-number*

   By default, no destination port number is specified.

   The destination port number must be the same as the port number of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.

6. Specify the source IP address for UDP packets.

   source ip *ip-address*

   By default, the source IP address of UDP packets is the primary IP address of their output interface.

   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no UDP packets can be sent out.

7. Specify the source port number for UDP packets.

   source port *port-number*

   By default, the system automatically selects an unused port as the source port for UDP packets.

   For successful probing, make sure the source port number specified in the command is not used by any service on the device. As a best practice, use the default port number.

8. Set the number of UDP packets sent per probe.

   probe packet-number *number*

   The default setting is 10.

9. Set the interval for sending UDP packets.

   probe packet-interval *interval*

   The default setting is 20 milliseconds.

10. Specify how long the NQA client waits for a response from the server before it regards the response times out.

    probe packet-timeout *timeout*

    The default setting is 3000 milliseconds.

11. (Optional.) Set the payload size for each UDP packet.

    data-size *size*

    The default payload size is 100 bytes.

12. (Optional.) Specify the payload fill string for UDP packets.

    data-fill *string*

    The default payload fill string is 00010203040506070809.

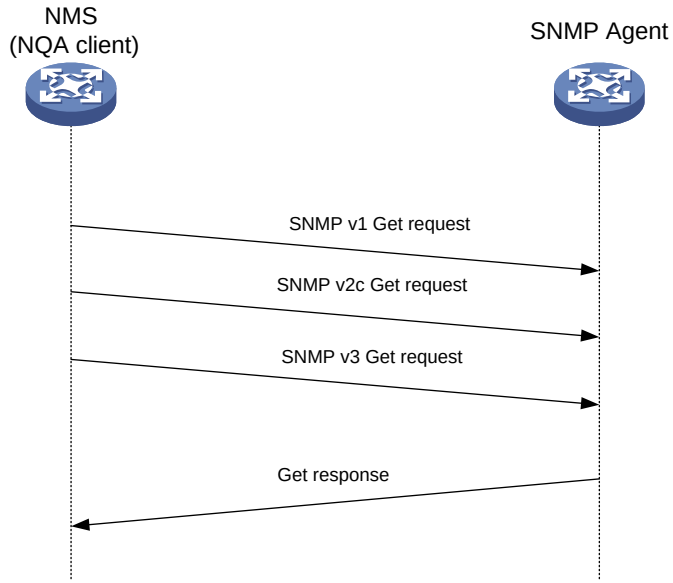
Configuring the SNMP operation

About this task

In an SNMP operation, the NQA client sends an SNMP Get request to the SNMP agent, and determines SNMP service availability on the SNMP agent based on whether it can receive a reply.

Currently, the widely used mainstream versions of SNMP on network devices are v1, v2c, and v3. During each operation, the NQA client will test all three versions of SNMP (v1, v2c, and v3). The operation is successful if a reply is received from any of these versions, as shown in Figure 7.

Figure 7 SNMP operation diagram



Procedure

1. Enter system view.

   system-view

2. Create an NQA operation and enter NQA operation view.

   nqa entry *admin-name operation-tag*

3. Specify the SNMP type and enter its view.

   type snmp

4. Specify the destination address for SNMP packets.

   destination ip *ip-address*

   By default, no destination IP address is specified.

5. Specify the source IP address for SNMP packets.

   source ip *ip-address*

   By default, the source IP address of SNMP packets is the primary IP address of their output interface.

   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no SNMP packets can be sent out.

6. Specify the source port number for SNMP packets.

   source port *port-number*

   By default, the system automatically selects an unused port as the source port for SNMP packets.

   For successful probing, make sure the source port number specified in the command is not used by any service on the device. As a best practice, use the default port number.

7. Specify the community name carried in the SNMPv1 and SNMPv2c packets.

   community read { **cipher** | **simple** } *community-name*

   By default, the SNMPv1 and SNMPv2c packets carry community name **public**.

   Make sure the specified community name is the same as the community name configured on the SNMP agent.

Configuring the TCP operation

About this task

The TCP operation measures the time for the NQA client to establish a TCP connection to a port on the NQA server.

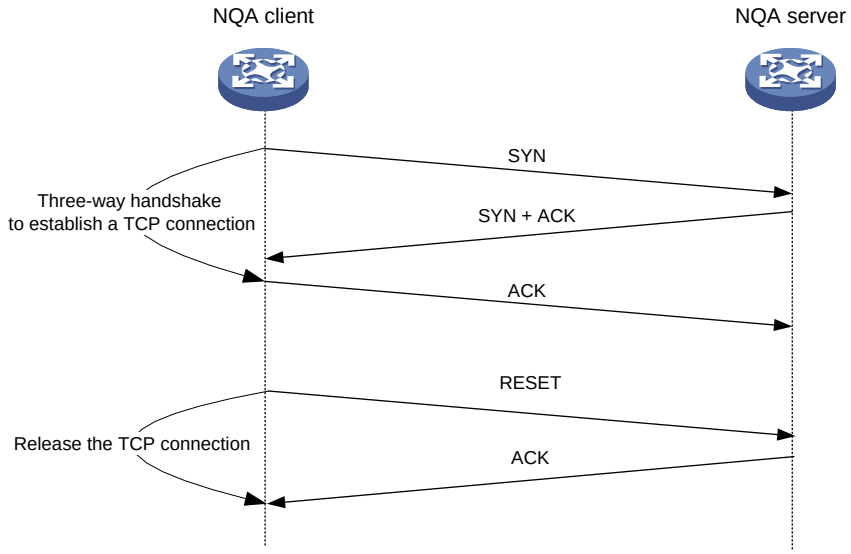
Do not initiate TCP operations too frequently. Doing so might occupy too many resources and affect normal services on the destination device.

As shown in Figure 8, a TCP operation involves the following steps:

- Establish a TCP connection via the three-way handshake mechanism.
- Release the TCP connection through RESET packets.

If a TCP connection can be successfully established and then released, the operation is successful. For TCP operation, a single probe refers to the establishment of one TCP connection.

Figure 8 TCP operation diagram



Restrictions and guidelines

The TCP operation requires both the NQA server and the NQA client. Before you perform a TCP operation, configure a TCP listening service on the NQA server. For more information about the TCP listening service configuration, see "Configuring the NQA server."

The TCP operation sets up a TCP connection per probe.

Procedure

1. Enter system view.
   `system-view`
2. Create an NQA operation and enter NQA operation view.
   `nqa entry` *`admin-name operation-tag`*
3. Specify the TCP type and enter its view.
   `type tcp`
4. Specify the destination address for TCP packets.
   `destination ip` *`ip-address`*
   By default, no destination IP address is specified.

The destination address must be the same as the IP address of the TCP listening service configured on the NQA server. To configure a TCP listening service on the server, use the **nqa server tcp-connect** command.

5. Specify the destination port for TCP packets.

   destination port *port-number*

   By default, no destination port number is configured.

   The destination port number must be the same as the port number of the TCP listening service configured on the NQA server. To configure a TCP listening service on the server, use the **nqa server tcp-connect** command.

6. Specify the source IP address for TCP packets.

   source ip *ip-address*

   By default, the source IP address of TCP packets is the primary IP address of their output interface.

   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no TCP packets can be sent out.

Configuring the UDP echo operation

About this task

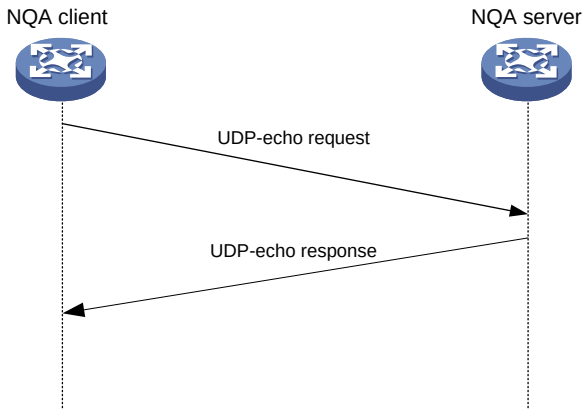
A UDP echo operation is typically used to test the connectivity and delay of a network. When you test the network connectivity and delay with UDP packets, you must enable NQA server on the peer end, and enable up the associated UDP port on the NQA server.

For a UDP echo operation, a probe refers to sending a probe packet.

As shown in Figure 9, a UDP echo operation is performed as follows:

1. The NQA client sends a UDP packet to the destination according to the configured probe time and frequency.
2. Upon receiving the UDP packet, the destination end sends a reply based on the packet.
3. Upon receiving the UDP packet, the NQA client calculates the time for the packet to reach the destination IP address and the packet loss rate, which reflects the network performance and status.

Figure 9 UDP echo operation diagram



Restrictions and guidelines

The UDP echo operation requires both the NQA server and the NQA client. Before you perform a UDP echo operation, configure a UDP listening service on the NQA server. For more information about the UDP listening service configuration, see "Configuring the NQA server."

The UDP echo operation sends a UDP packet to the destination device per probe.

Procedure

1. Enter system view.
   system-view
2. Create an NQA operation and enter NQA operation view.
   nqa entry *admin-name operation-tag*
3. Specify the UDP echo type and enter its view.
   type udp-echo
4. Specify the destination address for UDP packets.
   destination ip *ip-address*

   By default, no destination IP address is specified.

   The destination address must be the same as the IP address of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.
5. Specify the destination port number for UDP packets.
   destination port *port-number*

   By default, no destination port number is specified.

   The destination port number must be the same as the port number of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.
6. Specify the source IP address for UDP packets.
   source ip *ip-address*

   By default, the source IP address of UDP packets is the primary IP address of their output interface.

   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no UDP packets can be sent out.
7. Specify the source port number for UDP packets.
   source port *port-number*

   By default, the system automatically selects an unused port as the source port for UDP packets.

   For successful probing, make sure the source port number specified in the command is not used by any service on the device. As a best practice, use the default port number.
8. (Optional.) Set the payload size for each UDP packet.
   data-size *size*

   The default setting is 100 bytes.
9. (Optional.) Specify the payload fill string for UDP packets.
   data-fill *string*

   The default payload fill string is 00010203040506070809.

Configuring the UDP tracert operation

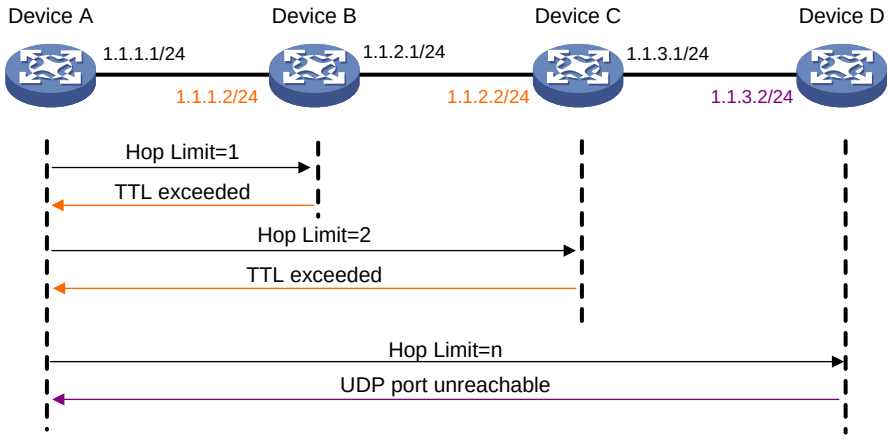
About this task

The UDP tracert operation determines the routing path from the source device to the destination device. A UDP tracert operation sends a probe packet to a hop along the path per probe.

As shown in Figure 10, the UDP tracert operation is implemented based on the ICMP protocol. The operation process is similar to that of a regular tracert as follows:

1. The source device (Device A) sends an IP packet to the destination device (Device D) with a TTL value of 1 and a UDP port number that no application on the destination can use.
2. The first hop (Device B, the Layer 3 device that first receives the packet) responds by sending a TTL-expired ICMP message (containing IP address 1.1.1.2 of the first hop) to the source. In this way, the source device obtains the IP address of the first Layer 3 device (1.1.1.2).
3. The source device sends a packet with a TTL value of 2 to the destination device.
4. The second hop (Device C) responds with a TTL-expired ICMP message. In this way, the source device obtains the IP address of the second Layer 3 device (1.1.2.2).
5. This process continues until the packet reaches the destination device. Because no application on the destination device uses the specified UDP port, the destination device returns a port-unreachable ICMP error message (containing IP address 1.1.3.2 of the destination device).
6. Upon receiving the port-unreachable ICMP error message, the source device determines that the packet has reached the destination, and obtains the path from the source to the destination (1.1.1.2; 1.1.2.2; 1.1.3.2).

Figure 10 UDP tracert operation diagram



Restrictions and guidelines

The UDP tracert operation is not supported on IPv6 networks. To determine the routing path that the IPv6 packets traverse from the source to the destination, use the **tracert ipv6** command. For more information about the command, see *Network Management and Monitoring Command Reference.*

Prerequisites

Before you configure the UDP tracert operation, you must perform the following tasks:

- Enable sending ICMP time exceeded messages on the intermediate devices between the source and destination devices. If the intermediate devices are H3C devices, use the **ip ttl-expires enable** command.
- Enable sending ICMP destination unreachable messages on the destination device. If the destination device is an H3C device, use the **ip unreachables enable** command.

  For more information about the **ip ttl-expires enable** and **ip unreachables enable** commands, see *Layer 3—IP Services Command Reference*.

Procedure

1. Enter system view.

   system-view
2. Create an NQA operation and enter NQA operation view.

nqa entry *admin-name operation-tag*

3. Specify the UDP tracert operation type and enter its view.

   type udp-tracert

4. Specify the destination device for the operation. Choose one option as needed:
   - Specify the destination device by its host name.

     destination host *host-name*

     By default, no destination host name is specified.
   - Specify the destination device by its IP address.

     destination ip *ip-address*

     By default, no destination IP address is specified.

5. Specify the destination port number for UDP packets.

   destination port *port-number*

   By default, the destination port number is 33434.

   This port number must be an unused number on the destination device, so that the destination device can reply with ICMP port unreachable messages.

6. Specify an output interface for UDP packets.

   out interface *interface-type interface-number*

   By default, the NQA client determines the output interface based on the routing table lookup.

7. Specify the source IP address for UDP packets.
   - Specify the IP address of the specified interface as the source IP address.

     source interface *interface-type interface-number*

     By default, the source IP address of UDP packets is the primary IP address of their output interface.
   - Specify the source IP address.

     source ip *ip-address*

     The specified source interface must be up. The source IP address must be the IP address of a local interface, and the local interface must be up. Otherwise, no probe packets can be sent out.

8. Specify the source port number for UDP packets.

   source port *port-number*

   By default, the system automatically selects an unused port as the source port for UDP packets.

   For successful probing, make sure the source port number specified in the command is not used by any service on the device. As a best practice, use the default port number.

9. Set the maximum number of consecutive probe failures.

   max-failure *times*

   The default setting is 5.

10. Set the initial TTL value for UDP packets.

    init-ttl *value*

    The default setting is 1.

11. (Optional.) Set the payload size for each UDP packet.

    data-size *size*

    The default setting is 100 bytes.

12. (Optional.) Enable the no-fragmentation feature.

    no-fragment enable

    By default, the no-fragmentation feature is disabled.

Configuring the voice operation

About this task

The voice operation measures VoIP network performance.

The voice operation works as follows:

1. The NQA client sends voice packets at sending intervals to the destination device (NQA server).

   The voice packets are of one of the following codec types:

   - G.711 A-law.
   - G.711 μ-law.
   - G.729 A-law.

2. The destination device time stamps each voice packet it receives and sends it back to the source.
3. Upon receiving the packet, the source device calculates the jitter and one-way delay based on the timestamp.

The voice operation sends a number of voice packets to the destination device per probe. The number of packets to send per probe is determined by using the **probe packet-number** command.

The following parameters that reflect VoIP network performance can be calculated by using the metrics gathered by the voice operation:

- **Calculated Planning Impairment Factor (ICPIF)**—Measures impairment to voice quality on a VoIP network. It is decided by packet loss and delay. A higher value represents a lower service quality.
- **Mean Opinion Scores (MOS)**—A MOS value can be evaluated by using the ICPIF value, in the range of 1 to 5. A higher value represents a higher service quality.

The evaluation of voice quality depends on users' tolerance for voice quality. For users with higher tolerance for voice quality, use the **advantage-factor** command to set an advantage factor. When the system calculates the ICPIF value, it subtracts the advantage factor to modify ICPIF and MOS values for voice quality evaluation.

The voice operation requires both the NQA server and the NQA client. Before you perform a voice operation, configure a UDP listening service on the NQA server. For more information about UDP listening service configuration, see "Configuring the NQA server."

Restrictions and guidelines

To ensure successful voice operations and avoid affecting existing services, do not perform the operations on well-known ports from 1 to 1023.

The **display nqa history** command does not display the results or statistics of the voice operation. To view the results or statistics of the voice operation, use the **display nqa result** or **display nqa statistics** command.

Before starting the operation, make sure the network devices are time synchronized by using NTP. For more information about NTP, see "Configuring NTP."

Procedure

1. Enter system view.

   system-view

2. Create an NQA operation and enter NQA operation view.

   nqa entry *admin-name operation-tag*

3. Specify the voice type and enter its view.

type voice

4. Specify the destination IP address for voice packets.

   destination ip *ip-address*

   By default, no destination IP address is configured.

   The destination IP address must be the same as the IP address of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.

5. Specify the destination port number for voice packets.

   destination port *port-number*

   By default, no destination port number is configured.

   The destination port number must be the same as the port number of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.

6. Specify the source IP address for voice packets.

   source ip *ip-address*

   By default, the source IP address of voice packets is the primary IP address of their output interface.

   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no voice packets can be sent out.

7. Specify the source port number for voice packets.

   source port *port-number*

   By default, the system automatically selects an unused port as the source port for voice packets.

   For successful probing, make sure the source port number specified in the command is not used by any service on the device. As a best practice, use the default port number.

8. Configure the basic voice operation parameters.
   - Specify the codec type.

     codec-type { **g711a** | **g711u** | **g729a** }

     By default, the codec type is G.711 A-law.
   - Set the advantage factor for calculating MOS and ICPIF values.

     advantage-factor *factor*

     By default, the advantage factor is 0.

9. Configure the probe parameters for the voice operation.
   - Set the number of voice packets to be sent per probe.

     probe packet-number *number*

     The default setting is 1000.
   - Set the interval for sending voice packets.

     probe packet-interval *interval*

     The default setting is 20 milliseconds.
   - Specify how long the NQA client waits for a response from the server before it regards the response times out.

     probe packet-timeout *timeout*

     The default setting is 5000 milliseconds.

10. Configure the payload parameters.

    a. Set the payload size for voice packets.

    data-size *size*

By default, the voice packet size varies by codec type. The default packet size is 172 bytes for G.711A-law and G.711 μ-law codec type, and 32 bytes for G.729 A-law codec type.

b. (Optional.) Specify the payload fill string for voice packets.

data-fill *string*

The default payload fill string is 00010203040506070809.

Configuring the DLSw operation

About this task

The DLSw operation measures the response time of a DLSw device.

An NQA DLSw operation initiates a TCP connection to the DLSw port of the peer device, and then determines the DLSw status (enabled or not) on the peer device based on whether the connection can be established. Similar to a TCP operation in implementation, an NQA DLSw operation can be considered as a TCP operation with a fixed destination port number.

The DLSw operation sets up a DLSw connection to the DLSw device per probe.

Restrictions and guidelines

For the DLSw operation to succeed, configure the **nqa server tcp-connect** command on the NQA server and make sure the port number for the TCP listening service is 2065.

Procedure

1. Enter system view.

   system-view

2. Create an NQA operation and enter NQA operation view.

   nqa entry *admin-name operation-tag*

3. Specify the DLSw type and enter its view.

   type dlsw

4. Specify the destination IP address for the probe packets.

   destination ip *ip-address*

   By default, no destination IP address is specified.

5. Specify the source IP address for the probe packets.

   source ip *ip-address*

   By default, the source IP address of the probe packets is the primary IP address of their output interface.

   The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

Configuring the path jitter operation

About this task

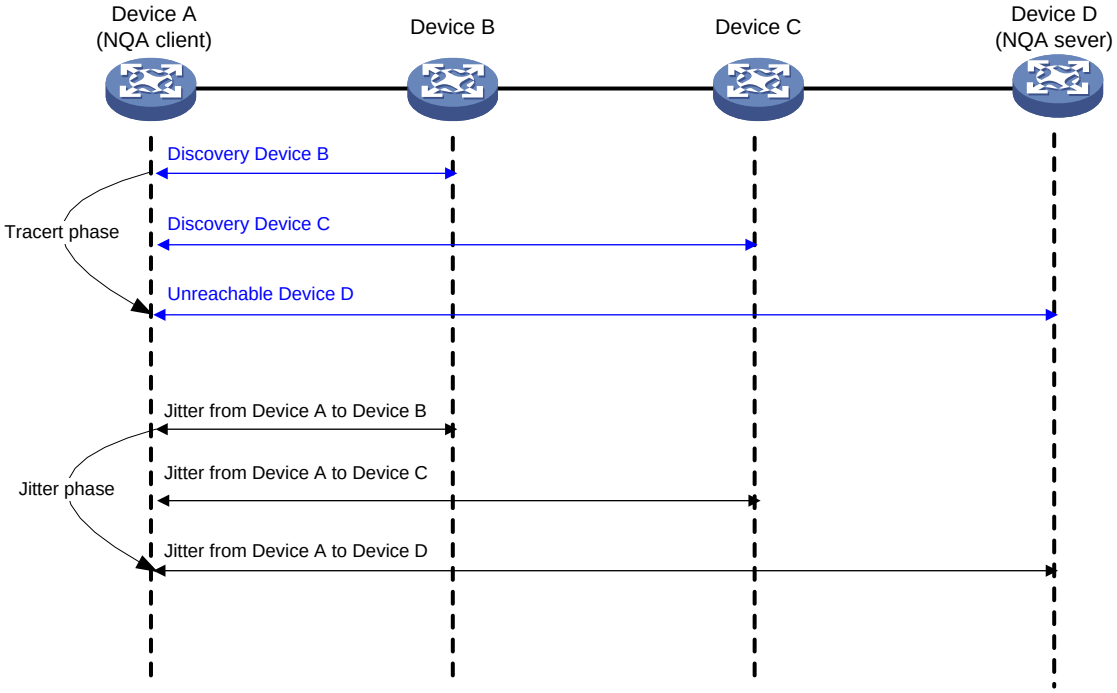
As a supplement to the UDP jitter operation, the path jitter operation is used in scenarios with significant jitter to further detect the quality of the network along the intermediate path. This can help you identify specific path segments with poor network performance.

As shown in Figure 11, a path jitter operation involves the following steps per probe:

1. The NQA client obtains the path to the destination through tracert. A maximum of 64 hops can be detected.

2. The NQA client sends a number of ICMP echo requests to each hop along the path to check for packet loss and calculate the delay and jitter time of the path. The path jitter operation records the average jitter, positive jitter, and negative jitter values for each path.

Figure 11 Path jitter operation diagram



Prerequisites

Before you configure the path jitter operation, you must perform the following tasks:

- Enable sending ICMP time exceeded messages on the intermediate devices between the source and destination devices. If the intermediate devices are H3C devices, use the **ip ttl-expires enable** command.
- Enable sending ICMP destination unreachable messages on the destination device. If the destination device is an H3C device, use the **ip unreachables enable** command.

  For more information about the **ip ttl-expires enable** and **ip unreachables enable** commands, see *Layer 3—IP Services Command Reference*.

Procedure

1. Enter system view.

   system-view
2. Create an NQA operation and enter NQA operation view.

   nqa entry *admin-name operation-tag*
3. Specify the path jitter type and enter its view.

   type path-jitter
4. Specify the destination IP address for ICMP echo requests.

   destination ip *ip-address*

   By default, no destination IP address is specified.
5. Specify the source IP address for ICMP echo requests.

   source ip *ip-address*

   By default, the source IP address of ICMP echo requests is the primary IP address of their output interface.

The source IP address must be the IP address of a local interface, and the interface must be up. Otherwise, no ICMP echo requests can be sent out.

6. Configure the probe parameters for the path jitter operation.
   a. Set the number of ICMP echo requests to be sent per probe.
      probe packet-number *number*
      The default setting is 10.
   b. Set the interval for sending ICMP echo requests.
      probe packet-interval *interval*
      The default setting is 20 milliseconds.
   c. Specify how long the NQA client waits for a response from the server before it regards the response times out.
      probe packet-timeout *timeout*
      The default setting is 3000 milliseconds.
7. (Optional.) Specify an LSR path.
   lsr-path *ip-address*&<1-8>
   By default, no LSR path is specified.
   The path jitter operation uses tracert to detect the LSR path to the destination, and sends ICMP echo requests to each hop on the LSR path.
8. Perform the path jitter operation only on the destination address.
   target-only
   By default, the path jitter operation is performed on each hop on the path to the destination.
9. (Optional.) Set the payload size for each ICMP echo request.
   data-size *size*
   The default setting is 100 bytes.
10. (Optional.) Specify the payload fill string for ICMP echo requests.
    data-fill *string*
    The default payload fill string is 00010203040506070809.

Configuring optional parameters for the NQA operation

Restrictions and guidelines

The parameter settings take effect only on the current operation.

The supported parameters vary by NQA operation type. For information about supported parameters, see NQA commands in *Network Management and Monitoring Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter the view of an existing NQA operation.
   nqa entry *admin-name operation-tag*
3. Specify an NQA operation type and enter its view.
   type { **dhcp** | **dlsw** | **dns** | | **ftp** | **http** | **icmp-echo** | **icmp-jitter** | **path-jitter** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **udp-tracert** | **voice** }
4. Configure a description for the operation.
   description *text*
   By default, no description is configured.

5. Set the interval at which the NQA operation repeats.

   frequency *interval*

   For a voice or path jitter operation, the default setting is 60000 milliseconds.

   For other types of operations, the default setting is 0 milliseconds, and only one operation is performed.

   When the interval expires, but the operation is not completed or is not timed out, the next operation does not start.

6. Specify the probe times.

   probe count *times*

   In an UDP tracert operation, the NQA client performs three probes to each hop to the destination by default.

   In other types of operations, the NQA client performs one probe to the destination per operation by default.

   This command is not available for the voice and path jitter operations. Each of these operations performs only one probe.

7. Set the probe timeout time.

   probe timeout *timeout*

   The default setting is 3000 milliseconds.

8. Set the maximum number of hops that the probe packets can traverse.

   ttl *value*

   The default setting is 30 for probe packets of the UDP tracert operation, and is 20 for probe packets of other types of operations.

9. Set the ToS value in the IP header of the probe packets.

   tos *value*

   The default setting is 0.

10. Enable the routing table bypass feature.

    route-option bypass-route

    By default, the routing table bypass feature is disabled.

    This command does not take effect if the destination address of the NQA operation is an IPv6 address.

11. Specify the VPN instance where the operation is performed.

    vpn-instance *vpn-instance-name*

    By default, the operation is performed on the public network.

Configuring the collaboration feature

About this task

The collaboration feature associates a detection module, the Track module, and an application module. You can use NQA as a detection module in the collaboration feature to monitor the NQA operation. If the number of consecutive NQA operation failures reaches the specified threshold, the Track-associated application module is triggered to take relevant actions.

Restrictions and guidelines

The collaboration feature is not available for the following types of operations:

- ICMP jitter operation.
- UDP jitter operation.
- UDP tracert operation.

- Voice operation.
- Path jitter operation.

Procedure

1. Enter system view.

   system-view
2. Enter the view of an existing NQA operation.

   nqa entry *admin-name operation-tag*
3. Configure a reaction entry.

   reaction *item-number* **checked-element probe-fail threshold-type consecutive** *consecutive-occurrences* **action-type trigger-only**

   You cannot modify the content of an existing reaction entry.
4. Return to system view.

   quit
5. Associate Track with NQA.

   For information about the configuration, see *High Availability Configuration Guide*.
6. Associate Track with an application module.

   For information about the configuration, see *High Availability Configuration Guide*.

Configuring threshold monitoring

About this task

This feature allows you to monitor the NQA operation running status.

An NQA operation supports the following threshold types:

- **average**—If the average value for the monitored performance metric either exceeds the upper threshold or goes below the lower threshold, a threshold violation occurs.
- **accumulate**—If the total number of times that the monitored performance metric is out of the specified value range reaches or exceeds the specified threshold, a threshold violation occurs.
- **consecutive**—If the number of consecutive times that the monitored performance metric is out of the specified value range reaches or exceeds the specified threshold, a threshold violation occurs.

Threshold violations for the average or accumulate threshold type are determined on a per NQA operation basis. The threshold violations for the consecutive type are determined from the time the NQA operation starts.

The following actions might be triggered:

- **none**—NQA displays results only on the terminal screen. It does not send traps to the NMS.
- **trap-only**—NQA displays results on the terminal screen, and meanwhile it sends traps to the NMS.

  To send traps to the NMS, the NMS address must be specified by using the **snmp-agent target-host** command. For more information about the command, see *Network Management and Monitoring Command Reference*.
- **trigger-only**—NQA displays results on the terminal screen, and meanwhile triggers other modules for collaboration.

In a reaction entry, configure a monitored element, a threshold type, and an action to be triggered to implement threshold monitoring.

The state of a reaction entry can be invalid, over-threshold, or below-threshold.

- Before an NQA operation starts, the reaction entry is in invalid state.
- If the threshold is violated, the state of the entry is set to over-threshold. Otherwise, the state of the entry is set to below-threshold.

Restrictions and guidelines

The threshold monitoring feature is not available for the path jitter operations.

Procedure

1. Enter system view.

   system-view

2. Enter the view of an existing NQA operation.

   nqa entry *admin-name operation-tag*

3. Enable sending traps to the NMS when specific conditions are met.

   reaction trap { **path-change** | **probe-failure** *consecutive-probe-failures* | **test-complete** | **test-failure** [*accumulate-probe-failures*] }

   By default, no traps are sent to the NMS.

   The ICMP jitter, UDP jitter, and voice operations support only the **test-complete** keyword.

   The following parameters are not available for the UDP tracert operation:

   - The **probe-failure** *consecutive-probe-failures* option.
   - The *accumulate-probe-failures* argument.

4. Configure threshold monitoring. Choose the options to configure as needed:

   - Monitor the operation duration.

     reaction *item-number* **checked-element probe-duration threshold-type** { **accumulate** *accumulate-occurrences* | **average** | **consecutive** *consecutive-occurrences* } **threshold-value** *upper-threshold lower-threshold* [**action-type** { **none** | **trap-only** }]

     This reaction entry is not supported in the ICMP jitter, UDP jitter, UDP tracert, or voice operations

   - Monitor failure times.

     reaction *item-number* **checked-element probe-fail threshold-type** { **accumulate** *accumulate-occurrences* | **consecutive** *consecutive-occurrences* } [**action-type** { **none** | **trap-only** }]

     This reaction entry is not supported in the ICMP jitter, UDP jitter, UDP tracert, or voice operations.

   - Monitor the number of consecutive operation failures.

     reaction *item-number* **checked-element test-fail threshold-type consecutive** *consecutive-occurrences* **threshold-value** *upper-threshold* **action-type** { **none** | **trap-only** }

     Only the ICMP echo, TCP/UDP echo, DHCP, DLSw, DNS, FTP, HTTP, and SNMP operations support this reaction entry.

   - Monitor the operation result change.

     reaction *item-number* **checked-element test-result-change threshold-value** *upper-threshold* **action-type** { **none** | **trap-only** }

     Only the ICMP echo operation supports this reaction entry.

   - Monitor the round-trip time.

     reaction *item-number* **checked-element rtt threshold-type** { **accumulate** *accumulate-occurrences* | **average** } **threshold-value** *upper-threshold lower-threshold* [**action-type** { **none** | **trap-only** }]

Only the ICMP jitter, UDP jitter, and voice operations support this reaction entry.

- Monitor packet loss.

  reaction *item-number* **checked-element packet-loss threshold-type accumulate** *accumulate-occurrences* [**action-type** { **none** | **trap-only** }]

  Only the ICMP jitter, UDP jitter, and voice operations support this reaction entry.

- Monitor the one-way jitter.

  reaction *item-number* **checked-element** { **jitter-ds** | **jitter-sd** } **threshold-type** { **accumulate** *accumulate-occurrences* | **average** } **threshold-value** *upper-threshold lower-threshold* [**action-type** { **none** | **trap-only** }]

  Only the ICMP jitter, UDP jitter, and voice operations support this reaction entry.

- Monitor the one-way delay.

  reaction *item-number* **checked-element** { **owd-ds** | **owd-sd** } **threshold-value** *upper-threshold lower-threshold* [**action-type** { **none** | **trap-only** }]

  Only the ICMP jitter, UDP jitter, and voice operations support this reaction entry.

- Monitor the ICPIF value.

  reaction *item-number* **checked-element icpif threshold-value** *upper-threshold lower-threshold* [**action-type** { **none** | **trap-only** }]

  Only the voice operation supports this reaction entry.

- Monitor the MOS value.

  reaction *item-number* **checked-element mos threshold-value** *upper-threshold lower-threshold* [**action-type** { **none** | **trap-only** }]

  Only the voice operation supports this reaction entry.

The DNS operation does not support the action of sending trap messages. For the DNS operation, the action type can only be **none**.

Configuring the NQA statistics collection feature

About this task

NQA forms statistics within the same collection interval as a statistics group. To display information about the statistics groups, use the **display nqa statistics** command.

When the maximum number of statistics groups is reached, the NQA client deletes the oldest statistics group to save a new one.

A statistics group is automatically deleted when its hold time expires.

Restrictions and guidelines

The NQA statistics collection feature is not available for the UDP tracert operations.

If you use the **frequency** command to set the interval to 0 milliseconds for an NQA operation, NQA does not generate any statistics group for the operation.

Procedure

1. Enter system view.

   system-view

2. Enter the view of an existing NQA operation.

   nqa entry *admin-name operation-tag*

3. Set the statistics collection interval.

statistics interval *interval*

The default setting is 60 minutes.

4. Set the maximum number of statistics groups that can be saved.

   statistics max-group *number*

   By default, the NQA client can save a maximum of two statistics groups for an operation.

   To disable the NQA statistics collection feature, set the *number* argument to 0.

5. Set the hold time of statistics groups.

   statistics hold-time *hold-time*

   The default setting is 120 minutes.

Configuring the saving of NQA history records

About this task

This task enables the NQA client to save NQA history records. You can use the **display nqa history** command to display the NQA history records.

Restrictions and guidelines

The NQA history record saving feature is not available for the following types of operations:

- ICMP jitter operation.
- UDP jitter operation.
- Voice operation.
- Path jitter operation.

Procedure

1. Enter system view.

   system-view

2. Enter the view of an existing NQA operation.

   nqa entry *admin-name operation-tag*

3. Enable the saving of history records for the NQA operation.

   history-record enable

   By default, this feature is enabled only for the UDP tracert operation.

4. Set the lifetime of history records.

   history-record keep-time *keep-time*

   The default setting is 120 minutes.

   A record is deleted when its lifetime is reached.

5. Set the maximum number of history records that can be saved.

   history-record number *number*

   The default setting is 50.

   When the maximum number of history records is reached, the system will delete the oldest record to save a new one.

Scheduling the NQA operation on the NQA client

About this task

The NQA operation runs between the specified start time and end time (the start time plus operation duration). If the specified start time is ahead of the system time, the operation starts immediately. If

both the specified start and end time are ahead of the system time, the operation does not start. To display the current system time, use the **display clock** command.

Restrictions and guidelines

You cannot enter the operation type view or the operation view of a scheduled NQA operation.

A system time adjustment does not affect started or completed NQA operations. It affects only the NQA operations that have not started.

Procedure

1. Enter system view.

   system-view

2. Specify the scheduling parameters for an NQA operation.

   nqa schedule *admin-name operation-tag* **start-time** { *hh:mm:ss* [*yyyy/mm/dd* | *mm/dd/yyyy*] | **now** } **lifetime** { *lifetime* | **forever** } [**recurring**]

Configuring NQA templates on the NQA client

Restrictions and guidelines

Some operation parameters for an NQA template can be specified by the template configuration or the feature (such as load balancing) that uses the template. When both are specified, the parameters in the template configuration take effect.

NQA template tasks at a glance

To configure NQA templates, perform the following tasks:

1. Perform at least one of the following tasks:
   - Configuring the ICMP template
   - Configuring the DNS template
   - Configuring the TCP template
   - Configuring the TCP half open template
   - Configuring the UDP template
   - Configuring the HTTP template
   - Configuring the HTTPS template
   - Configuring the FTP template
   - Configuring the RADIUS template
   - Configuring the SSL template
2. (Optional.) Configuring optional parameters for the NQA template

Configuring the ICMP template

About this task

A feature that uses the ICMP template performs the ICMP operation to measure the reachability of a destination device. The ICMP template is supported on both IPv4 and IPv6 networks.

Procedure

1. Enter system view.

system-view

2. Create an ICMP template and enter its view.

   nqa template icmp *name*

3. Specify the destination IP address for the operation.

   IPv4:

   destination ip *ip-address*

   IPv6:

   destination ipv6 *ipv6-address*

   By default, no destination IP address is configured.

4. Specify the source IP address for ICMP echo requests. Choose one option as needed:

   - Use the IP address of the specified interface as the source IP address.

     source interface *interface-type interface-number*

     By default, the primary IP address of the output interface is used as the source IP address of ICMP echo requests.

     The specified source interface must be up.

   - Specify the source IPv4 address.

     source ip *ip-address*

     By default, the primary IPv4 address of the output interface is used as the source IPv4 address of ICMP echo requests.

     The specified source IPv4 address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

   - Specify the source IPv6 address.

     source ipv6 *ipv6-address*

     By default, the primary IPv6 address of the output interface is used as the source IPv6 address of ICMP echo requests.

     The specified source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

5. Specify the next hop IP address for ICMP echo requests.

   IPv4:

   next-hop ip *ip-address*

   IPv6:

   next-hop ipv6 *ipv6-address*

   By default, no IP address of the next hop is configured.

   To avoid probe failure, make sure the IP address belongs to a directly connected next hop.

6. Configure the probe result sending on a per-probe basis.

   reaction trigger per-probe

   By default, the probe result is sent to the feature that uses the template after three consecutive failed or successful probes.

   If you execute the **reaction trigger per-probe** and **reaction trigger probe-pass** commands multiple times, the most recent configuration takes effect.

   If you execute the **reaction trigger per-probe** and **reaction trigger probe-fail** commands multiple times, the most recent configuration takes effect.

7. (Optional.) Set the payload size for each ICMP request.

   data-size *size*

   The default setting is 100 bytes.

8. (Optional.) Specify the payload fill string for ICMP echo requests.

   data-fill *string*

   The default payload fill string is the hexadecimal string 00010203040506070809.

Configuring the DNS template

About this task

A feature that uses the DNS template performs the DNS operation to determine the status of the server. The DNS template is supported on both IPv4 and IPv6 networks.

In DNS template view, you can specify the address expected to be returned. If the returned IP addresses include the expected address, the DNS server is valid and the operation succeeds. Otherwise, the operation fails.

Prerequisites

Create a mapping between the domain name and an address before you perform the DNS operation. For information about configuring the DNS server, see documents about the DNS server configuration.

Procedure

1. Enter system view.

   system-view

2. Create a DNS template and enter DNS template view.

   nqa template dns *name*

3. Specify the destination IP address for the probe packets.

   IPv4:

   destination ip *ip-address*

   IPv6:

   destination ipv6 *ipv6-address*

   By default, no destination address is specified.

4. Specify the destination port number for the probe packets.

   destination port *port-number*

   By default, the destination port number is 53.

5. Specify the source IP address for the probe packets.

   IPv4:

   source ip *ip-address*

   By default, the source IPv4 address of the probe packets is the primary IPv4 address of their output interface.

   The source IPv4 address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

   IPv6:

   source ipv6 *ipv6-address*

   By default, the source IPv6 address of the probe packets is the primary IPv6 address of their output interface.

   The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

6. Specify the source port number for the probe packets.

   source port *port-number*

By default, the system automatically selects an unused port as the source port for probe packets.

For successful probing, make sure the source port number specified in the command is not used by any service on the device. As a best practice, use the default port number.

7. Specify the domain name to be translated.

   resolve-target *domain-name*

   By default, no domain name is specified.

8. Specify the domain name resolution type.

   resolve-type { **A** | **AAAA** }

   By default, the type is type A.

   A type A query resolves a domain name to a mapped IPv4 address, and a type AAAA query to a mapped IPv6 address.

9. (Optional.) Specify the IP address that is expected to be returned.

   IPv4:

   expect ip *ip-address*

   IPv6:

   expect ipv6 *ipv6-address*

   By default, no expected IP address is specified.

Configuring the TCP template

About this task

A feature that uses the TCP template performs the TCP operation to test whether the NQA client can establish a TCP connection to a specific port on the server.

In TCP template view, you can specify the expected data to be returned. If you do not specify the expected data, the TCP operation tests only whether the client can establish a TCP connection to the server.

The TCP operation requires both the NQA server and the NQA client. Before you perform a TCP operation, configure a TCP listening service on the NQA server. For more information about the TCP listening service configuration, see "Configuring the NQA server."

Procedure

1. Enter system view.

   system-view

2. Create a TCP template and enter its view.

   nqa template tcp *name*

3. Specify the destination IP address for the probe packets.

   IPv4:

   destination ip *ip-address*

   IPv6:

   destination ipv6 *ipv6-address*

   By default, no destination IP address is specified.

   The destination address must be the same as the IP address of the TCP listening service configured on the NQA server. To configure a TCP listening service on the server, use the **nqa server tcp-connect** command.

4. Specify the destination port number for the operation.

   destination port *port-number*

By default, no destination port number is specified.

The destination port number must be the same as the port number of the TCP listening service configured on the NQA server. To configure a TCP listening service on the server, use the **nqa server tcp-connect** command.

5. Specify the source IP address for the probe packets.

   IPv4:

   source ip *ip-address*

   By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

   The source IP address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

   IPv6:

   source ipv6 *ipv6-address*

   By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

   The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

6. (Optional.) Specify the payload fill string for the probe packets.

   data-fill *string*

   By default, no payload fill string is configured for probe packets.

7. (Optional.) Configure the expected data.

   expect data *expression* [**offset** *number*]

   By default, no expected data is configured.

   The NQA client performs expect data check only when you configure both the **data-fill** and **expect-data** commands.

Configuring the TCP half open template

About this task

A feature that uses the TCP half open template performs the TCP half open operation to test whether the TCP service is available on the server. The TCP half open operation is used when the feature cannot get a response from the TCP server through an existing TCP connection.

In the TCP half open operation, the NQA client sends a TCP ACK packet to the server. If the client receives an RST packet, it considers that the TCP service is available on the server.

Procedure

1. Enter system view.

   system-view

2. Create a TCP half open template and enter its view.

   nqa template tcphalfopen *name*

3. Specify the destination IP address of the operation.

   IPv4:

   destination ip *ip-address*

   IPv6:

   destination ipv6 *ipv6-address*

   By default, no destination IP address is specified.

The destination address must be the same as the IP address of the TCP listening service configured on the NQA server. To configure a TCP listening service on the server, use the **nqa server tcp-connect** command.

4. Specify the source IP address for the probe packets.

   IPv4:

   source ip *ip-address*

   By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

   The source IPv4 address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

   IPv6:

   source ipv6 *ipv6-address*

   By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

   The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

5. Specify the next hop IP address for the probe packets.

   IPv4:

   next-hop ip *ip-address*

   IPv6:

   next-hop ipv6 *ipv6-address*

   By default, the IP address of the next hop is configured.

   To avoid probe failure, make sure the IP address belongs to a directly connected next hop.

6. Configure the probe result sending on a per-probe basis.

   reaction trigger per-probe

   By default, the probe result is sent to the feature that uses the template after three consecutive failed or successful probes.

   If you execute the **reaction trigger per-probe** and **reaction trigger probe-pass** commands multiple times, the most recent configuration takes effect.

   If you execute the **reaction trigger per-probe** and **reaction trigger probe-fail** commands multiple times, the most recent configuration takes effect.

Configuring the UDP template

About this task

A feature that uses the UDP template performs the UDP operation to test the following items:

- Reachability of a specific port on the NQA server.
- Availability of the requested service on the NQA server.

In UDP template view, you can specify the expected data to be returned. If you do not specify the expected data, the UDP operation tests only whether the client can receive the response packet from the server.

The UDP operation requires both the NQA server and the NQA client. Before you perform a UDP operation, configure a UDP listening service on the NQA server. For more information about the UDP listening service configuration, see "Configuring the NQA server."

Procedure

1. Enter system view.

system-view

2. Create a UDP template and enter its view.

   nqa template udp *name*

3. Specify the destination IP address of the operation.

   IPv4:

   destination ip *ip-address*

   IPv6:

   destination ipv6 *ipv6-address*

   By default, no destination IP address is specified.

   The destination address must be the same as the IP address of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.

4. Specify the destination port number for the operation.

   destination port *port-number*

   By default, no destination port number is specified.

   The destination port number must be the same as the port number of the UDP listening service configured on the NQA server. To configure a UDP listening service on the server, use the **nqa server udp-echo** command.

5. Specify the source IP address for the probe packets.

   IPv4:

   source ip *ip-address*

   By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

   The source IP address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

   IPv6:

   source ipv6 *ipv6-address*

   By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

   The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

6. Specify the payload fill string for the probe packets.

   data-fill *string*

   The default payload fill string is the hexadecimal string 00010203040506070809.

7. (Optional.) Set the payload size for the probe packets.

   data-size *size*

   The default setting is 100 bytes.

8. (Optional.) Configure the expected data.

   expect data *expression* [**offset** *number*]

   By default, no expected data is configured.

   Expected data check is performed only when both the **data-fill** command and the **expect data** command are configured.

Configuring the HTTP template

About this task

A feature that uses the HTTP template performs the HTTP operation to measure the time it takes the NQA client to obtain data from an HTTP server.

The expected data is checked only when the data is configured and the HTTP response contains the Content-Length field in the HTTP header.

The status code of the HTTP packet is a three-digit field in decimal notation, and it includes the status information for the HTTP server. The first digit defines the class of response.

Prerequisites

Before you perform the HTTP operation, you must configure the HTTP server.

Procedure

1. Enter system view.

   system-view

2. Create an HTTP template and enter its view.

   nqa template http *name*

3. Specify the destination URL for the HTTP template.

   url *url*

   By default, no destination URL is specified for an HTTP template.

   Enter the URL in one of the following formats:

   - http://*host*/*resource*
   - http://*host*:*port*/*resource*

   If the value for the *host* parameter is an IPv6 address, you must use a pair of square brackets ([]) to include the IPv6 address, for example, http://[2001::1]/test.

4. Specify an HTTP login username.

   username *username*

   By default, no HTTP login username is specified.

5. Specify an HTTP login password.

   password { **cipher** | **simple** } *string*

   By default, no HTTP login password is specified.

6. Specify the HTTP version.

   version { **v1.0** | **v1.1** }

   By default, HTTP 1.0 is used.

7. Specify the HTTP operation type.

   operation { **get** | **post** | **raw** }

   By default, the HTTP operation type is **get**.

   If you set the operation type to raw, the client pads the content configured in raw request view to the HTTP request to send to the HTTP server.

8. Configure the content of the HTTP raw request.

   a. Enter raw request view.

      raw-request

      Every time you enter raw request view, the previously configured raw request content is cleared.

   b. Enter or paste the request content.

By default, no request content is configured.

To ensure successful operations, make sure the request content does not contain command aliases configured by using the **alias** command. For more information about the **alias** command, see CLI commands in *Fundamentals Command Reference*.

c. Return to HTTP template view.

quit

The system automatically saves the configuration in raw request view before it returns to HTTP template view.

This step is required only when the operation type is set to **raw**.

9. Specify the source IP address for the probe packets.

IPv4:

source ip *ip-address*

By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

The source IPv4 address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

IPv6:

source ipv6 *ipv6-address*

By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

10. (Optional.) Configure the expected status codes.

expect status *status-list*

By default, no expected status code is configured.

11. (Optional.) Configure the expected data.

expect data *expression* [**offset** *number*]

By default, no expected data is configured.

Configuring the HTTPS template

About this task

A feature that uses the HTTPS template performs the HTTPS operation to measure the time it takes for the NQA client to obtain data from an HTTPS server.

The expected data is checked only when the expected data is configured and the HTTPS response contains the Content-Length field in the HTTPS header.

The status code of the HTTPS packet is a three-digit field in decimal notation, and it includes the status information for the HTTPS server. The first digit defines the class of response.

Prerequisites

Before you perform the HTTPS operation, configure the HTTPS server and the SSL client policy for the SSL client. For information about configuring SSL client policies, see *Security Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Create an HTTPS template and enter its view.

nqa template https *name*

3. Specify the destination URL for the HTTPS template.

   url *url*

   By default, no destination URL is specified for an HTTPS template.

   Enter the URL in one of the following formats:

   - https://*host*/*resource*
   - https://*host*:*port*/*resource*

   If the value for the *host* parameter is an IPv6 address, you must use a pair of square brackets ([]) to include the IPv6 address, for example, https://[2001::1]/test.

4. Specify an HTTPS login username.

   username *username*

   By default, no HTTPS login username is specified.

5. Specify an HTTPS login password.

   password { **cipher** | **simple** } *string*

   By default, no HTTPS login password is specified.

6. Specify an SSL client policy.

   ssl-client-policy *policy-name*

   By default, no SSL client policy is specified.

7. Specify the HTTPS version.

   version { **v1.0** | **v1.1** }

   By default, HTTPS 1.0 is used.

8. Specify the HTTPS operation type.

   operation { **get** | **post** | **raw** }

   By default, the HTTPS operation type is **get**.

   If you set the operation type to raw, the client pads the content configured in raw request view to the HTTPS request to send to the HTTPS server.

9. Configure the content of the HTTPS raw request.

   a. Enter raw request view.

   raw-request

   Every time you enter raw request view, the previously configured raw request content is cleared.

   b. Enter or paste the request content.

   By default, no request content is configured.

   To ensure successful operations, make sure the request content does not contain command aliases configured by using the **alias** command. For more information about the **alias** command, see CLI commands in *Fundamentals Command Reference*.

   c. Return to HTTPS template view.

   quit

   The system automatically saves the configuration in raw request view before it returns to HTTPS template view.

   This step is required only when the operation type is set to **raw**.

10. Specify the source IP address for the probe packets.

    IPv4:

    source ip *ip-address*

By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

The source IP address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

IPv6:

source ipv6 *ipv6-address*

By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

11. (Optional.) Configure the expected data.

expect data *expression* [**offset** *number*]

By default, no expected data is configured.

12. (Optional.) Configure the expected status codes.

expect status *status-list*

By default, no expected status code is configured.

Configuring the FTP template

About this task

A feature that uses the FTP template performs the FTP operation. The operation measures the time it takes the NQA client to transfer a file to or download a file from an FTP server.

Configure the username and password for the FTP client to log in to the FTP server before you perform an FTP operation. For information about configuring the FTP server, see *Fundamentals Configuration Guide*.

Procedure

1. Enter system view.

system-view

2. Create an FTP template and enter its view.

nqa template ftp *name*

3. Specify an FTP login username.

username *username*

By default, no FTP login username is specified.

4. Specify an FTP login password.

password { **cipher** | **simple** } *sting*

By default, no FTP login password is specified.

5. Specify the source IP address for the probe packets.

IPv4:

source ip *ip-address*

By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

The source IP address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

IPv6:

source ipv6 *ipv6-address*

By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

6. Set the data transmission mode.

   mode { **active** | **passive** }

   The default mode is **active**.

7. Specify the FTP operation type.

   operation { **get** | **put** }

   By default, the FTP operation type is **get**, which means obtaining files from the FTP server.

8. Specify the destination URL for the FTP template.

   url *url*

   By default, no destination URL is specified for an FTP template.

   Enter the URL in one of the following formats:

   - ftp://*host*/*filename*.
   - ftp://*host*:*port*/*filename*.

   If the value for the *host* parameter is an IPv6 address, you must use a pair of square brackets ([]) to include the IPv6 address, for example, ftp://[2001::1]/test. When you perform the **get** operation, the file name is required.

   When you perform the **put** operation, the *filename* argument does not take effect, even if it is specified. The file name for the **put** operation is determined by using the **filename** command.

9. Specify the name of a file to be transferred.

   filename *filename*

   By default, no file is specified.

   This task is required only for the **put** operation.

   The configuration does not take effect for the **get** operation.

Configuring the RADIUS template

About this task

A feature that uses the RADIUS template performs the RADIUS authentication operation to check the availability of the authentication service on the RADIUS server.

The RADIUS authentication operation workflow is as follows:

1. The NQA client sends an authentication request (Access-Request) to the RADIUS server. The request includes the username and the password. The password is encrypted by using the MD5 algorithm and the shared key.
2. The RADIUS server authenticates the username and password.
   - If the authentication succeeds, the server sends an Access-Accept packet to the NQA client.
   - If the authentication fails, the server sends an Access-Reject packet to the NQA client.
3. The NQA client determines the availability of the authentication service on the RADIUS server based on the response packet it received:
   - If an Access-Accept packet is received, the authentication service is available on the RADIUS server.
   - If an Access-Reject packet is received, the authentication service is not available on the RADIUS server.

Prerequisites

Before you configure the RADIUS template, specify a username, password, and shared key on the RADIUS server. For more information about configuring the RADIUS server, see AAA in *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Create a RADIUS template and enter its view.

   nqa template radius *name*

3. Specify the destination IP address of the operation.

   IPv4:

   destination ip *ip-address*

   IPv6:

   destination ipv6 *ipv6-address*

   By default, no destination IP address is specified.

4. Specify the destination port number for the operation.

   destination port *port-number*

   By default, the destination port number is 1812.

5. Specify a username.

   username *username*

   By default, no username is specified.

6. Specify a password.

   password { **cipher** | **simple** } *string*

   By default, no password is specified.

7. Specify a shared key for secure RADIUS authentication.

   key { **cipher** | **simple** } *string*

   By default, no shared key is specified for RADIUS authentication.

8. Specify the source IP address for the probe packets.

   IPv4:

   source ip *ip-address*

   By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

   The source IP address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

   IPv6:

   source ipv6 *ipv6-address*

   By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

   The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

Configuring the SSL template

About this task

A feature that uses the SSL template performs the SSL operation to measure the time required to establish an SSL connection to an SSL server.

Prerequisites

Before you configure the SSL template, you must configure the SSL client policy. For information about configuring SSL client policies, see *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Create an SSL template and enter its view.

   nqa template ssl *name*
3. Specify the destination IP address of the operation.

   IPv4:

   destination ip *ip-address*

   IPv6:

   destination ipv6 *ipv6-address*

   By default, no destination IP address is specified.
4. Specify the destination port number for the operation.

   destination port *port-number*

   By default, the destination port number is not specified.
5. Specify an SSL client policy.

   ssl-client-policy *policy-name*

   By default, no SSL client policy is specified.
6. Specify the source IP address for the probe packets.

   IPv4:

   source ip *ip-address*

   By default, the primary IPv4 address of the output interface is used as the source IPv4 address of the probe packets.

   The source IP address must be the IPv4 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

   IPv6:

   source ipv6 *ipv6-address*

   By default, the primary IPv6 address of the output interface is used as the source IPv6 address of the probe packets.

   The source IPv6 address must be the IPv6 address of a local interface, and the interface must be up. Otherwise, no probe packets can be sent out.

Configuring optional parameters for the NQA template

Restrictions and guidelines

Unless otherwise specified, the following optional parameters apply to all types of NQA templates.

The parameter settings take effect only on the current NQA template.

Procedure

1. Enter system view.
   system-view
2. Enter the view of an existing NQA template.
   nqa template { **dns** | **ftp** | **http** | **https** | **icmp** | **radius** | **ssl** | **tcp** | **tcphalfopen** | **udp** } *name*
3. Configure a description.
   description *text*
   By default, no description is configured.
4. Set the interval at which the NQA operation repeats.
   frequency *interval*
   The default setting is 5000 milliseconds.
   If the operation is not completed when the interval expires, the next operation does not start.
5. Set the probe timeout time.
   probe timeout *timeout*
   The default setting is 3000 milliseconds.
6. Set the TTL for the probe packets.
   ttl *value*
   The default setting is 20.
   This command is not available for the ARP template.
7. Set the ToS value in the IP header of the probe packets.
   tos *value*
   The default setting is 0.
   This command is not available for the ARP template.
8. Specify the VPN instance where the operation is performed.
   vpn-instance *vpn-instance-name*
   By default, the operation is performed on the public network.
9. Set the number of consecutive successful probes to determine a successful operation event.
   reaction trigger probe-pass *count*
   The default setting is 3.
   If the number of consecutive successful probes for an NQA operation is reached, the NQA client notifies the feature that uses the template of the successful operation event.
10. Set the number of consecutive probe failures to determine an operation failure.
    reaction trigger probe-fail *count*
    The default setting is 3.
    If the number of consecutive probe failures for an NQA operation is reached, the NQA client notifies the feature that uses the NQA template of the operation failure.

Display and maintenance commands for NQA

Execute **display** commands in any view.

Task	Command
Display history records of NQA	**display nqa history** [*admin-name*

Task	Command
operations.	*operation-tag*]
Display the current monitoring results of reaction entries.	**display nqa reaction counters** [*admin-name operation-tag* [*item-number*]]
Display the most recent result of the NQA operation.	**display nqa result** [*admin-name operation-tag*]
Display NQA server status.	**display nqa server status**
Display NQA statistics.	**display nqa statistics** [*admin-name operation-tag*]

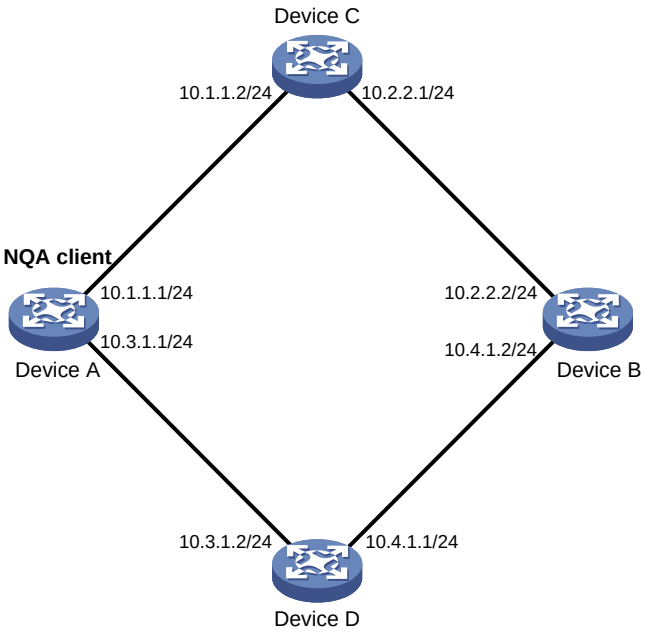
NQA configuration examples

Example: Configuring the ICMP echo operation

Network configuration

As shown in Figure 12, configure an ICMP echo operation on the NQA client (Device A) to test the round-trip time to Device B. The next hop of Device A is Device C.

Figure 12 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 12. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create an ICMP echo operation.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test1
[DeviceA-nqa-admin-test1] type icmp-echo
```

Specify 10.2.2.2 as the destination IP address of ICMP echo requests.

```
[DeviceA-nqa-admin-test1-icmp-echo] destination ip 10.2.2.2
```

Specify 10.1.1.2 as the next hop. The ICMP echo requests are sent through Device C to Device B.

```
[DeviceA-nqa-admin-test1-icmp-echo] next-hop ip 10.1.1.2
```

Configure the ICMP echo operation to perform 10 probes.

```
[DeviceA-nqa-admin-test1-icmp-echo] probe count 10
```

Set the probe timeout time to 500 milliseconds for the ICMP echo operation.

```
[DeviceA-nqa-admin-test1-icmp-echo] probe timeout 500
```

Configure the ICMP echo operation to repeat every 5000 milliseconds.

```
[DeviceA-nqa-admin-test1-icmp-echo] frequency 5000
```

Enable saving history records.

```
[DeviceA-nqa-admin-test1-icmp-echo] history-record enable
```

Set the maximum number of history records to 10.

```
[DeviceA-nqa-admin-test1-icmp-echo] history-record number 10
[DeviceA-nqa-admin-test1-icmp-echo] quit
```

Start the ICMP echo operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the ICMP echo operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the ICMP echo operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 10             Receive response times: 10
    Min/Max/Average round trip time: 2/5/3
    Square-Sum of round trip time: 96
    Last succeeded probe time: 2019-08-23 15:00:01.2
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the ICMP echo operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test) history records:
  Index      Response     Status           Time
  370        3            Succeeded        2019-08-23 15:00:01.2
  369        3            Succeeded        2019-08-23 15:00:01.2
  368        3            Succeeded        2019-08-23 15:00:01.2
  367        5            Succeeded        2019-08-23 15:00:01.2
  366        3            Succeeded        2019-08-23 15:00:01.2
  365        3            Succeeded        2019-08-23 15:00:01.2
  364        3            Succeeded        2019-08-23 15:00:01.1
  363        2            Succeeded        2019-08-23 15:00:01.1
  362        3            Succeeded        2019-08-23 15:00:01.1
  361        2            Succeeded        2019-08-23 15:00:01.1
```

The output shows that the packets sent by Device A can reach Device B through Device C. No packet loss occurs during the operation. The minimum, maximum, and average round-trip times are 2, 5, and 3 milliseconds, respectively.

Example: Configuring the ICMP jitter operation

Network configuration

As shown in Figure 13, configure an ICMP jitter operation to test the jitter between Device A and Device B.

Figure 13 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 13. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device A:

   # Create an ICMP jitter operation.

   ```
   <DeviceA> system-view
   [DeviceA] nqa entry admin test1
   [DeviceA-nqa-admin-test1] type icmp-jitter
   ```

   # Specify 10.2.2.2 as the destination address for the operation.

   ```
   [DeviceA-nqa-admin-test1-icmp-jitter] destination ip 10.2.2.2
   ```

   # Configure the operation to repeat every 1000 milliseconds.

   ```
   [DeviceA-nqa-admin-test1-icmp-jitter] frequency 1000
   [DeviceA-nqa-admin-test1-icmp-jitter] quit
   ```

   # Start the ICMP jitter operation.

   ```
   [DeviceA] nqa schedule admin test1 start-time now lifetime forever
   ```

   # After the ICMP jitter operation runs for a period of time, stop the operation.

   ```
   [DeviceA] undo nqa schedule admin test1
   ```

Verifying the configuration

Display the most recent result of the ICMP jitter operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 10              Receive response times: 10
    Min/Max/Average round trip time: 1/2/1
    Square-Sum of round trip time: 13
    Last packet received time: 2019-03-09 17:40:29.8
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packets out of sequence: 0
```

```
    Packets arrived late: 0
  ICMP-jitter results:
   RTT number: 10
    Min positive SD: 0                    Min positive DS: 0
    Max positive SD: 0                    Max positive DS: 0
    Positive SD number: 0                 Positive DS number: 0
    Positive SD sum: 0                    Positive DS sum: 0
    Positive SD average: 0                Positive DS average: 0
    Positive SD square-sum: 0             Positive DS square-sum: 0
    Min negative SD: 1                    Min negative DS: 2
    Max negative SD: 1                    Max negative DS: 2
    Negative SD number: 1                 Negative DS number: 1
    Negative SD sum: 1                    Negative DS sum: 2
    Negative SD average: 1                Negative DS average: 2
    Negative SD square-sum: 1             Negative DS square-sum: 4
    SD average: 1                         DS average: 2
  One way results:
    Max SD delay: 1                       Max DS delay: 2
    Min SD delay: 1                       Min DS delay: 2
    Number of SD delay: 1                 Number of DS delay: 1
    Sum of SD delay: 1                    Sum of DS delay: 2
    Square-Sum of SD delay: 1             Square-Sum of DS delay: 4
    Lost packets for unknown reason: 0
```

Display the statistics of the ICMP jitter operation.

```
[DeviceA] display nqa statistics admin test1
NQA entry (admin admin, tag test1) test statistics:
  NO. : 1
    Start time: 2019-03-09 17:42:10.7
    Life time: 156 seconds
    Send operation times: 1560          Receive response times: 1560
    Min/Max/Average round trip time: 1/2/1
    Square-Sum of round trip time: 1563
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packets out of sequence: 0
    Packets arrived late: 0
  ICMP-jitter results:
   RTT number: 1560
    Min positive SD: 1                    Min positive DS: 1
    Max positive SD: 1                    Max positive DS: 2
    Positive SD number: 18                Positive DS number: 46
    Positive SD sum: 18                   Positive DS sum: 49
    Positive SD average: 1                Positive DS average: 1
    Positive SD square-sum: 18            Positive DS square-sum: 55
    Min negative SD: 1                    Min negative DS: 1
```

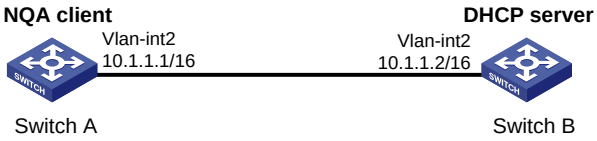
```
  Max negative SD: 1                    Max negative DS: 2
  Negative SD number: 24                Negative DS number: 57
  Negative SD sum: 24                   Negative DS sum: 58
  Negative SD average: 1                Negative DS average: 1
  Negative SD square-sum: 24            Negative DS square-sum: 60
  SD average: 16                        DS average: 2
One way results:
  Max SD delay: 1                       Max DS delay: 2
  Min SD delay: 1                       Min DS delay: 1
  Number of SD delay: 4                 Number of DS delay: 4
  Sum of SD delay: 4                    Sum of DS delay: 5
  Square-Sum of SD delay: 4             Square-Sum of DS delay: 7
  Lost packets for unknown reason: 0
```

Example: Configuring the DHCP operation

Network configuration

As shown in Figure 14, configure a DHCP operation to test the time required for Switch A to obtain an IP address from the DHCP server (Switch B).

Figure 14 Network diagram



Procedure

Create a DHCP operation.

```
<SwitchA> system-view
[SwitchA] nqa entry admin test1
[SwitchA-nqa-admin-test1] type dhcp
```

Specify the DHCP server address (10.1.1.2) as the destination address.

```
[SwitchA-nqa-admin-test1-dhcp] destination ip 10.1.1.2
```

Enable the saving of history records.

```
[SwitchA-nqa-admin-test1-dhcp] history-record enable
[SwitchA-nqa-admin-test1-dhcp] quit
```

Start the DHCP operation.

```
[SwitchA] nqa schedule admin test1 start-time now lifetime forever
```

After the DHCP operation runs for a period of time, stop the operation.

```
[SwitchA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the DHCP operation.

```
[SwitchA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 512/512/512
    Square-Sum of round trip time: 262144
```

```
    Last succeeded probe time: 2019-11-22 09:56:03.2
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the DHCP operation.

```
[SwitchA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
  Index      Response     Status           Time
  1          512          Succeeded        2019-11-22 09:56:03.2
```

The output shows that it took Switch A 512 milliseconds to obtain an IP address from the DHCP server.

Example: Configuring the DNS operation

Network configuration

As shown in Figure 15, configure a DNS operation to test whether Device A can perform address resolution through the DNS server and test the resolution time.

Figure 15 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 15. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create a DNS operation.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test1
[DeviceA-nqa-admin-test1] type dns
```

Specify the IP address of the DNS server (10.2.2.2) as the destination address.

```
[DeviceA-nqa-admin-test1-dns] destination ip 10.2.2.2
```

Specify **host.com** as the domain name to be translated.

```
[DeviceA-nqa-admin-test1-dns] resolve-target host.com
```

Enable the saving of history records.

```
[DeviceA-nqa-admin-test1-dns] history-record enable
[DeviceA-nqa-admin-test1-dns] quit
```

Start the DNS operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the DNS operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the DNS operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 62/62/62
    Square-Sum of round trip time: 3844
    Last succeeded probe time: 2019-11-10 10:49:37.3
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the DNS operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test) history records:
  Index      Response     Status           Time
  1          62           Succeeded        2019-11-10 10:49:37.3
```

The output shows that it took Device A 62 milliseconds to translate domain name **host.com** into an IP address.

Example: Configuring the FTP operation

Network configuration

As shown in Figure 16, configure an FTP operation to test the time required for Device A to upload a file to the FTP server. The login username and password are **admin** and **systemtest**, respectively. The file to be transferred to the FTP server is **config.txt**.

Figure 16 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 16. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create an FTP operation.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test1
[DeviceA-nqa-admin-test1] type ftp
```

Specify the URL of the FTP server.

```
[DeviceA-nqa-admin-test-ftp] url ftp://10.2.2.2
```

Specify 10.1.1.1 as the source IP address.

```
[DeviceA-nqa-admin-test1-ftp] source ip 10.1.1.1
```

Configure the device to upload file **config.txt** to the FTP server.

```
[DeviceA-nqa-admin-test1-ftp] operation put
[DeviceA-nqa-admin-test1-ftp] filename config.txt
```

Set the username to **admin** for the FTP operation.

```
[DeviceA-nqa-admin-test1-ftp] username admin
```

Set the password to **systemtest** for the FTP operation.

```
[DeviceA-nqa-admin-test1-ftp] password simple systemtest
```

Enable the saving of history records.

```
[DeviceA-nqa-admin-test1-ftp] history-record enable
[DeviceA-nqa-admin-test1-ftp] quit
```

Start the FTP operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the FTP operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the FTP operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 173/173/173
    Square-Sum of round trip time: 29929
    Last succeeded probe time: 2019-11-22 10:07:28.6
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to disconnect: 0
    Failures due to no connection: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the FTP operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
  Index      Response     Status           Time
  1          173          Succeeded        2019-11-22 10:07:28.6
```

The output shows that it took Device A 173 milliseconds to upload a file to the FTP server.

Example: Configuring the HTTP operation

Network configuration

As shown in Figure 17, configure an HTTP operation on the NQA client to test the time required to obtain data from the HTTP server.

Figure 17 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 17. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create an HTTP operation.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test1
[DeviceA-nqa-admin-test1] type http
```

Specify the URL of the HTTP server.

```
[DeviceA-nqa-admin-test1-http] url http://10.2.2.2/index.htm
```

Configure the HTTP operation to get data from the HTTP server.

```
[DeviceA-nqa-admin-test1-http] operation get
```

Configure the operation to use HTTP version 1.0.

```
[DeviceA-nqa-admin-test1-http] version v1.0
```

Enable the saving of history records.

```
[DeviceA-nqa-admin-test1-http] history-record enable
[DeviceA-nqa-admin-test1-http] quit
```

Start the HTTP operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the HTTP operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the HTTP operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 64/64/64
    Square-Sum of round trip time: 4096
    Last succeeded probe time: 2019-11-22 10:12:47.9
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to disconnect: 0
    Failures due to no connection: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the HTTP operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
  Index      Response     Status           Time
  1          64           Succeeded        2019-11-22 10:12:47.9
```

The output shows that it took Device A 64 milliseconds to obtain data from the HTTP server.

Example: Configuring the UDP jitter operation

Network configuration

As shown in Figure 18, configure a UDP jitter operation to test the jitter, delay, and round-trip time between Device A and Device B.

Figure 18 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 18. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device B:

   # Enable the NQA server.

   ```
   <DeviceB> system-view
   [DeviceB] nqa server enable
   ```

   # Configure a listening service to listen to UDP port 9000 on IP address 10.2.2.2.

   ```
   [DeviceB] nqa server udp-echo 10.2.2.2 9000
   ```

4. Configure Device A:

   # Create a UDP jitter operation.

   ```
   <DeviceA> system-view
   [DeviceA] nqa entry admin test1
   [DeviceA-nqa-admin-test1] type udp-jitter
   ```

   # Specify 10.2.2.2 as the destination address of the operation.

   ```
   [DeviceA-nqa-admin-test1-udp-jitter] destination ip 10.2.2.2
   ```

   # Set the destination port number to 9000.

   ```
   [DeviceA-nqa-admin-test1-udp-jitter] destination port 9000
   ```

   # Configure the operation to repeat every 1000 milliseconds.

   ```
   [DeviceA-nqa-admin-test1-udp-jitter] frequency 1000
   [DeviceA-nqa-admin-test1-udp-jitter] quit
   ```

   # Start the UDP jitter operation.

   ```
   [DeviceA] nqa schedule admin test1 start-time now lifetime forever
   ```

   # After the UDP jitter operation runs for a period of time, stop the operation.

   ```
   [DeviceA] undo nqa schedule admin test1
   ```

Verifying the configuration

Display the most recent result of the UDP jitter operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 10              Receive response times: 10
    Min/Max/Average round trip time: 15/32/17
    Square-Sum of round trip time: 3235
    Last packet received time: 2019-05-29 13:56:17.6
```

```
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packets out of sequence: 0
    Packets arrived late: 0
  UDP-jitter results:
   RTT number: 10
    Min positive SD: 4                    Min positive DS: 1
    Max positive SD: 21                   Max positive DS: 28
    Positive SD number: 5                 Positive DS number: 4
    Positive SD sum: 52                   Positive DS sum: 38
    Positive SD average: 10               Positive DS average: 10
    Positive SD square-sum: 754           Positive DS square-sum: 460
    Min negative SD: 1                    Min negative DS: 6
    Max negative SD: 13                   Max negative DS: 22
    Negative SD number: 4                 Negative DS number: 5
    Negative SD sum: 38                   Negative DS sum: 52
    Negative SD average: 10               Negative DS average: 10
    Negative SD square-sum: 460           Negative DS square-sum: 754
    SD average: 10                        DS average: 10
  One way results:
    Max SD delay: 15                      Max DS delay: 16
    Min SD delay: 7                       Min DS delay: 7
    Number of SD delay: 10                Number of DS delay: 10
    Sum of SD delay: 78                   Sum of DS delay: 85
    Square-Sum of SD delay: 666           Square-Sum of DS delay: 787
    SD lost packets: 0                  DS lost packets: 0
    Lost packets for unknown reason: 0
```

Display the statistics of the UDP jitter operation.

```
[DeviceA] display nqa statistics admin test1
NQA entry (admin admin, tag test1) test statistics:
  NO. : 1
    Start time: 2019-05-29 13:56:14.0
    Life time: 47 seconds
    Send operation times: 410           Receive response times: 410
    Min/Max/Average round trip time: 1/93/19
    Square-Sum of round trip time: 206176
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packets out of sequence: 0
    Packets arrived late: 0
  UDP-jitter results:
   RTT number: 410
```

```
  Min positive SD: 3                    Min positive DS: 1
  Max positive SD: 30                   Max positive DS: 79
  Positive SD number: 186               Positive DS number: 158
  Positive SD sum: 2602                 Positive DS sum: 1928
  Positive SD average: 13               Positive DS average: 12
  Positive SD square-sum: 45304         Positive DS square-sum: 31682
  Min negative SD: 1                    Min negative DS: 1
  Max negative SD: 30                   Max negative DS: 78
  Negative SD number: 181               Negative DS number: 209
  Negative SD sum: 181                  Negative DS sum: 209
  Negative SD average: 13               Negative DS average: 14
  Negative SD square-sum: 46994         Negative DS square-sum: 3030
  SD average: 9                         DS average: 1
One way results:
  Max SD delay: 46                      Max DS delay: 46
  Min SD delay: 7                       Min DS delay: 7
  Number of SD delay: 410               Number of DS delay: 410
  Sum of SD delay: 3705                 Sum of DS delay: 3891
  Square-Sum of SD delay: 45987         Square-Sum of DS delay: 49393
  SD lost packets: 0                  DS lost packets: 0
  Lost packets for unknown reason: 0
```

Example: Configuring the SNMP operation

Network configuration

As shown in Figure 19, configure an SNMP operation to test the time the NQA client uses to get a response from the SNMP agent.

Figure 19 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 19. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure the SNMP agent (Device B):

   # Set the SNMP version to **all**.

   ```
   <DeviceB> system-view
   [DeviceB] snmp-agent sys-info version all
   ```

   # Set the read community to **public**.

   ```
   [DeviceB] snmp-agent community read public
   ```

   # Set the write community to **private**.

   ```
   [DeviceB] snmp-agent community write private
   ```

4. Configure Device A:

   # Create an SNMP operation.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test1
[DeviceA-nqa-admin-test1] type snmp
```

Specify 10.2.2.2 as the destination IP address of the SNMP operation.

```
[DeviceA-nqa-admin-test1-snmp] destination ip 10.2.2.2
```

Enable the saving of history records.

```
[DeviceA-nqa-admin-test1-snmp] history-record enable
[DeviceA-nqa-admin-test1-snmp] quit
```

Start the SNMP operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the SNMP operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the SNMP operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 50/50/50
    Square-Sum of round trip time: 2500
    Last succeeded probe time: 2019-11-22 10:24:41.1
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the SNMP operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
  Index      Response     Status           Time
  1          50           Succeeded        2019-11-22 10:24:41.1
```

The output shows that it took Device A 50 milliseconds to receive a response from the SNMP agent.

Example: Configuring the TCP operation

Network configuration

As shown in Figure 20, configure a TCP operation to test the time required for Device A to establish a TCP connection with Device B.

Figure 20 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 20. (Details not shown.)

2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device B:

   # Enable the NQA server.

   ```
   <DeviceB> system-view
   [DeviceB] nqa server enable
   ```

   # Configure a listening service to listen to TCP port 9000 on IP address 10.2.2.2.

   ```
   [DeviceB] nqa server tcp-connect 10.2.2.2 9000
   ```

4. Configure Device A:

   # Create a TCP operation.

   ```
   <DeviceA> system-view
   [DeviceA] nqa entry admin test1
   [DeviceA-nqa-admin-test1] type tcp
   ```

   # Specify 10.2.2.2 as the destination IP address.

   ```
   [DeviceA-nqa-admin-test1-tcp] destination ip 10.2.2.2
   ```

   # Set the destination port number to 9000.

   ```
   [DeviceA-nqa-admin-test1-tcp] destination port 9000
   ```

   # Enable the saving of history records.

   ```
   [DeviceA-nqa-admin-test1-tcp] history-record enable
   [DeviceA-nqa-admin-test1-tcp] quit
   ```

   # Start the TCP operation.

   ```
   [DeviceA] nqa schedule admin test1 start-time now lifetime forever
   ```

   # After the TCP operation runs for a period of time, stop the operation.

   ```
   [DeviceA] undo nqa schedule admin test1
   ```

Verifying the configuration

Display the most recent result of the TCP operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 13/13/13
    Square-Sum of round trip time: 169
    Last succeeded probe time: 2019-11-22 10:27:25.1
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to disconnect: 0
    Failures due to no connection: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the TCP operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
  Index      Response     Status           Time
  1          13           Succeeded        2019-11-22 10:27:25.1
```

The output shows that it took Device A 13 milliseconds to establish a TCP connection to port 9000 on the NQA server.

Example: Configuring the UDP echo operation

Network configuration

As shown in Figure 21, configure a UDP echo operation on the NQA client to test the round-trip time to Device B. The destination port number is 8000.

Figure 21 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 21. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device B:

   # Enable the NQA server.

   ```
   <DeviceB> system-view
   [DeviceB] nqa server enable
   ```

   # Configure a listening service to listen to UDP port 8000 on IP address 10.2.2.2.

   ```
   [DeviceB] nqa server udp-echo 10.2.2.2 8000
   ```

4. Configure Device A:

   # Create a UDP echo operation.

   ```
   <DeviceA> system-view
   [DeviceA] nqa entry admin test1
   [DeviceA-nqa-admin-test1] type udp-echo
   ```

   # Specify 10.2.2.2 as the destination IP address.

   ```
   [DeviceA-nqa-admin-test1-udp-echo] destination ip 10.2.2.2
   ```

   # Set the destination port number to 8000.

   ```
   [DeviceA-nqa-admin-test1-udp-echo] destination port 8000
   ```

   # Enable the saving of history records.

   ```
   [DeviceA-nqa-admin-test1-udp-echo] history-record enable
   [DeviceA-nqa-admin-test1-udp-echo] quit
   ```

   # Start the UDP echo operation.

   ```
   [DeviceA] nqa schedule admin test1 start-time now lifetime forever
   ```

   # After the UDP echo operation runs for a period of time, stop the operation.

   ```
   [DeviceA] undo nqa schedule admin test1
   ```

Verifying the configuration

Display the most recent result of the UDP echo operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 25/25/25
    Square-Sum of round trip time: 625
    Last succeeded probe time: 2019-11-22 10:36:17.9
```

```
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the UDP echo operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
  Index      Response     Status           Time
  1          25           Succeeded        2019-11-22 10:36:17.9
```

The output shows that the round-trip time between Device A and port 8000 on Device B is 25 milliseconds.

Example: Configuring the UDP tracert operation

Network configuration

As shown in Figure 22, configure a UDP tracert operation to determine the routing path from Device A to Device B.

Figure 22 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 22. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Execute the **ip ttl-expires enable** command on the intermediate devices and execute the **ip unreachables enable** command on Device B.
4. Configure Device A:

   # Create a UDP tracert operation.

   ```
   <DeviceA> system-view
   [DeviceA] nqa entry admin test1
   [DeviceA-nqa-admin-test1] type udp-tracert
   ```

   # Specify 10.2.2.2 as the destination IP address.

   ```
   [DeviceA-nqa-admin-test1-udp-tracert] destination ip 10.2.2.2
   ```

   # Set the destination port number to 33434.

   ```
   [DeviceA-nqa-admin-test1-udp-tracert] destination port 33434
   ```

   # Configure Device A to perform three probes to each hop.

   ```
   [DeviceA-nqa-admin-test1-udp-tracert] probe count 3
   ```

   # Set the probe timeout time to 500 milliseconds.

   ```
   [DeviceA-nqa-admin-test1-udp-tracert] probe timeout 500
   ```

   # Configure the UDP tracert operation to repeat every 5000 milliseconds.

   ```
   [DeviceA-nqa-admin-test1-udp-tracert] frequency 5000
   ```

   # Specify GigabitEthernet 1/0/1 as the output interface for UDP packets.

```
[DeviceA-nqa-admin-test1-udp-tracert] out interface gigabitethernet 1/0/1
```

Enable the no-fragmentation feature.

```
[DeviceA-nqa-admin-test1-udp-tracert] no-fragment enable
```

Set the maximum number of consecutive probe failures to 6.

```
[DeviceA-nqa-admin-test1-udp-tracert] max-failure 6
```

Set the TTL value to 1 for UDP packets in the start round of the UDP tracert operation.

```
[DeviceA-nqa-admin-test1-udp-tracert] init-ttl 1
```

Start the UDP tracert operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the UDP tracert operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the UDP tracert operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 6              Receive response times: 6
    Min/Max/Average round trip time: 1/1/1
    Square-Sum of round trip time: 1
    Last succeeded probe time: 2019-09-09 14:46:06.2
  Extended results:
    Packet loss in test: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
Failures due to other errors: 0
  UDP-tracert results:
    TTL    Hop IP             Time
    1      3.1.1.1            2019-09-09 14:46:03.2
    2      10.2.2.2           2019-09-09 14:46:06.2
```

Display the history records of the UDP tracert operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
Index      TTL  Response  Hop IP           Status          Time
1          2    2         10.2.2.2         Succeeded       2019-09-09 14:46:06.2
1          2    1         10.2.2.2         Succeeded       2019-09-09 14:46:05.2
1          2    2         10.2.2.2         Succeeded       2019-09-09 14:46:04.2
1          1    1         3.1.1.1          Succeeded       2019-09-09 14:46:03.2
1          1    2         3.1.1.1          Succeeded       2019-09-09 14:46:02.2
1          1    1         3.1.1.1          Succeeded       2019-09-09 14:46:01.2
```

Example: Configuring the voice operation

Network configuration

As shown in Figure 23, configure a voice operation to test jitters, delay, MOS, and ICPIF between Device A and Device B.

Figure 23 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 23. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device B:

   # Enable the NQA server.

   ```
   <DeviceB> system-view
   [DeviceB] nqa server enable
   ```

   # Configure a listening service to listen to UDP port 9000 on IP address 10.2.2.2.

   ```
   [DeviceB] nqa server udp-echo 10.2.2.2 9000
   ```

4. Configure Device A:

   # Create a voice operation.

   ```
   <DeviceA> system-view
   [DeviceA] nqa entry admin test1
   [DeviceA-nqa-admin-test1] type voice
   ```

   # Specify 10.2.2.2 as the destination IP address.

   ```
   [DeviceA-nqa-admin-test1-voice] destination ip 10.2.2.2
   ```

   # Set the destination port number to 9000.

   ```
   [DeviceA-nqa-admin-test1-voice] destination port 9000
   [DeviceA-nqa-admin-test1-voice] quit
   ```

   # Start the voice operation.

   ```
   [DeviceA] nqa schedule admin test1 start-time now lifetime forever
   ```

   # After the voice operation runs for a period of time, stop the operation.

   ```
   [DeviceA] undo nqa schedule admin test1
   ```

Verifying the configuration

Display the most recent result of the voice operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1000          Receive response times: 1000
    Min/Max/Average round trip time: 31/1328/33
    Square-Sum of round trip time: 2844813
    Last packet received time: 2019-06-13 09:49:31.1
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
Packets out of sequence: 0
    Packets arrived late: 0
  Voice results:
```

```
   RTT number: 1000
    Min positive SD: 1                    Min positive DS: 1
    Max positive SD: 204                  Max positive DS: 1297
    Positive SD number: 257               Positive DS number: 259
    Positive SD sum: 759                  Positive DS sum: 1797
    Positive SD average: 2                Positive DS average: 6
    Positive SD square-sum: 54127         Positive DS square-sum: 1691967
    Min negative SD: 1                    Min negative DS: 1
    Max negative SD: 203                  Max negative DS: 1297
    Negative SD number: 255               Negative DS number: 259
    Negative SD sum: 759                  Negative DS sum: 1796
    Negative SD average: 2                Negative DS average: 6
    Negative SD square-sum: 53655         Negative DS square-sum: 1691776
    SD average: 2                         DS average: 6
  One way results:
    Max SD delay: 343                     Max DS delay: 985
    Min SD delay: 343                     Min DS delay: 985
    Number of SD delay: 1                 Number of DS delay: 1
    Sum of SD delay: 343                  Sum of DS delay: 985
    Square-Sum of SD delay: 117649        Square-Sum of DS delay: 970225
    SD lost packets: 0                   DS lost packets: 0
    Lost packets for unknown reason: 0
  Voice scores:
    MOS value: 4.38                       ICPIF value: 0
```

Display the statistics of the voice operation.

```
[DeviceA] display nqa statistics admin test1
NQA entry (admin admin, tag test1) test statistics:
  NO. : 1

    Start time: 2019-06-13 09:45:37.8
    Life time: 331 seconds
    Send operation times: 4000           Receive response times: 4000
    Min/Max/Average round trip time: 15/1328/32
    Square-Sum of round trip time: 7160528
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
Packets out of sequence: 0
    Packets arrived late: 0
  Voice results:
   RTT number: 4000
    Min positive SD: 1                    Min positive DS: 1
    Max positive SD: 360                  Max positive DS: 1297
    Positive SD number: 1030              Positive DS number: 1024
    Positive SD sum: 4363                 Positive DS sum: 5423
    Positive SD average: 4                Positive DS average: 5
```

```
  Positive SD square-sum: 497725          Positive DS square-sum: 2254957
  Min negative SD: 1                      Min negative DS: 1
  Max negative SD: 360                    Max negative DS: 1297
  Negative SD number: 1028                Negative DS number: 1022
  Negative SD sum: 1028                   Negative DS sum: 1022
  Negative SD average: 4                  Negative DS average: 5
  Negative SD square-sum: 495901          Negative DS square-sum: 5419
  SD average: 16                          DS average: 2
One way results:
  Max SD delay: 359                       Max DS delay: 985
  Min SD delay: 0                         Min DS delay: 0
  Number of SD delay: 4                   Number of DS delay: 4
  Sum of SD delay: 1390                   Sum of DS delay: 1079
  Square-Sum of SD delay: 483202          Square-Sum of DS delay: 973651
  SD lost packets: 0                    DS lost packets: 0
  Lost packets for unknown reason: 0
Voice scores:
  Max MOS value: 4.38                     Min MOS value: 4.38
  Max ICPIF value: 0                      Min ICPIF value: 0
```

Example: Configuring the DLSw operation

Network configuration

As shown in Figure 24, configure a DLSw operation to test the response time of the DLSw device.

Figure 24 Network diagram



Procedure

1. Configure Device A:

Assign IP addresses to interfaces, as shown in Figure 24. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create a DLSw operation.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test1
[DeviceA-nqa-admin-test1] type dlsw
```

Specify 10.2.2.2 as the destination IP address.

```
[DeviceA-nqa-admin-test1-dlsw] destination ip 10.2.2.2
```

Enable the saving of history records.

```
[DeviceA-nqa-admin-test1-dlsw] history-record enable
[DeviceA-nqa-admin-test1-dlsw] quit
```

Start the DLSw operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the DLSw operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

2. Configure Device B:

Assign IP addresses to interfaces. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Enable the NQA server.

```
<DeviceB> system-view
[DeviceB] nqa server enable
```

Configure a listening service to listen to TCP port 2065 on IP address 10.2.2.2. (For the DLSw operation, the port number for the TCP listening service must be 2065.)

```
[DeviceB] nqa server tcp-connect 10.2.2.2 2065
```

Verifying the configuration

Display the most recent result of the DLSw operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
    Send operation times: 1              Receive response times: 1
    Min/Max/Average round trip time: 19/19/19
    Square-Sum of round trip time: 361
    Last succeeded probe time: 2019-11-22 10:40:27.7
  Extended results:
    Packet loss ratio: 0%
    Failures due to timeout: 0
    Failures due to disconnect: 0
    Failures due to no connection: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
```

Display the history records of the DLSw operation.

```
[DeviceA] display nqa history admin test1
NQA entry (admin admin, tag test1) history records:
  Index      Response     Status           Time
  1          19           Succeeded        2019-11-22 10:40:27.7
```

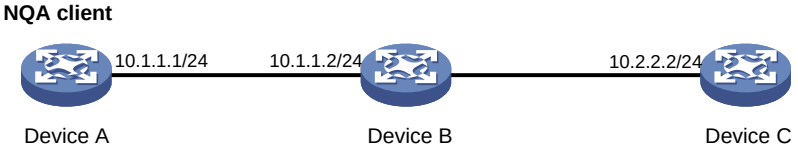
The output shows that the response time of the DLSw device is 19 milliseconds.

Example: Configuring the path jitter operation

Network configuration

As shown in Figure 25, configure a path jitter operation to test the round trip time and jitters from Device A to Device B and Device C.

Figure 25 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 25. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Execute the **ip ttl-expires enable** command on Device B and execute the **ip unreachables enable** command on Device C.

Create a path jitter operation.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test1
[DeviceA-nqa-admin-test1] type path-jitter
```

Specify 10.2.2.2 as the destination IP address of ICMP echo requests.

```
[DeviceA-nqa-admin-test1-path-jitter] destination ip 10.2.2.2
```

Configure the path jitter operation to repeat every 10000 milliseconds.

```
[DeviceA-nqa-admin-test1-path-jitter] frequency 10000
[DeviceA-nqa-admin-test1-path-jitter] quit
```

Start the path jitter operation.

```
[DeviceA] nqa schedule admin test1 start-time now lifetime forever
```

After the path jitter operation runs for a period of time, stop the operation.

```
[DeviceA] undo nqa schedule admin test1
```

Verifying the configuration

Display the most recent result of the path jitter operation.

```
[DeviceA] display nqa result admin test1
NQA entry (admin admin, tag test1) test results:
  Hop IP 10.1.1.2
    Basic Results
      Send operation times: 10              Receive response times: 10
      Min/Max/Average round trip time: 9/21/14
      Square-Sum of round trip time: 2419
    Extended Results
      Failures due to timeout: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packets out of sequence: 0
      Packets arrived late: 0
    Path-Jitter Results
      Jitter number: 9
        Min/Max/Average jitter: 1/10/4
      Positive jitter number: 6
        Min/Max/Average positive jitter: 1/9/4
        Sum/Square-Sum positive jitter: 25/173
      Negative jitter number: 3
        Min/Max/Average negative jitter: 2/10/6
        Sum/Square-Sum positive jitter: 19/153

  Hop IP 10.2.2.2
    Basic Results
```

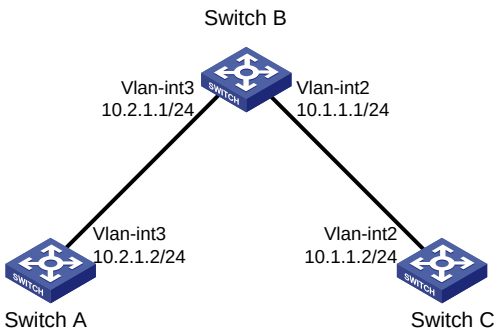
```
  Send operation times: 10              Receive response times: 10
  Min/Max/Average round trip time: 15/40/28
  Square-Sum of round trip time: 4493
Extended Results
  Failures due to timeout: 0
  Failures due to internal error: 0
  Failures due to other errors: 0
  Packets out of sequence: 0
  Packets arrived late: 0
Path-Jitter Results
  Jitter number: 9
    Min/Max/Average jitter: 1/10/4
  Positive jitter number: 6
    Min/Max/Average positive jitter: 1/9/4
    Sum/Square-Sum positive jitter: 25/173
  Negative jitter number: 3
    Min/Max/Average negative jitter: 2/10/6
    Sum/Square-Sum positive jitter: 19/153
```

Example: Configuring NQA collaboration

Network configuration

As shown in Figure 26, configure a static route to Switch C with Switch B as the next hop on Switch A. Associate the static route, a track entry, and an ICMP echo operation to monitor the state of the static route.

Figure 26 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 26. (Details not shown.)
2. On Switch A, configure a static route, and associate the static route with track entry 1.

   ```
   <SwitchA> system-view
   [SwitchA] ip route-static 10.1.1.2 24 10.2.1.1 track 1
   ```
3. On Switch A, configure an ICMP echo operation:

   # Create an NQA operation with administrator name **admin** and operation tag **test1**.

   ```
   [SwitchA] nqa entry admin test1
   ```

   # Configure the NQA operation type as ICMP echo.

   ```
   [SwitchA-nqa-admin-test1] type icmp-echo
   ```

   # Specify 10.2.1.1 as the destination IP address.

```
[SwitchA-nqa-admin-test1-icmp-echo] destination ip 10.2.1.1
```

Configure the operation to repeat every 100 milliseconds.

```
[SwitchA-nqa-admin-test1-icmp-echo] frequency 100
```

Create reaction entry 1. If the number of consecutive probe failures reaches 5, collaboration is triggered.

```
[SwitchA-nqa-admin-test1-icmp-echo] reaction 1 checked-element probe-fail
threshold-type consecutive 5 action-type trigger-only
[SwitchA-nqa-admin-test1-icmp-echo] quit
```

Start the ICMP operation.

```
[SwitchA] nqa schedule admin test1 start-time now lifetime forever
```

4. On Switch A, create track entry 1, and associate it with reaction entry 1 of the NQA operation.

```
[SwitchA] track 1 nqa entry admin test1 reaction 1
```

Verifying the configuration

Display information about all the track entries on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Positive
  Duration: 0 days 0 hours 0 minutes 0 seconds
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    NQA entry: admin test1
    Reaction: 1
```

Display brief information about active routes in the routing table on Switch A.

```
[SwitchA] display ip routing-table

Destinations : 13        Routes : 13

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.1.1.0/24         Static 60   0            10.2.1.1        Vlan3
10.2.1.0/24         Direct 0    0            10.2.1.2        Vlan3
10.2.1.0/32         Direct 0    0            10.2.1.2        Vlan3
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.255/32       Direct 0    0            10.2.1.2        Vlan3
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

The output shows that the static route with the next hop 10.2.1.1 is active, and the status of the track entry is positive.

Remove the IP address of VLAN-interface 3 on Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] undo ip address
```

Display information about all the track entries on Switch A.

```
[SwitchA] display track all
Track ID: 1
  State: Negative
  Duration: 0 days 0 hours 0 minutes 0 seconds
  Notification delay: Positive 0, Negative 0 (in seconds)
  Tracked object:
    NQA entry: admin test1
    Reaction: 1
```

Display brief information about active routes in the routing table on Switch A.

```
[SwitchA] display ip routing-table

Destinations : 12        Routes : 12

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
0.0.0.0/32          Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Direct 0    0            10.2.1.2        Vlan3
10.2.1.0/32         Direct 0    0            10.2.1.2        Vlan3
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.255/32       Direct 0    0            10.2.1.2        Vlan3
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/32        Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
127.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
224.0.0.0/4         Direct 0    0            0.0.0.0         NULL0
224.0.0.0/24        Direct 0    0            0.0.0.0         NULL0
255.255.255.255/32  Direct 0    0            127.0.0.1       InLoop0
```

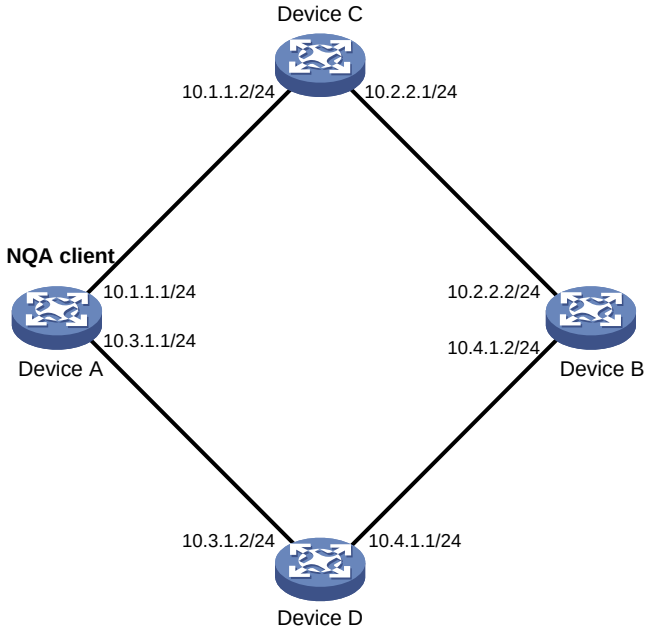
The output shows that the static route does not exist, and the status of the track entry is negative.

Example: Configuring the ICMP template

Network configuration

As shown in Figure 27, configure an ICMP template for a feature to perform the ICMP echo operation from Device A to Device B.

Figure 27 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 27. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create ICMP template **icmp**.

```
<DeviceA> system-view
[DeviceA] nqa template icmp icmp
```

Specify 10.2.2.2 as the destination IP address of ICMP echo requests.

```
[DeviceA-nqatplt-icmp-icmp] destination ip 10.2.2.2
```

Set the probe timeout time to 500 milliseconds for the ICMP echo operation.

```
[DeviceA-nqatplt-icmp-icmp] probe timeout 500
```

Configure the ICMP echo operation to repeat every 3000 milliseconds.

```
[DeviceA-nqatplt-icmp-icmp] frequency 3000
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-icmp-icmp] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

```
[DeviceA-nqatplt-icmp-icmp] reaction trigger probe-fail 2
```

Example: Configuring the DNS template

Network configuration

As shown in Figure 28, configure a DNS template for a feature to perform the DNS operation. The operation tests whether Device A can perform the address resolution through the DNS server.

Figure 28 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 28. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create DNS template **dns**.

```
<DeviceA> system-view
[DeviceA] nqa template dns dns
```

Specify the IP address of the DNS server (10.2.2.2) as the destination IP address.

```
[DeviceA-nqatplt-dns-dns] destination ip 10.2.2.2
```

Specify **host.com** as the domain name to be translated.

```
[DeviceA-nqatplt-dns-dns] resolve-target host.com
```

Set the domain name resolution type to type A.

```
[DeviceA-nqatplt-dns-dns] resolve-type A
```

Specify 3.3.3.3 as the expected IP address.

```
[DeviceA-nqatplt-dns-dns] expect ip 3.3.3.3
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-dns-dns] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

```
[DeviceA-nqatplt-dns-dns] reaction trigger probe-fail 2
```

Example: Configuring the TCP template

Network configuration

As shown in Figure 29, configure a TCP template for a feature to perform the TCP operation. The operation tests whether Device A can establish a TCP connection to Device B.

Figure 29 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 29. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device B:

   # Enable the NQA server.

```
<DeviceB> system-view
[DeviceB] nqa server enable
```

Configure a listening service to listen to TCP port 9000 on IP address 10.2.2.2.

```
[DeviceB] nqa server tcp-connect 10.2.2.2 9000
```

4. Configure Device A:

Create TCP template **tcp**.

```
<DeviceA> system-view
[DeviceA] nqa template tcp tcp
```

Specify 10.2.2.2 as the destination IP address.

```
[DeviceA-nqatplt-tcp-tcp] destination ip 10.2.2.2
```

Set the destination port number to 9000.

```
[DeviceA-nqatplt-tcp-tcp] destination port 9000
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-tcp-tcp] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

```
[DeviceA-nqatplt-tcp-tcp] reaction trigger probe-fail 2
```

Example: Configuring the TCP half open template

Network configuration

As shown in Figure 30, configure a TCP half open template for a feature to test whether Device B can provide the TCP service for Device A.

Figure 30 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 30. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device A:

Create TCP half open template **test**.

```
<DeviceA> system-view
[DeviceA] nqa template tcphalfopen test
```

Specify 10.2.2.2 as the destination IP address.

```
[DeviceA-nqatplt-tcphalfopen-test] destination ip 10.2.2.2
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-tcphalfopen-test] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

```
[DeviceA-nqatplt-tcphalfopen-test] reaction trigger probe-fail 2
```

Example: Configuring the UDP template

Network configuration

As shown in Figure 31, configure a UDP template for a feature to perform the UDP operation. The operation tests whether Device A can receive a response from Device B.

Figure 31 Network diagram



Procedure

1. Assign IP addresses to interfaces, as shown in Figure 31. (Details not shown.)
2. Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)
3. Configure Device B:

   # Enable the NQA server.

   ```
   <DeviceB> system-view
   [DeviceB] nqa server enable
   ```

   # Configure a listening service to listen to UDP port 9000 on IP address 10.2.2.2.

   ```
   [DeviceB] nqa server udp-echo 10.2.2.2 9000
   ```

4. Configure Device A:

   # Create UDP template **udp**.

   ```
   <DeviceA> system-view
   [DeviceA] nqa template udp udp
   ```

   # Specify 10.2.2.2 as the destination IP address.

   ```
   [DeviceA-nqatplt-udp-udp] destination ip 10.2.2.2
   ```

   # Set the destination port number to 9000.

   ```
   [DeviceA-nqatplt-udp-udp] destination port 9000
   ```

   # Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

   ```
   [DeviceA-nqatplt-udp-udp] reaction trigger probe-pass 2
   ```

   # Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

   ```
   [DeviceA-nqatplt-udp-udp] reaction trigger probe-fail 2
   ```

Example: Configuring the HTTP template

Network configuration

As shown in Figure 32, configure an HTTP template for a feature to perform the HTTP operation. The operation tests whether the NQA client can get data from the HTTP server.

Figure 32 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 32. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create HTTP template **http**.

```
<DeviceA> system-view
[DeviceA] nqa template http http
```

Specify **http://10.2.2.2/index.htm** as the URL of the HTTP server.

```
[DeviceA-nqatplt-http-http] url http://10.2.2.2/index.htm
```

Set the HTTP operation type to **get**.

```
[DeviceA-nqatplt-http-http] operation get
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-http-http] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

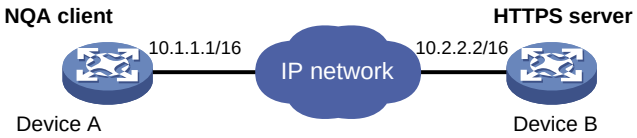
```
[DeviceA-nqatplt-http-http] reaction trigger probe-fail 2
```

Example: Configuring the HTTPS template

Network configuration

As shown in Figure 33, configure an HTTPS template for a feature to test whether the NQA client can get data from the HTTPS server (Device B).

Figure 33 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 33. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Configure an SSL client policy named **abc** on Device A, and make sure Device A can use the policy to connect to the HTTPS server. (Details not shown.)

Create HTTPS template **test**.

```
<DeviceA> system-view
[DeviceA] nqa template https https
```

Specify **http://10.2.2.2/index.htm** as the URL of the HTTPS server.

```
[DeviceA-nqatplt-https-https] url https://10.2.2.2/index.htm
```

Specify SSL client policy **abc** for the HTTPS template.

```
[DeviceA-nqatplt-https- https] ssl-client-policy abc
```

Set the HTTPS operation type to **get** (the default HTTPS operation type).

```
[DeviceA-nqatplt-https-https] operation get
```

Set the HTTPS version to 1.0 (the default HTTPS version).

```
 [DeviceA-nqatplt-https-https] version v1.0
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-https-https] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

```
[DeviceA-nqatplt-https-https] reaction trigger probe-fail 2
```

Example: Configuring the FTP template

Network configuration

As shown in Figure 34, configure an FTP template for a feature to perform the FTP operation. The operation tests whether Device A can upload a file to the FTP server. The login username and password are **admin** and **systemtest**, respectively. The file to be transferred to the FTP server is **config.txt**.

Figure 34 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 34. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Create FTP template **ftp**.

```
<DeviceA> system-view
[DeviceA] nqa template ftp ftp
```

Specify the URL of the FTP server.

```
[DeviceA-nqatplt-ftp-ftp] url ftp://10.2.2.2
```

Specify 10.1.1.1 as the source IP address.

```
[DeviceA-nqatplt-ftp-ftp] source ip 10.1.1.1
```

Configure the device to upload file **config.txt** to the FTP server.

```
[DeviceA-nqatplt-ftp-ftp] operation put
[DeviceA-nqatplt-ftp-ftp] filename config.txt
```

Set the username to **admin** for the FTP server login.

```
[DeviceA-nqatplt-ftp-ftp] username admin
```

Set the password to **systemtest** for the FTP server login.

```
[DeviceA-nqatplt-ftp-ftp] password simple systemtest
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-ftp-ftp] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

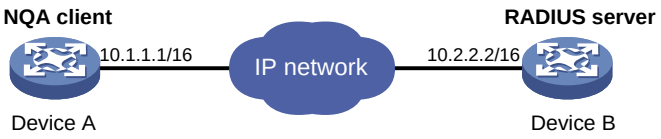
```
[DeviceA-nqatplt-ftp-ftp] reaction trigger probe-fail 2
```

Example: Configuring the RADIUS template

Network configuration

As shown in Figure 35, configure a RADIUS template for a feature to test whether the RADIUS server (Device B) can provide authentication service for Device A. The username and password are **admin** and **systemtest**, respectively. The shared key is **123456** for secure RADIUS authentication.

Figure 35 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 35. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Configure the RADIUS server. (Details not shown.)

Create RADIUS template **radius**.

```
<DeviceA> system-view
[DeviceA] nqa template radius radius
```

Specify 10.2.2.2 as the destination IP address of the operation.

```
[DeviceA-nqatplt-radius-radius] destination ip 10.2.2.2
```

Set the username to **admin**.

```
[DeviceA-nqatplt-radius-radius] username admin
```

Set the password to **systemtest**.

```
[DeviceA-nqatplt-radius-radius] password simple systemtest
```

Set the shared key to **123456** in plain text for secure RADIUS authentication.

```
[DeviceA-nqatplt-radius-radius] key simple 123456
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-radius-radius] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

```
[DeviceA-nqatplt-radius-radius] reaction trigger probe-fail 2
```

Example: Configuring the SSL template

Network configuration

As shown in Figure 36, configure an SSL template for a feature to test whether Device A can establish an SSL connection to the SSL server on Device B.

Figure 36 Network diagram



Procedure

Assign IP addresses to interfaces, as shown in Figure 36. (Details not shown.)

Configure static routes or a routing protocol to make sure the devices can reach each other. (Details not shown.)

Configure an SSL client policy named **abc** on Device A, and make sure Device A can use the policy to connect to the SSL server on Device B. (Details not shown.)

Create SSL template **ssl**.

```
<DeviceA> system-view
[DeviceA] nqa template ssl ssl
```

Set the destination IP address and port number to **10.2.2.2** and **9000**, respectively.

```
[DeviceA-nqatplt-ssl-ssl] destination ip 10.2.2.2
[DeviceA-nqatplt-ssl-ssl] destination port 9000
```

Specify SSL client policy **abc** for the SSL template.

```
[DeviceA-nqatplt-ssl-ssl] ssl-client-policy abc
```

Configure the NQA client to notify the feature of the successful operation event if the number of consecutive successful probes reaches 2.

```
[DeviceA-nqatplt-ssl-ssl] reaction trigger probe-pass 2
```

Configure the NQA client to notify the feature of the operation failure if the number of consecutive failed probes reaches 2.

```
[DeviceA-nqatplt-ssl-ssl] reaction trigger probe-fail 2
```

Contents

Configuring iNQA 1

About iNQA 1
iNQA benefits 1
Basic concepts 1
Application scenarios 4
Operating mechanism 6
Restrictions and guidelines: iNQA configuration 7
Instance restrictions and guidelines 7
Network and target flow requirements 7
Collaborating with other features 8
iNQA tasks at a glance 8
Prerequisites 9
Configuring a collector 9
Configuring parameters for a collector 9
Configuring a collector instance 9
Configuring an MP 11
Enabling packet loss measurement 11
Configuring the analyzer 12
Configuring parameters for the analyzer 12
Configuring an analyzer instance 12
Configuring an AMS 13
Enabling the measurement functionality 13
Configuring iNQA logging 13
Display and maintenance commands for iNQA collector 14
Display and maintenance commands for iNQA analyzer 14
iNQA configuration examples 15
Example: Configuring an end-to-end iNQA packet loss measurement 15
Example: Configuring an point-to-point iNQA packet loss measurement 19

Configuring iNQA

About iNQA

Intelligent Network Quality Analyzer (iNQA) is an H3C-proprietary protocol that allows you to measure network performance quickly in large-scale IP networks. In the current software version, iNQA supports measuring packet loss on forward, backward, and bidirectional flows. The packet loss data includes number of lost packets, packet loss rate, number of lost bytes, and byte loss rate. The measurement results help you to know when and where the packet loss occurs and the event severity level.

iNQA benefits

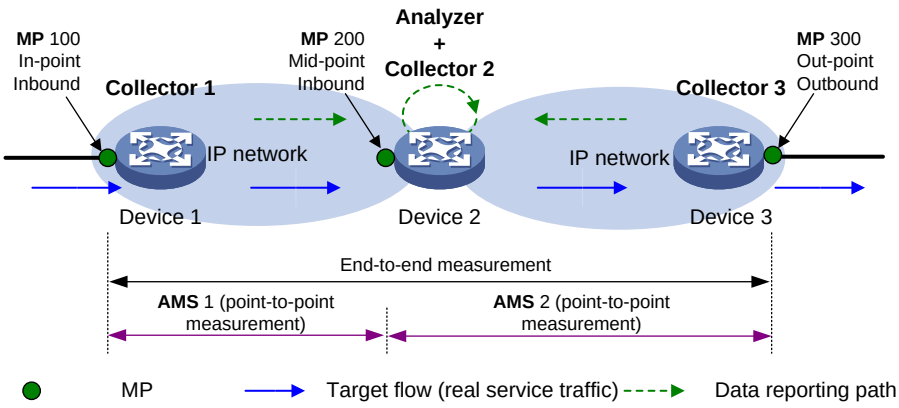
iNQA provides the following benefits:

- **True measurement results**—iNQA measures the service packets directly to calculate packet loss results, thus reflecting the real network quality.
- **Wide application range**—Applicable to Layer 2 network and Layer 3 IP network. iNQA supports the network-level and direct link measurement flexibly.
- **Fast fault location**—iNQA obtains the packet loss time, packet loss location, and number of lost packets in real time.
- **Applicable to different applications**—You can apply iNQA to multiple scenarios, such as point-to-point, point-to-multipoint, and multipoint-to-multipoint.

Basic concepts

Figure 1 shows the important iNQA concepts including MP, collector, analyzer, and AMS.

Figure 1 Network diagram



Collector

The collector manages MPs, collects data from MPs, and reports the data to the analyzer.

Analyzer

The analyzer collects the data from collector instances and summarizes the data.

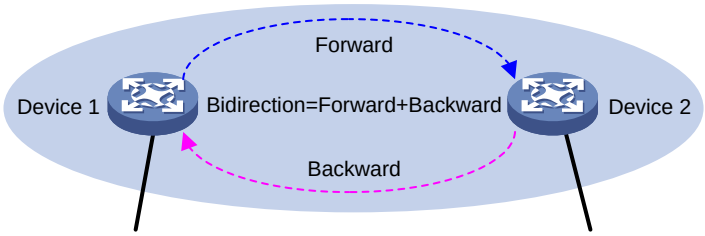
A device supports both the collector and analyzer functionalities. You can enable the collector and analyzer functionalities on different devices or the same device.

Target flow

A target flow is vital for iNQA measurement. You can specify a flow by using any combination of the following items: source IPv4 address/segment, destination IPv4 address/segment, protocol type, source port number, destination port number, and DSCP value. Using more items defines a more explicit flow and generates more accurate analysis data.

iNQA measures the flow according to the flow direction. After you define a forward flow, the flow in the opposite direction is a backward flow. The bidirectional flows refers to a forward flow and a backward flow. As shown in Figure 2, if you define the flow from Device 1 to Device 2 as the forward flow, then a flow from Device 2 to Device 1 is the backward flow. If you want to measure the packet loss on forward and backward flows between Device 1 and Device 2, you can specify bidirectional flows. The intermediate devices between the ingress and egress devices of the bidirectional flows can be the same or different.

Figure 2 Flow direction



MP

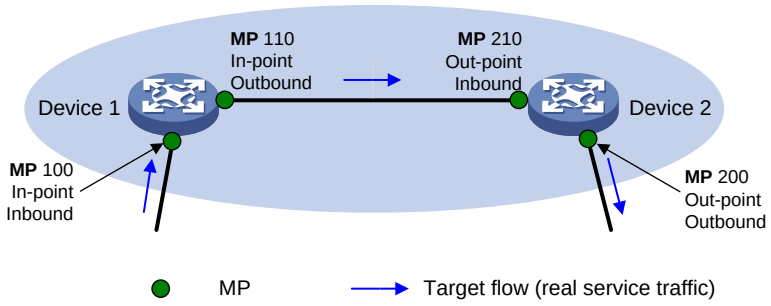
MP is a logical concept. An MP counts statistics and generates data for a flow. To measure packet loss on an interface on a collector, an MP must be bound to the interface.

An MP contains the following attributes:

- Measurement location of the flow.
  - An ingress point refers to the point that the flow enters the network.
  - An egress point refers to the point that the flow leaves the network.
  - A middle point refers to the point between an ingress point and egress point.
- Flow direction on the measurement point.

  A flow entering the MP is an inbound flow, and a flow leaving the MP is an outbound flow.
  - As shown in Figure 3, MP 100 is marked as in-point/inbound. MP 100 is the ingress point of the flow and the flow enters MP 100.
  - As shown in Figure 3, MP 110 is marked as in-point/outbound. MP 110 is the ingress point of the flow and the flow leaves MP 110.
  - As shown in Figure 3, MP 200 is marked as out-point/outbound. MP 200 is the egress point of the flow and the flow leaves MP 200.
  - As shown in Figure 3, MP 210 is marked as out-point/inbound. MP 210 is the egress point of the flow and the flow enters MP 210.

Figure 3 MP



AMS

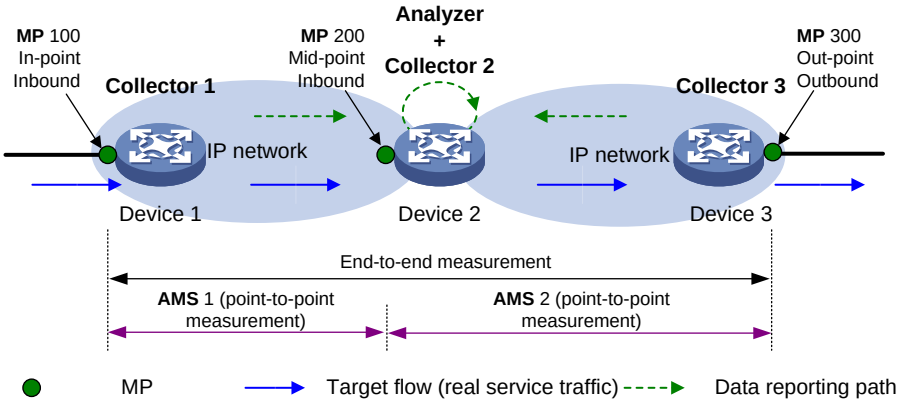
Configured on the analyzer, an AMS defines a measurement span for point-to-point performance measurement. You can configure multiple AMSs for an instance, and each AMS can be bound to MPs on any collector of the same instance. Therefore, iNQA can measure and summarize the data of the forward flow, backward flow, or bidirectional flows in any AMS.

Each AMS has an ingress MP group and egress MP group. The ingress MP group is the set of the ingress MPs in the AMS and the egress MP group is the set of the egress MPs.

As shown in Figure 4:

- To measure the packet loss between MP 100 and MP 300, AMS is not needed.
- If packet loss occurs between MP 100 and MP 300, configure AMS 1 and AMS 2 on the analyzer to locate on which span the packet loss occurs.
  - Bind MP 100 of collector 1 and MP 200 of collector 2 to AMS 1. Add MP 100 to the ingress MP group of the AMS 1 and MP 200 to the egress MP group of the AMS 1.
  - Bind MP 200 of collector 2 and MP 300 of collector 3 to AMS 2. Add MP 200 to the ingress MP group of the AMS 2 and MP 300 to the egress MP group of the AMS 2.

Figure 4 Network diagram



Instance

The instance allows measurement on a per-flow basis. In an instance, you can configure the target flow, flow direction, MPs, and measurement interval.

On the collector and analyzer, create an instance of the same ID for the same target flow. An instance can be bound to only one target flow. On the same device, you can configure multiple instances to measure and collect the packet loss rate of different target flows.

Flag bit

Flag bits, also called color bits, are used to distinguish target flows from unintended traffic flows.

iNQA uses ToS field bit 5 in the IPv4 packet header as the flag bit.

The ToS field consist of a 6-bit (bits 0 to 5) DSCP filed and a 2-bit (bits 6 to 7) ECN field. If bit 5 is used as the flag bit, to avoid inaccurate packet loss statistics, do not use it for DSCP purposes.

Application scenarios

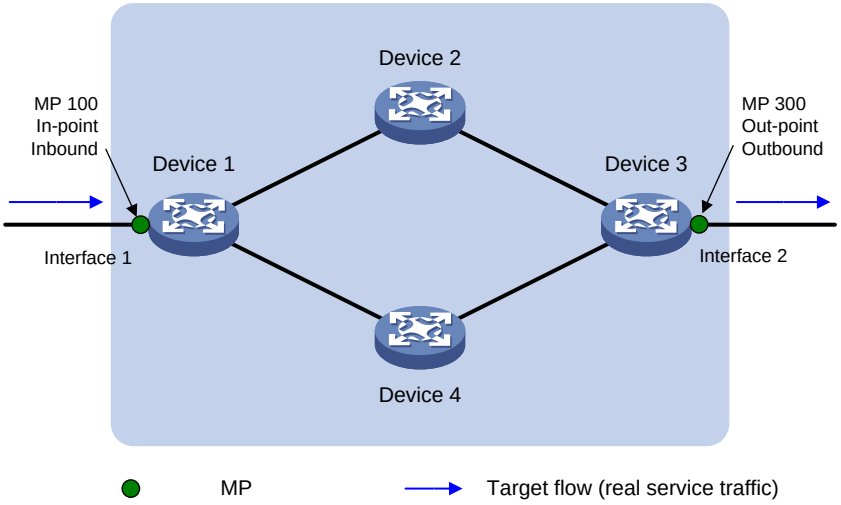
End-to-end packet loss measurement

iNQA measures whether packet loss occurs between the ingress points (where the target flow enters the IP network) and the egress points (where the flow leaves the network).

- Scenario 1: The target flow has only one ingress point and one egress point, and the two points are on the same network, for example, the local area network or the core network.

  As shown in Figure 5, to measure packet loss for the flow in the network, deploy iNQA on Device 1 and Device 3.

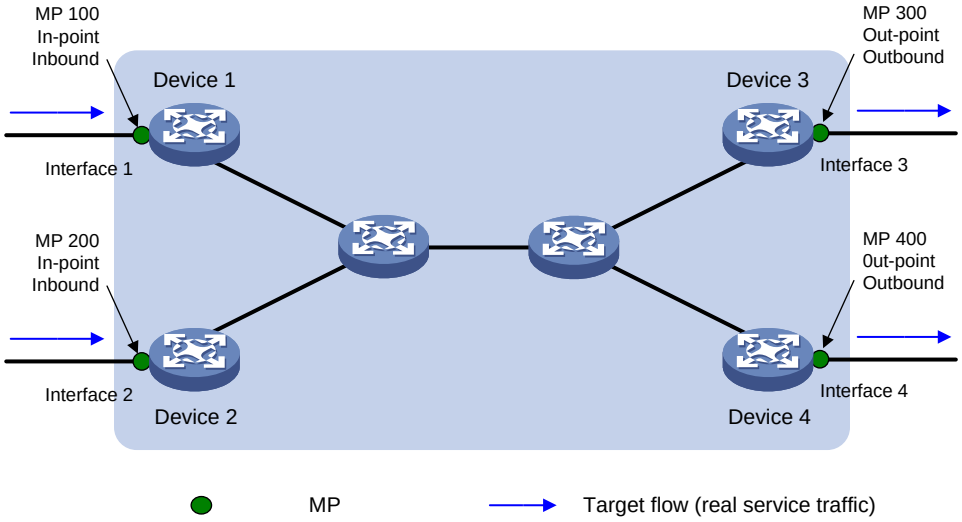
  Figure 5 Scenario 1



- Scenario 2: The target flow can have multiple ingress points and egress points, and the points are on the same network, for example, the local area network or the core network.

  As shown in Figure 6, to measure packet loss for the flow in the network, deploy iNQA on Device 1, Device 2, Device 3, and Device 4.

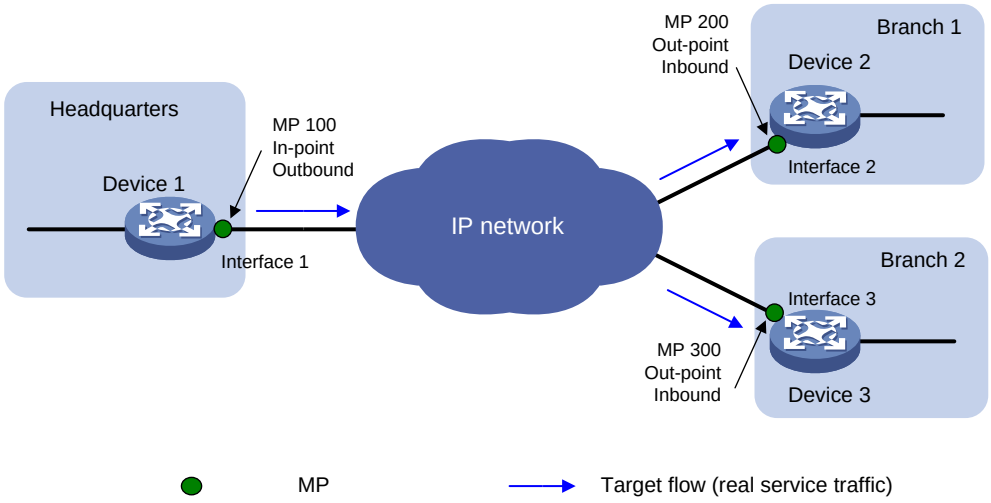
Figure 6 Scenario 2



- Scenario 3: The ingress points and the egress points of a target flow are on the different networks.

  As shown in Figure 7, to measure whether packet loss occurs when the flow crosses over the IP network, deploy iNQA on the egress devices of the headquarters and branches to measure.

  Figure 7 Scenario 3

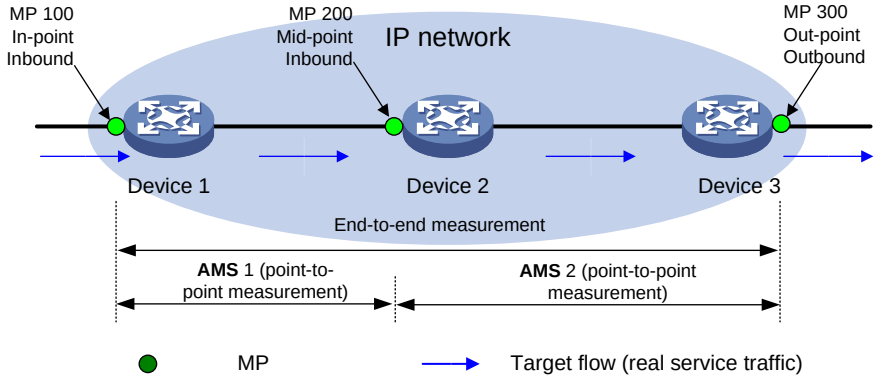


Point-to-point packet loss measurement

This measurement is based on atomic measurement spans (AMSs), which define smaller ranges than the end-to-end measurement. This method is helpful if you want to find out between which specific devices the packet loss occurs.

As shown in Figure 8, to configure point-to-point measurement, define AMSs, specify the flow direction in each AMS, and configure the ingress MP group and egress MP group for each AMS.

Figure 8 Point-to-point packet loss measurement



Operating mechanism

iNQA uses the model of multi-point collection and single-point calculation. Multiple collectors collect and report the packet data periodically and one analyzer calculates the data periodically.

Before starting the iNQA packet loss measurement, make sure all collectors are time synchronized through NTP or PTP. Therefore, all collectors can use the same measurement interval to color the flow and report the packet statistics to the analyzer. As a best practice, the analyzer and all collectors are time synchronized to facilitate management and maintenance. For more information about NTP, see "Configuring NTP." For more information about PTP, see "Configuring PTP."

To enable only packet loss measurement, you can synchronize time through NTP or PTP.

iNQA packet loss measurement mechanism

The number of incoming packets and that of outgoing packets in a network should be equal within a time period. If they are not equal, packet loss occurs in the network.

As shown in Figure 9, the network uses an external clock source and synchronizes time through PTP. The flow enters the network from MP 100, passes through MP 200, and leaves the network from MP 300. The devices where the flow passes are collectors and NTP clients, and the aggregation device is the analyzer and NTP server.

The iNQA packet loss measurement works as follows:

1. The analyzer synchronizes the time with all collectors through PTP.
2. The ingress MP on collector 1 identifies the target flow. It colors and decolors the packets in the flow alternatively at intervals, and periodically reports the packet statistics to the analyzer.
3. The middle MP on collector 2 identifies the target flow and reports the packet statistics to the analyzer periodically.
4. The egress MP on collector 3 identifies the target flow. It decolors the colored packets and reports the packet statistics to the analyzer periodically.
5. The analyzer calculates packet loss for the flow of the same period and same instance as follows:

   Number of lost packets = Number of incoming packets on the MP – Number of outgoing packets on the MP

   Packet loss rate = (Number of incoming packets on the MP – Number of outgoing packets on the MP) / Number of incoming packets on the MP

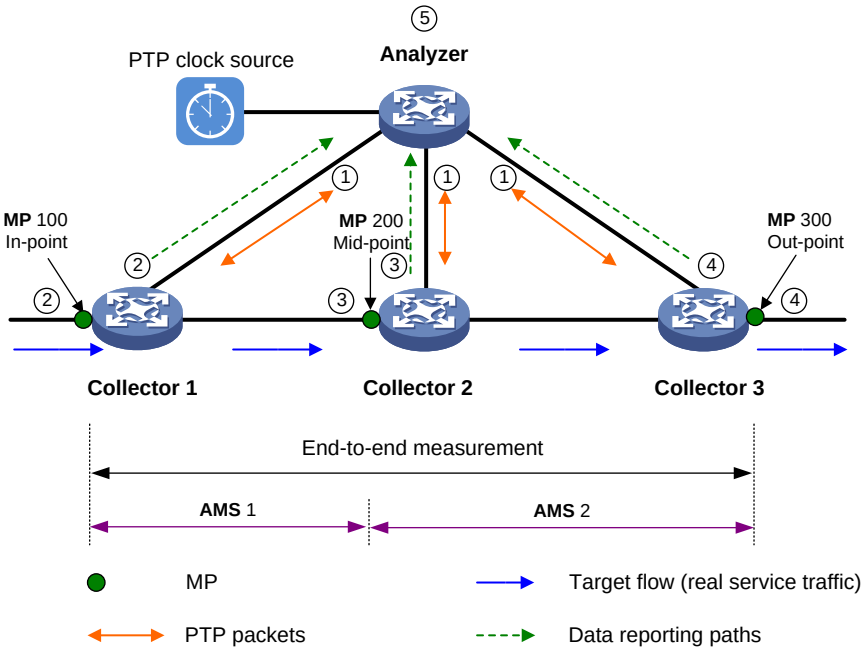
   The analyzer calculates the byte loss in the similar way the packet loss is calculated.

For the end-to-end measurement, the data from the ingress MP and egress MP is used.

For the point-to-point measurement, the analyzer calculates the result on a per-AMS basis.

- In AMS 1: Packet loss = Number of packets at MP 100 – Number of packets at MP 200
- In AMS 2: Packet loss = Number of packets at MP 200 – Number of packets at MP 300

Figure 9 Network diagram



Restrictions and guidelines: iNQA configuration

Instance restrictions and guidelines

If an analyzer instance is not correctly bound to a collector, or the bound collector does not report the data on time, the analyzer will uses 0 to calculate the packet loss rate. For example, if an analyzer instance is bound to only ingress MPs but no egress MPs, the packet loss rate is 100 % because the analyzer uses 0 as the packet count on egress MPs.

A collector is uniquely identified by its ID. For the same collector, the specified collector ID must be the same as the collector ID bound to an analyzer instance on the analyzer. The collector ID is the IPv4 address of the collector, and must be routable from the analyzer. As a best practice, configure the Router ID of the device as the collector ID.

An analyzer is uniquely identified by its ID. For the same analyzer, the specified analyzer ID must be the same as the analyzer ID bound with a collector instance on the collector. The analyzer ID is the IPv4 address of the analyzer, and must be routable from the collector. As a best practice, configure the Router ID of the device as the analyzer ID.

To measure the same flow, configure the same instance ID, the target flow attributes, and measurement interval on the analyzer and collectors.

Network and target flow requirements

To measure the same target flow, make sure the physical interfaces bound to all MPs are in the same type of networks. For example, the interfaces are all in IP networks or VXLAN networks.

- iNQA measures the numbers of packets and bytes, including the packet header and body parts, in the target flow. When a flow travels through different networks, the packet header might be

modified, making the packet count and byte count for the entire packet different at different MPs.

- iNQA identifies the target flow based on the value of a specific field in the packet header. If the packets are encapsulated or decapsulated during the transmission over different networks, iNQA cannot identify the target flow at different MPs for measurement.

The measured packets are known IPv4 unicast and multicast packets. For unknown IP unicast and broadcast packets, one packet entering an MP might be copied as multiple packets leaving the MP.

Collaborating with other features

iNQA on an aggregate interface might function incorrectly after a new member port joins the aggregation group and causes insufficient ACL resources. To solve the problem, follow either of the methods:

- Reduce the number of member ports.
- Use the **display qos-acl resource** command to view the QoS and ACL resource usage. Delete unnecessary settings to release ACL resources and then add a new member port again.

  For more information about the **display qos-acl resource** command, see ACL commands in *ACL and QoS Command Reference*.

With iNQA enabled on a Layer 2 aggregate interface, do not execute the **port s-mlag group** command to assign the Layer 2 aggregate interface to an S-MLAG group. For more information about the **port s-mlag group** command, see Ethernet link aggregation commands in *Layer 2—LAN Switching Command Reference*.

To specify ToS field bit 6 or 7 as the flag bit, do not enable ECN, and vice versa. For more information about ECN, see congestion avoidance configuration in *ACL and QoS Configuration Guide*.

If the device contains multiple MPs with different roles per iNQA instance, and the interfaces of the MPs are configured with QoS inbound rate limiting, the rate limiting feature cannot take effect on the interfaces. For more information about rate limiting, see rate limit configuration in *ACL and QoS Configuration Guide*.

iNQA tasks at a glance

To configure iNQA, perform the following tasks:

- Configuring a collector
  - Configuring parameters for a collector
  - Configuring a collector instance
  - Configuring an MP
  - Enabling packet loss measurement
- Configuring the analyzer
  - Configuring parameters for the analyzer
  - Configuring an analyzer instance
  - Configuring an AMS

    No AMSs are required in the end-to-end packet loss measurements. Configure an AMS in the point-to-point packet loss measurements.
  - Enabling the measurement functionality
- Configuring iNQA logging

Prerequisites

Before configuring iNQA, configure NTP to synchronize the clock between the analyzer and all collectors. For more information about NTP, see "Configuring NTP."

Configuring a collector

Configuring parameters for a collector

Restrictions and guidelines

If you enable packet loss measurement or latency measurement for a collector instance on the device, you cannot modify the packet loss or latency flag bit. To do that, you must first disable packet loss measurement or latency measurement for the collector instance. To view the packet loss measurement or latency measurement status for a collector instance, execute the **display inqa collector instance** command.

Procedure

1. Enter system view.

   system-view

2. Enable the collector functionality and enter its view.

   inqa collector

   By default, the collector functionality is disabled.

3. Specify a collector ID.

   collector id *collector-id*

   By default, no collector ID is specified.

   Make sure the specified collector ID must be an IPv4 address that can be routable from the analyzer.

4. Specify a ToS field bit to flag packet loss measurement.

   flag loss-measure tos-bit *tos-bit*

   By default, no ToS field bit is specified to flag packet loss measurement.

5. Bind an analyzer to the collector.

   analyzer *analyzer-id* [**udp-port** *port-number*] [**vpn-instance** *vpn-instance-name*]

   By default, no analyzer is bound to a collector.

   An analyzer specified in collector view is bound to all collector instances on the collector.

   An analyzer specified in collector instance view is bound to the specified collector instance. The analyzer ID in collector instance view takes precedence over that in collector view. A collector and a collector instance can be bound to only one analyzer. If you execute this command multiple times in the same view, the most recent configuration takes effect.

Configuring a collector instance

Restrictions and guidelines

To collect packet loss statistics on the same target flow passing through multiple collectors, create an instance on the analyzer and each collector, and make sure the instance IDs are the same.

A collector instance can monitor a maximum of two flows. Follow these guidelines when you specify the target flows:

- To monitor only one flow, specify the **forward** keyword. The collector instance does not support monitoring only one backward flow.
- To monitor two flows:
  - If the endpoint devices of the two flows are the same, specify the **bidirection** keyword. In addition, specify both the destination IPv4 address and source IPv4 address.
  - If the endpoint devices of the two flows are not identical, specify a forward flow and then a backward flow in the collector instance. Alternatively, you can create two collector instances and specify a forward flow for each collector instance.

For the flows to be monitored by different collector instances, the flow attributes must not be identical.

For a forward flow and backward flow monitored by a collector instance, the flow attributes cannot be all identical. If they have the same attributes except the direction, define them as bidirectional flows.

Procedure

1. Enter system view.

   system-view

2. Enter collector view.

   inqa collector

3. Create a collector instance and enter its view.

   instance *instance-id*

4. Bind the collector instance to an analyzer.

   analyzer *analyzer-id* [**udp-port** *port-number*] [**vpn-instance** *vpn-instance-name*]

   By default, no analyzer **is bound to a** collector instance.

   An analyzer specified in collector view is bound to all collector instances. An analyzer specified in collector instance view is bound to the specific collector instance. The analyzer ID in collector instance view takes precedence over that in collector view. A collector and a collector instance can be bound to only one analyzer. If you execute this command multiple times in the same view, the most recent configuration takes effect.

5. Specify a flow to be monitored by the collector instance.

   IPv4:

   flow { **backward** | **bidirection** | **forward** } { **destination-ip** *dest-ip-address* [*dest-mask-length*] | **dscp** *dscp-value* | **protocol** { { **tcp** | **udp** } { **destination-port** *dest-port-number1* [**to** *dest-port-number2*] | **source-port** *src-port-number1* [**to** *src-port-number2*] } * | *protocol-number* } | **source-ip** *src-ip-address* [*src-mask-length*] } * [**vxlan-transmit** | **vxlan-hybrid**]

   IPv6:

   flow { **backward** | **bidirection** | **forward** } **ipv6** { **destination-ipv6** *dest-ipv6-address* [*dest-prefix-length*] |**dscp** *dscp-value* | **protocol** { { **tcp** | **udp** } { **destination-port** *dest-port-number1* [**to** *dest-port-number2*] | **source-port** *src-port-number1* [**to** *src-port-number2*] } * | *protocol-number* } | **source-ipv6** *src-ipv6-address* [*src-prefix-length*] } * [**vxlan-transmit** | **vxlan-hybrid**]

   By default, no flow is specified for a collector instance.

6. Specify the measurement interval for the collector instance.

interval *interval*

By default, the measurement interval for a collector instance is 10 seconds.

Make sure the measurement interval for the same collector instance on all collectors are the same.

To modify the measurement interval for an enabled collector instance, first disable the measurement and then modify the measurement interval for the same collector instance in all collectors.

7. (Optional.) Configure a description for a collector instance.

   description *text*

   By default, a collector instance does not have a description.

Configuring an MP

1. Enter system view.

   system-view

2. Enter collector view.

   inqa collector

3. Create a collector instance and enter its view.

   instance *instance-id*

4. Configure an MP.

   mp *mp-id* { **in-point** | **mid-point** | **out-point** } **port-direction** { **inbound** | **outbound** }

   By default, no MP is configured.

5. Return to collector view.

   quit

6. Return to system view.

   quit

7. Enter interface view.

   interface *interface-type interface-number*

8. Bind an interface to the MP.

   inqa mp *mp-id*

   By default, no interface is bound to an MP.

Enabling packet loss measurement

About this task

Enable a packet loss measure for a collector instance as follows:

- Enable the fixed duration packet loss measurement to measure the network performance in a time period or to accurately locate the fault points for packet loss.
- Enable the continual packet loss measurement to avoid unperceived packet loss. Once starting, it does not stop unless you disable it manually.

Procedure

1. Enter system view.

   system-view

2. Enter collector view.

inqa collector

3. Create a collector instance and enter its view.
   instance *instance-id*
4. Enable packet loss measurement for the collector.
   - Enable the fixed duration packet loss measurement for the collector instance.
     loss-measure enable duration [*duration*]
     By default, the fixed duration packet loss measurement is disabled.
   - Enable continual packet loss measurement for the collector instance.
     loss-measure enable continual
     By default, the continual packet loss measurement is disabled.

Configuring the analyzer

Configuring parameters for the analyzer

1. Enter system view.
   system-view
2. Enable the analyzer functionality and enter its view.
   inqa analyzer
   By default, the analyzer functionality is disabled.
3. Specify an analyzer ID.
   analyzer id *analyzer-id*
   By default, no analyzer ID is specified.
   Make sure the analyzer ID must be an IPv4 address that can be routable from the collector.
4. (Optional.) Specify a UDP port for communication between the analyzer and the collectors.
   protocol udp-port *port-number*
   By default, the UDP port number used for communication between the analyzer and collectors is 53312.
   If the default UDP port number is used by other services, you can execute this command to specify a new UDP port number.

Configuring an analyzer instance

1. Enter system view.
   system-view
2. Enter analyzer view.
   inqa analyzer
3. Create an analyzer instance and enter its view.
   instance *instance-id*
4. (Optional.) Configure a description for an analyzer instance.
   description *text*
   By default, an analyzer instance does not have a description.
5. Bind a collector to the analyzer instance.
   collector *collector-id*

By default, no collector is bound to an analyzer instance.

Configuring an AMS

About this task

No AMSs are required in the end-to-end packet loss measurements. Configure an AMS in the point-to-point packet loss measurements.

Procedure

1. Enter system view.
   system-view
2. Enter analyzer view.
   inqa analyzer
3. Create an analyzer instance and enter its view.
   instance *instance-id*
4. Create an AMS and enter its view.
   - Create an AMS, enable packet loss measurement for it, and enter its view.
     ams *ams-id*
5. Specify the flow direction to be measured in the analyzer AMS.
   flow { **backward** | **bidirection** | **forward** }
   By default, no flow direction is specified to be measured in the analyzer AMS.
6. Add an MP to the ingress MP group for the AMS.
   in-group collector *collector-id* **mp** *mp-id*
   By default, the ingress MP group for an AMS does not have any MP.
7. Add an MP to the egress MP group for the AMS.
   out-group collector *collector-id* **mp** *mp-id*
   By default, the egress MP group for an AMS does not have any MP.

Enabling the measurement functionality

1. Enter system view.
   system-view
2. Enter analyzer view.
   inqa analyzer
3. Create an analyzer instance and enter its view.
   instance *instance-id*
4. Enable the measurement functionality of the analyzer instance.
   measure enable
   By default, the measurement functionality of an analyzer instance is disabled.

Configuring iNQA logging

About this task

iNQA calculates the packet loss rate periodically, and the analyzer generates a log on five consecutive crossings of a threshold.

- If the packet loss rate exceeds the upper limit for consecutive five measurement intervals, the analyzer sends a log to the information center.
- If the packet loss rate is less than the lower limit for consecutive five measurement intervals, the analyzer sends an event clearing log to the information center.

You can configure the information center to determine whether to output logs and the output destination. For more information about the information center, see "Configuring information center."

Procedure

1. Enter system view.
   system-view
2. Enter analyzer view.
   inqa analyzer
3. Create an analyzer instance and enter its view.
   instance *instance-id*
4. Configure packet loss logging for the analyzer instance.
   loss-measure alarm upper-limit *upper-limit* **lower-limit** *lower-limit*
   By default, the packet loss logging is not configured for an analyzer instance and the analyzer will not generate logs for threshold crossing events.

Display and maintenance commands for iNQA collector

Execute **display** commands in any view.

Task	Command
Display the collector configuration.	**display inqa collector**
Display the collector instance configuration.	**display inqa collector instance** { *instance-id* **| all** }

Display and maintenance commands for iNQA analyzer

Execute **display** commands in any view.

Task	Command
Display the analyzer configuration.	**display inqa analyzer**
Display the analyzer instance configuration.	**display inqa analyzer instance** { *instance-id* **| all** }
Display the AMS configuration in an analyzer instance.	**display inqa analyzer instance** *instance-id* **ams** { *ams-id* | **all** }
Display iNQA packet loss statistics.	**display inqa statistics loss instance** *instance-id* [**ams** *ams-id*]

iNQA configuration examples

Example: Configuring an end-to-end iNQA packet loss measurement

Network configuration

As shown in Figure 10, Video phone 1 sends a video data flow to Video phone 2 through an IP network. The Video phone 2 is experiencing an erratic video display problem.

- Enable the collector functionality on Device 1 and Device 2 and enable the analyzer functionality on Device 2.
- Define the flow from Device 1 to Device 2 as the forward flow.
- Set the packet loss upper limit and packet loss lower limit to 6% and 4%, respectively.

Figure 10 Network diagram

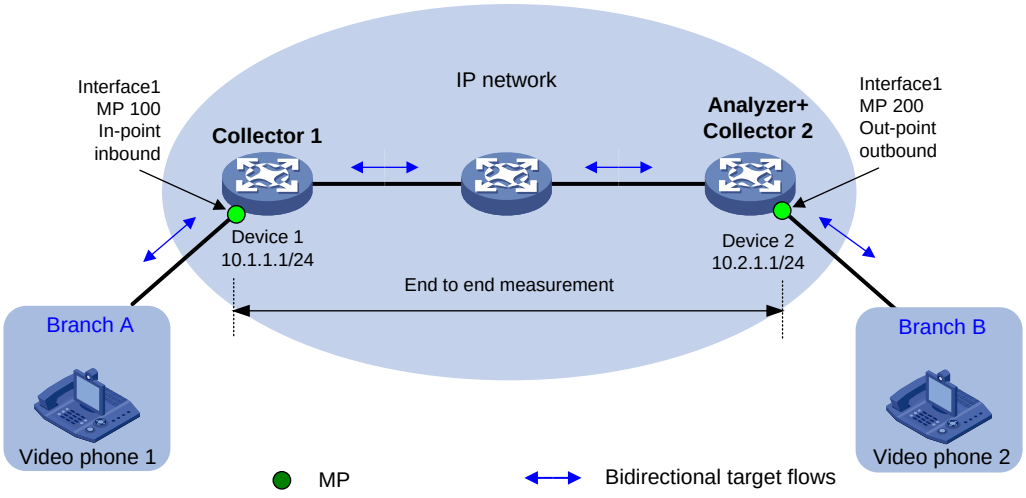


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

1. Assign 10.1.1.1 and 10.2.1.1 to Collector 1 and Collector 2, respectively.
2. Configure OSPF in the IP network to make Collector 1 and the analyzer can reach each other. (Details not shown.)
3. Configure NTP or PTP on Collector 1 and Collector 2 for clock synchronization. (Details not shown.)

Procedure

1. Configure Collector 1:

   # Specify 10.1.1.1 as the collector ID.

   ```
   <Collector1> system-view
   [Collector1] inqa collector
   [Collector1-inqa-collector] collector id 10.1.1.1
   ```

   # Bind the collector to the analyzer with ID 10.2.1.1.

```
[Collector1-inqa-collector] analyzer 10.2.1.1
```

Specify ToS field bit 5 as the flag bit.

```
[Collector1-inqa-collector] flag loss-measure tos-bit 5
```

Configure collector instance 1 to monitor the bidirectional flows entering the network from GigabitEthernet1/0/1 between 10.1.1.0 and 10.2.1.0.

```
[Collector1-inqa-collector] instance 1
[Collector1-inqa-collector-instance-1] flow bidirection source-ip 10.1.1.0 24
destination-ip 10.2.1.0 24
[Collector1-inqa-collector-instance-1] mp 100 in-point port-direction inbound
[Collector1-inqa-collector-instance-1] quit
[Collector1-inqa-collector] quit
[Collector1] interface gigabitethernet 1/0/1
[Collector1-GigabitEthernet1/0/1] inqa mp 100
[Collector1-GigabitEthernet1/0/1] quit
```

Enable continual packet loss measurement.

```
[Collector1] inqa collector
[Collector1-inqa-collector] instance 1
[Collector1-inqa-collector-instance-1] loss-measure enable continual
[Collector1-inqa-collector-instance-1] return
<Collector1>
```

2. Configure Collector 2 and the analyzer:

Specify 10.2.1.1 as the collector ID.

```
<AnalyzerColl2> system-view
[AnalyzerColl2] inqa collector
[AnalyzerColl2-inqa-collector] collector id 10.2.1.1
```

Bind the collector to the analyzer with ID 10.2.1.1.

```
[AnalyzerColl2-inqa-collector] analyzer 10.2.1.1
```

Specify ToS field bit 5 as the flag bit.

```
[AnalyzerColl2-inqa-collector] flag loss-measure tos-bit 5
```

Configure collector instance 1 to monitor the bidirectional flows entering the network from GigabitEthernet1/0/1 between 10.1.1.0 and 10.2.1.0.

```
[AnalyzerColl2-inqa-collector] instance 1
[AnalyzerColl2-inqa-collector-instance-1] flow bidirection source-ip 10.1.1.0 24
destination-ip 10.2.1.0 24
[AnalyzerColl2-inqa-collector-instance-1] mp 200 out-point port-direction outbound
[AnalyzerColl2-inqa-collector-instance-1] quit
[AnalyzerColl2-inqa-collector] quit
[AnalyzerColl2] interface gigabitethernet 1/0/1
[AnalyzerColl2-GigabitEthernet1/0/1] inqa mp 200
[AnalyzerColl2-GigabitEthernet1/0/1] quit
```

Enable continual packet loss measurement.

```
[AnalyzerColl2] inqa collector
[AnalyzerColl2-inqa-collector] instance 1
[AnalyzerColl2-inqa-collector-instance-1] loss-measure enable continual
[AnalyzerColl2-inqa-collector-instance-1] quit
[AnalyzerColl2-inqa-collector] quit
```

Specify 10.2.1.1 as the analyzer ID.

```
[AnalyzerColl2] inqa analyzer
```

```
[AnalyzerColl2-inqa-analyzer] analyzer id 10.2.1.1
```

Bind analyzer instance **1** to Collector 1 and Collector 2.

```
[AnalyzerColl2-inqa-analyzer] instance 1
[AnalyzerColl2-inqa-analyzer-instance-1] collector 10.1.1.1
[AnalyzerColl2-inqa-analyzer-instance-1] collector 10.2.1.1
```

Set the packet loss upper limit and lower limit to 6% and 4%, respectively.

```
[AnalyzerColl2-inqa-analyzer-instance-1] loss-measure alarm upper-limit 6
lower-limit 4
```

Enable the measurement functionality of analyzer instance **1**.

```
[AnalyzerColl2-inqa-analyzer-instance-1] measure enable
[AnalyzerColl2-inqa-analyzer-instance-1] return
<AnalyzerColl2>
```

Verifying the configuration

1. On Collector 1:

Display the collector configuration.

```
<Collector1> display inqa collector
Collector ID             : 10.1.1.1
Loss-measure flag        : 5
Delay-measure flag       : --
Analyzer ID              : 10.2.1.1
Analyzer UDP-port        : 53312
VPN-instance-name        : --
Current instance count   : 1
```

Display the configuration of collector instance **1**.

```
<Collector1> display inqa collector instance 1
Instance ID              : 1
Loss-measure status      : Enabled
Loss-measure duration    : --
Delay-measure status     : Disabled
Delay-measure duration   : --
Description              : --
Analyzer ID              : --
Analyzer UDP-port        : --
VPN-instance-name        : --
Interval                 : 10 sec
Flow configuration:
  flow bidirection source-ip 10.1.1.0 24 destination-ip 10.2.1.0 24
MP configuration:
  mp 100 in-point inbound, GE1/0/1
```

2. On Collector 2 and the analyzer:

Display the collector configuration.

```
<AnalyzerColl2> display inqa collector
Collector ID             : 10.2.1.1
Loss-measure flag        : 5
Delay-measure flag       : --
Analyzer ID              : 10.2.1.1
Analyzer UDP-port        : 53312
```

```
VPN-instance-name        : --
Current instance count   : 1
```

Display the configuration of collector instance **1**.

```
<AnalyzerColl2> display inqa collector instance 1
Instance ID              : 1
Loss-measure status      : Enabled
Loss-measure duration    : --
Delay-measure status     : Disabled
Delay-measure duration   : --
Description              : --
Analyzer ID              : --
Analyzer UDP-port        : --
VPN-instance-name        : --
Interval                 : 10 sec
Flow configuration:
  flow bidirection source-ip 10.1.1.0 24 destination-ip 10.2.1.0 24
MP configuration:
  mp 200 out-point outbound, GE1/0/1
```

Display the analyzer configuration.

```
<AnalyzerColl2> display inqa analyzer
Analyzer ID              : 10.2.1.1
Protocol UDP-port        : 53312
Current instance count   : 1
```

Display the configuration of analyzer instance **1**.

```
<AnalyzerColl2> display inqa analyzer instance 1
Instance ID              : 1
Status                   : Enable
Description              : --
Loss-alarm upper-limit        : 6.000000%
Loss-alarm lower-limit        : 4.000000%
Delay-alarm upper-limit        : --
Delay-alarm lower-limit        : --
Current AMS count        : 0
Collectors               : 10.1.1.1
                           10.2.1.1
```

Display iNQA packet loss statistics of analyzer instance **1**.

```
<AnalyzerColl2> display inqa statistics loss instance 1
Latest packet loss statistics for forward flow:
 Period     LostPkts             PktLoss%     LostBytes            ByteLoss%
 19122483   15                   15.000000%   1500                 15.000000%
 19122482   15                   15.000000%   1500                 15.000000%
 19122481   15                   15.000000%   1500                 15.000000%
 19122480   15                   15.000000%   1500                 15.000000%
 19122479   15                   15.000000%   1500                 15.000000%
 19122478   15                   15.000000%   1500                 15.000000%
Latest packet loss statistics for backward flow:
 Period     LostPkts             PktLoss%     LostBytes            ByteLoss%
 19122483   15                   15.000000%   1500                 15.000000%
```

19122482	15	15.000000%	1500	15.000000%
19122481	15	15.000000%	1500	15.000000%
19122480	15	15.000000%	1500	15.000000%
19122479	15	15.000000%	1500	15.000000%
19122478	15	15.000000%	1500	15.000000%

Example: Configuring an point-to-point iNQA packet loss measurement

As shown in Figure 11, Video phone 1 sends a video data flow to Video phone 2 through an IP network. The Video phone 2 is experiencing an erratic video display problem.

- Enable the collector functionality on Device 1, Device 2, and Device 3, and enable the analyzer functionality on Device 2.
- Define the flow from Device 1 to Device 3 as the forward flow.
- Set the packet loss upper limit and packet loss lower limit to 6% and 4%, respectively.
- Run the packet loss measurement for 15 minutes.

Figure 11 Network diagram

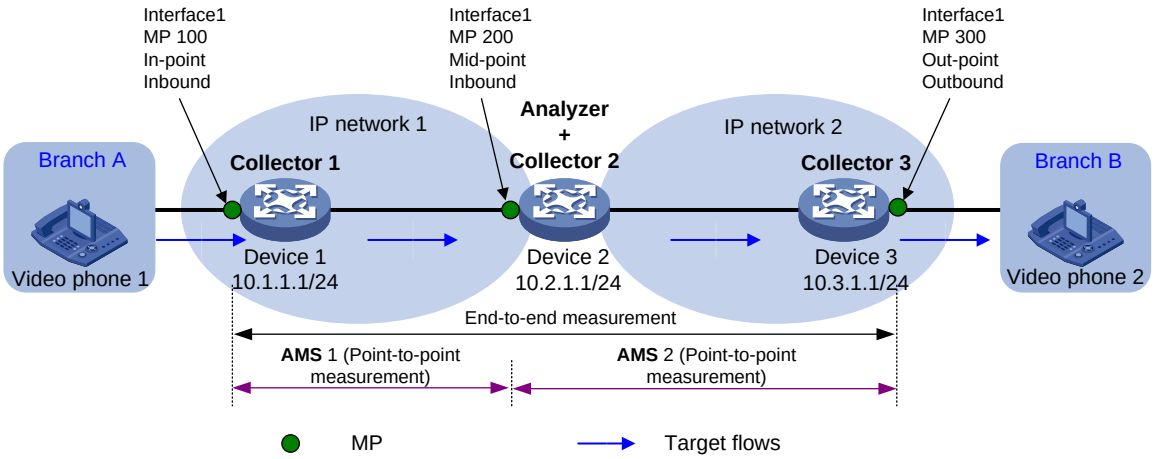


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

1. Assign 10.1.1.1, 10.2.1.1, and 10.3.1.1 to Collector 1, Collector 2, and Collector 3, respectively.
2. Configure OSPF in the IP network to make sure Collector 1, Collector 3, and the analyzer can reach each other. (Details not shown.)
3. Configure NTP or PTP on Collector 1, Collector 2, and Collector 3 for clock synchronization. (Details not shown.)

Procedure

1. Configure Collector 1:

   # Specify 10.1.1.1 as the collector ID.

   ```
   <Collector1> system-view
   [Collector1] inqa collector
   ```

```
[Collector1-inqa-collector] collector id 10.1.1.1
```

Bind the collector to the analyzer with ID 10.2.1.1.

```
[Collector1-inqa-collector] analyzer 10.2.1.1
```

Specify ToS field bit 5 as the flag bit.

```
[Collector1-inqa-collector] flag loss-measure tos-bit 5
```

Configure collector instance **1** to monitor the forward flow entering the network at GigabitEthernet1/0/1 from 10.1.1.0 and 10.3.1.0.

```
[Collector1-inqa-collector] instance 1
[Collector1-inqa-collector-instance-1] flow forward source-ip 10.1.1.0 24
destination-ip 10.3.1.0 24
[Collector1-inqa-collector-instance-1] mp 100 in-point port-direction inbound
[Collector1-inqa-collector-instance-1] quit
[Collector1-inqa-collector] quit
[Collector1] interface gigabitethernet 1/0/1
[Collector1-GigabitEthernet1/0/1] inqa mp 100
[Collector1-GigabitEthernet1/0/1] quit
```

Enable continual packet loss measurement for collector instance **1**.

```
[Collector1] inqa collector
[Collector1-inqa-collector] instance 1
[Collector1-inqa-collector-instance-1] loss-measure enable continual
[Collector1-inqa-collector-instance-1] return
<Collector1>
```

2. Configure Collector 2 and the analyzer:

Specify 10.2.1.1 as the collector ID.

```
<AnalyzerColl2> system-view
[AnalyzerColl2] inqa collector
[AnalyzerColl2-inqa-collector] collector id 10.2.1.1
```

Bind the collector to the analyzer with ID 10.2.1.1.

```
[AnalyzerColl2-inqa-collector] analyzer 10.2.1.1
```

Specify ToS field bit 5 as the flag bit.

```
[AnalyzerColl2-inqa-collector] flag loss-measure tos-bit 5
```

Configure collector instance **1** to monitor the forward flow entering the network at GigabitEthernet1/0/1 from 10.1.1.0 and 10.3.1.0.

```
[AnalyzerColl2-inqa-collector] instance 1
[AnalyzerColl2-inqa-collector-instance-1] flow forward source-ip 10.1.1.0 24
destination-ip 10.3.1.0 24
[AnalyzerColl2-inqa-collector-instance-1] mp 200 mid-point port-direction inbound
[AnalyzerColl2-inqa-collector-instance-1] quit
[AnalyzerColl2-inqa-collector] quit
[AnalyzerColl2] interface gigabitethernet 1/0/1
[AnalyzerColl2-GigabitEthernet1/0/1] inqa mp 200
[AnalyzerColl2-GigabitEthernet1/0/1] quit
```

Enable continual packet loss measurement for collector instance **1**.

```
[AnalyzerColl2] inqa collector
[AnalyzerColl2-inqa-collector] instance 1
[AnalyzerColl2-inqa-collector-instance-1] loss-measure enable continual
[AnalyzerColl2-inqa-collector-instance-1] quit
[AnalyzerColl2-inqa-collector] quit
```

Specify 10.2.1.1 as the analyzer ID.

```
[AnalyzerColl2] inqa analyzer
[AnalyzerColl2-inqa-analyzer] analyzer id 10.2.1.1
```

Bind analyzer instance **1** to Collector 1, Collector 2, and Collector 3.

```
[AnalyzerColl2-inqa-analyzer] instance 1
[AnalyzerColl2-inqa-analyzer-instance-1] collector 10.1.1.1
[AnalyzerColl2-inqa-analyzer-instance-1] collector 10.2.1.1
[AnalyzerColl2-inqa-analyzer-instance-1] collector 10.3.1.1
```

Configure AMS 1 to measure the packet loss rate for the forward flow from MP 100 to MP 200.

```
[AnalyzerColl2-inqa-analyzer-instance-1] ams 1
[AnalyzerColl2-inqa-analyzer-instance-1-ams-1] flow forward
[AnalyzerColl2-inqa-analyzer-instance-1-ams-1] in-group collector 10.1.1.1 mp 100
[AnalyzerColl2-inqa-analyzer-instance-1-ams-1] out-group collector 10.2.1.1 mp 200
[AnalyzerColl2-inqa-analyzer-instance-1-ams-1] quit
```

Configure AMS 2 to measure the packet loss rate for the forward flow from MP 200 to MP 300.

```
[AnalyzerColl2-inqa-analyzer-instance-1] ams 2
[AnalyzerColl2-inqa-analyzer-instance-1-ams-2] flow forward
[AnalyzerColl2-inqa-analyzer-instance-1-ams-2] in-group collector 10.2.1.1 mp 200
[AnalyzerColl2-inqa-analyzer-instance-1-ams-2] out-group collector 10.3.1.1 mp 300
[AnalyzerColl2-inqa-analyzer-instance-1-ams-2] quit
```

Set the packet loss upper limit and lower limit to 6% and 4%, respectively.

```
[AnalyzerColl2-inqa-analyzer-instance-1] loss-measure alarm upper-limit 6
lower-limit 4
```

Enable the measurement functionality of analyzer instance **1**.

```
[AnalyzerColl2-inqa-analyzer-instance-1] measure enable
[AnalyzerColl2-inqa-analyzer-instance-1] return
<AnalyzerColl2>
```

3. Configure Collector 3:

Specify 10.3.1.1 as the collector ID.

```
<Collector3> system-view
[Collector3] inqa collector
[Collector3-inqa-collector] collector id 10.3.1.1
```

Bind the collector to the analyzer with ID 10.2.1.1.

```
[Collector3-inqa-collector] analyzer 10.2.1.1
```

Specify ToS field bit 5 as the flag bit.

```
[Collector3-inqa-collector] flag loss-measure tos-bit 5
```

Configure collector instance **1** to monitor the forward flow entering the network at GigabitEthernet1/0/1 from 10.1.1.0 and 10.3.1.0.

```
[Collector3-inqa-collector] instance 1
[Collector3-inqa-collector-instance-1] flow forward source-ip 10.1.1.0 24
destination-ip 10.3.1.0 24
[Collector3-inqa-collector-instance-1] mp 300 out-point port-direction outbound
[Collector3-inqa-collector-instance-1] quit
[Collector3-inqa-collector] quit
[Collector3] interface gigabitethernet 1/0/1
[Collector3-GigabitEthernet1/0/1] inqa mp 300
[Collector3-GigabitEthernet1/0/1] quit
```

Enable continual packet loss measurement for collector instance **1**.

```
[Collector3] inqa collector
[Collector3-inqa-collector] instance 1
[Collector3-inqa-collector-instance-1] loss-measure enable continual
[Collector3-inqa-collector-instance-1] return
<Collector3>
```

Verifying the configuration

1. On Collector 1:

Display the collector configuration.

```
<Collector1> display inqa collector
Collector ID             : 10.1.1.1
Loss-measure flag        : 5
Delay-measure flag       : --
Analyzer ID              : 10.2.1.1
Analyzer UDP-port        : 53312
VPN-instance-name        : --
Current instance count   : 1
```

Display the configuration of collector instance **1**.

```
<Collector1> display inqa collector instance 1
Instance ID              : 1
Loss-measure status      : Enabled
Loss-measure duration    : --
Delay-measure status     : Disabled
Delay-measure duration   : --
Description              : --
Analyzer ID              : --
Analyzer UDP-port        : --
VPN-instance-name        : --
Interval                 : 10 sec
Flow configuration:
  flow forward source-ip 10.1.1.0 24 destination-ip 10.3.1.0 24
MP configuration:
  mp 100 in-point inbound, GE1/0/1
```

2. On Collector 2 and the analyzer:

Display the collector configuration.

```
<AnalyzerColl2> display inqa collector
Collector ID             : 10.2.1.1
Loss-measure flag        : 5
Delay-measure flag       : --
Analyzer ID              : 10.2.1.1
Analyzer UDP-port        : 53312
VPN-instance-name        : --
Current instance count   : 1
```

Display the configuration of collector instance **1**.

```
<AnalyzerColl2>  display inqa collector instance 1
Instance ID              : 1
Loss-measure status      : Enabled
Loss-measure duration    : --
```

```
Delay-measure status     : Disabled
Delay-measure duration   : --
Description              : --
Analyzer ID              : --
Analyzer UDP-port        : --
VPN-instance-name        : --
Interval                 : 10 sec
Flow configuration:
  flow forward source-ip 10.1.1.0 24 destination-ip 10.3.1.0 24
MP configuration:
  mp 200 mid-point inbound, GE1/0/1
```

Display analyzer configuration.

```
<AnalyzerColl2> display inqa analyzer
Analyzer ID              : 10.2.1.1
Protocol UDP-port        : 53312
Current instance count   : 1
```

Display the configuration of analyzer instance **1**.

```
<AnalyzerColl2> display inqa analyzer instance 1
Instance ID              : 1
Status                   : Enabled
Description              : --
Loss-alarm upper-limit   : 6.000000%
Loss-alarm lower-limit   : 4.000000%
Delay-alarm upper-limit  : --
Delay-alarm lower-limit  : --
Current AMS count        : 2
Collectors               : 10.1.1.1
                           10.2.1.1
                           10.3.1.1
```

Display the configuration of all AMSs in analyzer instance **1**.

```
<AnalyzerColl2> display inqa analyzer instance 1 ams all
AMS ID                   : 1
Flow direction           : Forward
Delay-measure status     : Disabled
In-group                 : collector 10.1.1.1 mp 100
Out-group                : collector 10.2.1.1 mp 200

AMS ID                   : 2
Flow direction           : Forward
In-group                 : collector 10.2.1.1 mp 200
Out-group                : collector 10.3.1.1 mp 300
```

Display iNQA packet loss statistics of analyzer instance **1**.

```
<AnalyzerColl2> display inqa statistics loss instance 1 ams 1
Latest packet loss statistics for forward flow:
```

Period	LostPkts	PktLoss%	LostBytes	ByteLoss%
19122483	15	15.000000%	1500	15.000000%
19122482	15	15.000000%	1500	15.000000%
19122481	15	15.000000%	1500	15.000000%

```
19122480   15                  15.000000%   1500                15.000000%
19122479   15                  15.000000%   1500                15.000000%
19122478   15                  15.000000%   1500                15.000000%
```

3. On Collector 3:

Display the collector configuration.

```
<Collector3> display inqa collector
Collector ID             : 10.3.1.1
Loss-measure flag        : 5
Delay-measure flag       : --
Analyzer ID              : 10.2.1.1
Analyzer UDP-port        : 53312
VPN-instance-name        : --
Current instance count   : 1
```

Display the configuration of collector instance **1**.

```
<Collector3> display inqa collector instance 1
Instance ID              : 1
Loss-measure status      : Enabled
Loss-measure duration    : --
Delay-measure status     : Disabled
Delay-measure duration   : --
Description              : --
Analyzer ID              : --
Analyzer UDP-port        : --
VPN-instance-name        : --
Interval                 : 10 sec
Flow configuration:
  flow forward source-ip 10.1.1.0 24 destination-ip 10.3.1.0 24
MP configuration:
  mp 300 out-point outbound, GE1/0/1
```

Contents

Configuring NTP 1
About NTP 1
NTP application scenarios 1
NTP working mechanism 1
NTP architecture 2
NTP association modes 3
NTP security 8
Protocols and standards 9
FIPS compliance 9
Restrictions and guidelines: NTP configuration 10
Client/server mode tasks at a glance 10
Symmetric active/passive mode tasks at a glance 11
Broadcast mode tasks at a glance 12
Multicast mode tasks at a glance 13
Enabling the NTP service 13
Configuring NTP association mode 14
Configuring NTP in client/server mode 14
Configuring NTP in symmetric active/passive mode 15
Configuring NTP in broadcast mode 16
Configuring NTP in multicast mode 18
Configuring the local clock as the reference source 19
Configuring access control rights 20
Configuring NTP authentication 21
Configuring NTP authentication in client/server mode 21
Configuring NTP authentication in symmetric active/passive mode 22
Configuring NTP authentication in broadcast mode 24
Configuring NTP authentication in multicast mode 25
Controlling NTP message sending and receiving 27
Specifying a source address for NTP messages 27
Disabling an interface from receiving NTP messages 28
Configuring the maximum number of dynamic associations 28
Setting a DSCP value for NTP packets 29
Controlling NTP time synchronization 29
Ignoring clock stratum level restrictions on NTP clients 29
Setting the maximum offset allowed for time synchronization through NTP 29
Specifying the NTP time-offset thresholds for log and trap outputs 30
Display and maintenance commands for NTP 30
NTP configuration examples 31
Example: Configuring NTP client/server association mode 31
Example: Configuring IPv6 NTP client/server association mode 32
Example: Configuring NTP symmetric active/passive association mode 33
Example: Configuring IPv6 NTP symmetric active/passive association mode 34
Example: Configuring NTP broadcast association mode 36
Example: Configuring NTP multicast association mode 38
Example: Configuring IPv6 NTP multicast association mode 40
Example: Configuring NTP authentication in client/server association mode 43
Example: Configuring NTP authentication in broadcast association mode 45
Configuring SNTP 48
About SNTP 48
Relationship between SNTP and NTP 48
SNTP operating mechanism 48
SNTP association modes 49
Protocols and standards 50
FIPS compliance 50
Restrictions and guidelines: SNTP configuration 50
Client/server mode tasks at a glance 50

Broadcast mode tasks at a glance ····· 51
Enabling the SNTP service ····· 51
Configuring SNTP in client/server mode ····· 51
Specifying an NTP server for the device ····· 51
Configuring SNTP authentication in client/server mode ····· 52
Configuring SNTP in broadcast mode ····· 53
Configuring the SNTP broadcast client ····· 53
Configuring SNTP authentication in broadcast mode ····· 54
Controlling SNTP time synchronization ····· 54
Ignoring clock stratum level restrictions on SNTP clients ····· 54
Setting the maximum offset allowed for time synchronization through SNTP ····· 55
Monitoring the time synchronization status ····· 55
Specifying the SNTP time-offset thresholds for log and trap outputs ····· 56
Display and maintenance commands for SNTP ····· 57
SNTP configuration examples ····· 57
Example: Configuring SNTP in client/server association mode ····· 57
Example: Configuring SNTP in broadcast association mode ····· 58

Configuring NTP

About NTP

NTP is used to synchronize system clocks among distributed time servers and clients on a network. NTP runs over UDP and uses UDP port 123.

NTP application scenarios

Various tasks, including network management, charging, auditing, and distributed computing depend on accurate and synchronized system time setting on the network devices. NTP is typically used in large networks to dynamically synchronize time among network devices.

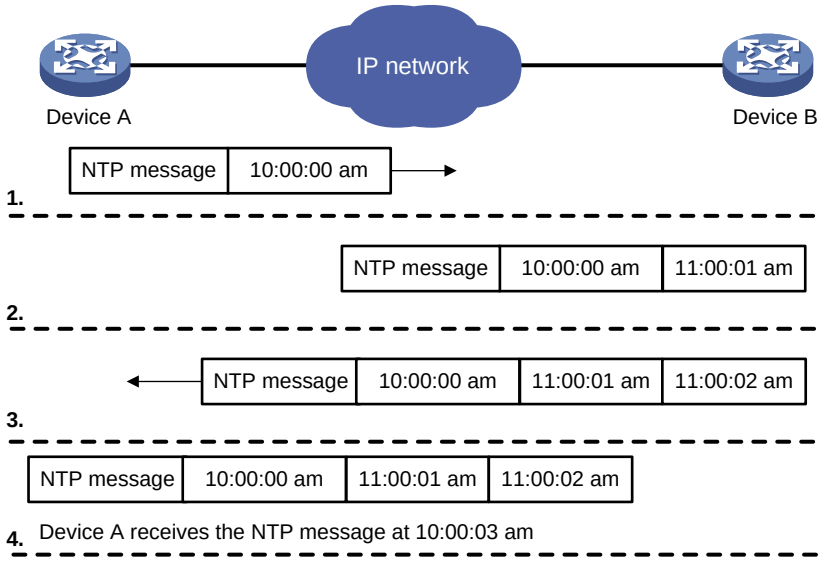
NTP guarantees higher clock accuracy than manual system clock setting. In a small network that does not require high clock accuracy, you can keep time synchronized among devices by changing their system clocks one by one.

NTP working mechanism

Figure 1 shows how NTP synchronizes the system time between two devices (Device A and Device B, in this example). Assume that:

- Prior to the time synchronization, the time is set to 10:00:00 am for Device A and 11:00:00 am for Device B.
- Device B is used as the NTP server. Device A is to be synchronized to Device B.
- It takes 1 second for an NTP message to travel from Device A to Device B, and from Device B to Device A.
- It takes 1 second for Device B to process the NTP message.

Figure 1 Basic work flow



The synchronization process is as follows:

2. Device A sends Device B an NTP message, which is timestamped when it leaves Device A. The time stamp is 10:00:00 am (T1).

3. When this NTP message arrives at Device B, Device B adds a timestamp showing the time when the message arrived at Device B. The timestamp is 11:00:01 am (T2).
4. When the NTP message leaves Device B, Device B adds a timestamp showing the time when the message left Device B. The timestamp is 11:00:02 am (T3).
5. When Device A receives the NTP message, the local time of Device A is 10:00:03 am (T4).

Up to now, Device A can calculate the following parameters based on the timestamps:

- The roundtrip delay of the NTP message: Delay = (T4 – T1) – (T3 – T2) = 2 seconds.
- Time difference between Device A and Device B: Offset = [(T2 – T1) + (T3 – T4)] /2 = 1 hour.

Based on these parameters, Device A can be synchronized to Device B.

This is only a rough description of the work mechanism of NTP. For more information, see the related protocols and standards.

NTP architecture

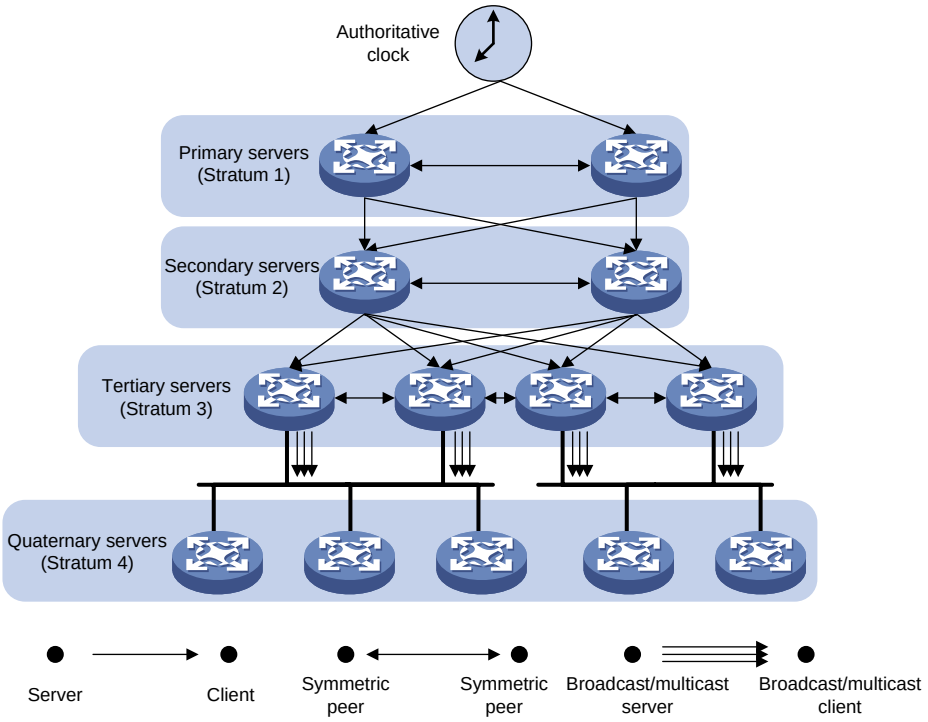
CAUTION:

When the device acts as an NTP client, the NTP server's clock stratum must be not smaller than 0 and not larger than 14. If the NTP server's clock stratum is larger than 14, the device will not synchronize with that server's clock. You can log in to the NTP server to change its clock stratum by executing the **`ntp-service refclock-master`** command.

- To check the stratum for an IPv4 NTP server, execute the **`display ntp-service sessions`** command on the device and check the **stra** field in the output.
- To check the stratum for an IPv6 NTP server, execute the **`display ntp-service ipv6 sessions`** command on the device and check the **Clock stratum** field in the output.

NTP network architecture

Figure 2 NTP network architecture



As shown in Figure 2, the NTP network is a hierarchical distributed system that contains multiple tiers of clocks to provide time synchronization services. This architecture helps reduce dependency on a single time source while providing redundancy and load balancing. The following are the key components of the NTP network:

- **Stratum 0**—The highest precision devices, such as atomic clocks, GPS, or radio clocks. They do not directly connect to the network but provide time to stratum 1 servers through wired or wireless connections.
- **Stratum 1**—Servers directly connected to stratum 0 devices. They act as primary time servers or reference clocks. Stratum 1 servers are typically used to provide precise time for public or private networks.
- **Stratum 2**—Servers connected to stratum 1 servers, receiving time from one or more stratum 1 servers and then providing time to lower-stratum servers or clients.
- **Stratum 3 through stratum 15**—Servers that receive time from servers of a higher stratum and that distribute this time down. Each stratum introduces some delay and jitter, but time synchronization accuracy can be achieved through complex algorithms and the use of multiple time sources.

If the devices in a network cannot synchronize to an authoritative time source, you can perform the following tasks:

- Select a device that has a relatively accurate clock from the network.
- Use the local clock of the device as the reference clock to synchronize other devices in the network.

NTP stratum

The NTP clock stratum represents the hierarchical level of an NTP server within the time synchronization structure. The stratum value is in the range of 1 to 15, with a larger value indicating a greater distance from the reference time source, which might result in reduced synchronization accuracy.

The clock stratum is typically determined automatically based on the upstream time source to which the server is connected. An NTP client only synchronizes with an NTP server whose clock stratum is less than or equal to its own, and the server’s stratum must be in the range of 0 to 14. Therefore, if you configure the device as an NTP server, you must set its clock stratum according to the network hierarchy.

NTP association modes

About NTP association modes

NTP supports the following association modes:

- Client/server mode
- Symmetric active/passive mode
- Broadcast mode
- Multicast mode

You can select one or more association modes for time synchronization. Table 1 provides detailed description for the four association modes. If the device acts as an NTP client and receives multiple NTP clock signals, it selects the optimal NTP clock for time synchronization.

In this document, an "NTP server" or a "server" refers to a device that operates as an NTP server in client/server mode. Time servers refer to all the devices that can provide time synchronization, including NTP servers, NTP symmetric peers, broadcast servers, and multicast servers.

Table 1 NTP association modes

Mode	Working process	Principle	Application scenario
Client/server	On the client, specify the IP address of the NTP server. A client sends a clock synchronization message to the NTP servers. Upon receiving the message, the servers automatically operate in server mode and send a reply. If the client can be synchronized to multiple time servers, it selects an optimal clock and synchronizes its local clock to the optimal reference source after receiving the replies from the servers.	A client can synchronize to a server, but a server cannot synchronize to a client.	As the NTP architecture diagram in "NTP architecture" shows, this mode is intended for scenarios where devices with a higher stratum value obtain time from a device with a lower stratum value.
Symmetric active/passive	On the symmetric active peer, specify the IP address of the symmetric passive peer. A symmetric active peer periodically sends clock synchronization messages to a symmetric passive peer. The symmetric passive peer automatically operates in symmetric passive mode and sends a reply. If the symmetric active peer can be synchronized to multiple time servers, it selects an optimal clock and synchronizes its local clock to the optimal reference source after receiving the replies from the servers.	A symmetric active peer and a symmetric passive peer can be synchronized to each other. If both of them are synchronized, the peer with a higher stratum is synchronized to the peer with a lower stratum.	As the NTP architecture diagram in "NTP architecture" shows, this mode is most often used between servers with the same stratum to operate as a backup for one another. If a server fails to communicate with all the servers of a lower stratum, the server can still synchronize to the servers of the same stratum.
Broadcast	A server periodically sends clock synchronization messages to the broadcast address 255.255.255.255. Clients listen to the broadcast messages from the servers to synchronize to the server according to the broadcast messages. When a client receives the first broadcast message, the client and the server start to exchange messages to calculate the network delay between them. Then, only the broadcast server sends clock synchronization messages.	A broadcast client can synchronize to a broadcast server, but a broadcast server cannot synchronize to a broadcast client.	A broadcast server sends clock synchronization messages to synchronize clients in the same subnet. As the NTP architecture diagram in "NTP architecture" shows, broadcast mode is intended for configurations involving one or a few servers and a potentially large client population. The broadcast mode has lower time accuracy than the client/server and symmetric active/passive modes because only the broadcast servers send clock synchronization messages.
Multicast	A multicast server periodically sends clock synchronization messages to the user-configured	A multicast client can synchronize to a multicast server, but a	A multicast server can provide time synchronization for clients

Mode	Working process	Principle	Application scenario
	multicast address. Clients listen to the multicast messages from servers and synchronize to the server according to the received messages.	multicast server cannot synchronize to a multicast client.	in the same subnet or in different subnets. The multicast mode has lower time accuracy than the client/server and symmetric active/passive modes.

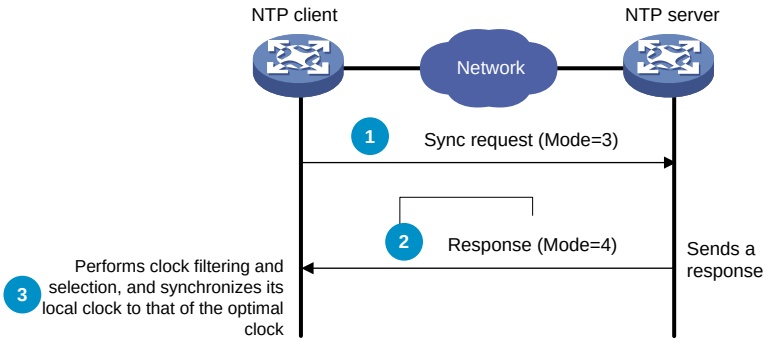
Operating mechanism for client/server association mode

NTP client/server mode allows an NTP server to maintain time information and allow NTP clients to communicate with the server to calibrate their clocks. In this mode, clients can accurately synchronize time with the server, ensuring time consistency across the network.

The time synchronization process in NTP client/server mode is as follows:

1. The client periodically sends time synchronization request messages to the server, with the mode field set to 3, indicating client mode. The client does not concern about the server's reachability or stratum.
2. Upon receiving the synchronization request from the client, the server fills the current time information in the response message and sets the mode field to 4, indicating server mode. Then, it sends the message back to the client. During this process, the server does not need to retain any state information about the client.
3. Upon receiving the response from the server, the client performs clock filtering and selection, and synchronizes its local clock to that of the optimal clock.

Figure 3 Operating mechanism for client/server association mode



Operating mechanism for symmetric active/passive mode

In symmetric active/passive mode, time synchronization is implemented through communication and negotiation between two nodes (peers). Both peers can act as the time source, with the more reliable and accurate time from both nodes as the optimal time.

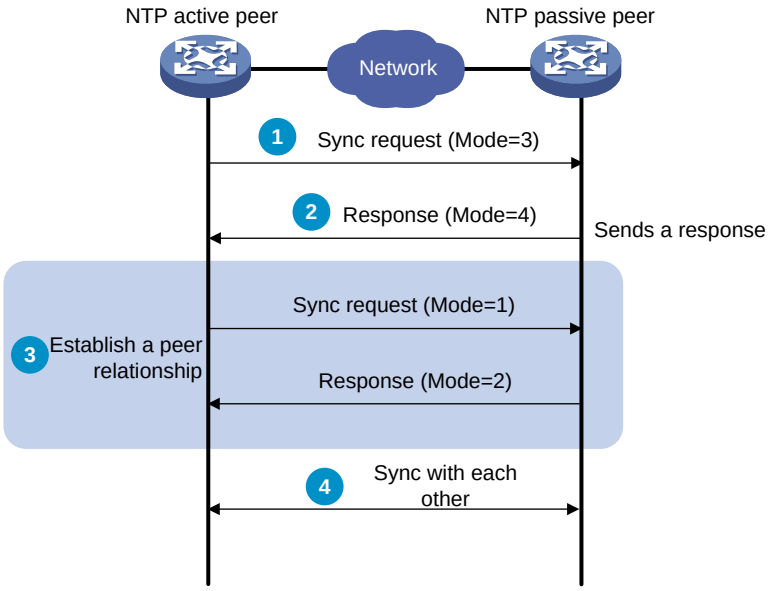
The time synchronization process in symmetric active/passive mode is as follows:

1. The symmetric active peer first sends a message with the mode field set to 1 (indicating an active peer), and the passive peer responds with a message with the mode field set to 2 (indicating a passive peer) to establish a peer relationship.
2. The peers send and receive time information for each other and use it to adjust their clocks. During this process, the two peers calculate the round-trip time delay and offset, and adjust their local clocks accordingly.

The peer with a smaller clock stratum will synchronize the clock of the peer with a greater clock stratum. The stratum indicates the distance from the standard time source. Nodes with a lower stratum are determined as closer to the standard time source, and therefore more reliable.

If the active peer can synchronize time with multiple time servers, it performs clock filtering and selection after receiving a response message. NTP uses a series of complex algorithms, such as clock filtering, offset prediction, and jitter estimation to select the best time sample and adjust the local time accordingly. Through these algorithms, the peers select the most stable and reliable time sample over a period as the basis for time synchronization.

Figure 4 Operating mechanism for symmetric active/passive mode



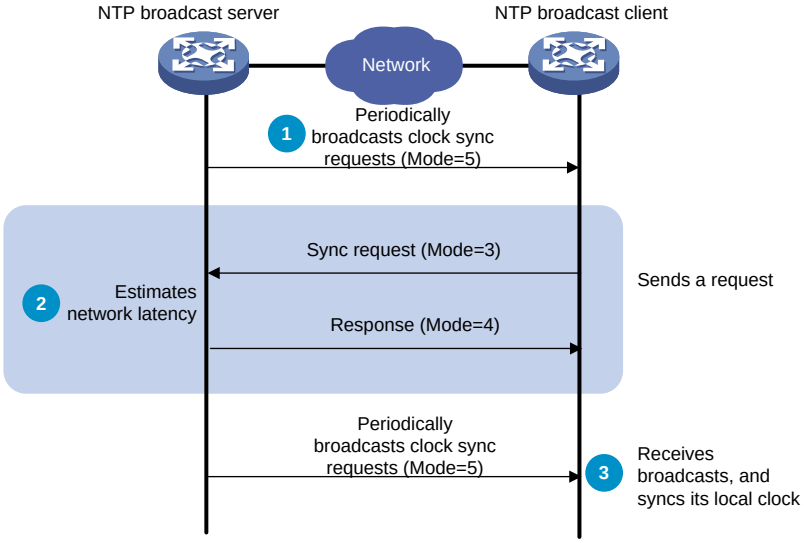
Operating mechanism for broadcast mode

NTP broadcast mode allows an NTP server to broadcast time information to the entire network. In this mode, any NTP clients listening to the broadcast can receive the time information and synchronize their system clocks accordingly.

The time synchronization process in NTP broadcast mode is as follows:

1. The broadcast server periodically sends clock synchronization request messages to broadcast address 255.255.255.255, with the mode field set to 5, indicating broadcast mode.
2. The broadcast client listens to the broadcast messages from the server. When the client receives the first broadcast message, the client and the server start to exchange messages to calculate the network delay between them.
3. The client enters broadcast mode, continues to listen for subsequent broadcast messages, and synchronizes its local clock based on the time information in these messages. Then, only the broadcast server sends clock synchronization messages.

Figure 5 Operating mechanism for broadcast mode



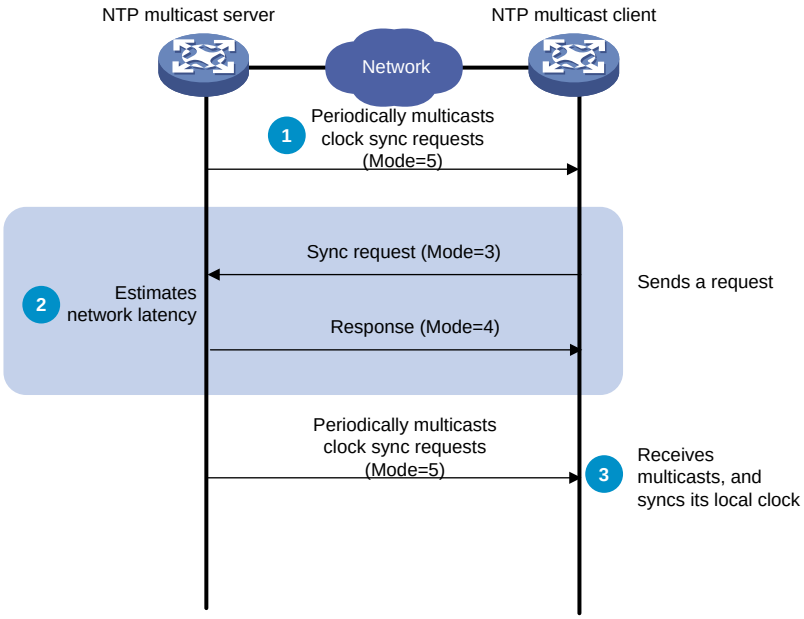
Operating mechanism for multicast mode

NTP multicast mode allows an NTP server to send time synchronization messages to user-configured IPv4 or IPv6 multicast addresses. A multicast server can provide time synchronization for clients in the same subnet or in different subnets. Only clients that have joined a multicast group can receive and process these messages.

The time synchronization process in NTP multicast mode is as follows:

1. The multicast server periodically sends clock synchronization request messages to the user-configured IPv4 or IPv6 multicast address, with the mode field set to 5, indicating multicast mode.
2. The broadcast client listens to the broadcast messages from the server. When the client receives the first broadcast message, the client and the server start to exchange messages to calculate the network delay between them.
3. The client enters multicast mode, continues to listen for subsequent multicast messages, and synchronizes its local clock based on the time information in these messages.

Figure 6 Operating mechanism for multicast mode



NTP security

To improve time synchronization security, NTP provides the access control and authentication functions.

NTP access control

You can control NTP access by using an ACL. The access rights are in the following order, from the least restrictive to the most restrictive:

- **Peer**—Allows time requests and NTP control queries (such as alarms, authentication status, and time server information) and allows the local device to synchronize itself to a peer device.
- **Server**—Allows time requests and NTP control queries, but does not allow the local device to synchronize itself to a peer device.
- **Synchronization**—Allows only time requests from a system whose address passes the access list criteria.
- **Query**—Allows only NTP control queries from a peer device to the local device.

When the device receives an NTP request, it matches the request against the access rights in order from the least restrictive to the most restrictive: **peer**, **server**, **synchronization**, and **query**.

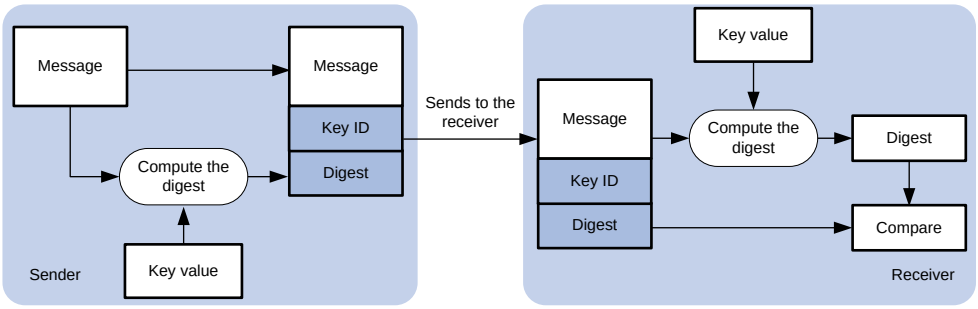
- If no NTP access control is configured, the **peer** access right applies.
- If the IP address of the peer device matches a **permit** statement in an ACL, the access right is granted to the peer device. If a **deny** statement or no ACL is matched, no access right is granted.
- If no ACL is specified for an access right or the ACL specified for the access right is not created, the access right is not granted.
- If none of the ACLs specified for the access rights is created, the **peer** access right applies.
- If none of the ACLs specified for the access rights contains rules, no access right is granted.

This feature provides minimal security for a system running NTP. A more secure method is NTP authentication.

NTP authentication

Use this feature to authenticate the NTP messages for security purposes. If an NTP message passes authentication, the device can receive it and get time synchronization information. If not, the device discards the message. This function makes sure the device does not synchronize to an unauthorized time server.

Figure 7 NTP authentication



As shown in Figure 7, NTP authentication is performed as follows:

2. The sender uses the key identified by the key ID to calculate a digest for the NTP message through the MD5/HMAC authentication algorithm. Then it sends the calculated digest together with the NTP message and key ID to the receiver.
3. Upon receiving the message, the receiver performs the following actions:
   a. Finds the key according to the key ID in the message.
   b. Uses the key and the MD5/HMAC authentication algorithm to calculate the digest for the message.
   c. Compares the digest with the digest contained in the NTP message.
      - If they are different, the receiver discards the message.
      - If they are the same and an NTP association is not required to be established, the receiver provides a response packet. For information about NTP associations, see "Configuring the maximum number of dynamic associations."
      - If they are the same and an NTP association is required to be established or has existed, the local device determines whether the sender is allowed to use the authentication ID. If the sender is allowed to use the authentication ID, the receiver accepts the message. If the sender is not allowed to use the authentication ID, the receiver discards the message.

Protocols and standards

- RFC 1305, *Network Time Protocol (Version 3) Specification, Implementation and Analysis*
- RFC 5905, *Network Time Protocol Version 4: Protocol and Algorithms Specification*

FIPS compliance

Higher security is required in FIPS mode. When the device operates in FIPS mode, the following restrictions apply for NTP:

- NTP can operate only in client/server mode and symmetric active/passive peer mode, and cannot operate in multicast or broadcast mode.
- Only NTPv3 and NTPv4 are supported. NTPv1 and NTPv2 are not supported.
- NTP supports the HMAC-SHA-1, HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 authentication algorithms and does not support the MD5 authentication algorithm.

For more information about FIPS mode, see *Security Configuration Guide*.

Restrictions and guidelines: NTP configuration

- You cannot configure both NTP and SNTP on the same device. NTP and SNTP communicate using UDP port 123 by default. If another service module uses this port, the device will fail to enable NTP and SNTP. If you enable NTP or SNTP, other service modules cannot use port 123.
- NTP is supported only on the following Layer 3 interfaces:
  - Layer 3 Ethernet interfaces.
  - Layer 3 Ethernet subinterfaces.
  - Layer 3 aggregate interfaces.
  - Layer 3 aggregate subinterfaces.
  - VLAN interfaces.
  - Tunnel interfaces.
- Do not configure NTP settings on an aggregate member port.
- The NTP service and SNTP service are mutually exclusive. You can only enable either NTP service or SNTP service at a time.
- To avoid frequent time changes or even synchronization failures, do not specify more than one reference source on a network.
- For correct time synchronization, make sure the time offset between the system time and the NTP clock source is less than 68 years.
- You must specify a Multitenant Device Context (MDC) or context in the **`clock protocol`** command for using NTP to obtain the time. For more information about the **`clock protocol`** command, see device management commands in *Fundamentals Command Reference*.

△ **CAUTION:**

When the device acts as an NTP client, the NTP server's clock stratum must be not smaller than 0 and not larger than 14. If the NTP server's clock stratum is larger than 14, the device will not synchronize with that server's clock. You can log in to the NTP server to change its clock stratum by executing the **`ntp-service refclock-master`** command.

- To check the stratum for an IPv4 NTP server, execute the **`display ntp-service sessions`** command on the device and check the **stra** field in the output.
- To check the stratum for an IPv6 NTP server, execute the **`display ntp-service ipv6 sessions`** command on the device and check the **Clock stratum** field in the output.

Client/server mode tasks at a glance

Tasks on the client

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. Configuring NTP in client/server mode
3. (Optional.) Configuring access control rights
4. (Optional.) Configuring NTP authentication in client/server mode
5. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages

- Configuring the maximum number of dynamic associations
- Setting a DSCP value for NTP packets

6. (Optional.) Controlling NTP time synchronization
   - Ignoring clock stratum level restrictions on NTP clients
   - Setting the maximum offset allowed for time synchronization through NTP
7. (Optional.) Specifying the NTP time-offset thresholds for log and trap outputs

Tasks on the server

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. Configuring NTP in client/server mode

   You must enable the NTP server. By default, the NTP server is enabled. You do not need to perform this step.
3. (Optional.) Configuring the local clock as the reference source
4. (Optional.) Configuring access control rights
5. (Optional.) Configuring NTP authentication in client/server mode
6. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages
   - Configuring the maximum number of dynamic associations
   - Setting a DSCP value for NTP packets

Symmetric active/passive mode tasks at a glance

Tasks on the active peer

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. Configuring NTP in symmetric active/passive mode
3. (Optional.) Configuring the local clock as the reference source
4. (Optional.) Configuring access control rights
5. (Optional.) Configuring NTP authentication in symmetric active/passive mode
6. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages
   - Configuring the maximum number of dynamic associations
   - Setting a DSCP value for NTP packets
7. (Optional.) Controlling NTP time synchronization
   - Ignoring clock stratum level restrictions on NTP clients
   - Setting the maximum offset allowed for time synchronization through NTP
8. (Optional.) Specifying the NTP time-offset thresholds for log and trap outputs

Tasks on the passive peer

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. (Optional.) Configuring the local clock as the reference source

3. (Optional.) Configuring access control rights
4. (Optional.) Configuring NTP authentication in symmetric active/passive mode
5. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages
   - Configuring the maximum number of dynamic associations
   - Setting a DSCP value for NTP packets
6. (Optional.) Controlling NTP time synchronization
   - Ignoring clock stratum level restrictions on NTP clients
   - Setting the maximum offset allowed for time synchronization through NTP
7. (Optional.) Specifying the NTP time-offset thresholds for log and trap outputs

Broadcast mode tasks at a glance

Tasks on the broadcast client

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. Configuring NTP authentication in broadcast mode
3. (Optional.) Configuring access control rights
4. Configuring NTP authentication in broadcast mode
5. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages
   - Configuring the maximum number of dynamic associations
   - Setting a DSCP value for NTP packets
6. (Optional.) Controlling NTP time synchronization
   - Ignoring clock stratum level restrictions on NTP clients
   - Setting the maximum offset allowed for time synchronization through NTP
7. (Optional.) Specifying the NTP time-offset thresholds for log and trap outputs

Tasks on the broadcast server

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. Configuring NTP in broadcast mode
3. (Optional.) Configuring the local clock as the reference source
4. (Optional.) Configuring access control rights
5. Configuring NTP authentication in broadcast mode
6. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages
   - Configuring the maximum number of dynamic associations
   - Setting a DSCP value for NTP packets

Multicast mode tasks at a glance

Tasks on the multicast client

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. Configuring NTP in multicast mode
3. (Optional.) Configuring access control rights
   - Configuring NTP authentication in multicast mode
4. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages
   - Configuring the maximum number of dynamic associations
   - Setting a DSCP value for NTP packets
5. (Optional.) Controlling NTP time synchronization
   - Ignoring clock stratum level restrictions on NTP clients
   - Setting the maximum offset allowed for time synchronization through NTP
6. (Optional.) Specifying the NTP time-offset thresholds for log and trap outputs

Tasks on the multicast server

1. Enabling the NTP service

   By default, the NTP service is enabled. You do not need to perform this step.
2. Configuring NTP in multicast mode
3. (Optional.) Configuring the local clock as the reference source
4. (Optional.) Configuring access control rights
5. (Optional.) Configuring NTP authentication in multicast mode
6. (Optional.) Controlling NTP message sending and receiving
   - Specifying a source address for NTP messages
   - Disabling an interface from receiving NTP messages
   - Configuring the maximum number of dynamic associations
   - Setting a DSCP value for NTP packets

Enabling the NTP service

Restrictions and guidelines

NTP and SNTP are mutually exclusive. Before you enable NTP, make sure SNTP is disabled.

Procedure

1. Enter system view.

   `system-view`
2. Enable the NTP service.

   `ntp-service enable`

   By default, the NTP service is disabled.

Configuring NTP association mode

Configuring NTP in client/server mode

About this task

By default, NTP uses UDP port 123 for communication. If another service module uses this port, the device will fail to enable NTP.

- When NTP service is enabled, the device will act as the NTP server or client and use UDP port 123 to provide NTP service. If another service module uses this port, the device will fail to enable NTP.
- For security purposes, the device will disable UDP port 123 and assign a currently free UDP port number to NTP if you disable IPv4 NTP server during operation of the NTP feature by executing the **undo ntp-service time-server enable** command. If IPv4 NTP server is enabled, the device will switch the NTP listening port back to 123.
- For security purposes, the device will disable UDP port 123 and assign a currently free UDP port number to NTP if you disable IPv6 NTP server during operation of the NTP feature by executing the **undo ntp-service ipv6 time-server enable** command. If IPv6 NTP server is enabled, the device will switch the IPv6 NTP listening port back to 123.

Restrictions and guidelines

To configure NTP in client/server mode, specify an NTP server for the client.

For a client to synchronize to an NTP server, make sure the server is synchronized by other devices or uses its local clock as the reference source.

If the stratum level of a server is higher than or equal to a client, the client will not synchronize to that server.

You can specify multiple servers for a client by executing the **ntp-service unicast-server** or **ntp-service ipv6 unicast-server** command multiple times.

Configuring the NTP client

1. Enter system view.

   system-view

2. Specify an NTP server for the device.

   IPv4:

   ntp-service unicast-server { *server-name* | *ip-address* } [**vpn-instance** *vpn-instance-name*] [**authentication-keyid** *keyid* | **maxpoll** *maxpoll-interval* | **minpoll** *minpoll-interval* | **priority** | **source** *interface-type interface-number* | **version** *number*] *

   IPv6:

   ntp-service ipv6 unicast-server { *server-name* | *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**authentication-keyid** *keyid* | **maxpoll** *maxpoll-interval* | **minpoll** *minpoll-interval* | **priority** | **source** *interface-type interface-number*] *

   By default, no NTP server is specified.

Configuring the NTP server

1. Enter system view.

   system-view

2. Enable NTP server.

   IPv4 network:

ntp-service time-server enable

By default, NTP server is enabled.

IPv6 network:

ntp-service ipv6 time-server enable

By default, IPv6 NTP server is enabled.

A device can synchronize the time of other devices through NTP only when it meets the following conditions:

- NTP server is enabled on the device by using the **ntp-service time-server enable** command.
- The device is permitted by the ACLs configured in the **ntp-service acl** or **ntp-service ipv6 acl** command on the other devices.

If you disable NTP server on the device, the device cannot synchronize the time of other devices through NTP.

Configuring NTP in symmetric active/passive mode

Restrictions and guidelines

To configure NTP in symmetric active/passive mode, specify a symmetric passive peer for the active peer.

For a symmetric passive peer to process NTP messages from a symmetric active peer, execute the **ntp-service enable** command on the symmetric passive peer to enable NTP.

For time synchronization between the symmetric active peer and the symmetric passive peer, make sure either or both of them are in synchronized state.

You can specify multiple symmetric passive peers by executing the **ntp-service unicast-peer** or **ntp-service ipv6 unicast-peer** command multiple times.

In IPv4 NTP symmetric peers mode, broadcast mode, or multicast mode, do not execute the **undo ntp-service time-server enable** command, no matter whether the device acts as an NTP client or server. If you fail to do so, UDP port 123 in the IPv4 protocol stack will be disabled, affecting NTP clock synchronization.

In IPv6 NTP symmetric peers mode, broadcast mode, or multicast mode, do not execute the **undo ntp-service ipv6 time-server enable** command, no matter whether the device acts as an IPv6 NTP client or server. If you fail to do so, UDP port 123 in the IPv6 protocol stack will be disabled, affecting IPv6 NTP clock synchronization.

Configuring an active peer

1. Enter system view.

   system-view

2. Specify a symmetric passive peer for the device.

   IPv4:

   ntp-service unicast-peer { *peer-name* | *ip-address* } [**vpn-instance** *vpn-instance-name*] [**authentication-keyid** *keyid* | **maxpoll** *maxpoll-interval* | **minpoll** *minpoll-interval* | **priority** | **source** *interface-type interface-number* | **version** *number*] *

   IPv6:

   ntp-service ipv6 unicast-peer { *peer-name* | *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**authentication-keyid** *keyid* | **maxpoll** *maxpoll-interval* | **minpoll** *minpoll-interval* | **priority** | **source** *interface-type interface-number*] *

By default, no symmetric passive peer is specified.

3. Enable NTP server.

   IPv4 network:

   ntp-service time-server enable

   By default, NTP server is enabled.

   IPv6 network:

   ntp-service ipv6 time-server enable

   By default, IPv6 NTP server is enabled.

   A device can synchronize the time of other devices through NTP only when it meets the following conditions:

   - NTP server is enabled on the device by using the **ntp-service time-server enable** command.
   - The device is permitted by the ACLs configured in the **ntp-service acl** or **ntp-service ipv6 acl** command on the other devices.

   If you disable NTP server on the device, the device cannot synchronize the time of other devices through NTP.

Configuring a passive peer

1. Enter system view.

   system-view

2. Enable NTP server.

   IPv4:

   ntp-service time-server enable

   By default, IPv4 NTP server is enabled.

   IPv6:

   ntp-service ipv6 time-server enable

   By default, IPv6 NTP server is enabled.

   A device can provide NTP synchronization to other devices only when it meets the following conditions:

   - NTP server is enabled on the device.
   - The device is permitted by the ACLs configured in the **ntp-service acl** or **ntp-service ipv6 acl** command on the other devices.

   If you disable NTP server on the device, the device cannot synchronize the time of other devices through NTP or synchronize time to the NTP server in symmetric peers, broadcast, or multicast mode.

Configuring NTP in broadcast mode

Restrictions and guidelines

To configure NTP in broadcast mode, you must configure an NTP broadcast client and an NTP broadcast server.

For a broadcast client to synchronize to a broadcast server, make sure the broadcast server is synchronized by other devices or uses its local clock as the reference source.

Configuring the broadcast client

1. Enter system view.

   system-view

2. Enable NTP server.

   IPv4:

   ntp-service time-server enable

   By default, IPv4 NTP server is enabled.

   IPv6:

   ntp-service ipv6 time-server enable

   By default, IPv6 NTP server is enabled.

   A device can provide NTP synchronization to other devices only when it meets the following conditions:

   - NTP server is enabled on the device.
   - The device is permitted by the ACLs configured in the **ntp-service acl** or **ntp-service ipv6 acl** command on the other devices.

   If you disable NTP server on the device, the device cannot synchronize the time of other devices through NTP or synchronize time to the NTP server in symmetric peers, broadcast, or multicast mode.

3. Enter interface view.

   interface *interface-type interface-number*

4. Configure the device to operate in broadcast client mode.

   ntp-service broadcast-client

   By default, the device does not operate in any NTP association mode.

   After you execute the command, the device receives NTP broadcast messages from the specified interface.

Configuring the broadcast server

1. Enter system view.

   system-view

2. Enable NTP server.

   IPv4 network:

   ntp-service time-server enable

   By default, NTP server is enabled.

   IPv6 network:

   ntp-service ipv6 time-server enable

   By default, IPv6 NTP server is enabled.

   A device can synchronize the time of other devices through NTP only when it meets the following conditions:

   - NTP server is enabled on the device by using the **ntp-service time-server enable** command.
   - The device is permitted by the ACLs configured in the **ntp-service acl** or **ntp-service ipv6 acl** command on the other devices.

   If you disable NTP server on the device, the device cannot synchronize the time of other devices through NTP or synchronize time to the NTP server in symmetric peers, broadcast, or multicast mode..

3. Enter interface view.

   interface *interface-type interface-number*

4. Configure the device to operate in NTP broadcast server mode.

   ntp-service broadcast-server [**authentication-keyid** *keyid* | **version** *number*] *

By default, the device does not operate in any NTP association mode.

After you execute the command, the device sends NTP broadcast messages from the specified interface.

Configuring NTP in multicast mode

Restrictions and guidelines

To configure NTP in multicast mode, you must configure an NTP multicast client and an NTP multicast server.

For a multicast client to synchronize to a multicast server, make sure the multicast server is synchronized by other devices or uses its local clock as the reference source.

Configuring a multicast client

1. Enter system view.

   system-view

2. Enable NTP server.

   IPv4:

   ntp-service time-server enable

   By default, IPv4 NTP server is enabled.

   IPv6:

   ntp-service ipv6 time-server enable

   By default, IPv6 NTP server is enabled.

   A device can provide NTP synchronization to other devices only when it meets the following conditions:

   - NTP server is enabled on the device.
   - The device is permitted by the ACLs configured in the **ntp-service acl** or **ntp-service ipv6 acl** command on the other devices.

   If you disable NTP server on the device, the device cannot synchronize the time of other devices through NTP or synchronize time to the NTP server in symmetric peers, broadcast, or multicast mode.

3. Enter interface view.

   interface *interface-type interface-number*

4. Configure the device to operate in multicast client mode.

   IPv4:

   ntp-service multicast-client [*ip-address*]

   IPv6:

   ntp-service ipv6 multicast-client *ipv6-address*

   By default, the device does not operate in any NTP association mode.

   After you execute the command, the device receives NTP multicast messages from the specified interface.

Configuring the multicast server

1. Enter system view.

   system-view

2. Enable NTP server.

   IPv4 network:

   ntp-service time-server enable

By default, NTP server is enabled.

IPv6 network:

ntp-service ipv6 time-server enable

By default, IPv6 NTP server is enabled.

A device can synchronize the time of other devices through NTP only when it meets the following conditions:

- NTP server is enabled on the device by using the **ntp-service time-server enable** command.
- The device is permitted by the ACLs configured in the **ntp-service acl** or **ntp-service ipv6 acl** command on the other devices.

If you disable NTP server on the device, the device cannot synchronize the time of other devices through NTP.

3. Enter interface view.

   interface *interface-type interface-number*

4. Configure the device to operate in multicast server mode.

   IPv4:

   ntp-service multicast-server [*ip-address*][**authentication-keyid** *keyid* | **ttl** *ttl-number* | **version** *number*]*

   IPv6:

   ntp-service ipv6 multicast-server *ipv6-address* [**authentication-keyid** *keyid* | **ttl** *ttl-number*] *

   By default, the device does not operate in any NTP association mode.

   After you execute the command, the device sends NTP multicast messages from the specified interface.

Configuring the local clock as the reference source

About this task

The clock stratum of an NTP server that obtains time synchronization from an authoritative clock (such as an atomic clock) is typically set to 1 and used as the primary time server to synchronize the clocks of other devices in the network. The NTP distance between devices and the primary time server in the network, which is the number of servers in the NTP synchronization chain, determines the stratum of the device clocks.

If the devices in a network cannot synchronize to an authoritative time source, you can perform the following tasks:

- Select a device that has a relatively accurate clock from the network.
- Use the local clock of the device as the reference clock to synchronize other devices in the network.

Restrictions and guidelines

Make sure the local clock can provide the time accuracy required for the network. After you configure the local clock as the reference source, the local clock is synchronized, and can operate as a time server to synchronize other devices in the network. If the local clock is incorrect, timing errors occur.

The system time reverts to the initial BIOS default after a cold reboot. The system clock stops and does not record the passing of time during a warm reboot. As a best practice, do not configure the local clock as the reference source or configure the device as a time server.

Devices differ in clock precision. As a best practice to avoid network flapping and clock synchronization failure, configure only one reference clock on the same network segment and make sure the clock has high precision.

Prerequisites

Before you configure this feature, adjust the local system time to ensure that it is accurate.

Procedure

1. Enter system view.
   system-view
2. Configure the local clock as the reference source.
   ntp-service refclock-master [*ip-address*] [*stratum*]
   By default, the device does not use the local clock as the reference source.

Configuring access control rights

Prerequisites

Before you configure the right for peer devices to access the NTP services on the local device, create and configure ACLs associated with the access right. For information about configuring an ACL, see *ACL and QoS Configuration Guide*.

Restrictions and guidelines

Table 2 Restrictions and guidelines for configuring NTP access control rights

Access control right	Whether the time can be synchronized (whether configurable on a client)	Whether can synchronize the time of other devices (whether configurable on a time server)	Whether allows control queries
Peer	Yes	Yes	Yes
Server	No	Yes	Yes
Synchronization	No	Yes	No
Query	No	No	Yes

The **ntp-service noquery enable** command and its undo form are used only to configure the device to disallow or allow control queries and do not disable or enable clock synchronization. If the **ntp-service noquery enable** command or its undo form and the **ntp-service acl** or **ntp-service ipv6 acl** command are both configured, the **ntp-service noquery enable** command or its undo form determines whether control queries are allowed.

Procedure

1. Enter system view.
   system-view
2. Configure the right for peer devices to access the NTP services on the local device.
   IPv4:
   ntp-service access { **peer** | **query** | **server** | **synchronization** } **acl** *ipv4-acl-number*
   IPv6:

ntp-service ipv6 { **peer** |**query** | **server** | **synchronization** } **acl** *ipv6-acl-number*

By default, the right for peer devices to access the NTP services on the local device is **peer**.

3. Disallow control queries for the local device.

   ntp-service noquery enable

   By default, control queries for the local device are allowed.

Configuring NTP authentication

Configuring NTP authentication in client/server mode

Restrictions and guidelines

To ensure a successful NTP authentication in client/server mode, configure the same authentication key ID, algorithm, and key on the server and client. Make sure the peer device is allowed to use the key ID for authentication on the local device.

NTP authentication results differ when different configurations are performed on client and server. For more information, see Table 3. (N/A in the table means that whether the configuration is performed or not does not make any difference.)

Table 3 NTP authentication results

Client			Server	
Enable NTP authentication	**Specify the server and key**	**Trusted key**	**Enable NTP authentication**	**Trusted key**
Successful authentication				
Yes	Yes	Yes	Yes	Yes
Failed authentication				
Yes	Yes	Yes	Yes	No
Yes	Yes	Yes	No	N/A
Yes	Yes	No	N/A	N/A
Authentication not performed				
Yes	No	N/A	N/A	N/A
No	N/A	N/A	N/A	N/A

Configuring NTP authentication for a client

1. Enter system view.

   system-view

2. Enable NTP authentication.

   ntp-service authentication enable

   By default, NTP authentication is disabled.

3. Configure an NTP authentication key.

   ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

By default, no NTP authentication key exists.

4. Configure the key as a trusted key.

   ntp-service reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.

5. Associate the specified key with an NTP server.

   IPv4:

   ntp-service unicast-server { *server-name* | *ip-address* } [**vpn-instance** *vpn-instance-name*] **authentication-keyid** *keyid*

   IPv6:

   ntp-service ipv6 unicast-server { *server-name* | *ipv6-address* } [**vpn-instance** *vpn-instance-name*] **authentication-keyid** *keyid*

Configuring NTP authentication for a server

1. Enter system view.

   system-view

2. Enable NTP authentication.

   ntp-service authentication enable

   By default, NTP authentication is disabled.

3. Configure an NTP authentication key.

   ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

   By default, no NTP authentication key exists.

4. Configure the key as a trusted key.

   ntp-service reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.

Configuring NTP authentication in symmetric active/passive mode

Restrictions and guidelines

To ensure a successful NTP authentication in symmetric active/passive mode, configure the same authentication key ID, algorithm, and key on the active peer and passive peer. Make sure the peer device is allowed to use the key ID for authentication on the local device.

NTP authentication results differ when different configurations are performed on active peer and passive peer. For more information, see Table 4. (N/A in the table means that whether the configuration is performed or not does not make any difference.)

Table 4 NTP authentication results

Active peer				Passive peer	
Enable NTP authentication	**Specify the peer and key**	**Trusted key**	**Stratum level**	**Enable NTP authentication**	**Trusted key**
Successful authentication					
Yes	Yes	Yes	N/A	Yes	Yes

Active peer				Passive peer	
Enable NTP authentication	**Specify the peer and key**	**Trusted key**	**Stratum level**	**Enable NTP authentication**	**Trusted key**
Failed authentication					
Yes	Yes	Yes	N/A	Yes	No
Yes	Yes	Yes	N/A	No	N/A
Yes	No	N/A	N/A	Yes	N/A
No	N/A	N/A	N/A	Yes	N/A
Yes	Yes	No	Larger than the passive peer	N/A	N/A
Yes	Yes	No	Smaller than the passive peer	Yes	N/A
Authentication not performed					
Yes	No	N/A	N/A	No	N/A
No	N/A	N/A	N/A	No	N/A
Yes	Yes	No	Smaller than the passive peer	No	N/A

Configuring NTP authentication for an active peer

1. Enter system view.

   system-view
2. Enable NTP authentication.

   ntp-service authentication enable

   By default, NTP authentication is disabled.
3. Configure an NTP authentication key.

   ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

   By default, no NTP authentication key exists.
4. Configure the key as a trusted key.

   ntp-service reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.
5. Associate the specified key with a passive peer.

   IPv4:

   ntp-service unicast-peer { *ip-address* | *peer-name* } [**vpn-instance** *vpn-instance-name*] **authentication-keyid** *keyid*

   IPv6:

   ntp-service ipv6 unicast-peer { *ipv6-address* | *peer-name* } [**vpn-instance** *vpn-instance-name*] **authentication-keyid** *keyid*

Configuring NTP authentication for a passive peer

1. Enter system view.

   system-view
2. Enable NTP authentication.

ntp-service authentication enable

By default, NTP authentication is disabled.

3. Configure an NTP authentication key.

ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

By default, no NTP authentication key exists.

4. Configure the key as a trusted key.

ntp-service reliable authentication-keyid *keyid*

By default, no authentication key is configured as a trusted key.

Configuring NTP authentication in broadcast mode

Restrictions and guidelines

To ensure a successful NTP authentication in broadcast mode, configure the same authentication key ID, algorithm, and key on the broadcast server and client. Make sure the peer device is allowed to use the key ID for authentication on the local device.

NTP authentication results differ when different configurations are performed on broadcast client and server. For more information, see Table 5. (N/A in the table means that whether the configuration is performed or not does not make any difference.)

Table 5 NTP authentication results

Broadcast server			Broadcast client	
Enable NTP authentication	**Specify the server and key**	**Trusted key**	**Enable NTP authentication**	**Trusted key**
Successful authentication				
Yes	Yes	Yes	Yes	Yes
Failed authentication				
Yes	Yes	Yes	Yes	No
Yes	Yes	Yes	No	N/A
Yes	Yes	No	Yes	N/A
Yes	No	N/A	Yes	N/A
No	N/A	N/A	Yes	N/A
Authentication not performed				
Yes	Yes	No	No	N/A
Yes	No	N/A	No	N/A
No	N/A	N/A	No	N/A

Configuring NTP authentication for a broadcast client

1. Enter system view.

system-view

2. Enable NTP authentication.

ntp-service authentication enable

By default, NTP authentication is disabled.

3. Configure an NTP authentication key.

   ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

   By default, no NTP authentication key exists.

4. Configure the key as a trusted key.

   ntp-service reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.

Configuring NTP authentication for a broadcast server

1. Enter system view.

   system-view

2. Enable NTP authentication.

   ntp-service authentication enable

   By default, NTP authentication is disabled.

3. Configure an NTP authentication key.

   ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

   By default, no NTP authentication key exists.

4. Configure the key as a trusted key.

   ntp-service reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.

5. Enter interface view.

   interface *interface-type interface-number*

6. Associate the specified key with the broadcast server.

   ntp-service broadcast-server authentication-keyid *keyid*

   By default, the broadcast server is not associated with a key.

Configuring NTP authentication in multicast mode

Restrictions and guidelines

To ensure a successful NTP authentication in multicast mode, configure the same authentication key ID, algorithm, and key on the multicast server and client. Make sure the peer device is allowed to use the key ID for authentication on the local device.

NTP authentication results differ when different configurations are performed on broadcast client and server. For more information, see Table 6. (N/A in the table means that whether the configuration is performed or not does not make any difference.)

Table 6 NTP authentication results

Multicast server			Multicast client	
Enable NTP authentication	**Specify the server and key**	**Trusted key**	**Enable NTP authentication**	**Trusted key**
Successful authentication				
Yes	Yes	Yes	Yes	Yes
Failed authentication				
Yes	Yes	Yes	Yes	No
Yes	Yes	Yes	No	N/A
Yes	Yes	No	Yes	N/A
Yes	No	N/A	Yes	N/A
No	N/A	N/A	Yes	N/A
Authentication not performed				
Yes	Yes	No	No	N/A
Yes	No	N/A	No	N/A
No	N/A	N/A	No	N/A

Configuring NTP authentication for a multicast client

1. Enter system view.

   system-view
2. Enable NTP authentication.

   ntp-service authentication enable

   By default, NTP authentication is disabled.
3. Configure an NTP authentication key.

   ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

   By default, no NTP authentication key exists.
4. Configure the key as a trusted key.

   ntp-service reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.

Configuring NTP authentication for a multicast server

1. Enter system view.

   system-view
2. Enable NTP authentication.

   ntp-service authentication enable

   By default, NTP authentication is disabled.
3. Configure an NTP authentication key.

   ntp-service authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

By default, no NTP authentication key exists.

4. Configure the key as a trusted key.

   ntp-service reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.

5. Enter interface view.

   interface *interface-type interface-number*

6. Associate the specified key with a multicast server.

   IPv4:

   ntp-service multicast-server [*ip-address*] **authentication-keyid** *keyid*

   IPv6:

   ntp-service ipv6 multicast-server *ipv6-multicast-address* **authentication-keyid** *keyid*

   By default, no multicast server is associated with the specified key.

Controlling NTP message sending and receiving

Specifying a source address for NTP messages

About this task

You can specify a source address for NTP messages directly or by specifying a source interface. If you specify a source interface for NTP messages, the device uses the IP address of the specified interface as the source address to send NTP messages.

Restrictions and guidelines

To prevent interface status changes from causing NTP communication failures, specify an interface that is always up as the source interface, a loopback interface for example.

When the device responds to an NTP request, the source IP address of the NTP response is always the IP address of the interface that has received the NTP request.

If you have specified the source interface for NTP messages in the **ntp-service unicast-server/ntp-service ipv6 unicast-server** or **ntp-service unicast-peer/ntp-service ipv6 unicast-peer** command, the IP address of the specified interface is used as the source IP address for NTP messages.

If you have configured the **ntp-service broadcast-server** or **ntp-service multicast-server/ntp-service ipv6 multicast-server** command in an interface view, the IP address of the interface is used as the source IP address for broadcast or multicast NTP messages.

Procedure

1. Enter system view.

   system-view

2. Specify the source address for NTP messages.

   IPv4:

   ntp-service source { *interface-type interface-number* | *ipv4-address* }

   IPv6:

   ntp-service ipv6 source *interface-type interface-number*

   By default, no source address is specified for NTP messages.

Disabling an interface from receiving NTP messages

About this task

When NTP is enabled, all interfaces by default can receive NTP messages. For security purposes, you can disable some of the interfaces from receiving NTP messages.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Disable the interface from receiving NTP packets.
   IPv4:
   undo ntp-service inbound enable
   IPv6:
   undo ntp-service ipv6 inbound enable
   By default, an interface receives NTP messages.

Configuring the maximum number of dynamic associations

About this task

Perform this task to restrict the number of dynamic associations to prevent dynamic associations from occupying too many system resources.

NTP has the following types of associations:

- **Static association**—A manually created association.
- **Dynamic association**—Temporary association created by the system during NTP operation. A dynamic association is removed if no messages are exchanged within about 12 minutes.

The following describes how an association is established in different association modes:

- **Client/server mode**—After you specify an NTP server, the system creates a static association on the client. The server simply responds passively upon the receipt of a message, rather than creating an association (static or dynamic).
- **Symmetric active/passive mode**—After you specify a symmetric passive peer on a symmetric active peer, static associations are created on the symmetric active peer, and dynamic associations are created on the symmetric passive peer.
- **Broadcast or multicast mode**—Static associations are created on the server, and dynamic associations are created on the client.

Restrictions and guidelines

A single device can have a maximum of 128 concurrent associations, including static associations and dynamic associations. After this command is configured, new associations will not be established if the maximum number of associations has been reached or exceeded.

Procedure

1. Enter system view.
   system-view
2. Configure the maximum number of dynamic sessions.
   ntp-service max-dynamic-sessions *number*
   By default, the maximum number of dynamic sessions is 100.

Setting a DSCP value for NTP packets

About this task

Configure this task to specify the sending precedence for NTP packets. This feature allows you to set a DSCP value for NTP packets when the device acts as an NTP or IPv6 NTP server or client.

Procedure

1. Enter system view.
   system-view
2. Set a DSCP value for NTP packets.
   IPv4:
   ntp-service dscp *dscp-value*
   IPv6:
   ntp-service ipv6 dscp *dscp-value*
   The default DSCP value is 48 for IPv4 packets and 56 for IPv6 packets.

Controlling NTP time synchronization

Ignoring clock stratum level restrictions on NTP clients

About this task

Application scenarios

Perform this task to control whether the device responds to time synchronization requests from NTP clients when it acts as an NTP server.

Operating mechanism

By default, when an NTP server receives a clock synchronization request from an NTP client, it checks the stratum level of the client. The server responds to the request and synchronizes the clock only if the client has a higher stratum level than the server. In some network environments, if a client has a stratum level lower than or equal to the server and changing the stratum level is not supported, you can perform this task to ignore the stratum level restrictions on NTP clients.

Procedure

1. Enter system view.
   system-view
2. Configure the device to not restrict the stratum level of NTP clients.
   ntp-service ignore clock-stratum
   By default, an NTP server restricts the stratum level of NTP clients when it synchronizes the clock with the NTP clients.

Setting the maximum offset allowed for time synchronization through NTP

About this task

Application scenarios

This feature controls whether the device synchronizes the time with an NTP server when it acts as an NTP client.

Operating mechanism

When the device acts as an NTP client, it subtracts the local time from the time provided by the NTP server to calculate the time offset. The device will adjust the local NTP time based on this offset. If this offset exceeds the configured maximum offset, the device will determine that the time from the NTP server is unreliable and will not synchronize the time from the NTP server.

Procedure

1. Enter system view.
   system-view
2. Set the maximum offset allowed for time synchronization through NTP.
   ntp-service time-offset-limit *limit-threshold*
   By default, no maximum offset is set for time synchronization through NTP.

Specifying the NTP time-offset thresholds for log and trap outputs

About this task

By default, the system synchronizes the NTP client's time to the server and outputs a log and a trap when the time offset exceeds 128 ms for multiple times.

After you set the NTP time-offset thresholds for log and trap outputs, the system synchronizes the client's time to the server when the time offset exceeds 128 ms for multiple times, but outputs logs and traps only when the time offset exceeds the specified thresholds, respectively.

Procedure

1. Enter system view.
   system-view
2. Specify the NTP time-offset thresholds for log and trap outputs.
   ntp-service time-offset-threshold { **log** *log-threshold* | **trap** *trap-threshold* } *
   By default, no NTP time-offset thresholds are set for log and trap outputs.

Display and maintenance commands for NTP

The **ntp-service noquery enable** command is mutually exclusive with the **display ntp-service trace** command. To use the **display ntp-service trace** command, you must configure the device to allow control queries.

Execute **display** commands in any view.

Task	Command
Display information about IPv6 NTP associations.	**display ntp-service ipv6 sessions** [**verbose**]
Display information about IPv4 NTP associations.	**display ntp-service sessions** [**verbose**]
Display information about NTP service status.	**display ntp-service status**
Display brief information about the NTP servers from the local device back to the primary NTP server.	**display ntp-service trace** [**source** *interface-type interface-number*]

NTP configuration examples

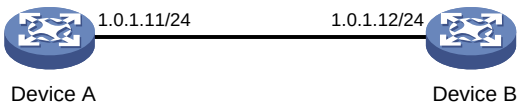
Example: Configuring NTP client/server association mode

Network configuration

As shown in Figure 8, perform the following tasks:

- Configure Device A's local clock as its reference source, with stratum level 2.
- Configure Device B to operate in client mode and specify Device A as the NTP server of Device B.

Figure 8 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Device A and Device B can reach each other, as shown in Figure 8. (Details not shown.)
2. Configure Device A:

   # Enable the NTP service.

   ```
   <DeviceA> system-view
   [DeviceA] ntp-service enable
   ```

   # Specify the local clock as the reference source, with stratum level 2.

   ```
   [DeviceA] ntp-service refclock-master 2
   ```
3. Configure Device B:

   # Enable the NTP service.

   ```
   <DeviceB> system-view
   [DeviceB] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [DeviceB] clock protocol ntp
   ```

   # Specify Device A as the NTP server of Device B.

   ```
   [DeviceB] ntp-service unicast-server 1.0.1.11
   ```

Verifying the configuration

Verify that Device B has synchronized its time with Device A, and the clock stratum level of Device B is 3.

```
[DeviceB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 1.0.1.11
 Local mode: client
 Reference clock ID: 1.0.1.11
 Leap indicator: 00
 Clock jitter: 0.000977 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00383 ms
```

```
 Root dispersion: 16.26572 ms
 Reference time: d0c6033f.b9923965  Wed, Dec 29 2019 18:58:07.724
 System poll interval: 64 s
```

Verify that an IPv4 NTP association has been established between Device B and Device A.

```
[DeviceB] display ntp-service sessions
      source          reference       stra reach poll  now offset  delay disper
********************************************************************************
[12345]1.0.1.11        127.127.1.0        2     1   64   15   -4.0 0.0038 16.262
Notes: 1 source(master), 2 source(peer), 3 selected, 4 candidate, 5 configured.
 Total sessions: 1
```

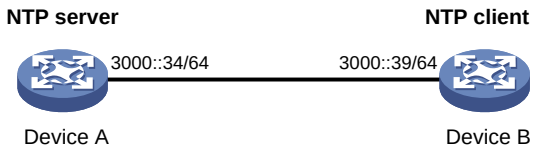
Example: Configuring IPv6 NTP client/server association mode

Network configuration

As shown in Figure 9, perform the following tasks:

- Configure Device A's local clock as its reference source, with stratum level 2.
- Configure Device B to operate in client mode and specify Device A as the IPv6 NTP server of Device B.

Figure 9 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Device A and Device B can reach each other, as shown in Figure 9. (Details not shown.)
2. Configure Device A:

   # Enable the NTP service.

   ```
   <DeviceA> system-view
   [DeviceA] ntp-service enable
   ```

   # Specify the local clock as the reference source, with stratum level 2.

   ```
   [DeviceA] ntp-service refclock-master 2
   ```

3. Configure Device B:

   # Enable the NTP service.

   ```
   <DeviceB> system-view
   [DeviceB] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [DeviceB] clock protocol ntp
   ```

   # Specify Device A as the IPv6 NTP server of Device B.

   ```
   [DeviceB] ntp-service ipv6 unicast-server 3000::34
   ```

Verifying the configuration

Verify that Device B has synchronized its time with Device A, and the clock stratum level of Device B is 3.

```
[DeviceB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3000::34
 Local mode: client
 Reference clock ID: 163.29.247.19
 Leap indicator: 00
 Clock jitter: 0.000977 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.02649 ms
 Root dispersion: 12.24641 ms
 Reference time: d0c60419.9952fb3e  Wed, Dec 29 2019 19:01:45.598
 System poll interval: 64 s
```

Verify that an IPv6 NTP association has been established between Device B and Device A.

```
[DeviceB] display ntp-service ipv6 sessions
Notes: 1 source(master), 2 source(peer), 3 selected, 4 candidate, 5 configured.

 Source: [12345]3000::34
 Reference: 127.127.1.0          Clock stratum: 2
 Reachabilities: 15              Poll interval: 64
 Last receive time: 19           Offset: 0.0
 Roundtrip delay: 0.0            Dispersion: 0.0

 Total sessions: 1
```

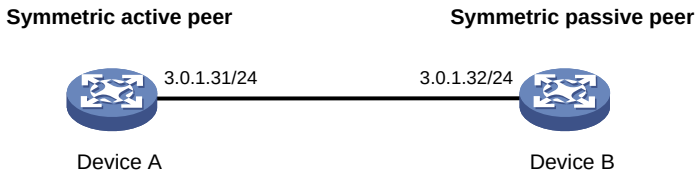
Example: Configuring NTP symmetric active/passive association mode

Network configuration

As shown in Figure 10, perform the following tasks:

- Configure Device A's local clock as its reference source, with stratum level 2.
- Configure Device A to operate in symmetric active mode and specify Device B as the passive peer of Device A.

Figure 10 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Device A and Device B can reach each other, as shown in Figure 10. (Details not shown.)
2. Configure Device B:

   # Enable the NTP service.

```
<DeviceB> system-view
[DeviceB] ntp-service enable
```

Specify NTP for obtaining the time.

```
[DeviceB] clock protocol ntp
```

3. Configure Device A:

Enable the NTP service.

```
<DeviceA> system-view
[DeviceA] ntp-service enable
```

Specify NTP for obtaining the time.

```
[DeviceA] clock protocol ntp
```

Specify the local clock as the reference source, with stratum level 2.

```
[DeviceA] ntp-service refclock-master 2
```

Configure Device B as its symmetric passive peer.

```
[DeviceA] ntp-service unicast-peer 3.0.1.32
```

Verifying the configuration

Verify that Device B has synchronized its time with Device A.

```
[DeviceB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3.0.1.31
 Local mode: sym_passive
 Reference clock ID: 3.0.1.31
 Leap indicator: 00
 Clock jitter: 0.000916 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00609 ms
 Root dispersion: 1.95859 ms
 Reference time: 83aec681.deb6d3e5  Wed, Jan  8 2019 14:33:11.081
 System poll interval: 64 s
```

Verify that an IPv4 NTP association has been established between Device B and Device A.

```
[DeviceB] display ntp-service sessions
       source          reference       stra reach poll  now offset  delay disper
**************************************************************************
   [12]3.0.1.31        127.127.1.0        2    62   64   34 0.4251 6.0882 1392.1
Notes: 1 source(master), 2 source(peer), 3 selected, 4 candidate, 5 configured.
 Total sessions: 1
```

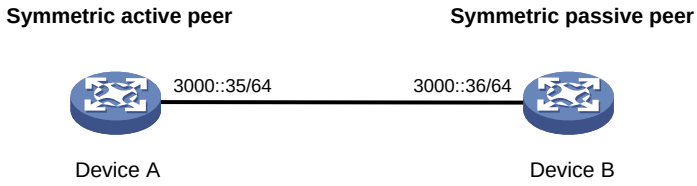
Example: Configuring IPv6 NTP symmetric active/passive association mode

Network configuration

As shown in Figure 11, perform the following tasks:

- Configure Device A's local clock as its reference source, with stratum level 2.
- Configure Device A to operate in symmetric active mode and specify Device B as the IPv6 passive peer of Device A.

Figure 11 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Device A and Device B can reach each other, as shown in Figure 11. (Details not shown.)
2. Configure Device B:

   # Enable the NTP service.

   ```
   <DeviceB> system-view
   [DeviceB] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [DeviceB] clock protocol ntp
   ```

3. Configure Device A:

   # Enable the NTP service.

   ```
   <DeviceA> system-view
   [DeviceA] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [DeviceA] clock protocol ntp
   ```

   # Specify the local clock as the reference source, with stratum level 2.

   ```
   [DeviceA] ntp-service refclock-master 2
   ```

   # Configure Device B as the IPv6 symmetric passive peer.

   ```
   [DeviceA] ntp-service ipv6 unicast-peer 3000::36
   ```

Verifying the configuration

Verify that Device B has synchronized its time with Device A.

```
[DeviceB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3000::35
 Local mode: sym_passive
 Reference clock ID: 251.73.79.32
 Leap indicator: 11
 Clock jitter: 0.000977 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.01855 ms
 Root dispersion: 9.23483 ms
 Reference time: d0c6047c.97199f9f  Wed, Dec 29 2019 19:03:24.590
 System poll interval: 64 s
```

Verify that an IPv6 NTP association has been established between Device B and Device A.

```
[DeviceB] display ntp-service ipv6 sessions
Notes: 1 source(master), 2 source(peer), 3 selected, 4 candidate, 5 configured.
```

```
Source:  [1234]3000::35
Reference: 127.127.1.0          Clock stratum: 2
Reachabilities: 15              Poll interval: 64
Last receive time: 19           Offset: 0.0
Roundtrip delay: 0.0            Dispersion: 0.0

Total sessions: 1
```

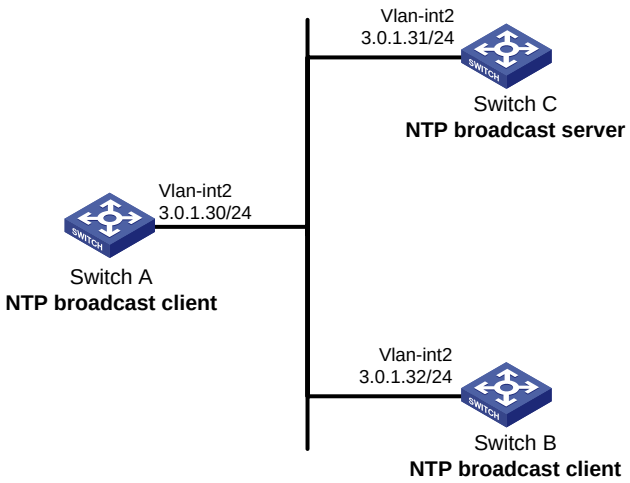
Example: Configuring NTP broadcast association mode

Network configuration

As shown in Figure 12, configure Switch C as the NTP server for multiple devices on the same network segment so that these devices synchronize the time with Switch C.

- Configure Switch C's local clock as its reference source, with stratum level 2.
- Configure Switch C to operate in broadcast server mode and send broadcast messages from VLAN-interface 2.
- Configure Switch A and Switch B to operate in broadcast client mode, and listen to broadcast messages on VLAN-interface 2.

Figure 12 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Switch A, Switch B, and Switch C can reach each other, as shown in Figure 12. (Details not shown.)
2. Configure Switch C:

   # Enable the NTP service.

   ```
   <SwitchC> system-view
   [SwitchC] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchC] clock protocol ntp
   ```

   # Specify the local clock as the reference source, with stratum level 2.

   ```
   [SwitchC] ntp-service refclock-master 2
   ```

   # Configure Switch C to operate in broadcast server mode and send broadcast messages from VLAN-interface 2.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server
```

3. Configure Switch A:

   # Enable the NTP service.

```
<SwitchA> system-view
[SwitchA] ntp-service enable
```

   # Specify NTP for obtaining the time.

```
[SwitchA] clock protocol ntp
```

   # Configure Switch A to operate in broadcast client mode and receive broadcast messages on VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ntp-service broadcast-client
```

4. Configure Switch B:

   # Enable the NTP service.

```
<SwitchB> system-view
[SwitchB] ntp-service enable
```

   # Specify NTP for obtaining the time.

```
[SwitchB] clock protocol ntp
```

   # Configure Switch B to operate in broadcast client mode and receive broadcast messages on VLAN-interface 2.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ntp-service broadcast-client
```

Verifying the configuration

The following procedure uses Switch A as an example to verify the configuration.

Verify that Switch A has synchronized to Switch C, and the clock stratum level is 3 on Switch A and 2 on Switch C.

```
[SwitchA-Vlan-interface2] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3.0.1.31
 Local mode: bclient
 Reference clock ID: 3.0.1.31
 Leap indicator: 00
 Clock jitter: 0.044281 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00229 ms
 Root dispersion: 4.12572 ms
 Reference time: d0d289fe.ec43c720  Sat, Jan  8 2019  7:00:14.922
 System poll interval: 64 s
```

Verify that an IPv4 NTP association has been established between Switch A and Switch C.

```
[SwitchA-Vlan-interface2] display ntp-service sessions
       source          reference       stra reach poll  now offset  delay disper
**************************************************************************
 [1245]3.0.1.31        127.127.1.0        2     1   64  519   -0.0 0.0022 4.1257
Notes: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured.
 Total sessions: 1
```

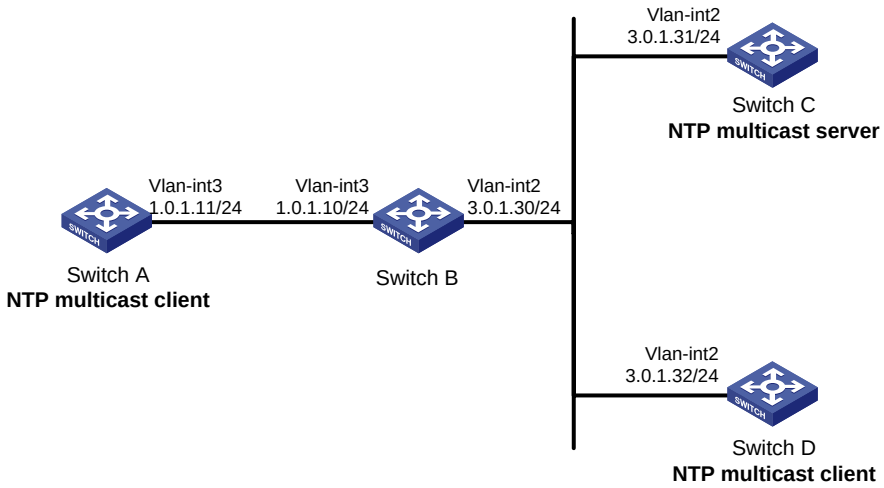
Example: Configuring NTP multicast association mode

Network configuration

As shown in Figure 13, configure Switch C as the NTP server for multiple devices on different network segments so that these devices synchronize the time with Switch C.

- Configure Switch C's local clock as its reference source, with stratum level 2.
- Configure Switch C to operate in multicast server mode and send multicast messages from VLAN-interface 2.
- Configure Switch A and Switch D to operate in multicast client mode and receive multicast messages on VLAN-interface 3 and VLAN-interface 2, respectively.

Figure 13 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure the switches can reach each other, as shown in Figure 13. (Details not shown.)
2. Configure Switch C:

   # Enable the NTP service.

   ```
   <SwitchC> system-view
   [SwitchC] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchC] clock protocol ntp
   ```

   # Specify the local clock as the reference source, with stratum level 2.

   ```
   [SwitchC] ntp-service refclock-master 2
   ```

   # Configure Switch C to operate in multicast server mode and send multicast messages from VLAN-interface 2.

   ```
   [SwitchC] interface vlan-interface 2
   [SwitchC-Vlan-interface2] ntp-service multicast-server
   ```

3. Configure Switch D:

   # Enable the NTP service.

   ```
   <SwitchD> system-view
   [SwitchD] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchD] clock protocol ntp
   ```

Configure Switch D to operate in multicast client mode and receive multicast messages on VLAN-interface 2.

```
[SwitchD] interface vlan-interface 2
[SwitchD-Vlan-interface2] ntp-service multicast-client
```

4. Verify the configuration:

Verify that Switch D has synchronized to Switch C, and the clock stratum level is 3 on Switch D and 2 on Switch C.

Switch D and Switch C are on the same subnet, so Switch D can receive the multicast messages from Switch C without being enabled with the multicast functions.

```
[SwitchD-Vlan-interface2] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3.0.1.31
 Local mode: bclient
 Reference clock ID: 3.0.1.31
 Leap indicator: 00
 Clock jitter: 0.044281 s
 Stability: 0.000 pps
 Clock precision: 2^-10
 Root delay: 0.00229 ms
 Root dispersion: 4.12572 ms
 Reference time: d0d289fe.ec43c720  Sat, Jan  8 2019  7:00:14.922
 System poll interval: 64 s
```

Verify that an IPv4 NTP association has been established between Switch D and Switch C.

```
[SwitchD-Vlan-interface2] display ntp-service sessions
      source          reference       stra reach poll  now offset  delay disper
**********************************************************************************
 [1245]3.0.1.31        127.127.1.0        2     1   64  519   -0.0 0.0022 4.1257
Notes: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured.
 Total sessions: 1
```

5. Configure Switch B:

Because Switch A and Switch C are on different subnets, you must enable the multicast functions on Switch B before Switch A can receive multicast messages from Switch C.

Enable IP multicast functions.

```
<SwitchB> system-view
[SwitchB] multicast routing
[SwitchB-mrib] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] pim dm
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/1
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] igmp enable
[SwitchB-Vlan-interface3] igmp static-group 224.0.1.1
[SwitchB-Vlan-interface3] quit
[SwitchB] igmp-snooping
```

```
[SwitchB-igmp-snooping] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] igmp-snooping static-group 224.0.1.1 vlan 3
```

6. Configure Switch A:

   # Enable the NTP service.

```
<SwitchA> system-view
[SwitchA] ntp-service enable
```

   # Specify NTP for obtaining the time.

```
[SwitchA] clock protocol ntp
```

   # Configure Switch A to operate in multicast client mode and receive multicast messages on VLAN-interface 3.

```
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ntp-service multicast-client
```

Verifying the configuration

Verify that Switch A has synchronized its time with Switch C, and the clock stratum level of Switch A is 3.

```
[SwitchA-Vlan-interface3] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3.0.1.31
 Local mode: bclient
 Reference clock ID: 3.0.1.31
 Leap indicator: 00
 Clock jitter: 0.165741 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00534 ms
 Root dispersion: 4.51282 ms
 Reference time: d0c61289.10b1193f  Wed, Dec 29 2019 20:03:21.065
 System poll interval: 64 s
```

Verify that an IPv4 NTP association has been established between Switch A and Switch C.

```
[SwitchA-Vlan-interface3] display ntp-service sessions
      source          reference       stra reach poll  now offset  delay disper
**************************************************************************
 [1234]3.0.1.31        127.127.1.0        2   247   64  381   -0.0 0.0053 4.5128
Notes: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured.
 Total sessions: 1
```

Example: Configuring IPv6 NTP multicast association mode

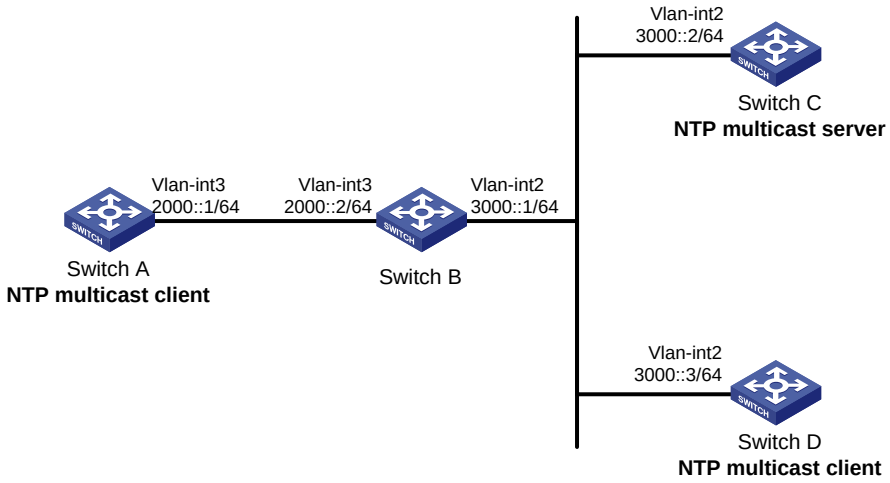
Network configuration

As shown in Figure 14, configure Switch C as the NTP server for multiple devices on different network segments so that these devices synchronize the time with Switch C.

- Configure Switch C's local clock as its reference source, with stratum level 2.
- Configure Switch C to operate in IPv6 multicast server mode and send IPv6 multicast messages from VLAN-interface 2.

- Configure Switch A and Switch D to operate in IPv6 multicast client mode and receive IPv6 multicast messages on VLAN-interface 3 and VLAN-interface 2, respectively.

Figure 14 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure the switches can reach each other, as shown in Figure 14. (Details not shown.)
2. Configure Switch C:

   # Enable the NTP service.

   ```
   <SwitchC> system-view
   [SwitchC] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchC] clock protocol ntp
   ```

   # Specify the local clock as the reference source, with stratum level 2.

   ```
   [SwitchC] ntp-service refclock-master 2
   ```

   # Configure Switch C to operate in IPv6 multicast server mode and send multicast messages from VLAN-interface 2.

   ```
   [SwitchC] interface vlan-interface 2
   [SwitchC-Vlan-interface2] ntp-service ipv6 multicast-server ff24::1
   ```

3. Configure Switch D:

   # Enable the NTP service.

   ```
   <SwitchD> system-view
   [SwitchD] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchD] clock protocol ntp
   ```

   # Configure Switch D to operate in IPv6 multicast client mode and receive multicast messages on VLAN-interface 2.

   ```
   [SwitchD] interface vlan-interface 2
   [SwitchD-Vlan-interface2] ntp-service ipv6 multicast-client ff24::1
   ```

4. Verify the configuration:

   # Verify that Switch D has synchronized its time with Switch C, and the clock stratum level of Switch D is 3.

   Switch D and Switch C are on the same subnet, so Switch D can Receive the IPv6 multicast messages from Switch C without being enabled with the IPv6 multicast functions.

```
[SwitchD-Vlan-interface2] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3000::2
 Local mode: bclient
 Reference clock ID: 165.84.121.65
 Leap indicator: 00
 Clock jitter: 0.000977 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00000 ms
 Root dispersion: 8.00578 ms
 Reference time: d0c60680.9754fb17  Wed, Dec 29 2019 19:12:00.591
 System poll interval: 64 s
```

Verify that an IPv6 NTP association has been established between Switch D and Switch C.

```
[SwitchD-Vlan-interface2] display ntp-service ipv6 sessions
Notes: 1 source(master), 2 source(peer), 3 selected, 4 candidate, 5 configured.

 Source:  [1234]3000::2
 Reference: 127.127.1.0          Clock stratum: 2
 Reachabilities: 111             Poll interval: 64
 Last receive time: 23           Offset: -0.0
 Roundtrip delay: 0.0            Dispersion: 0.0

 Total sessions: 1
```

5. Configure Switch B:

Because Switch A and Switch C are on different subnets, you must enable the IPv6 multicast functions on Switch B before Switch A can receive IPv6 multicast messages from Switch C.

Enable IPv6 multicast functions.

```
<SwitchB> system-view
[SwitchB] ipv6 multicast routing
[SwitchB-mrib6] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 pim dm
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/1
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] mld enable
[SwitchB-Vlan-interface3] mld static-group ff24::1
[SwitchB-Vlan-interface3] quit
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] mld-snooping static-group ff24::1 vlan 3
```

6. Configure Switch A:

Enable the NTP service.

```
<SwitchA> system-view
[SwitchA] ntp-service enable
```

Specify NTP for obtaining the time.

```
[SwitchA] clock protocol ntp
```

Configure Switch A to operate in IPv6 multicast client mode and receive IPv6 multicast messages on VLAN-interface 3.

```
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ntp-service ipv6 multicast-client ff24::1
```

Verifying the configuration

Verify that Switch A has synchronized to Switch C, and the clock stratum level is 3 on Switch A and 2 on Switch C.

```
[SwitchA-Vlan-interface3] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 System peer: 3000::2
 Local mode: bclient
 Reference clock ID: 165.84.121.65
 Leap indicator: 00
 Clock jitter: 0.165741 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00534 ms
 Root dispersion: 4.51282 ms
 Reference time: d0c61289.10b1193f  Wed, Dec 29 2019 20:03:21.065
 System poll interval: 64 s
```

Verify that an IPv6 NTP association has been established between Switch A and Switch C.

```
[SwitchA-Vlan-interface3] display ntp-service ipv6 sessions
Notes: 1 source(master), 2 source(peer), 3 selected, 4 candidate, 5 configured.

 Source:   [124]3000::2
 Reference: 127.127.1.0          Clock stratum: 2
 Reachabilities: 2               Poll interval: 64
 Last receive time: 71           Offset: -0.0
 Roundtrip delay: 0.0            Dispersion: 0.0

 Total sessions: 1
```

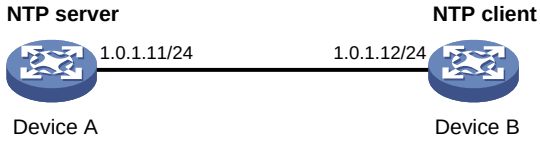
Example: Configuring NTP authentication in client/server association mode

Network configuration

As shown in Figure 15, perform the following tasks:

- Configure Device A's local clock as its reference source, with stratum level 2.
- Configure Device B to operate in client mode and specify Device A as the NTP server of Device B.
- Configure NTP authentication on both Device A and Device B.

Figure 15 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Device A and Device B can reach each other, as shown in Figure 15. (Details not shown.)
2. Configure Device A:

   # Enable the NTP service.

   ```
   <DeviceA> system-view
   [DeviceA] ntp-service enable
   ```

   # Specify the local clock as the reference source, with stratum level 2.

   ```
   [DeviceA] ntp-service refclock-master 2
   ```
3. Configure Device B:

   # Enable the NTP service.

   ```
   <DeviceB> system-view
   [DeviceB] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [DeviceB] clock protocol ntp
   ```

   # Enable NTP authentication on Device B.

   ```
   [DeviceB] ntp-service authentication enable
   ```

   # Create a plaintext authentication key, with key ID **42** and key value **aNiceKey**.

   ```
   [DeviceB] ntp-service authentication-keyid 42 authentication-mode md5 simple
   aNiceKey
   ```

   # Specify the key as a trusted key.

   ```
   [DeviceB] ntp-service reliable authentication-keyid 42
   ```

   # Specify Device A as the NTP server of Device B, and associate the server with key 42.

   ```
   [DeviceB] ntp-service unicast-server 1.0.1.11 authentication-keyid 42
   ```

   To enable Device B to synchronize its clock with Device A, enable NTP authentication on Device A.
4. Configure NTP authentication on Device A:

   # Enable NTP authentication.

   ```
   [DeviceA] ntp-service authentication enable
   ```

   # Create a plaintext authentication key, with key ID **42** and key value **aNiceKey**.

   ```
   [DeviceA] ntp-service authentication-keyid 42 authentication-mode md5 simple
   aNiceKey
   ```

   # Specify the key as a trusted key.

   ```
   [DeviceA] ntp-service reliable authentication-keyid 42
   ```

Verifying the configuration

Verify that Device B has synchronized its time with Device A, and the clock stratum level of Device B is 3.

```
[DeviceB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
```

```
 System peer: 1.0.1.11
 Local mode: client
 Reference clock ID: 1.0.1.11
 Leap indicator: 00
 Clock jitter: 0.005096 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00655 ms
 Root dispersion: 1.15869 ms
 Reference time: d0c62687.ab1bba7d  Wed, Dec 29 2019 21:28:39.668
 System poll interval: 64 s
```

Verify that an IPv4 NTP association has been established between Device B and Device A.

```
[DeviceB] display ntp-service sessions
       source          reference       stra reach poll  now offset  delay disper
********************************************************************************
 [1245]1.0.1.11        127.127.1.0        2     1   64  519   -0.0 0.0065    0.0
Notes: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured.
 Total sessions: 1
```

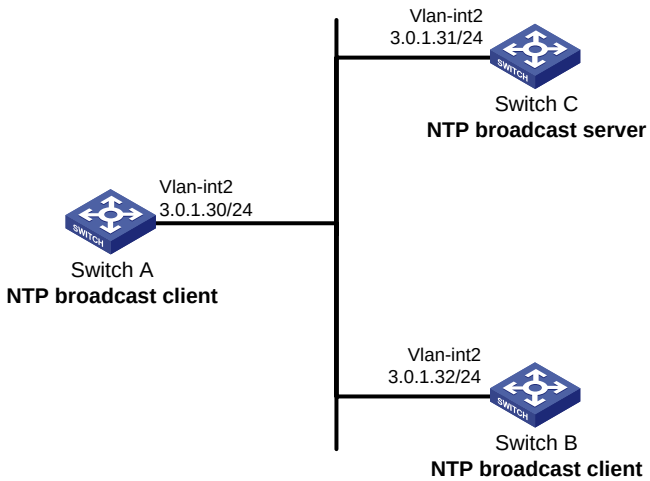
Example: Configuring NTP authentication in broadcast association mode

Network configuration

As shown in Figure 16, configure Switch C as the NTP server for multiple devices on the same network segment so that these devices synchronize the time with Switch C. Configure Switch A and Switch B to authenticate the NTP server.

- Configure Switch C's local clock as its reference source, with stratum level 3.
- Configure Switch C to operate in broadcast server mode and send broadcast messages from VLAN-interface 2.
- Configure Switch A and Switch B to operate in broadcast client mode and receive broadcast messages on VLAN-interface 2.
- Enable NTP authentication on Switch A, Switch B, and Switch C.

Figure 16 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Switch A, Switch B, and Switch C can reach each other, as shown in Figure 16. (Details not shown.)
2. Configure Switch A:

   # Enable the NTP service.

   ```
   <SwitchA> system-view
   [SwitchA] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchA] clock protocol ntp
   ```

   # Enable NTP authentication on Switch A. Create a plaintext NTP authentication key, with key ID of **88** and key value of **123456**. Specify it as a trusted key.

   ```
   [SwitchA] ntp-service authentication enable
   [SwitchA] ntp-service authentication-keyid 88 authentication-mode md5 simple 123456
   [SwitchA] ntp-service reliable authentication-keyid 88
   ```

   # Configure Switch A to operate in NTP broadcast client mode and receive NTP broadcast messages on VLAN-interface 2.

   ```
   [SwitchA] interface vlan-interface 2
   [SwitchA-Vlan-interface2] ntp-service broadcast-client
   ```

3. Configure Switch B:

   # Enable the NTP service.

   ```
   <SwitchB> system-view
   [SwitchB] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchB] clock protocol ntp
   ```

   # Enable NTP authentication on Switch B. Create a plaintext NTP authentication key, with key ID of **88** and key value of **123456**. Specify it as a trusted key.

   ```
   [SwitchB] ntp-service authentication enable
   [SwitchB] ntp-service authentication-keyid 88 authentication-mode md5 simple 123456
   [SwitchB] ntp-service reliable authentication-keyid 88
   ```

   # Configure Switch B to operate in broadcast client mode and receive NTP broadcast messages on VLAN-interface 2.

   ```
   [SwitchB] interface vlan-interface 2
   [SwitchB-Vlan-interface2] ntp-service broadcast-client
   ```

4. Configure Switch C:

   # Enable the NTP service.

   ```
   <SwitchC> system-view
   [SwitchC] ntp-service enable
   ```

   # Specify NTP for obtaining the time.

   ```
   [SwitchC] clock protocol ntp
   ```

   # Specify the local clock as the reference source, with stratum level 3.

   ```
   [SwitchC] ntp-service refclock-master 3
   ```

   # Configure Switch C to operate in NTP broadcast server mode and use VLAN-interface 2 to send NTP broadcast packets.

   ```
   [SwitchC] interface vlan-interface 2
   [SwitchC-Vlan-interface2] ntp-service broadcast-server
   [SwitchC-Vlan-interface2] quit
   ```

5. Verify the configuration:

NTP authentication is enabled on Switch A and Switch B, but not on Switch C, so Switch A and Switch B cannot synchronize their local clocks to Switch C.

```
[SwitchB-Vlan-interface2] display ntp-service status
 Clock status: unsynchronized
 Clock stratum: 16
 Reference clock ID: none
```

6. Enable NTP authentication on Switch C:

Enable NTP authentication on Switch C. Create a plaintext NTP authentication key, with key ID of **88** and key value of **123456**. Specify it as a trusted key.

```
[SwitchC] ntp-service authentication enable
[SwitchC] ntp-service authentication-keyid 88 authentication-mode md5 simple 123456
[SwitchC] ntp-service reliable authentication-keyid 88
```

Specify Switch C as an NTP broadcast server, and associate key **88** with Switch C.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server authentication-keyid 88
```

Verifying the configuration

Verify that Switch B has synchronized its time with Switch C, and the clock stratum level of Switch B is 4.

```
[SwitchB-Vlan-interface2] display ntp-service status
 Clock status: synchronized
 Clock stratum: 4
 System peer: 3.0.1.31
 Local mode: bclient
 Reference clock ID: 3.0.1.31
 Leap indicator: 00
 Clock jitter: 0.006683 s
 Stability: 0.000 pps
 Clock precision: 2^-23
 Root delay: 0.00127 ms
 Root dispersion: 2.89877 ms
 Reference time: d0d287a7.3119666f  Sat, Jan  8 2019  6:50:15.191
 System poll interval: 64 s
```

Verify that an IPv4 NTP association has been established between Switch B and Switch C.

```
[SwitchB-Vlan-interface2] display ntp-service sessions
      source          reference       stra reach poll  now offset  delay disper
**************************************************************************
 [1245]3.0.1.31        127.127.1.0        3     3   64   68   -0.0 0.0000    0.0
Notes: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured.
 Total sessions: 1
```

Configuring SNTP

About SNTP

SNTP is a simplified, client-only version of NTP specified in RFC 4330. It uses the same packet format and packet exchange procedure as NTP, but provides faster synchronization at the price of time accuracy.

Relationship between SNTP and NTP

The relationship between SNTP and NTP is as follows:

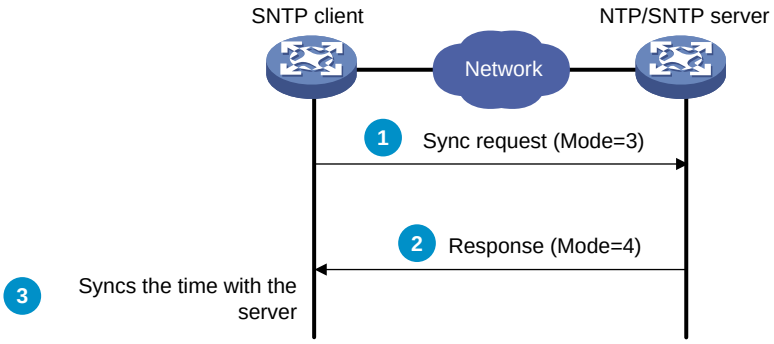
- **Purpose**—Both SNTP and NTP aim to synchronize the time of computer systems.
- **Protocol family**—SNTP is a subset of NTP and uses the same protocol family as NTP. SNTP uses the NTP message format.
- **Simplified version**—NTP provides additional features to enhance time synchronization accuracy, such as complex error detection and correction algorithms, dynamic server selection, load balancing, and control protocols. SNTP typically does not provide these complex features.
- **Precision**—NTP is designed to provide high-precision time synchronization on the network, typically at the millisecond level and even capable of reaching sub-millisecond levels. SNTP is suitable for scenarios where precision requirements are lower.
- **Compatibility**—In most cases, SNTP and NTP are compatible, because they use the same PDU format. This design allows interoperability between SNTP clients and NTP servers, and between NTP clients and SNTP servers.
- **Application scenarios**—SNTP, being a simplified version of NTP, is ideal for resource-constrained devices such as home routers and IP cameras, as well as applications where strict time accuracy is not critical. On the other hand, NTP is designed for use in scenarios demanding highly precise time synchronization, such as financial transaction systems, scientific research, and advanced network infrastructure.

SNTP operating mechanism

SNTP performs basic time synchronization by using some fields in NTP messages. It uses the same message exchange process as NTP but processes the received time information in a more simplified way, reducing complexity and resource requirements. The following is a typical SNTP time synchronization message exchange process:

1. The SNTP client generates an NTP message, sets the mode to client (mode 3), pads the original timestamp, and then sends it to the server.
2. Upon receiving the message, the NTP/SNTP server adds or updates the receive timestamp and transmit timestamp, sets the message mode to server (mode 4), and sends it back to the client as a response.
3. After receiving the response, the SNTP client logs the time when the response was received, extracts the transmission timestamp from the response, and adjusts the clock as needed.

Figure 17 SNTP operating mechanism



SNTP association modes

SNTP supports the following association modes:

- Client/server mode
- Broadcast mode

You can select one or more SNTP association modes for time synchronization. Table 7 provides detailed description for the association modes.

In this document, an "NTP server" or a "server" refers to a device that operates as an NTP server in client/server mode. Time servers refer to all the devices that can provide time synchronization, including NTP servers and NTP broadcast servers.

Table 7 SNTP association modes

Mode	Working process	Synchronization direction	Application scenario
Client/server	**1.** A client sends an NTP time synchronization request message to the NTP servers. **2.** After receiving the message, each NTP server automatically works in server mode and replies with a response message. **3.** The client filters and selects an optimal clock based on the received response messages and synchronizes time with that clock.	From the server to the client	As the NTP architecture diagram in "NTP architecture" shows, this mode is intended for scenarios where devices with a higher stratum value obtain time from a device with a lower stratum value.
Broadcast	**4.** A broadcast server periodically sends clock synchronization request messages to the broadcast address 255.255.255.255. **5.** Broadcast clients listen to the broadcast messages from the server and synchronize the time with the server based on the broadcast messages. When a broadcast client receives	From the broadcast server to the client.	As the NTP architecture diagram in "NTP architecture" shows, this mode is intended for scenarios where the same time server provides time synchronization for a large number of devices in the same subnet. The broadcast mode has lower time accuracy than the client/server and

Mode	Working process	Synchronization direction	Application scenario
	the first broadcast message, the client and the server start to exchange messages to calculate the network delay between them. Then, only the broadcast server sends clock synchronization messages.		symmetric active/passive modes because only the broadcast servers send clock synchronization messages.

Protocols and standards

RFC 4330, *Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI*

FIPS compliance

Higher security is required in FIPS mode. When the device operates in FIPS mode, the following restrictions apply for SNTP:

- SNTP can only use NTPv3 and NTPv4 packets and cannot use NTPv1 or NTPv2 packets.
- SNTP supports the HMAC-SHA-1, HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 authentication algorithms and does not support the MD5 authentication algorithm.

For more information about FIPS mode, see *Security Configuration Guide*.

Restrictions and guidelines: SNTP configuration

When you configure SNTP, follow these restrictions and guidelines:

- You cannot configure both NTP and SNTP on the same device.
- You must specify a Multitenant Device Context (MDC) or context in the **`clock protocol`** command for using NTP to obtain the time. For more information about the **`clock protocol`** command, see device management commands in *Fundamentals Configuration Guide*.
- You can configure SNTP only on one MDC or context.

Client/server mode tasks at a glance

Tasks on the client

1. Enabling the SNTP service
2. Configuring SNTP authentication in client/server mode

   The device can operate in client/server mode and broadcast mode simultaneously. If the device acts as a client and receives multiple NTP clock signals at the same time, it will select the optimal NTP clock for synchronization.

   - Specifying an NTP server for the device
   - (Optional.) Configuring SNTP authentication in client/server mode
3. (Optional.) Controlling SNTP time synchronization
   - Ignoring clock stratum level restrictions on SNTP clients
   - Setting the maximum offset allowed for time synchronization through SNTP
4. (Optional.) Monitoring the time synchronization status
5. (Optional.) Specifying the SNTP time-offset thresholds for log and trap outputs

Tasks on the server

1. Enabling the SNTP service
2. (Optional.) Configuring SNTP authentication in client/server mode

Broadcast mode tasks at a glance

Tasks on the broadcast client

1. Enabling the SNTP service
2. Configuring SNTP in broadcast mode

   The device can operate in client/server mode and broadcast mode simultaneously. If the device acts as a client and receives multiple NTP clock signals at the same time, it will select the optimal NTP clock for synchronization.

   - Configuring the SNTP broadcast client
   - (Optional.) Configuring SNTP authentication in broadcast mode
3. (Optional.) Controlling SNTP time synchronization
   - Ignoring clock stratum level restrictions on SNTP clients
   - Setting the maximum offset allowed for time synchronization through SNTP
4. (Optional.) Monitoring the time synchronization status
5. (Optional.) Specifying the SNTP time-offset thresholds for log and trap outputs

NOTE:

You can deploy an NTP broadcast server in the network to act as an SNTP broadcast server. This enables SNTP time synchronization in broadcast mode.

Enabling the SNTP service

Restrictions and guidelines

The NTP service and SNTP service are mutually exclusive. Before you enable SNTP, make sure NTP is disabled.

Procedure

1. Enter system view.

   `system-view`
2. Enable the SNTP service.

   `sntp enable`

   By default, the SNTP service is disabled.

Configuring SNTP in client/server mode

Specifying an NTP server for the device

About this task

In SNTP client/server mode, the device only supports acting as a client to synchronize the clock with the NTP server.

Restrictions and guidelines

To use an NTP server as the time source, make sure its clock has been synchronized. If the stratum level of the NTP server is greater than or equal to that of the client, the client does not synchronize with the NTP server.

Procedure

1. Enter system view.

   system-view

2. Specify an NTP server for the device.

   IPv4:

   sntp unicast-server { *server-name* | *ip-address* } [**vpn-instance** *vpn-instance-name*] [**authentication-keyid** *keyid* | **maxpoll** *maxpoll-interval* | **minpoll** *minpoll-interval* | **source** *interface-type interface-number* | **version** *number*] *

   IPv6:

   sntp ipv6 unicast-server { *server-name* | *ipv6-address* } [**vpn-instance** *vpn-instance-name*] [**authentication-keyid** *keyid* | **maxpoll** *maxpoll-interval* | **minpoll** *minpoll-interval* | **source** *interface-type interface-number*] *

   By default, no NTP server is specified for the device.

   You can specify multiple NTP servers for the client by repeating this step.

   To perform authentication, you need to specify the **authentication-keyid** *keyid* option.

Configuring SNTP authentication in client/server mode

About this task

SNTP authentication ensures that an SNTP client is synchronized only to an authenticated trustworthy NTP server.

Restrictions and guidelines

Enable authentication on both the NTP server and the SNTP client.

Use the same authentication key ID, algorithm, and key on the NTP server and SNTP client. Specify the key as a trusted key on both the NTP server and the SNTP client. For information about configuring NTP authentication on an NTP server, see "Configuring NTP."

On the SNTP client, associate the specified key with the NTP server. Make sure the server is allowed to use the key ID for authentication on the client.

With authentication disabled, the SNTP client can synchronize with the NTP server regardless of whether the NTP server is enabled with authentication.

Configuring SNTP authentication on the client

1. Enter system view.

   system-view

2. Enable SNTP authentication.

   sntp authentication enable

   By default, SNTP authentication is disabled.

3. Configure an SNTP authentication key.

sntp authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

By default, no SNTP authentication key exists.

4. Specify the key as a trusted key.

sntp reliable authentication-keyid *keyid*

By default, no trusted key is specified.

5. Associate the SNTP authentication key with an NTP server.

IPv4:

sntp unicast-server { *server-name* | *ip-address* } [**vpn-instance** *vpn-instance-name*] **authentication-keyid** *keyid*

IPv6:

sntp ipv6 unicast-server { *server-name* | *ipv6-address* } [**vpn-instance** *vpn-instance-name*] **authentication-keyid** *keyid*

By default, no NTP server is specified.

Configuring SNTP authentication on the server

1. Enter system view.

system-view

2. Enable SNTP authentication.

sntp authentication enable

By default, SNTP authentication is disabled.

3. Configure an SNTP authentication key.

sntp authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

By default, no SNTP authentication key exists.

4. Specify the key as a trusted key.

sntp reliable authentication-keyid *keyid*

By default, no trusted key is specified.

Configuring SNTP in broadcast mode

Configuring the SNTP broadcast client

About this task

In SNTP broadcast mode, the device supports only operating as an SNTP broadcast client to synchronize the clock with the NTP broadcast server.

Procedure

1. Enter system view.

system-view

2. Enter interface view.

interface *interface-type interface-number*

3. Configure the device to operate in SNTP broadcast client mode.

sntp broadcast-client

By default, the device does not operate in any SNTP association mode.

After you execute the command, the device receives NTP broadcast messages from the specified interface.

Configuring SNTP authentication in broadcast mode

About this task

In some networks where time synchronization security is required, you need to enable authentication when SNTP is used. SNTP authentication ensures that the client synchronizes time only with the authenticated server, which improves network security.

Restrictions and guidelines

- You must enable authentication on both the NTP broadcast server and SNTP broadcast client.
- You must configure the same authentication key (including key ID, authentication algorithm, and key value) for the NTP broadcast server and SNTP broadcast client and set the key as a trusted key. For information about configuring authentication on an NTP broadcast server, see "Configuring NTP authentication for a broadcast server."
- If SNTP authentication is not successfully enabled on the client, the client can synchronize with the server regardless of whether authentication is enabled on the server.

Procedure

1. Enter system view.

   system-view
2. Enable SNTP authentication.

   sntp authentication enable

   By default, SNTP authentication is disabled.
3. Configure an SNTP authentication key.

   sntp authentication-keyid *keyid* **authentication-mode** { **hmac-sha-1** | **hmac-sha-256** | **hmac-sha-384** | **hmac-sha-512** | **md5** } { **cipher** | **simple** } *string* [**acl** *ipv4-acl-number* | **ipv6 acl** *ipv6-acl-number*] *

   By default, no SNTP authentication key exists.
4. Configure the key as a trusted key.

   sntp reliable authentication-keyid *keyid*

   By default, no authentication key is configured as a trusted key.

Controlling SNTP time synchronization

Ignoring clock stratum level restrictions on SNTP clients

About this task

Application scenarios

Perform this task to control whether the device responds to time synchronization requests from SNTP clients when it acts as an SNTP server.

Operating mechanism

By default, when an SNTP server receives a clock synchronization request from an SNTP client, it checks the stratum level of the client. The server responds to the request and synchronizes the clock only if the client has a higher stratum level than the server. In some network environments, if a client

has a stratum level lower than or equal to the server and changing the stratum level is not supported, you can perform this task to ignore the stratum level restrictions on SNTP clients.

Procedure

1. Enter system view.

   system-view

2. Ignore restrictions on the stratum level of SNTP clients.

   sntp ignore clock-stratum

   By default, an SNTP server restricts the stratum level of SNTP clients when it synchronizes the clock with the SNTP clients.

Setting the maximum offset allowed for time synchronization through SNTP

About this task

Application scenarios

This feature controls whether the device synchronizes the time with an SNTP server when it acts as an SNTP client.

Operating mechanism

When the device acts as an SNTP client, it subtracts the local time from the time provided by the SNTP server to calculate the time offset. The device will adjust the local SNTP time based on this offset. If this offset exceeds the configured maximum offset, the device will determine that the time from the SNTP server is unreliable and will not synchronize the time from the SNTP server.

Procedure

1. Enter system view.

   system-view

2. Set the maximum offset allowed for time synchronization through SNTP.

   sntp time-offset-limit *limit-threshold*

   By default, no maximum offset is set for time synchronization through SNTP.

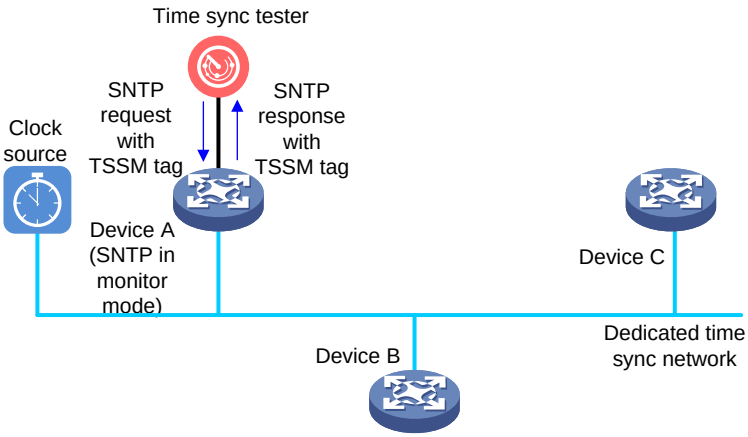
Monitoring the time synchronization status

About this task

Application scenarios

In networks where time is synchronized through non-SNTP methods (such as dedicated time synchronization networks and PTP), you can use this feature to monitor the time synchronization status. It can identify whether the difference between the time obtained by the device and the clock source is within the allowable range and whether the device's time is affected by any sudden changes in the clock source.

Figure 18 SNTP operating in monitor mode



Operating mechanism

The Time Synchronization Status Monitoring (TSSM) tag in SNTP packets is used to enhance the security and reliability of the time synchronization process. The TSSM tag is typically used in time synchronization applications with higher security levels to ensure that responses are made only to SNTP requests from known and trusted time sources.

When the device acts as an SNTP client, the packets it send do not carry the TSSM tag, and it can synchronize time with the SNTP server.

When SNTP is enabled, the SNTP server functionality is automatically enabled on the device. When the device operates as an SNTP server:

- By default, when the device receives an SNTP request, it processes it as a clock synchronization request regardless of whether it carries the TSSM flag, and responds with a packet that does not contain the TSSM flag.
- If this command is configured, the SNTP server will identify whether the received SNTP request contains the TSSM flag. If the request does not contain the TSSM flag, the device will discard the request. The device responds to the request only if the request contains the TSSM flag.

Restrictions and guidelines

After this feature is configured, the device can synchronize its clock as an SNTP client with upstream devices, but it cannot provide SNTP time synchronization services to downstream devices.

Procedure

1. Enter system view.

   `system-view`

2. Configure the device to respond only to SNTP requests that contain the TSSM tag.

   `sntp respond-to-withtssm`

   By default, the device processes all SNTP requests regardless of whether they carry the TSSM tag when it acts as an SNTP server.

Specifying the SNTP time-offset thresholds for log and trap outputs

About this task

By default, the system synchronizes the SNTP client's time to the server and outputs a log and a trap when the time offset exceeds 128 ms for multiple times.

After you set the SNTP time-offset thresholds for log and trap outputs, the system synchronizes the client's time to the server when the time offset exceeds 128 ms for multiple times, but outputs logs and traps only when the time offset exceeds the specified thresholds, respectively.

Procedure

1. Enter system view.

   system-view

2. Specify the SNTP time-offset thresholds for log and trap outputs.

   sntp time-offset-threshold { **log** *log-threshold* | **trap** *trap-threshold* } *

   By default, no SNTP time-offset thresholds are set for log and trap outputs.

Display and maintenance commands for SNTP

Execute **display** commands in any view.

Task	Command
Display information about all IPv6 SNTP associations.	**display sntp ipv6 sessions**
Display information about all IPv4 SNTP associations.	**display sntp sessions**

SNTP configuration examples

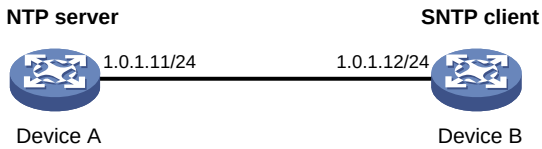
Example: Configuring SNTP in client/server association mode

Network configuration

As shown in Figure 19, perform the following tasks:

- Configure Device A's local clock as its reference source, with stratum level 2.
- Configure Device B to operate in SNTP client mode, and specify Device A as the NTP server.
- Configure NTP authentication on Device A and SNTP authentication on Device B.

Figure 19 Network diagram



Procedure

1. Assign an IP address to each interface, and make sure Device A and Device B can reach each other, as shown in Figure 19. (Details not shown.)
2. Configure Device A:

   # Enable the NTP service.

   ```
   <DeviceA> system-view
   [DeviceA] ntp-service enable
   ```

Specify NTP for obtaining the time.

```
[DeviceA] clock protocol ntp
```

Configure the local clock as the reference source, with stratum level 2.

```
[DeviceA] ntp-service refclock-master 2
```

Enable NTP authentication on Device A.

```
[DeviceA] ntp-service authentication enable
```

Configure a plaintext NTP authentication key, with key ID of **10** and key value of **aNiceKey**.

```
[DeviceA] ntp-service authentication-keyid 10 authentication-mode md5 simple
aNiceKey
```

Specify the key as a trusted key.

```
[DeviceA] ntp-service reliable authentication-keyid 10
```

3. Configure Device B:

Enable the SNTP service.

```
<DeviceB> system-view
[DeviceB] sntp enable
```

Specify NTP for obtaining the time.

```
[DeviceB] clock protocol ntp
```

Enable SNTP authentication on Device B.

```
[DeviceB] sntp authentication enable
```

Configure a plaintext authentication key, with key ID of **10** and key value of **aNiceKey**.

```
[DeviceB] sntp authentication-keyid 10 authentication-mode md5 simple aNiceKey
```

Specify the key as a trusted key.

```
[DeviceB] sntp reliable authentication-keyid 10
```

Specify Device A as the NTP server of Device B, and associate the server with key 10.

```
[DeviceB] sntp unicast-server 1.0.1.11 authentication-keyid 10
```

Verifying the configuration

Verify that an SNTP association has been established between Device B and Device A, and Device B has synchronized its time with Device A.

```
[DeviceB] display sntp sessions
SNTP server     Stratum   Version    Poll   Last receive time
1.0.1.11        2         4          64     Tue, May 17 2019  9:11:20.833 (Synced)
```

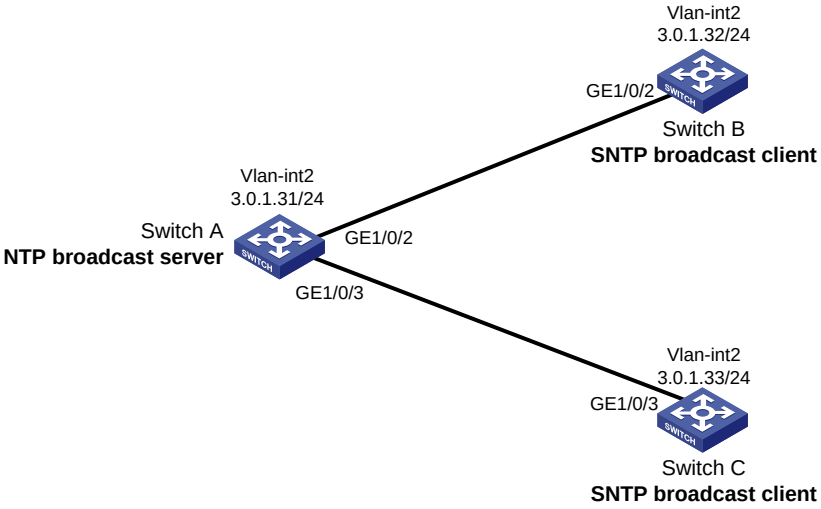
Example: Configuring SNTP in broadcast association mode

Network configuration

As shown in Figure 20, configure Switch A as the NTP broadcast server to provide NTP clock synchronization for other devices on the LAN.

- On Switch A, configure the local clock as the reference source, with stratum level 2.
- Configure Switch A to operate in broadcast server mode and send broadcast messages from VLAN-interface 2.
- Configure Switch B and Switch C to operate in broadcast client mode, and listen to broadcast messages on VLAN-interface 2.

Figure 20 Network diagram



Procedure

1. Configure Switch A:

Configure the device to use the local clock.

```
<SwitchA> system-view
[SwitchA] clock protocol none
```

Adjust the local system time to be correct.

```
[SwitchA] quit
<SwitchA> clock datetime 8:8:8 11/17/2022
```

Enable the NTP service.

```
<SwitchA> system-view
[SwitchA] ntp-service enable
```

Specify the local clock as the reference source, with stratum level 2.

```
[SwitchA] ntp-service refclock-master 2
```

Create VLAN 2 and assign GigabitEthernet1/0/2 that connects Switch B and GigabitEthernet1/0/3 that connects Switch C to VLAN 2.

```
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/2 gigabitethernet 1/0/3
[SwitchA-vlan2] quit
```

Configure Switch A to operate in broadcast server mode and send broadcast messages from VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 3.0.1.31 24
[SwitchA-Vlan-interface2] ntp-service broadcast-server
[SwitchA-Vlan-interface2] quit
```

2. Configure Switch B:

Enable SNTP

```
<SwitchB> system-view
[SwitchB] sntp enable
```

Specify using NTP to obtain the time.

```
[SwitchB] clock protocol ntp
```

Create VLAN 2 and assign GigabitEthernet1/0/2 that connects Switch A to VLAN 2.

```
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/2
[SwitchA-vlan2] quit
```

Configure Switch B to operate in SNTP broadcast client mode and receive broadcast messages on VLAN-interface 2.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 3.0.1.32 24
[SwitchB-Vlan-interface2] sntp broadcast-client
[SwitchB-Vlan-interface2] quit
```

3. Configure Switch C:

Enable SNTP.

```
<SwitchC> system-view
[SwitchC] sntp enable
```

Specify using NTP to obtain the time.

```
[SwitchC] clock protocol ntp
```

Create VLAN 2 and assign GigabitEthernet1/0/3 that connects Switch A to VLAN 2.

```
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/3
[SwitchA-vlan2] quit
```

Configure Switch C to operate in SNTP broadcast client mode and receive broadcast messages on VLAN-interface 2.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ip address 3.0.1.33 24
[SwitchC-Vlan-interface2] sntp broadcast-client
[SwitchC-Vlan-interface2] quit
```

Verifying the configuration

View SNTP session information on Switch B and Switch C. If information as follows is displayed, the device has established an session and synchronized the time with Switch A.

Switch B is used in this example.

```
[SwitchB] display sntp sessions
SNTP server     Stratum   Version    Poll   Last receive time
3.0.1.31        2         4          64     Thu, Nov  17 2022 08:08:40.442 (Synced)
```

View the system clock on Switch B and Switch C and verify that the time on Switch B and Switch C is consistent with that on Switch A.

Switch B is used in this example.

```
[SwitchB] display clock
08:08:45.120 UTC Thu 11/17/2022
```

Contents

Configuring PTP 1

About PTP 1
Basic concepts 1
Protocols and standards 4
PTP time synchronization mechanisms 5
Time synchronization basics 5
Grandmaster clock selection and master-member/subordinate relationship establishment 5
Optimal domain selection 7
Frequency synchronization mechanism 8
Phase synchronization mechanism 9
PTP support for IPv6 networks 11
Association between PTP and Track 11
Restrictions and guidelines: PTP configuration 11
PTP tasks at a glance 12
Configuring PTP (IEEE 1588 version 2) 12
Configuring PTP (SMPTE ST 2059-2) 13
Configuring PTP (AES67-2015) 13
Specifying PTP for obtaining the time 14
Creating a PTP instance 15
Specifying a PTP profile 15
Configuring clock nodes 16
Specifying a clock node type 16
Configuring an OC to operate only as a member clock 16
Specifying a PTP domain 17
Configuring the clock source whitelist 17
Enabling PTP globally 18
Enabling PTP on a port 18
Configuring PTP ports 19
Configuring the role of a PTP port 19
Configuring the mode for carrying timestamps 19
Specifying a delay measurement mechanism for a BC or an OC 20
Configuring one of the ports on a TC+OC clock as an OC-type port 21
Enabling unicast negotiation on PTP ports 21
Configuring PTP message transmission and receipt 22
Setting the interval for sending announce messages and the timeout multiplier for receiving announce messages 22
Setting the interval for sending Pdelay_Req messages 23
Setting the interval for sending Sync messages 23
Setting the minimum interval for sending Delay_Req messages 24
Setting the interval for sending Delay-Resp messages 24
Specifying the UDP transport protocol for PTP messages 25
Configuring a source IP address for multicast PTP messages transmitted over UDP 25
Configuring a destination IP address for unicast PTP messages transmitted over UDP 26
Configuring the MAC address for non-Pdelay messages 26
Setting a DSCP value for PTP messages transmitted over UDP 27
Specifying a VLAN tag for PTP messages 27
Disabling PTP path tracing 28
Adjusting and correcting clock synchronization 28
Setting the delay correction value 28
Calculating the TAI based on the UTC time 29
Configuring the PTP offset threshold and PTP synchronization suppressions 30
Configuring a priority for a clock 32
Configuring PTP logging 32
Enabling SNMP notification for the PTP module 33
About this task 33
Procedure 33
Display and maintenance commands for PTP 33

PTP configuration examples ····34
Example: Configuring PTP (IEEE 1588 version 2, IEEE 802.3/Ethernet transport, multicast transmission) ····34
Example: Configuring PTP (IEEE 1588 version 2, UDP transport, multicast transmission) ····37
Example: Configuring multiple PTP instances (IEEE 1588 version 2, multicast transmission)····41
Example: Configuring PTP (SMPTE ST 2059-2, UDP transport, multicast transmission) ····48
Example: Configuring PTP (AES67-2015, UDP transport, multicast transmission)····51

Configuring PTP

About PTP

Precision Time Protocol (PTP) provides time synchronization among devices with submicrosecond accuracy. It provides also precise frequency synchronization.

Basic concepts

PTP profile

IEEE 1588 is the basic PTP protocol. It defines the high-precision clock synchronization principle and the message interaction and processing specifications on the network. It was initially used for industrial automation and is now mainly used on bridged local area networks. 1588 has two versions: 1588v1 and 1588v2. 1588v1 can provide time synchronization only with sub-millisecond accuracy, and 1588v2 can provide time synchronization accuracy to the sub-microsecond level and can achieve both phase and frequency synchronization. 1588v2 is more widely used than 1588v1.

Based on 1588, PTP protocols including IEEE 802.1AS, ITU-T G.8275.1, and ITU-T G.8275.2 were developed to adapt to different application scenarios. This document focuses on the implementation mechanism and application scenarios of 1588v2.

- **IEEE 1588 version 2**—1588v2 defines high-accuracy clock synchronization mechanisms. It can be customized, enhanced, or tailored as needed. 1588v2 is the latest version.
- 802.1AS supports only point-to-point full-duplex Ethernet, IEEE 802.11, and IEEE 802.3 EPON links.
- **SMPTE ST 2059-2**—ST2059-2 is introduced based on IEEE 1588. It specifies a profile specifically for time synchronization of audio or video equipment in a professional broadcast environment. It includes a self-contained description of parameters, their default values, and permitted ranges.
- **AES67-2015**—AES67-2015 is introduced based on IEEE 1588. It specifies a profile specifically for time synchronization of professional equipment for broadcast, music production, and film and television post-production. It includes a self-contained description of parameters, their default values, and permitted ranges.

PTP domain

A PTP domain refers to a network or part of a network that is enabled with PTP. A PTP domain has only one reference clock called "grandmaster clock (GM)." All devices in the domain synchronize to the clock.

PTP instance

If the device belongs to multiple PTP domains, you must configure and associate a PTP instance with each PTP domain. You can configure PTP settings such as PTP profile and clock node type on a PTP instance. The settings take effect only on the PTP domain associated with the PTP instance. PTP instances are isolated from each other, allowing different PTP timing systems to run on a network without affecting each other.

Optimal PTP domain

Each PTP instance has a reference clock and clock information. For a device that has multiple PTP instances running simultaneously, you must select an optimal PTP instance and use the clock source traced by this PTP instance to synchronize the system time of the device. The domain with which the optimal PTP associates is the optimal domain.

Clock node and PTP port

A node in a PTP domain is a clock node. A port enabled with PTP is a PTP port. PTP defines the following types of basic clock nodes:

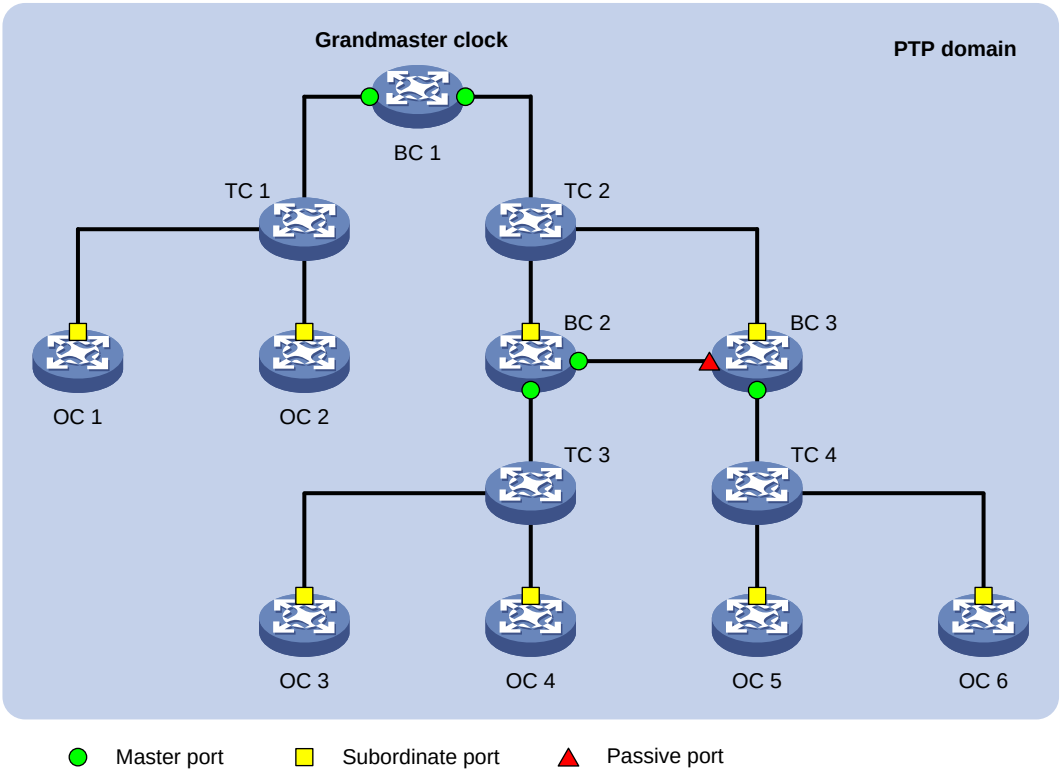
- **Ordinary Clock (OC)**—A PTP clock with a single PTP port in a PTP domain for time synchronization. It synchronizes time from its upstream clock node through the port. If an OC operates as the clock source, it sends synchronization time through a single PTP port to its downstream clock nodes.
- **Boundary Clock (BC)**—A clock with more than one PTP port in a PTP domain for time synchronization. A BC uses one of the ports to synchronize time from its upstream clock node. It uses the other ports to synchronize time to the relevant upstream clock nodes. If a BC operates as the clock source, such as BC 1 in Figure 1, it synchronizes time through multiple PTP ports to its downstream clock nodes.
- **Transparent Clock (TC)**—A TC does not keep time consistency with other clock nodes. A TC has multiple PTP ports. It forwards PTP messages among these ports and performs delay corrections for the messages, instead of performing time synchronization. TCs include the following types:
  - **End-to-End Transparent Clock (E2ETC)**—Forwards all PTP packets in the network and supports calculation of the delay on the entire link.
  - **Peer-to-Peer Transparent Clock (P2PTC)**—Forwards only Sync, Follow_Up, and Announce messages, terminates other PTP messages, and supports calculation of the delay on the entire link.

IMPORTANT:

When multiple PTP domains are configured on the device, the synchronization performance might fluctuate or decrease, and synchronization failure might occur because of limitation of hardware resources. OC and BC clock nodes do much more work than TCs. As a best practice, configure the device operating in multiple domains as an OC or BC clock node in a maximum of one PTP instance, to decrease calculations, minimize mutual influences between the domains, and optimize the synchronization performance.

Figure 1 shows the positions of these types of clock nodes in a PTP domain.

Figure 1 Clock nodes in a PTP domain



In addition to these basic types of clock nodes, PTP introduces hybrid clock nodes. For example, a TC+OC has multiple PTP ports in a PTP domain. One port is the OC type, and the others are the TC type.

A TC+OC forwards PTP messages through TC-type ports and performs delay corrections. In addition, it synchronizes time through its OC-type port. TC+OCs include these types: E2ETC+OC and P2PTC+OC.

Master-member/subordinate relationship

The master-member/subordinate relationship is automatically determined based on the Best Master Clock (BMC) algorithm. You can also manually specify a role for the clock nodes.

The master-member/subordinate relationship is defined as follows:

- **Master/Member node**—A master node sends a synchronization message, and a member node receives the synchronization message.
- **Master/Member clock**—The clock on a master node is a master clock (parent clock) The clock on a member node is a member clock.
- **Master/Subordinate port**—A master port sends a synchronization message, and a subordinate port receives the synchronization message. The master and subordinate ports can be on a BC or an OC.
- **Grant-port/request-port**—The IEEE 1588 version 2 PTP profile supports unicast negotiation. To configure unicast negotiation, you can specify PTP ports as request-ports or grant-ports. A grant-port grants and provides PTP service. A request-port requests and receives PTP service. Typically, a grant-port is in master state. A request-port is in slave state when selected as the port for time synchronization and in listening state when not selected.

A port that neither receives nor sends synchronization messages is a passive port.

Clock source type

A clock node supports the following clock sources:

- **Local clock source**—38.88 MHz clock signals generated by a crystal oscillator inside the clock monitoring module.
- **External clock source**—Clock signals generated by an external clock device. The signals are received and sent by a 1PPS/ToD port on the device. It is also called a ToD clock source.

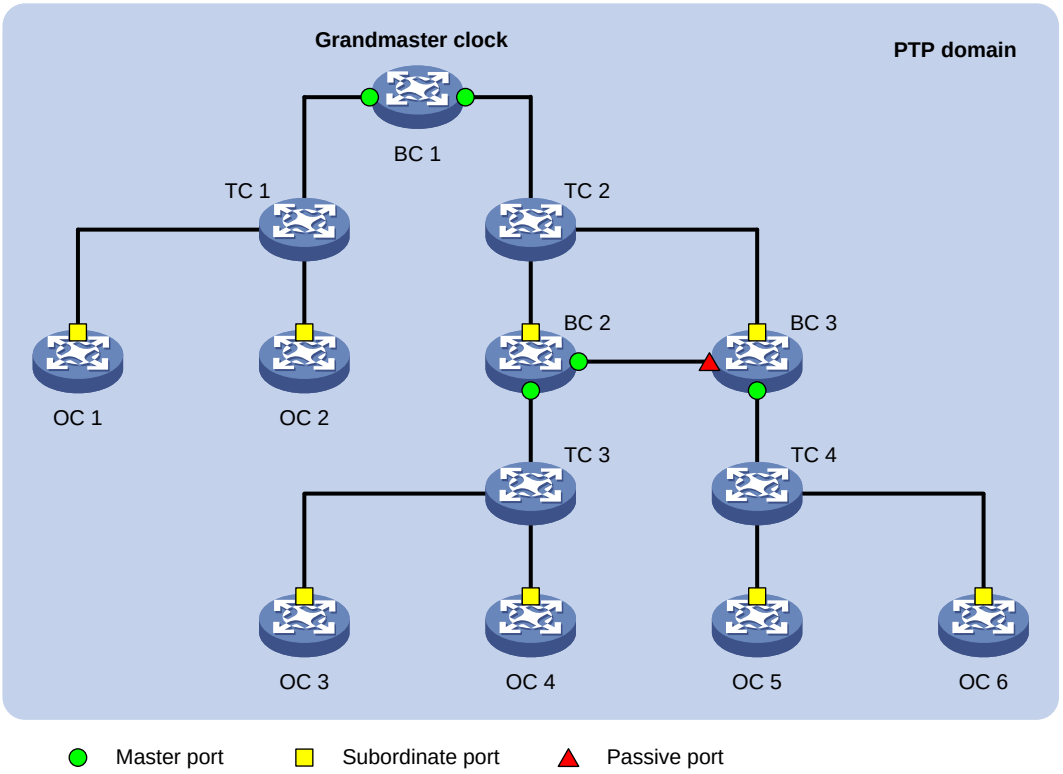
A clock node uses the local clock source, which is the 38.88 MHz clock signals generated by a crystal oscillator inside the clock monitoring module.

Grandmaster clock

As shown in Figure 2, the grandmaster clock (GM) is the ultimate source of time for clock synchronization in a PTP domain. It is elected automatically by the clock nodes in the PTP domain. The clock nodes exchange PTP messages and elect the GM by comparing the clock priority, time class, and time accuracy carried in the PTP messages.

You can also specify a GM manually.

Figure 2 Clock nodes in a PTP domain



Protocols and standards

- IEEE Std 1588-2008, *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, 2008*
- SMPTE ST 2059-2, *SMPTE Profile for Use of IEEE-1588 Precision Time Protocol in Professional Broadcast Applications*
- AES67-2015, *AES Standard for Audio Applications of Networks-High-Performance Streaming Audio-Over-IP Interoperability, 2015*

PTP time synchronization mechanisms

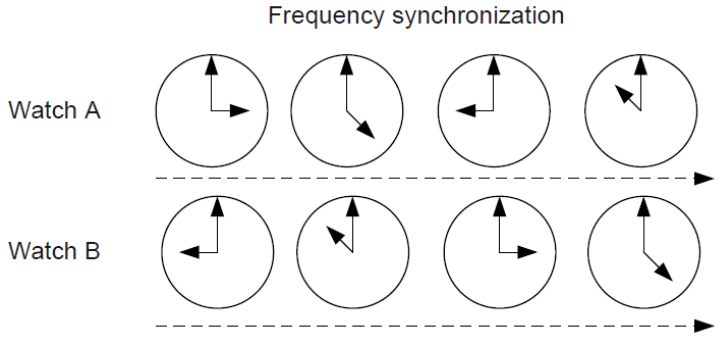
Time synchronization basics

Time synchronization includes frequency synchronization and phase synchronization. The synchronization requirements vary by networking environments.

Frequency synchronization

Frequency synchronization is also called clock synchronization. If two signals are the same in frequency or if they keep a constant phase difference, the two signals are frequency synchronized. As shown in Figure 3, the two watches have different time, but maintain a constant time difference (6 hours).

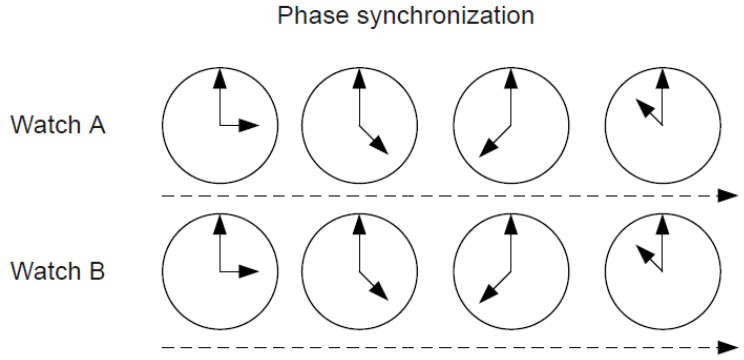
Figure 3 Frequency synchronization



Phase synchronization

Phase synchronization is also called time synchronization. If two signals are the same in frequency and phase (keep phase difference at 0), they are phase synchronized. As shown in Figure 4, the two watches have the same time at all times. Frequency synchronization is the prerequisite of phase synchronization.

Figure 4 Phase synchronization



Grandmaster clock selection and master-member/subordinate relationship establishment

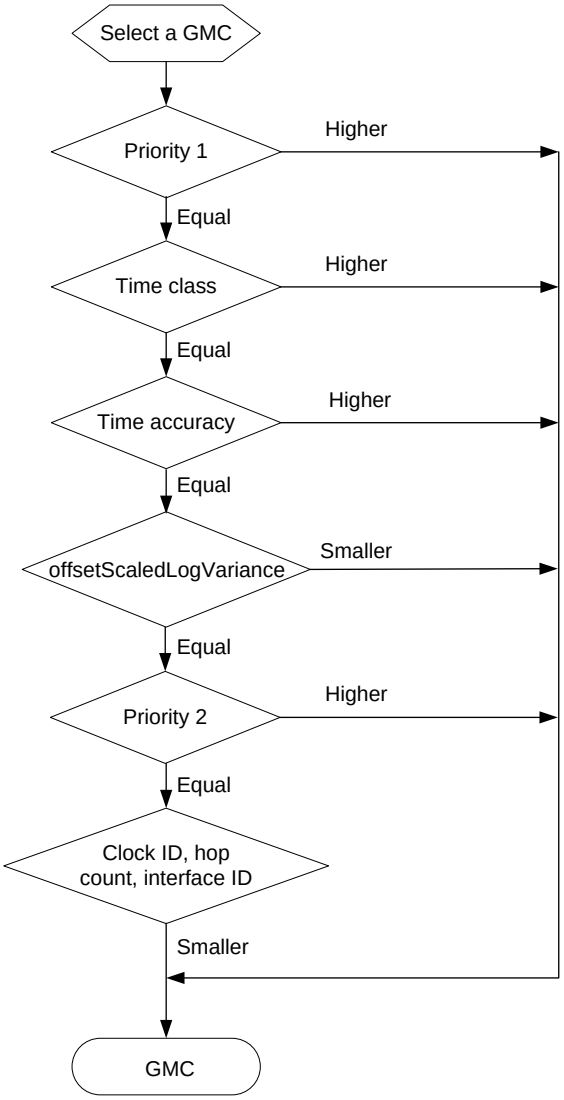
Grandmaster clock selection

A GM can be manually specified. It can also be elected through the BMC algorithm as follows:

1. The clock nodes in a PTP domain exchange announce messages and elect a GM by using the following rules in descending order:
   a. Clock node with higher priority 1.
   b. Clock node with higher time class.
   c. Clock node with higher time accuracy.
   d. Clock node with higher priority 2.
   e. Clock node with a smaller port ID (containing clock number and port number).

   The master nodes, member nodes, master ports, and subordinate ports are determined during the process. Then a spanning tree with the GM as the root is generated for the PTP domain.
2. The master node periodically sends announce messages to the member nodes. If the member nodes do not receive announce messages from the master node, they determine that the master node is invalid, and they start to elect another GM.

Figure 5 BMC comparison rule for GMC selection



Establishing master-member/subordinate relationships

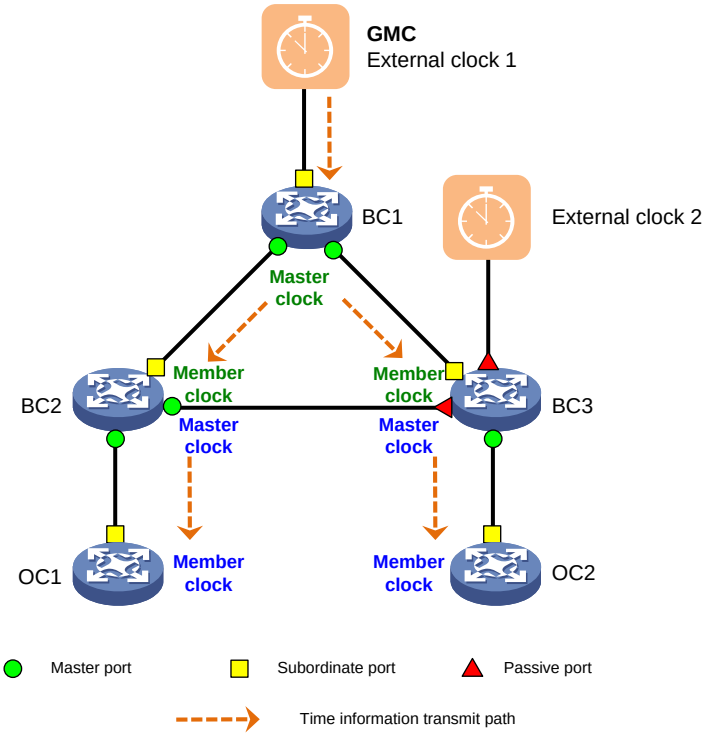
The master-member/subordinate relationships between the clock nodes (except for the TCs) on a PTP network are automatically determined based on the Best Master Clock (BMC) algorithm. You can also manually specify a role for the clock nodes.

The following master-member/subordinate relationships exist:

- **Master/Member node**—A master node propagates time synchronization information, and a member node receives time synchronization information.
- **Master/Member clock**—The clock on a master node is a master clock. The clock on a member node is a member clock.
- **Master/Subordinate port**—A master port propagates time synchronization, and a subordinate port receives time synchronization information. The master and subordinate ports can be on a BC or an OC. A port that neither propagates nor receives timing information is a passive port. Passive ports are used to prevent PTP messages from forming loops.

After the grandmaster clock is elected and master-member/subordinate relationships are established, a loop-free, all connected shortest path tree with the grandmaster clock as the root is established in the PTP domain, as shown in Figure 6. Time is synchronized between the clock nodes in the domain along the path tree.

Figure 6 PTP master-member/subordinate relationships



2.

Optimal domain selection

If only one PTP instance is configured on the device, this PTP instance is the optimal instance of the device. If multiple PTP instances are configured on the device, the optimal PTP instance (domain) is selected by using the following rules in descending order:

1. Activated instance (with a domain configured).
2. Instance on which the device is an OC or BC clock node.
3. Instance on which the clock source has a higher priority 1 value
4. Instance on which the clock source has a higher time class.
5. Instance on which the clock source has higher time accuracy.
6. Instance on which the clock source has a lower offset from the GM.
7. Instance on which the clock source has a higher priority 2 value.

8. Instance associated with a PTP domain that has a smaller domain ID.
9. The PTP instance on which the **`ptp active force-state`** command is configured has the lowest priority.

Frequency synchronization mechanism

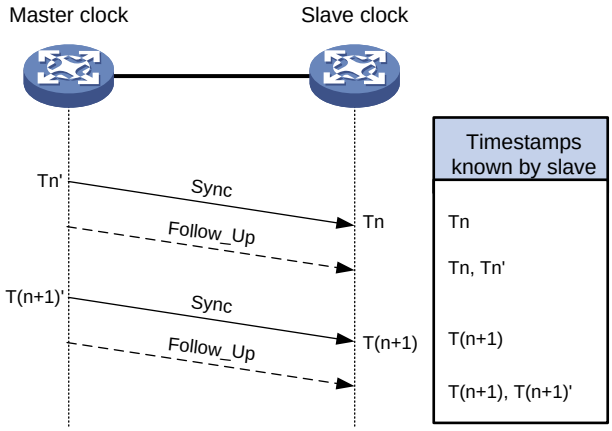
After the grandmaster clock is selected and master-member relationships are established, the master clock sends Sync messages to the member clocks periodically. Assume that the path delay between the master and member clock nodes remains unchanged, if the frequency of the member clock and the master clock are synchronized, the time offset between the master and member clock will be the same at the same intervals, which is T(n+1)–Tn = T(n+1)'–Tn', as shown in Figure 7.

- If T(n+1)–Tn > T(n+1)'–Tn', the member clock is faster in frequency than the master clock and will reduce its frequency.
- If T(n+1)–Tn < T(n+1)'–Tn', the member clock is slower in frequency than the master clock and will increase frequency faster.

The member clock uses the formula T(n+1)–Tn / T(n+1)'–Tn' to calculate its frequency ratio with the master clock and adjust its clock frequency accordingly.

PTP frequency synchronization can run in single-step mode or two-step mode. The figure below shows the most commonly used two-step frequency synchronization mode. In single-step mode, Tn' and T(n+1)' timestamps are carried in a Sync message, and no Follow_Up message is sent.

Figure 7 PTP frequency synchronization in two-step mode



PTP frequency synchronization can run in one of the following modes:

- **Single-step mode**—Uses Sync and Pdelay_Resp messages to carry timestamps, and does not send follow-up messages (Follow_Up or Pdelay_Resp_Follow_Up). The single-step mode has fewer interactive messages than the two-step mode, and occupies less network bandwidth, but it requires hardware chip support. In single-step mode, a timestamp is assigned before the message is sent. Therefore, the timestamp sent in single-step mode is the estimated message sending time that is not as accurate as the two-step mode.
- **Two-step mode**—Uses Follow_Up messages to carry Sync message transmit timestamps and Pdelay_Resp_Follow_Up messages to carry Pdelay_Req message receive and Pdelay_Resp message transmit timestamps. The timestamps sent in two-step mode is the actual sending time of the messages, so the two-step mode is more accurate than the single-step mode and can improve PTP time synchronization accuracy.

Phase synchronization mechanism

After master-member relationships are established between the clock nodes, the master and member clock nodes exchange PTP messages and record the message transmit and receive time. Based on the timestamps, each member clock calculates the path delay and time offset between them and the master clock and adjusts their time accordingly for time synchronization with the master clock.

PTP defines two path delay measurement mechanisms: Request_Response_ and Peer Delay, both of which are based on network symmetry.

Request_Response

The Request_Response mechanism measures the average path delay between the master and member clock nodes by using the PTP messages as shown in Figure 8. A TC between master and member clock nodes does not calculate the path delay. It forwards PTP messages and carries the Sync message residence time on it to the downstream clock node.

This mechanism can be implemented in one of the following two modes:

- **Two-step mode**—t1 is carried in the Follow_Up message as shown in Figure 8.
- **Single-step mode**—t1 is carried in the Sync message, and no Follow_Up message is sent.

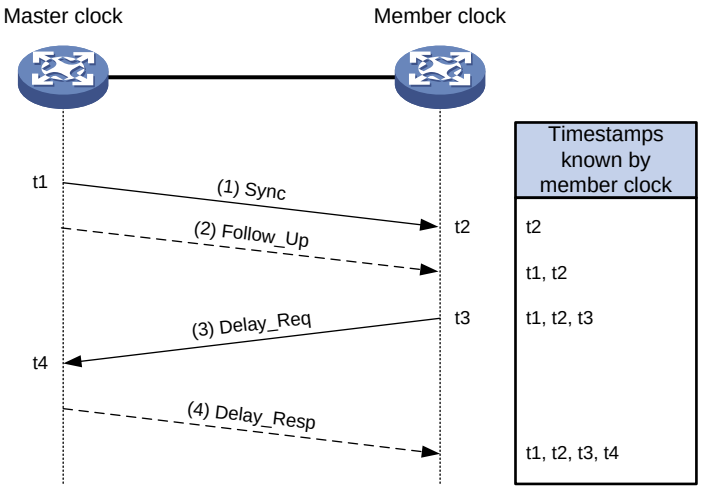
Figure 8 shows the Request_Response mechanism in two-step mode.

1. The master clock sends a Sync message to the member clock, and records the sending time t1. Upon receiving the message, the member clock records the receiving time t2.
2. After sending the Sync message, the master clock immediately sends a Follow_Up message that carries time t1.
3. The member clock sends a Delay_Req message to the master clock, and records the sending time t3. Upon receiving the message, the master clock records the receiving time t4.
4. The master clock returns a Delay_Resp message that carries time t4.

After this procedure, the member clock obtains all the four timestamps and can make the following calculations:

- Round-trip delay between the master and member clocks: (t2 – t1) + (t4 – t3)
- One-way delay between the master and member clocks: [(t2 – t1) + (t4 – t3)] / 2
- Offset between the member and master clocks: (t2 – t1) – [(t2 – t1) + (t4 – t3)] / 2 or [(t2 – t1) – (t4 – t3)] / 2

Figure 8 Request_Response mechanism (two-step node)



Peer Delay

The Peer Delay mechanism measures the average path delay between two clock nodes. The two clock nodes (BC, TC, or OC) implementing this mechanism send Pdelay messages to each other, and calculate the one-way link delay between them independently. The message interaction process and delay calculation method are identical on the two nodes. TCs that exist between master and member clock nodes divide the synchronization path into multiple links and participate in delay calculation. The link delays and Sync message residence time on the TCs are carried to downstream nodes.

This mechanism can be implemented in one of the following two modes:

- Two-step mode

  As shown in Figure 9, Pdelay messages include Pdelay_Req, Pdelay_Resp, and Pdelay_Resp_Follow_Up messages. t2 is carried in the Pdelay_Resp message, and t3 is carried in the Pdelay_Resp_Follow_Up message.

- Single-step mode:

  Pdelay messages include Pdelay_Req and Pdelay_Resp messages. t3 – t2 is carried in the Pdelay_Resp, and no Pdelay_Resp_Follow_Up message is sent.

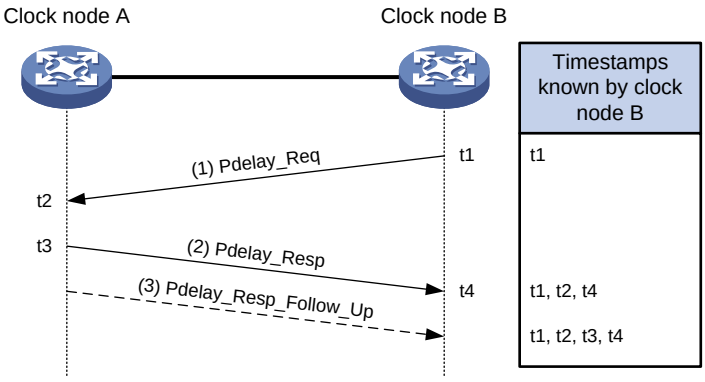
Figure 9 uses Clock node B as an example to describe the Peer Delay mechanism.

1. Clock node B sends a Pdelay_Req message to Clock node A, and records the sending time t1. Upon receiving the message, Clock node A records the receiving time t2.
2. Clock node A sends a Pdelay_Req message that carries t2 to Clock node B, and records the sending time t3. Upon receiving the message, Clock node B records the receiving time t4.
3. Clock node A immediately sends a Pdelay_Resp_Follow_Up message carrying t3 to Clock node B after sending the Pdelay_Req message.

After this procedure, Clock node B obtains all the four timestamps and can make the following calculations:

- Round-trip delay between Clock node A and Clock node B: (t2 – t1) + (t4 – t3)
- One-way delay between Clock node A and Clock node B: [(t2 – t1) + (t4 – t3)] / 2 = [(t4 – t1) – (t3 – t2)] / 2
- Time offset between the member clock and the master clock: Sync message receiving time on the member clock – Sync message sending time on the master clock – Total one-way delays on all links – Total Sync message residence time on all TCs.

Figure 9 Peer Delay mechanism (two-step mode)



PTP support for IPv6 networks

By default, PTP is encapsulated in Layer 2 packets and can be deployed in both IPv4 and IPv6 networks. If the **ptp transport-protocol** command is configured, PTP packets are transmitted using the UDP protocol. Support for PTP deployment in IPv6 networks depends on the device model. If the device supports PTP functionality in IPv6, to deploy PTP in IPv6 and use multicast to send PTP packets, use the **ptp source ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*] command to configure a source IPv6 address for PTP. To deploy PTP in IPv6 and use multicast to send PTP packets, use the **ptp source ipv6** *ipv6-address* [**vpn-instance** *vpn-instance-name*] command to configure a source IPv6 address for PTP. To use unicast to send PTP packets, use the **ptp unicast-destination ipv6** *ipv6-address* command to configure a destination IPv6 address for PTP.

Association between PTP and Track

If you associate a track entry with PTP, and then associate the track entry with an EAA policy, automatic monitoring of the PTP clock source and PTP clock state can be achieved. This enables the device to use PTP to synchronize the system clock when PTP is operating correctly, and use NTP to synchronize the system clock when PTP is not operating correctly.

After you configure this feature, the device creates a track entry associated with PTP. The state of the track entry depends on the state of the associated PTP object. The device supports associating Track with the following types of PTP objects:

- **Clock source**—Determines the state of the track entry based on the clock source state in the specified instance.
  - If the PTP clock source already exists in the specified PTP instance (that is, the device has a subordinate port, and the device is not a GM in the network), the track entry state is Positive.
  - If the PTP clock source is lost in the specified PTP instance (that is, the device is a GM in the network), the track entry state is Negative.
- **PTP time state**—Determines the state of the track entry based on the PTP time state.
  - If the PTP time state is Lock, the track entry state is Positive.
  - If the PTP time state is Unlock, the track entry state is Negative.

You can use the Boolean AND tracked list to monitor both the PTP clock source state and PTP time state. For more information about PTP, see "Configuring track" in *High Availability Configuration Guide*.

Restrictions and guidelines: PTP configuration

Before configuring PTP, determine the PTP profile and define the scope of the PTP domain and the role of every clock node.

In a PTP domain that runs the IEEE 1588 version 2 PTP profile, specify the BC or OC clock node type for the devices in the domain as a best practice. A TC clock node is mainly used for forwarding PTP messages. When receiving a PTP message, a TC multicasts the message from all its PTP interfaces except the inbound interface of the PTP message. To deploy a TC in the domain, plan the TC location and number and location of PTP interfaces on the TC manually to prevent PTP forwarding loops.

PTP tasks at a glance

Configuring PTP (IEEE 1588 version 2)

1. Specifying PTP for obtaining the time
2. (Optional.) Creating a PTP instance
3. Specifying a PTP profile

   Specify the IEEE 1588 version 2 PTP profile.
4. Configuring clock nodes
   - Specifying a clock node type
   - (Optional.) Configuring an OC to operate only as a member clock
5. Specifying a PTP domain
6. (Optional.) Configuring the clock source whitelist
7. Enabling PTP.

   For PTP to run on an interface, enable PTP globally and on the interface.
   - Enabling PTP globally
   - Enabling PTP on a port
8. Configuring PTP ports
   - (Optional.) Configuring the role of a PTP port
   - Configuring the mode for carrying timestamps
   - Specifying a delay measurement mechanism for a BC or an OC
   - Configuring one of the ports on a TC+OC clock as an OC-type port
   - (Optional.) Enabling unicast negotiation on PTP ports
9. (Optional.) Configuring PTP message transmission and receipt
   - Setting the interval for sending announce messages and the timeout multiplier for receiving announce messages
   - Setting the interval for sending Pdelay_Req messages
   - Setting the interval for sending Sync messages
   - Setting the minimum interval for sending Delay_Req messages
   - Setting the interval for sending Delay-Resp messages
10. (Optional.) Configuring parameters for PTP messages
    - Specifying the UDP transport protocol for PTP messages
    - Configuring a source IP address for multicast PTP messages transmitted over UDP
    - Configuring a destination IP address for unicast PTP messages transmitted over UDP
    - Configuring the MAC address for non-Pdelay messages
    - Setting a DSCP value for PTP messages transmitted over UDP
    - Specifying a VLAN tag for PTP messages
    - Disabling PTP path tracing
11. (Optional.) Adjusting and correcting clock synchronization
    - Setting the delay correction value
    - Calculating the TAI based on the UTC time
    - Configuring the PTP offset threshold and PTP synchronization suppressions
12. (Optional.) Configuring a priority for a clock
13. (Optional.) Configuring PTP logging

14. (Optional.) Enabling SNMP notification for the PTP module

Configuring PTP (SMPTE ST 2059-2)

1. Specifying PTP for obtaining the time
2. (Optional.) Creating a PTP instance
3. Specifying a PTP profile

   Specify the SMPTE ST 2059-2 PTP profile.
4. Configuring clock nodes
   - Specifying a clock node type
   - (Optional.) Configuring an OC to operate only as a member clock
5. (Optional.) Specifying a PTP domain
6. (Optional.) Configuring the clock source whitelist
7. Enabling PTP.

   For PTP to run on an interface, enable PTP globally and on the interface.
   - Enabling PTP globally
   - Enabling PTP on a port
8. Configuring PTP ports
   - (Optional.) Configuring the role of a PTP port
   - Configuring the mode for carrying timestamps
   - Specifying a delay measurement mechanism for a BC or an OC
9. (Optional.) Configuring PTP message transmission and receipt
   - Setting the interval for sending announce messages and the timeout multiplier for receiving announce messages
   - Setting the interval for sending Pdelay_Req messages
   - Setting the interval for sending Sync messages
   - Setting the minimum interval for sending Delay_Req messages
10. (Optional.) Configuring parameters for PTP messages
    - Specifying the UDP transport protocol for PTP messages
    - Configuring a source IP address for multicast PTP messages transmitted over UDP
    - Configuring a destination IP address for unicast PTP messages transmitted over UDP
    - Setting a DSCP value for PTP messages transmitted over UDP
    - Specifying a VLAN tag for PTP messages
    - Disabling PTP path tracing
11. (Optional.) Adjusting and correcting clock synchronization
    - Setting the delay correction value
    - Calculating the TAI based on the UTC time
    - Configuring the PTP offset threshold and PTP synchronization suppressions
12. (Optional.) Configuring a priority for a clock
13. (Optional.) Configuring PTP logging
14. (Optional.) Enabling SNMP notification for the PTP module

Configuring PTP (AES67-2015)

1. Specifying PTP for obtaining the time

2. (Optional.) Creating a PTP instance
3. Specifying a PTP profile

   Specify the AES67-2015 PTP profile.
4. Configuring clock nodes
   - Specifying a clock node type
   - (Optional.) Configuring an OC to operate only as a member clock
5. (Optional.) Specifying a PTP domain
6. (Optional.) Configuring the clock source whitelist
7. Enabling PTP.

   For PTP to run on an interface, enable PTP globally and on the interface.
   - Enabling PTP globally
   - Enabling PTP on a port
8. Configuring PTP ports
   - (Optional.) Configuring the role of a PTP port
   - Configuring the mode for carrying timestamps
   - Specifying a delay measurement mechanism for a BC or an OC
   - Configuring one of the ports on a TC+OC clock as an OC-type port
9. (Optional.) Configuring PTP message transmission and receipt
   - Setting the interval for sending announce messages and the timeout multiplier for receiving announce messages
   - Setting the interval for sending Pdelay_Req messages
   - Setting the interval for sending Sync messages
   - Setting the minimum interval for sending Delay_Req messages
10. (Optional.) Configuring parameters for PTP messages
    - Specifying the UDP transport protocol for PTP messages
    - Configuring a source IP address for multicast PTP messages transmitted over UDP
    - Configuring a destination IP address for unicast PTP messages transmitted over UDP
    - Setting a DSCP value for PTP messages transmitted over UDP
    - Specifying a VLAN tag for PTP messages
    - Disabling PTP path tracing
11. (Optional.) Adjusting and correcting clock synchronization
    - Setting the delay correction value
    - Calculating the TAI based on the UTC time
    - Configuring the PTP offset threshold and PTP synchronization suppressions
12. (Optional.) Configuring a priority for a clock
13. (Optional.) Configuring PTP logging
14. (Optional.) Enabling SNMP notification for the PTP module

Specifying PTP for obtaining the time

1. Enter system view.

   `system-view`
2. Specify PTP for obtaining the time.

   `clock protocol ptp`

By default, the device uses NTP to obtain the system time.

For more information about the **clock protocol** command, see device management commands in *Fundamentals Command Reference*.

Creating a PTP instance

About this task

A PTP instance is uniquely identified by its ID on a device. For easy identification and management, you can also set a name for a PTP instance.

Restrictions and guidelines

Do not set the same name for different PTP instances.

If you create PTP instances with the same ID but different names, the most recent configuration takes effect.

The default PTP instance with an ID of 0 has been created by default. You cannot delete the default PTP instance. PTP settings configured in system view take effect only on PTP instance 0. To configure settings for a PTP instance other than PTP instance 0, enter PTP instance view.

Procedure

1. Enter system view.

   system-view
2. Create a PTP instance.

   ptp instance *ptp-instance-id* [**name** *ptp-instance-name*]

   By default, PTP instance numbered **0** and named **default-instance** exists.

Specifying a PTP profile

Restrictions and guidelines

You must specify a PTP profile before configuring PTP settings. Changing the PTP profile restores all settings under the profile to the default.

Procedure

1. Enter system view.

   system-view
2. (Optional.) Enter PTP instance view.

   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.
3. Specify a PTP profile.

   ptp profile { **1588v2** | **aes67-2015** | **st2059-2** }

   By default, no PTP profile is configured, and PTP is not running on the device.

Configuring clock nodes

Specifying a clock node type

Restrictions and guidelines

Before you specify a clock node type, specify a PTP profile.

Changing the clock node type restores all settings to the default except for the PTP profile.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Enter PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
3. Specify a clock node type for the device.
   - IEEE 1588v2 PTP profile:
     ptp mode { **bc** | **e2etc** | **e2etc-oc** | **oc** | **p2ptc** | **p2ptc-oc** }
   - AES67-2015 or SMPTE ST 2059-2 PTP profile:
     ptp mode { **bc** | **e2etc** | **oc** | **p2ptc** }

   By default, no clock node type is specified.

Configuring an OC to operate only as a member clock

About this task

An OC can operate either as a master clock to send synchronization messages or as a member clock to receive synchronization messages. This task allows you to configure an OC to operate only as a member clock.

If an OC is operating only as a member clock, you can use the **ptp force-state** command to configure its PTP port as a master port or passive port.

Restrictions and guidelines

This task is applicable only to OCs.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Enter PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
3. Configure the OC to operate only as a member clock.
   ptp slave-only
   By default, an OC operates as a master or member clock.
   This command is mutually exclusive with the **ptp unicast-negotiate** command.

Specifying a PTP domain

About this task

Within a PTP domain, all devices follow the same rules to communicate with each other. Devices in different PTP domains cannot exchange PTP messages.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Enter PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
3. Specify a PTP domain for the device.
   ptp domain *value*
   By default, no PTP domain exists.
   Do not configure the same domain on different PTP instances.

Configuring the clock source whitelist

About this task

Application scenarios

In large network environments, especially when multiple grandmaster (GM) clocks exist, enabling the clock source whitelist can enhance network stability and security. This function restricts devices to synchronize time only from known and trusted clock sources on the whitelist, thereby avoiding malicious clock attacks and network time oscillations.

Operating mechanism

By default, when an interface with PTP enabled receives a PTP clock synchronization signal, it uses the Best Clock Master Algorithm (BCMA) to select a clock source from multiple clock signals. Devices within the same PTP clock synchronization network will use the time of this selected clock source as their reference and maintain time synchronization with it.

When the clock source whitelist is enabled and the whitelist entries are specified, only clock sources on the whitelist will participate in the clock source election and be considered for PTP time synchronization by the device. If the currently synchronized clock source is not on the whitelist, the device stops synchronizing with that clock source. The clock source whitelist configured in PTP instance view is only effective for that PTP instance, and the whitelist configured in system view is only effective for PTP instance 0. Each PTP instance independently elects its clock source based on its whitelist.

Restrictions and guidelines

Before enabling the clock source whitelist, make sure the whitelist includes all trusted clock sources, including the primary clock source and all backup clock sources.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Enter PTP instance view.
   ptp instance *ptp-instance-id*
   To configure the default instance with an instance ID of 0, skip this step.

Make sure the PTP domain is different for different instances.

3. Add a clock source to the clock source whitelist.
   ptp whitelist clockid *clock-id*
   By default, the clock source whitelist does not exist.
4. Enable the clock source whitelist.
   ptp whitelist clockid enable
   By default, the clock source whitelist is disabled.

Enabling PTP globally

Restrictions and guidelines

For PTP to run on an interface, you must enable PTP globally and on the interface.

Procedure

1. Enter system view.
   system-view
2. Enable PTP globally.
   ptp global enable
   By default, PTP is enabled globally.

Enabling PTP on a port

About this task

A port enabled with PTP becomes a PTP port.

Restrictions and guidelines

You can enable PTP on only one port on an OC.

To enable PTP on a Layer 3 Ethernet interface that has been assigned to a VPN instance, you must execute the **ptp source** command and specify the VPN instance if PTP messages are to be transmitted in multicast mode over UDP.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Enable PTP on the port.
   ptp enable
   By default, PTP is disabled on a port.

Configuring PTP ports

Configuring the role of a PTP port

About this task

You can configure the master, passive, or slave role for a PTP port.

For an OC that operates in **slave-only** mode, you can perform this task to change its PTP port role to master or slave.

Restrictions and guidelines

- You can configure one subordinate port on the device.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Configure the role of the PTP port.
   ptp force-state { **master** | **passive** | **slave** }
   By default, the PTP port role is automatically calculated through BMC.
   This command is mutually exclusive with the **ptp unicast-negotiate** command.
5. Return to system view.
   quit
6. (Optional.) Enter PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
7. Activate the port role configuration.
   ptp active force-state
   By default, the port role configuration is not activated.

Configuring the mode for carrying timestamps

About this task

Timestamps can be carried in either of the following modes:

- **Single-step mode**—The Sync message (in the Request_Response or Peer Delay mechanism) and Pdelay_Resp message (in the Peer Delay mechanism) carry their sending timestamps by themselves.
- **Two-step mode**—The Sync message (in the Request_Response or Peer Delay mechanism) and Pdelay_Resp message (in the Peer Delay mechanism) do not carry their sending timestamps by themselves. The subsequent messages carry their sending timestamps.

Procedure

1. Enter system view.

system-view

2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.
4. Configure the mode for carrying timestamps.
   ptp clock-step { **one-step** | **two-step** }

   By default, the two-step mode is used for carrying timestamps.

Specifying a delay measurement mechanism for a BC or an OC

About this task

PTP defines two transmission delay measurement mechanisms: Request_Response and Peer Delay. For correct communication, ports on the same link must share the same delay measurement mechanism.

Restrictions and guidelines

Follow these restrictions and guidelines when you perform this task:

- For the IEEE 1588 version 2, SMPTE ST 2059-2, or AES67-2015 PTP profile, you can configure this task only when the PTP clock node type is BC or OC.
- You cannot configure this task when the PTP clock node type is E2ETC, E2ETC+OC, P2PTC, or P2PTC+OC. The E2ETC and E2ETC+OC clock nodes support both the request-response and peer delay measurement mechanisms. The P2PTC clock node supports only the peer delay measurement mechanism.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.
4. Specify a delay measurement mechanism for a BC or an OC.
   ptp delay-mechanism { **e2e** | **p2p** }

   By default, when the PTP profile is IEEE 1588 version 2, AES67-2015, or SMPTE ST 2059-2, the request-response delay measurement mechanism applies.

   The **ptp delay-mechanism p2p** command is mutually exclusive with the **ptp unicast-negotiate** command.

Configuring one of the ports on a TC+OC clock as an OC-type port

About this task

All ports on a TC+OC (E2ETC+OC or P2PTC+OC) are TC-type ports by default. This feature allows you to configure one of the ports on a TC+OC clock as an OC-type port.

Restrictions and guidelines

This task is applicable only to E2ETC+OCs and P2PTC+OCs.

This task is not available for the SMPTE ST 2059-2 or AES67-2015 PTP profile.

For time synchronization accuracy, the OC-type port on an E2ETC+OC or P2PTC+OC must be specified as the master port.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Configure the port type as OC.
   ptp port-mode oc
   By default, the port type for all ports on a TC+OC is TC.

Enabling unicast negotiation on PTP ports

About this task

On a network where the IEEE 1588 version 2 PTP profile runs, you can use this task to establish master-subordinate relations between PTP ports by specifying the request-ports and grant-ports. A grant-port grants and provides PTP service. A request-port requests and receives PTP service. Typically, a grant-port is in master state. A request-port is in slave state when selected as the port for time synchronization and in listening state when not selected.

A request-port initiates a PTP link connection request to a specific grant-port. Then they negotiate PTP parameters and exchange PTP messages to synchronize the client time to the server. Plan the network in advance and specify the ports on the member (client) clock nodes as request-ports and the ports on the master (server) clock nodes as grant-ports.

Restrictions and guidelines

Only the IEEE 1588 version 2 profile supports this task.

Follow these restrictions and guidelines to configure unicast negotiation under the IEEE 1588 version 2 PTP profile:

- The **ptp transport-protocol udp** command must be configured for unicast negotiation to take effect.
- The E2ETC, P2PTC, and P2PTC-OC clock nodes do not support unicast negotiation
- The **ptp unicast-negotiate** command is mutually exclusive with the **ptp slave-only**, **ptp force-state**, and **ptp delay-mechanism p2p** commands.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter PTP instance view.
   ptp instance *ptp-instance-id*
   Skip this step for PTP instance 0.
4. Enable unicast negotiation and specify the port as a grant-port or request port.
   ptp unicast-negotiate { **grant-port** | **request-port** }

Configuring PTP message transmission and receipt

Setting the interval for sending announce messages and the timeout multiplier for receiving announce messages

About this task

A master node periodically sends announce messages at the specified interval. If a member node does not receive any announce message from the master node after the timeout expires, it determines that the master node is invalid. The timeout = timeout multiplier × interval at which the master node sends announce messages.

The interval at which the master node sends announce messages depends on whether unicast negotiation is enabled.

- If unicast negotiation is not enabled, the master node uses the value configured by using the **ptp announce-interval** command on its interface as the interval.
- If unicast negotiation is enabled, you must configure the interval on the request-port of the client. The request-port uses the configured interval to negotiate with the grant-port on the server for the interval at which the grant-port sends announce messages to the client. If the negotiation succeeds, the grant-port sends announce messages at the configured interval to the client. If the negotiation fails, the grant-port does not send announce messages to the client.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Set the interval for sending announce messages.
   ptp announce-interval *interval*
   The default settings vary by PTP profile.
   - **IEEE 1588 version 2 or AES67-2015**—The *interval* argument value is 1 and the interval for sending announce messages is 2 (2^1) seconds.

- **SMPTE ST 2059-2**—The *interval* argument value is –2 and the interval for sending announce messages is 1/4 (2^{-2}) seconds.

5. Set the number of intervals before a timeout occurs.

   ptp announce-timeout *multiple-value*

   By default, a timeout occurs when three intervals are reached.

Setting the interval for sending Pdelay_Req messages

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.

   interface *interface-type interface-number*

3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.

   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.

4. Set the interval for sending Pdelay_Req messages.

   ptp pdelay-req-interval *interval*

   By default, the *interval* argument value is 0 and the interval for sending peer delay request messages is 1 (2^0) second.

   For the SMPTE ST 2059-2 or AES67-2015 PTP profile, set the *interval* argument to a value in the range of **ptp syn-interval** *interval* to **ptp syn-interval** *interval* plus 5 as a best practice.

Setting the interval for sending Sync messages

About this task

The master node sends Sync messages to the member nodes periodically. The interval at which the master node sends Sync messages to member nodes depends on whether unicast negotiation is enabled.

- If unicast negotiation is not enabled, use this task to configure the interval on the master node.
- If unicast negotiation is enabled, you must configure the interval on the request-port of the client. The request-port uses the configured interval to negotiate with the grant-port on the server for the interval at which the grant-port sends Sync messages to the client. If the negotiation succeeds, the grant-port sends Sync messages at the configured interval to the client. If the negotiation fails, the grant-port does not send Sync messages to the client.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.

   interface *interface-type interface-number*

3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.

   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.

4. Set the interval for sending Sync messages.

   ptp syn-interval *interval*

   The default settings vary by PTP profile.

- **IEEE 1588 version 2**—The *interval* argument value is 0 and the interval for sending Sync messages is 1 (2^0) second.
- **IEEE 802.1AS, AES67-2015, or SMPTE ST 2059-2**—The *interval* argument value is –3 and the interval for sending Sync messages is 1/8 (2^{-3}) seconds.

Setting the minimum interval for sending Delay_Req messages

About this task

When receiving a Sync or Follow_Up message, an interface can send Delay_Req messages only when the minimum interval is reached. This task allows you to set the minimum interval for sending Delay_Req messages.

Restrictions and guidelines

In PTP multicast transport mode, this setting takes effect only when configured on the master clock. The master clock sends the value to a member clock through PTP messages to control the interval for the member clock to send Delay_Req messages. To view the interval, execute the **display ptp interface** command on the member clock.

In PTP unicast transport mode, this setting takes effect when configured on member clocks. It does not take effect when configured on the master clock.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Set the minimum interval for sending Delay_Req messages.
   ptp min-delayreq-interval *interval*
   The *interval* argument value is 0 and the minimum interval for sending delay request messages is 1 (2^0) second.
   For the SMPTE ST 2059-2 PTP profile, set the *interval* argument to a value in the range of **ptp syn-interval** *interval* to **ptp syn-interval** *interval* plus 5 as a best practice.

Setting the interval for sending Delay-Resp messages

About this task

This task, configured on a unicast negotiation request-port, specifies the interval at which the grant-port sends Delay_resp messages to the request-port. After receiving a Delay_req message from a request port, the grant-port responds by sending a Delay_resp message and starts a timer defined by this command. The grant-port will not send another Delay_resp message until it receives a Delay_req message after the timer expires.

Restrictions and guidelines

This task is available only for the IEEE 1588 version 2 PTP profile and is configured on unicast negotiation request-ports.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter PTP instance view.
   ptp instance *ptp-instance-id*
   Skip this step for PTP instance 0.
4. Set the interval for sending Delay_resp messages.
   ptp delay-resp-interval *interval*
   By default, the value of the *interval* argument is 0 and the minimum interval for sending Delay_resp messages is 1 (2^0) seconds.

Configuring parameters for PTP messages

Specifying the UDP transport protocol for PTP messages

About this task

PTP messages can be transported over IEEE 802.3/Ethernet or UDP.

Restrictions and guidelines

The SMPTE ST 2059-2 or AES67-2015 PTP profile supports only UDP transport of PTP messages and does not support this task.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Specify the UDP transport protocol for PTP messages.
   ptp transport-protocol udp
   By default, PTP messages are transported over IEEE 802.3/Ethernet.

Configuring a source IP address for multicast PTP messages transmitted over UDP

About this task

To transport multicast PTP messages over UDP, you must configure a source IP address for the messages.

Restrictions and guidelines

If both a source IP address for multicast PTP messages transmitted over UDP and a destination address for unicast PTP messages transmitted over UDP are configured, the system unicasts the messages.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Enter PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
3. Configure a source IP address for multicast PTP messages transmitted over UDP.
   ptp source { *ip-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]
   By default, no source IP address is configured for multicast PTP messages transmitted over UDP.

Configuring a destination IP address for unicast PTP messages transmitted over UDP

About this task

To transport unicast PTP messages over UDP, you must configure a destination IP address for the messages.

Restrictions and guidelines

If both a source IP address for multicast PTP messages transmitted over UDP and a destination address for unicast PTP messages transmitted over UDP are configured, the system unicasts the messages.

Prerequisites

Configure an IP address for the current interface, and make sure the interface and the peer PTP interface can reach each other.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Configure a destination IP address for unicast PTP messages transmitted over UDP.
   ptp unicast-destination { *ip-address* | **ipv6** *ipv6-address* }
   By default, no destination IP address is configured for unicast PTP messages transmitted over UDP.

Configuring the MAC address for non-Pdelay messages

About this task

Pdelay messages include Pdelay_Req, Pdelay_Resp, and Pdelay_Resp_Follow_Up messages. The destination MAC address of Pdelay messages is 0180-C200-000E by default, which cannot be modified. The destination MAC address of non-Pdelay messages is either 0180-C200-000E or 011B-1900-0000.

Restrictions and guidelines

This feature takes effect only when PTP messages are encapsulated in IEEE 802.3/Ethernet packets.

This task is not available for the SMPTE ST 2059-2 or AES67-2015 PTP profile.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.
   interface *interface-type interface-number*
4. Configure the destination MAC address for non-Pdelay messages.
   ptp destination-mac *mac-address*

   The default destination MAC address is 011B-1900-0000.

Setting a DSCP value for PTP messages transmitted over UDP

About this task

The DSCP value determines the sending precedence of PTP messages transmitted over UDP.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.
4. Set a DSCP value for PTP messages transmitted over UDP.
   ptp dscp *dscp*

   By default, the DSCP value is 56.

Specifying a VLAN tag for PTP messages

About this task

Perform this task to configure the VLAN ID and the 802.1p precedence in the VLAN tag carried by PTP messages.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view.

interface *interface-type interface-number*

3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Specify a VLAN tag for PTP messages.
   ptp vlan *vlan-id* [**dot1p** *dot1p-value*]
   By default, PTP messages do not have a VLAN tag.

Disabling PTP path tracing

About this task

PTP path tracing traces the clock nodes that the clock signals traverse from the GM to the device. The system can obtain complete path tracing information only when all clock nodes on the path are enabled with PTP path tracing. If a device on the path does not support PTP path tracing, disable this feature to prevent PTP intercommunication issues.

Procedure

1. Enter system view.
   system-view
2. Disable PTP path tracing.
   ptp path-trace disable
   By default, PTP path tracing is enabled.

Adjusting and correcting clock synchronization

Setting the delay correction value

About this task

PTP performs time synchronization based on the assumption that the delays in sending and receiving messages are the same. However, this is not practical. If you know the offset between the delays in sending and receiving messages, you can set the delay correction value for more accurate time synchronization.

Procedure

1. Enter system view.
   system-view
2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.
   interface *interface-type interface-number*
3. (Optional.) Assign the interface to a PTP instance and enter interface PTP instance view.
   ptp instance *ptp-instance-id*
   To configure settings for PTP instance 0, skip this step.
4. Set a delay correction value.
   ptp asymmetry-correction { **minus** | **plus** } *value*
   The default is 0 nanoseconds. Delay correction is not performed.

Calculating the TAI based on the UTC time

About this task

UTC and TAI

International Atomic Time (TAI) is a high-precision atomic coordinate time scale. Coordinated Universal Time (UTC) is based on TAI and adjusted by leap seconds for synchronization with astronomical time.

UTC is constantly compared with UT1 (astronomical time, reflecting the earth's rotation). The International Bureau of Weights and Measures (BIPM) will notify to add or subtract one second from the UTC time scale at the last minute on a specific date (typically June 30 or December 31) to ensure that the difference between UTC and UT1 is within one second.

The BIPM publishes the difference between UTC and TAI periodically. Till now, UTC is 37 seconds behind TAI.

Internet devices typically use the UTC time. To provide more accurate time, PTP also uses TAI. If the device acts as a subordinate clock node, it uses the UTC time and TAI time of the clock reference. The device locally has the following clock sources: one is UTC time (time generated by the local oscillator on the device MPU), and the other is TAI time (also known as PTP time, which is produced by the clock module/PTP chip). If the device acts as the clock reference, it will calculate the TAI time based on the UTC time, and synchronize the UTC and TAI time to the subordinate clocks. This task is used to calculate the TAI time based on the UTC time.

Adjustment of the deviation between UTC and TAI

TAI = Current UTC + cumulative offset + leap second that might occur in the future. To get the correct TAI time from the UTC time,

- Set the accurate UTC time on the device.
- Use the **ptp utc offset** command to set the cumulative offset of UTC relative to TAI as published by BIPM. This command takes effect immediately.
- If the BIPM notifies to add or subtract one second from the UTC time scale on a specific data after you use the **ptp utc offset** command to configure the cumulative offset of TAI relative to UTC, use the **ptp utc** { **leap59-date** | **leap61-date** } *date* command to change the UTC time. The TAI time will be changed accordingly.

UTC and TAI synchronization

When the device is used as the GM clock in a network, and services that rely on PTP time synchronization are deployed in the network (such as iFIT measurement and deterministic networking technology.), jumps in PTP time can affect service operation. Configuring the synchronization of TAI time using UTC time can reduce the jumps in PTP time and thereby minimize the impact of PTP time jumps on services.

When the device is used as the GM clock in the network, the device will perform the following by default:

1. The device uses its own UTC time as the reference clock source for the entire network.
2. After the network administrator enables the PTP function in the interface view, the device synchronizes UTC time to the clock module/PTP chip.
3. Once the clock module/PTP chip synchronizes with the UTC time of this device, the device uses the timing frequency of the clock module/PTP chip for timing and uses TAI time to synchronize PTP time to downstream devices.
4. The device automatically synchronizes UTC time to the clock module/PTP chip that has the PTP function enabled every 6 hours.
5. If the network administrator modifies the UTC time (for example, by configuring the **clock data-time**, **ptp utc offset, or ptp utc** { **leap59-date** | **leap61-date** } commands), the device immediately and automatically synchronizes the updated UTC time to

the clock module/PTP chip. This ensures that the UTC time of devices in the network is consistent with the UTC time of this device.

Because TAI time and UTC time are synchronized and then counted separately according to their respective frequencies, extremely subtle deviations might exist between TAI time and UTC time. Synchronizing TAI time with UTC time every 6 hours can cause jumps in TAI time, which in turn causes jumps across the entire network time, further affecting the operation of services that depend on PTP time synchronization.

To avoid this issue, you can synchronize TAI time to UTC time. The device will synchronize TAI time to UTC time every 6 hours, thus avoiding time jumps that affect the entire network.

Restrictions and guidelines

This setting takes effect only when it is configured on the master clock node and the local clock of the master clock node is the GM.

Configure this feature on both the master and member nodes for the new master clock to provide accurate TAI time after a master/member node switchover.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Enter PTP instance view.

   ptp instance *ptp-instance-id*

   Skip this step for PTP instance 0.

3. Set the cumulative offset of UTC relative to TAI.

   ptp utc offset *utc-offset*

   The default is 0 seconds.

4. Add or subtract one second from the UTC time scale at the last minute on the specified date.

   ptp utc { **leap59-date** | **leap61-date** } *date*

   By default, no second is added or subtracted from the UTC time.

5. Enable the device to synchronize the UTC time to the TAI time when the device acts as the GM clock.

   ptp refclock-master sync-utc

   By default, the device synchronizes the UTC time to the TAI time when it acts as the GM clock.

Configuring the PTP offset threshold and PTP synchronization suppressions

About this task

Application scenarios

This command is designed to prevent clock source jumps. With this command configured, the device can handle clock jumps effectively to ensure that time synchronization resumes after the clock stabilizes, avoiding errors caused by sudden clock changes.

Operating mechanism

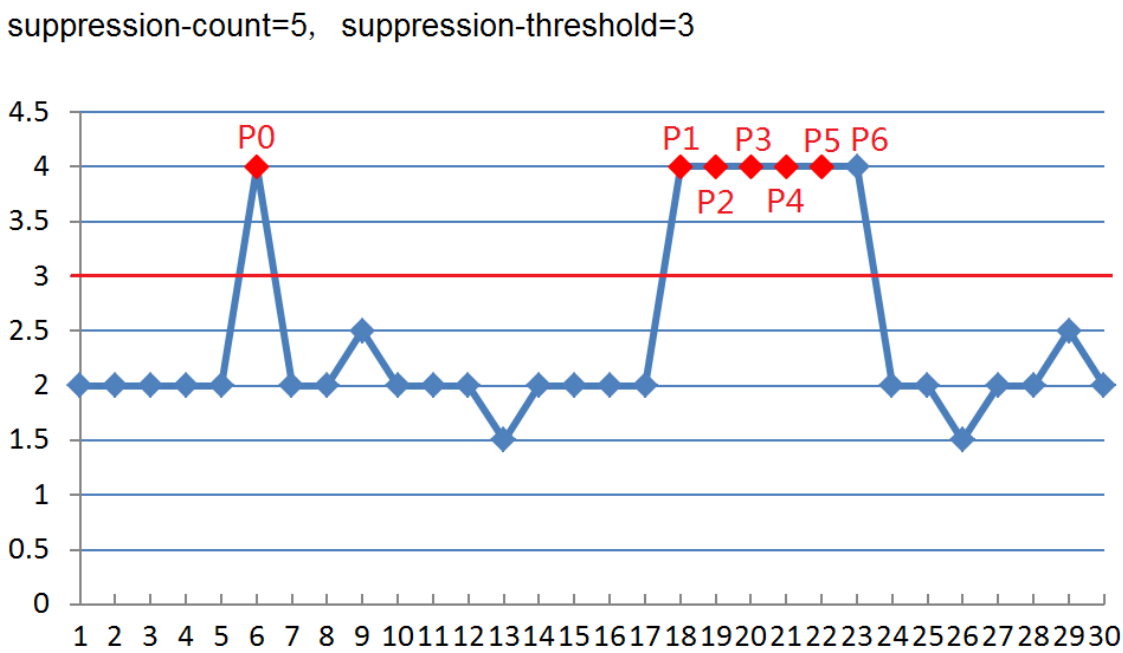
After master-member relationships are established between the clock nodes, the master and member clock nodes exchange PTP messages and record the message transmit and receive time. Based on the timestamps, each member clock calculates the path delay and then gets the time offset between them and the master clock and adjusts their time accordingly for time synchronization with the master clock.

With this command configured, the device determines that a clock jump has occurred and pauses PTP synchronization if the clock is stable but the offset of the current sample exceeds *threshold-value*. The device uses the following rules to determine whether to synchronize the clock in the future:

- **Normal synchronization**—If the offset of the next sample is equal to or smaller than *threshold-value*, PTP resumes time synchronization.
- **Pausing synchronization**—If subsequent offsets remain greater than *threshold-value*, PTP stops time synchronization until the suppression count reaches the preset value. The device then determines that the master clock is stable. When the device receives the next PTP signal, the device synchronizes the time regardless of whether the offset value exceeds *threshold-value*.

As shown in Figure 10, assume that the values for *suppression-count* and *threshold-value* are set to 5 and 3, respectively. When the offset is less than or equal to 3, synchronization occurs. When a sudden clock change causes the offset to exceed 3, synchronization of the clock signal is stopped until with a maximum of *suppression-count* clock signals being paused for synchronization. At time points P0 through P5, the device does not perform time synchronization, because the offset value is greater than 3. At time point P6, the device performs time synchronization not whether the offset value exceeds *threshold-value*.

Figure 10 PTP clock source jump prevention



Procedure

1. Enter system view.

   system-view

2. (Optional.) Enter PTP instance view.

   ptp instance *ptp-instance-id*

   Skip this step for PTP instance 0.

3. Configure the PTP offset threshold and the number of times PTP time synchronization is suppressed.

   ptp time-offset { **suppression-count** *suppression-count* | **suppression-threshold** *threshold-value* } *

By default, the PTP time offset threshold is 3000 nanoseconds and PTP time synchronization is suppressed twice.

Configuring a priority for a clock

About this task

Priorities for clocks are used to elect the GM. The smaller the priority value, the higher the priority.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Enter PTP instance view.

   ptp instance *ptp-instance-id*

   To configure settings for PTP instance 0, skip this step.

3. Configure the priority for the specified clock for GM election through BMC.

   ptp priority clock-source local { **priority1** *priority1* | **priority2** *priority2* }

   The default value varies by PTP profile:

   - **IEEE 1588 version 2, SMPTE ST 2059-2, or AES67-2015**—The priority 1 and priority 2 values are both 128.

Configuring PTP logging

About this task

PTP logs help monitor the clock source status. The following PTP logs are available:

- PTP log that indicates a lower time class

  Each PTP clock source has a class value. For a ToD clock source, you can set its class value by using the **ptp clock-source** command. The higher the value, the lower the class. When the class value of the clock source crosses the class threshold, the system outputs a log for notification.

- PTP log that indicates a higher time offset between the external reference clock and the PTP clock

  If the device uses an external reference clock, it periodically calculates the time offset between the external reference clock and the PTP clock. When the offset exceeds the threshold, the device outputs a log for notification

- PTP logs that indicate a higher time-offset-sum peak-to-peak value

  The PTP module calculates the time-offset-sum peak-to-peak value at specific intervals and compares the value with the threshold configured by this command. If the value is larger than the threshold, the system outputs a log for notification.

- PTP logs that indicate the time-locked or time-unlocked state

  When the time offset of the PTP reference clock crosses the PTP time locking threshold, the PTP time is put into unlocked state. The system outputs a time-unlocked log for notification. When the time offset of the PTP reference clock drops to or below the PTP time locking threshold, the PTP time is put into locked state. The system outputs a time-locked log for notification.

Procedure

1. Enter system view.

system-view

2. Configure the class threshold for the clock source.
   ptp alarm-threshold clock-source-class *class-value*
   By default, the class threshold for the clock source is 6.
   Only the IEEE 1588 version 2 PTP profile supports this configuration.
3. Configure the time-offset threshold between the external reference clock and the PTP clock.
   ptp alarm-threshold time-offset *time-offset-value*
   By default, the time-offset threshold is 500 ns between the external reference clock and the PTP clock.
   Only the IEEE 1588 version 2 PTP profile supports this configuration.
4. Set the time-offset-sum peak-to-peak threshold.
   ptp alarm-threshold time-offset-sum pk-pk *threshold-value*
   By default, the time-offset-sum peak-to-peak threshold is 500 ns.
   Only the IEEE 1588 version 2 PTP profile supports this configuration.
5. Set the PTP time locking and unlocking thresholds.
   ptp alarm-threshold { **time-lock** *lock-value* | **time-unlock** *unlock-value* } *
   By default, the PTP time locking threshold is 200 ns and unlocking threshold is 300 ns.

Enabling SNMP notification for the PTP module

About this task

To get informed of important events of the PTP module, enable SNMP notification for the PTP module.

The notifications generated for the PTP module will be sent to the SNMP module of the device. Configure SNMP notification parameters for the notifications to be output as needed.

For information about SNMP and SNMP notifications, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notification for the PTP module.
   snmp-agent trap enable ptp
   By default, SNMP notification is enabled for the PTP module.

Display and maintenance commands for PTP

Execute **display** commands in any view and the **reset** command in user view.

Task	Command
Display PTP clock information.	**display ptp clock** [**all** | **instance** *ptp-instance-id*]

Task	Command
Display the delay correction history.	**display ptp corrections** [**all** | **instance** *ptp-instance-id*]
Display information about foreign master nodes.	**display ptp foreign-masters-record** [**interface** *interface-type interface-number*] [**all** | **instance** *ptp-instance-id*]
Display PTP information on an interface.	**display ptp interface** [*interface-type interface-number*] [**all** | **instance** *ptp-instance-id*]
Display brief PTP information on all interfaces.	**display ptp interface brief**
Display brief information about the PTP synchronization path from the GM to the device.	**display ptp path-trace**
Display parent node information for the PTP device.	**display ptp parent** [**all** | **instance** *ptp-instance-id*]
Display historical role change information for PTP ports.	**display ptp port-history** [**interface** *interface-type interface-number*] [**all** | **instance** *ptp-instance-id*]
Display PTP statistics.	**display ptp statistics** [**interface** *interface-type interface-number*] [**all** | **instance** *ptp-instance-id*]
Display PTP clock time properties.	**display ptp time-property** [**all** | **instance** *ptp-instance-id*]
Clear PTP statistics.	**reset ptp statistics** [**interface** *interface-type interface-number*] [**all** | **instance** *ptp-instance-id*]

PTP configuration examples

Example: Configuring PTP (IEEE 1588 version 2, IEEE 802.3/Ethernet transport, multicast transmission)

Network configuration

As shown in Figure 11, configure PTP (IEEE 1588 version 2, IEEE 802.3/Ethernet transport, multicast transmission) to enable time synchronization between the devices.

- Specify the IEEE 1588 version 2 PTP profile and the IEEE 802.3/Ethernet transport protocol for PTP messages on Device A, Device B, and Device C.
- Assign Device A, Device B, and Device C to the same PTP domain. Specify the OC clock node type for Device A and Device C, and E2ETC clock node type for Device B. All clock nodes elect a GM through BMC in the PTP domain.
- Use the default Request_Response delay measurement mechanism on Device A and Device C.

Figure 11 Network diagram

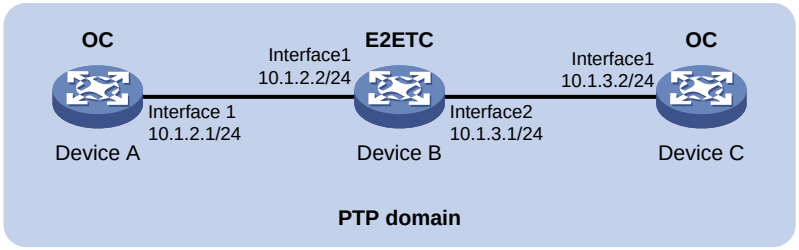


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

   # Specify the IEEE 1588 version 2 PTP profile.

   ```
   <DeviceA> system-view
   [DeviceA] ptp profile 1588v2
   ```

   # Specify the OC clock node type.

   ```
   [DeviceA] ptp mode oc
   ```

   # Create a PTP domain.

   ```
   [DeviceA] ptp domain 0
   ```

   # Enable PTP globally.

   ```
   [DeviceA] ptp global enable
   ```

   # Specify PTP for obtaining the time.

   ```
   [DeviceA] clock protocol ptp
   ```

   # Enable PTP on GigabitEthernet 1/0/1.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ptp enable
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

2. Configure Device B:

   # Specify the IEEE 1588 version 2 PTP profile.

   ```
   <DeviceB> system-view
   [DeviceB] ptp profile 1588v2
   ```

   # Specify the E2ETC clock node type.

   ```
   [DeviceB] ptp mode e2etc
   ```

   # Create a PTP domain.

   ```
   [DeviceB] ptp domain 0
   ```

   # Enable PTP globally.

   ```
   [DeviceB] ptp global enable
   ```

   # Specify PTP for obtaining the time.

   ```
   [DeviceB] clock protocol ptp
   ```

   # Enable PTP on GigabitEthernet 1/0/1.

   ```
   [DeviceB] interface gigabitethernet 1/0/1
   ```

```
[DeviceB-GigabitEthernet1/0/1] ptp enable
[DeviceB-GigabitEthernet1/0/1] quit
```

Enable PTP on GigabitEthernet 1/0/2.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ptp enable
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device C:

Specify the IEEE 1588 version 2 PTP profile.

```
<DeviceC> system-view
[DeviceC] ptp profile 1588v2
```

Specify the OC clock node type.

```
[DeviceC] ptp mode oc
```

Create a PTP domain.

```
[DeviceC] ptp domain 0
```

Enable PTP globally.

```
[DeviceC] ptp global enable
```

Specify PTP for obtaining the time.

```
[DeviceC] clock protocol ptp
```

Enable PTP on GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] ptp enable
[DeviceC-GigabitEthernet1/0/1] quit
```

Verifying the configuration

When the network is stable, perform the following tasks:

- Use the **`display ptp clock`** command to display PTP clock information.
- Use the **`display ptp interface brief`** command to display brief PTP running information for all PTP interfaces.

Display PTP clock information on Device A.

```
[DeviceA] display ptp clock
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : OC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0000
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 1
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                : 248
 Accuracy             : 254
 Offset (log variance) : 65535
Offset from master : 0 (ns)
```

```
Mean path delay    : 0 (ns)
Steps removed      : 0
Local clock time   : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device A.

```
[DeviceA] display ptp interface brief
Name       InstID    State         Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    0         Master        E2E              Two         0
```

Display PTP clock information on Device B.

```
[DeviceB] display ptp clock
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : E2ETC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0001
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 2
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                 : 248
 Accuracy              : 254
 Offset (log variance) : 65535
Offset from master  : N/A
Mean path delay     : N/A
Steps removed       : N/A
Local clock time    : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device B.

```
[DeviceB] display ptp interface brief
Name       InstID    State         Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    0         N/A           E2E              Two         0
GE1/0/2    0         N/A           E2E              Two         0
```

The command outputs show that Device A is elected as the GM and GigabitEthernet1/0/1 on Device A is the master port.

Example: Configuring PTP (IEEE 1588 version 2, UDP transport, multicast transmission)

Network configuration

As shown in Figure 12, configure PTP (IEEE 1588 version 2, UDP transport, multicast transmission) to enable time synchronization between the devices.

- Specify the IEEE 1588 version 2 PTP profile for Device A, Device B, and Device C.
- Specify multicast UDP transport for PTP messages.

- Assign Device A, Device B, and Device C to the same PTP domain. Specify the OC clock node type for Device A and Device C, and P2PTC clock node type for Device B. All clock nodes elect a GM through BMC in the PTP domain.
- Specify the peer delay measurement mechanism (p2p) for Device A and Device C.

Figure 12 Network diagram

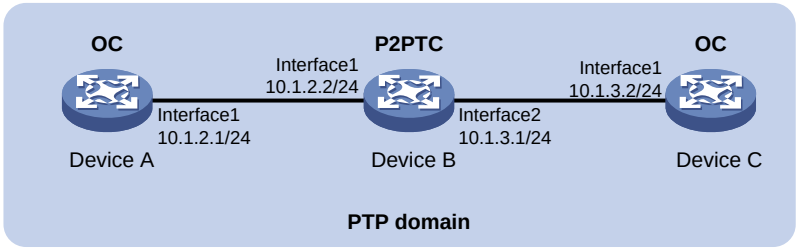


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Configure Device A:

   # Specify the IEEE 1588 version 2 PTP profile.

   ```
   <DeviceA> system-view
   [DeviceA] ptp profile 1588v2
   ```

   # Specify the OC clock node type.

   ```
   [DeviceA] ptp mode oc
   ```

   # Create a PTP domain.

   ```
   [DeviceA] ptp domain 0
   ```

   # Enable PTP globally.

   ```
   [DeviceA] ptp global enable
   ```

   # Configure the source IP address for multicast PTP messages transmitted over UDP.

   ```
   [DeviceA] ptp source 10.10.1.1
   ```

   # Specify PTP for obtaining the time.

   ```
   [DeviceA] clock protocol ptp
   ```

   # On GigabitEthernet 1/0/1, specify the UDP transport protocol for PTP messages, specify the delay measurement mechanism as **p2p**, and enable PTP.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ptp transport-protocol udp
   [DeviceA-GigabitEthernet1/0/1] ptp delay-mechanism p2p
   [DeviceA-GigabitEthernet1/0/1] ptp enable
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

2. Configure Device B:

   # Specify the IEEE 1588 version 2 PTP profile.

   ```
   <DeviceB> system-view
   [DeviceB] ptp profile 1588v2
   ```

   # Specify the P2PTC clock node type.

   ```
   [DeviceB] ptp mode p2ptc
   ```

Create a PTP domain.

```
[DeviceB] ptp domain 0
```

Enable PTP globally.

```
[DeviceB] ptp global enable
```

Configure the source IP address for multicast PTP messages transmitted over UDP.

```
[DeviceB] ptp source 10.10.2.1
```

Specify PTP for obtaining the time.

```
[DeviceB] clock protocol ptp
```

On GigabitEthernet 1/0/1, specify the UDP transport protocol for PTP messages and enable PTP.

```
[DeviceB] interface gigabitethernet 1/0/1
DeviceB-GigabitEthernet1/0/1] ptp transport-protocol udp
[DeviceB-GigabitEthernet1/0/1] ptp enable
[DeviceB-GigabitEthernet1/0/1] quit
```

On GigabitEthernet 1/0/2, specify the UDP transport protocol for PTP messages and enable PTP.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ptp transport-protocol udp
[DeviceB-GigabitEthernet1/0/2] ptp enable
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device C:

Specify the IEEE 1588 version 2 PTP profile.

```
<DeviceC> system-view
[DeviceC] ptp profile 1588v2
```

Specify the OC clock node type.

```
[DeviceC] ptp mode oc
```

Create a PTP domain.

```
[DeviceC] ptp domain 0
```

Enable PTP globally.

```
[DeviceC] ptp global enable
```

Configure the source IP address for multicast PTP messages transmitted over UDP.

```
[DeviceC] ptp source 10.10.3.1
```

Specify PTP for obtaining the time.

```
[DeviceC] clock protocol ptp
```

On GigabitEthernet 1/0/1, specify the UDP transport protocol for PTP messages, specify the delay measurement mechanism as **p2p**, and enable PTP.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] ptp transport-protocol udp
[DeviceC-GigabitEthernet1/0/1] ptp delay-mechanism p2p
[DeviceC-GigabitEthernet1/0/1] ptp enable
[DeviceC-GigabitEthernet1/0/1] quit
```

Verifying the configuration

When the network is stable, perform the following tasks:

- Use the **display ptp clock** command to display PTP clock information.
- Use the **display ptp interface brief** command to display brief PTP running information for all PTP interfaces.

Display PTP clock information on Device A.

```
[DeviceA] display ptp clock
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : OC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0000
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 1
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                 : 248
 Accuracy              : 254
 Offset (log variance) : 65535
Offset from master : 0 (ns)
Mean path delay    : 0 (ns)
Steps removed      : 0
Local clock time   : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device A.

```
[DeviceA] display ptp interface brief
Name       InstID     State         Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    0         Master        P2P              Two         0
```

Display PTP clock information on Device B.

```
[DeviceB] display ptp clock
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : P2PTC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0001
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 2
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                 : 248
 Accuracy              : 254
 Offset (log variance) : 65535
Offset from master  : N/A
Mean path delay     : N/A
Steps removed       : N/A
Local clock time    : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device B.

```
[DeviceB] display ptp interface brief
Name       InstID     State         Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    0          N/A           P2P              Two         0
GE1/0/2    0          N/A           P2P              Two         0
```

The command outputs show that Device A is elected as the GM and GigabitEthernet1/0/1 on Device A is the master port.

Example: Configuring multiple PTP instances (IEEE 1588 version 2, multicast transmission)

Network configuration

As shown in Figure 13, configure two PTP instances on a PTP network.

- PTP instance 1: Assign Device A, Device B, and Device C to PTP instance 1.
  - Specify the IEEE 1588 version 2 PTP profile and IEEE 802.3/Ethernet transport of PTP messages for the devices and assign the devices to PTP domain 1.
  - Specify the OC clock node type for Device A and Device C and E2ETC clock node type for Device B. All these clock nodes elect a GM through BMC.
  - Use the default Request_Response delay measurement mechanism on all clock nodes in PTP instance 1.
- PTP instance 2: Assign Device D, Device B, and Device E to PTP instance 2.
  - Specify the IEEE 1588 version 2 PTP profile and UDP transport of PTP messages for the devices and assign the devices to PTP domain 2.
  - Specify the OC clock node type for Device D and Device E and P2PTC clock node type for Device B. All the clock nodes elect a GM through BMC.
  - Use the peer delay measurement mechanism on Device D and Device E.

Figure 13 Network diagram

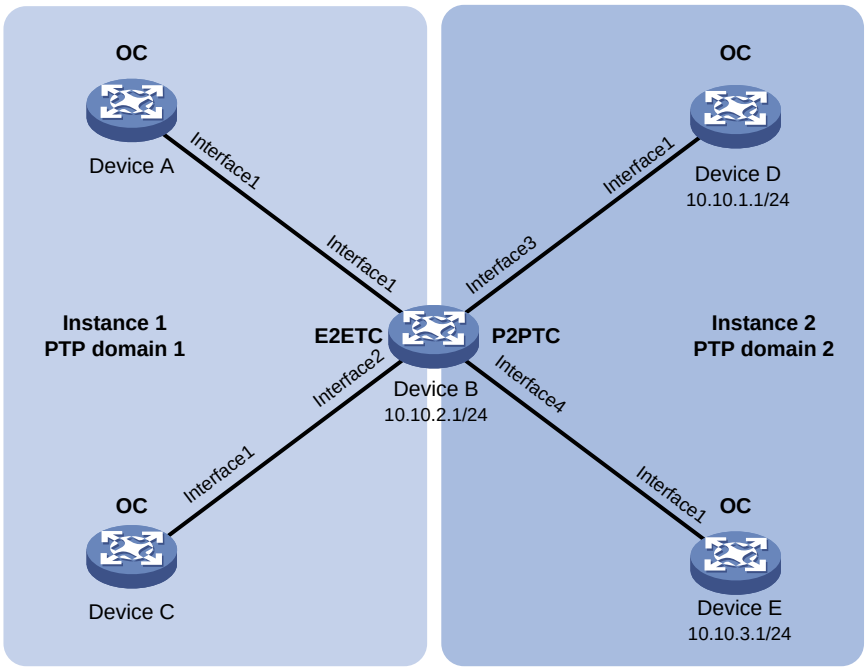


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

1. Configure PTP instance 1 (IEEE 802.3/Ethernet transport of PTP messages)
   a. Configure Device A:

      # Enable PTP globally.

      ```
      <DeviceA> system-view
      [DeviceA] ptp global enable
      ```

      # Create a PTP instance, specifying its ID as **1** and name as **ptp1**, and enter its view.

      ```
      [DeviceA] ptp instance 1 name ptp1
      ```

      # Specify the IEEE 1588 version 2 PTP profile.

      ```
      [DeviceA-ptp-instance-1] ptp profile 1588v2
      ```

      # Specify the OC clock node type.

      ```
      [DeviceA-ptp-instance-1] ptp mode oc
      ```

      # Specify PTP domain 1.

      ```
      [DeviceA-ptp-instance-1] ptp domain 1
      [DeviceA-ptp-instance-1] quit
      ```

      # Specify PTP for obtaining the time.

      ```
      [DeviceA] clock protocol ptp
      ```

      # Assign GigabitEthernet 1/0/1 to PTP instance 1, enter interface PTP instance view, and enable PTP.

      ```
      [DeviceA] interface gigabitethernet 1/0/1
      [DeviceA-GigabitEthernet1/0/1] ptp instance 1
      [DeviceA-GigabitEthernet1/0/1-ptp-instance-1] ptp enable
      [DeviceA-GigabitEthernet1/0/1-ptp-instance-1] quit
      [DeviceA-GigabitEthernet1/0/1] quit
      ```

   b. Configure Device B:

      # Enable PTP globally.

      ```
      <DeviceB> system-view
      [DeviceB] ptp global enable
      ```

      # Create a PTP instance, specifying its ID as **1** and name as **ptp1**, and enter its view.

      ```
      [DeviceB] ptp instance 1 name ptp1
      ```

      # Specify the IEEE 1588 version 2 PTP profile.

      ```
      [DeviceB-ptp-instance-1] ptp profile 1588v2
      ```

      # Specify the E2ETC clock node type.

      ```
      [DeviceB-ptp-instance-1] ptp mode e2etc
      ```

      # Specify PTP domain 1.

      ```
      [DeviceB-ptp-instance-1] ptp domain 1
      [DeviceB-ptp-instance-1] quit
      ```

      # Specify PTP for obtaining the time.

```
[DeviceB] clock protocol ptp
```

Assign GigabitEthernet 1/0/1 to PTP instance 1, enter interface PTP instance view, and enable PTP.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] ptp instance 1
[DeviceB-GigabitEthernet1/0/1-ptp-instance-1] ptp enable
[DeviceB-GigabitEthernet1/0/1-ptp-instance-1] quit
[DeviceB-GigabitEthernet1/0/1] quit
```

Assign GigabitEthernet 1/0/2 to PTP instance 1, enter interface PTP instance view, and enable PTP.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ptp instance 1
[DeviceB-GigabitEthernet1/0/2-ptp-instance-1] ptp enable
[DeviceB-GigabitEthernet1/0/2-ptp-instance-1] quit
[DeviceB-GigabitEthernet1/0/2] quit
```

c. Configure Device C:

Enable PTP globally.

```
<DeviceC> system-view
[DeviceC] ptp global enable
```

Create a PTP instance, specifying its ID as **1** and name as **ptp1**, and enter its view.

```
[DeviceC] ptp instance 1 name ptp1
[DeviceC-ptp-instance-1]
```

Specify the IEEE 1588 version 2 PTP profile.

```
[DeviceC-ptp-instance-1] ptp profile 1588v2
```

Specify the OC clock node type.

```
[DeviceC-ptp-instance-1] ptp mode oc
```

Specify PTP domain 1.

```
[DeviceC-ptp-instance-1] ptp domain 1
[DeviceC-ptp-instance-1] quit
```

Specify PTP for obtaining the time.

```
[DeviceC] clock protocol ptp
```

Assign GigabitEthernet 1/0/1 to PTP instance 1, enter interface PTP instance view, and enable PTP.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] ptp instance 1
[DeviceC-GigabitEthernet1/0/1-ptp-instance-1] ptp enable
[DeviceC-GigabitEthernet1/0/1-ptp-instance-1] quit
[DeviceC-GigabitEthernet1/0/1] quit
```

2. Configure PTP instance 2 (UDP transport of PTP messages)

a. Configure Device D:

Enable PTP globally.

```
<DeviceD> system-view
[DeviceD] ptp global enable
```

Create a PTP instance, specifying its ID as **2** and name as **ptp2**, and enter its view.

```
[DeviceD] ptp instance 2 name ptp2
```

Specify the IEEE 1588 version 2 PTP profile.

```
[DeviceD-ptp-instance-2] ptp profile 1588v2
```

Specify the OC clock node type.

```
[DeviceD-ptp-instance-1] ptp mode oc
```

Specify PTP domain 2.

```
[DeviceD-ptp-instance-2] ptp domain 2
```

Specify the source IP address for multicast transmission of PTP messages over UDP.

```
[DeviceD-ptp-instance-2] ptp source 10.10.1.1
[DeviceD-ptp-instance-2] quit
```

Specify PTP for obtaining the time.

```
[DeviceD] clock protocol ptp
```

Assign GigabitEthernet1/0/1 to PTP instance 2 and enter interface PTP instance view. Specify UDP transport of PTP messages and the peer delay measurement mechanism and enable PTP.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] ptp instance 2
[DeviceD-GigabitEthernet1/0/1-ptp-instance-2] ptp transport-protocol udp
[DeviceD-GigabitEthernet1/0/1-ptp-instance-2] ptp delay-mechanism p2p
[DeviceD-GigabitEthernet1/0/1-ptp-instance-2] ptp enable
[DeviceD-GigabitEthernet1/0/1-ptp-instance-2] quit
[DeviceD-GigabitEthernet1/0/1] quit
```

b. Configure Device B:

Create a PTP instance, specifying its ID as **2** and name as **ptp2**, and enter its view.

```
<DeviceB> system-view
[DeviceB] ptp instance 2 name ptp2
```

Specify the IEEE 1588 version 2 PTP profile.

```
[DeviceB-ptp-instance-2] ptp profile 1588v2
```

Specify the P2PTC clock node type.

```
[DeviceB-ptp-instance-2] ptp mode p2ptc
```

Specify PTP domain 2.

```
[DeviceB-ptp-instance-2] ptp domain 2
```

Specify the source IP address for multicast transmission of PTP messages over UDP.

```
[DeviceB-ptp-instance-2] ptp source 10.10.2.1
[DeviceB-ptp-instance-2] quit
```

Specify PTP for obtaining the time.

```
[DeviceB] clock protocol ptp
[DeviceB] clock protocol ptp mdc 1
```

Assign GigabitEthernet 1/0/3 to PTP instance and enter interface PTP instance view. Specify UDP transport of PTP messages and enable PTP.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] ptp instance 2
[DeviceB-GigabitEthernet1/0/3-ptp-instance-2] ptp transport-protocol udp
[DeviceB-GigabitEthernet1/0/3-ptp-instance-2] ptp enable
[DeviceB-GigabitEthernet1/0/3-ptp-instance-2] quit
[DeviceB-GigabitEthernet1/0/3] quit
```

Assign GigabitEthernet 1/0/4 to PTP instance and enter interface PTP instance view. Specify UDP transport of PTP messages and enable PTP.

```
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] ptp instance 2
[DeviceB-GigabitEthernet1/0/4-ptp-instance-2] ptp transport-protocol udp
```

```
[DeviceB-GigabitEthernet1/0/4-ptp-instance-2] ptp enable
[DeviceB-GigabitEthernet1/0/4-ptp-instance-2] quit
[DeviceB-GigabitEthernet1/0/4] quit
```

c. Configure Device E:

Enable PTP globally.

```
<DeviceE> system-view
[DeviceE] ptp global enable
```

Create a PTP instance, specifying its ID as **2** and name as **ptp2**, and enter its view.

```
[DeviceE] ptp instance 2 name ptp2
```

Specify the IEEE 1588 version 2 PTP profile.

```
[DeviceE-ptp-instance-2] ptp profile 1588v2
```

Specify the OC clock node type.

```
[DeviceE-ptp-instance-2] ptp mode oc
```

Specify PTP domain 2.

```
[DeviceE-ptp-instance-2] ptp domain 2
```

Specify the source IP address for multicast transmission of PTP messages over UDP.

```
[DeviceE-ptp-instance-2] ptp source 10.10.3.1
[DeviceE-ptp-instance-2] quit
```

Specify PTP for obtaining the time.

```
[DeviceE] clock protocol ptp
```

Assign GigabitEthernet 1/0/4 to PTP instance 2 and enter interface PTP instance view. Specify UDP transport of PTP messages and the peer delay measurement mechanism and enable PTP.

```
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] ptp instance 2
[DeviceE-GigabitEthernet1/0/1-ptp-instance-2] ptp transport-protocol udp
[DeviceE-GigabitEthernet1/0/1-ptp-instance-2] ptp delay-mechanism p2p
[DeviceE-GigabitEthernet1/0/1-ptp-instance-2] ptp enable
[DeviceE-GigabitEthernet1/0/1-ptp-instance-2] quit
[DeviceE-GigabitEthernet1/0/1] quit
```

Verifying the configuration

1. Verify the configuration of PTP instance 1.

When the network is stable, perform the following tasks to verify the configuration of PTP instance 1.

- Use the **`display ptp clock`** command to display PTP clock information.
- Use the **`display ptp interface brief`** command to display brief PTP running information for all PTP interfaces.

Display PTP clock information for PTP instance 1 on Device A.

```
[DeviceA] display ptp clock instance 1
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : OC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0000
Clock type          : Local
```

```
Clock domain        : 1
Number of PTP ports : 1
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                : 248
 Accuracy             : 254
 Offset (log variance) : 65535
Offset from master : 0 (ns)
Mean path delay    : 0 (ns)
Steps removed      : 0
Local clock time   : Tue Jul 23 09:41:50 2019
```

Display brief PTP running information for all PTP interfaces on Device A.

```
[DeviceA] display ptp interface brief
Name       InstID   State        Delay mechanism  Clock step  Asymmetry correction
GE1/0/1     1        Master        E2E              Two         0
```

Display PTP clock information for PTP instance 1 on Device B.

```
[DeviceB] display ptp clock instance 1
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : E2ETC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0001
Clock type          : Local
Clock domain        : 1
Number of PTP ports : 2
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                : 248
 Accuracy             : 254
 Offset (log variance) : 65535
Offset from master : N/A
Mean path delay    : N/A
Steps removed      : N/A
Local clock time   : Tue Jul 23 09:41:50 2019
```

Display brief PTP running information for all PTP interfaces on Device B.

```
[DeviceB] display ptp interface brief
Name      InstID    State        Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    1         N/A           E2E              Two         0
GE1/0/2    1         N/A           E2E              Two         0
GE1/0/3    2         N/A           P2P              Two         0
GE1/0/4    2         N/A           P2P              Two         0
```

The command output shows that Device A is elected as the GM, and GigabitEthernet1/0/1 on Device A sends time synchronization information to its downstreams as a master port.

2. Verify the configuration of PTP instance 2.

When the network is stable, perform the following tasks to verify the configuration of PTP instance 2:

- Use the **display ptp clock** command to display PTP clock information.
- Use the **display ptp interface brief** command to display brief PTP running information for all PTP interfaces.

Display PTP clock information for PTP instance 2 on Device D.

```
[DeviceD] display ptp clock instance 2
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : OC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0000
Clock type          : Local
Clock domain        : 2
Number of PTP ports : 1
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                : 248
 Accuracy             : 254
 Offset (log variance) : 65535
Offset from master : 0 (ns)
Mean path delay    : 0 (ns)
Steps removed      : 0
Local clock time   : Tue Jul 23 09:41:50 2019
```

Display brief PTP running information for all PTP interfaces on Device D.

```
[DeviceD] display ptp interface brief
Name      InstID    State        Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    2          Master        P2P              Two         0
```

Display PTP clock information for PTP instance 2 on Device B.

```
[DeviceB] display ptp clock instance 2
PTP global state    : Enabled
PTP profile         : IEEE 1588 Version 2
PTP mode            : P2PTC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0001
Clock type          : Local
Clock domain        : 2
Number of PTP ports : 2
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                : 248
 Accuracy             : 254
```

```
 Offset (log variance) : 65535
Offset from master : N/A
Mean path delay    : N/A
Steps removed      : N/A
Local clock time   : Tue Jul 23 09:41:50 2019
```

Display brief PTP running information for all PTP interfaces on Device B.

```
[DeviceB] display ptp interface brief
Name      InstID    State       Delay mechanism  Clock step  Asymmetry
correction
GE1/0/1     1           N/A           E2E               Two          0
GE1/0/2     1           N/A           E2E               Two          0
GE1/0/3     2           N/A           P2P               Two          0
GE1/0/4     2           N/A           P2P               Two          0
```

The command output shows that Device D is elected as the GM, and GigabitEthernet1/0/1 on Device D sends time synchronization information to its downstreams as a master port.

Example: Configuring PTP (SMPTE ST 2059-2, UDP transport, multicast transmission)

Network configuration

As shown in Figure 14, configure PTP (SMPTE ST 2059-2, UDP transport, multicast transmission) to enable time synchronization between Device A and Device C:

- Specify the SMPTE ST 2059-2 PTP profile and multicast UDP transport of PTP messages for Device A, Device B, and Device C.
- Specify the OC clock node type for Device A and Device C, and the P2PTC clock node type for Device B. All clock nodes elect a GM through BMC.
- Use the peer delay measurement mechanism on all clock nodes in the PTP domain.

Figure 14 Network diagram

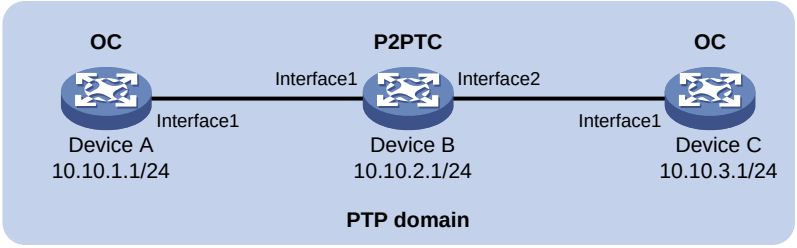


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

IMPORTANT:

The SMPTE ST 2059-2 PTP profile transports PTP messages over UDP rather than IEEE 802.3/Ethernet. The profile supports both multicast and unicast transmission modes.

1. Configure Device A:

   # Specify the SMPTE ST 2059-2 PTP profile.

   ```
   <DeviceA> system-view
   [DeviceA] ptp profile st2059-2
   ```

   # Specify the OC clock node type.

   ```
   [DeviceA] ptp mode oc
   ```

   # Create a PTP domain.

   ```
   [DeviceA] ptp domain 0
   ```

   # Enable PTP globally.

   ```
   [DeviceB] ptp global enable
   ```

   # Configure the source IP address for multicast PTP messages transmitted over UDP.

   ```
   [DeviceA] ptp source 10.10.1.1
   ```

   # Specify PTP for obtaining the time.

   ```
   [DeviceA] clock protocol ptp
   ```

   # On GigabitEthernet 1/0/1, specify the peer delay measurement mechanism and enable PTP.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ptp transport-protocol udp
   [DeviceA-GigabitEthernet1/0/1] ptp delay-mechanism p2p
   [DeviceA-GigabitEthernet1/0/1] ptp enable
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

2. Configure Device B:

   # Specify the SMPTE ST 2059-2 PTP profile.

   ```
   <DeviceB> system-view
   [DeviceB] ptp profile st2059-2
   ```

   # Specify the P2PTC clock node type.

   ```
   [DeviceB] ptp mode p2ptc
   ```

   # Create a PTP domain.

   ```
   [DeviceB] ptp domain 0
   ```

   # Enable PTP globally.

   ```
   [DeviceB] ptp global enable
   ```

   # Configure the source IP address for multicast PTP messages transmitted over UDP.

   ```
   [DeviceB] ptp source 10.10.2.1
   ```

   # Specify PTP for obtaining the time.

   ```
   [DeviceB] clock protocol ptp
   ```

   # Enable PTP on GigabitEthernet 1/0/1.

   ```
   [DeviceB] interface gigabitethernet 1/0/1
   DeviceB-GigabitEthernet1/0/1] ptp transport-protocol udp
   [DeviceB-GigabitEthernet1/0/1] ptp enable
   [DeviceB-GigabitEthernet1/0/1] quit
   ```

   # Enable PTP on GigabitEthernet 1/0/2.

   ```
   [DeviceB] interface gigabitethernet 1/0/2
   [DeviceB-GigabitEthernet1/0/2] ptp transport-protocol udp
   [DeviceB-GigabitEthernet1/0/2] ptp enable
   [DeviceB-GigabitEthernet1/0/2] quit
   ```

3. Configure Device C:

   # Specify the SMPTE ST 2059-2 PTP profile.

   ```
   <DeviceC> system-view
   ```

```
[DeviceC] ptp profile st2059-2
```

Specify the OC clock node type.

```
[DeviceC] ptp mode oc
```

Create a PTP domain.

```
[DeviceC] ptp domain 0
```

Enable PTP globally.

```
[DeviceC] ptp global enable
```

Configure the source IP address for multicast PTP messages transmitted over UDP.

```
[DeviceC] ptp source 10.10.3.1
```

Specify PTP for obtaining the time.

```
[DeviceC] clock protocol ptp
```

On GigabitEthernet 1/0/1, specify the peer delay measurement mechanism and enable PTP.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] ptp delay-mechanism p2p
[DeviceC-GigabitEthernet1/0/1] ptp enable
[DeviceC-GigabitEthernet1/0/1] quit
```

Verifying the configuration

When the network is stable, perform the following tasks to verify the PTP configuration:

- Use the **`display ptp clock`** command to display PTP clock information.
- Use the **`display ptp interface brief`** command to display brief PTP running information for all PTP interfaces.

Display PTP clock information on Device A.

```
[DeviceA] display ptp clock
PTP global state    : Enabled
PTP profile         : SMPTE ST 2059-2
PTP mode            : OC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0000
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 1
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                 : 248
 Accuracy              : 254
 Offset (log variance) : 65535
Offset from master  : 106368539000 (ns)
Mean path delay     : 2791000 (ns)
Steps removed       : 1
Local clock time    : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device A.

```
[DeviceA] display ptp interface brief
Name       InstID     State         Delay mechanism  Clock step  Asymmetry correction
```

```
GE1/0/1    0          Master        P2P             Two         0
```

Display PTP clock information on Device B.

```
[DeviceB] display ptp clock
PTP global state    : Enabled
PTP profile         : SMPTE ST 2059-2
PTP mode            : P2PTC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0001
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 2
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                 : 248
 Accuracy              : 254
 Offset (log variance) : 65535
Offset from master  : N/A
Mean path delay     : N/A
Steps removed       : N/A
Local clock time    : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device B.

```
[DeviceB] display ptp interface brief
Name       InstID     State         Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    0          N/A           P2P              Two         0
GE1/0/2    0          N/A           P2P              Two         0
```

The output shows that Device A is elected as the GM and GigabitEthernet1/0/1 on Device A sends time synchronization information to its downstreams as a master port.

Example: Configuring PTP (AES67-2015, UDP transport, multicast transmission)

Network configuration

As shown in Figure 15, configure PTP (AES67-2015, UDP transport, multicast transmission) to enable time synchronization between Device A and Device C:

- Specify the AES67-2015 PTP profile and multicast UDP transport of PTP messages for Device A, Device B, and Device C.
- Assign Device A, Device B, and Device C to the same PTP domain. Specify the OC clock node type for Device A and Device C, and the P2PTC clock node type for Device B. All clock nodes elect a GM through BMC.
- Use the peer delay measurement mechanism on all clock nodes in the PTP domain.

Figure 15 Network diagram

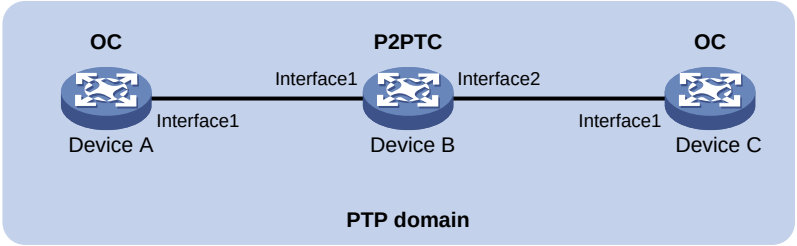


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

IMPORTANT:

The AES67-2015 PTP profile transports PTP messages over UDP rather than IEEE 802.3/Ethernet. The profile supports both multicast and unicast transmission modes.

1. Configure Device A:

   # Specify the AES67-2015 PTP profile.

   ```
   <DeviceA> system-view
   [DeviceA] ptp profile aes67-2015
   ```

   # Specify the OC clock node type.

   ```
   [DeviceA] ptp mode oc
   ```

   # Create a PTP domain.

   ```
   [DeviceA] ptp domain 0
   ```

   # Enable PTP globally.

   ```
   [DeviceB] ptp global enable
   ```

   # Configure the source IP address for multicast PTP messages transmitted over UDP.

   ```
   [DeviceA] ptp source 10.10.1.1
   ```

   # Specify PTP for obtaining the time.

   ```
   [DeviceA] clock protocol ptp
   ```

   # On GigabitEthernet 1/0/1, specify the peer delay measurement mechanism and enable PTP.

   ```
   [DeviceA] interface gigabitethernet 1/0/1
   [DeviceA-GigabitEthernet1/0/1] ptp transport-protocol udp
   [DeviceA-GigabitEthernet1/0/1] ptp delay-mechanism p2p
   [DeviceA-GigabitEthernet1/0/1] ptp enable
   [DeviceA-GigabitEthernet1/0/1] quit
   ```

2. Configure Device B:

   # Specify the AES67-2015 PTP profile.

   ```
   <DeviceB> system-view
   [DeviceB] ptp profile aes67-2015
   ```

   # Specify the P2PTC clock node type.

   ```
   [DeviceB] ptp mode p2ptc
   ```

   # Create a PTP domain.

```
[DeviceB] ptp domain 0
```

Enable PTP globally.

```
[DeviceB] ptp global enable
```

Configure the source IP address for multicast PTP messages transmitted over UDP.

```
[DeviceB] ptp source 10.10.2.1
```

Specify PTP for obtaining the time.

```
[DeviceB] clock protocol ptp
```

Enable PTP on GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
DeviceB-GigabitEthernet1/0/1] ptp transport-protocol udp
[DeviceB-GigabitEthernet1/0/1] ptp enable
[DeviceB-GigabitEthernet1/0/1] quit
```

Enable PTP on GigabitEthernet 1/0/2.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] ptp transport-protocol udp
[DeviceB-GigabitEthernet1/0/2] ptp enable
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device C:

Specify the AES67-2015 PTP profile.

```
<DeviceC> system-view
[DeviceC] ptp profile aes67-2015
```

Specify the OC clock node type.

```
[DeviceC] ptp mode oc
```

Create a PTP domain.

```
[DeviceC] ptp domain 0
```

Enable PTP globally.

```
[DeviceC] ptp global enable
```

Configure the source IP address for multicast PTP messages transmitted over UDP.

```
[DeviceC] ptp source 10.10.3.1
```

Specify PTP for obtaining the time.

```
[DeviceC] clock protocol ptp
```

On GigabitEthernet 1/0/1, specify the peer delay measurement mechanism and enable PTP.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] ptp delay-mechanism p2p
[DeviceC-GigabitEthernet1/0/1] ptp enable
[DeviceC-GigabitEthernet1/0/1] quit
```

Verifying the configuration

When the network is stable, perform the following tasks to verify the PTP configuration:

- Use the **`display ptp clock`** command to display PTP clock information.
- Use the **`display ptp interface brief`** command to display brief PTP running information for all PTP interfaces.

Display PTP clock information on Device A.

```
[DeviceA] display ptp clock
PTP global state    : Enabled
PTP profile         : AES67-2015
PTP mode            : OC
```

```
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0000
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 1
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                 : 248
 Accuracy              : 254
 Offset (log variance) : 65535
Offset from master  : 106368539000 (ns)
Mean path delay     : 2791000 (ns)
Steps removed       : 1
Local clock time    : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device A.

```
[DeviceA] display ptp interface brief
Name       InstID     State         Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    0          Master        P2P              Two         0
```

Display PTP clock information on Device B.

```
[DeviceB] display ptp clock
PTP global state    : Enabled
PTP profile         : AES67-2015
PTP mode            : P2PTC
PTP whitelist state : Disabled
Slave only          : No
Lock status         : Unlocked
Clock ID            : 000FE2-FFFE-FF0001
Clock type          : Local
Clock domain        : 0
Number of PTP ports : 2
Priority1     : 128
Priority2     : 128
Clock quality :
 Class                 : 248
 Accuracy              : 254
 Offset (log variance) : 65535
Offset from master  : N/A
Mean path delay     : N/A
Steps removed       : N/A
Local clock time   : Sun Jan 15 20:57:29 2019
```

Display brief PTP running information for all PTP interfaces on Device B.

```
[DeviceB] display ptp interface brief
Name       InstID     State         Delay mechanism  Clock step  Asymmetry correction
GE1/0/1    0          N/A           P2P              Two         0
GE1/0/2    0          N/A           P2P              Two         0
```

The output shows that Device A is elected as the GM and GigabitEthernet1/0/1 on Device A sends time synchronization information to its downstreams as a master port.

Contents

Configuring PoE 1

About PoE 1
PoE system 1
Protocols and standards 1
Restrictions and guidelines: PoE configuration 2
PoE configuration tasks at a glance 2
Prerequisites for configuring PoE 3
Enabling PoE 3
Enabling PoE for a PI 3
Enabling fast PoE for a PSE 4
Configuring PoE delay 4
Disabling PoE on shutdown interfaces 5
Allowing inrush currents drawn by PDs 5
Enabling PI power cycling upon a system warm reboot 6
Configuring PD detection 6
Enabling nonstandard PD detection 6
Configuring a PD detection mode 6
Configuring a PD disconnection detection mode 7
Configuring the PoE power 8
Configuring the maximum PI power 8
Configuring the PoE priority policy 8
Configuring the PI priority policy 8
Configuring PoE monitoring 9
Configuring PSE power monitoring 9
Configuring a PI by using a PoE profile 9
Restrictions and guidelines 9
Configuring a PoE profile 10
Applying a PoE profile 10
Upgrading PSE firmware in service 11
Enabling forced PoE power supply 11
Display and maintenance commands for PoE 12
PoE configuration examples 12
Example: Configuring PoE 12
Troubleshooting PoE 14
Failure to set the priority of a PI to critical 14
Failure to apply a PoE profile to a PI 14

Configuring PoE

About PoE

Power over Ethernet (PoE) enables a network device to supply power to terminals over twisted pair cables.

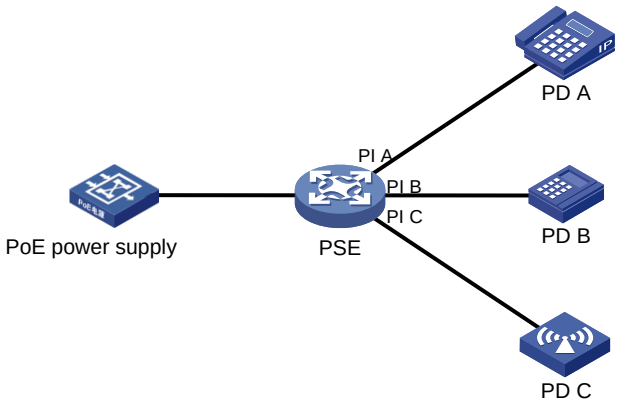
PoE system

As shown in Figure 1, a PoE system includes the following elements:

- **PoE power supply**—A PoE power supply provides power for the entire PoE system.
- **PSE**—A power sourcing equipment (PSE) supplies power to PDs. PSE devices are classified into single-PSE devices and multiple-PSE devices.
  - A single-PSE device has only one piece of PSE firmware.
  - A multiple-PSE device has multiple PSEs. A multiple-PSE device uses PSE IDs to identify different PSEs. To view the mapping between the ID and slot number of a PSE, execute the **`display poe device`** command.

  The device is a multiple-PSE device.
- **PI**—A power interface (PI) is a PoE-capable Ethernet interface on a PSE.
- **PD**—A powered device (PD) receives power from a PSE. PDs include IP telephones, APs, portable chargers, POS terminals, and Web cameras. You can also connect a PD to a redundant power source for reliability.

Figure 1 PoE system diagram



Protocols and standards

802.3af, 802.3at, and 802.3bt are the 3rd, 4th, and 5th editions of the PoE protocol established by the IEEE. The differences between the three PoE technologies lie in their power transmission capabilities and the number of wires used for power transmission.

- 802.3af, also known as the PoE standard, was released in 2003. It provides a maximum power output of 15.4 W and uses 2 pairs of wires for power transmission.

  The name of the 802.3af standard is *802.3af-2003, IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 3: Carrier Sense Multiple Access with Collision*

Detection (CSMA/CD) Access Method and Physical Layer Specifications - Data Terminal Equipment (DTE) Power Via Media Dependent Interface (MDI).

- 802.3at, also known as the PoE+ standard, was released in 2009. It provides a maximum power output of 30 W and uses 2 or 4 pairs of wires for power transmission. It is backward compatible with the 802.3af standard.

  The name of the 802.3at standard is *802.3at-2009, IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment 3: Data Terminal Equipment (DTE) Power via the Media Dependent Interface (MDI) Enhancements*.
- 802.3bt, also known as the PoE++ standard, was released in 2018. It provides a maximum power output of 90 W and is compatible with existing low-power PoE devices. This standard uses 4 pairs of wires for power transmission and supports PoE devices over long distances and with better performance. It is backward compatible with the 802.3af and 802.3at standards.

  The name of the 802.3bt standard is *802.3bt-2018 - IEEE Standard for Ethernet Amendment 2: Physical Layer and Management Parameters for Power over Ethernet over 4 pairs*.

Restrictions and guidelines: PoE configuration

△ **CAUTION:**

When a PI is IEEE 802.3at- or IEEE 802.3bt compliant, it supplies power to the PD only if the total remaining power of all PSEs is greater than or equal to the guard band. To obtain the total remaining power of all PSEs, execute the **`display poe interface`** without specifying an interface. The value of the **Remaining** field is the total remaining power of all PSEs. The guard band value depends on the PoE standard and software version.

You can configure settings for a PI through either of the following ways:

- Configure the settings directly on the PI.
- Configure a PoE profile and apply it to the PI. If you apply a PoE profile to multiple PIs, these PIs have the same PoE features. If you connect a PD to another PI, you can apply the PoE profile of the original PI to the new PI. This method relieves the task of configuring PoE on the new PI.

You can only use one way to configure a parameter for a PI. To use the other way to reconfigure a parameter, you must first remove the original configuration.

You must use the same configuration method for the **`poe max-power`** *`max-power`* and **`poe priority`** { **`critical`** | **`high`** | **`low`** } commands.

If a PI fails to supply power to a device connected to it, identify whether the device is a PD. If it is a PD, enable nonstandard PD protection for the device.

PoE configuration tasks at a glance

To configure PoE, perform the following tasks:

1. Enabling PoE
   a. Enabling PoE for a PI
2. (Optional.) Configuring PoE delay
3. (Optional.) Disabling PoE on shutdown interfaces
4. (Optional.) Allowing inrush currents drawn by PDs
5. (Optional.) Enabling PI power cycling upon a system warm reboot
6. (Optional.) Configuring PD detection

- Enabling nonstandard PD detection
- Configuring a PD disconnection detection mode

7. (Optional.) Configuring the PoE power
   - Configuring the maximum PI power
8. (Optional.) Configuring the PoE priority policy
   - Configuring the PI priority policy
9. (Optional.) Configuring PoE monitoring
   - Configuring PSE power monitoring
10. (Optional.) Configuring a PI by using a PoE profile

    To use a PoE profile to enable PoE and configure the priority and maximum power for a PI, see "Configuring a PI by using a PoE profile."
11. (Optional.) Upgrading PSE firmware in service
12. (Optional.) Enabling forced PoE power supply

Prerequisites for configuring PoE

Before you configure PoE, make sure the PoE power supply and PSEs are operating correctly.

Enabling PoE

Enabling PoE for a PI

About this task

After you enable PoE for a PI, the PI supplies power to the connected PD if the PI will not result in PSE power overload. PSE overload occurs when the sum of the power consumption of all PIs exceeds the maximum power of the PSE.

If the PI will result in PSE power overload, the following restrictions apply:

- If the PI priority policy is not enabled, the PI does not supply power to the connected PD.
- If the PI priority policy is enabled, whether the PDs can be powered depends on the priority of the PI.

For more information about the PI priority policy, see "Configuring the PI priority policy."

Restrictions and guidelines

A PI can supply power to a PD only when the PI and PD use the same power transmission mode.

Procedure

1. Enter system view.

   system-view
2. Enter PI view.

   interface *interface-type interface-number*
3. (Optional.) Configure a description for the PD connected to the PI.

   poe pd-description *text*

   By default, no description is configured for the PD connected to the PI.
4. (Optional.) Specify a PoE standard for the PI.

   poe standard { **at** | **bt** }

   By default, a PI complies with the IEEE 802.3bt PoE standard.

NOTE:

Support for this feature depends on the device model. Only some models support this feature.

5. Enable PoE for the PI.

   poe enable

   By default, when the device starts up with the initial configuration, PoE is disabled on PIs. When the device starts up with the factory defaults, PoE is enabled on PIs.

   For more information about the initial configuration and factory defaults, see configuration file management in *Fundamentals Configuration Guide*.

Enabling fast PoE for a PSE

About this task

This feature enables a PI on a PSE to supply power to PDs immediately after the PSE is powered on.

Procedure

1. Enter system view.

   system-view

2. Enable fast PoE for a PSE.

   Single-PSE device:

   poe fast-on enable

   Multiple-PSE device:

   poe fast-on enable pse *pse-id*

   By default, fast PoE is disabled on a PSE.

Configuring PoE delay

About this task

By default, the device executes the **poe enable** command and supplies power to an interface immediately when any one of the following conditions is met:

- The **poe enable** command is configured.
- The device reboots with the **poe enable** command in the configuration file.
- The interface comes up and the PoE module resumes PoE power supply to the interface.

This task creates a PoE delay timer after the **poe enable** command is executed and allows the PoE module to supply power to the PI only after the timer expires.

Restrictions and guidelines

The **undo poe enable** command is executed immediately upon configuration and is not affected by this task.

Procedure

1. Enter system view.

   system-view

2. Enable PoE delay.

   poe power-delay *time*

   By default, PoE delay is disabled.

Disabling PoE on shutdown interfaces

About this task

By default, the device continues supplying power to an interface after the interface is shut down by the **shutdown** command or by an upper layer module such as monitor link. As a result, the PD connected to the shutdown interface operates continuously but fails to access the network.

This task disables the PoE module from supplying power to an interface after the interface is shut down. After the interface comes up or you disable this task, the PoE module resumes power supply to the interface.

Restrictions and guidelines

The task does not power off an interface that has been shut down but is supplying power to a PD.

Procedure

1. Enter system view.
   system-view
2. Disable PoE on shutdown interfaces.
   poe track-shutdown
   By default, PoE is not disabled on shutdown interfaces.

Allowing inrush currents drawn by PDs

About this task

Inrush current might occur at PD startup and trigger PSE self-protection, As a result, the PSE stops supplying power to the PDs. To continue power supply to the PDs, configure this feature to allow inrush current drawn by PDs.

IEEE 802.3af and IEEE 802.3at define specifications for inrush current.

Restrictions and guidelines

CAUTION:
Inrush current might damage device components. Use this feature with caution.

Procedure

1. Enter system view.
   system-view
2. Enter PI view.
   interface *interface-type interface-number*
3. Allow inrush current drawn by PDs.
   poe high-inrush enable pse *pse-id*
   By default, inrush current drawn by PDs is not allowed.

Enabling PI power cycling upon a system warm reboot

About this task

During the system warm reboot process (upon execution of the **reboot** command), the PIs continue supplying power to the PDs but data connections between the PDs and the device are interrupted. After the system reboots, PDs might not re-initiate data connections with the device. Power cycling PIs upon a system warm reboot allows the PDs to re-establish data connections with the device after a warm reboot.

Procedure

1. Enter system view.
   system-view
2. Enter PI power cycling upon a system warm reboot.
   poe reset enable
   By default, PI power cycling upon a system warm reboot is disabled.

Configuring PD detection

Enabling nonstandard PD detection

About this task

PDs are classified into standard PDs and nonstandard PDs. Standard PDs are compliant with IEEE 802.3af and IEEE 802.3at. A PSE supplies power to a nonstandard PD only after nonstandard PD detection is enabled.

The device supports PSE-based and PI-based nonstandard PD detection. Enabling nonstandard PD detection for a PSE enables this feature for all PIs on the PSE. As a best practice for disabling nonstandard PD detection for all PIs successfully in one operation, disable this feature in both system view and interface view.

Procedure

1. Enter system view.
   system-view
2. Enable nonstandard PD detection. Choose one option as needed.
   - Enable nonstandard PD detection for the PSE.
     poe legacy enable pse *pse-id*
     By default, nonstandard PD detection is disabled for a PSE.
   - Execute the following commands in sequence to enable nonstandard PD detection for a PI:
     interface *interface-type interface-number*
     poe legacy enable
     By default, nonstandard PD detection is disabled for a PI.

Configuring a PD detection mode

About this task

The device detects PDs in one of the following modes:

- **None**—The device supplies power to PDs that are correctly connected to the device without causing short circuit.
- **Simple**—The device supplies power to PDs that comply with basic requirements of 802.3af or 802.3at.
- **Strict**—The device supplies power to PDs that comply with all requirements of 802.3af or 802.3at.
- **Loose**—The device supplies power to PDs that do not comply with basic requirements of 802.3af or 802.3at due to non-standard capacitance or high interference.

Restrictions and guidelines

CAUTION:

A non-PD device might be damaged when power is supplied to it. To avoid device damage, do not specify the **none** mode when the PI connects to a non-PD device.

For this feature to take effect on nonstandard PDs, you must enable detection for nonstandard PDs by using the **poe legacy enable** command.

Procedure

1. Enter system view.

   system-view
2. Enter PI view.

   interface *interface-type interface-number*
3. Configure the PD detection mode.

   poe detection-mode { **none** | **simple** | **strict** | **loose** }

   By default, the PD detection mode is strict.

NOTE:

Support for the **none** or **simple** detection mode depends on the device model. For some early models, the system prompts operation failure when you perform this task.

Configuring a PD disconnection detection mode

CAUTION:

If you change the PD disconnection detection mode when the device is running, the connected PDs are powered off. Be aware of the impact of this operation before you perform it.

About this task

A PSE detects PD disconnection in AC mode or DC mode. The AC mode uses less power than the DC mode.

Procedure

1. Enter system view.

   system-view
2. Configure a PD disconnection detection mode.

   poe disconnect { **ac** | **dc** }

   By default, the PD disconnection detection mode is dc.

NOTE:

Support for this feature depends on the device model. Some early models might not support this feature.

Configuring the PoE power

Configuring the maximum PI power

About this task

The maximum PI power is the maximum power that a PI can provide to the connected PD. If the PD requires more power than the maximum PI power, the PI does not supply power to the PD.

Procedure

1. Enter system view.

   system-view

2. Enter PI view.

   interface *interface-type interface-number*

3. Configure the maximum power for the PI.

   poe max-power *max-power*

   By default, the maximum power of a PI is the maximum power of the interface. The maximum interface power varies by device model. For more information, see the hardware information and specifications for the product.

Configuring the PoE priority policy

Configuring the PI priority policy

About this task

The PI priority policy enables the PSE to perform priority-based power allocation to PIs when PSE power overload occurs. The priority levels for PIs are critical, high, and low in descending order.

When PSE power overload occurs, the PSE supplies power to PDs as follows:

- If the PI priority policy is disabled, the PSE supplies power to PDs depending on whether you have configured the maximum PSE power.
  - If you have configured the maximum PSE power, the PSE does not supply power to the newly-added or existing PD that causes PSE power overload.
  - If you have not configured the maximum PSE power, the PSE works as follows:
    - If the PoE bank self-protection mechanism is triggered, power supply to newly connected PDs is prohibited. This mechanism reserves some power that is used when regular power is insufficient. If the reserved power is still insufficient, new PD connection is prohibited to avoid triggering the system power self-protection mechanism.
    - Triggering the system power self-protection mechanism causes the device to power cycle.
- If the PI priority policy is enabled, the PSE supplies power to PDs as follows:
  - If a PD being powered causes PSE power overload, the PSE stops supplying power to the PD.

- If a newly-added PD causes system power overload, the PSE supplies power to PDs in priority descending order of the PIs to which they are connected. If the newly-added PD and a PD being powered have the same priority, the PD being powered takes precedence. If multiple PIs being powered have the same priority, the PIs with smaller IDs takes precedence.

Restrictions and guidelines

Before you configure a PI with critical priority, make sure the remaining power from the maximum PSE power minus the maximum powers of the existing PIs with critical priority is greater than maximum power of the PI.

Configuration for a PI whose power is preempted remains unchanged.

Procedure

1. Enter system view.
   system-view
2. Enable the PI priority policy.
   poe pd-policy priority
   By default, the PI priority policy is disabled.
3. Enter PI view.
   interface *interface-type interface-number*
4. (Optional.) Configure a priority for the PI.
   poe priority { **critical** | **high** | **low** }
   By default, the priority for a PI is **low**.

Configuring PoE monitoring

Configuring PSE power monitoring

About this task

The system monitors PSE power utilization and sends notification messages when PSE power utilization exceeds or drops below the threshold. If PSE power utilization crosses the threshold multiple times in succession, the system sends notification messages only for the first crossing. For more information about the notification message, see "Configuring SNMP."

Procedure

1. Enter system view.
   system-view
2. Configure a power alarm threshold for a PSE.
   poe utilization-threshold *value* **pse** *pse-id*
   By default, the power alarm threshold for a PSE is 80%.

Configuring a PI by using a PoE profile

Restrictions and guidelines

To modify a PoE profile applied on a PI, first remove the PoE profile from the PI.

You can configure a PI either on the PI or by using a PoE profile. The **poe max-power** *max-power* and **poe priority** { **critical** | **high** | **low** } commands must be configured using the same method.

Configuring a PoE profile

1. Enter system view.
   system-view
2. Create a PoE profile and enter its view.
   poe-profile *profile-name* [*index*]
   By default, no PoE profiles exist.
3. Enable PoE.
   poe enable
   By default, PoE is disabled.
4. (Optional.) Configure the maximum PI power.
   poe max-power *max-power*
   By default, the maximum power of a PI is the maximum power of the interface. The maximum interface power varies by device model. For more information, see the hardware information and specifications for the product.
5. (Optional.) Configure a PI priority.
   poe priority { **critical** | **high** | **low** }
   The default priority is **low**.
   This command takes effect only after the PI priority policy is enabled.

Applying a PoE profile

Restrictions and guidelines

You can apply a PoE profile to multiple PIs in system view or a single PI in PI view. If you perform the operation in both views for the same PI, the most recent operation takes effect.

You can apply only one PoE profile to a PI.

Applying a PoE profile in system view

1. Enter system view.
   system-view
2. Apply a PoE profile to PIs.
   apply poe-profile { **index** *index* | **name** *profile-name* } **interface** *interface-range*
   By default, a PoE profile is not applied to a PI.

Applying a PoE profile in PI view

1. Enter system view.
   system-view
2. Enter PI view.
   interface *interface-type interface-number*
3. Apply the PoE profile to the interface.
   apply poe-profile { **index** *index* | **name** *profile-name* }
   By default, a PoE profile is not applied to a PI.

Upgrading PSE firmware in service

About this task

You can upgrade the PSE firmware in service in the following modes:

- **Refresh mode**—Updates the PSE firmware without deleting it. You can use the refresh mode in most cases.
- **Full mode**—Deletes the current PSE firmware and reloads a new one. Use the full mode if the PSE firmware is damaged and you cannot execute any PoE commands.

Restrictions and guidelines

If the PSE firmware upgrade fails because of interruption such as a device reboot, you can restart the device and upgrade it in full mode again. After the upgrade, restart the device manually for the new PSE firmware to take effect.

Procedure

1. Enter system view.
   system-view
2. Upgrade the PSE firmware in service.
   poe update { **full** | **refresh** } *filename* **pse** *pse-id*

Enabling forced PoE power supply

About this task

Before supplying power to a PD, the device performs a detection of the PD. It supplies power to the PD only after the PD passes the detection. If the PD fails the detection but the power provided by the device meets the PD specifications, you can configure this task to enable forced power supply to the PD.

When a PI is supplying power to PDs, the **poe force-power** command triggers the device to perform the following steps:

1. Immediately stop PoE power supply on the interface.
2. Wait for the interface to switch to forced power mode.
3. Automatically re-enable PoE power supply.

Restrictions and guidelines

This task enables the device to supply power to a PD directly without performing a detection of the PD. To avoid damaging the PD, make sure the power provided by the device meets the PD specifications before configuring this task.

If a PI is supplying PoE power, executing the **poe force-power** command forcibly re-enables PoE on that PI.

For a PoE model that supports combo interfaces, when all PIs are fully powered, power redistribution is triggered if you configure this feature on a combo interface and connect a PD to it. The system cuts power to all powered PIs, restores power to PIs enabled with forced PoE power supply, and then restores power to the remaining PIs in ascending of interface numbers.

Procedure

1. Enter system view.
   system-view
2. Enter PI view.
   interface *interface-type interface-number*

3. Enable forced PoE power supply.

   poe force-power

   By default, forced PoE power supply is disabled.

CAUTION:

This command enables the device to supply power to a PD directly without performing a detection of the PD. To avoid damaging the PD, make sure the power provided by the device meets the PD specifications before configuring this command.

Display and maintenance commands for PoE

Execute **display** commands in any view.

Task	Command
Display brief PSE information.	**display poe device** [**slot** *slot-number*]
Display the power supplying information for the specified PI.	**display poe interface** [*interface-type interface-number*]
Display power information for PIs.	**display poe interface power** [*interface-type interface-number*]
Display detailed PSE information.	**display poe pse** [*pse-id*]
Display the power supplying information for all PIs on a PSE.	**display poe pse** *pse-id* **interface**
Display power information for all PIs on a PSE.	**display poe pse** *pse-id* **interface power**
Display all information about the PoE profile.	**display poe-profile** [**index** *index* | **name** *profile-name*]
Display all information about the PoE profile applied to the specified PI.	**display poe-profile interface** *interface-type interface-number*

PoE configuration examples

Example: Configuring PoE

Network configuration

As shown in Figure 2, the device has PSE 4. Configure PoE to meet the following requirements:

- The device supplies power to IP phones and APs.
- The device preferentially guarantees power supply to IP phone B when PSE power overload occurs.
- Power supplied to AP 3 does not exceed 9000 milliwatts.

Figure 2 Network diagram

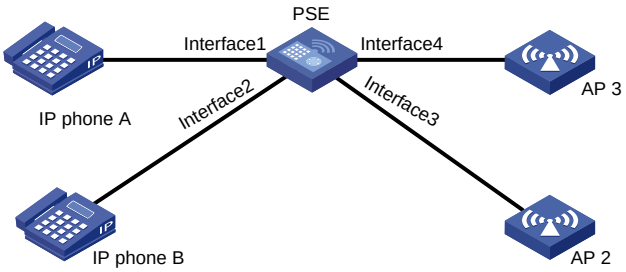


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

Enable the PI priority policy.

```
<PSE> system-view
[PSE] poe pd-policy priority
```

Enable PoE on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/3.

```
[PSE] interface gigabitethernet 1/0/1
[PSE-GigabitEthernet1/0/1] poe enable
[PSE-GigabitEthernet1/0/1] quit
[PSE] interface gigabitethernet 1/0/3
[PSE-GigabitEthernet1/0/3] poe enable
[PSE-GigabitEthernet1/0/3] quit
```

Enable PoE on GigabitEthernet 1/0/2, and set its power priority to **critical**.

```
[PSE] interface gigabitethernet 1/0/2
[PSE-GigabitEthernet1/0/2] poe enable
[PSE-GigabitEthernet1/0/2] poe priority critical
[PSE-GigabitEthernet1/0/2] quit
```

Enable PoE on GigabitEthernet 1/0/4, and set its maximum power to 9000 milliwatts.

```
[PSE] interface gigabitethernet 1/0/4
[PSE-GigabitEthernet1/0/4] poe enable
[PSE-GigabitEthernet1/0/4] poe max-power 9000
[PSE-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Verify that the device is supplying power correctly to the IP phones and APs, and the IP phones and APs are operating correctly.

Troubleshooting PoE

Failure to set the priority of a PI to critical

Symptom

Power supply priority configuration for a PI failed.

Solution

To resolve the issue:

1. Identify whether the remaining guaranteed power of the PSE is lower than the maximum power of the PI. If it is, increase the maximum PSE power or reduce the maximum power of the PI.
2. Identify whether the priority has been configured through other methods. If the priority has been configured, remove the configuration.
3. If the issue persists, contact H3C Support.

Failure to apply a PoE profile to a PI

Symptom

PoE profile application for a PI failed.

Solution

To resolve the issue:

1. Identify whether some settings in the PoE profile have been configured. If they have been configured, remove the configuration.
2. Identify whether the settings in the PoE profile meet the requirements of the PI. If they do not, modify the settings in the PoE profile.
3. Identify whether another PoE profile is already applied to the PI. If it is, remove the application.
4. If the issue persists, contact H3C Support.

Contents

Configuring SNMP 1

About SNMP 1
SNMP framework 1
MIB and view-based MIB access control 1
SNMP operations 2
Protocol versions 2
Access control modes 2
SNMP attack prevention 2
FIPS compliance 3
Restrictions and guidelines: SNMP configuration 4
SNMP tasks at a glance 4
Enabling the SNMP agent 4
Enabling SNMP versions 5
Configuring SNMP common parameters 5
Configuring an SNMPv1 or SNMPv2c community 7
About configuring an SNMPv1 or SNMPv2c community 7
Restrictions and guidelines for configuring an SNMPv1 or SNMPv2c community 7
Configuring an SNMPv1/v2c community by a community name 7
Configuring an SNMPv1/v2c community by creating an SNMPv1/v2c user 8
Configuring an SNMPv3 group and user 8
Restrictions and guidelines for configuring an SNMPv3 group and user 8
Configuring an SNMPv3 group and user in non-FIPS mode 9
Configuring an SNMPv3 group and user in FIPS mode 9
Configuring SNMP attack prevention 10
Enabling SNMP silence 10
Enabling SNMP IP blacklist 11
Configuring the SNMPv3 user denylist 11
Enabling MIB statistics caching for interfaces 11
Configuring SNMP notifications 12
About SNMP notifications 12
Enabling SNMP notifications 12
Configuring parameters for sending SNMP notifications 13
Examining the system configuration for changes 15
Configuring SNMP logging 15
Display and maintenance commands for SNMP 16
SNMP configuration examples 17
Example: Configuring SNMPv1/SNMPv2c 17
Example: Configuring SNMPv3 18

Configuring SNMP

About SNMP

Simple Network Management Protocol (SNMP) is used for a management station to access and operate the devices on a network, regardless of their vendors, physical characteristics, and interconnect technologies.

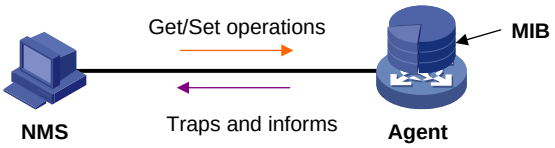
SNMP enables network administrators to read and set the variables on managed devices for state monitoring, troubleshooting, statistics collection, and other management purposes.

SNMP framework

The SNMP framework contains the following elements:

- **SNMP manager**—Works on an NMS to monitor and manage the SNMP-capable devices in the network. It can get and set values of MIB objects on an agent.
- **SNMP agent**—Works on a managed device to receive and handle requests from the NMS, and sends notifications to the NMS when events, such as an interface state change, occur.
- **Management Information Base (MIB)**—Specifies the variables (for example, interface status and CPU usage) maintained by the SNMP agent for the SNMP manager to read and set.

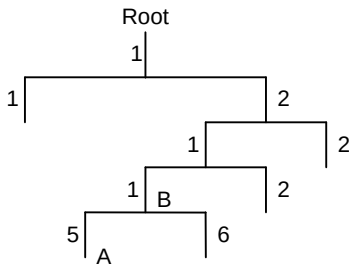
Figure 1 Relationship between NMS, agent, and MIB



MIB and view-based MIB access control

A MIB stores variables called "nodes" or "objects" in a tree hierarchy and identifies each node with a unique OID. An OID is a dotted numeric string that uniquely identifies the path from the root node to a leaf node. For example, object B in Figure 2 is uniquely identified by the OID {1.2.1.1}.

Figure 2 MIB tree



A MIB view represents a set of MIB objects (or MIB object hierarchies) with certain access privileges and is identified by a view name. The MIB objects included in the MIB view are accessible while those excluded from the MIB view are inaccessible.

A MIB view can have multiple view records each identified by a *view-name oid-tree* pair.

You control access to the MIB by assigning MIB views to SNMP groups or communities.

SNMP operations

SNMP provides the following basic operations:

- **Get**—NMS retrieves the value of an object node in an agent MIB.
- **Set**—NMS modifies the value of an object node in an agent MIB.
- **Notification**—SNMP notifications include traps and informs. The SNMP agent sends traps or informs to report events to the NMS. The difference between these two types of notification is that informs require acknowledgment but traps do not. Informs are more reliable but are also resource-consuming. Traps are available in SNMPv1, SNMPv2c, and SNMPv3. Informs are available only in SNMPv2c and SNMPv3.

Protocol versions

The device supports SNMPv1, SNMPv2c, and SNMPv3 in non-FIPS mode and supports only SNMPv3 in FIPS mode. An NMS and an SNMP agent must use the same SNMP version to communicate with each other.

- **SNMPv1**—Uses community names for authentication. To access an SNMP agent, an NMS must use the same community name as set on the SNMP agent. If the community name used by the NMS differs from the community name set on the agent, the NMS cannot establish an SNMP session to access the agent or receive traps from the agent.
- **SNMPv2c**—Uses community names for authentication. SNMPv2c is compatible with SNMPv1, but supports more operation types, data types, and error codes.
- **SNMPv3**—Uses a user-based security model (USM) to secure SNMP communication. You can configure authentication and privacy mechanisms to authenticate and encrypt SNMP packets for integrity, authenticity, and confidentiality.

Access control modes

SNMP uses the following modes to control access to MIB objects:

- **View-based Access Control Model**—VACM mode controls access to MIB objects by assigning MIB views to SNMP communities or users.
- **Role based access control**—RBAC mode controls access to MIB objects by assigning user roles to SNMP communities or users.
    - SNMP communities or users with predefined user role network-admin or level-15 have read and write access to all MIB objects.
    - SNMP communities or users with predefined user role network-operator have read-only access to all MIB objects.
    - SNMP communities or users with a user-defined user role have access rights to MIB objects as specified by the **`rule`** command.

RBAC mode controls access on a per MIB object basis, and VACM mode controls access on a MIB view basis. As a best practice to enhance MIB security, use the RBAC mode.

If you create the same SNMP community or user with both modes multiple times, the most recent configuration takes effect. For more information about RBAC, see *Fundamentals Command Reference*.

SNMP attack prevention

SNMP uses the following features to automatically detect and prevent SNMP attacks:

- **SNMP silence**—If an SNMP attack is detected, the device denies access from all NMSs during the 5-minute silence period.

- **SNMP IP address blacklist**—The device denies access from IP addresses in the blacklist during the locking period.
- **SNMP user blacklist**—The device denies access from SNMPv3 users in the blacklist during the locking period.

SNMP silence

SNMP silence enables the device to automatically detect and defend against SNMP attacks.

After you enable SNMP, the device automatically starts an SNMP silence timer and counts the number of packets that fail SNMP authentication within 1 minute.

- If the number of the packets is smaller than 100, the device restarts the timer and counting.
- If the number of the packets is equal to or greater than 100, the SNMP module enters a silence period of 4 to 5 minutes, during which the device does not respond to any SNMP packets. After the silence period expires, the SNMP module resumes normal operations and the device restarts the timer and counting.
- Whether SNMP silence must be enabled from the CLI depends on the device model.

SNMP IP blacklist

The SNMP IP blacklist feature uses the following mechanism to prevent SNMP attacks:

1. If an NMS fails to establish an SNMP connection to the device, the device adds the IP address of the NMS to the SNMP IP address blacklist, locks the IP address, and starts the observation period. The first locking period is 8 seconds and the observation period is 5 minutes. During the locking period, the IP address is not allowed to establish SNMP connections to the device.
2. After the first locking period expires, if the IP address fails to establish an SNMP connection to the device before the observation period expires, the IP address is locked for a second time and the observation period is restarted. The second locking period is 16 seconds and the observation period is 5 minutes.
3. After the second locking period expires, if the IP address fails to establish an SNMP connection to the device before the observation period expires, the IP address is locked for a third time and the observation period is restarted. The third locking period is 32 seconds and the observation period is 5 minutes.
4. After the third locking period expires, if the IP address fails to establish an SNMP connection to the device before the observation period expires, the IP address is locked for the fourth time and the observation period is restarted. The fourth locking period is 5 minutes and the observation period is 5 minutes.

If the IP address establishes an SNMP connection to the device during the observation period, the IP address will be unlocked automatically and deleted from the blacklist.

SNMP user denylist

If the number of consecutive authentication failures of an SNMPv3 user that attempts to access the device exceeds the maximum limit within the specified period of time, the system will add the user to the SNMPv3 user denylist. An SNMPv3 user in the denylist is not allowed to access the device. This feature allows you to prevent malicious attacks from unauthorized users effectively.

An SNMPv3 user added to the denylist will be removed automatically from the denylist after the locking time expires. If a user is added to the denylist mistakenly, you can manually remove it from the denylist.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Restrictions and guidelines: SNMP configuration

For device security, follow these restrictions and guidelines when the device is operating in weak security mode:

- As a best practice, do use SNMPv1 or SNMPv2c, and do not set the community name to public or private when the device is running in weak security mode.
- As a best practice, configure an authentication algorithm and an encryption algorithm.

For device security, follow these restrictions and guidelines when the device is operating in strong security mode:

- Do not use SNMPv1 or SNMPv2c, and do not set the community name to public or private when the device is running in strong security mode.
- Configure the SHA256 authentication algorithm or an authentication algorithm with higher security, and configure the AES128 encryption algorithm or an encryption algorithm with higher security

To display security risks in the system and the corresponding recommended actions, execute the **`display system security-risk`** command. To specify the enhanced security mode for the device, execute the **`enhanced-security mode enable`** command. By default, the device runs in weak security mode. For more information about security modes, see system security configuration in *Security Configuration Guide*.

SNMP tasks at a glance

The device can act as an SNMP NMS and SNMP agent at the same time.

When the device acts as an SNMP agent, perform the following tasks:

1. Enabling the SNMP agent
2. Enabling SNMP versions
3. Configuring SNMP basic parameters
   - (Optional.) Configuring SNMP common parameters
   - Configuring an SNMPv1 or SNMPv2c community
   - Configuring an SNMPv3 group and user
4. (Optional.) Configuring SNMP attack prevention
   - Enabling SNMP silence
   - Enabling SNMP IP blacklist
   - Configuring the SNMPv3 user denylist
5. (Optional.) Enabling MIB statistics caching for interfaces
6. (Optional.) Configuring SNMP notifications
7. (Optional.) Examining the system configuration for changes
8. (Optional.) Configuring SNMP logging

Enabling the SNMP agent

Restrictions and guidelines

The SNMP agent is enabled when you use any command that begins with **`snmp-agent`** except for the **`snmp-agent calculate-password`** command.

The SNMP agent will fail to be enabled when the port that the agent will listen on is used by another service. You can use the **snmp-agent port** command to specify a listening port. To view the UDP port use information, execute the **display udp verbose** command. For more information about the **display udp verbose** command, see IP performance optimization commands in *Layer 3—IP Services Configuration Guide*.

If you disable the SNMP agent, the SNMP settings do not take effect. The **display current-configuration** command does not display the SNMP settings and the SNMP settings will not be saved in the configuration file. For the SNMP settings to take effect, enable the SNMP agent.

Procedure

1. Enter system view.
   system-view
2. Enable the SNMP agent.
   snmp-agent
   By default, the SNMP agent is disabled.

Enabling SNMP versions

About this task

Before you use the SNMP feature, execute the **display snmp-agent sys-info** command to view the current SNMP version. For the device to receive SNMP messages with a specific version, you must specify an SNMP version. To communicate with each other, an NMS and an SNMP agent must use the same SNMP version.

Restrictions and guidelines

The device supports SNMPv1, SNMPv2c, and SNMPv3 in non-FIPS mode and supports only SNMPv3 in FIPS mode.

The community name and data carried in SNMPv1 and SNMPv2c messages are in plaintext form, putting the SNMP communication at risks. As a best practice, use SNMPv3.

To use SNMP notifications in IPv6, enable SNMPv2c or SNMPv3.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP versions.
   In non-FIPS mode:
   snmp-agent sys-info version { **all** | { **v1** | **v2c** | **v3** } * }
   In FIPS mode:
   snmp-agent sys-info version version v3
   By default, SNMPv3 is enabled.
   If you execute the command multiple times with different options, all the configurations take effect, but only one SNMP version is used by the agent and NMS for communication.

Configuring SNMP common parameters

Restrictions and guidelines

- SNMP engine ID

An SNMP engine ID uniquely identifies a device in an SNMP managed network. Make sure the local SNMP engine ID is unique within your SNMP managed network to avoid communication problems. By default, the device is assigned a unique SNMP engine ID.

If you have configured SNMPv3 users, change the local SNMP engine ID only when necessary. The change can void the SNMPv3 usernames and encrypted keys you have configured.

- SNMP listening port

  The SNMP agent will fail to be enabled when the port that the agent will listen on is used by another service. You can use the **snmp-agent port** command to change the SNMP listening port. As a best practice, execute the **display udp verbose** command to view the UDP port use information before specifying a new SNMP listening port. For more information about the **display udp verbose** command, see IP performance optimization commands in *Layer 3—IP Services Configuration Guide*.

- MIB view

  You can delete the default view by using the **undo snmp-agent mib-view** command. Be cautious with deleting the default MIB view. The operation blocks the access to any MIB object on the device from NMSs that use the default view. To restore the default view, restart the device.

  If you create a new MIB view, specify both permitted and excluded nodes by using the **snmp-agent mib-view** command. If the MIB view defines only excluded nodes, users bound to this view cannot access any device nodes.

  To bind MIB views, execute the **snmp-agent community** or **snmp-agent group** command.

Procedure

1. Enter system view.

   system-view

2. Specify an SNMP listening port.

   snmp-agent port *port-number*

   By default, the SNMP listening port is UDP port 161.

3. Set a local SNMP engine ID.

   snmp-agent local-engineid { *engineid* | **bridge-mac-based** }

   By default, the local SNMP engine ID is the company ID plus the device ID.

4. Set an engine ID for a remote SNMP entity.

   snmp-agent remote { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*] **engineid** *engineid*

   By default, no remote entity engine IDs exist.

   This step is required for the device to send SNMPv3 notifications to a host, typically NMS.

5. Create or update a MIB view.

   snmp-agent mib-view { **excluded** | **included** } *view-name oid-tree* [**mask** *mask-value*]

   By default, the MIB view **ViewDefault** is predefined. In this view, all the MIB objects in the **iso** subtree but the **snmpUsmMIB**, **snmpVacmMIB**, and **snmpModules.18** subtrees are accessible.

   Each *view-name oid-tree* pair represents a view record. If you specify the same record with different MIB sub-tree masks multiple times, the most recent configuration takes effect.

6. Configure an SNMP global ACL.

   snmp-agent acl [**ipv6**] { *acl-number* | **name** *acl-name* }

   By default, no SNMP global ACL is configured.

7. Configure the system management information.

- Configure the system contact.

  snmp-agent sys-info contact *sys-contact*

  By default, the system contact is **New H3C Technologies Co., Ltd.**.
- Configure the system location.

  snmp-agent sys-info location *sys-location*

  By default, the system location is **Hangzhou, China**.

8. Create an SNMP context.

   snmp-agent context *context-name*

   By default, no SNMP contexts exist.
9. Configure the maximum SNMP packet size (in bytes) that the SNMP agent can handle.

   snmp-agent packet max-size *byte-count*

   By default, an SNMP agent can process SNMP packets with a maximum size of 1500 bytes.
10. Set the DSCP value for SNMP responses.

    snmp-agent packet response dscp *dscp-value*

    By default, the DSCP value for SNMP responses is 0.

Configuring an SNMPv1 or SNMPv2c community

About configuring an SNMPv1 or SNMPv2c community

You can create an SNMPv1 or SNMPv2c community by using a community name or by creating an SNMPv1 or SNMPv2c user. After you create an SNMPv1 or SNMPv2c user, the system automatically creates a community by using the username as the community name.

Restrictions and guidelines for configuring an SNMPv1 or SNMPv2c community

SNMPv1 and SNMPv2c settings are not supported in FIPS mode.

Make sure the NMS and agent use the same SNMP community name.

Only users with the network-admin or level-15 user role can create SNMPv1 or SNMPv2c communities, users, or groups. Users with other user roles cannot create SNMPv1 or SNMPv2c communities, users, or groups even if these roles are granted access to related commands or commands of the SNMPv1 or SNMPv2c feature.

Configuring an SNMPv1/v2c community by a community name

1. Enter system view.

   system-view
2. Create an SNMPv1/v2c community. Choose one option as needed.
   - In VACM mode:

     snmp-agent community { **read** | **write** } [**simple** | **cipher**] *community-name* [**mib-view** *view-name*] [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

- In RBAC mode:

  snmp-agent community [**simple** | **cipher**] *community-name* **user-role** *role-name* [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

3. (Optional.) Map the SNMP community name to an SNMP context.

   snmp-agent community-map *community-name* **context** *context-name*

Configuring an SNMPv1/v2c community by creating an SNMPv1/v2c user

1. Enter system view.

   system-view

2. Create an SNMPv1/v2c group.

   snmp-agent group { **v1** | **v2c** } *group-name* [**notify-view** *view-name* | **read-view** *view-name* | **write-view** *view-name*] * [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

3. Add an SNMPv1/v2c user to the group.

   snmp-agent usm-user { **v1** | **v2c** } *user-name group-name* [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

   The system automatically creates an SNMP community by using the username as the community name.

4. (Optional.) Map the SNMP community name to an SNMP context.

   snmp-agent community-map *community-name* **context** *context-name*

Configuring an SNMPv3 group and user

Restrictions and guidelines for configuring an SNMPv3 group and user

Restrictions and guidelines for privilege configuration

Only users with the network-admin or level-15 user role can create SNMPv3 users or groups. Users with other user roles cannot create SNMPv3 users or groups even if these roles are granted access to related commands or commands of the SNMPv3 feature.

Restrictions and guidelines for security level configuration

All users in an SNMP group share the security level and access rights of the group.

SNMPv3 supports the following security levels, from highest to lowest:

- **Privacy**—Authentication and encryption.
- **Authentication**—Authentication but no encryption.
- **None**—No authentication or encryption.

When you create an SNMP user group or SNMPv3 user on the device, you can specify a security level. You can also specify a security level on the NMS. The following requirements must be met when you specify a security level:

- For an SNMP connection to be established successfully, make sure the security level you specify is the same or in descending order of security for the SNMP user, on the NMS, and for

the SNMP user group. If the security level is different, the security level specified on the NMS applies.

- If the security level specified on the NMS requires authentication, the security level of the SNMPv3 user must also require authentication, and the authentication password and algorithm configured on the device must match those specified on the NMS.
- If the security level specified on the NMS requires encryption, the security level of the SNMPv3 user must also require encryption, and the encryption password and algorithm configured on the device must match those specified on the NMS.

Configuring an SNMPv3 group and user in non-FIPS mode

1. Enter system view.

   system-view

2. Create an SNMPv3 group.

   snmp-agent group v3 *group-name* [**authentication** | **privacy**] [**notify-view** *view-name* | **read-view** *view-name* | **write-view** *view-name*] *[**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

3. (Optional.) Calculate the encrypted form for the key in plaintext form.

   snmp-agent calculate-password *plain-password* **mode** { **3desmd5** | **3dessha** | **3dessha224** | **3dessha256** | **3dessha384** | **3dessha512** | **aes192md5** | **aes192sha** | **aes192sha224** | **aes192sha256** | **aes192sha384** | **aes192sha512** | **aes256md5** | **aes256sha** | **aes256sha224** | **aes256sha256** | **aes256sha384** | **aes256sha512** | **md5** | **sha** | **sha224** | **sha256** | **sha384** | **sha512** } { **local-engineid** | **specified-engineid** *engineid* }

4. Create an SNMPv3 user. Choose one option as needed.

   - In VACM mode:

     snmp-agent usm-user v3 *user-name group-name* [**remote** { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]] [{ **cipher** | **simple** } **authentication-mode** { **md5** | **sha** | **sha224** | **sha256** | **sha384** | **sha512** } *auth-password* [**privacy-mode** { **3des** | **aes128** | **aes192** | **aes256** | **des56** } *priv-password*]] [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

   - In RBAC mode:

     snmp-agent usm-user v3 *user-name* **user-role** *role-name* [**remote** { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]] [{ **cipher** | **simple** } **authentication-mode** { **md5** | **sha** | **sha224** | **sha256** | **sha384** | **sha512** } *auth-password* [**privacy-mode** { **3des** | **aes128** | **aes192** | **aes256** | **des56** } *priv-password*]] [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

   To send notifications to an SNMPv3 NMS, you must specify the **remote** keyword.

5. (Optional.) Assign a user role to the SNMPv3 user created in RBAC mode.

   snmp-agent usm-user v3 *user-name* **user-role** *role-name*

   By default, an SNMPv3 user has the user role assigned to it at its creation.

Configuring an SNMPv3 group and user in FIPS mode

1. Enter system view.

system-view

2. Create an SNMPv3 group.

   snmp-agent group v3 *group-name* { **authentication** | **privacy** } [**notify-view** *view-name* | **read-view** *view-name* | **write-view** *view-name*] * [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

3. (Optional.) Calculate the encrypted form for the key in plaintext form.

   snmp-agent calculate-password *plain-password* **mode** { **aes192sha** | **aes192sha224** | **aes192sha256** | **aes192sha384** | **aes192sha512** | **aes256sha** | **aes256sha224** | **aes256sha256** | **aes256sha384** | **aes256sha512** | **sha** | **sha224** | **sha256** | **sha384** | **sha512** } { **local-engineid** | **specified-engineid** *engineid* }

4. Create an SNMPv3 user. Choose one option as needed.

   - In VACM mode:

     snmp-agent usm-user v3 *user-name group-name* [**remote** { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]] { **cipher** | **simple** } **authentication-mode** { **sha** | **sha224** | **sha256** | **sha384** | **sha512** } *auth-password* [**privacy-mode** { **aes128** | **aes192** | **aes256** } *priv-password*] [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

   - In RBAC mode:

     snmp-agent usm-user v3 *user-name* **user-role** *role-name* [**remote** { *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]] { **cipher** | **simple** } **authentication-mode** { **sha** | **sha224** | **sha256** | **sha384** | **sha512** } *auth-password* [**privacy-mode** { **aes128** | **aes192** | **aes256** } *priv-password*] [**acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* } | **acl ipv6** { *ipv6-acl-number* | **name** *ipv6-acl-name* }] *

   To send notifications to an SNMPv3 NMS, you must specify the **remote** keyword.

5. (Optional.) Assign a user role to the SNMPv3 user created in RBAC mode.
6. **snmp-agent usm-user v3** *user-name* **user-role** *role-name*

   By default, an SNMPv3 user has the user role assigned to it at its creation.

Configuring SNMP attack prevention

Both SNMP silence and SNMP blacklist can be used to prevent SNMP attacks. You can configure the two features simultaneously. However, during the SNMP silence period, the SNMP blacklist feature does not take effect.

Enabling SNMP silence

1. Enter system view.

   system-view

2. Enable SNMP silence.

   snmp-agent silence enable

   By default, SNMP silence is disabled.

Enabling SNMP IP blacklist

1. Enter system view.
   system-view
2. Enable SNMP IP blacklist.
   snmp-agent blacklist ip-block enable
   By default, SNMP IP blacklist is disabled.

Configuring the SNMPv3 user denylist

Adding an SNMPv3 user to the denylist

1. Enter system view.
   system-view
2. Enable SNMPv3 user denylist.
   snmp-agent denylist user enable
   By default, SNMPv3 user denylist is disabled.
3. Specify the maximum number of consecutive authentication failures within the specified period of time for adding an SNMPv3 user to the denylist.
   snmp-agent denylist user authfailure *times* **period** *period-time*
   By default, an SNMPv3 user is added to the denylist if the user has five consecutive authentication failure within 5 minutes.
4. Specify the period of time that an SNMPv3 user is kept in the denylist.
   snmp-agent denylist user timeout *time*
   By default, an SNMPv3 user is kept in the denylist for 5 minutes.

Immediately removing an SNMPv3 user from the denylist

To remove an SNMPv3 from the denylist, execute the following command in user view:

snmp-agent denylist user activate *user-name*

Enabling MIB statistics caching for interfaces

About this task

Enabling MIB statistics caching for interfaces can improve the data acquisition efficiency of network devices in a high query frequency environment and reduce their processing load. This feature enables the device to directly retrieve previously retrieved data from the cache within the cache duration, avoiding redundant calculations and enhancing performance.

With this feature enabled, the device can directly retrieve interface statistics from the cache without recalculation when the statistics are retrieved from the iftable and ifxtable through SNMP again before the end of the cache duration. This practice reduces the processing load on the device and enhances the overall data retrieval response speed.

Restrictions and guidelines

Do not set a too long cache duration. In environments with low network traffic or low query frequency, you can disable MIB statistics caching for interfaces to prevent outdated data from affecting the accuracy of network management.

Prerequisites

Before you configure this command, make sure SNMP is enabled on the device. Additionally, you must assess network traffic and query frequency to set an appropriate cache duration to ensure data accuracy.

Procedure

1. Enter system view.

   system-view

2. Enable MIB statistics caching for interfaces and set the cache duration.

   snmp-agent cache if-mib [**cache-duration** *cache-duration*]

   By default, MIB statistics caching for interfaces is enabled and the cache duration is 5 seconds.

   As a best practice to optimize read performance without sacrificing too much data timeliness, set a short cache duration, for example, 30 seconds in network environments with a high query frequency.

Configuring SNMP notifications

About SNMP notifications

The SNMP agent sends notifications (traps and informs) to inform the NMS of significant events, such as link state changes and user logins or logouts. After you enable notifications for a module, the module sends the generated notifications to the SNMP agent. The SNMP agent sends the received notifications as traps or informs based on the current configuration. Unless otherwise stated, the **trap** keyword in the command line includes both traps and informs.

Enabling SNMP notifications

Restrictions and guidelines

Enable an SNMP notification only if necessary. SNMP notifications are memory-intensive and might affect device performance.

To generate linkUp or linkDown notifications when the link state of an interface changes, you must perform the following tasks:

- Enable linkUp or linkDown notification globally by using the **snmp-agent trap enable standard** [**linkdown** | **linkup**] * command.
- Enable linkUp or linkDown notification on the interface by using the **enable snmp trap updown** command.

After you enable notifications for a module, whether the module generates notifications also depends on the configuration of the module. For more information, see the configuration guide for each module.

To use SNMP notifications in IPv6, enable SNMPv2c or SNMPv3.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications.

   snmp-agent trap enable [**configuration** | *protocol* | **sgcc** [**ucmacchangetrap** | **useropertrap** | **usertrap**] | **standard** [**authentication** | **coldstart** | **linkdown** | **linkup** | **warmstart**] * | **system**]

By default, configuration change notifications, SNMP standard notifications, SGCC private MIB notifications, and system notifications are enabled. For the SNMP notifications enabling status for other modules, see the module guide.

For the device to send SNMP notifications for a protocol, first enable the protocol.

3. Enter interface view.

   interface *interface-type interface-number*

4. Enable link state notifications.

   enable snmp trap updown

   By default, link state notifications are enabled.

Configuring parameters for sending SNMP notifications

About this task

You can configure the SNMP agent to send notifications as traps or informs to a host, typically an NMS, for analysis and management. Traps are less reliable and use fewer resources than informs, because an NMS does not send an acknowledgment when it receives a trap.

When network congestion occurs or the destination is not reachable, the SNMP agent buffers notifications in a queue. You can set the queue size and the notification lifetime (the maximum time that a notification can stay in the queue). When the queue size is reached, the system discards the new notification it receives. If modification of the queue size causes the number of notifications in the queue to exceed the queue size, the oldest notifications are dropped for new notifications. A notification is deleted when its lifetime expires.

You can extend standard linkUp/linkDown notifications to include interface description and interface type, but must make sure the NMS supports the extended SNMP messages.

Configuring the parameters for sending SNMP traps

1. Enter system view.

   system-view

2. Configure a target host.

   In non-FIPS mode:

   snmp-agent target-host trap address udp-domain { *ipv4-target-host* | **ipv6** *ipv6-target-host* } [**udp-port** *port-number*] [**dscp** *dscp-value*] [**vpn-instance** *vpn-instance-name*] **params** { **cipher-securityname** *cipher-security-string* [**v1** | **v2c**] | **securityname** *security-string* [**v1** | **v2c** | **v3** [**authentication** | **privacy**]] }

   In FIPS mode:

   snmp-agent target-host trap address udp-domain { *ipv4-target-host* | **ipv6** *ipv6-target-host* } [**udp-port** *port-number*] [**dscp** *dscp-value*] [**vpn-instance** *vpn-instance-name*] **params securityname** *security-string* **v3** { **authentication** | **privacy** }

   By default, no target host is configured.

   For the configuration to take effect, do not specify both the **v1** and **ipv6** keywords in the command, because SNNMPv1 does not support IPv6.

3. (Optional.) Configure a source address for sending traps.

   snmp-agent trap source *interface-type* { *interface-number* | *interface-number.subnumber* }

   By default, SNMP uses the IP address of the outgoing routed interface as the source IP address.

Configuring the parameters for sending SNMP informs

1. Enter system view.

   system-view

2. Configure a target host.

   In non-FIPS mode:

   snmp-agent target-host inform address udp-domain { *ipv4-target-host* | **ipv6** *ipv6-target-host* } [**udp-port** *port-number*] [**vpn-instance** *vpn-instance-name*] **params** { **cipher-securityname** *cipher-security-string* **v2c** | **securityname** *security-string* { **v2c** | **v3** [**authentication** | **privacy**] } }

   In FIPS mode:

   snmp-agent target-host inform address udp-domain { *ipv4-target-host* | **ipv6** *ipv6-target-host* } [**udp-port** *port-number*] [**vpn-instance** *vpn-instance-name*] **params securityname** *security-string* **v3** { **authentication** | **privacy** }

   By default, no target host is configured.

   Only SNMPv2c and SNMPv3 support inform packets.

3. (Optional.) Configure a source address for sending informs.

   snmp-agent inform source *interface-type* { *interface-number* | *interface-number.subnumber* }

   By default, SNMP uses the IP address of the outgoing routed interface as the source IP address.

Configuring common parameters for sending notifications

1. Enter system view.

   system-view

2. (Optional.) Enable extended linkUp/linkDown notifications.

   snmp-agent trap if-mib link extended

   By default, the SNMP agent sends standard linkUp/linkDown notifications.

   If the NMS does not support extended linkUp/linkDown notifications, do not use this command.

3. Enable the device to generate authentication failure alarm messages in extended format.

   snmp-agent trap snmpv2-mib authenticationfailure extended

   By default, the device sends authentication failure alarm messages in standard format, instead of extended format.

   This command takes effect only when authentication failure alarm is enabled in SNMP standard alarms. By default, authentication failure alarm is enabled in SNMP standard alarms.

4. (Optional.) Set the notification queue size.

   snmp-agent trap queue-size *size*

   By default, the notification queue can hold 100 notification messages.

5. (Optional.) Set the notification lifetime.

   snmp-agent trap life *seconds*

   The default notification lifetime is 120 seconds.

6. (Optional.) Specify the notification format as **CMCC**.

   snmp-agent trap format cmcc

   By default, the SNMP notifications are in general format.

7. (Optional.) Configure the device to carry its serial number in the notifications sent an NMS.

   snmp-agent trap withsn

By default, the device does not carry its serial number in the notifications sent an NMS.

8. Enable SNMP notification buffering and set the buffering interval.

   snmp-agent trapbuffer enable [**interval** *interval*]

   By default, SNMP notification buffering is enabled, and the buffering interval is 1 minute.

Examining the system configuration for changes

About this task

The SNMP module examines the system running configuration, startup configuration, and next-startup configuration file for changes periodically and generates a log if any change is found. If SNMP notifications for configuration changes has been enabled, the system generates also an SNMP notification.

Procedure

1. Enter system view.

   system-view

2. Set the interval at which the SNMP module examines the system configuration for changes.

   snmp-agent configuration-examine interval *interval*

   By default, the SNMP module examines the system configuration for changes at intervals of 600 seconds.

3. Enable SNMP notifications for system configuration changes.

   snmp-agent trap enable configuration

   By default, SNMP notifications is enabled for system configuration changes.

Configuring SNMP logging

About this task

The SNMP agent logs Get requests, Set requests, Set responses, SNMP notifications, and SNMP authentication failures, but does not log Get responses.

- **Get operation**—The agent logs the IP address of the NMS, name of the accessed node, and node OID.
- **Set operation**—The agent logs the NMS' IP address, name of accessed node, node OID, variable value, and error code and index for the Set operation.
- **Notification tracking**—The agent logs the SNMP notifications after sending them to the NMS.
- **SNMP authentication failure**—The agent logs related information when an NMS fails to be authenticated by the agent.

The SNMP module sends these logs to the information center. You can configure the information center to output these messages to certain destinations, such as the console and the log buffer. The total output size for the node field (MIB node name) and the value field (value of the MIB node) in each log entry is 1024 bytes. If this limit is exceeded, the information center truncates the data in the fields. For more information about the information center, see "Configuring the information center."

Restrictions and guidelines

Enable SNMP logging only if necessary. SNMP logging is memory-intensive and might impact device performance.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP logging.

   snmp-agent log { **all** | **authfail** | **get-operation** | **set-operation** }

   By default, SNMP logging is enabled for set operations and disabled for SNMP authentication failures and get operations.

3. Enable SNMP notification logging.

   snmp-agent trap log

   By default, SNMP notification logging is disabled.

Display and maintenance commands for SNMP

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display SNMPv1 or SNMPv2c community information. (This command is not supported in FIPS mode.)	**display snmp-agent community** [**read** | **write**]
Display SNMP contexts.	**display snmp-agent context** [*context-name*]
Display SNMP group information.	**display snmp-agent group** [*group-name*]
Display the local engine ID.	**display snmp-agent local-engineid**
Display SNMP MIB node information.	**display snmp-agent mib-node** [**details** | **index-node** | **trap-node** | **verbose**]
Display MIB view information.	**display snmp-agent mib-view** [**exclude** | **include** | **viewname** *view-name*]
Display remote engine IDs.	**display snmp-agent remote** [{ *ipv4-address* | **ipv6** *ipv6-address* } [**vpn-instance** *vpn-instance-name*]]
Display SNMP agent statistics.	**display snmp-agent statistics**
Display SNMP agent system information.	**display snmp-agent sys-info** [**contact** | **location** | **version**] *
Display basic information about the notification queue.	**display snmp-agent trap queue**
Display SNMP notifications drop records.	**display snmp-agent trapbuffer drop**
Display SNMP notifications sending records.	**display snmp-agent trapbuffer send**
Display the SNMP notifications enabling status for modules.	**display snmp-agent trap-list**
Display SNMPv3 user information.	**display snmp-agent usm-user** [**engineid** *engineid* | **username** *user-name* | **group** *group-name*] *
Clear all records from the SNMP trap buffer.	**reset snmp-agent trapbuffer**

SNMP configuration examples

Example: Configuring SNMPv1/SNMPv2c

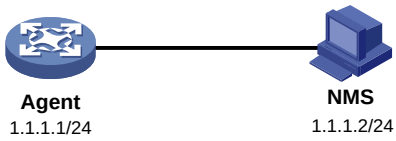
The device does not support this configuration example in FIPS mode.

The configuration procedure is the same for SNMPv1 and SNMPv2c. This example uses SNMPv1.

Network configuration

As shown in Figure 3, the NMS (1.1.1.2/24) uses SNMPv1 to manage the SNMP agent (1.1.1.1/24), and the agent automatically sends notifications to report events to the NMS.

Figure 3 Network diagram



Procedure

1. Configure the SNMP agent:

   # Assign IP address **1.1.1.1/24** to the agent and make sure the agent and the NMS can reach each other. (Details not shown.)

   # Specify SNMPv1, and create read-only community **public** and read and write community **private**.

   ```
   <Agent> system-view
   [Agent] snmp-agent sys-info version v1
   [Agent] snmp-agent community read public
   [Agent] snmp-agent community write private
   ```

   # Configure contact and physical location information for the agent.

   ```
   [Agent] snmp-agent sys-info contact Mr.Wang-Tel:3306
   [Agent] snmp-agent sys-info location telephone-closet,3rd-floor
   ```

   # Enable SNMP notifications, specify the NMS at 1.1.1.2 as an SNMP trap destination, and use **public** as the community name. (To make sure the NMS can receive traps, specify the same SNMP version in the **`snmp-agent target-host`** command as is configured on the NMS.)

   ```
   [Agent] snmp-agent trap enable
   [Agent] snmp-agent target-host trap address udp-domain 1.1.1.2 params securityname
   public v1
   ```

2. Configure the SNMP NMS:
   - Specify SNMPv1.
   - Create read-only community **public**, and create read and write community **private**.
   - Set the timeout timer and maximum number of retries as needed.

   For information about configuring the NMS, see the NMS manual.

   NOTE:

   The SNMP settings on the agent and the NMS must match.

Verifying the configuration

Try to get the MTU value of the NULL0 interface from the agent. The attempt succeeds.

```
Send request to 1.1.1.1/161 ...
```

```
Protocol version: SNMPv1
Operation: Get
Request binding:
1: 1.3.6.1.2.1.2.2.1.4.135471
Response binding:
1: Oid=ifMtu.135471 Syntax=INT Value=1500
Get finished
```

Use a wrong community name to get the value of a MIB node on the agent. You can see an authentication failure trap on the NMS.

```
1.1.1.1/2934 V1 Trap = authenticationFailure
SNMP Version = V1
Community = public
Command = Trap
Enterprise = 1.3.6.1.4.1.43.1.16.4.3.50
GenericID = 4
SpecificID = 0
Time Stamp = 8:35:25.68
```

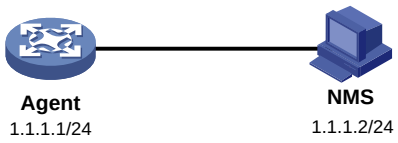
Example: Configuring SNMPv3

Network configuration

As shown in Figure 4, the NMS (1.1.1.2/24) uses SNMPv3 to monitor and manage the agent (1.1.1.1/24). The agent automatically sends notifications to report events to the NMS. The default UDP port 162 is used for SNMP notifications.

The NMS and the agent perform authentication when they establish an SNMP session. The authentication algorithm is **SHA-1** and the authentication key is **123456TESTauth&!**. The NMS and the agent also encrypt the SNMP packets between them by using the **AES** algorithm and encryption key **123456TESTencr&!**.

Figure 4 Network diagram



Configuring SNMPv3 in RBAC mode

1. Configure the agent:

   # Assign IP address **1.1.1.1/24** to the agent and make sure the agent and the NMS can reach each other. (Details not shown.)

   #Create user role **test**, and assign **test** read-only access to the objects under the **snmpMIB** node (OID:1.3.6.1.6.3.1), including the **linkUp** and **linkDown** objects.

   ```
   <Agent> system-view
   [Agent] role name test
   [Agent-role-test] rule 1 permit read oid 1.3.6.1.6.3.1
   ```

   # Assign user role **test** read-only access to the **system** node (OID:1.3.6.1.2.1.1) and read-write access to the **interfaces** node(OID:1.3.6.1.2.1.2).

   ```
   [Agent-role-test] rule 2 permit read oid 1.3.6.1.2.1.1
   [Agent-role-test] rule 3 permit read write oid 1.3.6.1.2.1.2
   [Agent-role-test] quit
   ```

Create SNMPv3 user **RBACtest**. Assign user role **test** to **RBACtest.** Set the authentication algorithm to **SHA-1**, authentication key to **123456TESTauth&!**, encryption algorithm to **AES**, and encryption key to **123456TESTencr&!**.

```
[Agent] snmp-agent usm-user v3 RBACtest user-role test simple authentication-mode sha
123456TESTauth&! privacy-mode aes128 123456TESTencr&!
```

#Configure contact and physical location information for the agent.

```
[Agent] snmp-agent sys-info contact Mr.Wang-Tel:3306
[Agent] snmp-agent sys-info location telephone-closet,3rd-floor
```

#Enable notifications on the agent. Specify the NMS at 1.1.1.2 as the notification destination, and **RBACtest** as the username.

```
[Agent] snmp-agent trap enable
[Agent] snmp-agent target-host trap address udp-domain 1.1.1.2 params
securitynameRBACtest v3 privacy
```

2. Configure the NMS:

- Specify SNMPv3.
- Create SNMPv3 user **RBACtest**.
- Enable authentication and encryption. Set the authentication algorithm to **SHA-1**, authentication key to **123456TESTauth&!**, encryption algorithm to **AES**, and encryption key to **123456TESTencr&!**.
- Set the timeout timer and maximum number of retries.

For information about configuring the NMS, see the NMS manual.

NOTE:

The SNMP settings on the agent and the NMS must match.

Configuring SNMPv3 in VACM mode

1. Configure the agent:

Assign IP address 1.1.1.1/24 to the agent, and make sure the agent and the NMS can reach each other. (Details not shown.)

Create SNMPv3 group **managev3group** and assign **managev3group** read-only access to the objects under the **snmpMIB** node (OID: 1.3.6.1.2.1.2.2) in the **test** view, including the **linkUp** and **linkDown** objects.

```
<Agent> system-view
[Agent] undo snmp-agent mib-view ViewDefault
[Agent] snmp-agent mib-view included test snmpMIB
[Agent] snmp-agent group v3 managev3group privacy read-view test
```

#Assign SNMPv3 group **managev3group** read-write access to the objects under the **system** node (OID: 1.3.6.1.2.1.1) and **interfaces** node (OID:1.3.6.1.2.1.2) in the **test** view.

```
[Agent] snmp-agent mib-view included test 1.3.6.1.2.1.1
[Agent] snmp-agent mib-view included test 1.3.6.1.2.1.2
[Agent] snmp-agent group v3 managev3group privacy read-view test write-view test
```

Add user **VACMtest** to SNMPv3 group **managev3group**, and set the authentication algorithm to **SHA-1**, authentication key to **123456TESTauth&!**, encryption algorithm to **AES**, and encryption key to **123456TESTencr&!**.

```
[Agent] snmp-agent usm-user v3 VACMtest managev3group simple authentication-mode sha
123456TESTauth&! privacy-mode aes128 123456TESTencr&!
```

Configure contact and physical location information for the agent.

```
[Agent] snmp-agent sys-info contact Mr.Wang-Tel:3306
[Agent] snmp-agent sys-info location telephone-closet,3rd-floor
```

Enable notifications on the agent. Specify the NMS at 1.1.1.2 as the trap destination, and **VACMtest** as the username.

```
[Agent] snmp-agent trap enable
[Agent] snmp-agent target-host trap address udp-domain 1.1.1.2 params VACMtest v3 privacy
```

2. Configure the SNMP NMS:
   - Specify SNMPv3.
   - Create SNMPv3 user VACMtest.
   - Enable authentication and encryption. Set the authentication algorithm to **SHA-1**, authentication key to **123456TESTauth&!**, encryption algorithm to **AES**, and encryption key to **123456TESTencr&!**.
   - Set the timeout timer and maximum number of retries.

   For information about configuring the NMS, see the NMS manual.

NOTE:

The SNMP settings on the agent and the NMS must match.

Verifying the configuration

- Use username **RBACtest** to access the agent.

  # Retrieve the value of the **sysName** node. The value **Agent** is returned.

  # Set the value for the **sysName** node to **Sysname**. The operation fails because the NMS does not have write access to the node.

  # Shut down or bring up an interface on the agent. The NMS receives linkUP (OID: 1.3.6.1.6.3.1.1.5.4) or linkDown (OID: 1.3.6.1.6.3.1.1.5.3) notifications.

- Use username **VACMtest** to access the agent.

  # Retrieve the value of the **sysName** node. The value **Agent** is returned.

  # Set the value for the **sysName** node to **Sysname**. The operation succeeds.

  # Shut down or bring up an interface on the agent. The NMS receives **linkUP** (OID: 1.3.6.1.6.3.1.1.5.4) or **linkDown** (OID: 1.3.6.1.6.3.1.1.5.3) notifications.

Contents

Configuring RMON 1

About RMON 1
RMON working mechanism 1
RMON groups 1
Sample types for the alarm group and the private alarm group 3
Protocols and standards 3
Configuring the RMON statistics function 3
About the RMON statistics function 3
Creating an RMON Ethernet statistics entry 3
Creating an RMON history control entry 3
Configuring the RMON alarm function 4
Display and maintenance commands for RMON 5
RMON configuration examples 6
Example: Configuring the Ethernet statistics function 6
Example: Configuring the history statistics function 6
Example: Configuring the alarm function 7

Configuring RMON

About RMON

Remote Network Monitoring (RMON) is an SNMP-based network management protocol. It enables proactive remote monitoring and management of network devices.

RMON working mechanism

RMON can periodically or continuously collect traffic statistics for an Ethernet port and monitor the values of MIB objects on a device. When a value reaches the threshold, the device automatically logs the event or sends a notification to the NMS. The NMS does not need to constantly poll MIB variables and compare the results.

RMON uses SNMP notifications to notify NMSs of various alarm conditions. SNMP reports function and interface operating status changes such as link up, link down, and module failure to the NMS.

RMON groups

Among standard RMON groups, the device implements the statistics group, history group, event group, alarm group, probe configuration group, and user history group. The Comware system also implements a private alarm group, which enhances the standard alarm group. The probe configuration group and user history group are not configurable from the CLI. To configure these two groups, you must access the MIB.

Statistics group

The statistics group samples traffic statistics for monitored Ethernet interfaces and stores the statistics in the Ethernet statistics table (ethernetStatsTable). The statistics include:

- Number of collisions.
- CRC alignment errors.
- Number of undersize or oversize packets.
- Number of broadcasts.
- Number of multicasts.
- Number of bytes received.
- Number of packets received.

The statistics in the Ethernet statistics table are cumulative sums.

History group

The history group periodically samples traffic statistics on interfaces and saves the history samples in the history table (etherHistoryTable). The statistics include:

- Bandwidth utilization.
- Number of error packets.
- Total number of packets.

The history table stores traffic statistics collected for each sampling interval.

Event group

The event group controls the generation and notifications of events triggered by the alarms defined in the alarm group and the private alarm group. The following are RMON alarm event handling methods:

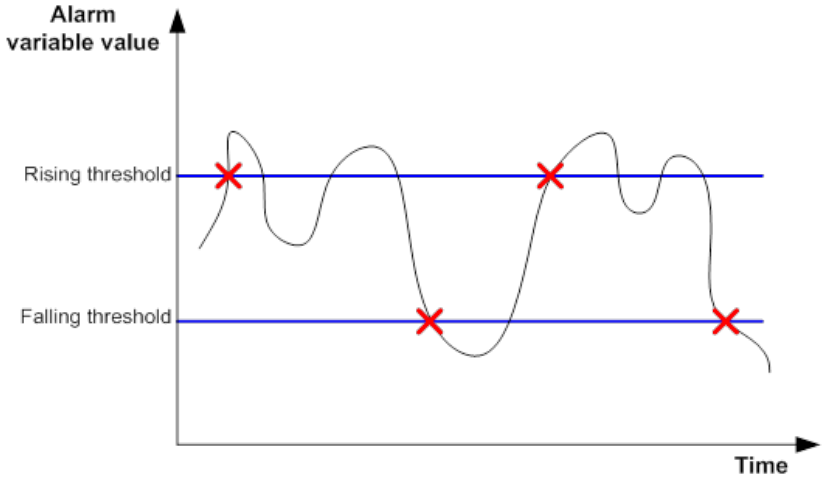
- **Log**—Logs event information (including event time and description) in the event log table so the management device can get the logs through SNMP.
- **Trap**—Sends an SNMP notification when the event occurs.
- **Log-Trap**—Logs event information in the event log table and sends an SNMP notification when the event occurs.
- **None**—Takes no actions.

Alarm group

The RMON alarm group monitors alarm variables, such as the count of incoming packets (etherStatsPkts) on an interface. After you create an alarm entry, the RMON agent samples the value of the monitored alarm variable regularly. If the value of the monitored variable is greater than or equal to the rising threshold, a rising alarm event is triggered. If the value of the monitored variable is smaller than or equal to the falling threshold, a falling alarm event is triggered. The event group defines the action to take on the alarm event.

If an alarm entry crosses a threshold multiple times in succession, the RMON agent generates an alarm event only for the first crossing. For example, if the value of a sampled alarm variable crosses the rising threshold multiple times before it crosses the falling threshold, only the first crossing triggers a rising alarm event, as shown in Figure 1.

Figure 1 Rising and falling alarm events



Private alarm group

The private alarm group enables you to perform basic math operations on multiple variables, and compare the calculation result with the rising and falling thresholds.

The RMON agent samples variables and takes an alarm action based on a private alarm entry as follows:

1. Samples the private alarm variables in the user-defined formula.
2. Processes the sampled values with the formula.
3. Compares the calculation result with the predefined thresholds, and then takes one of the following actions:
   - Triggers the event associated with the rising alarm event if the result is equal to or greater than the rising threshold.
   - Triggers the event associated with the falling alarm event if the result is equal to or less than the falling threshold.

If a private alarm entry crosses a threshold multiple times in succession, the RMON agent generates an alarm event only for the first crossing. For example, if the value of a sampled alarm variable

crosses the rising threshold multiple times before it crosses the falling threshold, only the first crossing triggers a rising alarm event.

Sample types for the alarm group and the private alarm group

The RMON agent supports the following sample types:

- **absolute**—RMON compares the value of the monitored variable with the rising and falling thresholds at the end of the sampling interval.
- **delta**—RMON subtracts the value of the monitored variable at the previous sample from the current value, and then compares the difference with the rising and falling thresholds.

Protocols and standards

- RFC 4502, *Remote Network Monitoring Management Information Base Version 2*
- RFC 2819, *Remote Network Monitoring Management Information Base Status of this Memo*

Configuring the RMON statistics function

About the RMON statistics function

RMON implements the statistics function through the Ethernet statistics group and the history group.

The Ethernet statistics group provides the cumulative statistic for a variable from the time the statistics entry is created to the current time.

The history group provides statistics that are sampled for a variable for each sampling interval. The history group uses the history control table to control sampling, and it stores samples in the history table.

Creating an RMON Ethernet statistics entry

Restrictions and guidelines

The index of an RMON statistics entry must be globally unique. If the index has been used by another interface, the creation operation fails.

You can create only one RMON statistics entry for an Ethernet interface.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Create an RMON Ethernet statistics entry.
   rmon statistics *entry-number* [**owner** *text*]

Creating an RMON history control entry

Restrictions and guidelines

You can configure multiple history control entries for one interface, but you must make sure their entry numbers and sampling intervals are different.

You can create a history control entry successfully even if the specified bucket size exceeds the available history table size. RMON will set the bucket size as closely to the expected bucket size as possible.

Procedure

1. Enter system view.

   system-view

2. Enter Ethernet interface view.

   interface *interface-type interface-number*

3. Create an RMON history control entry.

   rmon history *entry-number* **buckets** *number* **interval** *interval* [**owner** *text*]

   By default, no RMON history control entries exist.

   You can create multiple RMON history control entries for an Ethernet interface.

Configuring the RMON alarm function

Restrictions and guidelines

When you create a new event, alarm, or private alarm entry, follow these restrictions and guidelines:

- The entry must not have the same set of parameters as an existing entry.
- The maximum number of entries is not reached.

Table 1 shows the parameters to be compared for duplication and the entry limits.

Table 1 RMON configuration restrictions

Entry	Parameters to be compared	Maximum number of entries
Event	• Event description (**description** *string*) • Event type (**log**, **trap**, **logtrap**, or **none**) • Community name (*security-string*)	60
Alarm	• Alarm variable (*alarm-variable*) • Sampling interval (*sampling-interval*) • Sample type (**absolute** or **delta**) • Rising threshold (*threshold-value1*) • Falling threshold (*threshold-value2*)	60
Private alarm	• Alarm variable formula (*prialarm-formula*) • Sampling interval (*sampling-interval*) • Sample type (**absolute** or **delta**) • Rising threshold (*threshold-value1*) • Falling threshold (*threshold-value2*)	50

Prerequisites

To send notifications to the NMS when an alarm is triggered, configure the SNMP agent as described in "Configuring SNMP" before configuring the RMON alarm function.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Create an RMON event entry.

   rmon event *entry-number* [**description** *string*] { **log** | **log-trap** *security-string* | **none** | **trap** *security-string* } [**owner** *text*]

   By default, no RMON event entries exist.

3. Create an RMON alarm entry.

   - Create an RMON alarm entry.

     rmon alarm *entry-number alarm-variable sampling-interval* { **absolute** | **delta** } [**startup-alarm** { **falling** | **rising** | **rising-falling** }] **rising-threshold** *threshold-value1 event-entry1* **falling-threshold** *threshold-value2 event-entry2* [**owner** *text*]

   - Create an RMON private alarm entry.

     rmon prialarm *entry-number prialarm-formula prialarm-des sampling-interval* { **absolute** | **delta** } [**startup-alarm** { **falling** | **rising** | **rising-falling** }] **rising-threshold** *threshold-value1 event-entry1* **falling-threshold** *threshold-value2 event-entry2* **entrytype** { **forever** | **cycle** *cycle-period* } [**owner** *text*]

   By default, no RMON alarm entries or RMON private alarm entries exist.

   You can associate an alarm with an event that has not been created yet. The alarm will trigger the event only after the event is created.

Display and maintenance commands for RMON

Execute **display** commands in any view.

Task	Command
Display RMON alarm entries.	**display rmon alarm** [*entry-number*]
Display RMON event entries.	**display rmon event** [*entry-number*]
Display log information for event entries.	**display rmon eventlog** [*entry-number*]
Display RMON history control entries and history samples.	**display rmon history** [*interface-type interface-number*]
Display RMON private alarm entries.	**display rmon prialarm** [*entry-number*]
Display RMON statistics.	**display rmon statistics** [*interface-type interface-number*]

RMON configuration examples

Example: Configuring the Ethernet statistics function

Network configuration

As shown in Figure 2, create an RMON Ethernet statistics entry on the device to gather cumulative traffic statistics for GigabitEthernet 1/0/1.

Figure 2 Network diagram

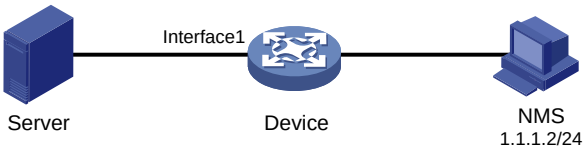


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Create an RMON Ethernet statistics entry for GigabitEthernet 1/0/1.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] rmon statistics 1 owner user1
```

Verifying the configuration

Display statistics collected for GigabitEthernet 1/0/1.

```
<Sysname> display rmon statistics gigabitethernet 1/0/1
EtherStatsEntry 1 owned by user1 is VALID.
  Interface : GigabitEthernet1/0/1<ifIndex.3>
  etherStatsOctets         : 21657     , etherStatsPkts          : 307
  etherStatsBroadcastPkts  : 56        , etherStatsMulticastPkts : 34
  etherStatsUndersizePkts  : 0         , etherStatsOversizePkts  : 0
  etherStatsFragments      : 0         , etherStatsJabbers       : 0
  etherStatsCRCAlignErrors : 0         , etherStatsCollisions    : 0
  etherStatsDropEvents (insufficient resources): 0
  Incoming packets by size:
  64     : 235       ,  65-127  : 67        ,  128-255  : 4
  256-511: 1         ,  512-1023: 0         ,  1024-1518: 0
```

Get the traffic statistics from the NMS through SNMP. (Details not shown.)

Example: Configuring the history statistics function

Network configuration

As shown in Figure 3, create an RMON history control entry on the device to sample traffic statistics for GigabitEthernet 1/0/1 every minute.

Figure 3 Network diagram

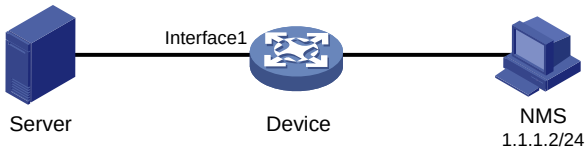


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Create an RMON history control entry to sample traffic statistics every minute for GigabitEthernet 1/0/1. Retain a maximum of eight samples for the interface in the history statistics table.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] rmon history 1 buckets 8 interval 60 owner user1
```

Verifying the configuration

Display the history statistics collected for GigabitEthernet 1/0/1.

```
[Sysname-GigabitEthernet1/0/1] display rmon history
HistoryControlEntry 1 owned by user1 is VALID
  Sampled interface     : GigabitEthernet1/0/1<ifIndex.3>
  Sampling interval     : 60(sec) with 8 buckets max
  Sampling record 1 :
    dropevents        : 0         , octets               : 834
    packets           : 8         , broadcast packets    : 1
    multicast packets : 6         , CRC alignment errors : 0
    undersize packets : 0         , oversize packets     : 0
    fragments         : 0         , jabbers              : 0
    collisions        : 0         , utilization          : 0
  Sampling record 2 :
    dropevents        : 0         , octets               : 962
    packets           : 10        , broadcast packets    : 3
    multicast packets : 6         , CRC alignment errors : 0
    undersize packets : 0         , oversize packets     : 0
    fragments         : 0         , jabbers              : 0
    collisions        : 0         , utilization          : 0
```

Get the traffic statistics from the NMS through SNMP. (Details not shown.)

Example: Configuring the alarm function

Network configuration

As shown in Figure 4, configure the device to monitor the incoming traffic statistic on GigabitEthernet 1/0/1, and send RMON alarms when either of the following conditions is met:

- The 5-second delta sample for the traffic statistic crosses the rising threshold (100).
- The 5-second delta sample for the traffic statistic drops below the falling threshold (50).

Figure 4 Network diagram

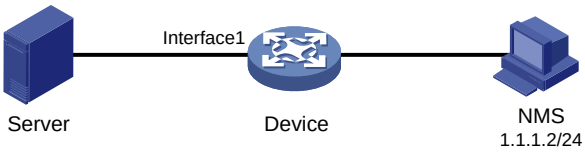


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Configure the SNMP agent (the device) with the same SNMP settings as the NMS at 1.1.1.2. This example uses SNMPv1, read community **public**, and write community **private**.

```
<Sysname> system-view
[Sysname] snmp-agent
[Sysname] snmp-agent community read public
[Sysname] snmp-agent community write private
[Sysname] snmp-agent sys-info version v1
[Sysname] snmp-agent trap enable
[Sysname] snmp-agent trap log
[Sysname] snmp-agent target-host trap address udp-domain 1.1.1.2 params securityname
public
```

Create an RMON Ethernet statistics entry for GigabitEthernet 1/0/1.

```
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] rmon statistics 1 owner user1
[Sysname-GigabitEthernet1/0/1] quit
```

Create an RMON event entry and an RMON alarm entry to send SNMP notifications when the delta sample for 1.3.6.1.2.1.16.1.1.1.4.1 exceeds 100 or drops below 50.

```
[Sysname] rmon event 1 trap public owner user1
[Sysname] rmon alarm 1 1.3.6.1.2.1.16.1.1.1.4.1 5 delta rising-threshold 100 1
falling-threshold 50 1 owner user1
```

NOTE:

The string 1.3.6.1.2.1.16.1.1.1.4.1 is the object instance for GigabitEthernet 1/0/1. The digits before the last digit (1.3.6.1.2.1.16.1.1.1.4) represent the object for total incoming traffic statistics. The last digit (1) is the RMON Ethernet statistics entry index for GigabitEthernet 1/0/1.

Verifying the configuration

Display the RMON alarm entry.

```
<Sysname> display rmon alarm 1
AlarmEntry 1 owned by user1 is VALID.
  Sample type                    : delta
  Sampled variable               : 1.3.6.1.2.1.16.1.1.1.4.1<etherStatsOctets.1>
  Sampling interval (in seconds) : 5
  Rising threshold               : 100(associated with event 1)
  Falling threshold              : 50(associated with event 1)
  Alarm sent upon entry startup  : risingOrFallingAlarm
```

```
  Latest value                   : 0
```

Display statistics for GigabitEthernet 1/0/1.

```
<Sysname> display rmon statistics gigabitethernet 1/0/1
EtherStatsEntry 1 owned by user1 is VALID.
  Interface : GigabitEthernet1/0/1<ifIndex.3>
  etherStatsOctets         : 57329     , etherStatsPkts          : 455
  etherStatsBroadcastPkts  : 53        , etherStatsMulticastPkts : 353
  etherStatsUndersizePkts  : 0         , etherStatsOversizePkts  : 0
  etherStatsFragments      : 0         , etherStatsJabbers       : 0
  etherStatsCRCAlignErrors : 0         , etherStatsCollisions    : 0
  etherStatsDropEvents (insufficient resources): 0
  Incoming packets by size :
  64     : 7         ,  65-127  : 413       ,  128-255  : 35
  256-511: 0         ,  512-1023: 0         ,  1024-1518: 0
```

The NMS receives the notification when the alarm is triggered.

Contents

Using Ansible for automated configuration management 1

About Ansible 1
Ansible network architecture 1
Basic Ansible concepts 2
Basic Ansible concepts 3
How Ansible works 4
Prerequisites for Ansible 4
Device setup examples for management with Ansible 4
Example: Setting up the device for management with Ansible 4

Using Ansible for automated configuration management

About Ansible

Ansible is an automation tool used for remote deployment on a large scale of devices that can be accessed through SSH. Deployed with Ansible, the system can perform large-scale and automated system O&M operations such as cluster configuration and application deployment. The flexible Ansible usage and replicable deployment parameters significantly reduce duplicated effort of O&M personnel and improve the O&M efficiency.

Ansible is programmed in Python and uses SSH to connect to network devices for centralized configuration and management.

Ansible network architecture

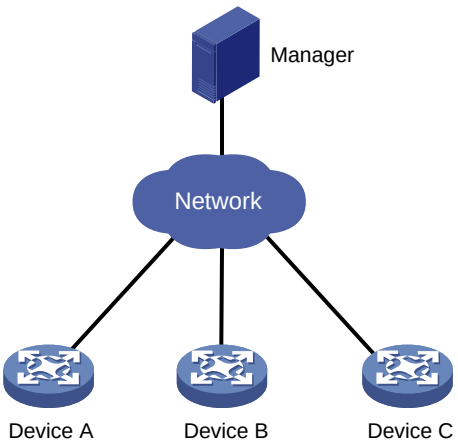
As shown in Figure 1, an Ansible system consists of the following elements:

- **Manager**—A host installed with Ansible. For more information about installing Ansible, see the Ansible official website.

  The manager supports remote configuration such as creating VLANs. Not all services modules are configurable through Ansible. To identify the service modules that you can configure by using Ansible, access the Comware 7 Python library.
- **Managed device**—A managed device acts as an SSH server and does not require any agent software to be installed. Once an SSH connection is established between the manager and the device, the device automatically executes the commands deployed by the manager.

  H3C devices can act as managed devices.

Figure 1 Ansible network architecture



Basic Ansible concepts

Figure 2 Basic Ansible concepts

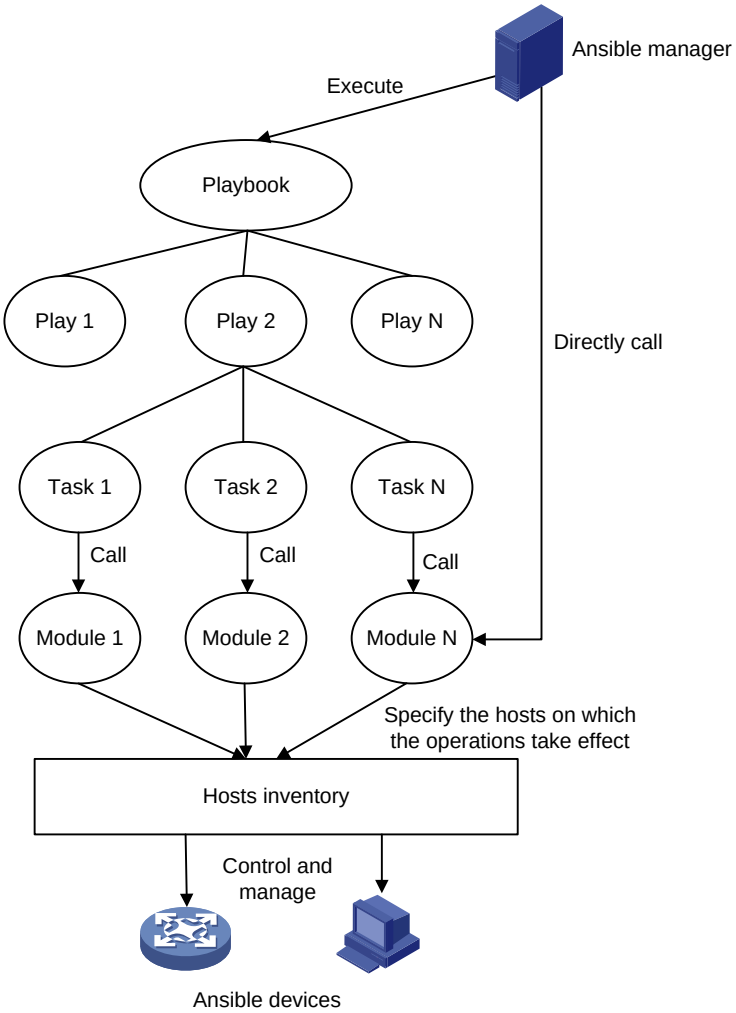


Figure 3 shows the basic Ansible concepts:

- **Playbook—**An automated script written in YAML format. A playbook contains all configuration to be deployed to the devices and is executed in the order of the content. It avoids repeated and large-scale input on the manager, which is crucial for Ansible automation.
- **Play**—A play contains all operations targeted at a group of hosts. A playbook can contain multiple plays.
- **Task**—A basic unit that describes the operation and configuration to be executed on the target hosts. Each operation corresponds to a task. For example, creating VLAN 10 and specifying an IP address for VLAN-interface 10 correspond to two separate tasks. A play can contain multiple tasks.
- **Module**—A module can be a Python or Bash script. Ansible can perform specific operations on the target hosts and return the results only after tasks call modules. Modules can be called in the playbook or directly called by command lines on the manager. As a best practice to avoid repeated execution of commands, use the playbook to call modules.
- **Hosts inventory**—An important configuration file that informs Ansible of hosts to be managed and categorizes them. Target hosts of the operations written in the playbook must be defined in the hosts inventory.

Basic Ansible concepts

Figure 3 Basic Ansible concepts

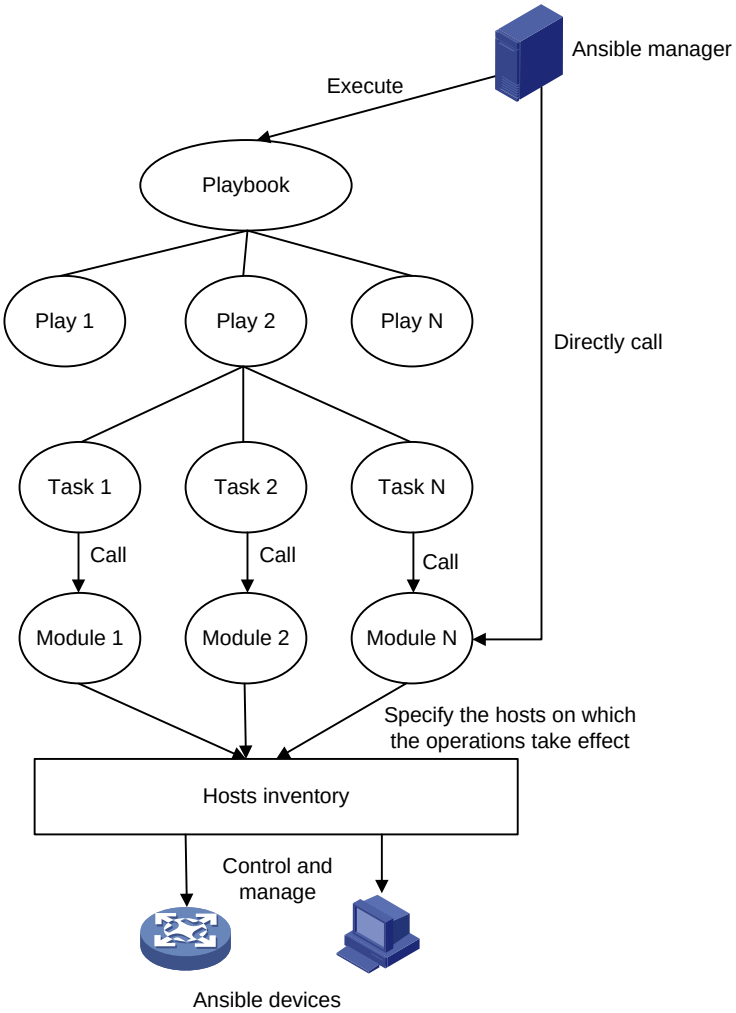


Figure 3 shows the basic Ansible concepts:

- **Playbook—**An automated script written in YAML format. A playbook contains all configuration to be deployed to the devices and is executed in the order of the content. It avoids repeated and large-scale input on the manager, which is crucial for Ansible automation.
- **Play**—A play contains all operations targeted at a group of hosts. A playbook can contain multiple plays.
- **Task**—A basic unit that describes the operation and configuration to be executed on the target hosts. Each operation corresponds to a task. For example, creating VLAN 10 and specifying an IP address for VLAN-interface 10 correspond to two separate tasks. A play can contain multiple tasks.
- **Module**—A module can be a Python or Bash script. Ansible can perform specific operations on the target hosts and return the results only after tasks call modules. Modules can be called in the playbook or directly called by command lines on the manager. As a best practice to avoid repeated execution of commands, use the playbook to call modules.
- **Hosts inventory**—An important configuration file that informs Ansible of hosts to be managed and categorizes them. Target hosts of the operations written in the playbook must be defined in the hosts inventory.

How Ansible works

1. Install Ansible on the manager and write the operations to be executed by the devices in the playbook.
2. The manager (SSH client) initiates SSH connections to the devices (SSH servers).
3. The manager executes the playbook to deploy commands to the target devices.
4. The devices automatically execute the received commands.

Prerequisites for Ansible

Before you use Ansible to configure the device, complete the following tasks:

- Configure NTP or PTP or manually configure the system time on the Ansible server and the device to synchronize their system time. For more information about time synchronization, see device management configuration in *Fundamentals Configuration guide* and NTP and PTP configuration in *Network Management and Monitoring Configuration Guide*.
- Configure the device as an SSH server. For more information about SSH configuration, see *Security Configuration Guide*.

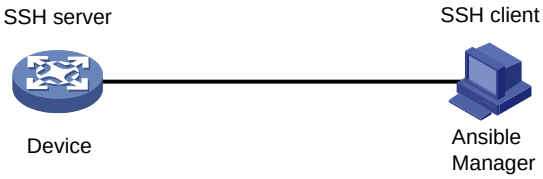
Device setup examples for management with Ansible

Example: Setting up the device for management with Ansible

Network configuration

As shown in Figure 4, enable SSH server on the device and use the Ansible manager to manage the device over SSH.

Figure 4 Network diagram



Prerequisites

1. Assign IP addresses to the device and manager. Make sure they have connectivity to each other.
2. Configure NTP or PTP or manually configure the system time on both the device and manager so they use the same system time. (Details not shown.)

Configuring the device as an SSH server

Create local key pairs. For more information, see SSH configuration in *Security Configuration Guide*.

Create a local user named **abc** and set the password for the user.

```
<Device> system-view
[Device]local-user abc
[Device-luser-manage-abc] password simple 123456TESTplat&!
```

Assign the network-admin user role to the user and authorize the user to use SSH, HTTP, and HTTPS services.

```
[Device-luser-manage-abc] authorization-attribute user-role network-admin
[Device-luser-manage-abc] service-type ssh http https
[Device-luser-manage-abc] quit
```

Enable NETCONF over SSH.

```
[Device] netconf ssh server enable
```

Enable scheme authentication for SSH login and assign the network-admin user role to the login users.

```
[Device] line vty 0 63
[Device-line-vty0-63] authentication-mode scheme
[Device-line-vty0-63] user-role network-admin
[Device-line-vty0-63] quit
```

Enable the SSH server.

```
[Device] ssh server enable
```

Authorize SSH user abc to use all service types, including SCP, SFTP, Stelnet, and NETCONF. Set the authentication method to password.

```
[Device] ssh user abc service-type all authentication-type password
```

Enable the SFTP server or SCP server.

- If the device supports SFTP, enable the SFTP server.
  ```
  [Device] sftp server enable
  ```
- If the device does not support SFTP, enable the SCP server.
  ```
  [Device] scp server enable
  ```

Configuring the Ansible manager

In this example, the Ubuntu operating system is installed on the manager. To use a manager installed with another Linux system or Unix-like system, see the Ansible documentation official website at https://docs.ansible.com.

To configure the Ansible manager:

1. Configure the default hosts inventory:

   The default hosts inventory file path is **/etc/ansible/hosts**. The manager must write the host information of the device (such as the IP address or domain name) in this file, for example:

   ```
   192.168.1.1    Device
   ```

   For more information about the hosts inventory writing rules and editing the file path, see the Ansible documentation official website at https://docs.ansible.com.

2. Use the PiP tool to install Ansible.

   # Check the Python version.

   ```
   $ python --version
   ```

   # If Python is not installed, install it first.

   ```
   $ sudo apt-get update
   $ sudo apt-get install python3
   ```

   # Check the PiP version.

   ```
   $ pip --version
   ```

   # If PiP is not installed, install it first.

   ```
   $ sudo apt-get update
   $ sudo apt-get install python-pip
   ```

   # Install Ansible.

```
$ sudo pip install markupsafe
$ sudo pip install ansible
```

After installation, check the Ansible version.

```
$ ansible -- version
```

3. Install the Python library.

NOTE:

When Ansible executes scripts, the called modules must be implemented via Python. Install the Python library to obtain the code that enables the modules called by the playbook to operate correctly.

Download the up-to-date Comware Python library source code from GitHub.

```
$ sudo git clone https://github.com/HPENetworking/pyhpecw7
```

Install the Comware Python library.

```
$ cd pyhpecw7-master
$ sudo chmod 777 setup.py
$ sudo python setup.py install
```

NOTE:

If a message displays a missing component, use the **`$ sudo apt-get install`** *xxx* command to install the component. *xxx* represents the name of a component.

Download the Ansible module.

```
$ sudo git clone https://github.com/HPENetworking/ansible-hpe-cw7
```

Use the **`find`** command to search for the **namespaces.py** file from the root directory. If the file is found and the file path contains **pyhpecw7**, open the file and edit the content as follows:

```
HPDATA="http://www.h3c.com/netconf/data:1.0"
HPDATA_C='{'+HPDATA+'}'
NETCONFBASE="urn:ietf:params:xml:ns:netconf:base:1.0"
NETCONFBASE_C='{'+NETCONFBASE+'}'
HPCONFIG="http://www.h3c.com/netconf/data:1.0"
HPCONFIG_C='{'+HPCONFIG+'}'
HPACTION="http://www.h3c.com/netconf/data:1.0"
HPACTION_C='{'+HPACTION+'}'
```

4. Configure the SSH service.

Install an SSH server.

```
$ sudo apt-get install ssh
```

Edit the sshd configuration file. Change the value for the **PermitRootLogin** field to **Yes**.

```
$ sudo vi sshd_config
```

Restart the SSH service.

```
$ sudo service ssh restart
```

5. Define the hosts inventory variables.

In the **ansible-hpe-cw7-master/hosts** file, change the username and password that the manager use to log in to the device through SSH. The manager uses the new username and password to log in to all hosts.

```
[all:vars]
```

```
username=abc
password=123456TESTplat&!
```

In the **ansible-hpe-cw7-master/hosts** file, change the **[switches]** field to **Device**. All host names of **[switches]** belong to the same group. Ansible can execute the same task on all hosts in the group.

```
[switches]
Device
```

6. Write the playbook.

Create the **vlans.yml** file and write the following code in the file:

```
---
- name: VLAN automation with Ansible on Devices
  hosts: switches
  gather_facts: no
  connection: local

  tasks:
    - name: ensure VLAN 10 exists
      comware_vlan: vlanid=10 name=VLAN10 descr=LOCAL_SEGMENT username={{ username }}
password={{ password }} hostname={{ inventory_hostname }}

    - name: ensure VLAN 20 exists
      comware_vlan: vlanid=20 state=present username={{ username }}
password={{ password }} hostname={{ inventory_hostname }}
```

7. Execute the script.

Grant the permission to execute the **vlans.yml** file.

```
$ sudo chmod 777 vlans.yml
```

Use the **`ansible-playbook`** command to execute the **vlans.yml** file.

```
$ sudo ansible-playbook -i hosts vlans.yml

PLAY [VLAN automation with Ansible on Devices] *********************************

TASK [ensure VLAN 10 exists] ***************************************************
changed: [Device]

TASK [ensure VLAN 20 exists] ***************************************************
ok: [Device]

PLAY RECAP *********************************************************************
Device                     : ok=2    changed=1     unreachable=0    failed=0
```

The output shows that the script created VLAN 10 and detected that VLAN 20 already existed.

Contents

Configuring EPA 1
About EPA 1
Application scenarios 1
Endpoint identification methods 2
Static endpoint identification 2
Restrictions and guidelines: EPA configuration 2
Configuring static endpoint identification 3
Disabling EPA logging 3
Display and maintenance commands for EPA 4
EPA configuration examples 4
Example: Configuring static endpoint identification 4

Configuring EPA

About EPA

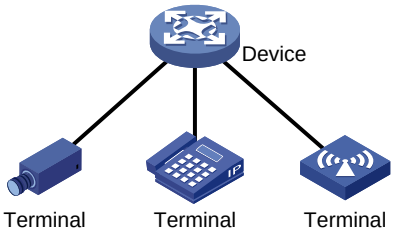
Endpoint Analysis (EPA) allows you to monitor associations and disassociations of endpoints (for example, cameras and IP phones) connecting to an H3C device.

Application scenarios

Non-SmartMC networking

As shown in Figure 1, the device configured with EPA monitors associations and disassociations of endpoints connecting to it. The device can be a standalone device or an IRF fabric. For an IRF fabric, EPA monitors endpoints connecting to all members. The collected association and disassociation information will be reported to the master device for processing.

Figure 1 Non-SmartMC networking

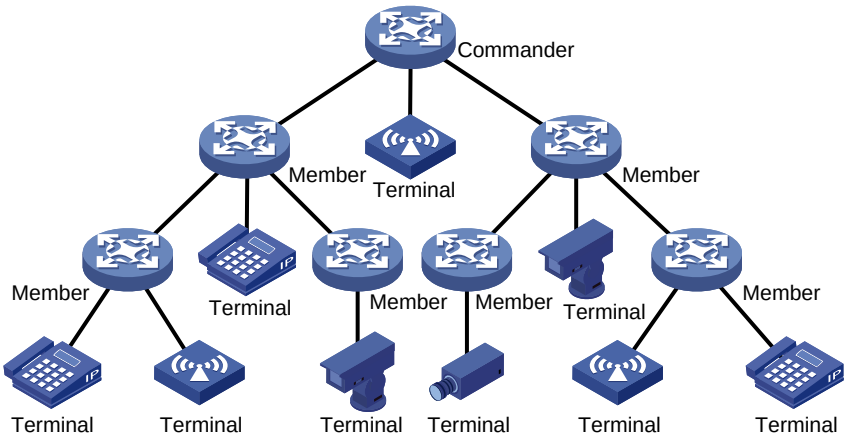


SmartMC networking

In a SmartMC network as shown in Figure 2, EPA settings are configured only on the commander. The members report association and disassociation information about their associated endpoints to the commander. The commander records such information about all associated endpoints in the network, and deploys EPA settings to the members.

You can view EPA information in the entire network on the commander.

Figure 2 SmartMC networking



Endpoint identification methods

EPA supports static endpoint identification. This method identifies associations and disassociations of endpoints by matching MAC addresses of the endpoints. This method requires manual configuration of endpoint MAC addresses on the device and is applicable to networks with a small number of endpoints. This method can only identify cameras and cannot identify the vendor and operating system of cameras.

Static endpoint identification

EPA monitors endpoint associations and disassociations by monitoring the generation and aging of MAC address entries learned by devices. A device configured with EPA can monitor only endpoints in the same subnet as the endpoint access port on the device.

For more information about MAC address entries, see *Layer 2—LAN Switching Configuration Guide*.

Collecting endpoint association information

If a device configured with EPA learns a new MAC address entry, it compares the MAC address and VLAN ID with the configured EPA monitor rules.

- If a match is found, the device determines that a monitored endpoint came online.
  - In a non-SmartMC network, the device records the endpoint association event locally.
  - In a SmartMC network, if the device is the commander, it records the endpoint association event locally. If the device is a member, it reports the event to the commander.
- If no match is found, the device determines that the endpoint is not a monitored endpoint and does not record the endpoint association event.

Collecting endpoint disassociation information

When the MAC address entry of an endpoint ages out, the device determines that the endpoint went offline.

- In a non-SmartMC network, the device records the endpoint disassociation event locally for 7 days.
- In a SmartMC network, the commander records the endpoint disassociation event for 7 days.

Synchronizing information in a SmartMC network

In a SmartMC network, endpoint monitor rules are configured on the commander and deployed to all members by the commander. The rules take effect on both the commander and members. If a member detects an association or disassociation event of a monitored endpoint, it reports the event to the commander for statistics collection and analysis.

You can view endpoint associations and disassociations that occurred in the entire network from the commander. For more information about SmartMC, see "Configuring SmartMC."

Restrictions and guidelines: EPA configuration

Do not use the **mac-address dynamic** command to configure dynamic MAC address entries for monitored endpoints. If you do so, the system might fail to identify endpoint association events.

When you configure EPA in a SmartMC network, follow these restrictions and guidelines:

- Make sure all devices in the network support EPA.
- You can configure endpoint monitor rules only on the commander.
- Configure the same aging time for MAC address entries on all devices in the SmartMC network. Otherwise, endpoint association and disassociation analysis on the commander

might be inaccurate. For more information about MAC address entries, see *Layer 2—LAN Switching Configuration Guide*.

- To view endpoint association and disassociation events in a SmartMC network, execute the **display epa monitor-information** command on the commander instead of a member. If you execute the command on a member, the command displays only association events of endpoints connecting to the member.

Configuring static endpoint identification

Restrictions and guidelines

When you configure static endpoint identification rules, follow these restrictions and guidelines:

- To configure multiple rules to monitor an endpoint in different VLANs, make sure the specified VLAN ranges in these rules do not overlap with each other.
- As a best practice to ensure the optimal EPA performance, specify the VLANs in which an endpoint will be monitored.
- If you configure a rule to monitor an endpoint in all VLANs, make sure the endpoint will not come online from over 10 VLANs.
- You can configure a maximum of 1024 monitor rules. As a best practice to ensure the optimal EPA performance, do not configure over 512 monitor rules.
- You can specify a rule ID when creating a monitor rule. If you do not specify the ID, the system assigns the smallest available ID to the rule.
- You cannot execute the **epa monitor-rule** command multiple times to edit an existing rule. To edit an existing rule, use the **undo epa monitor-rule** command to delete the rule and then create the rule again.

Procedure

1. Enter system view.

   system-view

2. Create a static endpoint identification rule.

   epa monitor-rule [*monitor-rule-id*] **mac** *mac-address* [**mask** *mac-mask*] [**vlan** *vlan-id*]

   By default, no static endpoint identification rules exist.

Disabling EPA logging

About this task

By default, the EPA module logs endpoint associations and disassociations. If a monitored endpoint comes online or goes offline frequently, the device will generate a large number of log entries. In this case, to avoid affecting device performance, disable EPA logging as a best practice.

Procedure

1. Enter system view.

   system-view

2. Disable EPA logging.

   epa online-offline-log disable

   By default, EPA logging is enabled.

Display and maintenance commands for EPA

Execute **display** commands in any view.

Task	Command
Display endpoint association and disassociation information detected by EPA.	**display epa monitor-information** [**online** | **offline**] [**device** *device-id* | **mac** *mac-address* [**vlan** *vlan-id*]] [**verbose**]

EPA configuration examples

Example: Configuring static endpoint identification

Network configuration

As shown in Figure 3, the SmartMC network is established manually. Configure the devices in SmartMC network so EPA can automatically identify camera 1 and camera 2 when they access the SmartMC network.

Figure 3 Network diagram

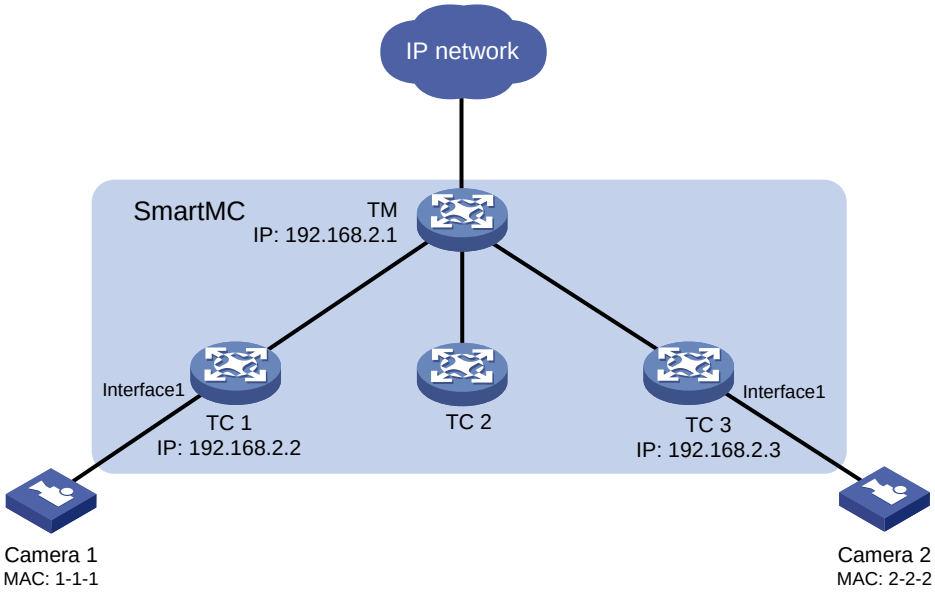


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Configuring the TM

1. Configure SmartMC:

   # Configure VLAN-interface 1.

   ```
   [TM] interface vlan-interface 1
   [TM-Vlan-interface1] ip address 192.168.2.2 24
   ```

```
[TM-Vlan-interface1] quit
```

Enable the Telnet service.

```
[TM] telnet server enable
```

Enable NETCONF over SOAP over HTTP.

```
[TM] netconf soap http enable
```

Enable SSH-based NETCONF over SSH.

```
[TM] netconf ssh server enable
```

Enable LLDP globally.

```
[TM] lldp global enable
```

Create a user. Set the username to **admin** and password to **hello12345**, add the **telnet**, **http**, **https**, and **ssh** service types, and authorize the user to use the **network-admin** user role.

```
[TM] local-user admin
[TM-luser-manage-admin] password simple hello12345
[TM-luser-manage-admin] service-type telnet http https ssh
[TM-luser-manage-admin] authorization-attribute user-role network-admin
[TM-luser-manage-admin] quit
```

Set scheme authentication for VTY user lines 0 to 63.

```
[TM] line vty 0 63
[TM-line-vty0-63] authentication-mode scheme
[TM-line-vty0-63] quit
```

Enable SmartMC, set the device role to commander, and set the username to **admin** and the password (plaintext) to **hello12345**.

```
[TM] smartmc tm username admin password simple hello12345 enable
```

2. Configure static endpoint identification rules on the commander.

```
[TM] epa monitor-rule mac 1-1-1
[TM] epa monitor-rule mac 2-2-2
```

Configuring TC 1

1. Configure SmartMC:

Configure VLAN-interface 1.

```
<TC1> system-view
[TC1] interface vlan-interface 1
[TC1-Vlan-interface1] ip address 192.168.2.1 24
[TC1-Vlan-interface1] quit
```

Enable the Telnet service.

```
[TC1] telnet server enable
```

Enable NETCONF over SOAP over HTTP.

```
[TC1] netconf soap http enable
```

Enable SSH-based NETCONF over SSH.

```
[TC1] netconf ssh server enable
```

Enable LLDP globally.

```
[TC1] lldp global enable
```

Create a user named **admin**.

```
[TC1] local-user admin
```

Lower password complexity requirements. For more information about these commands, see password control commands in *Security Command Reference*.

```
[TC1-luser-manage-admin] password-control length 4
[TC1-luser-manage-admin] password-control composition type-number 1 type-length 1
```

```
[TC1-luser-manage-admin] undo password-control complexity user-name check
```

Set the username and password to **admin**, add the **telnet**, **http**, **https**, and **ssh** service types, and authorize the user to use the **network-admin** user role.

```
[TC1-luser-manage-admin] password simple admin
[TC1-luser-manage-admin] service-type telnet http https ssh
[TC1-luser-manage-admin] authorization-attribute user-role network-admin
[TC1-luser-manage-admin] quit
```

Configure the admin user to support public key authentication when logging in through SSH, and specify the public key name as **-smartmc-publickey-**.

```
[TC1] ssh user admin service-type netconf authentication-type publickey assign publickey -smartmc-publickey-
```

Set scheme authentication for VTY user lines 0 to 63.

```
[TC1] line vty 0 63
[TC1-line-vty0-63] authentication-mode scheme
[TC1-line-vty0-63] quit
```

Enable SmartMC and set the device role to member.

```
[TC1] smartmc tc enable
```

Configuring TC 3

Configure TC 3 in the same way TC 1 is configured. (Details not shown.)

Verifying the configuration

On TC 1, display the online and offline information of endpoints.

```
[TC1] display epa monitor-information verbose
Auto identification : Disabled
Local device type   : SmartMC TC
Local device ID     : 3acc-58e2-0100
Total endpoints identified by EPA: 1

Access device ID: 3acc-58e2-0100
  MAC address   : 0001-0001-0001           VLAN  : 1
  IP address    : -
  Interface     : GigabitEthernet1/0/1     Status: Online
  Category      : Camera                   Vendor: -
  OS            : -
```

On the TM, display the online and offline information of endpoints.

```
[TM] display epa monitor-information verbose
Auto identification : Disabled
Local device type   : SmartMC TM
Local device ID     : 90bc-1b85-0300
Total endpoints identified by EPA: 2

Access device ID: 3acc-58e2-0100
  MAC address   : 0001-0001-0001           VLAN  : 1
  IP address    : 192.168.2.11
  Interface     : GigabitEthernet1/0/1     Status: Online
  Category      : Camera                   Vendor: -
  OS            : -
```

```
Access device ID: 3acc-58e2-0300
  MAC address   : 0002-0002-0002           VLAN  : 1
  IP address    : 192.168.2.12
  Interface     : GigabitEthernet1/0/1     Status: Online
  Category      : Camera                   Vendor: -
  OS            : -
```

Contents

Configuring CWMP 1

About CWMP 1
CWMP network framework 1
Basic CWMP functions 1
How CWMP works 3
Restrictions and guidelines: CWMP configuration 5
CWMP tasks at a glance 5
Enabling CWMP from the CLI 6
Configuring ACS attributes 6
About ACS attributes 6
Configuring the preferred ACS attributes 6
Configuring the default ACS attributes from the CLI 7
Configuring CPE attributes 8
About CPE attributes 8
Specifying an SSL client policy for HTTPS connection to ACS 8
Configuring ACS authentication parameters 8
Configuring the provision code 9
Configuring the CWMP connection interface 9
Configuring autoconnect parameters 10
Setting the close-wait timer 11
Enabling NAT traversal for the CPE 11
Display and maintenance commands for CWMP 11
CWMP configuration examples 12
Example: Configuring CWMP 12
Restrictions and guidelines: CWMP ACS configuration 1
Configuring ACS authentication parameters 1
About CWMP CPEs 1

Configuring CWMP

About CWMP

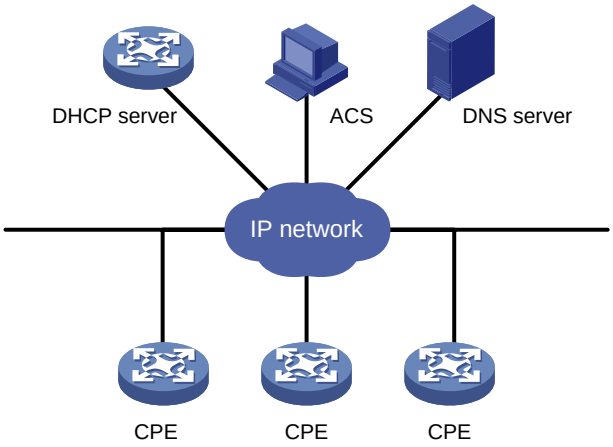
CPE WAN Management Protocol (CWMP), also called "TR-069," is a DSL Forum technical specification for remote management of network devices.

The protocol was initially designed to provide remote autoconfiguration through a server for large numbers of dispersed end-user devices in a network. CWMP can be used on different types of networks, including Ethernet.

CWMP network framework

Figure 1 shows a basic CWMP network framework.

Figure 1 CWMP network framework



A basic CWMP network includes the following network elements:

- **ACS**—Autoconfiguration server, the management device in the network.
- **CPE**—Customer premises equipment, the managed device in the network.
- **DNS server**—Domain name system server. CWMP defines that the ACS and the CPE use URLs to identify and access each other. DNS is used to resolve the URLs.
- **DHCP server**—Assigns ACS attributes along with IP addresses to CPEs when the CPEs are powered on. DHCP server is optional in CWMP. With a DHCP server, you do not need to configure ACS attributes manually on each CPE. The CPEs can contact the ACS automatically when they are powered on for the first time.

The device is operating as a CPE in the CWMP framework.

Basic CWMP functions

You can autoconfigure and upgrade CPEs in bulk from the ACS.

Autoconfiguration

You can create configuration files for different categories of CPEs on the ACS. Based on the device models and serial numbers of the CPEs, the ACS verifies the categories of the CPEs and issues the associated configuration to them.

The following are methods available for the ACS to issue configuration to the CPE:

- Transfers the configuration file to the CPE, and specifies the file as the next-startup configuration file. At a reboot, the CPE starts up with the ACS-specified configuration file.
- Runs the configuration in the CPE's RAM. The configuration takes effect immediately on the CPE. For the running configuration to survive a reboot, you must save the configuration on the CPE.

CPE software management

The ACS can manage CPE software upgrade.

When the ACS finds a software version update, the ACS notifies the CPE to download the software image file from a specific location. The location can be the URL of the ACS or an independent file server.

If the CPE successfully downloads the software image file and the file is validated, the CPE notifies the ACS of a successful download. If the CPE fails to download the software image file or the file is invalidated, the CPE notifies the ACS of an unsuccessful download.

Data backup

The ACS can require the CPE to upload a configuration file or log file to a specific location. The destination location can be the ACS or a file server.

CPE status and performance monitoring

The ACS can monitor the status and performance of CPEs. Table 1 shows the available CPE status and performance objects for the ACS to monitor.

Table 1 CPE status and performance objects available for the ACS to monitor

Category	Objects	Remarks
Device information	Manufacturer ManufacturerOUI SerialNumber HardwareVersion SoftwareVersion	N/A
Operating status and information	DeviceStatus UpTime	N/A
Configuration file	ConfigFile	Local configuration file stored on CPE for upgrade. The ACS can issue configuration to the CPE by transferring a configuration file to the CPE or running the configuration in CPE's RAM.
CWMP settings	ACS URL	URL address of the ACS to which the CPE initiates a CWMP connection. This object is also used for main/backup ACS switchover.
	ACS username ACS password	When the username and password of the ACS are changed, the ACS changes the ACS username and password on the CPE to the new username and password. When a main/backup ACS switchover occurs, the main ACS also changes the ACS username and password to the backup ACS username and password.
	PeriodicInformEnable	Whether to enable or disable the periodic Inform feature.
	PeriodicInformInterval	Interval for periodic connection from the CPE

Category	Objects	Remarks
		to the ACS for configuration and software update.
	PeriodicInformTime	Scheduled time for connection from the CPE to the ACS for configuration and software update.
	ConnectionRequestURL (CPE URL)	N/A
	ConnectionRequestUsername (CPE username) ConnectionRequestPassword (CPE password)	CPE username and password for authentication from the ACS to the CPE.

How CWMP works

RPC methods

CWMP uses remote procedure call (RPC) methods for bidirectional communication between CPE and ACS. The RPC methods are encapsulated in HTTP or HTTPS.

Table 2 shows the primary RPC methods used in CWMP.

Table 2 RPC methods

RPC method	Description
Get	The ACS obtains the values of parameters on the CPE.
Set	The ACS modifies the values of parameters on the CPE.
Inform	The CPE sends an Inform message to the ACS for the following purposes: • Initiates a connection to the ACS. • Reports configuration changes to the ACS. • Periodically updates CPE settings to the ACS.
Download	The ACS requires the CPE to download a configuration or software image file from a specific URL for software or configuration update.
Upload	The ACS requires the CPE to upload a file to a specific URL.
Reboot	The ACS reboots the CPE remotely for the CPE to complete an upgrade or recover from an error condition.

Autoconnect between ACS and CPE

The CPE automatically initiates a connection to the ACS when one of the following events occurs:

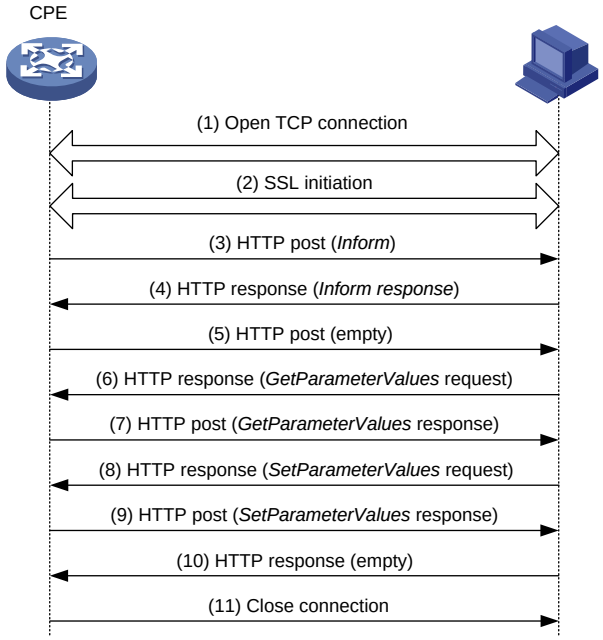
- ACS URL change. The CPE initiates a connection request to the new ACS URL.
- CPE startup. The CPE initiates a connection to the ACS after the startup.
- Timeout of the periodic Inform interval. The CPE re-initiates a connection to the ACS at the Inform interval.
- Expiration of the scheduled connection initiation time. The CPE initiates a connection to the ACS at the scheduled time.

CWMP connection establishment

Step 1 through step 5 in Figure 2 show the procedure of establishing a connection between the CPE and the ACS.

1. After obtaining the basic ACS parameters, the CPE initiates a TCP connection to the ACS.
2. If HTTPS is used, the CPE and the ACS initialize SSL for a secure HTTP connection.
3. The CPE sends an Inform message in HTTPS to initiate a CWMP session.
4. After the CPE passes authentication, the ACS returns an Inform response to establish the session.
5. After sending all requests, the CPE sends an empty HTTP post message.

Figure 2 CWMP connection establishment

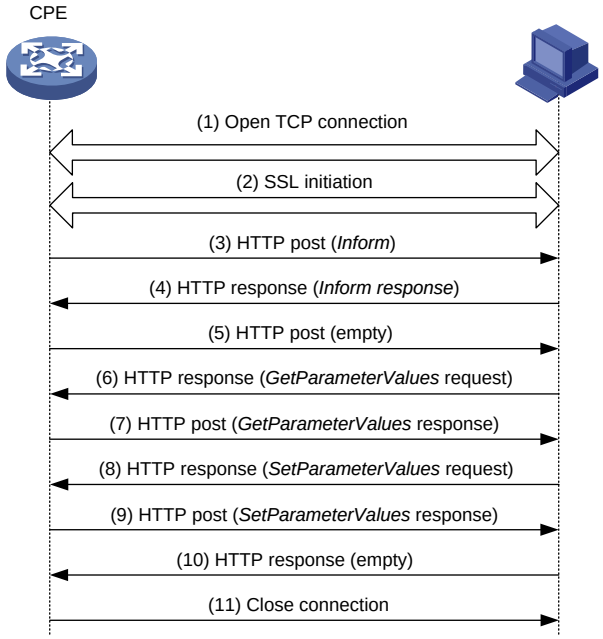


Main/backup ACS switchover

Typically, two ACSs are used in a CWMP network for consecutive monitoring on CPEs. When the main ACS needs to reboot, it points the CPE to the backup ACS. Step 6 through step 11 in Figure 3 show the procedure of a main/backup ACS switchover.

1. Before the main ACS reboots, it queries the ACS URL set on the CPE.
2. The CPE replies with its ACS URL setting.
3. The main ACS sends a Set request to change the ACS URL on the CPE to the backup ACS URL.
4. After the ACS URL is modified, the CPE sends a response.
5. The main ACS sends an empty HTTP message to notify the CPE that it has no other requests.
6. The CPE closes the connection, and then initiates a new connection to the backup ACS URL.

Figure 3 Main and backup ACS switchover



Restrictions and guidelines: CWMP configuration

CWMP is available only in IPv4 networks. If you configure CWMP in an IPv6 network, the feature does not take effect.

You can configure ACS and CPE attributes from the CPE's CLI, the DHCP server, or the ACS. For an attribute, the CLI- and ACS-assigned values have higher priority than the DHCP-assigned value. The CLI- and ACS-assigned values overwrite each other, whichever is assigned later.

This document only describes configuring ACS and CPE attributes from the CLI and DHCP server. For more information about configuring and using the ACS, see ACS documentation.

CWMP tasks at a glance

To configure CWMP, perform the following tasks:

1. Enabling CWMP from the CLI

   You can also enable CWMP from a DHCP server.
2. Configuring ACS attributes
   a. Configuring the preferred ACS attributes
   b. (Optional.) Configuring the default ACS attributes from the CLI
3. Configuring CPE attributes
   a. Specifying an SSL client policy for HTTPS connection to ACS

      This task is required when the ACS uses HTTPS for secure access.
   b. (Optional.) Configuring ACS authentication parameters
   c. (Optional.) Configuring the provision code
   d. (Optional.) Configuring the CWMP connection interface
   e. (Optional.) Configuring autoconnect parameters
   f. (Optional.) Setting the close-wait timer

g. (Optional.) Enabling NAT traversal for the CPE

Enabling CWMP from the CLI

1. Enter system view.
   system-view
2. Enter CWMP view.
   cwmp
3. Enable CWMP.
   cwmp enable
   By default, CWMP is disabled.

Configuring ACS attributes

About ACS attributes

You can configure two sets of ACS attributes for the CPE: preferred and default.

- The preferred ACS attributes are configurable from the CPE's CLI, the DHCP server, and ACS.
- The default ACS attributes are configurable only from the CLI.

If the preferred ACS attributes are not configured, the CPE uses the default ACS attributes for connection establishment.

Configuring the preferred ACS attributes

Assigning ACS attributes from the DHCP server

The DHCP server in a CWMP network assigns the following information to CPEs:

- IP addresses for the CPEs.
- DNS server address.
- ACS URL and ACS login authentication information.

This section introduces how to use DHCP option 43 to assign the ACS URL and ACS login authentication username and password. For more information about DHCP and DNS, see *Layer 3—IP Services Configuration Guide*.

If the DHCP server is an H3C device, you can configure DHCP option 43 by using the **option 43 hex 01***length URL username password* command.

- *length*—A hexadecimal number that indicates the total length of the *length*, *URL*, *username*, and *password* arguments, including the spaces between these arguments. No space is allowed between the **01** keyword and the length value.
- *URL*—ACS URL.
- *username*—Username for the CPE to authenticate to the ACS.
- *password*—Password for the CPE to authenticate to the ACS.

NOTE:

The ACS URL, username and password must use the hexadecimal format and be space separated.

The following example configures the ACS address as **https://169.254.76.31:7547/acs**, username as **1234**, and password as **5678**:

```
<Sysname> system-view
[Sysname] dhcp server ip-pool 0
[Sysname-dhcp-pool-0] option 43 hex
012368747470733A2F2F3136392E3235342E37362E33313A3735343720313233342035363738
```

Table 3 Hexadecimal forms of the ACS attributes

Attribute	Attribute value	Hexadecimal form
Length	39 characters	27
ACS URL	https://169.254.76.31:7547/acs	68747470733A2F2F3136392E3235342E37362E33313A3735343720 **NOTE:** The two ending digits (20) represent the space.
ACS connect username	1234	3132333420 **NOTE:** The two ending digits (20) represent the space.
ACS connect password	5678	35363738

Configuring the preferred ACS attributes from the CLI

1. Enter system view.
   system-view
2. Enter CWMP view.
   cwmp
3. Configure the preferred ACS URL.
   cwmp acs url *url*
   By default, no preferred ACS URL has been configured.
4. Configure the username for authentication to the preferred ACS URL.
   cwmp acs username *username*
   By default, no username has been configured for authentication to the preferred ACS URL.
5. (Optional.) Configure the password for authentication to the preferred ACS URL.
   cwmp acs password { **cipher** | **simple** } *string*
   By default, no password has been configured for authentication to the preferred ACS URL.

Configuring the default ACS attributes from the CLI

1. Enter system view.
   system-view
2. Enter CWMP view.
   cwmp
3. Configure the default ACS URL.
   cwmp acs default url *url*
   By default, no default ACS URL has been configured.
4. Configure the username for authentication to the default ACS URL.
   cwmp acs default username *username*
   By default, no username has been configured for authentication to the default ACS URL.

5. (Optional.) Configure the password for authentication to the default ACS URL.

   cwmp acs default password { **cipher** | **simple** } *string*

   By default, no password has been configured for authentication to the default ACS URL.

Configuring CPE attributes

About CPE attributes

You can configure the following CPE attributes only from the CPE's CLI.

- CWMP connection interface.
- NAT traversal.
- Maximum number of connection retries.
- SSL client policy for HTTPS connection to ACS.

For other CPE attribute values, you can assign them to the CPE from the CPE's CLI or the ACS. The CLI- and ACS-assigned values overwrite each other, whichever is assigned later.

Specifying an SSL client policy for HTTPS connection to ACS

About this task

This task is required when the ACS uses HTTPS for secure access. CWMP uses HTTP or HTTPS for data transmission. When HTTPS is used, the ACS URL begins with **https://**. You must specify an SSL client policy for the CPE to authenticate the ACS for HTTPS connection establishment.

Prerequisites

Before you perform this task, first create an SSL client policy. For more information about configuring SSL client policies, see *Security Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter CWMP view.

   cwmp

3. Specify an SSL client policy.

   ssl client-policy *policy-name*

   By default, no SSL client policy is specified.

Configuring ACS authentication parameters

About this task

To protect the CPE against unauthorized access, configure a CPE username and password for ACS authentication. When an ACS initiates a connection to the CPE, the ACS must provide the correct username and password.

Procedure

1. Enter system view.

   system-view

2. Enter CWMP view.

cwmp

3. Configure the username for authentication to the CPE.

   cwmp cpe username *username*

   By default, no username has been configured for authentication to the CPE.

4. (Optional.) Configure the password for authentication to the CPE.

   cwmp cpe password { **cipher** | **simple** } *string*

   By default, no password has been configured for authentication to the CPE.

   The password setting is optional. You can specify only a username for authentication.

Configuring the provision code

About this task

The ACS can use the provision code to identify services assigned to each CPE. For correct configuration deployment, make sure the same provision code is configured on the CPE and the ACS. For information about the support of your ACS for provision codes, see the ACS documentation.

Procedure

1. Enter system view.

   system-view

2. Enter CWMP view.

   cwmp

3. Configure the provision code.

   cwmp cpe provision-code *provision-code*

   The default provision code is **PROVISIONINGCODE**.

Configuring the CWMP connection interface

About this task

The CWMP connection interface is the interface that the CPE uses to communicate with the ACS. To establish a CWMP connection, the CPE sends the IP address of this interface in the Inform messages, and the ACS replies to this IP address.

Typically, the CPE selects the CWMP connection interface automatically. If the CWMP connection interface is not the interface that connects the CPE to the ACS, the CPE fails to establish a CWMP connection with the ACS. In this case, you need to manually set the CWMP connection interface.

Procedure

1. Enter system view.

   system-view

2. Enter CWMP view.

   cwmp

3. Specify the interface that connects to the ACS as the CWMP connection interface.

   cwmp cpe connect interface *interface-type interface-number*

   By default, no CWMP connection interface is specified.

Configuring autoconnect parameters

About this task

You can configure the CPE to connect to the ACS periodically, or at a scheduled time for configuration or software update.

The CPE retries a connection automatically when one of the following events occurs:

- The CPE fails to connect to the ACS. The CPE considers a connection attempt as having failed when the close-wait timer expires. This timer starts when the CPE sends an Inform request. If the CPE fails to receive a response before the timer expires, the CPE resends the Inform request.
- The connection is disconnected before the session on the connection is completed.

To protect system resources, limit the number of retries that the CPE can make to connect to the ACS.

Configuring the periodic Inform feature

1. Enter system view.
   `system-view`
2. Enter CWMP view.
   `cwmp`
3. Enable the periodic Inform feature.
   `cwmp cpe inform interval enable`
   By default, this function is disabled.
4. Set the Inform interval.
   `cwmp cpe inform interval` *`interval`*
   By default, the CPE sends an Inform message to start a session every 600 seconds.

Scheduling a connection initiation

1. Enter system view.
   `system-view`
2. Enter CWMP view.
   `cwmp`
3. Schedule a connection initiation.
   `cwmp cpe inform time` *`time`*
   By default, no connection initiation has been scheduled.

Setting the maximum number of connection retries

1. Enter system view.
   `system-view`
2. Enter CWMP view.
   `cwmp`
3. Set the maximum number of connection retries.
   `cwmp cpe connect retry` *`retries`*
   By default, the CPE retries a failed connection until the connection is established.

Setting the close-wait timer

About this task

The close-wait timer specifies the following:

- The maximum amount of time the CPE waits for the response to a session request. The CPE determines that its session attempt has failed when the timer expires.
- The amount of time the connection to the ACS can be idle before it is terminated. The CPE terminates the connection to the ACS if no traffic is sent or received before the timer expires.

Procedure

1. Enter system view.
   `system-view`
2. Enter CWMP view.
   `cwmp`
3. Set the close-wait timer.
   `cwmp cpe wait timeout` *`seconds`*
   By default, the close-wait timer is 30 seconds.

Enabling NAT traversal for the CPE

About this task

For the connection request initiated from the ACS to reach the CPE, you must enable NAT traversal on the CPE when a NAT gateway resides between the CPE and the ACS.

The NAT traversal feature complies with RFC 3489 Simple Traversal of UDP Through NATs (STUN). The feature enables the CPE to discover the NAT gateway, and obtain an open NAT binding (a public IP address and port binding) through which the ACS can send unsolicited packets. The CPE sends the binding to the ACS when it initiates a connection to the ACS. For the connection requests sent by the ACS at any time to reach the CPE, the CPE maintains the open NAT binding.

Procedure

1. Enter system view.
   `system-view`
2. Enter CWMP view.
   `cwmp`
3. Enable NAT traversal.
   `cwmp cpe stun enable`
   By default, NAT traversal is disabled on the CPE.

Display and maintenance commands for CWMP

Execute **`display`** commands in any view.

Task	Command
Display CWMP configuration.	**`display cwmp configuration`**
Display the current status of CWMP.	**`display cwmp status`**

CWMP configuration examples

Example: Configuring CWMP

Network configuration

As shown in Figure 4, use H3C IMC BIMS as the ACS to bulk-configure the devices (CPEs), and assign ACS attributes to the CPEs from the DHCP server.

The configuration files for the CPEs in equipment rooms A and B are **configure1.cfg** and **configure2.cfg**, respectively.

Figure 4 Network diagram

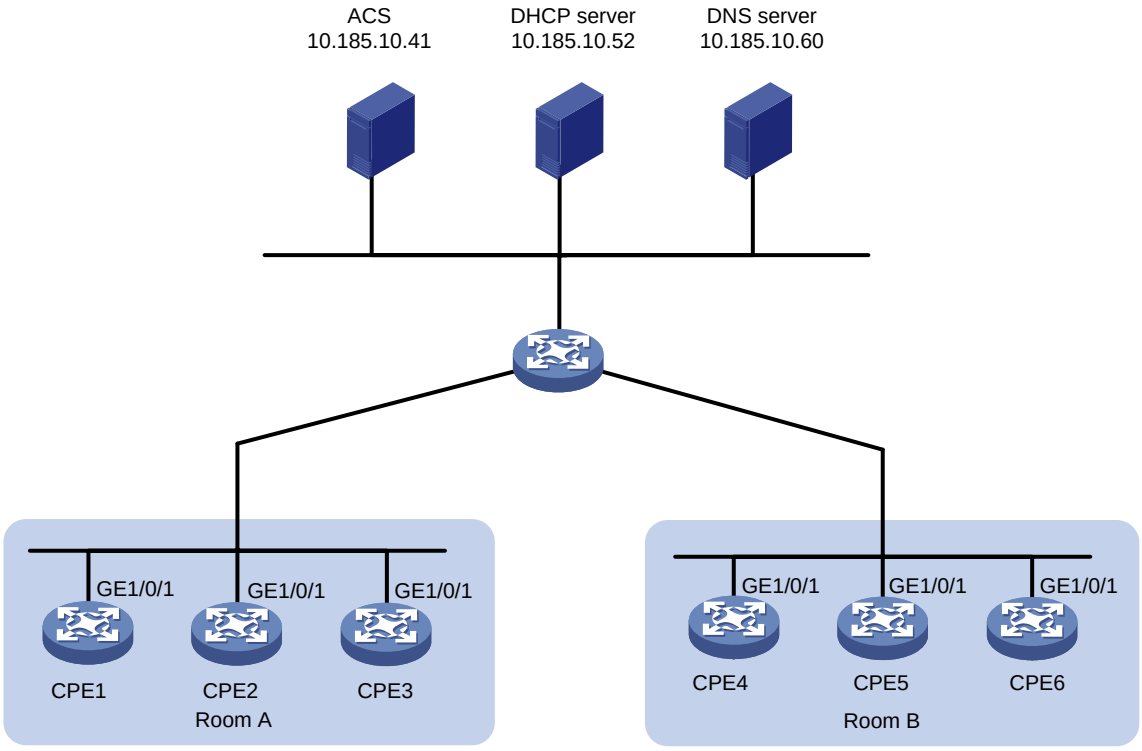


Table 4 shows the ACS attributes for the CPEs to connect to the ACS.

Table 4 ACS attributes

Item	Setting
Preferred ACS URL	https://10.185.10.41:9090
ACS username	admin
ACS password	12345

Table 5 lists serial numbers of the CPEs.

Table 5 CPE list

Room	Device	Serial number
A	CPE 1	210231A95YH10C000045
	CPE 2	210235AOLNH12000010

Room	Device	Serial number
	CPE 3	210235AOLNH12000015
B	CPE 4	210235AOLNH12000017
	CPE 5	210235AOLNH12000020
	CPE 6	210235AOLNH12000022

Configuring the ACS

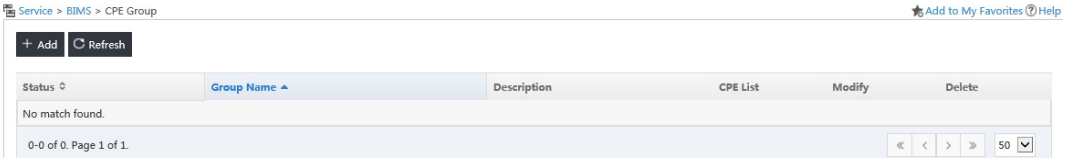
Figures in this section are for illustration only.

To configure the ACS:

1. Log in to the ACS:
   a. Launch a Web browser on the ACS configuration terminal.
   b. In the address bar of the Web browser, enter the ACS URL and port number. This example uses **https://10.185.10.41:8080/imc**.
   c. On the login page, enter the ACS login username and password, and then click **Login**.
2. Create a CPE group for each equipment room:
   a. Select **Service** > **BIMS** > **CPE Group** from the top navigation bar.

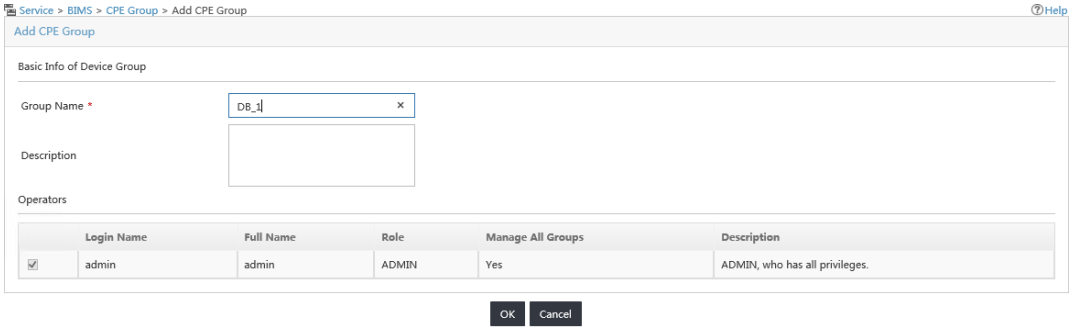
      The **CPE Group** page appears.

      Figure 5 CPE Group page

   b. Click **Add**.
   c. Enter a username, and then click **OK**.

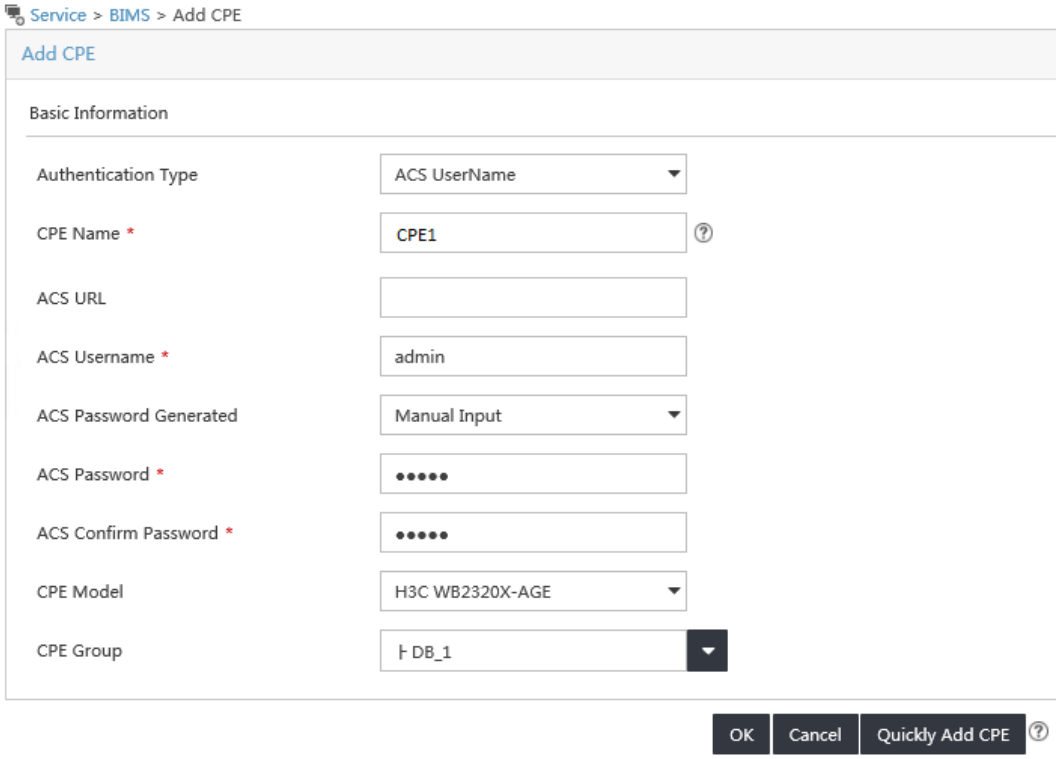
      Figure 6 Adding a CPE group

   d. Repeat the previous two steps to create a CPE group for CPEs in Room B.
3. Add CPEs to the CPE group for each equipment room:
   a. Select **Service** > **BIMS** > **Resource Management** > **Add CPE** from the top navigation bar.
   b. On the **Add CPE** page, configure the following parameters:
      - **Authentication Type**—Select **ACS UserName**.
      - **CPE Name**—Enter a CPE name.
      - **ACS Username**—Enter **admin**.
      - **ACS Password Generated**—Select **Manual Input**.

- **ACS Password**—Enter a password for ACS authentication.
- **ACS Confirm Password**—Re-enter the password.
- **CPE Model**—Select the CPE model.
- **CPE Group**—Select the CPE group.

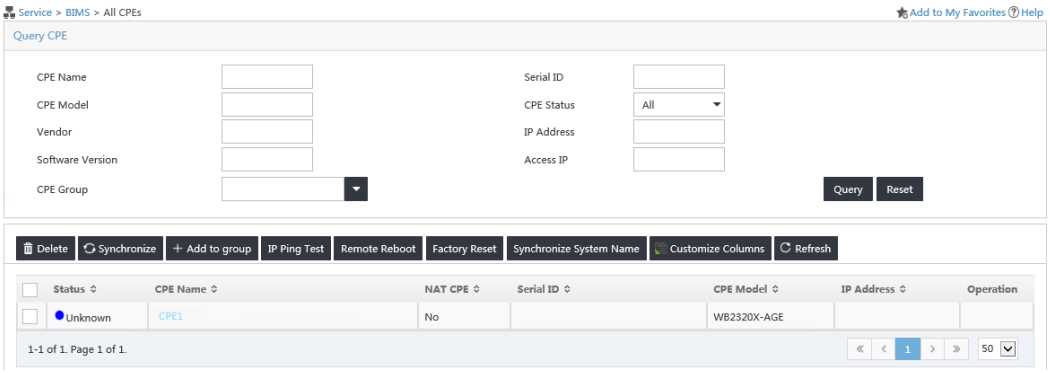
Figure 7 Adding a CPE



c. Click **OK**.

d. Verify that the CPE has been added successfully from the **All CPEs** page.

Figure 8 Viewing CPEs

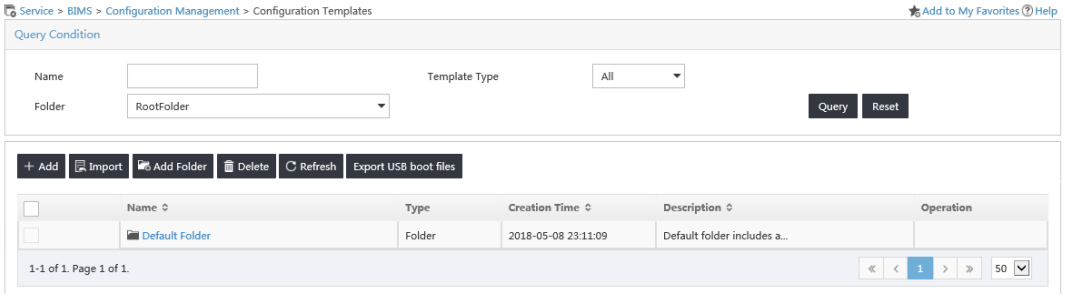


e. Repeat the previous steps to add CPE 2 and CPE 3 to the CPE group for Room A, and add CPEs in Room B to the CPE group for Room B.

4. Configure a configuration template for each equipment room:

a. Select **Service** > **BIMS** > **Configuration Management** > **Configuration Templates** from the top navigation bar.

Figure 9 Configuration Templates page



b. Click **Import**.

c. Select a source configuration file, select **Configuration Segment** as the template type, and then click **OK**.

The created configuration template will be displayed in the **Configuration Template** list after a successful file import.

IMPORTANT:

If the first command in the configuration template file is **`system-view`**, make sure no characters exist in front of the command.

Figure 10 Importing a configuration template

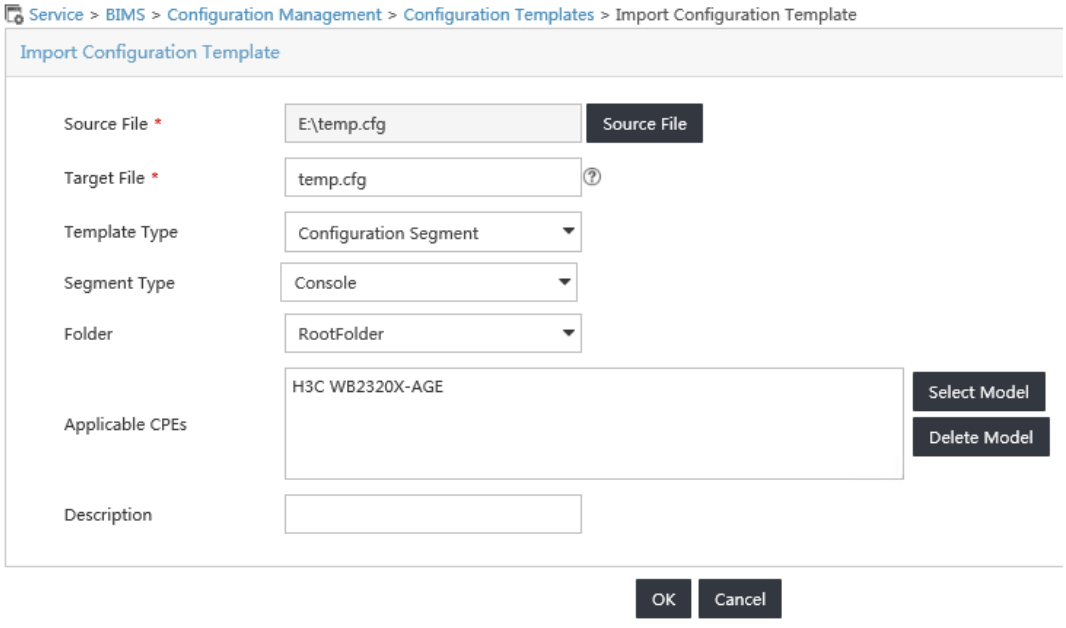
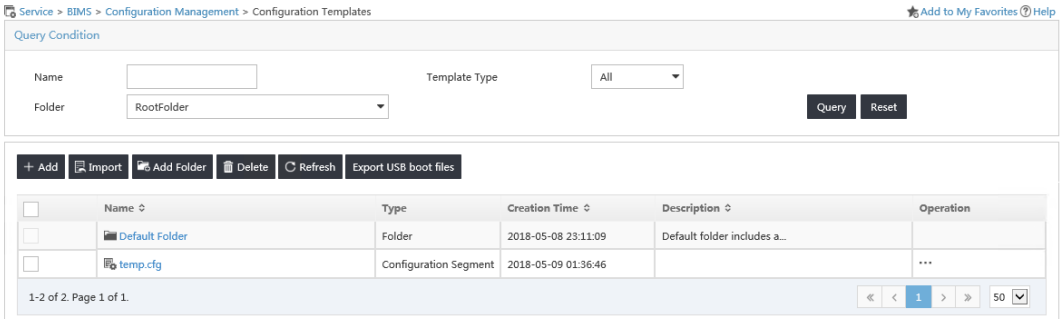


Figure 11 Configuration Template list

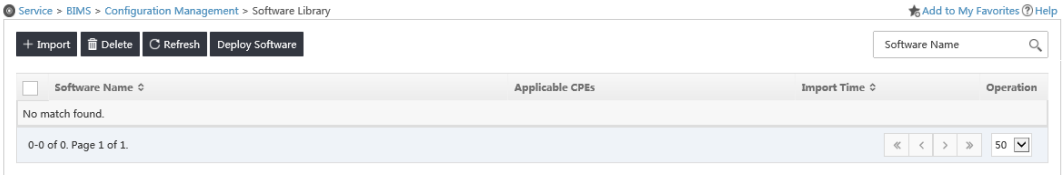


d. Repeat the previous steps to configure a configuration template for Room B.

5. Add software library entries:

a. Select **Service** > **BIMS** > **Configuration Management** > **Software Library** from the top navigation bar.

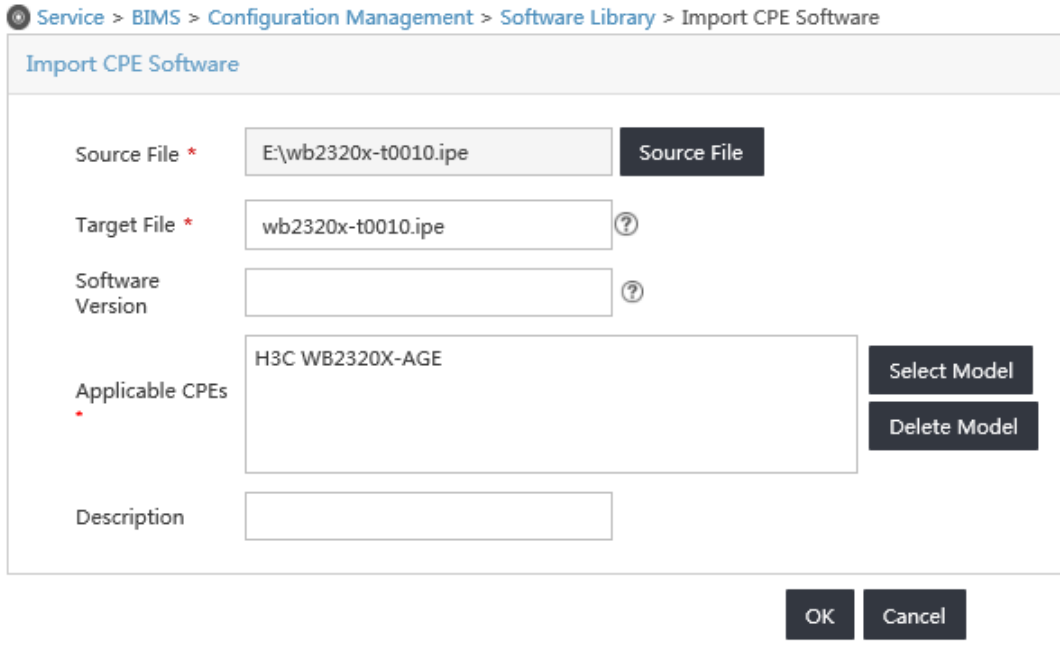
Figure 12 Software Library page



b. Click **Import**.

c. Select a source file, and then click **OK**.

Figure 13 Importing CPE software

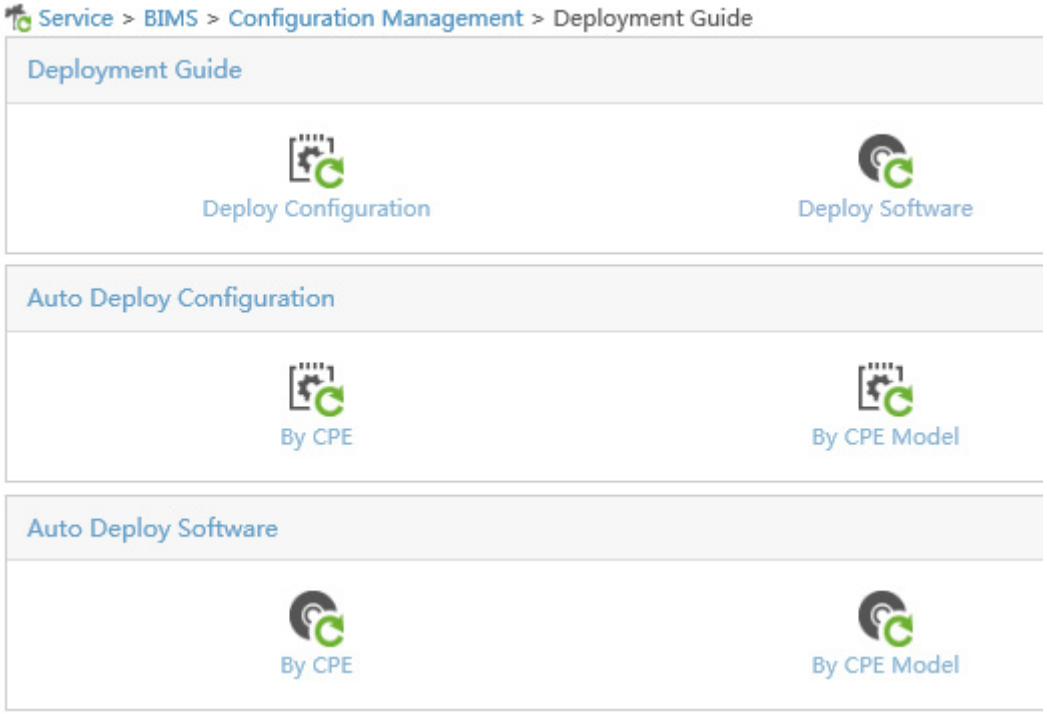


d. Repeat the previous steps to add software library entries for CPEs of different models.

6. Create an auto-deployment task for each equipment room:

a. Select **Service** > **BIMS** > **Configuration Management** > **Deployment Guide** from the top navigation bar.

Figure 14 Deployment Guide

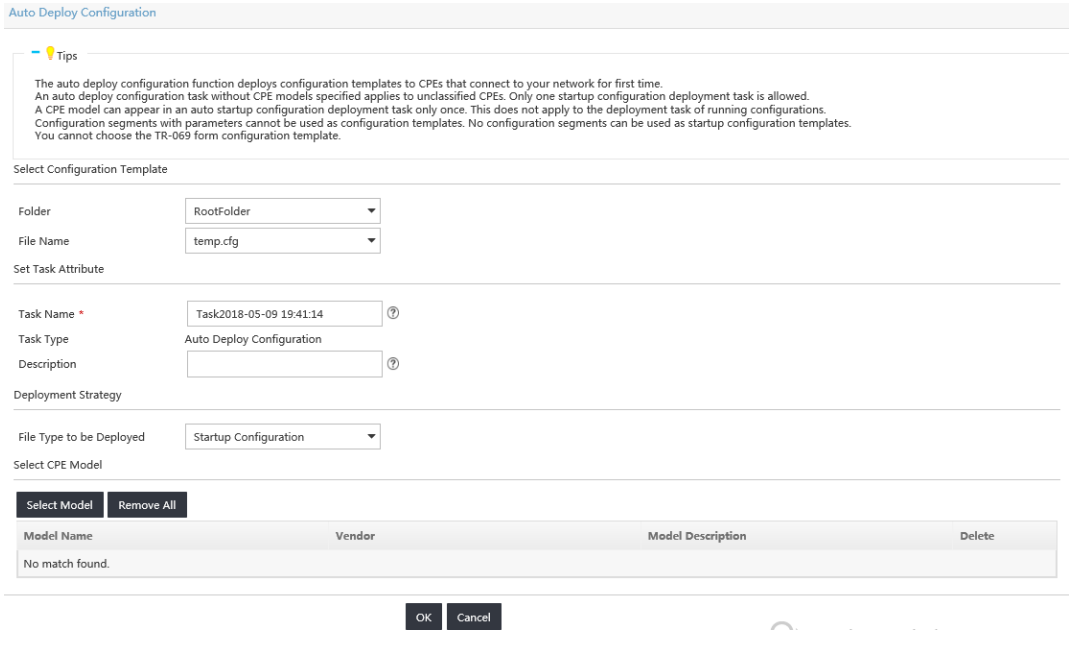


b. Click **By CPE Model** from the **Auto Deployment Configuration** field.

c. Select a configuration template, select **Startup Configuration** from the **File Type to be Deployed** list, and click **Select Model** to select CPEs in Room A. Then, click **OK**.

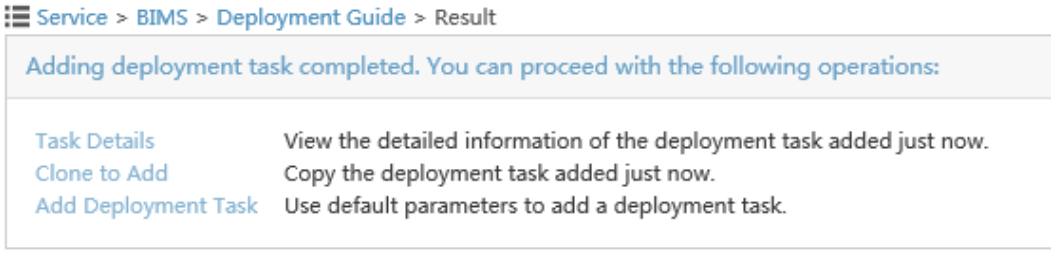
You can search for CPEs by CPE group.

Figure 15 Auto deployment configuration



d. Click **OK** on the **Auto Deploy Configuration** page.

Figure 16 Operation result



e. Repeat the previous steps to add a deployment task for CPEs in Room B.

Configuring the DHCP server

In this example, an H3C device is operating as the DHCP server.

1. Configure an IP address pool to assign IP addresses and DNS server address to the CPEs. This example uses subnet 10.185.10.0/24 for IP address assignment.

Enable DHCP.

```
<DHCP_server> system-view
[DHCP_server] dhcp enable
```

Enable DHCP server on VLAN-interface 1.

```
[DHCP_server] interface vlan-interface 1
[DHCP_server-Vlan-interface1] dhcp select server
[DHCP_server-Vlan-interface1] quit
```

Exclude the DNS server address 10.185.10.60 and the ACS IP address 10.185.10.41 from dynamic allocation.

```
[DHCP_server] dhcp server forbidden-ip 10.185.10.41
[DHCP_server] dhcp server forbidden-ip 10.185.10.60
```

Create DHCP address pool 0.

```
[DHCP_server] dhcp server ip-pool 0
```

Assign subnet 10.185.10.0/24 to the address pool, and specify the DNS server address 10.185.10.60 in the address pool.

```
[DHCP_server-dhcp-pool-0] network 10.185.10.0 mask 255.255.255.0
[DHCP_server-dhcp-pool-0] dns-list 10.185.10.60
```

2. Configure DHCP Option 43 to contain the ACS URL, username, and password in hexadecimal format.

```
[DHCP_server-dhcp-pool-0] option 43 hex
013B68747470733A2F2F6163732E6461746162617365 3A393039302F6163732076696 36B7920313233
33435
```

Configuring the DNS server

Map https://acs.database:9090 to https://10.185.1.41:9090 on the DNS server. For more information about DNS configuration, see DNS server documentation.

Connecting the CPEs to the network

Connect CPE 1 to the network, and then power on the CPE. (Details not shown.)

Log in to CPE 1 and configure its VLAN-interface 2 to use DHCP for IP address acquisition and assign interface GigabitEthernet 1/0/1 to VLAN 2.

```
<CPE1> system-view
[CPE1] vlan 2
[CPE1-vlan2] quit
```

```
[CPE1] interface gigabitethernet 1/0/1
[CPE1-GigabitEthernet1/0/1] port access vlan 2
[CPE1-GigabitEthernet1/0/1] quit
[CPE1] interface vlan-interface 2
[CPE1-Vlan-interface2] ip address dhcp-alloc
```

Repeat the previous steps to configure the other CPEs.

Verifying the configuration

Execute the **display current-configuration** command to verify that the running configurations on CPEs are the same as the configurations issued by the ACS.

Restrictions and guidelines: CWMP ACS configuration

To avoid ACS and CPE conflicts, do not configure both ACS and CPE features on the same device.

Configuring ACS authentication parameters

About this task

Application scenarios

CWMP supports mutual authentication. When the ACS initiates a connection with a CPE, the CPE can authenticate the ACS. When a CPE initiates a connection with the ACS, the ACS can authenticate the CPE. Perform this task to configure CPEs to authenticate the ACS.

Operating mechanism

When the ACS initiates a CWMP connection with a CPE, the ACS sends the username or username + password to the CPE for authentication. The CPE verifies whether the ACS is legal based on the configured username and password in this feature. If the authentication succeeds, the CWMP connection is established. If the authentication fails, the CWMP connection cannot be established.

Restrictions and guidelines

By default, CPEs do not authenticate the ACS. For the CPE to successfully authenticate the ACS, make sure the ACS and CPEs are configured with the same authentication parameters.

Procedure

1. Enter system view.
   system-view
2. Enter CWMP view.
   cwmp
3. Configure the username for the CPE to authenticate the ACS.
   cwmp cpe username *username*
   By default, no username has been configured for a CPE to authenticate the ACS.
4. (Optional.) Configure the password for the CPE to authenticate the ACS.
   cwmp cpe password { **cipher** | **simple** } *string*
   By default, no password has been configured for a CPE to authenticate the ACS.
   The password setting is optional. You can specify only a username for authentication.

About CWMP CPEs

After you configure a device as a CWMP CPE, the device automatically initiates a CWMP connection to the ACS. The ACS can manage the CPE through the connection.

Contents

Configuring EAA 1

About EAA 1
EAA framework 1
Elements in a monitor policy 2
EAA environment variables 3
Restrictions and guidelines 4
EAA tasks at a glance 4
Configuring a user-defined EAA environment variable 5
Configuring a monitor policy from the CLI 5
Tasks at a glance 5
Creating a CLI-defined policy 5
Setting the size for the EAA-monitored log buffer 5
Configuring automatic email sending parameters 6
Configuring an event 7
Configuring an action 8
Assigning a user role to a CLI-defined policy 8
Configuring the policy action runtime 9
Enabling the CLI-defined monitor policy 9
Configuring a monitor policy by using Tcl 9
Configuring a monitor policy by using Python 10
Suspending monitor policies 11
Display and maintenance commands for EAA 11
EAA configuration examples 12
Example: Configuring a CLI event monitor policy by using Tcl 12
Example: Configuring a CLI event monitor policy from the CLI 13
Configuring an automatic email sending action for a CLI-defined monitor policy 14
Example: Configuring a track event monitor policy from the CLI 16
Example: Configuring a CLI event monitor policy with EAA environment variables from the CLI 17

Configuring EAA

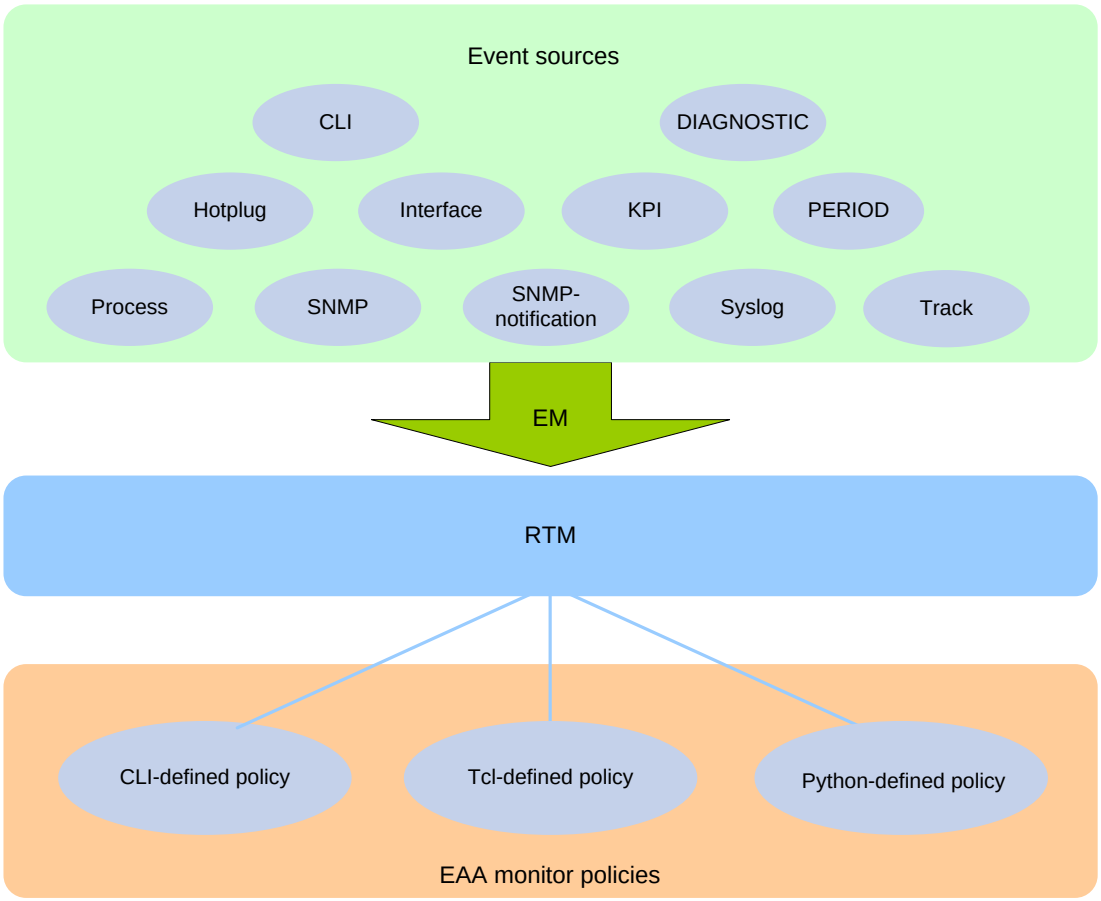
About EAA

Embedded Automation Architecture (EAA) is a monitoring framework that enables you to self-define monitored events and actions to take in response to an event. It allows you to create monitor policies by using the CLI or Tcl scripts.

EAA framework

EAA framework includes a set of event sources, a set of event monitors, a real-time event manager (RTM), and a set of user-defined monitor policies, as shown in Figure 1.

Figure 1 EAA framework



Event sources

Event sources are software or hardware modules that trigger events (see EAA framework).

For example, the CLI module triggers an event when you enter a command. The Syslog module (the information center) triggers an event when it receives a log message.

Event monitors

EAA creates one event monitor to monitor the system for the event specified in each monitor policy. An event monitor notifies the RTM to run the monitor policy when the monitored event occurs.

RTM

RTM manages the creation, state machine, and execution of monitor policies.

EAA monitor policies

A monitor policy specifies the event to monitor and actions to take when the event occurs.

The following types of EAA monitor policies are available:

- **CLI-defined EAA monitor policy**—Policy configured from the CLI.
- **Tcl-defined EAA monitor policy**—Policy defined by a Tcl script.
- **Python-defined EAA monitor policy**—Policy defined by a Python script.

Elements in a monitor policy

Elements in an EAA monitor policy include event, action, user role, and runtime.

Event

Table 1 shows types of events that EAA can monitor.

Table 1 Monitored events

Event type	Description
CLI	CLI event occurs in response to monitored operations performed at the CLI. For example, a command is entered, a question mark (?) is entered, or the **Tab** key is pressed to complete a command.
Hotplug	Hot-swapping event occurs when the monitored member device joins or leaves the IRF fabric.
Interface	Each interface event is associated with two user-defined thresholds: start and restart. An interface event occurs when the monitored interface traffic statistic crosses the start threshold in the following situations: • The statistic crosses the start threshold for the first time. • The statistic crosses the start threshold each time after it crosses the restart threshold.
period	Periodic event occurs when the specified interval expires.
process	Process event occurs in response to a state change of the monitored process (such as an exception, shutdown, start, or restart). Both manual and automatic state changes can cause the event to occur.
SNMP	Each SNMP event is associated with two user-defined thresholds: start and restart. SNMP event occurs when the monitored MIB variable's value crosses the start threshold in the following situations: • The monitored variable's value crosses the start threshold for the first time. • The monitored variable's value crosses the start threshold each time after it crosses the restart threshold.
SNMP-Notification	SNMP-Notification event occurs when the monitored MIB variable's value in an SNMP notification matches the specified condition. For example, the broadcast traffic rate on an Ethernet interface reaches or exceeds 30%.
Syslog	Syslog event occurs when the information center receives the monitored log within a specific period. NOTE: The log that is generated by the EAA RTM does not trigger the monitor policy to run.
Track	Track event occurs when the state of the track entry changes from Positive to Negative or from Negative to Positive. If you specify multiple track entries for a policy, EAA triggers the policy only when the state of all the track entries changes from Positive

Event type	Description
	(Negative) to Negative (Positive). If you set a suppress time for a policy, the timer starts when the policy is triggered. The system does not process the messages that report the track entry state change from Positive (Negative) to Negative (Positive) until the timer times out.

Action

You can create a series of order-dependent actions to take in response to the event specified in the monitor policy.

The following are available actions:

- Executing a command.
- Sending an email.
- Executing a reboot without saving the running configuration.
- Enabling an active/standby switchover.
- Sending a log.

User role

For EAA to execute an action in a monitor policy, you must assign the policy the user role that has access to the action-specific commands and resources. If EAA lacks access to an action-specific command or resource, EAA does not perform the action and all the subsequent actions.

For example, a monitor policy has four actions numbered from 1 to 4. The policy has user roles that are required for performing actions 1, 3, and 4. However, it does not have the user role required for performing action 2. When the policy is triggered, EAA executes only action 1.

For more information about user roles, see RBAC in *Fundamentals Configuration Guide*.

Runtime

The runtime limits the amount of time that the monitor policy runs its actions from the time it is triggered. This setting prevents a policy from running its actions permanently to occupy resources.

EAA environment variables

EAA environment variables decouple the configuration of action arguments from the monitor policy so you can modify a policy easily.

An EAA environment variable is defined as a <*variable_name variable_value*> pair and can be used in different policies. When you define an action, you can enter a variable name with a leading dollar sign (**$***variable_name*). EAA will replace the variable name with the variable value when it performs the action.

To change the value for an action argument, modify the value specified in the variable pair instead of editing each affected monitor policy.

EAA environment variables include system-defined variables and user-defined variables.

System-defined variables

System-defined variables are provided by default, and they cannot be created, deleted, or modified by users. System-defined variable names start with an underscore (_) sign. The variable values are set automatically depending on the event setting in the policy that uses the variables.

System-defined variables include the following types:

- **Public variable**—Available for any events.

- **Event-specific variable**—Available only for a type of event. For example, the CLI event-specific variable is _cmd. When an CLI event is triggered, the value of _cmd is the command line matching the monitor policy.

Table 2 shows all system-defined variables.

Table 2 System-defined EAA environment variables by event type

Event	Variable name and description
Any event	_event_id: Event ID. _event_type: Event type. _event_type_string: Event type description. _event_time: Time when the event occurs. _event_severity: Severity level of an event.
CLI	_cmd: Commands that are matched.
Hotplug	_slot: ID of the member device that joins or leaves the IRF fabric
Hotplug	_subslot: ID of the subslot where subcard hot-swapping occurs.
Interface	_ifname: Interface name.
Period	_period_value: Interval.
Process	_process_name: Process name.
SNMP	_oid: OID of the MIB variable where an SNMP operation is performed. _oid_value: Value of the MIB variable.
SNMP-Notification	_oid: OID that is included in the SNMP notification.
Syslog	_syslog_pattern: Log message content.

User-defined variables

You can use user-defined variables for all types of events.

User-defined variable names can contain digits, characters, and the underscore sign (_), except that the underscore sign cannot be the leading character.

Restrictions and guidelines

Make sure the actions in different policies do not conflict. Policy execution result will be unpredictable if policies that conflict in actions are running concurrently.

You can assign the same policy name to a CLI-defined policy, a Tcl-defined policy, or a Python-defined policy. However, you cannot assign the same name to policies that are the same type.

A monitor policy supports only one event and runtime. If you configure multiple events for a policy, the most recent one takes effect.

A monitor policy supports a maximum of 64 valid user roles. User roles added after this limit is reached do not take effect.

EAA tasks at a glance

To configure EAA, perform the following tasks:

- (Optional.) Configuring a user-defined EAA environment variable
- Perform one of the following tasks to configure an EAA policy:

    - Configuring a monitor policy from the CLI
    - Configuring a monitor policy by using Tcl
    - Configuring a monitor policy by using Python
- (Optional.) Suspending monitor policies

Configuring a user-defined EAA environment variable

About this task

Configure user-defined EAA environment variables so that you can use them when creating EAA monitor policies.

Procedure

1. Enter system view.
   system-view
2. Configure a user-defined EAA environment variable.
   rtm environment *var-name var-value*
   For the system-defined variables, see the system-defined EAA environment variables table in "System-defined variables."

Configuring a monitor policy from the CLI

Tasks at a glance

1. Creating a CLI-defined policy
2. (Optional.) Setting the size for the EAA-monitored log buffer
3. (Optional.) Configuring automatic email sending parameters
4. Configuring an event
5. Configuring an action
6. Assigning a user role to a CLI-defined policy
7. Configuring the policy action runtime
8. Enabling the CLI-defined monitor policy

Creating a CLI-defined policy

1. Enter system view.
   system-view
2. Create a CLI-defined policy and enter its view.
   rtm cli-policy *policy-name*

Setting the size for the EAA-monitored log buffer

About this task

This task is required only when a syslog event is specified for a CLI-defined policy.

After you specify and enable a syslog event EAA monitor policy, the system copies a log to the EAA-monitored log buffer upon generation of the log. The EAA module compares the log in the buffer with the log event. If they match, the monitor policy is triggered.

Under normal circumstances, you can use the default size setting for the EAA-monitored log buffer. However, when a feature is abnormal or multiple debugging functions are enabled, a large number of logs might be generated in a short time, causing some logs to be discarded before they are matched. In this case, you can increase the size of the EAA-monitored log buffer according to the device memory usage.

Procedure

1. Enter system view.

   system-view

2. Set the size for the EAA-monitored log buffer.

   rtm event syslog buffer-size *buffer-size*

   By default, the EAA-monitored log buffer can store a maximum of 50000 logs.

Configuring automatic email sending parameters

About this task

Automatic email sending allows EAA to generate and send an email of the specified content to the designated receivers when the specific event occurs. You can use this feature to collect and send device O & M information periodically to the administrators.

To enable email automatic sending, configure the following settings:

- Configure an event for the monitor policy (see "Configuring an event"). For example, execute the **event period** command to configure a periodic event.
- Configure automatic email sending parameters, including the email username, email password, and email server domain name.
- Configure actions for the monitor policy (see "Configuring an action").
  - Add a CLI action to the monitory policy to collect device O & M information. The information will be sent out by the email in an attachment file.
  - Add an email sending action to the monitor policy. You can specify the email receivers, subject, and content by using the **action email** command. Environment variables can be used in the email subject or content for the system to send traps or logs to the administrators.
- Assign a user role to the CLI-defined monitor policy (see "Assigning a user role to a CLI-defined policy").
- Configure the policy action runtime (see "Configuring the policy action runtime").
- Enable the CLI-defined monitor policy (see "Enabling the CLI-defined monitor policy").

Restrictions and guidelines

As a best practice, pay attention only to important alarms and O & M data, and control the email sending frequency within the acceptable range (typically a maximum of 100 emails per minute). If the mail sending frequency exceeds the acceptable range of the email server, the emails generated after the limit is reached will fail to be sent.

Procedure

1. Enter system view.

   system-view

2. Specify the email server domain name.

   rtm email domain *domain-name*

By default, no email server domain name exists.

3. Specify the email username and password.

   rtm email username *username* **password** { **cipher** | **simple** } *password*

   By default, no email username or password exists.

   For successful email sending, make sure the email username and password are consistent with those registered with the email server.

4. Set the maximum email size.

   rtm email max-size *size*

   By default, the maximum email size is 2 MB.

   For an email to be sent, make sure the email size does not exceed the value set by this command and the maximum email size specified on the email server.

Configuring an event

Restrictions and guidelines

A monitor policy supports only one event. If you configure multiple events for a policy, the most recent one takes effect.

Procedure

1. Enter system view.

   system-view

2. Create a CLI-defined policy and enter its view.

   rtm cli-policy *policy-name*

3. Configure an event for the policy. Choose one option as needed:

   - Configure a CLI event.

     event cli { **async** [**skip**] | **sync** } **mode** { **execute** | **help** | **tab** } **pattern** *regular-exp*

   - Configure a hotplug event.

     event hotplug [**insert** | **remove**] **slot** *slot-number* [**subslot** *subslot-number*]

   - Configure an interface event.

     event interface *interface-list* **monitor-obj** *monitor-obj* **start-op** *start-op* **start-val** *start-val* **restart-op** *restart-op* **restart-val** *restart-val* [**interval** *interval*]

   - Configure a periodic event.

     event period *interval*

     For the actions of a periodic event monitor policy to be executed completely, the interval configured for the periodic event must be equal to or larger than the action running time.

   - Configure a process event.

     event process { **exception** | **restart** | **shutdown** | **start** } [**name** *process-name* [**instance** *instance-id*]] [**slot** *slot-number*]

   - Configure an SNMP event.

     event snmp oid *oid* **monitor-obj** { **get** | **next** } **start-op** *start-op* **start-val** *start-val* **restart-op** *restart-op* **restart-val** *restart-val* [**interval** *interval*]

   - Configure an SNMP-Notification event.

     event snmp-notification oid *oid* **oid-val** *oid-val* **op** *op* [**drop**]

- Configure a Syslog event.

  event syslog priority { *priority* | **all** } **msg** *msg* **occurs** *times* **period** *period*

- Configure a track event.

  event track *track-list* **state { negative**|**positive** }[**suppress-time** *suppress-time*]

By default, a monitor policy does not contain an event.

Configuring an action

Restrictions and guidelines

You can configure a series of actions to be executed in response to the event specified in a monitor policy. EAA executes the actions in ascending order of action IDs. When you add actions to a policy, you must make sure the execution order is correct. If two actions have the same ID, the most recent one takes effect.

Procedure

1. Enter system view.

   system-view

2. Create a CLI-defined policy and enter its view.

   rtm cli-policy *policy-name*

3. Configure the actions to take when the event occurs. Choose the options to configure as needed:

   - Configure a CLI action.

     action *number* **cli** *command-line*

   - Configure an email sending action.

     action *number* **email receiver** *rx-list* **subject** *subject* [**content** *content*]

   - Configure a reboot action.

     action *number* **reboot** [**slot** *slot-number* [**subslot** *subslot-number*]]

   - Configure an active/standby switchover action.

     action *number* **switchover**

   - Configure a logging action.

     action *number* **syslog priority** *priority* **facility** *local-number* **msg** *msg-body*

   By default, a monitor policy does not contain any actions.

Assigning a user role to a CLI-defined policy

Restrictions and guidelines

An EAA policy cannot have both the **security-audit** user role and any other user roles. Any previously assigned user roles are automatically removed when you assign the **security-audit** user role to the policy. The previously assigned **security-audit** user role is automatically removed when you assign any other user roles to the policy.

Procedure

1. Enter system view.

   system-view

2. Create a CLI-defined policy and enter its view.
   rtm cli-policy *policy-name*
3. (Optional.) Assign a user role to the policy.
   user-role *role-name*
   By default, a monitor policy contains user roles that its creator had at the time of policy creation.

Configuring the policy action runtime

Restrictions and guidelines

For the actions of a periodic event monitor policy to be executed completely, the action runtime must be equal to or smaller than the interval specified for the periodic event.

Procedure

1. Enter system view.
   system-view
2. Create a CLI-defined policy and enter its view.
   rtm cli-policy *policy-name*
3. (Optional.) Configure the policy action runtime.
   running-time *time*
   The default policy action runtime is 20 seconds.
   If you configure multiple action runtimes for a policy, the most recent one takes effect.

Enabling the CLI-defined monitor policy

About this task

A CLI-defined monitor policy takes effect only after you enable it.

Procedure

1. Enter system view.
   system-view
2. Create a CLI-defined policy and enter its view.
   rtm cli-policy *policy-name*
3. Enable the CLI-defined policy.
   commit

Configuring a monitor policy by using Tcl

About this task

A Tcl script contains two parts: Line 1 and the other lines.

- Line 1
  Line 1 defines the event, user roles, and policy action runtime. After you create and enable a Tcl monitor policy, the device immediately parses, delivers, and executes Line 1.
  Line 1 must use the following format:
  ::platformtools::rtm::event_register *event-type arg1 arg2 arg3* ... **user-role** *role-name1* | [**user-role** *role-name2* | [...]] [**running-time** *running-time*]

- The *arg1 arg2 arg3* … arguments represent event matching rules. If an argument value contains spaces, use double quotation marks ("") to enclose the value. For example, "a b c."
- The configuration requirements for the *event-type*, *user-role*, and *running-time* arguments are the same as those for a CLI-defined monitor policy.

- The other lines

  From the second line, the Tcl script defines the actions to be executed when the monitor policy is triggered. You can use multiple lines to define multiple actions. The system executes these actions in sequence. The following actions are available:
  - Standard Tcl commands.
  - EAA-specific Tcl actions:
    - switchover (**::platformtools::rtm::action switchover**)
    - syslog (**::platformtools::rtm::action syslog priority** *priority* **facility** *local-number* **msg** *msg-body*). For more information about these arguments, see EAA commands in *Network Management and Monitoring Command Reference*.
  - Commands supported by the device.

Restrictions and guidelines

To revise the Tcl script of a policy, you must suspend all monitor policies first, and then resume the policies after you finish revising the script. The system cannot execute a Tcl-defined policy if you edit its Tcl script without first suspending these policies.

Procedure

1. Download the Tcl script file to the device by using FTP or TFTP.

   For more information about using FTP and TFTP, see *Fundamentals Configuration Guide*.
2. Create and enable a Tcl monitor policy.
   a. Enter system view.

      system-view

   b. Create a Tcl-defined policy and bind it to the Tcl script file.

      rtm tcl-policy *policy-name tcl-filename*

   By default, no Tcl policies exist.

   Make sure the script file is saved on all IRF member devices. This practice ensures that the policy can run correctly after a master/subordinate switchover occurs or the member device where the script file resides leaves the IRF.

Configuring a monitor policy by using Python

About this task

A Python script contains two parts: Line 1 and the other lines.

- Line 1

  Line 1 defines the event, user roles, and policy action runtime. After you create and enable a Python monitor policy, the device immediately parses, delivers, and executes Line 1.

  Line 1 must use the following format:

  ::comware::rtm::event_register *event-type arg1 arg2 arg3* … **user-role** *role-name1* | [**user-role** *role-name2* | […]] [**running-time** *running-time*]
  - The *arg1 arg2 arg3* … arguments represent event matching rules. If an argument value contains spaces, use double quotation marks ("") to enclose the value. For example, "a b c."

- The configuration requirements for the *event-type*, *user-role*, and *running-time* arguments are the same as those for a CLI-defined monitor policy.

- The other lines

  From the second line, the Python script defines the actions to be executed when the monitor policy is triggered. For the Python APIs supported by the device, see Python configuration in *Network Programmability Configuration Guide*.

Restrictions and guidelines

To revise the Python script of a policy, you must suspend the policy first, and then resume the policy after you finish revising the script. The system cannot execute a Python-defined policy if you edit its Python script without first suspending it.

Procedure

1. Download the Python script file to the device by using FTP or TFTP.

   For more information about using FTP and TFTP, see *Fundamentals Configuration Guide*.

2. Create and enable a Python monitor policy.

   a. Enter system view.

   system-view

   b. Create a Python-defined policy and bind it to the Python script file.

   rtm tcl-policy *policy-name tcl-filename*

   By default, no Python policies exist.

   Make sure the script file is saved on all IRF member devices. This practice ensures that the policy can run correctly after a master/subordinate switchover occurs or the member device where the script file resides leaves the IRF.

Suspending monitor policies

About this task

This task suspends all CLI-defined and Tcl-defined monitor policies. If a policy is running when you perform this task, the system suspends the policy after it executes all the actions.

Restrictions and guidelines

To restore the operation of the suspended policies, execute the **undo rtm scheduler suspend** command.

Procedure

1. Enter system view.

   system-view

2. Suspend monitor policies.

   rtm scheduler suspend

Display and maintenance commands for EAA

Execute **display** commands except for the **display this** command in any view.

Task	Command
Display the running configuration of all CLI-defined monitor policies.	**display current-configuration**

Task	Command
Display user-defined EAA environment variables.	**display rtm environment** [*var-name*]
Display EAA monitor policies.	**display rtm policy** { **active** | **registered** [**verbose**] } [*policy-name*]
Display history execution records for EAA monitor policies.	**display rtm policy history** [*policy-name*]
Display the running configuration of a CLI-defined monitor policy in CLI-defined monitor policy view.	**display this**

EAA configuration examples

Example: Configuring a CLI event monitor policy by using Tcl

Network configuration

As shown in Figure 2, use Tcl to create a monitor policy on the Device. This policy must meet the following requirements:

- EAA sends the log message "rtm_tcl_test is running" when a command that contains the **display this** string is entered.
- The system executes the command only after it executes the policy successfully.

Figure 2 Network diagram



Procedure

Edit a Tcl script file (rtm_tcl_test.tcl, in this example) for EAA to send the message "rtm_tcl_test is running" when a command that contains the **display this** string is executed.

```
::platformtools::rtm::event_register cli sync mode execute pattern display this user-role network-admin
::platformtools::rtm::action syslog priority 1 facility local4 msg rtm_tcl_test is running
```

Download the Tcl script file from the TFTP server at **1.2.1.1**.

```
<Sysname> tftp 1.2.1.1 get rtm_tcl_test.tcl
```

Create Tcl-defined policy **test** and bind it to the Tcl script file.

```
<Sysname> system-view
[Sysname] rtm tcl-policy test rtm_tcl_test.tcl
[Sysname] quit
```

Verifying the configuration

Execute the **display rtm policy registered** command to verify that a Tcl-defined policy named **test** is displayed in the command output.

```
<Sysname> display rtm policy registered
Total number: 1
```

```
Type  Event             TimeRegistered       PolicyName
TCL   CLI               Aug 01 09:47:12 2019 test
```

Enable the information center to output log messages to the current monitoring terminal.

```
<Sysname> terminal monitor
The current terminal is enabled to display logs.
<Sysname> system-view
[Sysname] info-center enable
Information center is enabled.
[Sysname] quit
```

Execute the **display this** command. Verify that the system displays an "rtm_tcl_test is running" message and a message that the policy is being executed successfully.

```
[Sysname] display this
%Aug  1 09:50:04:634 2019 Sysname RTM/1/RTM_ACTION: rtm_tcl_test is running
%Aug  1 09:50:04:636 2019 Sysname RTM/6/RTM_POLICY_SUCCESS: TCL policy test is running
successfully.
#
return
```

Example: Configuring a CLI event monitor policy from the CLI

Network configuration

Configure a policy from the CLI to monitor the event that occurs when a question mark (?) is entered at the command line that contains letters and digits.

When the event occurs, the system executes the command and sends the log message "hello world" to the information center.

Procedure

Create CLI-defined policy **test** and enter its view.

```
<Sysname> system-view
[Sysname] rtm cli-policy test
```

Add a CLI event that occurs when a question mark (?) is entered at any command line that contains letters and digits.

```
[Sysname-rtm-test] event cli async mode help pattern [a-zA-Z0-9]
```

Add an action that sends the message "hello world" with a priority of 4 from the logging facility **local3** when the event occurs.

```
[Sysname-rtm-test] action 0 syslog priority 4 facility local3 msg "hello world"
```

Add an action that enters system view when the event occurs.

```
[Sysname-rtm-test] action 2 cli system-view
```

Add an action that creates VLAN 2 when the event occurs.

```
[Sysname-rtm-test] action 3 cli vlan 2
```

Set the policy action runtime to 2000 seconds.

```
[Sysname-rtm-test] running-time 2000
```

Specify the **network-admin** user role for executing the policy.

```
[Sysname-rtm-test] user-role network-admin
```

Enable the policy.

```
[Sysname-rtm-test] commit
```

Verifying the configuration

Execute the **display rtm policy registered** command to verify that a CLI-defined policy named **test** is displayed in the command output.

```
[Sysname-rtm-test] display rtm policy registered
Total number: 1
Type  Event      TimeRegistered       PolicyName
CLI   CLI        Aug  1 14:56:50 2019 test
```

Enable the information center to output log messages to the current monitoring terminal.

```
[Sysname-rtm-test] return
<Sysname> terminal monitor
The current terminal is enabled to display logs.
<Sysname> system-view
[Sysname] info-center enable
Information center is enabled.
[Sysname] quit
```

Enter a question mark (?) at a command line that contains a letter **d**. Verify that the system displays a "hello world" message and a message that the policy is being executed successfully on the terminal screen.

```
<Sysname> d?
  debugging
  delete
  diagnostic-logfile
  dir
  display

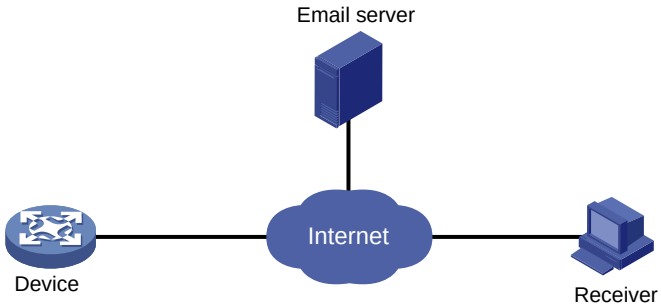
<Sysname>d%Aug  1 14:57:20:218 2019 Sysname RTM/4/RTM_ACTION: "hello world"
%Aug  1 14:58:11:170 2019 Sysname RTM/6/RTM_POLICY_SUCCESS: CLI policy test is running
successfully.
```

Configuring an automatic email sending action for a CLI-defined monitor policy

Network configuration

Configure a policy from the CLI to enable the system to send an email from a mailbox (username **test** and password **123**) to receivers test1@example.com and test2@example.com each hour to report brief interface information.

Figure 3 Network diagram



Procedure

Specify **example.com** as the email server domain name. Set the email username to **test** and password to **123**.

```
<Sysname> system-view
[Sysname] rtm email domain example.com
[Sysname] rtm email username test password simple 123
```

Create CLI-defined policy **test** and enter its view.

```
[Sysname] rtm cli-policy test
```

Create a periodic event and set the interval to 60 minutes.

```
[Sysname-rtm-test] event period 60
```

Add a CLI action to the policy, enabling the system to display brief interface information when the event occurs.

```
[Sysname-rtm-test] action 1 cli display interface brief
```

Add an email sending action to the policy, enabling the system to send an email with subject "**Interface information**", content "**This is the interface brief information.**", and an attachment file that contains brief interface information to test1@example.com and test2@example.com when the event occurs.

```
[Sysname-rtm-test] action 2 email receiver test1@example.com,test2@example.com subject
interface content "This is the interface brief information."
```

Enable the policy.

```
[Sysname-rtm-test] commit
```

Verifying the configuration

Execute the **display rtm policy registered** command to verify that a CLI-defined policy named **test** with event type **PERIOD** is displayed in the command output.

```
[Sysname-rtm-test] display rtm policy registered
Total number: 1
Type  Event      TimeRegistered       PolicyName
CLI   PERIOD     Sep 21 11:43:09 2019 test
```

Verify that the receivers receive the emails after the interval expires.

Verify that logs are output after the interval expires indicating that the device succeeded in sending the email and the policy was executed successfully.

```
[Sysname-rtm-test] return
<Sysname> terminal monitor
The current terminal is enabled to display logs.
<Sysname> system-view
[Sysname] info-center enable
Information center is enabled.
[Sysname] quit
<Sysname>%Sep 30 09:22:03:224 2020 Sysname RTM/6/RTM_EMAIL_SUCCESS: Succeed in sending
an email with the subject interface to test1@example.com,test2@example.com.
<Sysname>%Sep 30 09:22:03:237 2020 Sysname RTM/6/RTM_POLICY_SUCCESS: CLI policy test is
running successfully.
```

Example: Configuring a track event monitor policy from the CLI

Network configuration

As shown in Figure 4, Device A has established BGP sessions with Device D and Device E. Traffic from Device D and Device E to the Internet is forwarded through Device A.

Configure a CLI-defined EAA monitor policy on Device A to disconnect the sessions with Device D and Device E when GigabitEthernet 1/0/1 connected to Device C is down. In this way, traffic from Device D and Device E to the Internet can be forwarded through Device B.

Figure 4 Network diagram

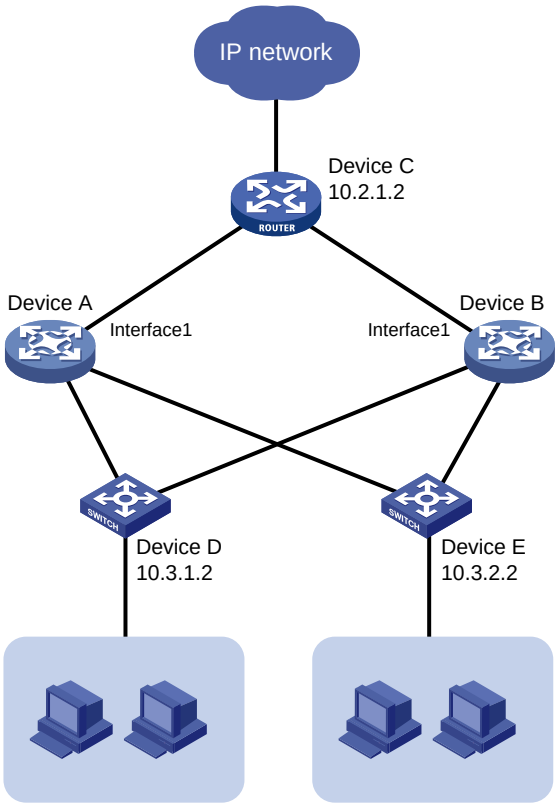


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

Display BGP peer information for Device A.

```
<DeviceA> display bgp peer ipv4

 BGP local router ID: 1.1.1.1
 Local AS number: 100
 Total number of peers: 3                 Peers in established state: 3

  * - Dynamically created peer
```

```
  Peer                    AS  MsgRcvd  MsgSent OutQ PrefRcv Up/Down  State

  10.2.1.2                200       13       16    0       0 00:16:12 Established
  10.3.1.2                300       13       16    0       0 00:10:34 Established
  10.3.2.2                300       13       16    0       0 00:10:38 Established
```

Create track entry **1** and associate it with the link state of GigabitEthernet 1/0/1.

```
<Device A> system-view
[Device A] track 1 interface gigabitethernet 1/0/1
```

Configure a CLI-defined EAA monitor policy so that the system automatically disables session establishment with Device D and Device E when GigabitEthernet 1/0/1 is down.

```
[Device A] rtm cli-policy test
[Device A-rtm-test] event track 1 state negative
[Device A-rtm-test] action 0 cli system-view
[Device A-rtm-test] action 1 cli bgp 100
[Device A-rtm-test] action 2 cli peer 10.3.1.2 ignore
[Device A-rtm-test] action 3 cli peer 10.3.2.2 ignore
[Device A-rtm-test] user-role network-admin
[Device A-rtm-test] commit
[Device A-rtm-test] quit
```

Verifying the configuration

Shut down GigabitEthernet 1/0/1.

```
[Device A] interface gigabitethernet 1/0/1
[Device A-GigabitEthernet1/0/1] shutdown
```

Execute the **`display bgp peer ipv4`** command on Device A to display BGP peer information to verify that no peer exists in established state.

Example: Configuring a CLI event monitor policy with EAA environment variables from the CLI

Network configuration

Define an environment variable to match the IP address 1.1.1.1.

Configure a policy from the CLI to monitor the event that occurs when a command line that contains **loopback0** is executed. In the policy, use the environment variable for IP address assignment.

When the event occurs, the system performs the following tasks:

- Creates the Loopback 0 interface.
- Assigns 1.1.1.1/24 to the interface.
- Sends the matching command line to the information center.

Procedure

Configure an EAA environment variable for IP address assignment. The variable name is **loopback0IP**, and the variable value is **1.1.1.1**.

```
<Sysname> system-view
[Sysname] rtm environment loopback0IP 1.1.1.1
```

Create the CLI-defined policy **test** and enter its view.

```
[Sysname] rtm cli-policy test
```

Add a CLI event that occurs when a command line that contains **loopback0** is executed.

```
[Sysname-rtm-test] event cli async mode execute pattern loopback0
```

Add an action that enters system view when the event occurs.

```
[Sysname-rtm-test] action 0 cli system-view
```

Add an action that creates the interface Loopback 0 and enters loopback interface view.

```
[Sysname-rtm-test] action 1 cli interface loopback 0
```

Add an action that assigns the IP address 1.1.1.1 to Loopback 0. The **loopback0IP** variable is used in the action for IP address assignment.

```
[Sysname-rtm-test] action 2 cli ip address $loopback0IP 24
```

Add an action that sends the matching **loopback0** command with a priority of 0 from the logging facility **local7** when the event occurs.

```
[Sysname-rtm-test] action 3 syslog priority 0 facility local7 msg $_cmd
```

Specify the **network-admin** user role for executing the policy.

```
[Sysname-rtm-test] user-role network-admin
```

Enable the policy.

```
[Sysname-rtm-test] commit
[Sysname-rtm-test] return
<Sysname>
```

Verifying the configuration

Enable the information center to output log messages to the current monitoring terminal.

```
<Sysname> terminal monitor
The current terminal is enabled to display logs.
<Sysname> terminal log level debugging
<Sysname> system-view
[Sysname] info-center enable
Information center is enabled.
```

Execute the **loopback0** command. Verify that the system displays a "loopback0" message and a message that the policy is being executed successfully on the terminal screen.

```
[Sysname] interface loopback0
[Sysname-LoopBack0]%Aug  1 09:46:10:592 2019 Sysname RTM/7/RTM_ACTION: interface
loopback0
%Aug  1 09:46:10:613 2019 Sysname RTM/6/RTM_POLICY_SUCCESS: CLI policy test is running
successfully.
```

Verify that Loopback 0 has been created and its IP address is 1.1.1.1.

```
<Sysname-LoopBack0> display interface loopback brief
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP         Description
Loop0                UP   UP(s)    1.1.1.1

<Sysname-LoopBack0>
```

Contents

Monitoring and maintaining processes 1

About monitoring and maintaining processes 1
Process monitoring and maintenance tasks at a glance 1
Monitoring and maintaining processes 1
Monitoring and maintaining user processes 2
About monitoring and maintaining user processes 2
Configuring core dump 2
Display and maintenance commands for user processes 3
Monitoring and maintaining kernel threads 4
Configuring kernel thread deadloop detection 4
Configuring kernel thread starvation detection 4
Display and maintenance commands for kernel threads 5

Monitoring and maintaining processes

About monitoring and maintaining processes

The system software of the device is a full-featured, modular, and scalable network operating system based on the Linux kernel. The system software features run the following types of independent processes:

- **User process**—Runs in user space. Most system software features run user processes. Each process runs in an independent space so the failure of a process does not affect other processes. The system automatically monitors user processes. The system supports preemptive multithreading. A process can run multiple threads to support multiple activities. Whether a process supports multithreading depends on the software implementation.
- **Kernel thread**—Runs in kernel space. A kernel thread executes kernel code. It has a higher security level than a user process. If a kernel thread fails, the system breaks down. You can monitor the running status of kernel threads.

Process monitoring and maintenance tasks at a glance

To monitor and maintain processes, perform the following tasks:

- Monitoring and maintaining user processes
  - Monitoring and maintaining processes
  - The commands in this section apply to both user processes and kernel threads.
  - Monitoring and maintaining user processes

    The commands in this section apply only to user processes.
- Monitoring and maintaining kernel threads
  - Monitoring and maintaining processes
  - The commands in this section apply to both user processes and kernel threads.
  - Monitoring and maintaining kernel threads
  - The commands in this section apply only to kernel threads.

Monitoring and maintaining processes

About this task

The commands in this section apply to both user processes and kernel threads. You can use the commands for the following purposes:

- Display the overall memory usage.
- Display the running processes and their memory and CPU usage.
- Locate abnormal processes.

If a process consumes excessive memory or CPU resources, the system identifies the process as an abnormal process.

- If an abnormal process is a user process, troubleshoot the process as described in "Monitoring and maintaining user processes."

- If an abnormal process is a kernel thread, troubleshoot the process as described in "Monitoring and maintaining kernel threads."

Procedure

Execute the following commands in any view.

Task	Command
Display memory usage.	**display memory** [**summary**] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display process state information.	**display process** [**all** | **job** *job-id* | **name** *process-name*] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display CPU usage for all processes.	**display process cpu** [**slot** *slot-number* [**cpu** *cpu-number*]]
Monitor process running state.	**monitor process** [**dumbtty**] [**iteration** *number*] [**slot** *slot-number* [**cpu** *cpu-number*]]
Monitor thread running state.	**monitor thread** [**dumbtty**] [**iteration** *number*] [**slot** *slot-number* [**cpu** *cpu-number*]]

For more information about the **display memory** command, see *Fundamentals Command Reference*.

Monitoring and maintaining user processes

About monitoring and maintaining user processes

Use this feature to monitor abnormal user processes and locate problems.

Configuring core dump

About this task

The core dump feature enables the system to generate a core dump file each time a process crashes until the maximum number of core dump files is reached. A core dump file stores information about the process. You can send the core dump files to H3C Support to troubleshoot the problems.

Restrictions and guidelines

Core dump files consume storage resources. Enable core dump only for processes that might have problems.

Procedure

Execute the following commands in user view:

1. (Optional.) Specify the directory for saving core dump files.

   exception filepath *directory*

   By default, the directory for saving core dump files is the root directory of the default file system. For more information about the default file system, see file system management in *Fundamentals Configuration Guide*.

2. Enable core dump for a process and specify the maximum number of core dump files, or disable core dump for a process.

   process core { **maxcore** *value* | **off** } { **job** *job-id* | **name** *process-name* }

By default, a process generates a core dump file for the first exception and does not generate any core dump files for subsequent exceptions.

Display and maintenance commands for user processes

Execute **display** commands in any view and other commands in user view.

Task	Command
Display context information for process exceptions.	**display exception context** [**count** *value*] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display the core dump file directory.	**display exception filepath** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display log information for all user processes.	**display process log** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display memory usage for all user processes.	**display process memory** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display memory allocation failure information for a user process.	**display process memory alloc-failure job** *job-id* [**slot** *slot-number* [**cpu** *cpu-number*]]
Display heap memory usage for a user process.	**display process memory heap job** *job-id* [**verbose**] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display information about used memory blocks for a user process.	**display process memory fragment free job** *job-id* [**slot** *slot-number* [**cpu** *cpu-number*]]
Display heap memory usage for a user process.	**display process memory fragment used** *used-block* **job** *job-id* [**slot** *slot-number* [**cpu** *cpu-number*]]
Display memory content starting from a specified memory block for a user process.	**display process memory heap job** *job-id* **address** *starting-address* **length** *memory-length* [**slot** *slot-number* [**cpu** *cpu-number*]]
Display the addresses of memory blocks with a specified size used by a user process.	**display process memory heap job** *job-id* **size** *memory-size* [**offset** *offset-size*] [**slot** *slot-number* [**cpu** *cpu-number*]]
Clear context information for process exceptions.	**reset exception context** [**slot** *slot-number* [**cpu** *cpu-number*]]

Monitoring and maintaining kernel threads

Configuring kernel thread deadloop detection

About this task

Kernel threads share resources. If a kernel thread monopolizes the CPU, other threads cannot run, resulting in a deadloop.

This feature enables the device to detect deadloops. If a thread occupies the CPU for a specific interval, the device determines that a deadloop has occurred and generates a deadloop message.

Restrictions and guidelines

CAUTION:

Change kernel thread deadloop detection settings only under the guidance of H3C Support. Inappropriate configuration can cause system breakdown. As a best practice, leave the default unchanged.

Procedure

1. Enter system view.

   system-view

2. Enable kernel thread deadloop detection.

   monitor kernel deadloop enable [**slot** *slot-number* [**cpu** *cpu-number* [**core** *core-number*&<1-64>]]]

   By default, kernel thread deadloop detection is enabled.

3. (Optional.) Set the threshold for identifying a kernel thread deadloop.

   monitor kernel deadloop time *time* [**slot** *slot-number* [**cpu** *cpu-number*]]

   By default, the threshold for identifying a kernel thread deadloop is 25 seconds.

4. (Optional.) Exclude a kernel thread from kernel thread deadloop detection.

   monitor kernel deadloop exclude-thread *tid* [**slot** *slot-number* [**cpu** *cpu-number*]]

   When enabled, kernel thread deadloop detection monitors all kernel threads by default.

5. (Optional.) Specify the action to be taken in response to a kernel thread deadloop.

   monitor kernel deadloop action { **reboot** | **record-only** } [**slot** *slot-number* [**cpu** *cpu-number*]]

   The default action is **reboot**.

Configuring kernel thread starvation detection

About this task

Starvation occurs when a thread is unable to access shared resources.

Kernel thread starvation detection enables the system to detect and report thread starvation. If a thread is not executed within a specific interval, the system determines that a starvation has occurred and generates a starvation message.

Thread starvation does not impact system operation. A starved thread can automatically run when certain conditions are met.

Restrictions and guidelines

CAUTION:

Configure kernel thread starvation detection only under the guidance of H3C Support. Inappropriate configuration can cause system breakdown. As a best practice, leave the default unchanged.

Procedure

1. Enter system view.

   system-view

2. Enable kernel thread starvation detection.

   monitor kernel starvation enable [**slot** *slot-number* [**cpu** *cpu-number*]]

   By default, kernel thread starvation detection is disabled.

3. (Optional.) Set the threshold for identifying a kernel thread starvation.

   monitor kernel starvation time *time* [**slot** *slot-number* [**cpu** *cpu-number*]]

   By default, the threshold for identifying a kernel thread starvation is 120 seconds.

4. (Optional.) Exclude a kernel thread from kernel thread starvation detection.

   monitor kernel starvation exclude-thread *tid* [**slot** *slot-number* [**cpu** *cpu-number*]]

   When enabled, kernel thread starvation detection monitors all kernel threads by default.

Display and maintenance commands for kernel threads

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display kernel thread deadloop detection configuration.	**display kernel deadloop configuration** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display kernel thread deadloop information.	**display kernel deadloop** *show-number* [*offset*] [**verbose**] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display kernel thread exception information.	**display kernel exception** *show-number* [*offset*] [**verbose**] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display kernel thread reboot information.	**display kernel reboot** *show-number* [*offset*] [**verbose**] [**slot** *slot-number* [**cpu** *cpu-number*]]
Display kernel thread starvation detection configuration.	**display kernel starvation configuration** [**slot** *slot-number* [**cpu** *cpu-number*]]
Display kernel thread starvation information.	**display kernel starvation** *show-number* [*offset*] [**verbose**] [**slot** *slot-number* [**cpu** *cpu-number*]]
Clear kernel thread deadloop information.	**reset kernel deadloop** [**slot** *slot-number* [**cpu** *cpu-number*]]
Clear kernel thread exception information.	**reset kernel exception** [**slot** *slot-number* [**cpu** *cpu-number*]]

Task	Command
Clear kernel thread reboot information.	**reset kernel reboot** [**slot** *slot-number* [**cpu** *cpu-number*]]
Clear kernel thread starvation information.	**reset kernel starvation** [**slot** *slot-number* [**cpu** *cpu-number*]]

Contents

Configuring samplers 1
About sampler 1
Creating a sampler 1
Display and maintenance commands for a sampler 1

Configuring samplers

About sampler

A sampler selects a packet from sequential packets and sends the packet to other service modules for processing. Sampling is useful when you want to limit the volume of traffic to be analyzed. The sampled data is statistically accurate and sampling decreases the impact on the forwarding capacity of the device.

The device supports random sampling mode.

In random sampling mode, any packet might be selected from sequential packets in each sampling.

Creating a sampler

1. Enter system view.
   system-view
2. Create a sampler.
   sampler *sampler-name* **mode random packet-interval n-power** *rate*
   By default, no samplers exist.

Display and maintenance commands for a sampler

Execute **display** commands in any view.

Task	Command
Display configuration information about the sampler.	**display sampler** [*sampler-name*] [**slot** *slot-number*]

Contents

Configuring port mirroring 2
About port mirroring 2
Terminology 2
Port mirroring classification 3
Local port mirroring (SPAN) 3
Layer 2 remote port mirroring (RSPAN) 3
Layer 3 remote port mirroring (ERSPAN) 4
Restrictions and guidelines: Port mirroring configuration 5
Configuring local port mirroring (SPAN) 5
Restrictions and guidelines for local port mirroring configuration 5
Local port mirroring tasks at a glance 6
Creating a local mirroring group 6
Configuring mirroring sources 6
Configuring the monitor port 7
Configuring Layer 2 remote port mirroring (RSPAN) 8
Restrictions and guidelines for Layer 2 remote port mirroring configuration 8
Layer 2 remote port mirroring with egress port configuration task list 8
Creating a remote destination group 9
Configuring the monitor port 9
Configuring the remote probe VLAN 10
Assigning the monitor port to the remote probe VLAN 10
Creating a remote source group 10
Configuring mirroring sources 11
Configuring the egress port 12
Configuring Layer 3 remote port mirroring (ERSPAN in encapsulation parameter mode) 13
Restrictions and guidelines for Layer 3 remote port mirroring in ERSPAN mode configuration 13
Layer 3 remote port mirroring tasks at a glance 13
Creating a local mirroring group on the source device 13
Configuring mirroring sources 13
Configuring the monitor port 14
Display and maintenance commands for port mirroring 15
Port mirroring configuration examples 15
Example: Configuring local port mirroring (SPAN in source port mode) 15
Example: Configuring local port mirroring (SPAN in source VLAN mode) 16
Example: Configuring local port mirroring (SPAN in source CPU mode) 18
Example: Configuring Layer 2 remote port mirroring (RSPAN with egress port) 19
Example: Configuring Layer 3 remote port mirroring (ERSPAN in encapsulation parameter mode) 21
Configuring flow mirroring 23
About flow mirroring 23
Types of flow-mirroring traffic to an interface 23
Flow mirroring SPAN or RSPAN 23
Flow mirroring ERSPAN 24
Restrictions and guidelines: Flow mirroring configuration 25
Flow mirroring tasks at a glance 25
Configuring a traffic class 26
Configuring a traffic behavior 26
Configuring a QoS policy 27
Applying a QoS policy 27
Applying a QoS policy to an interface 27
Applying a QoS policy to a VLAN 28
Applying a QoS policy globally 28
Flow mirroring configuration examples 29
Example: Configuring flow mirroring 29

Configuring port mirroring

About port mirroring

Port mirroring copies the packets passing through a port, VLAN, or CPU to a port that connects to a data monitoring device for packet analysis.

Terminology

The following terms are used in port mirroring configuration.

Mirroring source

The mirroring sources can be one or more monitored ports (called source ports), VLANs (called source VLANs), or CPUs (called source CPUs).

Packets passing through mirroring sources are copied to a port connecting to a data monitoring device for packet analysis. The copies are called mirrored packets.

Source device

The device where the mirroring sources reside is called a source device.

Mirroring destination

The mirroring destination connects to a data monitoring device and is the destination port (also known as the monitor port) of mirrored packets. Mirrored packets are sent out of the monitor port to the data monitoring device.

A monitor port might receive multiple copies of a packet when it monitors multiple mirroring sources. For example, two copies of a packet are received on Port A when the following conditions exist:

- Port A is monitoring bidirectional traffic of Port B and Port C on the same device.
- The packet travels from Port B to Port C.

Destination device

The device where the monitor port resides is called the destination device.

Mirroring direction

The mirroring direction specifies the direction of the traffic that is copied on a mirroring source.

- **Inbound**—Copies packets received.
- **Outbound**—Copies packets sent.
- **Bidirectional**—Copies packets received and sent.

Mirroring group

Port mirroring is implemented through mirroring groups. Mirroring groups can be classified into local mirroring groups, remote source groups, and remote destination groups.

Reflector port, egress port, and remote probe VLAN

Reflector ports, remote probe VLANs, and egress ports are used for Layer 2 remote port mirroring. The remote probe VLAN is a dedicated VLAN for transmitting mirrored packets to the destination device. Both the reflector port and egress port reside on a source device and send mirrored packets to the remote probe VLAN.

On port mirroring devices, all ports except source, destination, reflector, and egress ports are called common ports.

Egress port and remote probe VLAN

Remote probe VLANs and egress ports are used for Layer 2 remote port mirroring. The remote probe VLAN is a dedicated VLAN for transmitting mirrored packets to the destination device. The egress port resides on a source device and send mirrored packets to the remote probe VLAN.

On port mirroring devices, all ports except source, destination, and egress ports are called common ports.

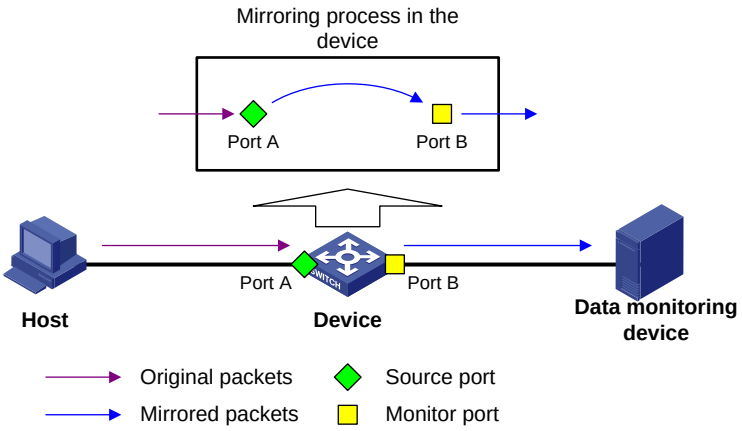
Port mirroring classification

Port mirroring includes the following types:

- **Local port mirroring**—Also known as Switch Port Analyzer (SPAN). In local port mirroring, the source device is directly connected to a data monitoring device. The source device acts as the destination device to forward mirrored packets to the data monitoring device.
- **Remote port mirroring**—In remote port mirroring, the source device is not directly connected to a data monitoring device. The source device copies mirrored packets to the destination device, which forwards them to the data monitoring device. The source device and destination device are connected through a Layer 2 network. Therefore, remote port mirroring is also called Layer 2 remote port mirroring or remote SPAN (RSPAN).

Local port mirroring (SPAN)

Figure 1 Local port mirroring implementation



As shown in Figure 1, the source port (Port A) and the monitor port (Port B) reside on the same device. Packets received on Port A are copied to Port B. Port B then forwards the packets to the data monitoring device for analysis.

Layer 2 remote port mirroring (RSPAN)

In Layer 2 remote port mirroring, the mirroring sources and destination reside on different devices and are in different mirroring groups.

A remote source group is a mirroring group that contains the mirroring sources. A remote destination group is a mirroring group that contains the mirroring destination. Intermediate devices are the devices between the source device and the destination device.

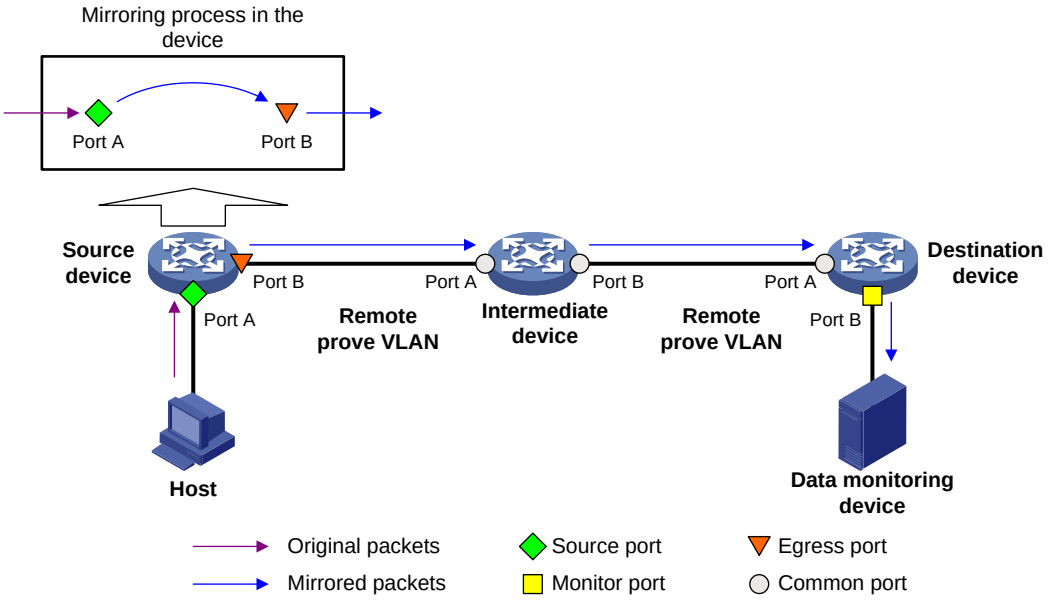
Layer 2 remote port mirroring can be implemented through the egress port method.

Egress port method

In Layer 2 remote port mirroring that uses the egress port method, packets are mirrored as follows:

1. The source device copies packets received on the mirroring sources to the egress port.
2. The egress port forwards the mirrored packets to the intermediate devices.
3. The intermediate devices flood the mirrored packets in the remote probe VLAN and transmit the mirrored packets to the destination device.
4. Upon receiving the mirrored packets, the destination device determines whether the ID of the mirrored packets is the same as the remote probe VLAN ID. If the two VLAN IDs match, the destination device forwards the mirrored packets to the data monitoring device through the monitor port.

Figure 2 Layer 2 remote port mirroring implementation through the egress port method



Layer 3 remote port mirroring (ERSPAN)

Layer 3 remote port mirroring encapsulates mirrored packets in GRE packets with a protocol number of 0x88BE and routes the packets to the remote monitoring device.

The switch series supports implementing Layer 3 remote port mirroring in encapsulation parameter mode.

Layer 3 remote port mirroring in encapsulation parameter mode

To implement Layer 3 remote port mirroring in encapsulation parameter mode, perform the following tasks:

1. On the source device, create a local mirroring group and configure the mirroring sources, the monitor port, and the encapsulation parameters for mirrored packets.
2. On all devices from source to destination, configure a unicast routing protocol to ensure Layer 3 reachability between the devices.

Create a local mirroring group on the source device, and specify the source ports and monitor port for the local mirroring group. When configuring the monitor port, specify the following encapsulation parameters for mirrored packets:

- Monitoring device IP address as destination IP address.
- Monitor port IP address as source IP address.

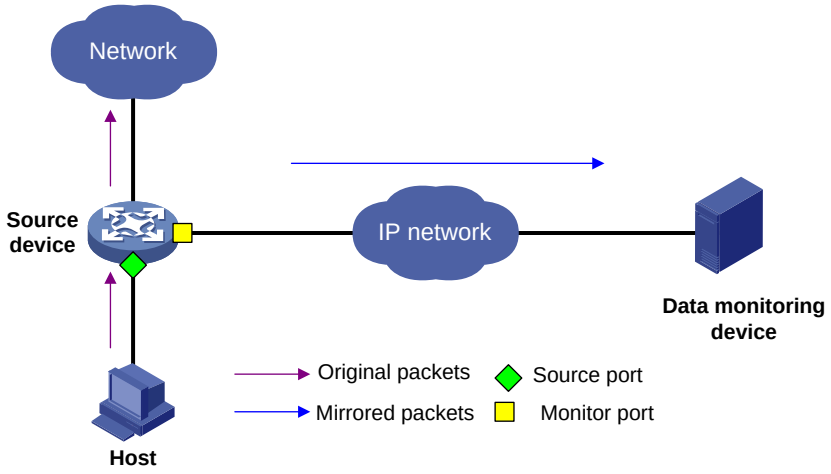
As shown in Figure 3, Layer 3 remote port mirroring in encapsulation parameter mode works as follows:

3. The source device copies a packet passing through a source port.

4. The source device encapsulates the copied packet with the specified encapsulation parameters, monitoring device IP as destination IP and monitor port IP as source IP.
5. The encapsulated packet is routed to the monitoring device through the IP network.
6. The monitoring device decapsulates the packet and analyzes the packet contents.

   The packet sent to the monitoring device through Layer 3 remote port mirroring in encapsulation parameter mode is encapsulated. In this mode, make sure the monitoring device supports decapsulating packets.

Figure 3 Layer 3 remote port mirroring in encapsulation parameter mode



Restrictions and guidelines: Port mirroring configuration

The device supports a maximum of nine mirroring groups: four mirroring groups with ports as mirroring sources, four mirroring groups with VLANs as mirroring sources, and one mirroring group with CPUs as mirroring sources.

In a mirroring group, if an aggregation member port is configured as a source port, and an aggregate interface or aggregation member port is configured as the monitor port, make sure one of the corresponding aggregation groups operate in static aggregation mode. For more about the static aggregation mode, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*.

When a mirroring group contains multiple monitor ports and the monitor ports are aggregation member ports, the mirrored packets will be sent to the related aggregate interfaces. Aggregation member ports not included in the mirroring group will also receive the mirrored packets.

Configuring local port mirroring (SPAN)

Restrictions and guidelines for local port mirroring configuration

A local mirroring group takes effect only after it is configured with the monitor port and mirroring sources.

For port mirroring to work correctly, do not assign a source port to a source VLAN.

Local port mirroring tasks at a glance

To configure local port mirroring, perform the following tasks:

1. Configuring mirroring sources

   Choose one of the following tasks:

   - Configuring source ports
   - Configuring source VLANs
   - Configuring source CPUs
2. Configuring the monitor port

Creating a local mirroring group

1. Enter system view.

   system-view
2. Create a local mirroring group.

   mirroring-group *group-id* **local**

   By default, no local mirroring group is created.

Configuring mirroring sources

Restrictions and guidelines for mirroring source configuration

When you configure source ports for a local mirroring group, follow these restrictions and guidelines:

- A mirroring group can contain multiple source ports.
- A port can act as a source port for only one mirroring group.
- A source port cannot be configured as an egress or monitor port.

When you configure source VLANs for a local mirroring group, follow these restrictions and guidelines:

- A mirroring group can contain multiple source VLANs.
- A VLAN can be configured as a source VLAN for only one local mirroring group.

A local mirroring group can contain multiple source CPUs.

Configuring source ports

- Configure source ports in system view:
  a. Enter system view.

     system-view
  b. Configure source ports for a local mirroring group.

     mirroring-group *group-id* **mirroring-port** *interface-list* { **both** | **inbound** | **outbound** }

     By default, no source port is configured for a local mirroring group.
- Configure source ports in interface view:
  c. Enter system view.

     system-view
  d. Enter interface view.

interface *interface-type interface-number*

e. Configure the port as a source port for a local mirroring group.

mirroring-group *group-id* **mirroring-port** { **both** | **inbound** | **outbound** }

By default, a port does not act as a source port for any local mirroring groups.

Configuring source VLANs

1. Enter system view.

system-view

2. Configure source VLANs for a local mirroring group.

mirroring-group *group-id* **mirroring-vlan** *vlan-list* { **both** | **inbound** | **outbound** }

By default, no source VLAN is configured for a local mirroring group.

Configuring source CPUs

1. Enter system view.

system-view

2. Configure source CPUs for a local mirroring group.

mirroring-group *group-id* **mirroring-cpu slot** *slot-number-list* { **both** | **inbound** | **outbound** }

By default, no source CPU is configured for a local mirroring group.

Configuring the monitor port

Restrictions and guidelines

Do not enable the spanning tree feature on the monitor port.

For a Layer 2 aggregate interface configured as the monitor port of a mirroring group, do not configure its member ports as source ports of the mirroring group.

For a Layer 2 aggregate interface configured as the monitor port of a mirroring group, do not assign its member ports to a source VLAN of the mirroring group.

Use a monitor port only for port mirroring, so the data monitoring device receives only the mirrored traffic.

The monitor port cannot be any member port of an aggregation group.

When you configure a member port of a dynamic aggregation group as the monitor port for a mirroring group, this command does not take effect.

Procedure

- Configure the monitor port in system view:

  f. Enter system view.

  system-view

  g. Configure the monitor port for a local mirroring group.

  mirroring-group *group-id* **monitor-port** *interface-list*

  By default, no monitor port is configured for a local mirroring group.

- Configure the monitor port in interface view:

  h. Enter system view.

  system-view

  i. Enter interface view.

interface *interface-type interface-number*

j. Configure the port as the monitor port for a mirroring group.
mirroring-group *group-id* **monitor-port**

By default, a port does not act as the monitor port for any local mirroring groups.

Configuring Layer 2 remote port mirroring (RSPAN)

Restrictions and guidelines for Layer 2 remote port mirroring configuration

To ensure successful traffic mirroring, configure devices in the order of the destination device, the intermediate devices, and the source device.

For port mirroring to work correctly, do not assign a source port, monitor port, or egress port to a source VLAN.

If intermediate devices exist, configure the intermediate devices to allow the remote probe VLAN to pass through.

For a mirrored packet to successfully arrive at the remote destination device, make sure its VLAN ID is not removed or changed.

Do not configure both MVRP and Layer 2 remote port mirroring. Otherwise, MVRP might register the remote probe VLAN with incorrect ports, which would cause the monitor port to receive undesired copies. For more information about MVRP, see *Layer 2—LAN Switching Configuration Guide*.

To monitor the bidirectional traffic of a source port, disable MAC address learning for the remote probe VLAN on the source, intermediate, and destination devices. For more information about MAC address learning, see *Layer 2—LAN Switching Configuration Guide*.

Layer 2 remote port mirroring with egress port configuration task list

Configuring the destination device

1. Creating a remote destination group
2. Configuring the monitor port
3. Configuring the remote probe VLAN
4. Assigning the monitor port to the remote probe VLAN

Configuring the source device

1. Creating a remote source group
2. Configuring mirroring sources

   Choose one of the following tasks:
   - Configuring source ports
   - Configuring source VLANs
   - Configuring source CPUs
3. Configuring the egress port
4. Configuring the remote probe VLAN

Creating a remote destination group

Restrictions and guidelines

Perform this task on the destination device only.

Procedure

1. Enter system view.
   system-view
2. Create a remote destination group.
   mirroring-group *group-id* **remote-destination**

Configuring the monitor port

Restrictions and guidelines for monitor port configuration

Perform this task on the destination device only.

Do not enable the spanning tree feature on the monitor port.

For a Layer 2 aggregate interface configured as the monitor port of a mirroring group, do not configure its member ports as source ports of the mirroring group.

For a Layer 2 aggregate interface configured as the monitor port of a mirroring group, do not assign its member ports to a source VLAN of the mirroring group.

Use a monitor port only for port mirroring, so the data monitoring device receives only the mirrored traffic.

A monitor port can belong to only one mirroring group.

An aggregation group member port cannot be configured as the monitor port for a mirroring group.

When you configure a member port of a dynamic aggregation group as the monitor port for a mirroring group, this command does not take effect.

Configuring the monitor port in system view

1. Enter system view.
   system-view
2. Configure the monitor port for a remote destination group.
   mirroring-group *group-id* **monitor-port** *interface-list*
   By default, no monitor port is configured for a remote destination group.

Configuring the monitor port in interface view

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure the port as the monitor port for a remote destination group.
   mirroring-group *group-id* **monitor-port**
   By default, a port does not act as the monitor port for any remote destination groups.

Configuring the remote probe VLAN

Restrictions and guidelines

This task is required on the both the source and destination devices.

Only an existing static VLAN can be configured as a remote probe VLAN.

When a VLAN is configured as a remote probe VLAN, use the remote probe VLAN for port mirroring exclusively.

Configure the same remote probe VLAN for the remote source group and the remote destination group.

Procedure

1. Enter system view.
   system-view
2. Configure the remote probe VLAN for the remote source or destination group.
   mirroring-group *group-id* **remote-probe vlan** *vlan-id*
   By default, no remote probe VLAN is configured for a remote source or destination group.

Assigning the monitor port to the remote probe VLAN

Restrictions and guidelines

Perform this task on the destination device only.

Procedure

1. Enter system view.
   system-view
2. Enter the interface view of the monitor port.
   interface *interface-type interface-number*
3. Assign the port to the remote probe VLAN.
   - Assign an access port to the remote probe VLAN.
     port access vlan *vlan-id*
   - Assign a trunk port to the remote probe VLAN.
     port trunk permit vlan *vlan-id*
   - Assign a hybrid port to the remote probe VLAN.
     port hybrid vlan *vlan-id* { **tagged** | **untagged** }

   For more information about the **port access vlan**, **port trunk permit vlan**, and **port hybrid vlan** commands, see *Layer 2—LAN Switching Command Reference*.

Creating a remote source group

Restrictions and guidelines

Perform this task on the source device only.

Procedure

1. Enter system view.
   system-view
2. Create a remote source group.

mirroring-group *group-id* **remote-source**

Configuring mirroring sources

Restrictions and guidelines for mirroring source configuration

Perform this task on the source device only.

When you configure source ports for a remote source group, follow these restrictions and guidelines:

- Do not assign a source port of a mirroring group to the remote probe VLAN of the mirroring group.
- A mirroring group can contain multiple source ports.
- A source port cannot be configured as a monitor port, or egress port.

When you configure source VLANs for a remote source group, follow these restrictions and guidelines:

- A remote source group can contain multiple source VLANs.
- A VLAN can be configured as the source VLAN for only one mirroring group.

Configuring source ports

- Configure source ports in system view:
  k. Enter system view.
  system-view
  l. Configure source ports for a remote source group.
  mirroring-group *group-id* **mirroring-port** *interface-list* { **both** | **inbound** | **outbound** }
  By default, no source port is configured for a remote source group.
- Configure source ports in interface view:
  m. Enter system view.
  system-view
  n. Enter interface view.
  interface *interface-type interface-number*
  o. Configure the port as a source port for a remote source group.
  mirroring-group *group-id* **mirroring-port** { **both** | **inbound** | **outbound** }
  By default, a port does not act as a source port for any remote source groups.

Configuring source VLANs

1. Enter system view.
   system-view
2. Configure source VLANs for a remote source group.
   mirroring-group *group-id* **mirroring-vlan** *vlan-list* { **both** | **inbound** | **outbound** }
   By default, no source VLAN is configured for a remote source group.

Configuring source CPUs

1. Enter system view.
   system-view
2. Configure source CPUs for a remote source group.

mirroring-group *group-id* **mirroring-cpu slot** *slot-number-list* { **both** | **inbound** | **outbound** }

By default, no source CPU is configured for a remote source group.

Configuring the egress port

Restrictions and guidelines for egress port configuration

Perform this task on the source device only.

Disable the following features on the egress port:

- Spanning tree.
- 802.1X.
- IGMP snooping.
- Static ARP.
- MAC address learning.

A port of an existing mirroring group cannot be configured as an egress port.

A mirroring group supports only one egress port.

Configuring the egress port in system view

1. Enter system view.

   system-view
2. Configure the egress port for a remote source group.

   mirroring-group *group-id* **monitor-egress** *interface-type interface-number*

   By default, no egress port is configured for a remote source group.
3. Enter the egress port view.

   interface *interface-type interface-number*
4. Assign the egress port to the remote probe VLAN.
   - Assign a trunk port to the remote probe VLAN.

     port trunk permit vlan *vlan-id*
   - Assign a hybrid port to the remote probe VLAN.

     port hybrid vlan *vlan-id* { **tagged** | **untagged** }

   For more information about the **port trunk permit vlan** and **port hybrid vlan** commands, see *Layer 2—LAN Switching Command Reference*.

Configuring the egress port in interface view

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Configure the port as the egress port for a remote source group.

   mirroring-group *group-id* **monitor-egress**

   By default, a port does not act as the egress port for any remote source groups.

Configuring Layer 3 remote port mirroring (ERSPAN in encapsulation parameter mode)

Restrictions and guidelines for Layer 3 remote port mirroring in ERSPAN mode configuration

To implement ERSPAN in encapsulation parameter mode, perform the following tasks:

1. On the source device, create a local mirroring group and configure the source ports or CPUs, the monitor port, and the encapsulation parameters for mirrored packets.
2. On all devices from source to destination, configure a unicast routing protocol to ensure Layer 3 reachability between the devices.

For Layer 3 remote port mirroring to work correctly, do not assign a source port or monitor port to a source VLAN.

For ERSPAN in encapsulation parameter mode, the data monitoring device must be able to remove the outer headers to obtain the original mirrored packets for analysis.

Layer 3 remote port mirroring tasks at a glance

To configure Layer 3 remote port mirroring in ERSPAN mode, perform the following tasks:

1. Creating a local mirroring group on the source device
2. Configuring mirroring sources

   Choose one of the following tasks:
   - Configuring source ports
   - Configuring source VLANs
   - Configuring source CPUs
3. Configuring the monitor port

Creating a local mirroring group on the source device

1. Enter system view.

   system-view
2. Create a local mirroring group.

   mirroring-group *group-id* **local**

   By default, no local mirroring group exists.

Configuring mirroring sources

Restrictions and guidelines for mirroring source configuration

When you configure source ports for the local mirroring group, follow these restrictions and guidelines:

- A source port cannot be configured as a egress port, or monitor port.

When you configure source VLANs for the local mirroring group, follow these restrictions and guidelines:

- To monitor the packets (incoming, outgoing, or both) of a VLAN passing through the source device, specify the VLAN as a source VLAN.
- A VLAN can act as a source VLAN for only one mirroring group.
- A local mirroring group can contain multiple source VLANs.

A local mirroring group can contain multiple source CPUs.

Configuring source ports

- Configure source ports in system view:
  a. Enter system view.
     system-view
  b. Configure source ports for a local mirroring group.
     mirroring-group *group-id* **mirroring-port** *interface-list* { **both** | **inbound** | **outbound** }
     By default, no source port is configured for a local mirroring group.
- Configure source ports in interface view:
  a. Enter system view.
     system-view
  b. Enter interface view.
     interface *interface-type interface-number*
  c. Configure the port as a source port for a local mirroring group.
     mirroring-group *group-id* **mirroring-port** { **both** | **inbound** | **outbound** }

Configuring source VLANs

1. Enter system view.
   system-view
2. Configure source VLANs for a local mirroring group.
   mirroring-group *group-id* **mirroring-vlan** *vlan-list* { **both** | **inbound** | **outbound** }
   By default, no source VLAN is configured for a local mirroring group.

Configuring source CPUs

1. Enter system view.
   system-view
2. Configure source CPUs for a local mirroring group.
   mirroring-group *group-id* **mirroring-cpu slot** *slot-number-list* { **both** | **inbound** | **outbound** }
   By default, no source CPU is configured for a local mirroring group.

Configuring the monitor port

Restrictions and guidelines

Do not enable the spanning tree feature on the monitor port.

Use a monitor port only for port mirroring, so the data monitoring device receives only the mirrored traffic.

An aggregation group member port cannot be configured as the monitor port for a mirroring group.

Procedure

- Configure the monitor port in system view:
  a. Enter system view.
     system-view
  b. Configure the monitor port in a local mirroring group and specify the encapsulation parameters.
     mirroring-group *group-id* **monitor-port** *interface-list* { **destination-ip** *destination-ip-address* **source-ip** *source-ip-address* | **destination-ipv6** *destination-ipv6-address* **source-ipv6** *source-ipv6-address* } [**dscp** *dscp-value* | **vlan** *vlan-id* | **vrf-instance** *vrf-name*] *
     By default, no monitor port is configured for a local mirroring group.
- Configure the monitor port in interface view:
  a. Enter system view.
     system-view
  b. Enter interface view.
     interface *interface-type interface-number*
  c. Specify the port as the monitor port in a local mirroring group and configure the encapsulation parameters in a local mirroring group.
     mirroring-group *group-id* **monitor-port** { **destination-ip** *destination-ip-address* **source-ip** *source-ip-address* | **destination-ipv6** *destination-ipv6-address* **source-ipv6** *source-ipv6-address* } [**dscp** *dscp-value* | **vlan** *vlan-id* | **vrf-instance** *vrf-name*] *
     By default, a port does not act as the monitor port for any local mirroring groups.

Display and maintenance commands for port mirroring

Execute **display** commands in any view.

Task	Command
Display mirroring group information.	**display mirroring-group** { *group-id* | **all** | **local** | **remote-destination** | **remote-source** }

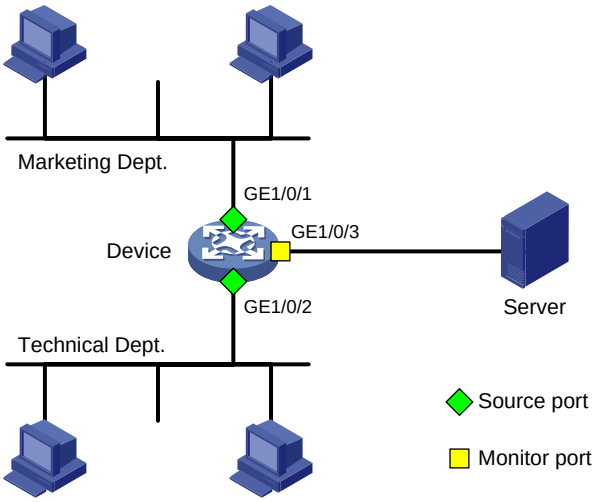
Port mirroring configuration examples

Example: Configuring local port mirroring (SPAN in source port mode)

Network configuration

As shown in Figure 4, configure local port mirroring in source port mode to enable the server to monitor the bidirectional traffic of the two departments.

Figure 4 Network diagram



Procedure

Create local mirroring group 1.

```
<Device> system-view
[Device] mirroring-group 1 local
```

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as source ports for local mirroring group 1.

```
[Device] mirroring-group 1 mirroring-port gigabitethernet 1/0/1 gigabitethernet 1/0/2
both
```

Configure GigabitEthernet 1/0/3 as the monitor port for local mirroring group 1.

```
[Device] mirroring-group 1 monitor-port gigabitethernet 1/0/3
```

Disable the spanning tree feature on the monitor port (GigabitEthernet 1/0/3).

```
[Device] interface gigabitethernet 1/0/3
[Device-GigabitEthernet1/0/3] undo stp enable
[Device-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Verify the mirroring group configuration.

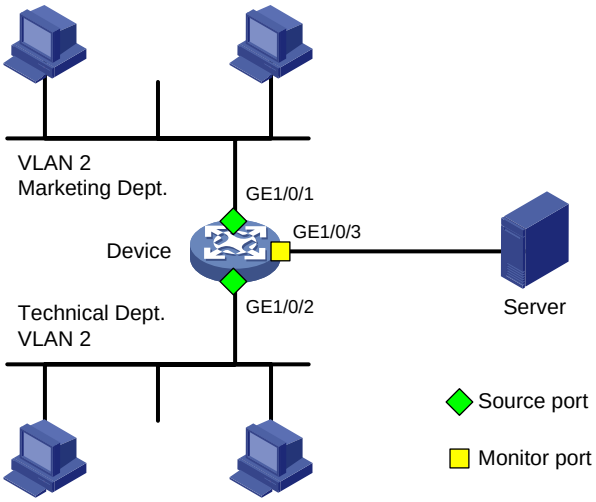
```
[Device] display mirroring-group all
Mirroring group 1:
    Type: Local
    Status: Active
    Mirroring port:
        GigabitEthernet1/0/1  Both
        GigabitEthernet1/0/2  Both
    Monitor port: GigabitEthernet1/0/3
```

Example: Configuring local port mirroring (SPAN in source VLAN mode)

Network configuration

As shown in Figure 5, configure local port mirroring in source VLAN mode to enable the server to monitor the bidirectional traffic of the two departments.

Figure 5 Network diagram



Procedure

Create local mirroring group 1.

```
<Device> system-view
[Device] mirroring-group 1 local
```

Create VLAN 2, and assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to VLAN 2.

```
[Device] vlan 2
[Device-vlan2] port gigabitethernet 1/0/1 gigabitethernet 1/0/2
[Device-vlan2] quit
```

Configure VLAN 2 as a source VLAN for local mirroring group 1.

```
[Device] mirroring-group 1 mirroring-vlan 2 both
```

Configure GigabitEthernet 1/0/3 as the monitor port for local mirroring group 1.

```
[Device] mirroring-group 1 monitor-port gigabitethernet 1/0/3
```

Disable the spanning tree feature on the monitor port (GigabitEthernet 1/0/3).

```
[Device] interface gigabitethernet 1/0/3
[Device-GigabitEthernet1/0/3] undo stp enable
[Device-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Verify the mirroring group configuration.

```
[Device] display mirroring-group all
Mirroring group 1:
    Type: Local
    Status: Active
    Mirroring VLAN:
        2  Both
    Monitor port: GigabitEthernet1/0/3
```

Example: Configuring local port mirroring (SPAN in source CPU mode)

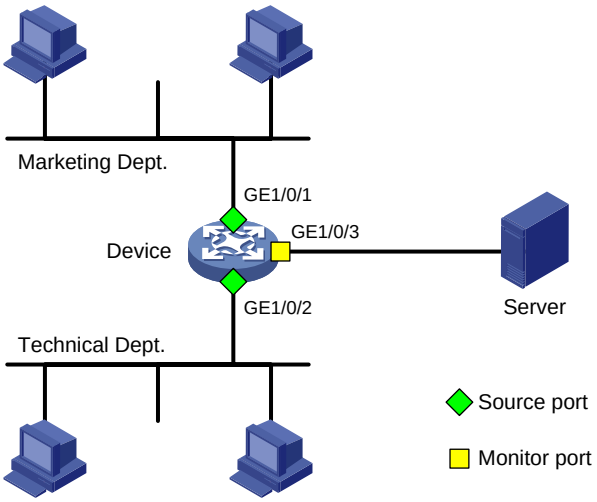
Network configuration

As shown in Figure 6, GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are located on the card in slot 1.

Configure local port mirroring in source CPU mode to enable the server to monitor all packets matching the following criteria:

- Received and sent by the Marketing Department and the Technical Department.
- Processed by the CPU in slot 1 of the device.

Figure 6 Network diagram



Procedure

Create local mirroring group 1.

```
<Device> system-view
[Device] mirroring-group 1 local
```

Configure the CPU in slot 1 of the device as a source CPU for local mirroring group 1.

```
[Device] mirroring-group 1 mirroring-cpu slot 1 both
```

Configure GigabitEthernet 1/0/3 as the monitor port for local mirroring group 1.

```
[Device] mirroring-group 1 monitor-port gigabitethernet 1/0/3
```

Disable the spanning tree feature on the monitor port (GigabitEthernet 1/0/3).

```
[Device] interface gigabitethernet 1/0/3
[Device-GigabitEthernet1/0/3] undo stp enable
[Device-GigabitEthernet1/0/3] quit
```

Verifying the configuration

Verify the mirroring group configuration.

```
[Device] display mirroring-group all
Mirroring group 1:
    Type: Local
    Status: Active
    Mirroring CPU:
```

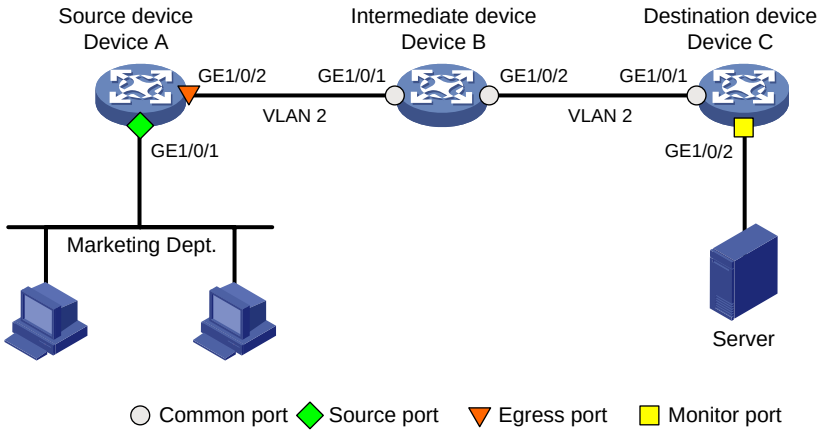
```
    Slot 1  Both
Monitor port: GigabitEthernet1/0/3
```

Example: Configuring Layer 2 remote port mirroring (RSPAN with egress port)

Network configuration

On the Layer 2 network shown in Figure 7, configure Layer 2 remote port mirroring to enable the server to monitor the bidirectional traffic of the Marketing Department.

Figure 7 Network diagram



Procedure

1. Configure Device C (the destination device):

   # Configure GigabitEthernet 1/0/1 as a trunk port, and assign the port to VLAN 2.

   ```
   <DeviceC> system-view
   [DeviceC] interface gigabitethernet 1/0/1
   [DeviceC-GigabitEthernet1/0/1] port link-type trunk
   [DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 2
   [DeviceC-GigabitEthernet1/0/1] quit
   ```

   # Create a remote destination group.

   ```
   [DeviceC] mirroring-group 2 remote-destination
   ```

   # Create VLAN 2.

   ```
   [DeviceC] vlan 2
   ```

   # Disable MAC address learning for VLAN 2.

   ```
   [DeviceC-vlan2] undo mac-address mac-learning enable
   [DeviceC-vlan2] quit
   ```

   # Configure VLAN 2 as the remote probe VLAN for the mirroring group.

   ```
   [DeviceC] mirroring-group 2 remote-probe vlan 2
   ```

   # Configure GigabitEthernet 1/0/2 as the monitor port for the mirroring group.

   ```
   [DeviceC] interface gigabitethernet 1/0/2
   [DeviceC-GigabitEthernet1/0/2] mirroring-group 2 monitor-port
   ```

   # Disable the spanning tree feature on GigabitEthernet 1/0/2.

   ```
   [DeviceC-GigabitEthernet1/0/2] undo stp enable
   ```

   # Assign GigabitEthernet 1/0/2 to VLAN 2 as an access port.

```
[DeviceC-GigabitEthernet1/0/2] port access vlan 2
[DeviceC-GigabitEthernet1/0/2] quit
```

2. Configure Device B (the intermediate device):

Create VLAN 2.

```
<DeviceB> system-view
[DeviceB] vlan 2
```

Disable MAC address learning for VLAN 2.

```
[DeviceB-vlan2] undo mac-address mac-learning enable
[DeviceB-vlan2] quit
```

Configure GigabitEthernet 1/0/1 as a trunk port, and assign the port to VLAN 2.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 2
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 as a trunk port, and assign the port to VLAN 2.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 2
[DeviceB-GigabitEthernet1/0/2] quit
```

3. Configure Device A (the source device):

Create a remote source group.

```
<DeviceA> system-view
[DeviceA] mirroring-group 1 remote-source
```

Create VLAN 2.

```
[DeviceA] vlan 2
```

Disable MAC address learning for VLAN 2.

```
[DeviceA-vlan2] undo mac-address mac-learning enable
[DeviceA-vlan2] quit
```

Configure VLAN 2 as the remote probe VLAN of the mirroring group.

```
[DeviceA] mirroring-group 1 remote-probe vlan 2
```

Configure GigabitEthernet 1/0/1 as a source port for the mirroring group.

```
[DeviceA] mirroring-group 1 mirroring-port gigabitethernet 1/0/1 both
```

Configure GigabitEthernet 1/0/2 as the egress port for the mirroring group.

```
[DeviceA] mirroring-group 1 monitor-egress gigabitethernet 1/0/2
```

Configure GigabitEthernet 1/0/2 as a trunk port, and assign the port to VLAN 2.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 2
```

Disable the spanning tree feature on the port.

```
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] quit
```

Verifying the configuration

Verify the mirroring group configuration on Device C.

```
[DeviceC] display mirroring-group all
Mirroring group 2:
    Type: Remote destination
```

```
    Status: Active
    Monitor port: GigabitEthernet1/0/2
    Remote probe VLAN: 2
```

Verify the mirroring group configuration on Device A.

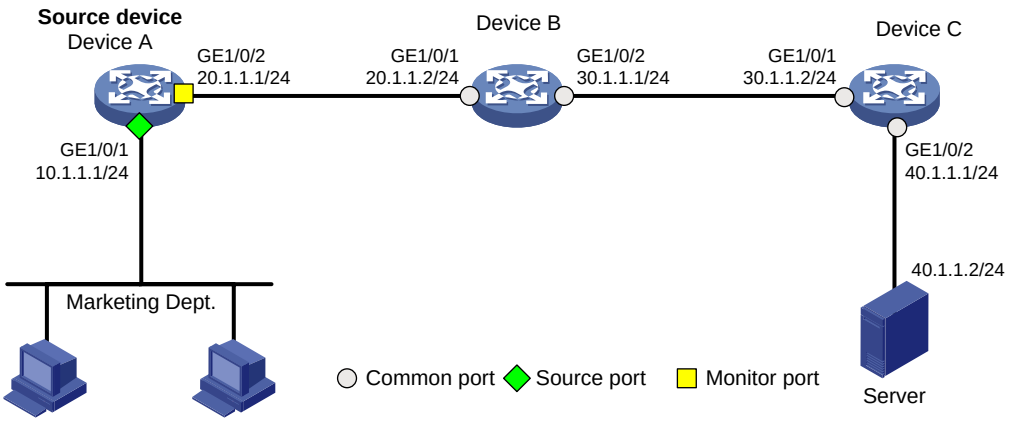
```
[DeviceA] display mirroring-group all
Mirroring group 1:
    Type: Remote source
    Status: Active
    Mirroring port:
        GigabitEthernet1/0/1  Both
    Monitor egress port: GigabitEthernet1/0/2
    Remote probe VLAN: 2
```

Example: Configuring Layer 3 remote port mirroring (ERSPAN in encapsulation parameter mode)

Network configuration

On a Layer 3 network shown in Figure 8, configure Layer 3 remote port mirroring in encapsulation parameter mode to enable the server to monitor the bidirectional traffic of the Marketing Department.

Figure 8 Network diagram



Procedure

1. Configure IP addresses for the interfaces as shown in Figure 8. (Details not shown.)
2. Configure Device A (the source device):

   # Enable the OSPF protocol.

   ```
   [DeviceA] ospf 1
   [DeviceA-ospf-1] area 0
   [DeviceA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
   [DeviceA-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
   [DeviceA-ospf-1-area-0.0.0.0] quit
   [DeviceA-ospf-1] quit
   ```

   # Create local mirroring group 1.

   ```
   [DeviceA] mirroring-group 1 local
   ```

   # Configure GigabitEthernet 1/0/1 as a source port.

```
[DeviceA] mirroring-group 1 mirroring-port gigabitethernet 1/0/1 both
```

Configure GigabitEthernet 1/0/2 as the monitor port. Specify the destination and source IP addresses for mirrored packets as 40.1.1.2 and 20.1.1.1, respectively.

```
[DeviceA] mirroring-group 1 monitor-port gigabitethernet 1/0/2 destination-ip
40.1.1.2 source-ip 20.1.1.1
```

3. Enable the OSPF protocol on Device B.

```
<DeviceB> system-view
[DeviceB] ospf 1
[DeviceB-ospf-1] area 0
[DeviceB-ospf-1-area-0.0.0.0] network 20.1.1.0 0.0.0.255
[DeviceB-ospf-1-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[DeviceB-ospf-1-area-0.0.0.0] quit
[DeviceB-ospf-1] quit
```

4. Enable the OSPF protocol on Device C.

```
[DeviceC] ospf 1
[DeviceC-ospf-1] area 0
[DeviceC-ospf-1-area-0.0.0.0] network 30.1.1.0 0.0.0.255
[DeviceC-ospf-1-area-0.0.0.0] network 40.1.1.0 0.0.0.255
[DeviceC-ospf-1-area-0.0.0.0] quit
[DeviceC-ospf-1] quit
```

Verifying the configuration

Verify the mirroring group configuration on Device A.

```
[DeviceA] display mirroring-group all
Mirroring group 1:
    Type: Local
    Status: Active
    Mirroring port:
        GigabitEthernet1/0/1  Both
    Monitor port: GigabitEthernet1/0/2
                  Encapsulation: Destination IP address 40.1.1.2
                                 Source IP address 20.1.1.1
                                 Destination MAC address 000f-e241-5e5b
```

Configuring flow mirroring

About flow mirroring

Flow mirroring copies packets matching a class to a destination for packet analyzing and monitoring. It is implemented through QoS.

To implement flow mirroring through QoS, perform the following tasks:

- Define traffic classes and configure match criteria to classify packets to be mirrored. Flow mirroring allows you to flexibly classify packets to be analyzed by defining match criteria.
- Configure traffic behaviors to mirror the matching packets to the specified destination.

You can configure an action to mirror the matching packets to one of the following destinations:

- **Interface**—The matching packets are copied to an interface and then forwarded to a data monitoring device for analysis.
- **CPU**—The matching packets are copied to the CPU of an IRF member device. The CPU analyzes the packets or delivers them to upper layers.

For more information about QoS policies, traffic classes, and traffic behaviors, see *ACL and QoS Configuration Guide*.

Types of flow-mirroring traffic to an interface

Depending on whether the mirroring source and mirroring destination are on the same device, flow-mirroring traffic to an interface includes the following types:

- **Flow mirroring SPAN**—Flow-mirrors traffic to a local interface.
- **Flow mirroring RSPAN**—Flow-mirrors traffic to an interface, and then forwards traffic to a remote Layer 2 interface based on the VLAN of the mirrored traffic or through QoS traffic redirecting.
- **Flow mirroring ERSPAN**—Encapsulates traffic in GRE packets with protocol number 0x88BE (ERSPANv2) and routes the traffic to a remote monitoring device at Layer 3.

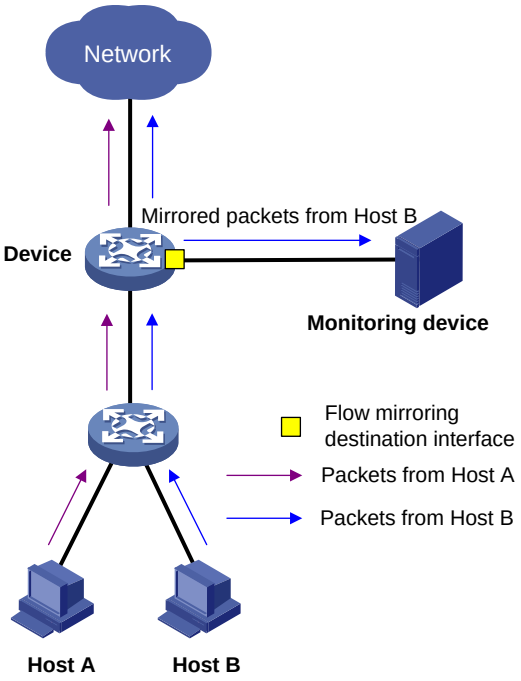
Flow mirroring SPAN or RSPAN

For flow mirroring SPAN, configure a QoS policy on the source device. Configure the QoS policy as follows:

1. Configure a traffic class to match packets.
2. Configure a traffic behavior to flow-mirror traffic to an interface without specifying the **`destination-ip`** or **`source-ip`** keyword.
3. Associate the traffic class with the traffic behavior.

When the device receives a matching packet, the device sends one copy of the packet to the interface specified by the traffic behavior. The interface forwards the mirrored packet to the monitoring device.

Figure 9 Flow mirroring SPAN



To implement RSPAN, forward the mirrored packet to a remote Layer 2 interface based on the VLAN of the mirrored packet or through QoS traffic redirecting.

Flow mirroring ERSPAN

Flow mirroring ERSPAN can be implemented in encapsulation parameter mode or monitoring group mode.

On all devices from source to destination, configure a unicast routing protocol to ensure Layer 3 reachability between the devices.

Encapsulation parameter mode

In this mode, configure a QoS policy on the source device. Configure the QoS policy as follows:

1. Configure a traffic class to match packets.
2. Configure a traffic behavior to flow-mirror traffic to an interface.
3. Associate the traffic class with the traffic behavior.

You can configure flow-mirroring traffic to an interface in one of the following modes:

- **Directly specifying an outgoing interface**—In this mode, specify both the outgoing interface and encapsulation parameters. The device encapsulates packets with the specified parameters and then forwards packets out of the specified interface.
- **Specifying an outgoing interface through route lookup**—In this mode, specify only encapsulation parameters without specifying an outgoing interface. The device looks up a route for the encapsulated mirrored packets based on the source IP address and destination IP address of the encapsulated packets. The outgoing interface of the route is a destination interface of the mirrored packets.

  In this mode, you can use the load sharing function of a routing protocol to forward mirrored packets to multiple destination interfaces.

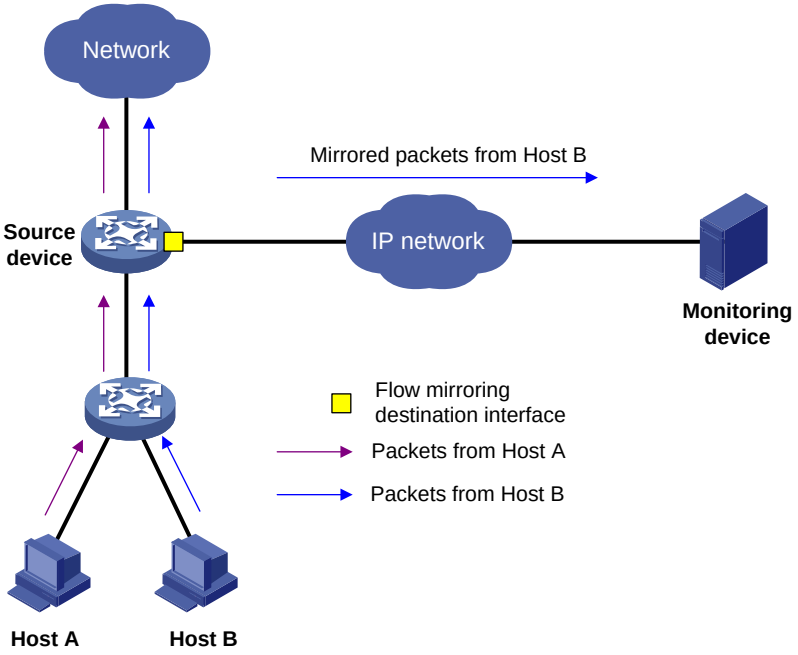
As shown in Figure 10, flow mirroring ERSPAN in encapsulation parameter mode works as follows:

4. The source device copies a matching packet.

5. The source device encapsulates the packet with the specified ERSPAN encapsulation parameters.
6. The source device forwards the packet in either of the following methods:
   - Forwards the mirrored packets out of the specified outgoing interface.
   - Looks up a route for the encapsulated mirrored packet based on the source IP address and destination IP address of the encapsulated packet.
7. The encapsulated packet is routed to the monitoring device.
8. The monitoring device decapsulates the packet and analyzes the packet contents.

   The packet sent to the monitoring device through flow mirroring in this mode is encapsulated. In this mode, make sure the monitoring device supports decapsulating packets.

Figure 10 Flow mirroring ERSPAN in encapsulation parameter mode



Restrictions and guidelines: Flow mirroring configuration

For information about the configuration commands except the **`mirror-to`** command, see *ACL and QoS Command Reference*.

Flow mirroring tasks at a glance

To configure flow mirroring, perform the following tasks:

1. Configuring a traffic class

   A traffic class defines the criteria that filters the traffic to be mirrored.
2. Configuring a traffic behavior

   A traffic behavior specifies mirroring destinations.
3. Configuring a QoS policy
4. Applying a QoS policy

Choose one of the following tasks:

- Applying a QoS policy to an interface
- Applying a QoS policy to a VLAN
- Applying a QoS policy globally

Configuring a traffic class

1. Enter system view.

   system-view

2. Create a class and enter class view.

   traffic classifier *classifier-name* [**operator** { **and** | **or** }]

3. Configure match criteria.

   if-match *match-criteria*

   By default, no match criterion is configured in a traffic class.

4. (Optional.) Display traffic class information.

   display traffic classifier user-defined [*classifier-name*] [**slot** *slot-number*]

   For more information about this command, see *ACL and QoS Command Reference*.

Configuring a traffic behavior

Procedure

1. Enter system view.

   system-view

2. Configure a monitoring group. This step is required for mirroring traffic to a monitoring group.

   d. Create a monitoring group.

   monitoring-group *group-id*

   e. Assign ports to the monitoring group.

   Syntax I:

   monitoring-port *interface-list* [{ **destination-ip** *destination-ip-address* **source-ip** *source-ip-address* | **destination-ipv6** *destination-ipv6-address* **source-ipv6** *source-ipv6-address* } [**dscp** *dscp-value* | **vlan** *vlan-id* | **vrf-instance** *vrf-name*] * [**destination-mac** *mac-address*]]

   Syntax II:

   monitoring-port { **destination-ip** *destination-ip-address* **source-ip** *source-ip-address* | **destination-ipv6** *destination-ipv6-address* **source-ipv6** *source-ipv6-address* } [**dscp** *dscp-value* | **vlan** *vlan-id* | **vrf-instance** *vrf-name*] * [**destination-mac** *mac-address*]

   By default, a monitoring group does not contain any port.

   If you specify encapsulation parameters when flow-mirroring traffic to a monitoring group, the encapsulation parameters for flow-mirroring traffic to the monitoring group member ports must be the same as the encapsulation parameters of the first member port.

   f. Return to system view.

   quit

3. Create a traffic behavior and enter traffic behavior view.

traffic behavior *behavior-name*

4. Configure mirroring destinations for the traffic behavior. Choose one option as needed:
   - Mirror traffic to interfaces.

     S**yntax I:**

     mirror-to interface *interface-type interface-number* [**backup-interface** *interface-type interface-number*] [**truncation**] [{ **destination-ip** *destination-ip-address* **source-ip** *source-ip-address* | **destination-ipv6** *destination-ipv6-address* **source-ipv6** *source-ipv6-address* } [**dscp** *dscp-value* | **vlan** *vlan-id* | **vrf-instance** *vrf-name*] *] [**destination-mac** *mac-address*]]

     Syntax II:

     mirror-to interface { **destination-ip** *destination-ip-address* **source-ip** *source-ip-address* | **destination-ipv6** *destination-ipv6-address* **source-ipv6** *source-ipv6-address* }[**truncation**] [**dscp** *dscp-value* | **vlan** *vlan-id* | **vrf-instance** *vrf-name*] * [**destination-mac** *mac-address*]

     By default, no mirroring actions exist to mirror traffic to interfaces.

     You cannot flow-mirror traffic to tunnel interfaces.
   - Mirror traffic to the CPU.

     mirror-to cpu

     By default, no mirroring actions exist to mirror traffic to the CPU.
5. (Optional.) Display traffic behavior configuration.

   display traffic behavior user-defined [*behavior-name*] [**slot** *slot-number*]

   For more information about this command, see *ACL and QoS Command Reference*.

Configuring a QoS policy

1. Enter system view.

   system-view
2. Create a QoS policy and enter QoS policy view.

   qos [**mirroring**] **policy** *policy-name*
3. Associate a class with a traffic behavior in the QoS policy.

   classifier *classifier-name* **behavior** *behavior-name*

   By default, no traffic behavior is associated with a class.
4. (Optional.) Display QoS policy configuration.

   display qos policy user-defined mirroring [*policy-name* [**classifier** *classifier-name*]] [**slot** *slot-number*]

   For more information about this command, see *ACL and QoS Command Reference*.

Applying a QoS policy

Applying a QoS policy to an interface

Restrictions and guidelines

You can apply a QoS policy to an interface to mirror the traffic of the interface.

A policy can be applied to multiple interfaces.

In one traffic direction of an interface, only one QoS policy can be applied.

The device does not support mirroring traffic of an aggregate interface.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Apply a policy to the interface.
   qos apply [**mirroring**] **policy** *policy-name* { **inbound** | **outbound** }
4. (Optional.) Display the QoS policy applied to the interface.
   display qos [**mirroring**] **policy interface** [*interface-type interface-number* [**inbound** | **outbound**] [**slot** *slot-number*]
   For more information about this command, see *ACL and QoS Command Reference*.

Applying a QoS policy to a VLAN

Restrictions and guidelines

You can apply a QoS policy to a VLAN to mirror the traffic on all ports in the VLAN.

Procedure

1. Enter system view.
   system-view
2. Apply a QoS policy to a VLAN.
   qos vlan-policy *policy-name* **vlan** *vlan-id-list* { **inbound** | **outbound** }
3. (Optional.) Display the QoS policy applied to the VLAN.
   display qos vlan-policy { **name** *policy-name* | **vlan** [*vlan-id*] } [**inbound** | **outbound**] [**slot** *slot-number*]
   For more information about this command, see *ACL and QoS Command Reference*.

Applying a QoS policy globally

Restrictions and guidelines

You can apply a QoS policy globally to mirror the traffic on all ports.

Procedure

1. Enter system view.
   system-view
2. Apply a QoS policy globally.
   qos apply [**mirroring**] **policy** *policy-name* **global** { **inbound** | **outbound** }
3. (Optional.) Display global QoS policies.
   display qos [**mirroring**] **policy global** [**inbound** | **outbound**] [**slot** *slot-number*]
   For more information about this command, see *ACL and QoS Command Reference*.

Flow mirroring configuration examples

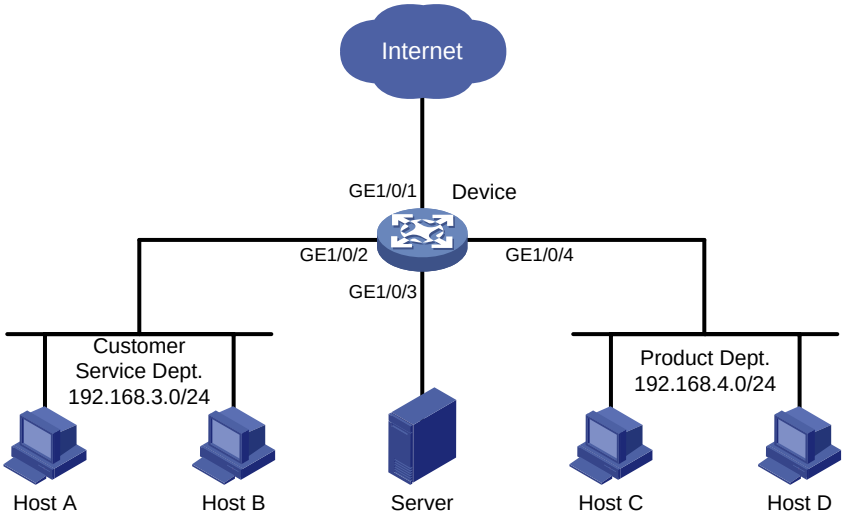
Example: Configuring flow mirroring

Network configuration

As shown in Figure 11, configure flow mirroring so that the server can monitor the following traffic:

- All traffic that the Product Department sends to access the Internet.
- IP traffic that the Product Department sends to the Customer Service Department during working hours (8:00 to 18:00) on weekdays.

Figure 11 Network diagram



Procedure

Create working hour range **work**, in which working hours are from 8:00 to 18:00 on weekdays.

```
<Device> system-view
[Device] time-range work 8:00 to 18:00 working-day
```

Create IPv4 advanced ACL 3000 to allow packets from the Product Department to access the Internet and the Customer Service Department during working hours.

```
[Device] acl advanced 3000
[Device-acl-ipv4-adv-3000] rule permit tcp source 192.168.4.0 0.0.0.255 destination-port eq www
[Device-acl-ipv4-adv-3000] rule permit ip source 192.168.4.0 0.0.0.255 destination 192.168.3.0 0.0.0.255 time-range work
[Device-acl-ipv4-adv-3000] quit
```

Create traffic class **tech_c**, and configure the match criterion as ACL 3000.

```
[Device] traffic classifier tech_c
[Device-classifier-tech_c] if-match acl 3000
[Device-classifier-tech_c] quit
```

Create traffic behavior **tech_b**, configure the action of mirroring traffic to GigabitEthernet 1/0/3.

```
[Device] traffic behavior tech_b
[Device-behavior-tech_b] mirror-to interface gigabitethernet 1/0/3
[Device-behavior-tech_b] quit
```

Create QoS policy **tech_p**, and associate traffic class **tech_c** with traffic behavior **tech_b** in the QoS policy.

```
[Device] qos policy tech_p
[Device-qospolicy-tech_p] classifier tech_c behavior tech_b
[Device-qospolicy-tech_p] quit
```

Apply QoS policy **tech_p** to the incoming packets of GigabitEthernet 1/0/4.

```
[Device] interface gigabitethernet 1/0/4
[Device-GigabitEthernet1/0/4] qos apply policy tech_p inbound
[Device-GigabitEthernet1/0/4] quit
```

Verifying the configuration

Verify that the server can monitor the following traffic:

- All traffic sent by the Product Department to access the Internet.
- IP traffic that the Product Department sends to the Customer Service Department during working hours on weekdays.

(Details not shown.)

Contents

Configuring sFlow 1

About sFlow 1
Protocols and standards 1
Restrictions and guidelines: sFlow configuration 1
Configuring basic sFlow information 2
Configuring flow sampling 2
Configuring counter sampling 3
Display and maintenance commands for sFlow 3
sFlow configuration examples 4
Example: Configuring sFlow 4
Troubleshooting sFlow 5
The remote sFlow collector cannot receive sFlow packets 5

Configuring sFlow

About sFlow

sFlow is a traffic monitoring technology.

As shown in Figure 1, the sFlow system involves an sFlow agent embedded in a device and a remote sFlow collector. The sFlow agent collects interface counter information and packet information and encapsulates the sampled information in sFlow packets. When the sFlow packet buffer is full, or the aging timer (fixed to 1 second) expires, the sFlow agent performs the following actions:

- Encapsulates the sFlow packets in the UDP datagrams.
- Sends the UDP datagrams to the specified sFlow collector.

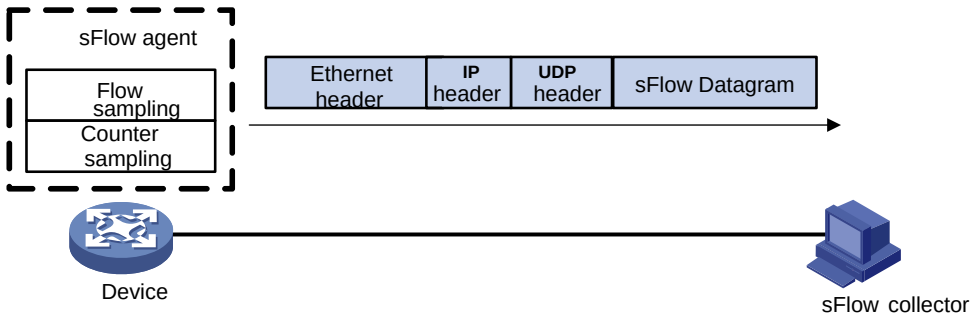
The sFlow collector analyzes the information and displays the results. One sFlow collector can monitor multiple sFlow agents.

sFlow provides the following sampling mechanisms:

- **Flow sampling**—Obtains packet information.
- **Counter sampling**—Obtains interface counter information.

sFlow can use flow sampling and counter sampling at the same time.

Figure 1 sFlow system



Protocols and standards

- RFC 3176, *InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks*
- sFlow.org, *sFlow Version 5*

Restrictions and guidelines: sFlow configuration

When you configure flow sampling or counter sampling on Layer 2 Ethernet interfaces in VLAN 1, the total bandwidth of the single or more Layer 2 Ethernet interfaces to be sampled cannot exceed 22 Gbps.

Configuring basic sFlow information

Restrictions and guidelines

As a best practice, manually configure an IP address for the sFlow agent. The device periodically checks whether the sFlow agent has an IP address. If the sFlow agent does not have an IP address, the device automatically selects an IPv4 address for the sFlow agent but does not save the IPv4 address in the configuration file.

Only one IP address can be configured for the sFlow agent on the device, and a newly configured IP address overwrites the existing one.

Procedure

1. Enter system view.

   system-view

2. Configure an IP address for the sFlow agent.

   sflow agent { **ip** *ipv4-address* | **ipv6** *ipv6-address* }

   By default, no IP address is configured for the sFlow agent.

3. Configure the sFlow collector information.

   sflow collector *collector-id* [**vpn-instance** *vpn-instance-name*] { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**datagram-size** *size* | **description** *string* | **dscp** *dscp-value* | **port** *port-number* | **time-out** *seconds*] *

   By default, no sFlow collector information is configured.

4. Specify the source IP address of sFlow packets.

   sflow source { **ip** *ipv4-address* | **ipv6** *ipv6-address* } *

   By default, the source IP address is determined by routing.

Configuring flow sampling

About this task

Perform this task to configure flow sampling on an Ethernet interface. The sFlow agent performs the following tasks:

1. Samples packets on that interface according to the configured parameters.
2. Encapsulates the packets into sFlow packets.
3. Encapsulates the sFlow packets in the UDP packets and sends the UDP packets to the specified sFlow collector.

Configuring flow sampling on an Ethernet interface

1. Enter system view.

   system-view

2. Enter Layer 2 Ethernet interface view or Layer 3 Ethernet interface view.

   interface *interface-type interface-number*

3. (Optional.) Set the flow sampling mode.

   sflow sampling-mode random

   By default, random sampling is used.

4. Enable flow sampling and specify the number of packets out of which flow sampling samples a packet on the interface.

   sflow sampling-rate *rate*

By default, flow sampling is disabled.

As a best practice, set the sampling interval to 2^n that is greater than or equal to 8192, for example, 32768.

5. (Optional.) Set the flow sampling direction.

   sflow sampling-direction { **both** | **inbound** | **outbound** }

   By default, the flow sampling direction is bidirectional.

6. (Optional.) Set the maximum number of bytes (starting from the packet header) that flow sampling can copy per packet.

   sflow flow max-header *length*

   The default setting is 128 bytes.

   As a best practice, use the default setting.

7. Specify the sFlow instance and sFlow collector for flow sampling.

   sflow flow [**instance** *instance-id*] **collector** *collector-id*

   By default, no sFlow instance or sFlow collector is specified for flow sampling.

Configuring counter sampling

About this task

Perform this task to configure counter sampling on an Ethernet interface. The sFlow agent performs the following tasks:

1. Periodically collects the counter information on that interface.
2. Encapsulates the counter information into sFlow packets.
3. Encapsulates the sFlow packets in the UDP packets and sends the UDP packets to the specified sFlow collector.

Procedure

1. Enter system view.

   system-view

2. Enter Layer 2/Layer 3 Ethernet interface view.

   interface *interface-type interface-number*

3. Enable counter sampling and set the counter sampling interval.

   sflow counter interval *interval*

   By default, counter sampling is disabled.

4. Specify the sFlow instance and sFlow collector for counter sampling.

   sflow counter [**instance** *instance-id*] **collector** *collector-id*

   By default, no sFlow instance or sFlow collector is specified for counter sampling.

Display and maintenance commands for sFlow

Execute **display** commands in any view.

Task	Command
Display sFlow configuration.	**display sflow**

sFlow configuration examples

Example: Configuring sFlow

Network configuration

As shown in Figure 2, perform the following tasks:

- Configure flow sampling in random mode and counter sampling on GigabitEthernet 1/0/1 of the device to monitor traffic on the port.
- Configure the device to send sampled information in sFlow packets through GigabitEthernet 1/0/3 to the sFlow collector.

Figure 2 Network diagram

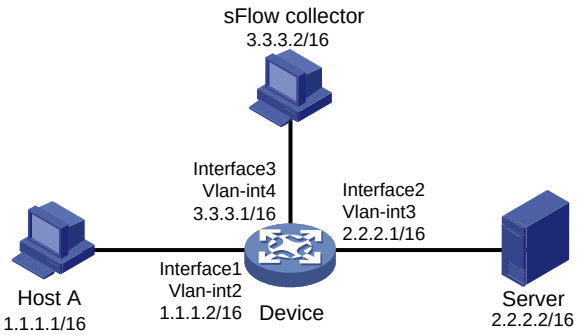


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

1. Configure VLANs, assign Ethernet interfaces to VLANs, and configure IP addresses and subnet masks for VLAN interfaces, as shown in Figure 2. (Details not shown.)
2. Configure the sFlow agent and configure information about the sFlow collector:

   # Configure the IP address for the sFlow agent.

   ```
   <Device> system-view
   [Device] sflow agent ip 3.3.3.1
   ```

   # Configure information about the sFlow collector. Specify the sFlow collector ID as 1, IP address as 3.3.3.2, port number as 6343 (default), and description as **netserver**.

   ```
   [Device] sflow collector 1 ip 3.3.3.2 description netserver
   ```

3. Configure counter sampling:

   # Enable counter sampling and set the counter sampling interval to 120 seconds on GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] sflow counter interval 120
   ```

   # Specify sFlow collector 1 for counter sampling.

   ```
   [Device-GigabitEthernet1/0/1] sflow counter collector 1
   ```

4. Configure flow sampling:

Enable flow sampling and set the flow sampling mode to random and sampling interval to 32767.

```
[Device-GigabitEthernet1/0/1] sflow sampling-mode random
[Device-GigabitEthernet1/0/1] sflow sampling-rate 32767
```

Specify sFlow collector 1 for flow sampling.

```
[Device-GigabitEthernet1/0/1] sflow flow collector 1
```

Verifying the configuration

Verify the following items:

- GigabitEthernet 1/0/1 enabled with sFlow is active.
- The counter sampling interval is 120 seconds.
- The flow sampling interval is 32767 (one packet is sampled from every 32767 packets).

```
[Device-GigabitEthernet1/0/1] display sflow
sFlow datagram version: 5
Global information:
Agent IP: 3.3.3.1(CLI)
Source address:
Collector information:
ID    IP              Port  Aging      Size VPN-instance Description       Dscp
1     3.3.3.2         6343  N/A        1400              netserver         0
Port counter sampling information:
Interface   Instance   CID   Interval(s)
GE1/0/1     1          1     120
Port flow sampling information:
Interface   Instance   FID   MaxHLen Rate       Mode      Status
GE1/0/1        1          1     128     32767      Random    Active
```

Troubleshooting sFlow

The remote sFlow collector cannot receive sFlow packets

Symptom

The remote sFlow collector cannot receive sFlow packets.

Analysis

The possible reasons include:

- The sFlow collector is not specified.
- sFlow is not configured on the interface.
- The IP address of the sFlow collector specified on the sFlow agent is different from that of the remote sFlow collector.
- No IP address is configured for the Layer 3 interface that sends sFlow packets.
- An IP address is configured for the Layer 3 interface that sends sFlow packets. However, the UDP datagrams with this source IP address cannot reach the sFlow collector.
- The physical link between the device and the sFlow collector fails.
- The sFlow collector is bound to a non-existent VPN.
- The length of an sFlow packet is less than the sum of the following two values:
  - The length of the sFlow packet header.

- The number of bytes that flow sampling can copy per packet.

Solution

To resolve the problem:

1. Use the **`display sflow`** command to verify that sFlow is correctly configured.
2. Verify that a correct IP address is configured for the device to communicate with the sFlow collector.
3. Verify that the physical link between the device and the sFlow collector is up.
4. Verify that the VPN bound to the sFlow collector already exists.
5. Verify that the length of an sFlow packet is greater than the sum of the following two values:
   - The length of the sFlow packet header.
   - The number of bytes (as a best practice, use the default setting) that flow sampling can copy per packet.

Contents

Configuring the information center 1

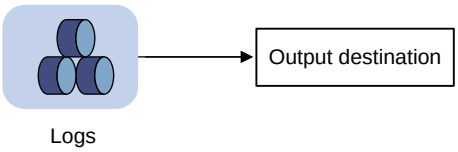
About the information center 1
Log types 1
Log levels 1
Log destinations 2
Default output rules for logs 2
Default output rules for diagnostic logs 2
Default output rules for security logs 2
Default output rules for hidden logs 3
Default output rules for trace logs 3
Log formats and field descriptions 3
FIPS compliance 7
Information center tasks at a glance 7
Managing standard system logs 7
Managing hidden logs 7
Managing security logs 7
Managing diagnostic logs 8
Managing trace logs 8
Enabling the information center 8
Outputting logs to various destinations 9
Outputting logs to the console 9
Outputting logs to the monitor terminal 9
Outputting logs to log hosts 10
Outputting logs to the log buffer 11
Saving logs to the log file 12
Setting the minimum storage period 13
About setting the minimum storage period 13
Procedure 14
Enabling synchronous information output 14
Configuring log suppression 14
Enabling duplicate log suppression 14
Configuring log suppression for a module 16
Disabling an interface from generating link up or link down logs 17
Enabling SNMP notifications for system logs 17
Managing security logs 18
Saving security logs to the security log file 18
Managing the security log file 18
Performing integrity check for a log file 19
Saving diagnostic logs to the diagnostic log file 19
Setting the maximum size of the trace log file 20
Display and maintenance commands for information center 21
Information center configuration examples 21
Example: Outputting logs to the console 21
Example: Outputting logs to a UNIX log host 22
Example: Outputting logs to a Linux log host 23

Configuring the information center

About the information center

The information center on the device receives logs generated by source modules and outputs logs to different destinations according to log output rules. Based on the logs, you can monitor device performance and troubleshoot network problems.

Figure 1 Information center diagram



Log types

Logs are classified into the following types:

- **Standard system logs**—Record common system information. Unless otherwise specified, the term "logs" in this document refers to standard system logs.
- **Diagnostic logs**—Record debugging messages.
- **Security logs**—Record security information, such as authentication and authorization information.
- **Hidden logs**—Record log information not displayed on the terminal, such as input commands.
- **Trace logs**—Record system tracing and debugging messages, which can be viewed only after the devkit package is installed.

Log levels

Logs are classified into eight severity levels from 0 through 7 in descending order. The information center outputs logs with a severity level that is higher than or equal to the specified level. For example, if you specify a severity level of 6 (informational), logs that have a severity level from 0 to 6 are output.

Table 1 Log levels

Severity value	Level	Description
0	Emergency	The system is unusable. For example, the system authorization has expired.
1	Alert	Action must be taken immediately. For example, traffic on an interface exceeds the upper limit.
2	Critical	Critical condition. For example, the device temperature exceeds the upper limit, the power module fails, or the fan tray fails.
3	Error	Error condition. For example, the link state changes.
4	Warning	Warning condition. For example, an interface is disconnected, or the memory resources are used up.
5	Notification	Normal but significant condition. For example, a terminal logs in to the device, or the device reboots.

Severity value	Level	Description
6	Informational	Informational message. For example, a command or a ping operation is executed.
7	Debugging	Debugging message.

Log destinations

The system outputs logs to the following destinations: console, monitor terminal, log buffer, log host, and log file. Log output destinations are independent and you can configure them after enabling the information center. One log can be sent to multiple destinations.

Default output rules for logs

A log output rule specifies the source modules and severity level of logs that can be output to a destination. Logs matching the output rule are output to the destination. Table 2 shows the default log output rules.

Table 2 Default output rules

Destination	Log source modules	Output switch	Severity
Console	All supported modules	Enabled	Debugging
Monitor terminal	All supported modules	Disabled	Debugging
Log host	All supported modules	Enabled	Informational
Log buffer	All supported modules	Enabled	Informational
Log file	All supported modules	Enabled	Informational

Default output rules for diagnostic logs

Diagnostic logs can only be output to the diagnostic log file, and cannot be filtered by source modules and severity levels. Table 3 shows the default output rule for diagnostic logs.

Table 3 Default output rule for diagnostic logs

Destination	Log source modules	Output switch	Severity
Diagnostic log file	All supported modules	Enabled	Debugging

Default output rules for security logs

Security logs can only be output to the security log file, and cannot be filtered by source modules and severity levels. Table 4 shows the default output rule for security logs.

Table 4 Default output rule for security logs

Destination	Log source modules	Output switch	Severity
Security log file	All supported modules	Disabled	Debugging

Default output rules for hidden logs

Hidden logs can be output to the log host, the log buffer, and the log file. Table 5 shows the default output rules for hidden logs.

Table 5 Default output rules for hidden logs

Destination	Log source modules	Output switch	Severity
Log host	All supported modules	Enabled	Informational
Log buffer	All supported modules	Enabled	Informational
Log file	All supported modules	Enabled	Informational

Default output rules for trace logs

Trace logs can only be output to the trace log file, and cannot be filtered by source modules and severity levels. Table 6 shows the default output rules for trace logs.

Table 6 Default output rules for trace logs

Destination	Log source modules	Output switch	Severity
Trace log file in the memory	All supported modules	Enabled	Debugging
Trace log file on the storage device	All supported modules	Enabled	Error

Log formats and field descriptions

Log formats

The format of logs varies by output destinations. The following table shows the original format of log information, which might be different from what you see. The actual format varies by the log resolution tool used.

Table 7 Log formats

Output destination	Format
Console, monitor terminal, log buffer, or log file	Prefix Timestamp Sysname Module/Level/Mnemonic: Content **Example**: `%Nov 24 14:21:43:502 2019 Sysname SHELL/5/SHELL_LOGIN: VTY logged in from 192.168.1.26`

Log host	**Non-customized format**:
	<PRI>Timestamp Sysname %%vvModule/Level/Mnemonic: Location; Content
	Example:
	`<190>Nov 24 16:22:21 2019 Sysname %%10SHELL/5/SHELL_LOGIN: -DevIP=1.1.1.1; VTY logged in from 192.168.1.26<190>Nov 24 16:22:21 2016 Sysname %%10 SHELL/5/SHELL_LOGIN: -DevIP=1.1.1.1; VTY logged in from 192.168.1.26`
	CMCC format:
	<PRI>Timestamp Sysname %vvModule/Level/Mnemonic: Location; Content
	`<189>Oct 9 14:59:04 2019 Sysname %10SHELL/5/SHELL_LOGIN: VTY logged in from 192.168.1.21`

	RFC 5424 format:
	\<PRI\>Version Timestamp Sysname Module - Mnemonic [SysLoc][AppLoc] Content
	`<190>1 2020-09-25T10:42:22.257Z Sysname SHELL – SHELL_LOGIN [SysLoc@25506 VendorID=%%10 Level=6][AppLoc@25506 Line=vty0 IPAddr=192.168.1.21 User=**] VTY logged in from 192.168.1.21`
	SGCC format:
	\<PRI\> Timestamp Sysname Devtype Content
	`<189> 2019-03-18 15:39:45 Sysname FW 0 2 admin 10.1.1.1`
	Unicom format:
	\<PRI\>Timestamp Hostip vvModule/Level/Serial_number: Content
	Example:
	`<189>Oct 13 16:48:08 2019 10.1.1.1 10SHELL/5/210231a64jx073000020: VTY logged in from 192.168.1.21`

Log field description

Table 8 Log field description

Field	Description
Prefix (information type)	A log to a destination other than the log host has an identifier in front of the timestamp: • An identifier of percent sign (%) indicates a log with a level equal to or higher than informational. • An identifier of asterisk (*) indicates a debugging log or a trace log. • An identifier of caret (^) indicates a diagnostic log.
PRI (priority)	A log destined for the log host has a priority identifier in front of the timestamp. The priority is calculated by using this formula: facility*8+level, where: • **facility** is the facility name. Facility names local0 through local7 correspond to values 16 through 23. The facility name can be configured using the **`info-center loghost`** command. It is used to identify log sources on the log host, and to query and filter the logs from specific log sources. • **level** is in the range of 0 to 7. See Table 1 for more information about severity levels.
Timestamp	Records the time when the log was generated. Logs sent to the log host and those sent to the other destinations have different timestamp precisions, and their timestamp formats are configured with different commands. For more information, see Table 9 and Table 10.
Hostip	Source IP address of the log. If the **`info-center loghost source`** command is configured, this field displays the IP address of the specified source interface. Otherwise, this field displays the sysname. This field exists only in logs that are sent to the log host in unicom format.
Serial number	Serial number of the device that generated the log. This field exists only in logs that are sent to the log host in unicom format.
Sysname (host name or host IP address)	The sysname is the host name or IP address of the device that generated the log. You can use the **`sysname`** command to modify the name of the device.
%% (vendor ID)	Indicates that the information was generated by an H3C device. This field exists only in logs sent to the log host.
vv (version information)	Identifies the version of the log, and has a value of 10. This field exists only in logs that are sent to the log host.

Field	Description
Module	Specifies the name of the module that generated the log. You can enter the **`info-center source ?`** command in system view to view the module list.
Level	Identifies the level of the log. See Table 1 for more information about severity levels.
Mnemonic	Describes the content of the log. It contains a string of up to 32 characters.
Location	Optional field that identifies the log sender. This field exists only in logs that are sent to the log host in non-customized or CMCC format. The field contains the following information: • **DevIp**—IP address of the log sender. • **Slot**—Member ID of the IRF member device that sent the log.
SysLoc	Optional field that identifies the system information of the logs. This field exists only in logs that are sent to the log host in RFC 5424 format. The field contains the following information: • **SysLoc@**—Vendor number of the device that generates logs. • **VendorID**—Vendor ID. Currently, the fixed value is 10. • **Level**—Log level. • **Slot**—Member ID of the IRF member device that sent the log.
AppLoc	Optional field that identifies the module information of the logs. This field exists only in logs that are sent to the log host in RFC 5424 format. The field contains the following information: • **AppLoc@**—Vendor number of the device that generates logs. • **Line**—User line used by the current user for login. • **IPAddr**—IP address used for user login. • **User**—Username for login.
Devtype	Device type. This field exists only in logs that are sent to the log host in SGCC format.
Content	Provides the content of the log.

Table 9 Timestamp precisions and configuration commands

Item	Destined for the log host	Destined for the console, monitor terminal, log buffer, and log file
Precision	Seconds (default) or milliseconds	Milliseconds
Command used to set the timestamp format	**`info-center timestamp loghost`**	**`info-center timestamp`**

Table 10 Description of the timestamp parameters

Timestamp parameters	Description
boot	Time that has elapsed since system startup, in the format of xxx.yyy. xxx represents the higher 32 bits, and yyy represents the lower 32 bits, of milliseconds elapsed. Logs that are sent to all destinations other than a log host support this parameter. **Example**: `%0.109391473 Sysname FTPD/5/FTPD_LOGIN: User ftp (192.168.1.23) has logged in successfully.` 0.109391473 is a timestamp in the **boot** format.
date	Current date and time. • For logs output to a log host, the timestamp can be in the format of MMM DD hh:mm:ss YYYY (accurate to seconds) or MMM DD hh:mm:ss.ms YYYY (accurate to milliseconds). • For logs output to other destinations, the timestamp is in the format of MMM DD hh:mm:ss:ms YYYY. All logs support this parameter. **Example**: `%May 30 05:36:29:579 2019 Sysname FTPD/5/FTPD_LOGIN: User ftp (192.168.1.23) has logged in successfully.` May 30 05:36:29:579 2019 is a timestamp in the **date** format in logs sent to the console.
iso	Timestamp format stipulated in ISO 8601, accurate to seconds (default) or milliseconds. Only logs that are sent to a log host support this parameter. **Example**: `<189>2019-05-30T06:42:44 Sysname %%10FTPD/5/FTPD_LOGIN: User ftp (192.168.1.23) has logged in successfully.` 2019-05-30T06:42:44 is a timestamp in the **iso** format accurate to seconds. A timestamp accurate to milliseconds is like 2019-05-30T06:42:44.708.
none	No timestamp is included. All logs support this parameter. **Example**: `% Sysname FTPD/5/FTPD_LOGIN: User ftp (192.168.1.23) has logged in successfully.` No timestamp is included.
no-year-date	Current date and time without year or millisecond information, in the format of MMM DD hh:mm:ss. Only logs that are sent to a log host support this parameter. **Example**: `<189>May 30 06:44:22 Sysname %%10FTPD/5/FTPD_LOGIN: User ftp (192.168.1.23) has logged in successfully.` May 30 06:44:22 is a timestamp in the **no-year-date** format.

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

Information center tasks at a glance

Managing standard system logs

1. Enabling the information center
2. Outputting logs to various destinations

   Choose the following tasks as needed:
   - Outputting logs to the console
   - Outputting logs to the monitor terminal
   - Outputting logs to log hosts
   - Outputting logs to the log buffer
   - Saving logs to the log file
3. (Optional.) Setting the minimum storage period
4. (Optional.) Enabling synchronous information output
5. (Optional.) Configuring log suppression

   Choose the following tasks as needed:
   - Enabling duplicate log suppression
   - Configuring log suppression for a module
   - Disabling an interface from generating link up or link down logs
6. (Optional.) Enabling SNMP notifications for system logs
7. (Optional.) Performing integrity check for a log file

Managing hidden logs

1. Enabling the information center
2. Outputting logs to various destinations

   Choose the following tasks as needed:
   - Outputting logs to log hosts
   - Outputting logs to the log buffer
   - Saving logs to the log file
3. (Optional.) Setting the minimum storage period
4. (Optional.) Configuring log suppression

   Choose the following tasks as needed:
   - Enabling duplicate log suppression
   - Configuring log suppression for a module

Managing security logs

1. Enabling the information center

2. (Optional.) Configuring log suppression
   Choose the following tasks as needed:
   - Enabling duplicate log suppression
   - Configuring log suppression for a module
3. Managing security logs
   - Saving security logs to the security log file
   - Managing the security log file
4. (Optional.) Performing integrity check for a log file

Managing diagnostic logs

1. Enabling the information center
2. (Optional.) Configuring log suppression
   Choose the following tasks as needed:
   - Enabling duplicate log suppression
   - Configuring log suppression for a module
3. Saving diagnostic logs to the diagnostic log file

Managing trace logs

1. Enabling the information center
2. (Optional.) Configuring log suppression
   Choose the following tasks as needed:
   - Enabling duplicate log suppression
   - Configuring log suppression for a module

Enabling the information center

About this task

The information center can output logs only after it is enabled.

Procedure

1. Enter system view.
   `system-view`
2. Enable the information center.
   `info-center enable`
   The information center is enabled by default.

Outputting logs to various destinations

Outputting logs to the console

Restrictions and guidelines

The **terminal monitor**, **terminal debugging**, and **terminal logging** commands take effect only for the current connection between the terminal and the device. If a new connection is established, the default is restored.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Configure an output rule for sending logs to the console.
   info-center source { *module-name* | **default** } **console** { **deny** | **level** *severity* }
   For information about the default output rules, see "Default output rules for logs."
3. (Optional.) Configure the timestamp format.
   info-center timestamp { **boot** | **date** | **none** }
   The default timestamp format is **date**.
4. Return to user view.
   quit
5. Enable log output to the console.
   terminal monitor
   By default, log output to the console is enabled.
6. Enable output of debugging messages to the console.
   terminal debugging
   By default, output of debugging messages to the console is disabled.
   This command enables output of debugging-level log messages to the console.
7. Set the lowest severity level of logs that can be output to the console.
   terminal logging level *severity*
   The default setting is 6 (informational).

Outputting logs to the monitor terminal

About this task

Monitor terminals refer to terminals that log in to the device through the VTY line.

Restrictions and guidelines

The **terminal monitor**, **terminal debugging**, and **terminal logging** commands take effect only for the current connection between the terminal and the device. If a new connection is established, the default is restored.

Procedure

1. Enter system view.
   system-view
2. (Optional.) Configure an output rule for sending logs to the monitor terminal.

info-center source { *module-name* | **default** } **monitor** { **deny** | **level** *severity* }

For information about the default output rules, see "Default output rules for logs."

3. (Optional.) Configure the timestamp format.

   info-center timestamp { **boot** | **date** | **none** }

   The default timestamp format is **date**.

4. Return to user view.

   quit

5. Enable log output to the monitor terminal.

   terminal monitor

   By default, log output to the monitor terminal is disabled.

6. Enable output of debugging messages to the monitor terminal.

   terminal debugging

   By default, output of debugging messages to the monitor terminal is disabled.

   This command enables output of debugging-level log messages to the monitor terminal.

7. Set the lowest level of logs that can be output to the monitor terminal.

   terminal logging level *severity*

   The default setting is 6 (informational).

Outputting logs to log hosts

About this task

You can use this feature to configure output of device-generated logs to specified log hosts.

When configuring this feature, you can specify the following information for a log host:

- **Log output filters**—Only logs that meet the specified filters are sent to the log host. This prevents logs that you are not concerned about from being sent to the log host and occupying network bandwidth and resources on the log host.
- **Log format**—The device encapsulates logs in the specified format so that the log host can parse the logs according to the corresponding rules.
- **Using TCP to send logs to enhance security**—If the TCP connection is disconnected, the device saves the logs locally in a file for user query.

1. Enter system view.

   system-view

2. (Optional.) Configure a log output filter or a log output rule. Choose one option as needed:
   - Configure a log output filter.

     info-center filter *filter-name* { *module-name* [**mnemonic** *mnemonic-value*] | **default** } { **deny** | **level** *severity* }

     You can create multiple log output filters. When specifying a log host, you can apply a log output filter to the log host to control log output.
   - Configure a log output rule for the log host output destination.

     info-center source { *module-name* | **default** } **loghost** { **deny** | **level** *severity* }

     For information about the default log output rules for the log host output destination, see "Default output rules for logs."

   The system chooses the settings to control log output to a log host in the following order:

   a. Log output filter applied to the log host.

b. Log output rules configured for the log host output destination by using the **info-center source** command.

c. Default log output rules (see "Default output rules for logs").

3. (Optional.) Specify a source IP address for logs sent to log hosts.

   info-center loghost source *interface-type interface-number*

   By default, the source IP address of logs sent to log hosts is the primary IP address of their outgoing interfaces.

4. (Optional.) Specify the format in which logs are output to log hosts.

   info-center format { **cmcc** | **rfc5424** | **sgcc** | **unicom** }

   By default, logs are output to log hosts in non-customized format.

5. (Optional.) Add the device serial number to the location field of logs sent to log hosts.

   info-center loghost locate-info with-sn

   By default, the device does not add the device serial number to the location field of logs sent to log hosts.

6. (Optional.) Configure the timestamp format.

   info-center timestamp loghost { **date** [**with-milliseconds**] | **iso** [**with-milliseconds** | **with-timezone**] * | **no-year-date** | **none** }

   The default timestamp format is **date**.

7. Specify a log host and configure related parameters.

   info-center loghost [**vpn-instance** *vpn-instance-name*] { *hostname* | *ipv4-address* | **ipv6** *ipv6-address* } [**port** *port-number*] [**dscp** *dscp-value*] [**facility** *local-number*] [**filter** *filter-name*] [**format rfc5424-json-v1 tcp** [**end-with-length**] **ssl-client-policy** *policy-name* [**directory** *dir-name* **size-quota** *size*]]

   By default, no log hosts or related parameters are specified.

   The value for the *port-number* argument must be the same as the value configured on the log host. Otherwise, the log host cannot receive logs.

Outputting logs to the log buffer

About this task

This feature enables log output to the log buffer. You can use the **display logbuffer** command to display the log buffer information and the buffered logs.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Configure an output rule for sending logs to the log buffer.

   info-center source { *module-name* | **default** } **logbuffer** { **deny** | **level** *severity* }

   For information about the default output rules, see "Default output rules for logs."

3. (Optional.) Configure the timestamp format.

   info-center timestamp { **boot** | **date** | **none** }

   The default timestamp format is **date**.

4. Enable log output to the log buffer.

   info-center logbuffer

   By default, log output to the log buffer is enabled.

5. (Optional.) Set the maximum log buffer size.

   info-center logbuffer size *buffersize*

   By default, a maximum of 512 logs can be buffered.

Saving logs to the log file

About this task

The log file feature enables log output to the log file. The system saves logs from the log file buffer to the log file. After saving logs to the log file, the system clears the log file buffer.

Log saving from the log file buffer to the log file is triggered by the following situations:

- Periodic saving.
- A manual immediate saving of buffered logs to the log file.
- The log file buffer is full.

A log file is automatically created when needed and has a maximum capacity. The log file capacity restrictions are as follows:

- When no log file space is available, the device stops saving new logs.
- When no log file space or storage device space is available, the system creates new log files as needed. The log files are named as **logfile1.log**, **logfile2.log**, and so on.

  When **logfile1.log** is full, the system compresses **logfile1.log** as **logfile1.log.gz** and creates a new log file named **logfile2.log**. The process repeats until the last log file is full.

  After the last log file is full, the device repeats the following process:

  a. The device locates the oldest compressed log file **logfileX.log.gz** and creates a new file using the same name (**logfileX.log**).

  b. When **logfileX.log** is full, the device compresses the log file as **logfileX.log.gz** to replace the existing file **logfileX.log.gz**.

TIP:

- Clean up the storage space of the device regularly to ensure sufficient storage space for the log file feature.
- As a best practice, back up the log files regularly to avoid loss of important logs.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Configure an output rule for sending logs to the log file.

   info-center source { *module-name* | **default** } **logfile** { **deny** | **level** *severity* }

   For information about the default output rules, see "Default output rules for logs."

3. (Optional.) Save log messages in XML format to log files.

   info-center logfile xml-format

   By default, log messages are saved in TXT format to log files.

   This command saves the log messages that are generated on the device in XML format to log files for structured storage and parsing by the log host.

4. Enable the log file feature.

   info-center logfile enable

   By default, the log file feature is enabled.

5. (Optional.) Enable log file overwrite-protection.

   info-center logfile overwrite-protection [**all-port-powerdown**]

   By default, log file overwrite-protection is disabled.

   Log file overwrite-protection is supported only in FIPS mode.

6. (Optional.) Set the maximum log file size.

   info-center logfile size-quota *size*

   By default, the maximum log file size is 5 MB.

7. (Optional.) Set the alarm threshold for log file usage.

   info-center logfile alarm-threshold *usage*

   The default alarm threshold for log file usage ratio is 80%. When the usage ratio of the log file reaches 80%, the system outputs a message to inform the user.

   Setting the alarm threshold value to 0 disables the log file usage alarm feature.

8. (Optional.) Specify the log file directory.

   info-center logfile directory *dir-name*

   The default log file directory is **flash:/logfile**.

   This command cannot survive an IRF reboot or a master/subordinate switchover.

9. Save logs in the log file buffer to the log file. Choose one or both options as needed:
   - Configure the automatic log file saving interval.

     info-center logfile frequency *freq-sec*

     The default log file saving interval is 86400 seconds.
   - Immediately save logs from the log file buffer to the log file.

     logfile save

     This command is available in any view.

Setting the minimum storage period

About setting the minimum storage period

Use this feature to set the minimum storage period for logs in the log files and log buffer. This feature ensures that logs will not be overwritten by new logs during a set period of time.

For logs

By default, when the log buffer is full, new logs will automatically overwrite the oldest logs. After the minimum storage period is set, the system identifies the storage period of a log to determine whether to delete the log. The system current time minus a log's generation time is the log's storage period.

- If the storage period of a log is shorter than or equal to the minimum storage period, the system does not delete the log. The new log will not be saved.
- If the storage period of a log is longer than the minimum storage period, the system deletes the log to save the new log.

For log files

By default, when the last log file is full, the device locates the oldest compressed log file **logfileX.log.gz** and creates a new file using the same name (**logfileX.log**). After the minimum storage period is set, the system identifies the storage period of the compressed log file before creating a new log file with the same name. The system current time minus the log file's last modification time is the log file's storage period.

- If the storage period of the compressed log file is shorter than or equal to the minimum storage period, the system stops saving new logs.

- If the storage period of the compressed log file is longer than the minimum storage period, the system creates a new file to save new logs.

Procedure

1. Enter system view.
   system-view
2. Set the minimum storage period for logs in the log files and log buffer.
   info-center syslog min-age *min-age*
   By default, the minimum storage period is not set.

Enabling synchronous information output

About this task

System log output interrupts ongoing configuration operations, obscuring previously entered commands. Synchronous information output shows the obscured commands. It also provides a command prompt in command editing mode, or a [Y/N] string in interaction mode so you can continue your operation from where you were stopped.

Procedure

1. Enter system view.
   system-view
2. Enable synchronous information output.
   info-center synchronous
   By default, synchronous information output is disabled.

Configuring log suppression

Enabling duplicate log suppression

About this task

The information center on the device outputs logs generated by service modules. The device identifies logs that have the same module name, level, mnemonic, location, and text as duplicate logs.

In some scenarios, for example, ARP attack or link failure, the service modules will generate a large volume of duplicate logs during a short period of time. Recording and output of consecutive duplicate logs wastes system and network resources. To resolve this issue, you can enable duplicate log suppression.

With this feature enabled, when a service module generates a log, the information center outputs the log and starts the duplicate log suppression timer. The suppression period of the duplicate log suppression timer varies by phase. The suppression periods in phase 1, 2, and later phases are 30 seconds, 2 minutes, and 10 minutes, respectively.

After you enable duplicate log suppression, the system starts suppression upon outputting a log:

- If only duplicate logs of the log are received during the suppression period of a phase, the information center does not output the duplicate logs. When the suppression period of the phase expires, the information center outputs the suppressed log and the number of times the log is suppressed, and starts the next suppression phase.

- If a different log is received during the suppression period of a phase, the information center performs the following operations:
  - Stops suppression on the log, and outputs the suppressed log and the number of times the log is suppressed.
  - Outputs the different log and starts phase-1 suppression for that log.
- If no log is received within the suppression period of any phase, the information center deletes the duplicate log suppression timer, stops suppression on the log, and does not output any log.

Procedure

1. Enter system view.

   system-view

2. Enable duplicate log suppression.

   info-center logging suppress duplicates

   By default, duplicate log suppression is disabled.

Examples

The following example uses SHELL_CMD logs to verify the duplicate log suppression feature. After the user executes a command on the device, the information center receives a SHELL_CMD log generated by the shell module, encapsulates the log, and then outputs the log to the log buffer.

1. Verify the suppression effect in phases 1, 2, 3, and 4 of a log (with suppression period of 30 seconds, 2 minutes, 10 minutes, and 10 minutes):

   # In a lab environment, continuously execute the **display logbuffer** command for 25 minutes.

   # View the output logs in the log buffer.

```
<Sysname> display logbuffer
Log buffer: Enabled
Max buffer size: 1024
Actual buffer size: 512
Dropped messages: 0
Overwritten messages: 0
Current messages: 5
%Jul 20 13:01:20:615 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer
%Jul 20 13:01:50:718 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 2 times in last 30 seconds.
%Jul 20 13:03:50:732 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 5 times in last 2 minutes.
%Jul 20 13:13:50:830 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 10 times in last 10 minutes.
%Jul 20 13:23:50:211 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 6 times in last 10 minutes.
```

   The output shows the following information:

   - The information center received the log **SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**; Command is display logbuffer**.
   - In phase 1, the log was suppressed twice by the information center.
   - In phase 2, the log was suppressed five times by the information center.
   - In phase 3, the log was suppressed 10 times by the information center.
   - In phase 4, the log was suppressed six times by the information center.

2. Continue to verify how duplicate log suppression works when a different log is received during the suppression period of a log:

Execute the **display logbuffer** command three times, and then execute the **display interface brief** command.

View the output logs in the log buffer.

```
<Sysname> display logbuffer
Log buffer: Enabled
Max buffer size: 1024
Actual buffer size: 512
Dropped messages: 0
Overwritten messages: 0
Current messages: 5
%Jul 20 13:01:20:615 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer
%Jul 20 13:01:50:718 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 2 times in last 30 seconds.
%Jul 20 13:03:50:732 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 5 times in last 2 minutes.
%Jul 20 13:13:50:830 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 10 times in last 10 minutes.
%Jul 20 13:23:50:211 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 6 times in last 10 minutes.
%Jul 20 13:24:56:205 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display logbuffer This message repeated 3 times in last 1 minute 6 seconds.
%Jul 20 13:25:41:205 2022 Sysname SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**;
Command is display interface brief.
<Sysname>
```

The output shows the following information:

- The information center stopped suppression for the log **SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**; Command is display logbuffer**.
- The information center output the log **SHELL/6/SHELL_CMD: -Line=con0-IPAddr=**-User=**; Command is display interface brief**, and started suppression for it.

Configuring log suppression for a module

About this task

This feature suppresses output of logs. You can use this feature to filter out the logs that you are not concerned with.

Perform this task to configure a log suppression rule to suppress output of all logs or logs with a specific mnemonic value for a module.

Procedure

1. Enter system view.

   system-view

2. Configure a log suppression rule for a module.

   info-center logging suppress module *module-name* **mnemonic** { **all** | *mnemonic-value* }

   By default, the device does not suppress output of any logs from any modules.

Disabling an interface from generating link up or link down logs

About this task

By default, an interface generates link up or link down log information when the interface state changes. In some cases, you might want to disable certain interfaces from generating this information. For example:

- You are concerned about the states of only some interfaces. In this case, you can use this function to disable other interfaces from generating link up and link down log information.
- An interface is unstable and continuously outputs log information. In this case, you can disable the interface from generating link up and link down log information.

Use the default setting in normal cases to avoid affecting interface status monitoring.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.

   interface *interface-type interface-number*
3. Disable the interface from generating link up or link down logs.

   undo enable log updown

   By default, an interface generates link up and link down logs when the interface state changes.

Enabling SNMP notifications for system logs

About this task

This feature enables the device to send an SNMP notification for each log message it outputs. The device encapsulates the logs in SNMP notifications and then sends them to the SNMP module and the log trap buffer.

You can configure the SNMP module to send received SNMP notifications in SNMP traps or informs to remote hosts. For more information, see "Configuring SNMP."

To view the traps in the log trap buffer, access the MIB corresponding to the log trap buffer.

Procedure

1. Enter system view.

   system-view
2. Enable SNMP notifications for system logs.

   snmp-agent trap enable syslog

   By default, the device does not send SNMP notifications for system logs.
3. Set the maximum number of traps that can be stored in the log trap buffer.

   info-center syslog trap buffersize *buffersize*

   By default, the log trap buffer can store a maximum of 1024 traps.

Managing security logs

Saving security logs to the security log file

About this task

Security logs are very important for locating and troubleshooting network problems. Generally, security logs are output together with other logs. It is difficult to identify security logs among all logs.

To solve this problem, you can save security logs to the security log file without affecting the current log output rules.

After you enable the security log file feature, the system processes security logs as follows:

1. Outputs security logs to the security log file buffer.
2. Saves logs from the security log file buffer to the security log file at the specified interval.

   If you have the security-audit role, you can also manually save security logs to the security log file.
3. Clears the security log file buffer immediately after the security logs are saved to the security log file.

Restrictions and guidelines

The device supports only one security log file. The system will overwrite old logs with new logs when the security log file is full. To avoid security log loss, you can set an alarm threshold for the security log file usage ratio. When the alarm threshold is reached, the system outputs a message to inform you of the alarm. You can log in to the device with the security-audit user role and back up the security log file to prevent the loss of important data.

Procedure

1. Enter system view.

   system-view
2. Enable the security log file feature.

   info-center security-logfile enable

   By default, the security log file feature is disabled.
3. Set the interval at which the system saves security logs.

   info-center security-logfile frequency *freq-sec*

   The default security log file saving interval is 86400 seconds.
4. (Optional.) Set the maximum size for the security log file.

   info-center security-logfile size-quota *size*

   The default maximum security log file size is 10 MB.
5. (Optional.) Set the alarm threshold of the security log file usage.

   info-center security-logfile alarm-threshold *usage*

   By default, the alarm threshold of the security log file usage ratio is 80. When the usage of the security log file reaches 80%, the system will send a message.

Managing the security log file

Restrictions and guidelines

To use the security log file management commands, you must have the security-audit user role. For information about configuring the security-audit user role, see AAA in *Security Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Change the directory of the security log file.
   info-center security-logfile directory *dir-name*
   By default, the security log file is saved in the **seclog** directory in the root directory of the storage device.
   The device uses the default security log file directory after an IRF reboot or a master/subordinate switchover.
3. Manually save all logs in the security log file buffer to the security log file.
   security-logfile save
   This command is available in any view.
4. (Optional.) Display the summary of the security log file.
   display security-logfile summary
   This command is available in any view.

Performing integrity check for a log file

About this task

When the information center saves a log message to the log file, it uses the device master key to calculate a digest for the log file and saves the digest in the log file. The digest is used to perform integrity check for the log file. You can use the default device master key or a master key configured by using the **set master-key** command. For more information about the master key, see public key management in *Security Configuration Guide*.

When this feature is configured, the information center uses the master key to recalculate a digest for the log file and then compares the digest with that saved in the log file. If the two digests are consistent, the log file passes the integrity check. If they do not match, the log file fails the integrity check.

Restrictions and guidelines

Currently, only standard system log files and security log files support integrity check, while diagnostic log files and trace log files do not support integrity check.

Procedure

To perform integrity check for a log file, execute the following command in user view:

check logfile *file-path*

For successful execution of this command, the file extension must be **.log**, **.log.gz**, **.xml**, or **.xml.gz**.

Saving diagnostic logs to the diagnostic log file

About this task

By default, the diagnostic log file feature saves diagnostic logs from the diagnostic log file buffer to the diagnostic log file at the specified saving interval. You can also manually trigger an immediate saving of diagnostic logs to the diagnostic log file. After saving diagnostic logs to the diagnostic log file, the system clears the diagnostic log file buffer.

The device supports multiple diagnostic log files. Each diagnostic log file has a maximum capacity. The diagnostic log files are named as **diagfile1.log**, **diagfile2.log**, and so on. When **diagfile1.log** is

full, the system compresses **diagfile1.log** as **diagfile1.log.gz** and creates a new diagnostic log file named **diagfile2.log**. The process repeats until the last diagnostic log file is full.

After the last diagnostic log file is full, the device repeats the following process:

1. The device locates the oldest compressed diagnostic log file **diagfileX.log.gz** and creates a new file using the same name (**diagfileX.log**).
2. When **diagfileX.log** is full, the device compresses the log file as **diagfileX.log.gz** to replace the existing file **diagfileX.log.gz**.

To avoid loss of important logs, back up the diagnostic log files regularly.

Procedure

1. Enter system view.
   system-view
2. Enable the diagnostic log file feature.
   info-center diagnostic-logfile enable
   By default, the diagnostic log file feature is enabled.
3. (Optional.) Set the maximum diagnostic log file size.
   info-center diagnostic-logfile quota *size*
   The default maximum diagnostic log file size is 5 MB.
4. (Optional.) Specify the diagnostic log file directory.
   info-center diagnostic-logfile directory *dir-name*
   The default diagnostic log file directory is **flash:/diagfile**.
   The device uses the default diagnostic log file directory after an IRF reboot or a master/subordinate switchover.
5. Save diagnostic logs in the diagnostic log file buffer to the diagnostic log file. Choose one option as needed:
   - Configure the automatic diagnostic log file saving interval.
     info-center diagnostic-logfile frequency *freq-sec*
     The default diagnostic log file saving interval is 86400 seconds.
   - Manually save diagnostic logs to the diagnostic log file.
     diagnostic-logfile save
     This command is available in any view.

Setting the maximum size of the trace log file

About this task

The device has only one trace log file. When the trace log file is full, the device overwrites the oldest trace logs with new ones.

Procedure

1. Enter system view.
   system-view
2. Set the maximum size for the trace log file.
   info-center trace-logfile quota *size*
   The default maximum size of the trace log file is 10 MB.

Display and maintenance commands for information center

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display the diagnostic log file configuration.	**display diagnostic-logfile summary**
Display the information center configuration.	**display info-center**
Display information about log output filters.	**display info-center filter** [*filter-name*]
Display log buffer information and buffered logs.	**display logbuffer** [**reverse**] [**level** *severity* | **size** *buffersize* | **slot** *slot-number*] * [**last-mins** *mins*]
Display the log buffer summary.	**display logbuffer summary** [**level** *severity* | **slot** *slot-number*] *
Display the content of the log file buffer.	**display logfile buffer**
Display the log file configuration.	**display logfile summary**
Display the content of the security log file buffer. (To execute this command, you must have the security-audit user role.)	**display security-logfile buffer**
Display summary information of the security log file.	**display security-logfile summary**
Clear the log buffer.	**reset logbuffer**

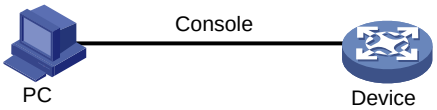
Information center configuration examples

Example: Outputting logs to the console

Network configuration

Configure the device to output to the console FTP logs that have a minimum severity level of **warning**.

Figure 2 Network diagram



Procedure

Enable the information center.

```
<Device> system-view
[Device] info-center enable
```

Disable log output to the console.

```
[Device] info-center source default console deny
```

To avoid output of unnecessary information, disable all modules from outputting log information to the specified destination (**console** in this example) before you configure the output rule.

Configure an output rule to output to the console FTP logs that have a minimum severity level of **warning**.

```
[Device] info-center source ftp console level warning
[Device] quit
```

Enable log output to the console.

```
<Device> terminal logging level 6
<Device> terminal monitor
The current terminal is enabled to display logs.
```

Now, if the FTP module generates logs, the information center automatically sends the logs to the console, and the console displays the logs.

Example: Outputting logs to a UNIX log host

Network configuration

Configure the device to output to the UNIX log host FTP logs that have a minimum severity level of **informational**.

Figure 3 Network diagram



Procedure

1. Make sure the device and the log host can reach each other. (Details not shown.)
2. Configure the device:

   # Enable the information center.

   ```
   <Device> system-view
   [Device] info-center enable
   ```

   # Specify log host 1.2.0.1/16 with **local4** as the logging facility.

   ```
   [Device] info-center loghost 1.2.0.1 facility local4
   ```

   # Disable log output to the log host.

   ```
   [Device] info-center source default loghost deny
   ```

   To avoid output of unnecessary information, disable all modules from outputting logs to the specified destination (**loghost** in this example) before you configure an output rule.

   # Configure an output rule to output to the log host FTP logs that have a minimum severity level of **informational**.

   ```
   [Device] info-center source ftp loghost level informational
   ```

3. Configure the log host:

   The log host configuration procedure varies by the vendor of the UNIX operating system. The following shows an example:

   a. Log in to the log host as a root user.

   b. Create a subdirectory named **Device** in directory **/var/log/**, and then create file **info.log** in the **Device** directory to save logs from the device.

```
# mkdir /var/log/Device
# touch /var/log/Device/info.log
```

c. Edit file **syslog.conf** in directory **/etc/** and add the following contents:

```
# Device configuration messages
local4.info /var/log/Device/info.log
```

In this configuration, **local4** is the name of the logging facility that the log host uses to receive logs. The value **info** indicates the **informational** severity level. The UNIX system records the log information that has a minimum severity level of **informational** to file **/var/log/Device/info.log**.

NOTE:

Follow these guidelines while editing file **/etc/syslog.conf**:

- Comments must be on a separate line and must begin with a pound sign (#).
- No redundant spaces are allowed after the file name.
- The logging facility name and the severity level specified in the **/etc/syslog.conf** file must be identical to those configured on the device by using the **`info-center loghost`** and **`info-center source`** commands. Otherwise, the log information might not be output to the log host correctly.

d. Display the process ID of **syslogd**, kill the **syslogd** process, and then restart **syslogd** by using the **–r** option to validate the configuration.

```
# ps -ae | grep syslogd
147
# kill -HUP 147
# syslogd -r &
```

Now, the device can output FTP logs to the log host, which stores the logs to the specified file.

Example: Outputting logs to a Linux log host

Network configuration

Configure the device to output to the Linux log host 1.2.0.1/16 FTP logs that have a minimum severity level of **informational**.

Figure 4 Network diagram



Procedure

1. Make sure the device and the log host can reach each other. (Details not shown.)
2. Configure the device:

   # Enable the information center.

   ```
   <Device> system-view
   [Device] info-center enable
   ```

   # Specify log host 1.2.0.1/16 with **local5** as the logging facility.

   ```
   [Device] info-center loghost 1.2.0.1 facility local5
   ```

   # Disable log output to the log host.

```
[Device] info-center source default loghost deny
```

To avoid outputting unnecessary information, disable all modules from outputting log information to the specified destination (**loghost** in this example) before you configure an output rule.

Configure an output rule to enable output to the log host FTP logs that have a minimum severity level of **informational**.

```
[Device] info-center source ftp loghost level informational
```

3. Configure the log host:

The log host configuration procedure varies by the vendor of the Linux operating system. The following shows an example:

a. Log in to the log host as a root user.

b. Create a subdirectory named **Device** in directory **/var/log/**, and create file **info.log** in the **Device** directory to save logs from the device.

```
# mkdir /var/log/Device
# touch /var/log/Device/info.log
```

c. Edit file **syslog.conf** in directory **/etc/** and add the following contents:

```
# Device configuration messages
local5.info /var/log/Device/info.log
```

In this configuration, **local5** is the name of the logging facility that the log host uses to receive logs. The value **info** indicates the **informational** severity level. The Linux system will store the log information with a severity level equal to or higher than **informational** to file **/var/log/Device/info.log**.

NOTE:

Follow these guidelines while editing file **/etc/syslog.conf**:

- Comments must be on a separate line and must begin with a pound sign (#).
- No redundant spaces are allowed after the file name.
- The logging facility name and the severity level specified in the **/etc/syslog.conf** file must be identical to those configured on the device by using the **`info-center loghost`** and **`info-center source`** commands. Otherwise, the log information might not be output to the log host correctly.

d. Display the process ID of **syslogd**, kill the **syslogd** process, and then restart **syslogd** by using the **-r** option to validate the configuration.

Make sure the **syslogd** process is started with the **-r** option on the Linux log host.

```
# ps -ae | grep syslogd
147
# kill -9 147
# syslogd -r &
```

Now, the device can output FTP logs to the log host, which stores the logs to the specified file.

Contents

Configuring the packet capture 1

About packet capture 1
Packet capture modes 1
Filter rule elements 1
Building a capture filter rule 2
Capture filter rule keywords 2
Capture filter rule operators 3
Capture filter rule expressions 4
Building a display filter rule 5
Display filter rule keywords 6
Display filter rule operators 7
Display filter rule expressions 8
Restrictions and guidelines: Packet capture 9
Configuring local packet capture 9
Configuring remote packet capture 9
Configuring feature image-based packet capture 9
Restrictions and guidelines 9
Prerequisites 10
Saving captured packets to a file 10
Displaying specific captured packets 10
Stopping packet capture 10
Displaying the contents in a packet file 10
Display and maintenance commands for packet capture 11
Packet capture configuration examples 11
Example: Configuring remote packet capture 11
Example: Configuring feature image-based packet capture 13

Configuring the packet capture

About packet capture

The packet capture feature captures packets. It can display the captured packets in real time, or save the captured packets to a .pcap file for future analysis.

Packet capture modes

The device supports the following packet capture modes: local packet capture, remote packet capture, and feature image-based packet capture.

Local packet capture

Local packet capture saves captured packets to a remote file on an FTP server or to a local file.

Remote packet capture

Remote packet capture sends captured packets to a remote packet analyzer (for example, Wireshark).

IMPORTANT:

The remote packet analyzer must support the Remote Capture Protocol (RPCAP).

Feature image-based packet capture

Feature image-based packet capture saves the captured packets to a local file or displays the captured packets on the terminal. This mode can also display contents of .pcap and .pcapng files.

Feature image-based packet capture requires you to install a specific image called the packet feature image.

Only feature image-based packet capture requires the packet feature image.

Filter rule elements

Packet capture supports using a capture filter rule to filter packets to be captured or using a display filter rule to filter packets to be displayed.

A filter rule is represented by a filter expression. A filter expression contains a keyword string or multiple keyword strings that are connected by operators.

Keywords include the following types:

- **Qualifiers**—Fixed keyword strings. To use a qualifier, you must enter the qualifier literally as shown.
- **Variables**—Values assigned in the required format.

Operators include the following types:

- **Logical operators**—Perform logical operations, such as the AND operation.
- **Arithmetic operators**—Perform arithmetic operations, such as the ADD operation.
- **Relational operators**—Indicate the relation between keyword strings. For example, the **=** operator indicates equality.

For more information about capture and display filters, go to the following websites:

- https://wiki.wireshark.org/CaptureFilters

- https://wiki.wireshark.org/DisplayFilters

Building a capture filter rule

Capture filter rule keywords

Qualifiers

Table 1 Qualifiers for capture filter rules

Category	Description	Examples
Protocol	Matches a protocol. If you do not specify a protocol qualifier, the filter matches any supported protocols.	• **arp**—Matches ARP. • **icmp**—Matches ICMP. • **ip**—Matches IPv4. • **ip6**—Matches IPv6. • **tcp**—Matches TCP. • **udp**—Matches UDP.
Direction	Matches packets based on its source or destination location (an IP address or port number). If you do not specify a direction qualifier, the **src or dst** qualifier applies. For example, **port 23** is equivalent to **src or dst port 23**.	• **src**—Matches the source IP address field. • **dst**—Matches the destination IP address field. • **src or dst**—Matches the source or destination IP address field.
Type	Specifies the direction type. The **host** qualifier applies if you do not specify any type qualifier. For example, **src 2.2.2.2** is equivalent to **src host 2.2.2.2**.	• **host**—Matches the IP address of a host. • **net**—Matches an IP subnet. • **port**—Matches a service port number. • **portrange**—Matches a service port range.
Others	Any other qualifiers than the previously described qualifiers.	• **broadcast**—Matches broadcast packets. • **multicast**—Matches multicast and broadcast packets. • **less**—Matches packets that are less than or equal to a specific size. • **greater**—Matches packets that are greater than or equal to a specific size. • **len**—Matches the packet length. • **vlan**—Matches VLAN packets.

Variables

A capture filter variable must be modified by one or more qualifiers.

The **broadcast**, **multicast**, and all protocol qualifiers cannot modify variables. The other qualifiers must be followed by variables.

Table 2 Variable types for capture filter rules

Variable type	Description	Examples
Integer	Represented in binary, octal, decimal, or hexadecimal notation.	The **port 23** expression matches traffic sent to or from port number 23.

Variable type	Description	Examples
Integer range	Represented by hyphenated integers.	The **portrange 100-200** expression matches traffic sent to or from any ports in the range of 100 to 200.
IPv4 address	Represented in dotted decimal notation.	The **src 1.1.1.1** expression matches traffic sent from the IPv4 host at 1.1.1.1.
IPv6 address	Represented in colon hexadecimal notation.	The **dst host 1::1** expression matches traffic sent to the IPv6 host at 1::1.
IPv4 subnet	Represented by an IPv4 network ID or an IPv4 address with a mask.	Both of the following expressions match traffic sent to or from the IPv4 subnet 1.1.1.0/24: • **src 1.1.1**. • **src net 1.1.1.0/24**.
IPv6 network segment	Represented by an IPv6 address with a prefix length.	The **dst net 1::/64** expression matches traffic sent to the IPv6 network 1::/64.

Capture filter rule operators

Logical operators

Logical operators are left associative. They group from left to right. The **not** operator has the highest priority. The **and** and **or** operators have the same priority.

Table 3 Logical operators for capture filter rules

Nonalphanumeric symbol	Alphanumeric symbol	Description
!	not	Reverses the result of a condition. Use this operator to capture traffic that matches the opposite value of a condition. For example, to capture non-HTTP traffic, use **not port 80**.
&&	and	Joins two conditions. Use this operator to capture traffic that matches both conditions. For example, to capture non-HTTP traffic that is sent to or from 1.1.1.1, use **host 1.1.1.1 and not port 80**.
||	or	Joins two conditions. Use this operator to capture traffic that matches either of the conditions. For example, to capture traffic that is sent to or from 1.1.1.1 or 2.2.2.2, use **host 1.1.1.1 or host 2.2.2.2**.

Arithmetic operators

Table 4 Arithmetic operators for capture filter rules

Nonalphanumeric symbol	Description
+	Adds two values.
-	Subtracts one value from another.
*	Multiplies one value by another.

Nonalphanumeric symbol	Description
/	Divides the value on the left by the value on the right.
&	Returns the result of the bitwise AND operation on two integral values in binary form.
|	Returns the result of the bitwise OR operation on two integral values in binary form.
<<	Performs the bitwise left shift operation on the operand to the left of the operator. The right-hand operand specifies the number of bits to shift.
>>	Performs the bitwise right shift operation on the operand to the left of the operator. The right-hand operand specifies the number of bits to shift.
[ ]	Specifies a byte offset relative to a protocol layer. This offset indicates the byte where the matching begins. You must enclose the offset value in the brackets and specify a protocol qualifier. For example, **ip[6]** matches the seventh byte of payload in IPv4 packets (the byte that is six bytes away from the beginning of the IPv4 payload).

Relational operators

Table 5 Relational operators for capture filter rules

Nonalphanumeric symbol	Description
=	Equal to. For example, **ip[6]=0x1c** matches an IPv4 packet if its seventh byte of payload is equal to 0x1c.
!=	Not equal to. For example, **len!=60** matches a packet if its length is not equal to 60 bytes.
>	Greater than. For example, **len>100** matches a packet if its length is greater than 100 bytes.
<	Less than. For example, **len<100** matches a packet if its length is less than 100 bytes.
>=	Greater than or equal to. For example, **len>=100** matches a packet if its length is greater than or equal to 100 bytes.
<=	Less than or equal to. For example, **len<=100** matches a packet if its length is less than or equal to 100 bytes.

Capture filter rule expressions

Logical expression

Use this type of expression to capture packets that match the result of logical operations.

Logical expressions contain keywords and logical operators. For example:

- **not port 23 and not port 22**—Captures packets with a port number that is not 23 or 22.
- **port 23 or icmp**—Captures packets with a port number 23 or ICMP packets.

In a logical expression, a qualifier can modify more than one variable connected by its nearest logical operator. For example, to capture packets sourced from IPv4 address 192.168.56.1 or IPv4 network 192.168.27, use either of the following expressions:

- **src 192.168.56.1 or 192.168.27**.
- **src 192.168.56.1 or src 192.168.27**.

The *expr relop expr* expression

Use this type of expression to capture packets that match the result of arithmetic operations.

This expression contains keywords, arithmetic operators (*expr*), and relational operators (*relop*). For example, **len+100>=200** captures packets that are greater than or equal to 100 bytes.

The *proto*[*expr*:*size*] expression

Use this type of expression to capture packets that match the result of arithmetic operations on a number of bytes relative to a protocol layer.

This type of expression contains the following elements:

- *proto*–Specifies a protocol layer.
- []—Performs arithmetic operations on a number of bytes relative to the protocol layer.
- *expr*—Specifies the arithmetic expression.
- *size*—Specifies the byte offset. This offset indicates the number of bytes relative to the protocol layer. The operation is performed on the specified bytes. The offset is set to 1 byte if you do not specify an offset.

For example, **ip[0]&0xf !=5** captures an IP packet if the result of ANDing the first byte with 0x0f is not 5.

To match a field, you can specify a field name for *expr*:*size*. For example, **icmp[icmptype]=0x08** captures ICMP packets that contain a value of 0x08 in the Type field.

The vlan *vlan_id* expression

Use this type of expression to capture 802.1Q tagged VLAN traffic.

This type of expression contains the **vlan** *vlan_id* keywords and logical operators. The *vlan_id* variable is an integer that specifies a VLAN ID. For example, **vlan 1 and ip** captures IPv4 packets in VLAN 1.

To capture packets of a VLAN, set a capture filter as follows:

- To capture tagged packets that are permitted on the interface, you must use the **vlan** *vlan_id* expression prior to any other expressions. For example, use the **vlan 3 and src 192.168.1.10 and dst 192.168.1.1** expression to capture packets of VLAN 3 that are sent from 192.168.1.10 to 192.168.1.1.
- To capture an untagged packet that is received on the interface, do not add **vlan** *vlan_id* to the capture filter expression if the device does not add a VLAN tag to the packet header.

Building a display filter rule

A display filter rule only identifies the packets to display. It does not affect which packets to save in a file.

Display filter rule keywords

Qualifiers

Table 6 Qualifiers for display filter rules

Category	Description	Examples
Protocol	Matches a protocol. If you do not specify a protocol qualifier, the filter matches any supported protocols.	• **eth**—Matches Ethernet. • **ftp**—Matches FTP. • **http**—Matches HTTP. • **icmp**—Matches ICMP. • **ip**—Matches IPv4. • **ipv6**—Matches IPv6. • **tcp**—Matches TCP. • **telnet**—Matches Telnet. • **udp**—Matches UDP.
Packet field	Matches a field in packets by using a dotted string in the *protocol.field*[*.level1-subfield*]…[*.leveln-subfield*] format.	• **tcp.flags.syn**—Matches the SYN bit in the flags field of TCP. • **tcp.port**—Matches the source or destination port field of TCP.

Variables

A packet field qualifier requires a variable.

Table 7 Variable types for display filter rules

Variable type	Description
Integer	Represented in binary, octal, decimal, or hexadecimal notation. For example, to display IP packets that are less than or equal to 1500 bytes, use one of the following expressions: • **ip.len le 1500**. • **ip.len le 02734**. • **ip.len le 0x436**.
Boolean	This variable type has two values: true or false. This variable type applies if you use a packet field string alone to identify the presence of a field in a packet. • If the field is present, the match result is true. The filter displays the packet. • If the field is not present, the match result is false. The filter does not display the packet. For example, to display TCP packets that contain the SYN field, use **tcp.flags.syn**.
MAC address (6 bytes)	Uses colons (:), dots (.), or hyphens (-) to break up the MAC address into two or four segments. For example, to display packets that contain a destination MAC address of ffff.ffff.ffff, use one of the following expressions: • **eth.dst==ff:ff:ff:ff:ff:ff**. • **eth.dst==ff-ff-ff-ff-ff-ff**. • **eth.dst ==ffff.ffff.ffff**.
IPv4 address	Represented in dotted decimal notation. For example: • To display IPv4 packets that are sent to or from 192.168.0.1, use **ip.addr==192.168.0.1**.

Variable type	Description
	• To display IPv4 packets that are sent to or from 129.111.0.0/16, use **ip.addr==129.111.0.0/16**.
IPv6 address	Represented in colon hexadecimal notation. For example: • To display IPv6 packets that are sent to or from 1::1, use **ipv6.addr==1::1**. • To display IPv6 packets that are sent to or from 1::/64, use **ipv6.addr==1::/64**.
String	Character string. For example, to display HTTP packets that contain the string **HTTP/1.1** for the request version field, use **http.request version=="HTTP/1.1"**.

Display filter rule operators

Logical operators are left associative. They group from left to right. The [ ] operator has the highest priority. The **not** operator has the highest priority. The **and** and **or** operators have the same priority.

Logical operators

Table 8 Logical operators for display filter rules

Nonalphanumeric symbol	Alphanumeric symbol	Description
[ ]	No alphanumeric symbol is available.	Used with protocol qualifiers. For more information, see "The proto[…] expression."
!	not	Displays packets that do not match the condition connected to this operator.
&&	and	Joins two conditions. Use this operator to display traffic that matches both conditions.
||	or	Joins two conditions. Use this operator to display traffic that matches either of the conditions.

Relational operators

Table 9 Relational operators for display filter rules

Nonalphanumeric symbol	Alphanumeric symbol	Description
==	eq	Equal to. For example, **ip.src==10.0.0.5** displays packets with the source IP address as 10.0.0.5.
!=	ne	Not equal to. For example, **ip.src!=10.0.0.5** displays packets whose source IP address is not 10.0.0.5.
>	gt	Greater than. For example, **frame.len>100** displays frames with a length greater than 100 bytes.
<	lt	Less than.

Nonalphanumeric symbol	Alphanumeric symbol	Description
		For example, **frame.len<100** displays frames with a length less than 100 bytes.
>=	ge	Greater than or equal to. For example, **frame.len ge 0x100** displays frames with a length greater than or equal to 256 bytes.
<=	le	Less than or equal to. For example, **frame.len le 0x100** displays frames with a length less than or equal to 256 bytes.

Display filter rule expressions

Logical expression

Use this type of expression to display packets that match the result of logical operations.

Logical expressions contain keywords and logical operators. For example, **ftp or icmp** displays all FTP packets and ICMP packets.

Relational expression

Use this type of expression to display packets that match the result of comparison operations.

Relational expressions contain keywords and relational operators. For example, **ip.len<=28** displays IP packets that contain a value of 28 or fewer bytes in the length field.

Packet field expression

Use this type of expression to display packets that contain a specific field.

Packet field expressions contain only packet field strings. For example, **tcp.flags.syn** displays all TCP packets that contain the SYN bit field.

The *proto*[…] expression

Use this type of expression to display packets that contain specific field values.

This type of expression contains the following elements:

- *proto*–Specifies a protocol layer or packet field.
- […]—Matches a number of bytes relative to a protocol layer or packet field. Values for the bytes to be matched must be a hexadecimal integer string. The expression in brackets can use the following formats:
  - [*n*:*m*]—Matches a total of *m* bytes after an offset of *n* bytes from the beginning of the specified protocol layer or field. To match only 1 byte, you can use both [n] and [n:1] formats. For example, **eth.src[0:3]==00:00:83** matches an Ethernet frame if the first three bytes of its source MAC address are 0x00, 0x00, and 0x83. The **eth.src[2] == 83** expression matches an Ethernet frame if the third byte of its source MAC address is 0x83.
  - [*n*-*m*]—Matches a total of (*m*-*n*+*1*) bytes, starting from the (*n*+*1)*th byte relative to the beginning of the specified protocol layer or packet field. For example, **eth.src[1-2]==00:83** matches an Ethernet frame if the second and third bytes of its source MAC address are 0x00 and 0x83, respectively.

Restrictions and guidelines: Packet capture

To capture packets forwarded through chips, first configure a traffic behavior to mirror the traffic to the CPU.

To capture packets forwarded by the CPU, enable packet capture directly.

The packet capture feature can capture only packets of a frame length equal to or less than 9196 bytes. If the frame length of a packet exceeds 9196 bytes, the system does not capture the packet.

Configuring local packet capture

To configure local packet capture, execute the following command in user view:

packet-capture local interface *interface-type interface-number* [**capture-filter** *capt-expression* | **limit-frame-size** *bytes* | **autostop filesize** *kilobytes* | **autostop duration** *seconds*] * **write** { *filepath* | **url** *url* [**username** *username* [**password** { **cipher** | **simple** } *string*]] }

The packet capture is executed in the background. After issuing this command, you can continue to configure other commands.

Configuring remote packet capture

About this task

To capture packets on an interface and send them to a remote packet analyzer (for example, Wireshark), perform this task. With remote packet capture, the analyzer must establish an RPCAP connection to the default or specified RPCAP service port on the device to receive captured packets.

To make sure a packet analyzer can establish an RPCAP connection to receive packets captured from an interface on the device, use the following procedure to configure remote packet capture:

1. Configure remote packet capture for the interface before you connect the packet analyzer to the device.
2. Configure the device as a server on the packet analyzer and connect the packet analyzer to the device. When you configure the device as a server, specify its IP address and make sure the RPCAP port number is the same as that configured on the device for the interface.

Procedure

1. Configure remote packet capture for an interface in user view:

   packet-capture remote interface *interface-type interface-number* [**port** *port*]
2. Connect the packet analyzer (for example, Wireshark) to the device. For more information about the packet analyzer, see its documentation.

Configuring feature image-based packet capture

Restrictions and guidelines

After configuring feature image-based packet capture, you cannot configure any other commands at the CLI until the capture finishes or is stopped.

There might be a delay for the capture to stop because of heavy traffic.

Prerequisites

1. Use the **display boot-loader** command to check whether the packet capture feature image is installed.
2. If the image is not installed, install the image by using the **boot-loader**, **install**, or **issu** command series.
3. Log out of the device and then log in again.

For more information about the commands, see *Fundamentals Command Reference*.

Saving captured packets to a file

To configure feature image-based packet capture and save the captured packets to a file, execute the following command in user view:

packet-capture interface *interface-type interface-number* [**capture-filter** *capt-expression* | **limit-captured-frames** *limit* | **limit-frame-size** *bytes* | **autostop filesize** *kilobytes* | **autostop duration** *seconds* | **autostop files** *numbers* | **capture-ring-buffer filesize** *kilobytes* | **capture-ring-buffer duration** *seconds* | **capture-ring-buffer files** *numbers*] * **write** *filepath* [**raw** | { **brief** | **verbose** }] *

Displaying specific captured packets

To configure feature image-based packet capture and display specific packet data, execute the following command in user view:

packet-capture interface *interface-type interface-number* [**capture-filter** *capt-expression* | **display-filter** *disp-expression* | **limit-captured-frames** *limit* | **limit-frame-size** *bytes* | **autostop duration** *seconds*] * [**raw** | { **brief** | **verbose** }] *

Stopping packet capture

About this task

Use this task to manually stop packet capture.

Procedure

Choose one option as needed:

- Stop local or remote packet capture.

  packet-capture stop

  Execute this command in user view.
- Stop feature image-based packet capture.

  Press **Ctrl+C**.

Displaying the contents in a packet file

About this task

Use this task to display the contents of a .pcap or .pcapng file on the device. Alternatively, you can transfer the file to a PC and use a packet analyzer (for example, Wireshark) to display the file content.

Prerequisites

1. Use the **display boot-loader** command to check whether the packet capture feature image is installed.
2. If the image is not installed, install the image by using **boot-loader**, **install**, or **issu** commands.
3. Log out of the device and then log in again.

For more information about the commands, see *Fundamentals Command Reference*.

Restrictions and guidelines

To stop displaying the contents, press **Ctrl+C**.

Procedure

To display the contents in a local packet file, execute the following command in user view:

packet-capture read *filepath* [**display-filter** *disp-expression*] [**raw** | { **brief** | **verbose** }] *

Display and maintenance commands for packet capture

Execute **display** commands in any view.

Task	Command
Display status information about local or remote packet capture.	**display packet-capture status**

Packet capture configuration examples

Example: Configuring remote packet capture

Network configuration

As shown in Figure 1, capture packets forwarded through the CPU or chips on Layer 2 interface GigabitEthernet 1/0/1. Use Wireshark to display the captured packets.

Figure 1 Network diagram

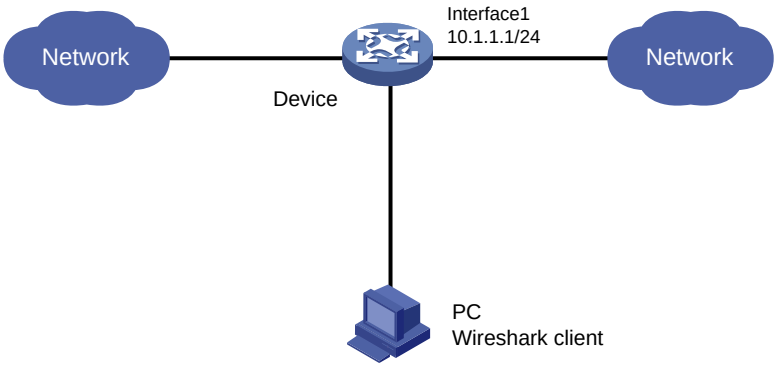


Table 10 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Configure the device:

   # Apply a QoS policy to the incoming direction of GigabitEthernet 1/0/1 to capture packets destined for the 20.1.1.0/16 network that are forwarded through chips.

   a. Create an IPv4 advanced ACL to match packets that are sent to the 20.1.1.0/16 network.

   ```
   <Device> system-view
   [Device] acl advanced 3000
   [Device-acl-ipv4-adv-3000] rule permit ip destination 20.1.1.0 255.255.0.0
   [Device-acl-ipv4-adv-3000] quit
   ```

   b. Configure a traffic behavior to mirror traffic to the CPU.

   ```
   [Device] traffic behavior behavior1
   [Device-behavior-behavior1] mirror-to cpu
   [Device-behavior-behavior1] quit
   ```

   c. Configure a traffic class to use the ACL to match traffic.

   ```
   [Device] traffic classifier classifier1
   [Device-classifier-class1] if-match acl 3000
   [Device-classifier-class1] quit
   ```

   d. Configure a QoS policy. Associate the traffic class with the traffic behavior.

   ```
   [Device] qos policy user1
   [Device-qospolicy-user1] classifier classifier1 behavior behavior1
   [Device-qospolicy-user1] quit
   ```

   e. Apply the QoS policy to the incoming direction of GigabitEthernet 1/0/1.

   ```
   [Device] interface gigabitethernet 1/0/1
   [Device-GigabitEthernet1/0/1] qos apply policy user1 inbound
   [Device-GigabitEthernet1/0/1] quit
   [Device] quit
   ```

   # Configure remote packet capture on GigabitEthernet 1/0/1. Set the RPCAP service port number to 2014.

   ```
   <Device> packet-capture remote interface gigabitethernet 1/0/1 port 2014
   ```

2. Configure Wireshark:

   a. Start Wireshark on the PC and select **Capture** > **Options**.

   b. Select **Remote** from the **Interface** list.

   c. Enter the IP address of the device **10.1.1.1** and the RPCAP service port number **2014**. Make sure there are routes available between the IP address and the PC.

   d. Click **OK** and then click **Start**.

   The captured packets are displayed on the Wireshark.

NOTE:

Options and option placement available with Wireshark might differ depending on its version.

Example: Configuring feature image-based packet capture

Network configuration

As shown in Figure 2, capture incoming IP packets of VLAN 3 on Layer 2 interface GigabitEthernet 1/0/1 that meet the following conditions:

- Sent from 192.168.1.10 or 192.168.1.11 to 192.168.1.1.
- Forwarded through the CPU or chips.

Figure 2 Network diagram

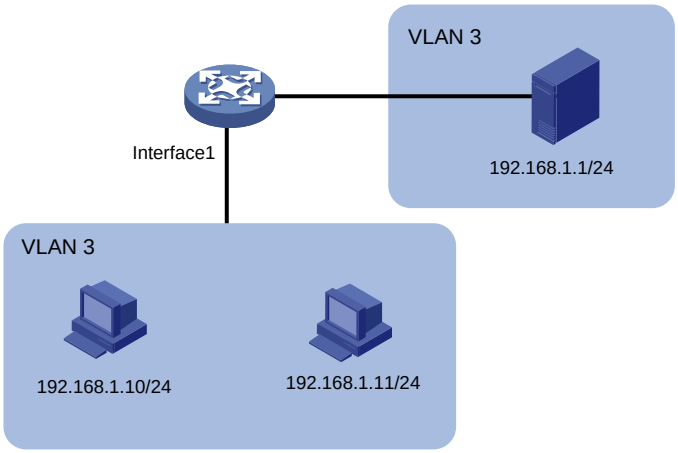


Table 11 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Install the packet capture feature.

   # Identify versions of the boot and system images running on the device. Prepare a packet capture feature image that is compatible with the current boot and system images.

```
<Device> display version
H3C Comware Software, Version 7.1.070, Demo 01
Copyright (c) 2004-2019 New H3C Technologies Co., Ltd. All rights reserved.
H3C S5590-28T8XC-HI uptime is 0 weeks, 0 days, 5 hours, 33 minutes
Last reboot reason : Cold reboot
Boot image: flash:/boot-01.bin
Boot image version: 7.1.070, Demo 01
  Compiled Oct 20 2019 16:00:00
System image: flash:/system-01.bin
System image version: 7.1.070, Demo 01
  Compiled Oct 20 2019 16:00:00
...
```

   # Download the packet capture feature image to the device. In this example, the **packet-capture-01.bin** image is stored on the TFTP server at 192.168.1.1.

```
<Device> tftp 192.168.1.1 get packet-capture-01.bin
Press CTRL+C to abort.
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
```

```
                                 Dload  Upload   Total   Spent    Left  Speed
100 11.3M    0 11.3M    0     0   155k      0 --:--:--  0:01:14 --:--:--  194k
Writing file...Done.
```

Install the packet capture feature image on all IRF member devices and commit the software change. In this example, there are two IRF member devices.

```
<Device> install activate feature flash:/packet-capture-01.bin slot 1
Verifying the file flash:/packet-capture-01.bin on slot 1....Done.
Identifying the upgrade methods....Done.
Upgrade summary according to following table:

flash:/packet-capture-01.bin
  Running Version             New Version
  None                        Demo 01

  Slot                        Upgrade Way
  1                           Service Upgrade
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...................Done.
<Device> install activate feature flash:/packet-capture-01.bin slot 2
Verifying the file flash:/packet-capture-01.bin on slot 2....Done.
Identifying the upgrade methods....Done.
Upgrade summary according to following table:

flash:/packet-capture-01.bin
  Running Version             New Version
  None                        Demo 01

  Slot                        Upgrade Way
  2                           Service Upgrade
Upgrading software images to compatible versions. Continue? [Y/N]:y
This operation might take several minutes, please wait...................Done.
<Device> install commit
This operation will take several minutes, please wait.....................Done.
```

Log out and then log in to the device again so you can execute the **packet-capture interface** and **packet-capture read** commands.

2. Apply a QoS policy to the incoming direction of GigabitEthernet 1/0/1 to capture packets from 192.168.1.10 or 192.168.1.11 to 192.168.1.1 that are forwarded through chips.

Create an IPv4 advanced ACL to match packets that are sent from 192.168.1.10 or 192.168.1.11 to 192.168.1.1.

```
<Device> system-view
[Device] acl advanced 3000
[Device-acl-ipv4-adv-3000] rule permit ip source 192.168.1.10 0 destination
192.168.1.1 0
[Device-acl-ipv4-adv-3000] rule permit ip source 192.168.1.11 0 destination
192.168.1.1 0
[Device-acl-ipv4-adv-3000] quit
```

Configure a traffic behavior to mirror traffic to the CPU.

```
[Device] traffic behavior behavior1
[Device-behavior-behavior1] mirror-to cpu
```

```
[Device-behavior-behavior1] quit
```

Configure a traffic class to use the ACL to match traffic.

```
[Device] traffic classifier classifier1
[Device-classifier-class1] if-match acl 3000
[Device-classifier-class1] quit
```

Configure a QoS policy. Associate the traffic class with the traffic behavior.

```
[Device] qos policy user1
[Device-qospolicy-user1] classifier classifier1 behavior behavior1
[Device-qospolicy-user1] quit
```

Apply the QoS policy to the incoming direction of GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] qos apply policy user1 inbound
[Device-GigabitEthernet1/0/1] quit
[Device] quit
```

3. Enable packet capture.

Capture incoming traffic on GigabitEthernet 1/0/1. Set the maximum number of captured packets to 10. Save the captured packets to the **flash:/a.pcap** file.

```
<Device> packet-capture interface gigabitethernet 1/0/1 capture-filter "src
192.168.1.10 or 192.168.1.11 and dst 192.168.1.1" limit-captured-frames 10 write
flash:/a.pcap
Capturing on 'GigabitEthernet1/0/1'
10
```

Verifying the configuration

Telnet to 192.168.1.1 from 192.168.1.10. (Details not shown.)

Display the contents in the packet file on the device.

```
<Device> packet-capture read flash:/a.pcap
  1  0.000000 192.168.1.10 -> 192.168.1.1 TCP 62 6325 > telnet [SYN] Seq=0 Win=65535 Len=0
MSS=1460 SACK_PERM=1
  2  0.000061 192.168.1.10 -> 192.168.1.1 TCP 60 6325 > telnet [ACK] Seq=1 Ack=1 Win=65535
Len=0
  3   0.024370 192.168.1.10 -> 192.168.1.1 TELNET 60 Telnet Data ...
  4   0.024449 192.168.1.10 -> 192.168.1.1 TELNET 78 Telnet Data ...
  5   0.025766 192.168.1.10 -> 192.168.1.1 TELNET 65 Telnet Data ...
  6   0.035096 192.168.1.10 -> 192.168.1.1 TELNET 60 Telnet Data ...
  7   0.047317 192.168.1.10 -> 192.168.1.1 TCP 60 6325 > telnet [ACK] Seq=42 Ack=434
Win=65102 Len=0
  8   0.050994 192.168.1.10 -> 192.168.1.1 TCP 60 6325 > telnet [ACK] Seq=42 Ack=436
Win=65100 Len=0
  9   0.052401 192.168.1.10 -> 192.168.1.1 TCP 60 6325 > telnet [ACK] Seq=42 Ack=438
Win=65098 Len=0
 10   0.057736 192.168.1.10 -> 192.168.1.1 TCP 60 6325 > telnet [ACK] Seq=42 Ack=440
Win=65096 Len=0
```

Contents

Configuring VCF fabric 1

About VCF fabric 1
VCF fabric topology 1
Automated VCF fabric deployment 2
Process of automated VCF fabric deployment 2
Template file 3
VCF fabric task at a glance 4
Configuring automated VCF fabric deployment 4
Enabling VCF fabric topology discovery 5
Configuring automated underlay network deployment 6
Restrictions and guidelines 6
Specify the template file for automated underlay network deployment 6
Specifying the role of the device in the VCF fabric 6
Configuring the device as a master spine node 7
Setting the NETCONF username and password 7
Pausing automated underlay network deployment 8
Display and maintenance commands for VCF fabric 8

Configuring VCF fabric

About VCF fabric

Based on OpenStack Networking, the Virtual Converged Framework (VCF) solution provides virtual network services from Layer 2 to Layer 7 for cloud tenants. This solution breaks the boundaries between the network, cloud management, and terminal platforms and transforms the IT infrastructure to a converged framework to accommodate all applications. It also implements automated topology discovery and automated deployment of underlay networks to reduce the administrators' workload and speed up network deployment and upgrade.

VCF fabric topology

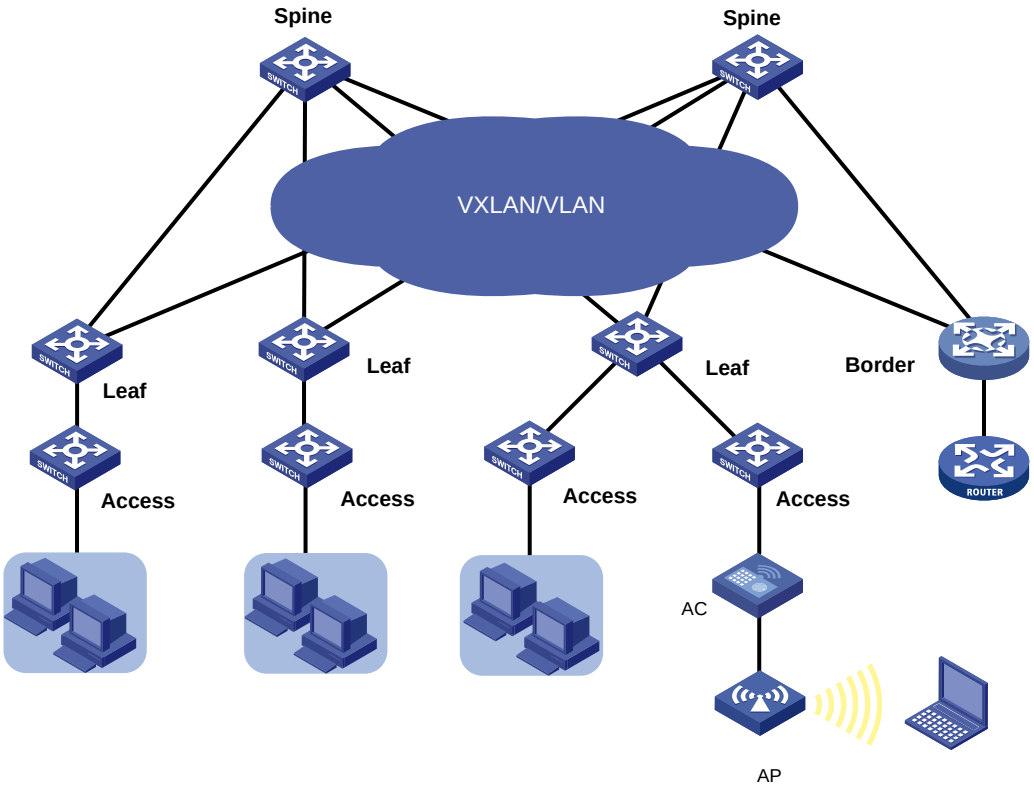
VCF fabric topology for a campus network

In a campus VCF fabric, a device has one of the following roles:

- **Spine node**—Connects to leaf nodes.
- **Leaf node**—Connects to access nodes.
- **Access node**—Connects to an upstream leaf node and downstream terminal devices. Cascading of access nodes is supported.
- **Border node**—Located at the border of a VCF fabric to provide access to the external network.

Spine nodes and leaf nodes form a large Layer 2 network, which can be a VLAN, a VXLAN with a centralized IP gateway, or a VXLAN with distributed IP gateways. For more information about centralized IP gateways and distributed IP gateways, see *VXLAN Configuration Guide.*

Figure 1 VCF fabric topology for a campus network



Automated VCF fabric deployment

VCF provides the following features to ease deployment:

- Automated topology discovery.

  In a VCF fabric, each device uses LLDP to collect local topology information from directly-connected peer devices. The local topology information includes connection interfaces, roles, MAC addresses, and management interface addresses of the peer devices. If multiple spine nodes exist in an IPv4 VCF fabric, the master spine node collects the topology for the entire network. In an IPv6 VCF fabric, the controller collects the topology automatically.

- Automated underlay network deployment.

  Automated underlay network deployment sets up a Layer 3 underlay network (a physical Layer 3 network) for users. It is implemented by automatically executing configurations (such as IRF configuration and Layer 3 reachability configurations) in user-defined template files.

Process of automated VCF fabric deployment

The device finishes automated VCF fabric deployment as follows:

1. Starts up without loading configuration and then obtains an IP address, the IP address of the TFTP server, and a template file name from the DHCP server.
2. Determines the name of the template file to be downloaded based on the device role and the template file name obtained from the DHCP server. For example, 1_leaf**.template** represents a template file for leaf nodes.
3. Downloads the template file from the TFTP server.
4. Parses the template file and performs the following operations:

- Deploys static configurations that are independent from the VCF fabric topology.
- Deploys dynamic configurations according to the VCF fabric topology.

  The topology process notifies the automation process of creation, deletion, and status change of neighbors. Based on the topology information, the automation process completes role discovery, IRF fabric setup, and automatic aggregation.

Template file

A template file contains the following contents:

- **System-predefined variables**—The variable names cannot be edited, and the variable values are set by the VCF topology discovery feature.
- **User-defined variables**—The variable names and values are defined by the user. The following are examples of user-defined variables:
  - IPv4:

    ```
    #USERDEF
    _username = aaa
    _password = hello12345
    _rbacUserRole = network-admin
    _loghost_ip = 172.16.1.136
    …
    ```
  - IPv6:

    ```
    #USERDEF
    _username = aaa
    _password = hello12345
    _rbacUserRole = network-admin
    _loghost_ip = 200:0:0:0:0:0:0:210
    _OOB = True
    _lagg_enable = false
    _lagg_mode = dynamic
    …
    ```
- **Static configurations**—Static configurations are independent from the VCF fabric topology and can be directly executed. The following are examples of static configurations:

  ```
  #STATICCFG
  #
   clock timezone beijing add 08:00:00
  #
   lldp global enable
  #
   stp global enable
  #
  ```
- **Dynamic configurations**—Dynamic configurations are dependent on the VCF fabric topology. The device first obtains the topology information through LLDP and then executes dynamic configurations. The following are examples of dynamic configurations:

  ```
  #
  interface $$_underlayIntfDown
   port link-mode route
   ip address unnumbered interface LoopBack0
   ospf 1 area 0.0.0.0
   ospf network-type p2p
  ```

```
 lldp management-address arp-learning
 lldp tlv-enable basic-tlv management-address-tlv interface LoopBack0
#
```

VCF fabric task at a glance

To configure a VCF fabric, perform the following tasks:

- Configuring automated VCF fabric deployment

  No tasks are required to be made on the device for automated VCF fabric deployment. However, you must make related configuration on the DHCP server and the TFTP server so the device can download and parse a template file to complete automated VCF fabric deployment.
- (Optional.) Adjust VCF fabric deployment

  If the device cannot obtain or parse the template file to complete automated VCF fabric deployment, choose the following tasks as needed:
  - Enabling VCF fabric topology discovery
  - Configuring automated underlay network deployment

Configuring automated VCF fabric deployment

Restrictions and guidelines

On a data center network, if the template file contains software version information, the device first compares the software version with the current software version. If the two versions are inconsistent, the device downloads the new software version to perform software upgrade. After restarting up, the device executes the configurations in the template file.

On a campus network, links between two access nodes cascaded through GigabitEthernet interfaces and links between leaf nodes and access nodes are automatically aggregated. For links between spine nodes and leaf nodes, the **`trunk permit vlan`** command is automatically executed.

On a campus network where multiple leaf nodes that have an access node attached form an IRF fabric, make sure only one link exists between each leaf node and its connected access node.

Do not perform link migration when devices in the VCF fabric are in the process of coming online or powering down after the automated VCF fabric deployment finishes. A violation might cause link-related configuration fails to update.

The version format of a template file for automated VCF fabric deployment is *x*.*y*. Only the *x* part is examined during a version compatibility check. For successful automated deployment, make sure *x* in the version of the template file to be used is not greater than *x* in the supported version. To display the supported version of the template file for automated VCF fabric deployment, use the **`display vcf-fabric underlay template-version`** command.

If the template file does not include IRF configurations, the device does not save the configurations after executing all configurations in the template file. To save the configurations, use the **`save`** command.

Two devices with the same role can automatically set up an IRF fabric only when the IRF physical interfaces on the devices are connected.

Two IRF member devices in an IRF fabric use the following rules to elect the IRF master during automated VCF fabric deployment:

- If the uptime of both devices is shorter than two hours, the device with the higher bridge MAC address becomes the IRF master.

- If the uptime of one device is equal to or longer than two hours, that device becomes the IRF master.
- If the uptime of both devices are equal to or longer than two hours, the IRF fabric cannot be set up. You must manually reboot one of the member devices. The rebooted device will become the IRF subordinate.

If the IRF member ID of a device is not 1, the IRF master might reboot during automatic IRF fabric setup.

Procedure

1. Finish the underlay network planning (such as IP address assignment, reliability design, and routing deployment) based on user requirements.
2. Configure the DHCP server.

   Configure the IP address of the device, the IP address of the TFTP server, and names of template files saved on the TFTP server. For more information, see the user manual of the DHCP server.
3. Configure the TFTP server.

   Create template files and save the template files to the TFTP server.

   The H3C DR1000 ADDC or ADCAM network management software can automatically create template files and save the files to the TFTP server. If no H3C DR1000 ADDC or ADCAM is available on the network, you must manually create template files and save the files to the TFTP server. For more information about template files, see "Template file."
4. (Optional.) Configure the NTP server.
5. Connect the device to the VCF fabric and start the device.

   After startup, the device uses a management Ethernet interface or VLAN-interface 1 to connect to the fabric management network. Then, it downloads the template file corresponding to its device role and parses the template file to complete automated VCF fabric deployment.
6. (Optional.) Save the deployed configuration.

   If the template file does not include IRF configurations, the device will not save the configurations after executing all configurations in the template file. To save the configurations, use the **save** command. For more information about this command, see configuration file management commands in *Fundamentals Command Reference*.

Enabling VCF fabric topology discovery

1. Enter system view.

   system-view
2. Enable LLDP globally.

   lldp global enable

   By default, LLDP is disabled globally.

   You must enable LLDP globally before you enable VCF fabric topology discovery, because the device needs LLDP to collect topology data of directly-connected devices.
3. Enable VCF fabric topology discovery.

   vcf-fabric topology enable

   By default, VCF fabric topology discovery is disabled.

Configuring automated underlay network deployment

Restrictions and guidelines

After automated underlay network deployment starts, you cannot use the CLI commands to modify the configuration. A violation might cause automated deployment failure. If you want to modify the configuration, first cancel or pause automated deployment. Additionally, do not start automated deployment after the modification. To cancel automated deployment, use the **undo vcf-fabric underlay autoconfigure** command. To pause automated deployment, use the **vcf-fabric underlay pause** command.

The underlay network can be an IPv4 or IPv6 network.

Specify the template file for automated underlay network deployment

1. Enter system view.
   system-view
2. Specify the template file for automated underlay network deployment.
   vcf-fabric underlay autoconfigure *template*
   By default, no template file is specified for automated underlay network deployment.

Specifying the role of the device in the VCF fabric

About this task

Perform this task to change the role of the device in the VCF fabric.

Restrictions and guidelines

If the device completes automated underlay network deployment by automatically downloading and parsing a template file, reboot the device after you change the device role. In this way, the device can obtain the template file corresponding to the new role and complete the automated underlay network deployment.

To use devices that have come online after automated deployment to form an IRF fabric, make sure all member devices in the IRF fabric have the same VCF fabric role.

Procedure

1. Enter system view.
   system-view
2. Specify the role of the device in the VCF fabric.
   vcf-fabric role { **access** | **aggr** | **leaf** | **spine** }
   The default setting varies by device model.
3. Return to system view.
   quit
4. Reboot the device.
   reboot
   For the new role to take effect, you must reboot the device.

Configuring the device as a master spine node

About this task

If multiple spine nodes exist on an IPv4 VCF fabric, you must configure a device as the master spine node to collect the topology for the entire VCF fabric network.

Restrictions and guidelines

In an IPv6 VCF fabric, the controller collects the topology automatically. You do not need to specify a master spine node.

Procedure

1. Enter system view.

   system-view

2. Configure the device as a master spine node.

   vcf-fabric spine-role master

   By default, the device is not a master spine node.

Setting the NETCONF username and password

About this task

During automated underlay network deployment, a spine node establishes NETCONF sessions with other devices in the underlay network to collect the global topology. If a user changes the username or password on a device in the undelay network, the NETCONF session between the device and the spine node is disconnected. Thun, the spine node fails to sense the topology change. To resolve the issue, perform this task to change the NETCONF username or password on the spine node.

Restrictions and guidelines

This feature is available on spine nodes. Changing the NETCONF username or password will disconnect NETCONF sessions between the spine node and leaf nodes. If multiple spine nodes exist in the VCF fabric, changing the NETCONF username or password on the master spine node will also disconnect NETCONF sessions between the master spine node and other spine nodes.

Make sure all devices in the underlay network use the same username and password for NETCONF session establishment.

Procedure

1. Enter system view.

   system-view

2. Set the NETCONF username.

   vcf-fabric underlay netconf-username *username*

   By default, the device uses the NETCONF username defined in the template file for automated VCF fabric deployment.

3. Set the NETCONF password.

   vcf-fabric underlay netconf-password { **cipher** | **simple** } *string*

   By default, the device uses the NETCONF password defined in the template file for automated VCF fabric deployment.

Pausing automated underlay network deployment

About this task

If you pause automated underlay network deployment, the VCF fabric will save the current status of the device. It will not respond to new LLDP events, set up the IRF fabric, aggregate links, or discover uplink or downlink interfaces.

Perform this task if all devices in the VCF fabric complete automated deployment and new devices are to be added to the VCF fabric.

Procedure

1. Enter system view.
   `system-view`
2. Pause automated underlay network deployment.
   `vcf-fabric underlay pause`
   By default, automated underlay network deployment is not paused.

Display and maintenance commands for VCF fabric

Execute **`display`** commands in any view.

Task	Command
Display the role of the device in the VCF fabric.	**`display vcf-fabric role`**
Display VCF fabric topology information.	**`display vcf-fabric topology`**
Display information about automated underlay network deployment.	**`display vcf-fabric underlay autoconfigure`**
Display the supported version and the current version of the template file for automated VCF fabric provisioning.	**`display vcf-fabric underlay template-version`**

Contents

Configuring NetStream 1

About NetStream 1
NetStream architecture 1
NetStream flow aging 2
NetStream data export 3
NetStream filtering 5
NetStream sampling 5
Protocols and standards 5
Restrictions and guidelines: NetStream configuration 5
NetStream tasks at a glance 6
Enabling NetStream 6
Enabling NetStream on an interface 6
Configuring NetStream filtering 6
Setting the NetStream flow classification key fields 7
Configuring NetStream sampling 8
Configuring the NetStream data export format 8
Configuring the refresh rate for NetStream version 9 or version 10 template 9
Configuring NetStream flow aging 10
Configuring periodical flow aging 10
Configuring the NetStream data export 10
Configuring the NetStream traditional data export 10
Configuring the NetStream aggregation data export 11
Enabling SNMP notifications for NetStream 11
Display and maintenance commands for NetStream 12
NetStream configuration examples 12
Example: Configuring NetStream traditional data export 12
Example: Configuring NetStream aggregation data export 15

Configuring session-based NetStream 18

About session-based NetStream 18
Session-based NetStream aggregation modes 18
Session-based NetStream data export 18
Configuring session-based NetStream settings 18
Display and maintenance commands for session-based NetStream 19
Session-based NetStream configuration examples 19
Example: Configuring session-based NetStream data export 19

Configuring NetStream

About NetStream

NetStream is an accounting technology that provides statistics on a per-flow basis. An IPv4 flow is defined by the following 7-tuple elements:

- Destination IP address.
- Source IP address.
- Destination port number.
- Source port number.
- Protocol number.
- ToS.
- Inbound or outbound interface.

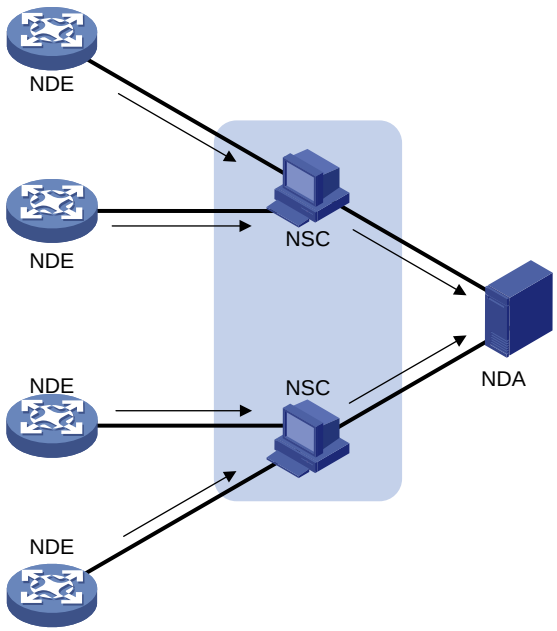
NetStream architecture

A typical NetStream system includes the following elements:

- **NetStream data exporter**—A device configured with NetStream. The NDE provides the following functions:
  - Classifies traffic flows by using the 7-tuple elements.
  - Collects data from the classified flows.
  - Aggregates the data.
  - Exports the data to the NSC.
- **NetStream collector**—A program running on an operating system. The NSC parses the packets received from the NDEs, and saves the data to its database.
- **NetStream data analyzer**—A network traffic analyzing tool. Based on the data in NSC, the NDA generates reports for traffic billing, network planning, and attack detection and monitoring. The NDA can collect data from multiple NSCs. Typically, the NDA features a Web-based system for easy operation.

NSC and NDA are typically integrated into a NetStream server.

Figure 1 NetStream system



NetStream flow aging

NetStream uses flow aging to enable the NDE to export NetStream data to NetStream servers. NetStream creates a NetStream entry for each flow for storing the flow statistics in the cache.

When a flow is aged out, the NDE performs the following operations:

- Exports the summarized data to NetStream servers in a specific format.
- Clears NetStream entry information in the cache.

NetStream supports the following flow aging methods:

- Periodical aging.
- Forced aging.

Periodical aging

Periodical aging uses the following methods:

- **Inactive flow aging**—A flow is inactive if no packet arrives for the NetStream entry within the inactive flow aging timer. When the timer expires, the following events occur:
  - The inactive flow entry is aged out.
  - The statistics of the flow are sent to NetStream servers and are cleared in the cache. The statistics can no longer be displayed by using the **`display ip netstream cache`** command.

  This method ensures that inactive flow entries are cleared from the cache in a timely manner so new entries can be cached.
- **Active flow aging**—A flow is active if packets arrive for the NetStream entry within the active flow aging timer. When the timer expires, the statistics of the active flow are exported to NetStream servers. The device continues to collect active flow statistics.

  This method periodically exports the statistics of active flows to NetStream servers.

Forced aging

To implement forced aging, execute the **`reset ip netstream statistics`** command to clear the NetStream cache immediately and export the cached entries to NetStream servers.

NetStream data export

Traditional data export

Traditional NetStream collects the statistics of each flow and exports the statistics to NetStream servers.

This method consumes more bandwidth and CPU than the aggregation method, and it requires a large cache size.

Aggregation data export

NetStream aggregation merges the flow statistics according to the aggregation criteria of an aggregation mode, and it sends the summarized data to NetStream servers. The NetStream aggregation data export uses less bandwidth than the traditional data export.

Table 1 lists the available aggregation modes. In each mode, the system merges statistics for multiple flows into statistics for one aggregate flow if each aggregation criterion is of the same value. The system records the statistics for the aggregate flow. These aggregation modes work independently and can take effect concurrently.

For example, when the aggregation mode configured on the NDE is protocol-port, NetStream aggregates the statistics of flow entries by protocol number, source port, and destination port. Four NetStream entries record four TCP flows with the same destination address, source port, and destination port, but with different source addresses. In the aggregation mode, only one NetStream aggregation entry is created and sent to NetStream servers.

Table 1 NetStream aggregation modes

Aggregation mode	Aggregation criteria
AS aggregation	• Source AS number • Destination AS number • Inbound interface index • Outbound interface index
Protocol-port aggregation	• Protocol number • Source port • Destination port
Source-prefix aggregation	• Source AS number • Source address mask length • Source prefix (source network address) • Inbound interface index
Destination-prefix aggregation	• Destination AS number • Destination address mask length • Destination prefix (destination network address) • Outbound interface index
Prefix aggregation	• Source AS number • Destination AS number • Source address mask length • Destination address mask length • Source prefix • Destination prefix • Inbound interface index • Outbound interface index
Prefix-port aggregation	• Source prefix • Destination prefix • Source address mask length

Aggregation mode	Aggregation criteria
	• Destination address mask length • ToS • Protocol number • Source port • Destination port • Inbound interface index • Outbound interface index
ToS-AS aggregation	• ToS • Source AS number • Destination AS number • Inbound interface index • Outbound interface index
ToS-source-prefix aggregation	• ToS • Source AS number • Source prefix • Source address mask length • Inbound interface index
ToS-destination-prefix aggregation	• ToS • Destination AS number • Destination address mask length • Destination prefix • Outbound interface index
ToS-prefix aggregation	• ToS • Source AS number • Source prefix • Source address mask length • Destination AS number • Destination address mask length • Destination prefix • Inbound interface index • Outbound interface index
ToS-protocol-port aggregation	• ToS • Protocol type • Source port • Destination port • Inbound interface index • Outbound interface index
ToS-BGP-nexthop	• ToS • BGP next hop • Outbound interface index

If packets are not forwarded according to the BGP routing table, the AS number or BGP next hop cannot be obtained.

NetStream export formats

NetStream exports data in UDP datagrams in one of the following formats:

- **Version 5**—Exports original statistics collected based on the 7-tuple elements and does not support the NetStream aggregation data export. The packet format is fixed and cannot be extended.

- **Version 8**—Supports the NetStream aggregation data export. The packet format is fixed and cannot be extended.
- **Version 9**—Based on a template that can be configured according to the template formats defined in RFCs. Version 9 supports exporting the NetStream aggregation data and collecting statistics about BGP next hop packets.
- **Version 10**—Similar to version 9. The difference between version 9 and version 10 is that version 10 export format is compliant with the IPFIX standard.

NetStream filtering

NetStream filtering uses an ACL to identify packets. Whether NetStream collects data for identified packets depends on the action in the matching rule.

- NetStream collects data for packets that match permit rules in the ACL.
- NetStream does not collect data for packets that match deny rules in the ACL.

For more information about ACL, see *ACL and QoS Configuration Guide*.

NetStream sampling

NetStream sampling collects statistics on fewer packets and is useful when the network has a large amount of traffic. NetStream on sampled traffic lessens the impact on the device's performance. For more information about sampling, see "Configuring samplers."

Protocols and standards

RFC 5101, *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information*

Restrictions and guidelines: NetStream configuration

If the following features (listed in descending order of priority) are configured, only the feature with the highest priority takes effect:

- NetAnalysis for UDP traffic.
- NetAnalysis unified flow analytics.
- FGLB adaptive routing.
- IPv4 NetStream and IPv6 NetStream.
- MOD and flow group in delay monitoring mode.

For more information about NetAnalysis for UDP traffic and NetAnalysis unified flow analytics, see "Configuring NetAnalysis." For more information about FGLB adaptive routing, see adaptive routing in *Layer 3—IP Routing Configuration Guide*. For more information about MOD and flow groups, see MOD configuration and flow group configuration in *Telemetry Configuration Guide*.

For SF interface modules, if you configure NetStream for incoming traffic on the input interface and outgoing traffic on the output interface for a flow, the NetStream configuration takes effect in only one direction.

NetStream tasks at a glance

To configure NetStream, perform the following tasks:

1.
2. Enabling NetStream
3. (Optional.) Setting the NetStream flow classification key fields
4. (Optional.) Configuring NetStream sampling
5. (Optional.) Configuring the NetStream data export format
6. (Optional.) Configuring the refresh rate for NetStream version 9 or version 10 template
7. (Optional.) Configuring NetStream flow aging
   - Configuring periodical flow aging
8. Configuring the NetStream data export
   a. Configuring the NetStream traditional data export
   b. (Optional.) Configuring the NetStream aggregation data export
9. (Optional.) Enabling SNMP notifications for NetStream

Enabling NetStream

Enabling NetStream on an interface

Restrictions and guidelines

Interface-specific NetStream configuration and interface-specific NetStream filtering are mutually exclusive. Do not use them simultaneously.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable NetStream on the interface.
   ip netstream { **inbound** | **outbound** }
   By default, NetStream is disabled on an interface.

Configuring NetStream filtering

About this task

Perform this task to enable NetStream on an interface and to collect traffic statistics about traffic matching an ACL.

NetStream filtering uses an ACL to identify packets.

- To enable NetStream to collect statistics for specific flows, use the ACL permit statements to identify these flows.
- To disable NetStream from collecting statistics for specific flows, use the ACL deny statements to identify these flows.

Restrictions and guidelines

Interface-specific NetStream configuration and interface-specific NetStream filtering are mutually exclusive. Do not use them simultaneously.

When NetStream filtering and sampling are both configured, packets are filtered first, and then the permitted packets are sampled.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable NetStream filtering on the interface.
   ip netstream { **inbound** | **outbound** } **filter acl** *acl-number*

   By default, NetStream filtering is disabled. NetStream collects statistics of all IPv4 packets passing through the interface.

Setting the NetStream flow classification key fields

About this task

By default, the hardware classifies packets into NetStream flows based on these key fields: source and destination IP addresses, source and destination MAC addresses, and protocol. Packets matching the same key fields are grouped into one flow and sent to NetStream regularly for further tracking.

You can set the flow classification key fields to suit specific traffic monitoring requirements. For example, if you are concerned with only the traffic destination, you can set the destination IP address as the flow classification key field. All packets destined for the same IP address are classified into one flow.

Restrictions and guidelines

To collect traffic flows based on the 7-tuple, configure the following key fields for NetStream flow classification:

- **destination-ip**.
- **destination-port**.
- **ip-tos**.
- **protocol**.
- **source-ip**.
- **source-port**.

If you do not follow this restriction, NetStream will set these fields to fixed values for all flows.

If you execute the command multiple times, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Set the flow classification key fields.

netstream key [**destination-ip** | **destination-mac** | **destination-port** | **flow-label** | **ip-tos protocol** | **source-ip** | **source-mac** | **source-port**] *

By default, no flow classification key field is set.

Configuring NetStream sampling

Restrictions and guidelines

By default, NetStream collects all data of target flows. If the flow traffic is heavy, NetStream is resource-consuming and can cause high CPU usage, which impacts the device forwarding performance. NetStream sampling is helpful to decrease the NetStream traffic volume. If the collected statistics can basically reflect the network status, you can enable this feature and set a proper sampling rate. The higher the sampling rate, the less the impact on device performance.

If NetStream sampling and filtering are both configured, packets are filtered first, and then the permitted packets are sampled.

This feature is available only on Layer 2 and Layer 3 Ethernet interfaces. It is not available on Layer 3 Ethernet subinterfaces and other types of interfaces.

Procedure

1. Enter system view.
   system-view
2. Create a sampler.
   sampler *sampler-name* **mode random packet-interval n-power** *rate*
   For more information about a sampler, see "Configuring samplers."
3. Enter interface view.
   interface *interface-type interface-number*
4. Enable NetStream sampling.
   ip netstream { **inbound** | **outbound** } **sampler** *sampler-name*
   By default, NetStream sampling is disabled.

Configuring the NetStream data export format

About this task

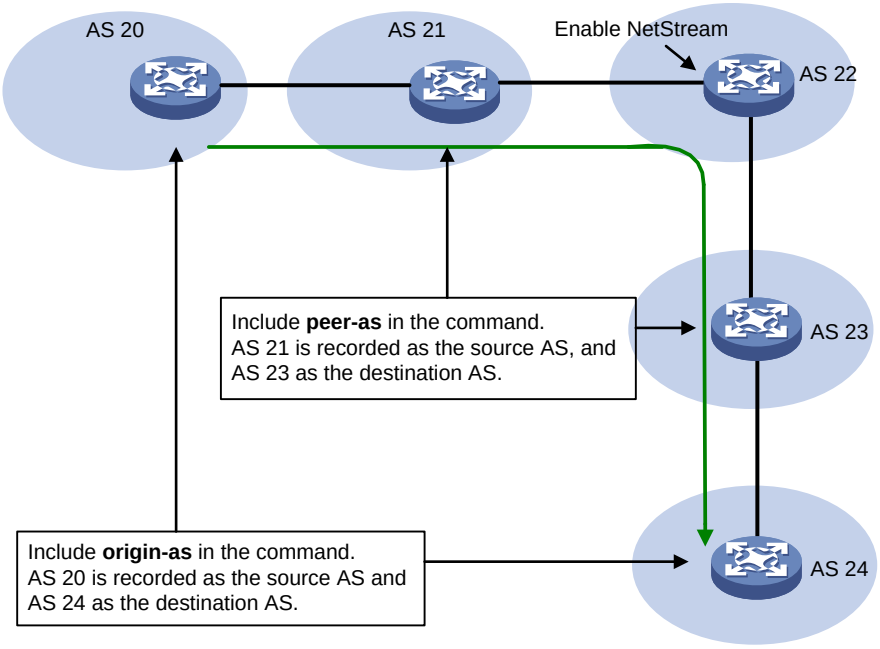
When you configure the NetStream data export format, you can also specify the following settings:

- Whether or not to export the BGP next hop and serial number information.
  Only version 9 and version 10 formats support exporting the BGP next hop and serial number information.
- How to export the autonomous system (AS) information: **origin-as** or **peer-as**.
  - **origin-as**—Records the original AS numbers for the flow source and destination.
  - **peer-as**—Records the peer AS numbers for the flow source and destination.

For example, as shown in Figure 2, a flow starts at AS 20, passes AS 21 through AS 23, and then reaches AS 24. NetStream is enabled on the device in AS 22.

- Specify the **origin-as** keyword to export AS 20 as the source AS and AS 24 as the destination AS.
- Specify the **peer-as** keyword to export AS 21 as the source AS and AS 23 as the destination AS.

Figure 2 Recorded AS information varies by different keyword configurations



Procedure

1. Enter system view.

   ```
   system-view
   ```

2. Configure the NetStream data export format, and configure the AS, BGP next hop, and serial number export attributes. Choose one option as needed:
   - Set NetStream data export format to version 5 and configure the AS export attribute.

     ```
     ip netstream export version 5 { origin-as | peer-as }
     ```

   - Set NetStream data export format to version 9 or version 10 and configure the AS, BGP export, and serial number attributes.

     ```
     ip netstream export version { 9 | 10 } { origin-as | peer-as }
     [ bgp-nexthop ] [ serial-number ]
     ```

   By default:
   - NetStream data export uses the version 9 format.
   - The peer AS numbers for the flow source and destination are exported.
   - The BGP next hop and serial number information is not exported.

Configuring the refresh rate for NetStream version 9 or version 10 template

About this task

Version 9 and version 10 are template-based and support user-defined formats. A NetStream device must send the template to NetStream servers regularly to update the template on the servers.

For a NetStream server to use the correct version 9 or version 10 template, configure the time-based or packet count-based refresh rate. If both settings are configured, the template is sent when either of the conditions is met.

Procedure

1. Enter system view.
   system-view
2. Configure the refresh rate for the NetStream version 9 or version 10 template.
   ip netstream export template refresh-rate { **packet** *packets* | **time** *minutes* }
   By default, the packet count-based refresh rate is 20 packets, and the time-based refresh interval is 30 minutes.

Configuring NetStream flow aging

Configuring periodical flow aging

1. Enter system view.
   system-view
2. Set the aging timer for active flows.
   ip netstream timeout active *minutes*
   By default, the aging timer for active flows is 5 minutes.
3. Set the aging timer for inactive flows.
   ip netstream timeout inactive *seconds*
   By default, the aging timer for inactive flows is 300 seconds.

Configuring the NetStream data export

Configuring the NetStream traditional data export

1. Enter system view.
   system-view
2. Specify a destination host for NetStream traditional data export.
   ip netstream export host *ip-address udp-port* [**vpn-instance** *vpn-instance-name*]
   By default, no destination host is specified.
3. (Optional.) Specify the source interface for NetStream data packets sent to NetStream servers.
   ip netstream export source interface *interface-type interface-number*
   By default, NetStream data packets take the IP address of their output interface (interface that is connected to the NetStream device) as the source IP address.
   As a best practice, connect the management Ethernet interface to a NetStream server, and configure the interface as the source interface.
4. (Optional.) Limit the data export rate.
   ip netstream export rate *rate*
   By default, the data export rate is not limited.

Configuring the NetStream aggregation data export

About this task

NetStream aggregation is implemented by software.

Restrictions and guidelines

Configurations in NetStream aggregation mode view apply only to the NetStream aggregation data export, and those in system view apply to the NetStream traditional data export. If configurations in NetStream aggregation mode view are not provided, the configurations in system view apply to the NetStream aggregation data export.

If the version 5 format is configured to export NetStream data, NetStream aggregation data export uses the version 8 format.

Procedure

1. Enter system view.

   system-view

2. Specify a NetStream aggregation mode and enter its view.

   ip netstream aggregation { **as** | **destination-prefix** | **prefix** | **prefix-port** | **protocol-port** | **source-prefix** | **tos-as** | **tos-bgp-nexthop** | **tos-destination-prefix** | **tos-prefix** | **tos-protocol-port** | **tos-source-prefix** }

   By default, no NetStream aggregation mode is configured.

3. Enable the NetStream aggregation mode.

   enable

   By default, all NetStream aggregation modes are disabled.

4. Specify a destination host for NetStream aggregation data export.

   ip netstream export host *ip-address udp-port* [**vpn-instance** *vpn-instance-name*]

   By default, no destination host is specified.

   If you expect only NetStream aggregation data, specify the destination host only in the related NetStream aggregation mode view.

5. (Optional.) Specify the source interface for NetStream data packets sent to NetStream servers.

   ip netstream export source interface *interface-type interface-number*

   By default, no source interface is specified for NetStream data packets. The packets take the IP address of the output interface as the source IP address.

   Source interfaces in different NetStream aggregation mode views can be different.

   If no source interface is configured in NetStream aggregation mode view, the source interface configured in system view applies.

Enabling SNMP notifications for NetStream

About this task

Use this feature to report important events for NetStream through SNMP notifications. The generated alarm information will be sent to the SNMP module on the device and you can configure the SNMP notification sending parameters to determine the export attributes of the alarm information. For more information, see "Configuring SNMP."

Procedure

1. Enter system view.

system-view

2. Enable SNMP notifications for NetStream.

   snmp-agent trap enable ip netstream [**entry-limit**]

   By default, SNMP notifications are enabled for NetStream.

Display and maintenance commands for NetStream

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display NetStream entry information.	**display ip netstream cache** [**verbose**] [**type** { **ip** | **ipl2** | **l2** }] [**destination** *destination-ip* | **interface** *interface-type interface-number* | **source** *source-ip*] * [**slot** *slot-number*]
Display information about the NetStream data export.	**display ip netstream export**
Display NetStream template information.	**display ip netstream template** [**slot** *slot-number*]
Age out and export all NetStream data, and clear the cache.	**reset ip netstream statistics**

NetStream configuration examples

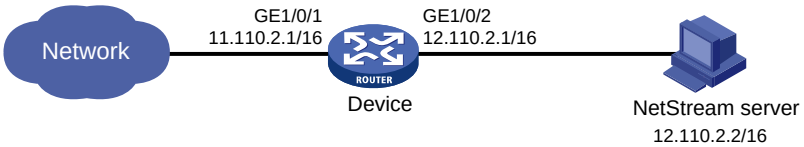
Example: Configuring NetStream traditional data export

Network configuration

As shown in Figure 3, configure NetStream on the device to collect statistics on packets passing through the device.

- Enable NetStream for incoming and outgoing traffic on GigabitEthernet 1/0/1.
- Configure the device to export NetStream traditional data to UDP port 5000 of the NetStream server.

Figure 3 Network diagram



Procedure

Assign an IP address to each interface, as shown in Figure 3. (Details not shown.)

Enable NetStream for incoming and outgoing traffic on GigabitEthernet 1/0/1.

```
<Device> system-view
[Device] interface gigabitethernet 1/0/1
```

```
[Device-GigabitEthernet1/0/1] ip netstream inbound
[Device-GigabitEthernet1/0/1] ip netstream outbound
[Device-GigabitEthernet1/0/1] quit
```

Specify 12.110.2.2 as the IP address of the destination host and UDP port 5000 as the export destination port number.

```
[Device] ip netstream export host 12.110.2.2 5000
```

Verifying the configuration

Display NetStream entry information.

```
[Device] display ip netstream cache
IP NetStream cache information:
  Active flow timeout             : 30 min
  Inactive flow timeout           : 300 sec
  Inactive flow timeout           : 30 sec
  Max number of entries           : 16384
  IP active flow entries          : 2
  MPLS active flow entries        : 0
  L2 active flow entries          : 0
  IPL2 active flow entries        : 0
  IP flow entries counted         : 0
  MPLS flow entries counted       : 0
  L2 flow entries counted         : 0
  IPL2 flow entries counted       : 0
  Last statistics resetting time  : Never

IP packet size distribution (11 packets in total):

 1-32   64   96  128  160  192  224  256  288  320  352  384  416  448  480
 .000 .000 .909 .000 .000 .090 .000 .000 .000 .000 .000 .000 .000 .000 .000

  512  544  576 1024 1536 2048 2560 3072 3584 4096 4608 >4608
 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000

 Protocol          Total  Packets    Flows  Packets Active(sec) Idle(sec)
                   Flows  /sec       /sec   /flow   /flow       /flow
-------------------------------------------------------------------------------

Type DstIP(Port)            SrcIP(Port)           Pro ToS If(Direct)  Pkts
     DstMAC(VLAN)           SrcMAC(VLAN)
     TopLblType(IP/MASK)    Lbl-Exp-S-List
-------------------------------------------------------------------------------
IP   10.1.1.1 (21)         100.1.1.2(1024)         1   0   GE1/0/1(I)  5
IP   100.1.1.2 (1024)      10.1.1.1 (21)           1   0   GE1/0/1(O)  5
```

Display information about the NetStream data export.

```
[Device] display ip netstream cache
  IP active flow entries          : 2
  MPLS active flow entries        : 0
  L2 active flow entries          : 0
  IPL2 active flow entries        : 0
```

```
  IP flow entries counted         : 0
  MPLS flow entries counted       : 0
  L2 flow entries counted         : 0
  IPL2 flow entries counted       : 0
  Last statistics resetting time  : Never

IP packet size distribution (11 packets in total):

 1-32   64   96  128  160  192  224  256  288  320  352  384  416  448  480
 .000 .000 .909 .000 .000 .090 .000 .000 .000 .000 .000 .000 .000 .000 .000

  512  544  576 1024 1536 2048 2560 3072 3584 4096 4608 >4608
 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000

 Protocol          Total  Packets    Flows  Packets Active(sec) Idle(sec)
                   Flows  /sec       /sec   /flow   /flow       /flow
-------------------------------------------------------------------------------

Type DstIP(Port)            SrcIP(Port)            Pro ToS If(Direct)  Pkts
     DstMAC(VLAN)           SrcMAC(VLAN)
     TopLblType(IP/MASK)    Lbl-Exp-S-List
-------------------------------------------------------------------------------
IP   10.1.1.1 (21)         100.1.1.2(1024)         1   0   GE1/0/1(I)  5
IP   100.1.1.2 (1024)      10.1.1.1 (21)           1   0   GE1/0/1(O)  5
```

Display information about the NetStream data export.

```
[Device] display ip netstream export
IP export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 12.110.2.2 (5000)
  Version 5 exported flow number                 : 0
  Version 5 exported UDP datagram number (failed) : 0 (0)
  Version 9 exported flow number                 : 10
  Version 9 exported UDP datagram number (failed) : 10 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

L2 export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 12.110.2.2 (5000)
  Version  9 exported flow number                : 0
  Version  9 exported UDP datagram number (failed): 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

IPL2 export information:
  Flow source interface                          : Not specified
```

```
Flow destination VPN instance                  : Not specified
Flow destination IP address (UDP)              : 12.110.2.2 (5000)
Version  9 exported flow number                : 10
Version  9 exported UDP datagram number (failed): 10 (0)
Version 10 exported flow number                : 0
Version 10 exported UDP datagram number (failed): 0 (0)
```

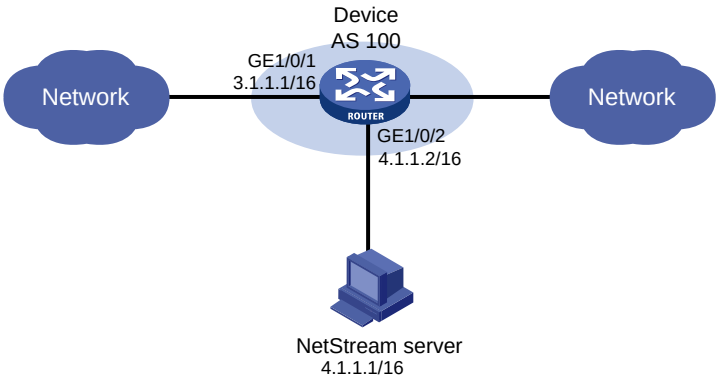
Example: Configuring NetStream aggregation data export

Network configuration

As shown in Figure 4, all routers in the network are running EBGP. Configure NetStream on the device to meet the following requirements:

- Use version 5 format to export NetStream traditional data to port 5000 of the NetStream server.
- Perform NetStream aggregation in the modes of protocol-port, source-prefix, destination-prefix, and prefix.
- Export the aggregation data of different modes to 4.1.1.1, with UDP ports 3000, 4000, 6000, and 7000.

Figure 4 Network diagram



Procedure

Assign an IP address to each interface, as shown in Figure 4. (Details not shown.)

Specify version 5 format to export NetStream traditional data and record the original AS numbers for the flow source and destination.

```
<Device> system-view
[Device] ip netstream export version 5 origin-as
```

Enable NetStream for incoming and outgoing traffic on GigabitEthernet 1/0/1.

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] ip netstream inbound
[Device-GigabitEthernet1/0/1] ip netstream outbound
[Device-GigabitEthernet1/0/1] quit
```

Specify 4.1.1.1 as the IP address of the destination host and UDP port 5000 as the export destination port number.

```
[Device] ip netstream export host 4.1.1.1 5000
```

Set the aggregation mode to protocol-port, and specify the destination host for the aggregation data export.

```
[Device] ip netstream aggregation protocol-port
```

```
[Device-ns-aggregation-protport] enable
[Device-ns-aggregation-protport] ip netstream export host 4.1.1.1 3000
[Device-ns-aggregation-protport] quit
```

Set the aggregation mode to source-prefix, and specify the destination host for the aggregation data export.

```
[Device] ip netstream aggregation source-prefix
[Device-ns-aggregation-srcpre] enable
[Device-ns-aggregation-srcpre] ip netstream export host 4.1.1.1 4000
[Device-ns-aggregation-srcpre] quit
```

Set the aggregation mode to destination-prefix, and specify the destination host for the aggregation data export.

```
[Device] ip netstream aggregation destination-prefix
[Device-ns-aggregation-dstpre] enable
[Device-ns-aggregation-dstpre] ip netstream export host 4.1.1.1 6000
[Device-ns-aggregation-dstpre] quit
```

Set the aggregation mode to prefix, and specify the destination host for the aggregation data export.

```
[Device] ip netstream aggregation prefix
[Device-ns-aggregation-prefix] enable
[Device-ns-aggregation-prefix] ip netstream export host 4.1.1.1 7000
[Device-ns-aggregation-prefix] quit
```

Verifying the configuration

Display information about the NetStream data export.

```
[Device] display ip netstream export
protocol-port aggregation export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 4.1.1.1 (3000)
  Version 8 exported flow number                 : 2
  Version 8 exported UDP datagram number (failed) : 2 (0)
  Version 9 exported flow number                 : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

source-prefix aggregation export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 4.1.1.1 (4000)
  Version 8 exported flow number                 : 2
  Version 8 exported UDP datagram number (failed) : 2 (0)
  Version 9 exported flow number                 : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

destination-prefix aggregation export information:
  Flow source interface                          : Not specified
```

```
  Flow destination VPN instance                   : Not specified
  Flow destination IP address (UDP)               : 4.1.1.1 (6000)
  Version 8 exported flow number                  : 2
  Version 8 exported UDP datagram number (failed) : 2 (0)
  Version 9 exported flow number                  : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                 : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

prefix aggregation export information:
  Flow source interface                           : Not specified
  Flow destination VPN instance                   : Not specified
  Flow destination IP address (UDP)               : 4.1.1.1 (7000)
  Version 8 exported flow number                  : 2
  Version 8 exported UDP datagram number (failed) : 2 (0)
  Version 9 exported flow number                  : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                 : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

IP export information:
  Flow source interface                           : Not specified
  Flow destination VPN instance                   : Not specified
  Flow destination IP address (UDP)               : 4.1.1.1 (5000)
  Version 5 exported flow number                  : 10
  Version 5 exported UDP datagram number (failed) : 10 (0)
  Version 9 exported flow number                  : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                 : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

L2 export information:
  Flow source interface                           : Not specified
  Flow destination VPN instance                   : Not specified
  Flow destination IP address (UDP)               : 12.110.2.2 (5000)
  Version  9 exported flow number                 : 0
  Version  9 exported UDP datagram number (failed): 0 (0)
  Version 10 exported flow number                 : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

IPL2 export information:
  Flow source interface                           : Not specified
  Flow destination VPN instance                   : Not specified
  Flow destination IP address (UDP)               : 12.110.2.2 (5000)
  Version  9 exported flow number                 : 0
  Version  9 exported UDP datagram number (failed): 0 (0)
  Version 10 exported flow number                 : 0
  Version 10 exported UDP datagram number (failed): 0 (0)
```

Configuring session-based NetStream

About session-based NetStream

Session-based NetStream provides statistics for session-based services and exports the statistics in NetStream version 9 format to NetStream servers.

Session-based NetStream aggregation modes

Session-based NetStream aggregates session statistics according to the aggregation criteria of an aggregation mode and exports the statistics to NetStream servers.

The system merges statistics for multiple sessions if each aggregation criterion is of the same value. The system supports only the session aggregating mode in the current software version.

Table 2 Session-based NetStream aggregation mode

Aggregation mode	Aggregation criteria
Session aggregation	• Source MAC address. • Destination MAC address. • Source IP address. • Destination IP address. • Source port number. • Destination port number. • Protocol type.

Session-based NetStream data export

Session-based NetStream uses an aging mechanism to export session statistics to NetStream servers.

Upon receiving a session packet, the device creates a NetStream entry for the session and collects data for the session. The NetStream data for the session is cleared and exported to the NetStream servers when either of the following conditions is met:

- The aging timer of the inactive session expires. The session is inactive if no packet of the session arrives within its aging timer.
- The aging timer of the NetStream entry expires.

If the NetStream cache has insufficient storage space but none of the above conditions is met, the device stops generating NetStream entries for new sessions. Statistics collection for existing sessions continues.

Configuring session-based NetStream settings

1. Enter system view.
   `system-view`
2. Enable session-based NetStream.
   `session-based netstream enable`
   By default, session-based NetStream is disabled.
3. (Optional.) Set the aging timer for session-based NetStream entries.

session-based netstream timeout *minutes*

By default, the aging timer for session-based NetStream entries is 5 minutes.

4. Specify a destination host for session-based NetStream data export.

   session-based netstream export host *ip-address udp-port* [**vpn-instance** *vpn-instance-name*]

   By default, no destination host is specified for session-based NetStream data export.

5. (Optional.) Specify a source IP address for session-based NetStream packets.

   session-based netstream export source ip *ip-address*

   By default, the source IP address of session-based NetStream packets is the primary IP address of the output interface.

Display and maintenance commands for session-based NetStream

Execute **display** commands in any view.

Task	Command
Display session-based NetStream statistics in the cache.	**display session-based netstream aggregation-cache session** [**slot** *slot-number*]

Session-based NetStream configuration examples

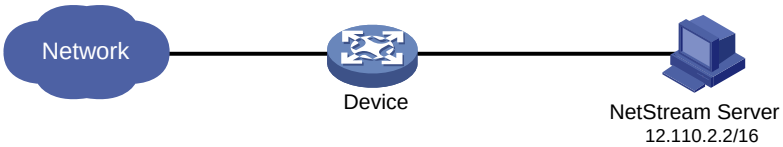
Example: Configuring session-based NetStream data export

Network configuration

As shown in Figure 5:

1. Enable session-based NetStream in application aggregation mode on the device.
2. Configure the device to export the NetStream statistics to UDP port 5000 on the NetStream server.

Figure 5 Network diagram



Procedure

1. Enable session-based NetStream.

   ```
   <Device> system-view
   [Device] session-based netstream enable
   ```

2. Specify 12.110.2.2 as the IP address of the destination host and UDP port 5000 as the export destination port number.

```
[Device] session-based netstream export host 12.110.2.2 5000
```

Verifying the configuration

Display session-based NetStream statistics in session aggregation mode.

```
<Device> display session-based netstream aggregation-cache session
Active entries                                          :1
Timeout time for session-based NetStream entries        :5
Session-timeout time for session-based NetStream entries :400

SrcMAC          DstMAC          SrcIP          DstIP         SrcPort   DstPort
InPkts          InBytes         OutPkts        OutBytes      Pro       LogType
0001-0001-0001  0002-0002-0002  192.168.1.1    192.168.2.1   4000      5000
4               240             4              240           6         0x1
```

Contents

Configuring cloud connections 1

About cloud connections 1
Multiple subconnections 1
Cloud connection establishment 1
Configuring a cloud connection 2
Unbinding the device from the Oasis server 3
Display and maintenance commands for cloud connections 4
Cloud connection configuration examples 4
Example: Configuring a cloud connection 4

Configuring cloud connections

About cloud connections

A cloud connection is a management tunnel established between a local device and the H3C cloud server. It enables you to manage the local device from the cloud server without accessing the network where the device resides.

Multiple subconnections

After a local device establishes a connection with the cloud server, service modules on the local device can establish multiple subconnections with the microservices on the cloud server. These subconnections are independent from each other and provide separate communication channels for different services. This mechanism avoids interference among different services.

Cloud connection establishment

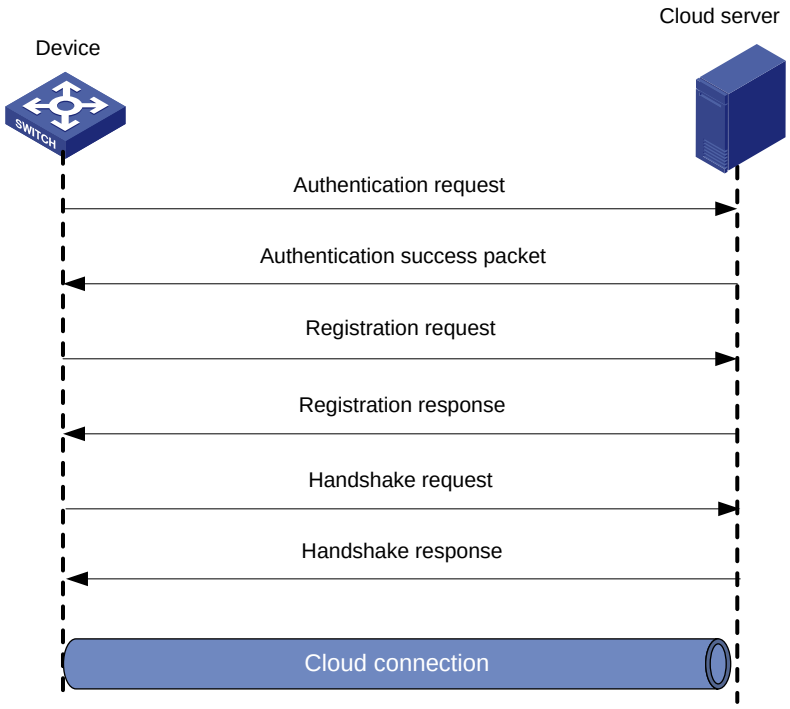
The cloud connection is established as follows:

1. The device sends an authentication request to the cloud server.
2. The cloud server sends an authentication success packet to the device.

   The device passes the authentication only if the serial number of the device has been added to the cloud server. If the authentication fails, the cloud server sends an authentication failure packet to the device.
3. The device sends a registration request to the cloud server.
4. The cloud server sends a registration response to the device.

   The registration response contains the uniform resource locator (URL) used to establish a cloud connection.
5. The device uses the URL to send a handshake request (changing the protocol from HTTP to WebSocket) to the cloud server.
6. The cloud server sends a handshake response to the device to finish establishing the cloud connection.
7. The device sends a subconnection URL request to the cloud server.
8. The cloud server sends a response that contains the microservice list and all subconnection URLs to the device.
9. The service module registers on the cloud connection management module.
10. The cloud connection management module sends the corresponding URLs to the service module.
11. The service module establishes subconnections with the cloud server.

Figure 1 Establishing a cloud connection



Configuring a cloud connection

About this task

For a device to establish a cloud connection to the cloud server, perform either of the following tasks:

- Specify the domain name of the cloud server on the device through CLI.
- Configure the device as a DHCP client and the cloud server as the DHCP server. The device obtains the IP address of the DHCP server and parses the option 253 field in the DHCP packets to obtain the IPv4 domain name of the cloud server. It can parse the option 65253 field in the DHCPv6 packets to obtain the IPv6 domain name of the cloud server. For more information about the option 253 field and the option 65253 field, see DHCP configuration and DHCPv6 configuration in *Layer—3 IP Services Configuration Guide*.

To establish cloud connections to the cloud server, a password is required. A device can use either of the following methods to obtain the password for establishing cloud connections to the cloud server:

- Execute the **`cloud-management server password`** command on the device to specify the password for establishing cloud connections to the cloud server.
- Configure the device as a DHCP client and the cloud server as the DHCP server. The device obtains the IP address of the DHCP server and parses the option 252 field in the DHCP packets to obtain the password for connection to the cloud server. For more information about the option 252 field, see DHCP configuration in *Layer 3—IP Services Configuration Guide*.

After establishing the cloud connection, the local device sends keepalive packets to the cloud server periodically. If the device does not receive a response from the cloud server within three keepalive intervals, the device sends a registration request to re-establish the cloud connection.

Restrictions and guidelines

The domain name obtained through DHCP has a higher priority than the domain name configured manually.

If a device obtains the domain name of the cloud server through DHCP after establishing a cloud connection to the cloud server with the manually configured domain name, the device performs the following tasks:

- If the automatically obtained and manually configured domain names are identical, the device retains the cloud connection.
- If the automatically obtained and manually configured domain names are different, the device tears down the cloud connection and then establishes a cloud connection to the cloud server with the automatically obtained domain name.

The password obtained through DHCP has a higher priority than the password configured manually.

If a device obtains the password for connection to the cloud server through DHCP after establishing a cloud connection to the cloud server with the manually configured password, the device performs the following tasks:

- If the automatically obtained and manually configured passwords are identical, the device retains the cloud connection.
- If the automatically obtained and manually configured passwords are different, the device tears down the cloud connection and then establishes a cloud connection to the cloud server with the automatically obtained password.

Prerequisites

Add the serial number of the device to be managed to the cloud server.

Configure DNS to ensure that the domain name of the cloud server can be translated into an IP address.

Procedure

1. Enter system view.

   system-view

2. Configure the domain name of the cloud server.
   - **Configure the IPv4 domain name of the cloud server.**

     cloud-management server { **domain** *domain-name* [**vpn-instance** *vpn-instance-name*] [**source-ip** *ip-address*] | **port** *port-number* }
   - **Configure the IPv6 domain name of the cloud server.**

     cloud-management server ipv6 domain *domain-name* [**vpn-instance** *vpn-instance-name*] [**source-ip** *ip-address*]

   By default, the domain name of the cloud server is not configured.
3. (Optional) Set the keepalive interval.

   cloud-management keepalive *interval*

   By default, the keepalive interval is 180 seconds.
4. (Optional) Specify the TCP port number used to establish cloud connections.

   cloud-management server port *port-number*

   By default, the TCP port number used to establish cloud connections is 443.
5. (Optional.) Set the password for establishing cloud connections to the cloud server.

   cloud-management server password { **cipher** | **simple** } *string*

   By default, no password is set for establishing cloud connections to the cloud server.

Unbinding the device from the Oasis server

About this task

A device can be registered on the Oasis server by only one user.

To register a device that has been registered by another user, you need to take the following steps:

1. Obtain a verification code for device unbinding from the Oasis server.
2. Execute the command on the device for sending the verification code to the Oasis server.
3. Register the device on the Oasis server.

Procedure

1. Enter system view.
   `system-view`
2. Send the verification code for device unbinding to the Oasis server.
   `cloud-management unbinding-code` *code*

Display and maintenance commands for cloud connections

Execute **`display`** commands in any view and **`reset`** commands in user view.

Task	Command
Display cloud connection state information.	**`display cloud-management state`**
Re-establish the cloud connection to the cloud server.	**`reset cloud-management tunnel`**

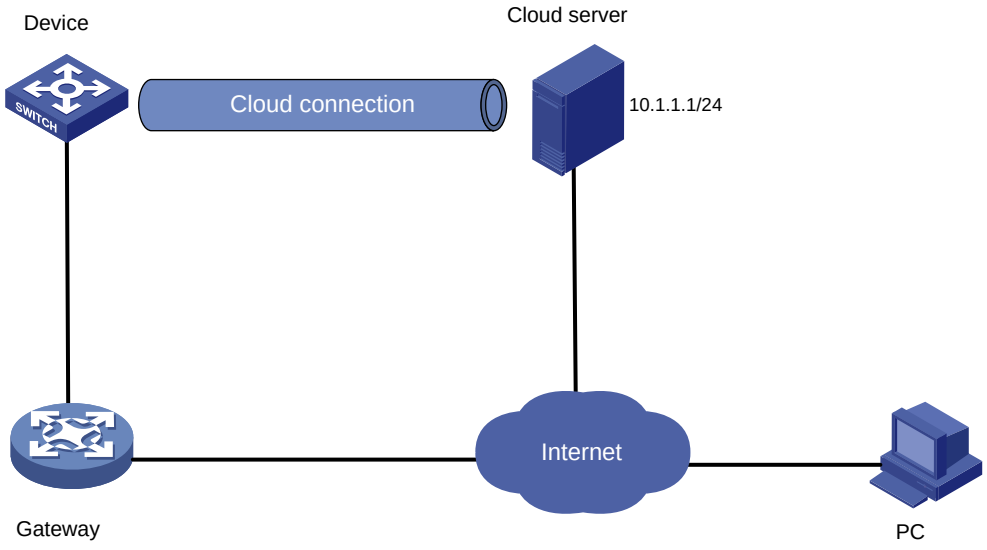
Cloud connection configuration examples

Example: Configuring a cloud connection

Network configuration

As shown in Figure 2, configure the device to establish a cloud connection with the cloud server.

Figure 2 Network diagram



Procedure

1. Configure IP addresses for interfaces, and configure a routing protocol to ensure that the devices can reach each other. (Details not shown.)
2. Log in to the cloud server to add the serial number of the device to the server. (Details not shown.)

   The IP address of the cloud server is 10.1.1.1/24 and the domain name is **example.com**.
3. Configure the domain name of the cloud server as **example.com**.

   ```
   <Device> system-view
   [Device] cloud-management server domain example.com
   ```

   # Map IP address 10.1.1.1 to host name **example.com**.

   ```
   [Device] ip host example.com 10.1.1.1
   ```

Verifying the configuration

Verify that the device and the cloud server have established a cloud connection.

```
<Device> display cloud-management state
Cloud connection state   : Established
Device state             : Request_success
Cloud server address     : 10.1.1.1
Cloud server domain name : example.com
Cloud server port        : 443
Connected at             : Wed Jan  27 14:18:40 2019
Duration                 : 00d 00h 02m 01s
Process state            : Message received
Failure reason           : N/A
```

Contents

Configuring IPv6 NetStream 1

About IPv6 NetStream 1
IPv6 NetStream architecture 1
IPv6 NetStream flow aging 2
IPv6 NetStream data export 3
IPv6 NetStream filtering 4
IPv6 NetStream sampling 4
Protocols and standards 4
Restrictions and guidelines: IPv6 NetStream configuration 4
IPv6 NetStream tasks at a glance 4
Enabling IPv6 NetStream 5
Enabling IPv6 NetStream on an interface 5
Configuring IPv6 NetStream filtering 5
Configuring IPv6 NetStream sampling 6
Configuring the IPv6 NetStream data export format 6
Configuring the refresh rate for IPv6 NetStream version 9 or version 10 template 8
Configuring IPv6 NetStream flow aging 8
Configuring periodical flow aging 8
Configuring the IPv6 NetStream data export 8
Configuring the IPv6 NetStream traditional data export 8
Configuring the IPv6 NetStream aggregation data export 9
Enabling SNMP notifications for IPv6 NetStream 10
Display and maintenance commands for IPv6 NetStream 10
IPv6 NetStream configuration examples 10
Example: Configuring IPv6 NetStream traditional data export 10
Example: Configuring IPv6 NetStream aggregation data export 12

Configuring IPv6 NetStream

About IPv6 NetStream

IPv6 NetStream is an accounting technology that provides statistics on a per-flow basis. An IPv6 flow is defined by the following 8-tuple elements:

- Destination IPv6 address.
- Source IPv6 address.
- Destination port number.
- Source port number.
- Protocol number.
- Traffic class.
- Flow label.
- Input or output interface.

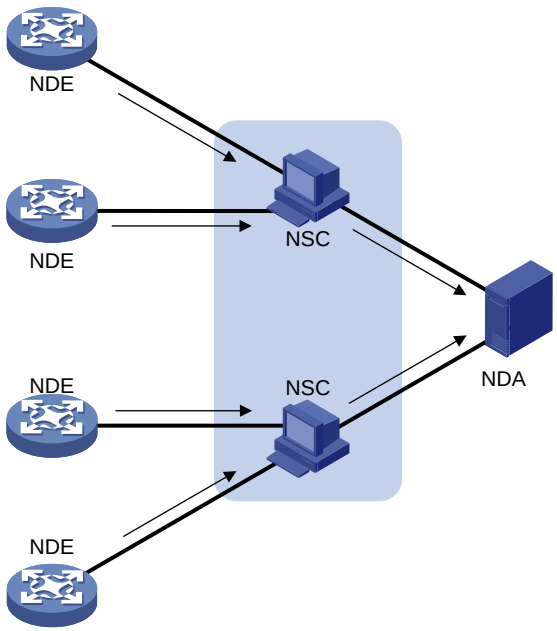
IPv6 NetStream architecture

A typical IPv6 NetStream system includes the following elements:

- **NetStream data exporter**—A device configured with IPv6 NetStream. The NDE provides the following functions:
  - Classifies traffic flows by using the 8-tuple elements.
  - Collects data from the classified flows.
  - Aggregates the data.
  - Exports the data to the NSC.
- **NetStream collector**—A program running in a Unix or Windows operating system. The NSC parses the packets received from the NDEs, and saves the data to its database.
- **NetStream data analyzer**—A network traffic analyzing tool. Based on the data in NSC, the NDA generates reports for traffic billing, network planning, and attack detection and monitoring. The NDA can collect data from multiple NSCs. Typically, the NDA features a Web-based system for easy operation.

NSC and NDA are typically integrated into a NetStream server.

Figure 1 IPv6 NetStream system



IPv6 NetStream flow aging

IPv6 NetStream uses flow aging to enable the NDE to export IPv6 NetStream data to NetStream servers. IPv6 NetStream creates an IPv6 NetStream entry for each flow for storing the flow statistics in the cache.

When a flow is aged out, the NDE performs the following operations:

- Exports the summarized data to NetStream servers in a specific format.
- Clears IPv6 NetStream entry information in the cache.

IPv6 NetStream supports the following flow aging methods:

- Periodical aging.
- Forced aging.

Periodical aging

Periodical aging uses the following methods:

- **Inactive flow aging**—A flow is inactive if no packet arrives for the IPv6 NetStream entry within the inactive flow aging timer. When the timer expires, the following events occur:
  - The inactive flow entry is aged out.
  - The statistics of the flow are sent to NetStream servers and are cleared in the cache. The statistics can no longer be displayed by using the **`display ipv6 netstream cache`** command.

  This method ensures that inactive flow entries are cleared from the cache in a timely manner so new entries can be cached.
- **Active flow aging**—A flow is active if packets arrive for the IPv6 NetStream entry within the active flow aging timer. When the timer expires, the statistics of the active flow are exported to NetStream servers. The device continues to collect its statistics, which can be displayed by using the **`display ipv6 netstream cache`** command.

  The active flow aging method periodically exports the statistics of active flows to NetStream servers.

Forced aging

To implement forced aging, execute the **reset ipv6 netstream statistics** command to clear the IPv6 NetStream cache immediately and export the cached entries to NetStream servers.

IPv6 NetStream data export

Traditional data export

IPv6 NetStream collects the statistics of each flow and exports the statistics to NetStream servers.

This method consumes a lot of bandwidth and CPU usage, and requires a large cache size. In addition, you do not need all of the data in most cases.

Aggregation data export

An IPv6 NetStream aggregation mode merges the flow statistics according to the aggregation criteria of the aggregation mode, and it sends the summarized data to NetStream servers. The IPv6 NetStream aggregation data export uses less bandwidth than the traditional data export.

Table 1 lists the available IPv6 NetStream aggregation modes. In each mode, the system merges multiple flows with the same values for all aggregation criteria into one aggregate flow. The system records the statistics for the aggregate flow. These aggregation modes work independently and can take effect concurrently.

Table 1 IPv6 NetStream aggregation modes

Aggregation mode	Aggregation criteria
AS aggregation	• Source AS number • Destination AS number • Input interface index • Output interface index
Protocol-port aggregation	• Protocol number • Source port • Destination port
Source-prefix aggregation	• Source AS number • Source mask • Source prefix (source network address) • Input interface index
Destination-prefix aggregation	• Destination AS number • Destination mask • Destination prefix (destination network address) • Output interface index
Source-prefix and destination-prefix aggregation	• Source AS number • Source mask • Source prefix (source network address) • Input interface index • Destination AS number • Destination mask • Destination prefix (destination network address) • Output interface index
BGP-nexthop	• BGP next hop • Output interface index

If IPv6 packets are not forwarded according to the BGP routing table, the AS number or BGP next hop cannot be obtained.

IPv6 NetStream data export format

IPv6 NetStream exports data in the version 9 or version 10 format.

Both formats are template-based and support exporting the IPv6 NetStream aggregation data and collecting statistics about BGP next hop.

The version 10 export format is compliant with the IPFIX standard.

IPv6 NetStream filtering

IPv6 NetStream filtering uses an ACL to identify packets. Whether IPv6 NetStream collects data for identified packets depends on the action in the matching rule.

- IPv6 NetStream collects data for packets that match permit rules in the ACL.
- IPv6 NetStream does not collect data for packets that match deny rules in the ACL.

For more information about ACLs, see *ACL and QoS Configuration Guide*.

IPv6 NetStream sampling

IPv6 NetStream sampling collects statistics on fewer packets and is useful when the network has a large amount of traffic. IPv6 NetStream on sampled traffic lessens the impact on the device's performance. For more information about sampling, see "Configuring samplers."

Protocols and standards

RFC 5101, *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information*

Restrictions and guidelines: IPv6 NetStream configuration

If the following features (listed in descending order of priority) are configured, only the feature with the highest priority takes effect:

- NetAnalysis for UDP traffic.
- NetAnalysis unified flow analytics.
- FGLB adaptive routing.
- IPv4 NetStream and IPv6 NetStream.
- MOD and flow group in delay monitoring mode.

For more information about NetAnalysis for UDP traffic and NetAnalysis unified flow analytics, see "Configuring NetAnalysis." For more information about FGLB adaptive routing, see adaptive routing in *Layer 3—IP Routing Configuration Guide*. For more information about MOD and flow groups, see MOD configuration and flow group configuration in *Telemetry Configuration Guide*.

IPv6 NetStream tasks at a glance

To configure IPv6 NetStream, perform the following tasks:

1. Enabling IPv6 NetStream
2. (Optional.) Configuring IPv6 NetStream sampling
3. (Optional.) Configuring the IPv6 NetStream data export format

4. (Optional.) Configuring the refresh rate for IPv6 NetStream version 9 or version 10 template
5. (Optional.) Configuring IPv6 NetStream flow aging
   - Configuring periodical flow aging
6. Configuring the IPv6 NetStream data export
   a. Configuring the IPv6 NetStream traditional data export
   b. (Optional.) Configuring the IPv6 NetStream aggregation data export
7. (Optional.) Enabling SNMP notifications for IPv6 NetStream

Enabling IPv6 NetStream

Enabling IPv6 NetStream on an interface

Restrictions and guidelines

Interface-specific IPv6 NetStream configuration and interface-specific IPv6 NetStream filtering are mutually exclusive. Do not use them simultaneously.

This feature is available only on Layer 2 and Layer 3 Ethernet interfaces. It is not available on Layer 3 Ethernet subinterfaces and other types of interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Enable IPv6 NetStream on the interface.
   ipv6 netstream { **inbound** | **outbound** }
   By default, IPv6 NetStream is disabled on an interface.

Configuring IPv6 NetStream filtering

About this task

Perform this task to enable IPv6 NetStream on an interface and to collect traffic statistics about traffic matching an ACL.

IPv6 NetStream filtering uses an ACL to identify packets.

- To enable IPv6 NetStream to collect statistics for specific flows, use the ACL permit statements to identify these flows.
- To disable IPv6 NetStream from collecting statistics for specific flows, use the ACL deny statements to identify these flows.

Restrictions and guidelines

If IPv6 NetStream filtering and sampling are both configured, IPv6 packets are filtered first, and then the permitted packets are sampled.

Interface-specific IPv6 NetStream configuration and interface-specific IPv6 NetStream filtering are mutually exclusive. Do not use them simultaneously.

This feature is available only on Layer 2 and Layer 3 Ethernet interfaces. It is not available on Layer 3 Ethernet subinterfaces and other types of interfaces.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Configure IPv6 NetStream filtering on the interface.
   ipv6 netstream { **inbound** | **outbound** } **filter acl** *acl-number*
   By default, IPv6 NetStream filtering is disabled. IPv6 NetStream collects statistics of all IPv6 packets passing through the interface.

Configuring IPv6 NetStream sampling

Restrictions and guidelines

By default, IPv6 NetStream collects all data of target flows. If the flow traffic is heavy, IPv6 NetStream is resource-consuming and can cause high CPU usage, which impacts the device forwarding performance. IPv6 NetStream sampling is helpful to decrease the IPv6 NetStream traffic volume. If the collected statistics can basically reflect the network status, you can enable this feature and set a proper sampling rate. The higher the sampling rate, the less impact on device performance.

If IPv6 NetStream sampling and filtering are both configured, IPv6 packets are filtered first, and then the permitted packets are sampled.

This feature is available only on Layer 2 and Layer 3 Ethernet interfaces. It is not available on Layer 3 Ethernet subinterfaces and other types of interfaces.

Procedure

1. Enter system view.
   system-view
2. Create a sampler.
   sampler *sampler-name* **mode random packet-interval n-power** *rate*
   For more information about samplers, see "Configuring samplers."
3. Enter interface view.
   interface *interface-type interface-number*
4. Configure IPv6 NetStream sampling.
   ipv6 netstream { **inbound** | **outbound** } **sampler** *sampler-name*
   By default, IPv6 NetStream sampling is disabled.

Configuring the IPv6 NetStream data export format

About this task

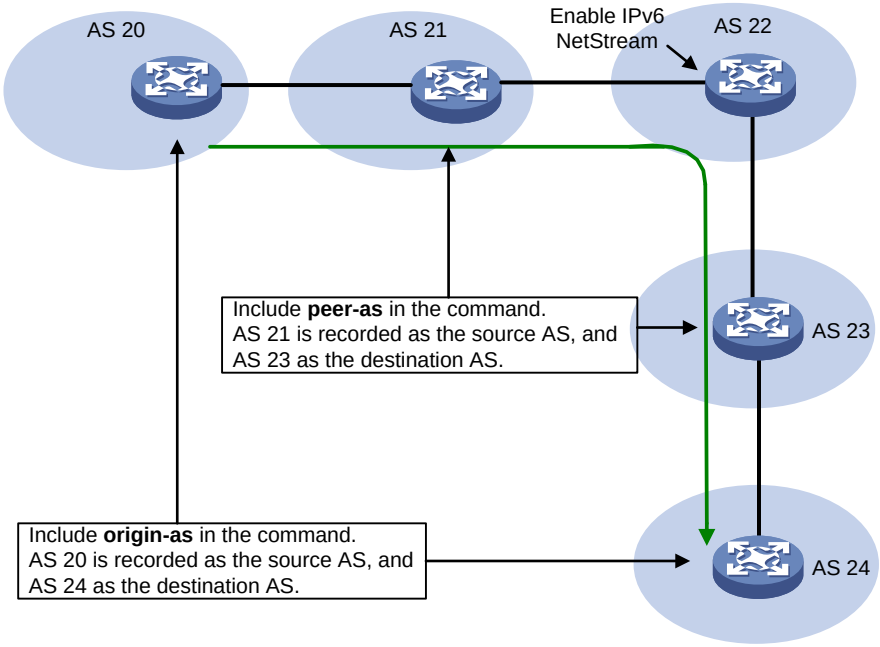
When you configure the IPv6 NetStream data export format, you can also specify the following settings:

- Whether or not to export the BGP next hop and serial number information.
- How to export the autonomous system (AS) information: **origin-as** or **peer-as**.
  - **origin-as**—Records the original AS numbers for the flow source and destination.
  - **peer-as**—Records the peer AS numbers for the flow source and destination.

For example, as shown in Figure 2, a flow starts at AS 20, passes AS 21 through AS 23, and then reaches AS 24. IPv6 NetStream is enabled on the device in AS 22.

- Specify the **origin-as** keyword to export AS 20 as the source AS and AS 24 as the destination AS.
- Specify the **peer-as** keyword to export AS 21 as the source AS and AS 23 as the destination AS.

Figure 2 Recorded AS information varies by different keyword configurations



Procedure

1. Enter system view.

   system-view

2. Configure the IPv6 NetStream data export format, and configure the AS, BGP next hop, and serial number export attributes.
   - Configure the version 9 format.

     ipv6 netstream export version 9 { **origin-as** | **peer-as** } [**bgp-nexthop**] [**serial-number**]
   - Configure the version 10 format.

     ipv6 netstream export version 10 [**origin-as** | **peer-as**] [**bgp-nexthop**] [**serial-number**]

   By default:
   - The version 9 format is used to export IPv6 NetStream data.
   - The peer AS numbers for the flow source and destination are exported.
   - The BGP next hop and serial number information is not exported.

Configuring the refresh rate for IPv6 NetStream version 9 or version 10 template

About this task

Version 9 and version 10 are template-based and support user-defined formats. An IPv6 NetStream device must send the updated template to NetStream servers regularly, because the servers do not permanently save templates.

For a NetStream server to use the correct version 9 or version 10 template, configure the time-based or packet count-based refresh rate. If both settings are configured, the template is sent when either of the conditions is met.

Procedure

1. Enter system view.

   system-view

2. Configure the refresh rate for the IPv6 NetStream version 9 or version 10 template.

   ipv6 netstream export template refresh-rate { **packet** *packets* | **time** *minutes* }

   By default, the packet count-based refresh rate is 20 packets, and the time-based refresh interval is 30 minutes.

Configuring IPv6 NetStream flow aging

Configuring periodical flow aging

1. Enter system view.

   system-view

2. Set the aging timer for active flows.

   ipv6 netstream timeout active *minutes*

   By default, the aging timer for active flows is 5 minutes.

3. Set the aging timer for inactive flows.

   ipv6 netstream timeout inactive *seconds*

   By default, the aging timer for inactive flows is 300 seconds.

Configuring the IPv6 NetStream data export

Configuring the IPv6 NetStream traditional data export

1. Enter system view.

   system-view

2. Specify a destination host for IPv6 NetStream traditional data export.

   ipv6 netstream export host { *ipv4-address* | *ipv6-address* } *udp-port* [**vpn-instance** *vpn-instance-name*]

   By default, no destination host is specified.

3. (Optional.) Specify the source interface for IPv6 NetStream data packets sent to the NetStream servers.

ipv6 netstream export source interface *interface-type interface-number*

By default, no source interface is specified for IPv6 NetStream data packets. The packets take the IPv6 address of the output interface (interface that is connected to the NetStream server) as the source IPv6 address.

As a best practice, connect the management Ethernet interface to a NetStream server, and configure the interface as the source interface.

4. (Optional.) Limit the IPv6 NetStream data export rate.

   ipv6 netstream export rate *rate*

   By default, the data export rate is not limited.

Configuring the IPv6 NetStream aggregation data export

About this task

IPv6 NetStream aggregation is implemented by software.

Restrictions and guidelines

Configurations in IPv6 NetStream aggregation mode view apply only to the IPv6 NetStream aggregation data export. Configurations in system view apply to the IPv6 NetStream traditional data export. When no configuration in IPv6 NetStream aggregation mode view is provided, the configurations in system view apply to the IPv6 NetStream aggregation data export.

Procedure

1. Enter system view.

   system-view

2. Specify an IPv6 NetStream aggregation mode and enter its view.

   ipv6 netstream aggregation { **as** | **bgp-nexthop** | **destination-prefix** | **prefix** | **protocol-port** | **source-prefix** }

   By default, no IPv6 NetStream aggregation mode is specified.

3. Enable the IPv6 NetStream aggregation mode.

   enable

   By default, the IPv6 NetStream aggregation is disabled.

4. Specify a destination host for IPv6 NetStream aggregation data export.

   ipv6 netstream export host { *ipv4-address* | *ipv6-address* } *udp-port* [**vpn-instance** *vpn-instance-name*]

   By default, no destination host is specified.

   If you expect only IPv6 NetStream aggregation data, specify the destination host only in the related IPv6 NetStream aggregation mode view.

5. (Optional.) Specify the source interface for IPv6 NetStream data packets sent to the NetStream servers.

   ipv6 netstream export source interface *interface-type interface-number*

   By default, no source interface is specified for IPv6 NetStream data packets. The packets take the IPv6 address of the output interface as the source IPv6 address.

   You can configure different source interfaces in different IPv6 NetStream aggregation mode views.

   If no source interface is configured in IPv6 NetStream aggregation mode view, the source interface configured in system view applies.

Enabling SNMP notifications for IPv6 NetStream

About this task

Use this feature to report important events for IPv6 NetStream through SNMP notifications. The generated alarm information will be sent to the SNMP module on the device and you can configure the SNMP notification sending parameters to determine the export attributes of the alarm information. For more information, see "Configuring SNMP."

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for IPv6 NetStream.
   snmp-agent trap enable ipv6 netstream [**entry-limit**]
   By default, SNMP notifications are enabled for IPv6 NetStream.

Display and maintenance commands for IPv6 NetStream

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display IPv6 NetStream entry information.	**display ipv6 netstream cache** [**verbose**] [**type** { **ip** | **ipl2** | **l2** }] [**destination** *destination-ipv6* | **interface** *interface-type interface-number* | **source** *source-ipv6*] * [**slot** *slot-number*]
Display information about the IPv6 NetStream data export.	**display ipv6 netstream export**
Display IPv6 NetStream template information.	**display ipv6 netstream template** [**slot** *slot-number*]
Age out, export all IPv6 NetStream data, and clear the cache.	**reset ipv6 netstream statistics**

IPv6 NetStream configuration examples

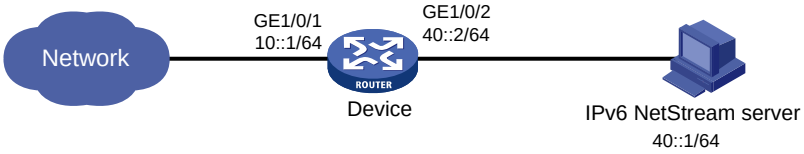
Example: Configuring IPv6 NetStream traditional data export

Network configuration

As shown in Figure 3, configure IPv6 NetStream on the device to collect statistics on packets passing through the device.

- Enable IPv6 NetStream for incoming and outgoing traffic on GigabitEthernet 1/0/1.
- Configure the device to export the IPv6 NetStream traditional data to UDP port 5000 of the NetStream server.

Figure 3 Network diagram



Procedure

Assign an IP address to each interface, as shown in Figure 3. (Details not shown.)

Enable IPv6 NetStream for incoming and outgoing traffic on GigabitEthernet 1/0/1.

```
<Device> system-view
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] ipv6 netstream inbound
[Device-GigabitEthernet1/0/1] ipv6 netstream outbound
[Device-GigabitEthernet1/0/1] quit
```

Specify 40::1 as the IP address of the destination host and UDP port 5000 as the export destination port number.

```
[Device] ipv6 netstream export host 40::1 5000
```

Verifying the configuration

Display information about IPv6 NetStream entries.

```
<Device> display ipv6 netstream cache
IPv6 NetStream cache information:
  Active flow timeout             : 5 min
  Inactive flow timeout           : 300 sec
  Max number of entries           : 16384
  IPv6 active flow entries        : 2
  MPLS active flow entries        : 0
  IPL2 active flow entries        : 0
  IPv6 flow entries counted       : 10
  MPLS flow entries counted       : 0
  IPL2 flow entries counted       : 0
  Last statistics resetting time  : 01/01/2019 at 00:01:02

IPv6 packet size distribution (1103746 packets in total):
1-32   64   96  128  160  192  224  256  288  320  352  384  416  448  480
.249 .694 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000

512  544  576 1024 1536 2048 2560 3072 3584 4096 4608 >4608
.000 .000 .027 .000 .027 .000 .000 .000 .000 .000 .000 .000

Protocol          Total Packets    Flows  Packets Active(sec) Idle(sec)
                  Flows /sec       /sec   /flow   /flow       /flow
-------------------------------------------------------------------------------
TCP-Telnet      2656855     372        4       86        49         27
TCP-FTP         5900082      86        9        9        11         33
TCP-FTPD        3200453    1006        5      193        45         33
TCP-WWW       546778274   11170      887       12         8         32
TCP-other      49148540    3752       79       47        30         32
```

```
UDP-DNS       117240379     570      190        3         7         34
UDP-other      45502422    2272       73       30         8         37
ICMP           14837957     125       24        5        12         34
IP-other          77406       5        0       47        52         27

 Type DstIP(Port)        SrcIP(Port)      Pro TC  FlowLbl If(Direct)   Pkts
      DstMAC(VLAN)       SrcMAC(VLAN)
      TopLblType(IP/MASK)Lbl-Exp-S-List
------------------------------------------------------------------------------
IP   2001::1(1024)      2002::1(21)        6   0   0x0     GE1/0/1(I)  42996
IP   2002::1(21)        2001::1(1024)      6   0   0x0     GE1/0/1(O)  42996
```

Display information about the IPv6 NetStream data export.

```
[Device] display ipv6 netstream export
IPv6 export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 40::1 (5000)
  Version 9 exported flow number                 : 10
  Version 9 exported UDP datagram number (failed) : 10 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

IPL2 export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 40::1 (5000)
  Version 9 exported flow number                 : 10
  Version 9 exported UDP datagram number (failed) : 10 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)
```

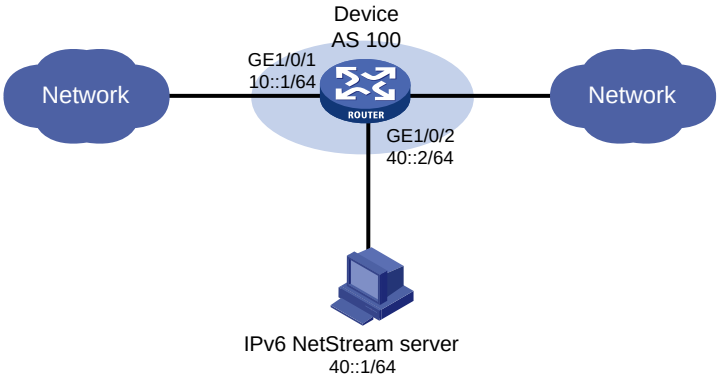
Example: Configuring IPv6 NetStream aggregation data export

Network configuration

As shown in Figure 4, all routers in the network are running IPv6 EBGP. Configure IPv6 NetStream on the device to meet the following requirements:

- Export the IPv6 NetStream traditional data to port 5000 of the NetStream server.
- Perform the IPv6 NetStream aggregation in the modes of protocol-port, source-prefix, destination-prefix, and prefix.
- Export the aggregation data of different modes to the UDP ports 3000, 4000, 6000, and 7000.

Figure 4 Network diagram



Procedure

Assign an IP address to each interface, as shown in Figure 4. (Details not shown.)

Enable IPv6 NetStream for incoming and outgoing traffic on GigabitEthernet 1/0/1.

```
<Device> system-view
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] ipv6 netstream inbound
[Device-GigabitEthernet1/0/1] ipv6 netstream outbound
[Device-GigabitEthernet1/0/1] quit
```

Specify 40::1 as the IP address of the destination host and UDP port 5000 as the export destination port number.

```
[Device] ipv6 netstream export host 40::1 5000
```

Set the aggregation mode to protocol-port, and specify the destination host for the aggregation data export.

```
[Device] ipv6 netstream aggregation protocol-port
[Device-ns6-aggregation-protport] enable
[Device-ns6-aggregation-protport] ipv6 netstream export host 40::1 3000
[Device-ns6-aggregation-protport] quit
```

Set the aggregation mode to source-prefix, and specify the destination host for the aggregation data export.

```
[Device] ipv6 netstream aggregation source-prefix
[Device-ns6-aggregation-srcpre] enable
[Device-ns6-aggregation-srcpre] ipv6 netstream export host 40::1 4000
[Device-ns6-aggregation-srcpre] quit
```

Set the aggregation mode to destination-prefix, and specify the destination host for the aggregation data export.

```
[Device] ipv6 netstream aggregation destination-prefix
[Device-ns6-aggregation-dstpre] enable
[Device-ns6-aggregation-dstpre] ipv6 netstream export host 40::1 6000
[Device-ns6-aggregation-dstpre] quit
```

Set the aggregation mode to prefix, and specify the destination host for the aggregation data export.

```
[Device] ipv6 netstream aggregation prefix
[Device-ns6-aggregation-prefix] enable
[Device-ns6-aggregation-prefix] ipv6 netstream export host 40::1 7000
[Device-ns6-aggregation-prefix] quit
```

Verifying the configuration

Display information about the IPv6 NetStream data export.

```
[Device] display ipv6 netstream export
protocol-port aggregation export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 40::1 (3000)
  Version 9 exported flow number                 : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

source-prefix aggregation export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 40::1 (4000)
  Version 9 exported flow number                 : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

destination-prefix aggregation export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 40::1 (6000)
  Version 9 exported flow number                 : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

prefix aggregation export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 40::1 (7000)
  Version 9 exported flow number                 : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)

IPv6 export information:
  Flow source interface                          : Not specified
  Flow destination VPN instance                  : Not specified
  Flow destination IP address (UDP)              : 40::1 (5000)
  Version 9 exported flow number                 : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                : 0
  Version 10 exported UDP datagram number (failed): 0 (0)
```

```
IPL2 export information:
  Flow source interface                            : Not specified
  Flow destination VPN instance                    : Not specified
  Flow destination IP address (UDP)                : 40::1 (5000)
  Version 9 exported flow number                   : 0
  Version 9 exported UDP datagram number (failed) : 0 (0)
  Version 10 exported flow number                  : 0
  Version 10 exported UDP datagram number (failed): 0 (0)
```

Contents

Managing device performance 1

About performance management 1
Restrictions and guidelines: Performance management 1
Performance management tasks at a glance 1
Configuring performance statistics tasks 1
Enabling performance statistics 2
Display and maintenance commands for performance management configuration 3
Performance management configuration examples 3
Example: Configuring performance management (gRPC) 3
Example: Configuring performance management (NETCONF) 5

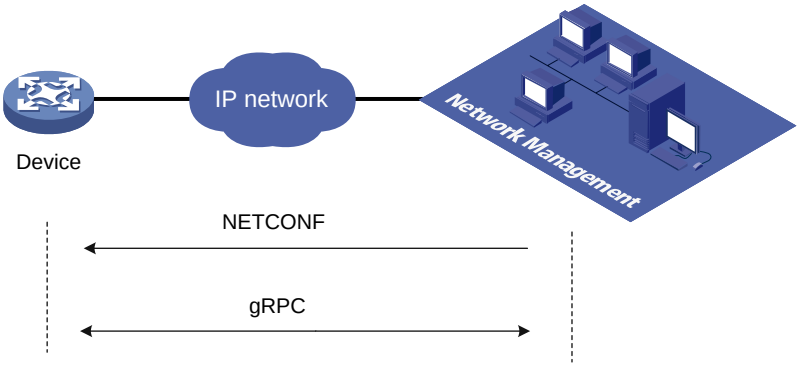
Managing device performance

About performance management

The performance management feature samples device performance and collects performance statistics, such as packet drop statistics in the inbound and outbound directions on an interface. You can use commands on the device to display the statistics or configure the device to transfer the statistics to a network management server through gRPC or NETCONF.

For more information about gRPC, see gRPC configuration in *Telemetry Configuration Guide*.

Figure 1 Performance management diagram



Restrictions and guidelines: Performance management

This feature supports collecting only statistics for traffic on interfaces in the current software version.

Performance management tasks at a glance

To manage device performance, perform the following tasks:

1. Configuring performance statistics tasks
2. Enabling performance statistics

Configuring performance statistics tasks

About this task

A performance statistics task periodically samples performance and collects performance statistics.

A performance statistics task contains the following parameters:

- **Instance**—Object for which you want to collect performance statistics. The current software version supports only physical interfaces and tunnel interfaces.
- **Performance sampling interval**—Intervals at which the device samples performance data.
- **Performance statistics interval**—Intervals at which the device collects performance statistics, including the peak values, valley values, and average values.

Restrictions and guidelines

Changing the sampling interval or statistics interval for a performance statistics task deletes all performance statistics for the task and recollects the performance statistics.

Changing the statistics interval also changes the sampling interval to the default corresponding to the new statistics interval.

Deleting a performance statistics task terminates the task and deletes all performance statistics for the task.

Procedure

1. Enter system view.

   system-view

2. Enter performance management view.

   performance-management

3. Create a performance statistics task and enter its view.

   statistics task *task-name*

   By default, no performance statistics tasks exist in the system.

   The device supports a maximum of 16 performance statistics task.

4. Bind a performance management instance to the performance statistics task.

   binding instance-type *instance-type* { **all** { **physical** | **tunnel** } | **instance** { *instance-name* } &<1-8> }

   By default, no performance management instances are bound to a performance statistics task.

5. Set the performance statistics interval.

   statistics interval *interval*

   The default statistics interval is 15 minutes.

6. Set the performance sampling interval.

   sample interval *interval*

   The default sampling interval and the possible sampling interval values vary by statistics interval, as shown in the following table:

Statistics interval (minutes)	Default sampling interval (seconds)	Available sampling interval values (seconds)
5	5	5, 10, 15, and 30
10	5	5, 10, 15, 30, and 60
15	5	5, 10, 15, 30, and 60
30	60	5, 10, 15, 30, and 60
60	300	5, 10, 15, 30, 60, and 300
1440	900	300, 600, 900, 1800, and 3600

Enabling performance statistics

1. Enter system view.

   system-view

2. Enter performance management view.

performance-management

3. Enable performance statistics.
   statistics enable
   By default, performance statistics is disabled.

Display and maintenance commands for performance management configuration

Execute **display** commands in any view.

Task	Command
Display the performance statistics for a performance statistics task.	**display performance-management statistics** *task-name* **data** *index* [**instance-type** *instance-type-name* [**measure** *measure-name* |**instance** { *instance-name* } &<1-8>] *]
Display performance statistics task information.	**display performance-management task** [*task-name*]

Performance management configuration examples

Example: Configuring performance management (gRPC)

Network configuration

As shown in Figure 2, configure Device B as a gRPC client to report performance statistics for GigabitEthernet 1/0/1 to the gRPC server. Set the sampling interval to 5 seconds and the statistics interval to 15 minutes.

Figure 2 Network diagram

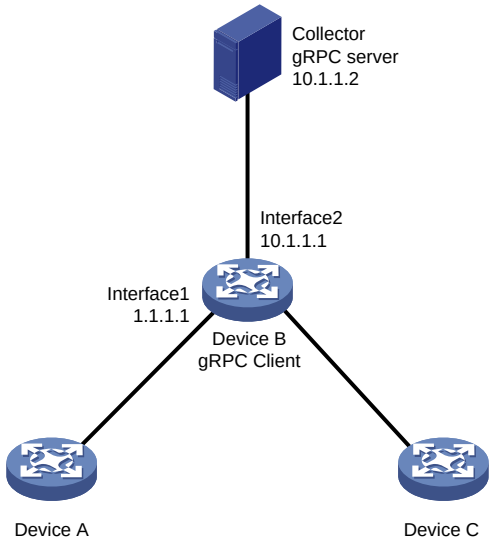


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Assign IP addresses to interfaces and configure routes. Make sure the network connection is available between the device and the server. (Details not shown.)
2. Configure Device B as a gRPC client:

   # Enable the gRPC service.

   ```
   <DeviceB> system-view
   [DeviceB] grpc enable
   ```

   # Create a sensor group named **test**, and add sensor path **ifmgr/devicecapabilities/**.

   ```
   [DeviceB] telemetry
   [DeviceB-telemetry] sensor-group test
   [DeviceB-telemetry-sensor-group-test] sensor path ifmgr/devicecapabilities
   [DeviceB-telemetry-sensor-group-test] quit
   ```

   # Create a destination group named **collector1**. Specify a collector that uses IPv4 address 10.1.1.2 and port number 50050.

   ```
   [DeviceB-telemetry] destination-group collector1
   [DeviceB-telemetry-destination-group-collector1] ipv4-address 10.1.1.2 port 50050
   [DeviceB-telemetry-destination-group-collector1] quit
   ```

   # Configure a subscription named **A** to bind sensor group **test** with destination group **collector1**. Set the sampling interval to 900 seconds.

   ```
   [DeviceB-telemetry] subscription A
   [DeviceB-telemetry-subscription-A] sensor-group test sample-interval 900
   [DeviceB-telemetry-subscription-A] destination-group collector1
   [DeviceB-telemetry-subscription-A] quit
   [DeviceB-telemetry] quit
   ```

3. Configure performance management for interface GigabitEthernet 1/0/1.

   # Enter performance management view.

   ```
   <DeviceB> system-view
   [DeviceB] performance-management
   ```

   # Create a performance statistics task named **task1** and enter its view.

   ```
   [DeviceB-pm] statistics task task1
   ```

   # Bind instance GigabitEthernet 1/0/1 to the task.

   ```
   [DeviceB-pm-task-task1] binding instance-type interface instance gigabitethernet
   1/0/1
   ```

   # Set the statistics interval to 15 minutes.

   ```
   [DeviceB-pm-task-task1] statistics interval 15
   ```

   # Set the sampling interval to 5 seconds.

   ```
   [DeviceB-pm-task-task1] sample interval 5
   [DeviceB-pm-task-task1] quit
   ```

   # Enable performance statistics.

   ```
   [DeviceB-pm] statistics enable
   ```

Verifying the configuration

Verify that the gRPC server receives performance statistics for GigabitEthernet 1/0/1 from Device B at 900-second intervals. (Details not shown.)

Example: Configuring performance management (NETCONF)

Network configuration

As shown in Figure 3, configure Device B as a NETCONF server to report performance statistics for GigabitEthernet 1/0/1 to the NETCONF client. Set the sampling interval to 5 seconds and the statistics interval to 15 minutes.

Figure 3 Network diagram

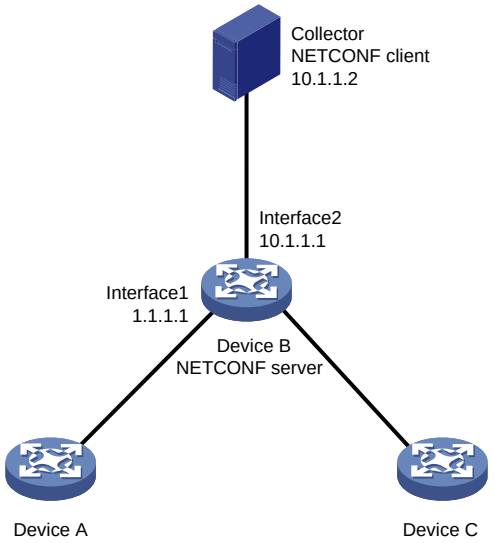


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Assign IP addresses to interfaces and configure routes. Make sure the network connection is available between the device and the server. (Details not shown.)
2. Configure performance management for interface GigabitEthernet 1/0/1.

   # Enter performance management view.

   ```
   <DeviceB> system-view
   [DeviceB] performance-management
   ```

   # Create a performance statistics task named **task1** and enter its view.

   ```
   [DeviceB-pm] statistics task task1
   ```

   # Bind instance GigabitEthernet 1/0/1 to the task.

   ```
   [DeviceB-pm-task-task1] binding instance-type interface instance gigabitethernet 1/0/1
   ```

   # Set the statistics interval to 15 minutes.

```
[DeviceB-pm-task-task1] statistics interval 15
```

Set the sampling interval to 5 seconds.

```
[DeviceB-pm-task-task1] sample interval 5
[DeviceB-pm-task-task1] quit
```

Enable performance statistics.

```
[DeviceB-pm] statistics enable
```

3. Configure the NETCONF settings on Device B. (Details not shown.) For more information, see "Configuring NETCONF."

Verifying the configuration

Verify that the NETCONF client receives performance statistics for GigabitEthernet 1/0/1 from Device B at intervals. (Details not shown.)

Contents

Configuring eMDI 1

About eMDI 1
eMDI implementation 1
Monitored items 2
eMDI collaboration with service modules 6
Restrictions and guidelines 7
Prerequisites 7
Configure eMDI 7
Configuring an eMDI instance 7
Starting an eMDI instance 9
Stopping an eMDI instance 9
Enabling SNMP notifications for eMDI 9
Display and maintenance commands for eMDI 10
eMDI configuration examples 10
Example: Configuring an eMDI instance for a UDP data flow 10

Configuring eMDI

About eMDI

Enhanced Media Delivery Index (eMDI) is a solution to audio and video service quality monitoring and fault locating. It is intended to solve problems caused by packet loss, packet sequence errors, and jitters.

eMDI monitors and analyzes specific TCP or RTP packets on each node of an IP network in real time, providing data for you to quickly locate network faults.

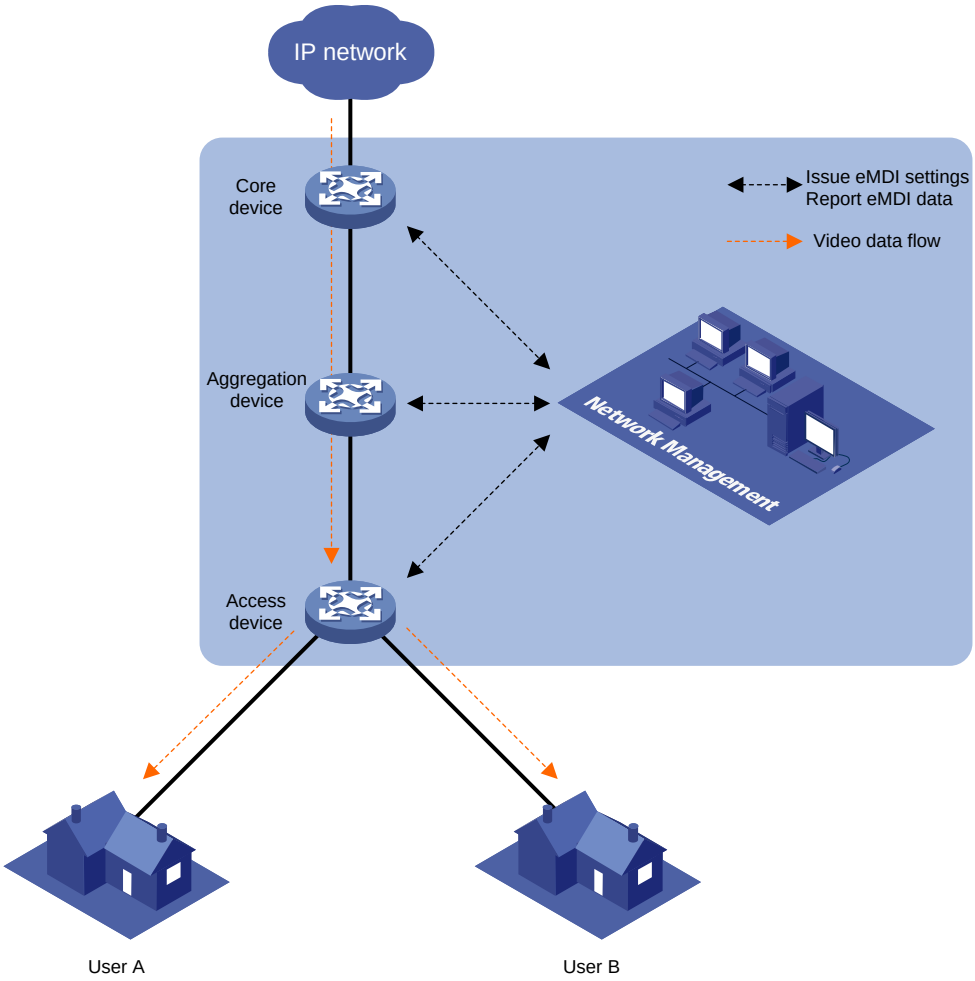
eMDI implementation

The eMDI feature is implemented on the basis of eMDI instances. For an eMDI instance, you can configure parameters, including the target data flow, monitoring period, alarm thresholds, and lifetime. Each eMDI instance monitors one data flow as defined by its parameters. It inspects fields of packets, periodically collects statistics, and reports to the NMS.

Figure 1 shows the eMDI deployment and operating process.

1. Upon service quality deterioration, you can obtain the data flow characteristics from the headers of interested packets, such as the source and destination IP addresses.
2. You deploy eMDI on each device that the data flow traverses by using the NMS or CLI.
3. The devices monitor the data flows, collect statistics, and report monitoring data to the NMS.
4. The NMS collects and analyzes the monitoring data and displays the analysis results graphically.
5. Based on the analysis results, you can quickly locate the network faults.

Figure 1 eMDI network diagram



Monitored items

Table 1 and Table 2 list the monitored items for TCP packets and UDP packets. RTP packets are carried in UDP packets.

Table 1 Monitored items for TCP packets

Item	Description	Calculation formula	Remarks
Rate	Packet input rate, in bps or pps.	• Rate (bps) = Total length of packets received within a monitoring period / Monitoring period • Rate (pps) = Number of packets received within a monitoring period / Monitoring period	This item is intended for unidirectional flows.

Item	Description	Calculation formula	Remarks
Upstream Packet Lost Ratio (UPLR)	Upstream packet loss rate.	UPLR = Number of lost upstream packets / (Number of received packets + Number of lost upstream packets)	If no packets are lost, the sequence number of the current packet plus the length of the current packet is the expected sequence number of the next packet. If the sequence number of the next packet is greater than the expected sequence number, packet loss has occurred on an upstream device. The number of lost packets can be calculated based on the average packet size. This item is intended for unidirectional flows.
Downstream Packet Lost Ratio (DPLR)	Downstream packet loss rate.	• Number of lost downstream packets = Total number of lost packets – Number of lost upstream packets • DPLR = Number of lost downstream packets / Number of received packets	If no packets are lost, the sequence number of the current packet plus the length of the current packet is the expected sequence number of the next packet. If the sequence number of the next packet is smaller than the expected sequence number, the packet is considered to be a retransmitted packet. The number of retransmitted packets is considered to be the total number of lost packets. This item is intended for unidirectional flows.
Downstream average Round Trip Time (DRTT)	Average downstream round trip time, in microseconds.	DRTT = T2 – T1	The timestamp for a received non-retransmitted packet is recorded as T1. The sequence number of the packet plus the length of the packet is the expected sequence number of the next packet. If the sequence number of an ACK packet received from a downstream device is greater than or equal to the expected number, the timestamp of the ACK packet is recorded as T2. DRTT is equal to the average value of the differences between the T2 values and the corresponding T1 values. This item is intended for bidirectional flows.

Item	Description	Calculation formula	Remarks
Upstream average Round Trip Time (URTT)	Average upstream round trip time, in microseconds.	URTT = T2 – T1	T1 indicates the timestamp when the device finds a SYN packet sent by the client for a connection. T2 indicates the timestamp when the device finds a SYN ACK packet returned by the server. Because URTT calculation depends on TCP control packets, it might not be calculated within a monitoring period. This item is intended for bidirectional flows.

Table 2 Monitored items for UDP packets

Item	Description	Calculation formula	Remarks
RTP Lost Rate (RTP-LR)	RTP packet loss rate.	RTP-LR = Number of lost packets / (Number of received packets + Number of lost packets – Number of out-of-order packets)	If the difference between the sequence number of current RTP packet and the greatest sequence number of earlier RTP packets is greater than 1, packet loss has occurred. If the sequence number of the current RTP packet is smaller than the sequence numbers of all earlier RTP packets, the packet is an out-of-order packet. This item is intended for unidirectional flows.
RTP Sequence Error Ratio (RTP-SER)	RTP packet sequence error rate.	RTP-SER = Number of out-of-order packets / (Number of received packets + Number of lost packets – Number of out-of-order packets)	This item is intended for unidirectional flows.
Jitter	RTP packet jitter, in microseconds.	Jitter = T2 – T1	T1 represents the interval between the time when the sender sends the first packet and the time when the sender sends the last packet. T2 represents the interval between the time when the receiver receives the first packet and the time when the receiver receives the last packet. This item is intended for unidirectional flows.

Item	Description	Calculation formula	Remarks
RTP Lost Packet (RTP-LP)	Maximum number of continuously lost RTP packets.	Maximum number of continuously lost RTP packets within a monitoring period.	The network administrators or network management applications can monitor the RTP-LP values to achieve the following purposes: • Determine the network state to promptly respond to network failures and ensure the smooth operation of critical services. For example, if the RTP-LP value increases rapidly, it means that traffic bursts might exist in the network. In this case, you can identify whether these bursts result from an attack and consider rate-limiting the bursts. If the RTP-LP value exceeds the empirical or normal value, it means network congestion or other failures exist. In this case, you can adjust packet retransmission policies, increase bandwidth, or activate backup links to alleviate bandwidth pressure and reduce packet loss effectively. • Quickly locate network failures by monitoring RTP-LP values in sections when service packets travel through various network devices from source to destination. This method helps identify the area where packet loss occurs, narrows down the failure detection range, and increases the troubleshooting efficiency. This item is intended for unidirectional flows.

Item	Description	Calculation formula	Remarks
RTP Effective Loss Factor (RTP-ELF)	RTP effective loss factor for Forward Error Correct (FEC).	RTP-ELF=Total number of packet loss windows in a period / Total number of sliding windows in that period.	The RTP-ELF value is the ratio of the total number of packet loss windows to the total number of sliding windows within a monitoring period. The RTP-ELF value is calculated based on the window size and the predefined packet loss threshold. When the number of service packets sampled within the monitoring period on the device reaches the sliding window size, the first window is generated and the sliding window count becomes 1. Each time a packet is received subsequently, one packet slides back to create a new window, and the sliding window count increments by one. Each window identifies whether the number of lost packets exceeds the predefined threshold. If the threshold is exceeded, the window is considered a packet lost window and the packet loss window count increments by 1. This process continues until the monitoring period ends and the window stops sliding. At this point, the device calculates the RTP-ELF value based on the total number of sliding windows and the total number of packet loss windows during the monitoring period. If the total number of packet loss windows exceeds 0 within a monitoring period, the period is considered to have a transmission failure.

eMDI collaboration with service modules

eMDI instances include static and dynamic types. Static instances are created by using the **instance** command, while dynamic instances are created by service modules through calling eMDI. For example, the Service Quality Analysis (SQA) module is automatically associated with eMDI. SQA treats service traffic as the target data flow and uses the 5-tuple of service packets to create a dynamic eMDI instance. Then, the module calculates the MOS parameter value of the service traffic based on the monitoring results of the dynamic eMDI instance.

The device supports running multiple instances simultaneously. Use the **display emdi instance** command to view information about the current running instances.

Restrictions and guidelines

eMDI is required on each node that monitored data flows traverse.

An eMDI instance can monitor only one data flow. The data flows for different eMDI instances cannot be the same or overlapped.

The following data flows are overlapped:

- **`flow ipv4 tcp source 1.1.1.1 destination 2.2.2.2`**
- **`flow ipv4 tcp source 1.1.1.1 destination 2.2.2.2 destination-port 20`**

For these two flows, if you have configured one data flow for an instance, the system will not allow you to configure the other data flow for a different instance.

The following data flows are not overlapped:

- **`flow ipv4 tcp source 1.1.1.1 destination 2.2.2.2 destination-port 10`**
- **`flow ipv4 tcp source 1.1.1.1 destination 2.2.2.2 destination-port 20`**

The device will detect a configuration conflict and fail to create the later instance if either of the following conditions exists for a dynamic instance created by a service module and a static instance manually created by an administrator:

- The target data flow parameters of the dynamic instance match those of the static instance.
- The target data flow parameters overlap for the dynamic instance and the static instance.

As a best practice, adjust the static instance parameters if dynamic instance creation fails. This ensures successful dynamic instance creation and allows the service module to monitor relevant values through eMDI.

The **`clock-rate`** keyword for the **`flow ipv4 udp`** command does not result in UDP flow overlapping.

A started eMDI instance automatically stops upon a master/subordinate or active/standby switchover or when the eMDI process restarts. You must execute the **`start`** command again to start the instance.

To modify the configuration for an eMDI instance that is already started, stop the instance first by executing the **`stop`** command.

Incomplete eMDI data might cause deviations in the monitoring results. The eMDI data is incomplete in the following situations:

- Not all packets of the monitored data flow pass through the device.
- All packets of the monitored data flow pass through the device, but not all of the packets are forwarded by the same member device.

Prerequisites

Identify the characteristics of the data flows to be monitored so you can specify data flows for eMDI instances.

Configure eMDI

Configuring an eMDI instance

1. Enter system view.

   `system-view`

2. Enable eMDI and enter its view.

   emdi

   By default, eMDI is disabled.

3. Create an eMDI instance and enter its view.

   instance *instance-name*

   By default, no eMDI instance exists.

4. (Optional.) Configure a description for the eMDI instance.

   description *text*

   By default, an eMDI instance does not have a description.

5. Specify the data flow to be monitored by using the eMDI instance.

   Choose one option as needed:

   - Specify a TCP data flow.

     flow ipv4 tcp source *source-ip-address* **destination** *destination-ip-address* [**destination-port** *destination-port-number* | **source-port** *source-port-number* | **vlan** *vlan-id* | **vni** *vxlan-id*] *

   - Specify a UDP data flow.

     flow ipv4 udp source *source-ip-address* **destination** *destination-ip-address* [**destination-port** *destination-port-number* | **source-port** *source-port-number* | **vlan** *vlan-id* | **vni** *vxlan-id* | **pt** *pt-value* | **clock-rate** *clock-rate-value*] *

   By default, no data flow is specified for an eMDI instance.

   Specifying more options defines a more granular data flow and provides more helpful monitoring data and analysis results. For example, to monitor the data from 10.0.0.1 to 10.0.1.1/100, specify the destination port number. If you do not specify the destination port number, the instance monitors the flow that the first received matching packet represents, which might cause deviations on the monitoring results.

6. (Optional.) Set the alarm thresholds for the eMDI instance.

   alarm { **dplr** | **rtp-lr** | **rtp-ser** | **uplr** } **threshold** *threshold-value*

   By default, the alarm thresholds for an eMDI instance are all 100.

   A **TCP data flow** supports only the **dplr** and **uplr keywords and** a **UDP data flow supports only the rtp-lr and rtp-ser keywords.**

7. (Optional.) Set the eMDI alarm logging threshold, that is, the number of consecutive alarms to be suppressed before logging the event.

   alarm suppression times *times-value*

   By default, the eMDI alarm logging threshold is 3.

8. (Optional.) Specify the sliding window size and packet loss threshold for the UDP data flow to be monitored.

   fec { **window** *window-size* | **threshold** *threshold-value* } *

   By default, the sliding window size and packet loss threshold for the UDP data flow is 100 and 5, respectively.

9. (Optional.) Specify the eMDI instance lifetime.

   lifetime { **seconds** *seconds* | **minutes** *minutes* | **hours** *hours* | **days** *days* }

   By default, the eMDI instance lifetime is 1 hour.

10. (Optional.) Specify the monitoring period for the eMDI instance.

    monitor-period *period-value*

By default, the monitoring period for an eMDI instance is 60 seconds.

Starting an eMDI instance

1. Enter system view.
   system-view
2. Enter eMDI view.
   emdi
3. Enter eMDI instance view.
   instance *instance-name*
4. Start the eMDI instance.
   Start

Stopping an eMDI instance

To stop a single eMDI instance:

1. Enter system view.
   system-view
2. Enter eMDI view.
   emdi
3. Enter eMDI instance view.
   instance *instance-name*
4. Stop the eMDI instance.
   Stop

To stop all eMDI instances:

5. Enter system view.
   system-view
6. Enter eMDI view.
   emdi
7. Stop all eMDI instances.
   stop all

Enabling SNMP notifications for eMDI

About this task

With SNMP notifications enabled for the eMDI module, this module can generate alarms to report important events of this module. The generated alarms are sent to the SNMP module of the device. You can specify how the alarms are output through setting the alarm output parameters in SNMP.

For more information about configuring SNMP, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable SNMP notifications for the eMDI module.

snmp-agent trap enable emdi

By default, SNMP notifications are enabled for eMDI.

Display and maintenance commands for eMDI

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display eMDI instance information.	**display emdi instance** [**name** *instance-name* | **id** *instance-id*] [**verbose**]
Display eMDI instance resource information.	**display emdi resource**
Display monitoring statistics for an eMDI instance.	**display emdi statistics** { **instance-name** *instance-name* | **instance-id** *instance-id* } [**number** *number*] [**abnormal** | **verbose**]
Clear eMDI statistics.	**reset emdi statistics** [**instance-name** *instance-name* | **instance-id** *instance-id-value*]

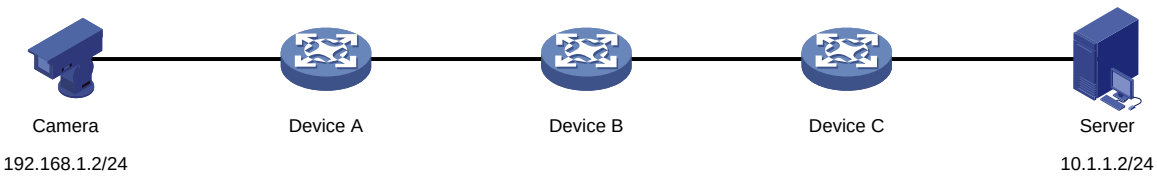
eMDI configuration examples

Example: Configuring an eMDI instance for a UDP data flow

Network configuration

As shown in Figure 2, the camera sends an RTP video data flow to the server through Device A, Device B, and Device C. The server is experiencing an erratic video display problem. Configure eMDI on Device A, Device B, and Device C to locate the faulty link or device.

Figure 2 Network diagram



Prerequisites

Assign IP addresses to the devices. Make sure all devices can reach each other.

Procedure

1. Configure Device A:

   # Enable eMDI.

   ```
   <DeviceA> system-view
   [DeviceA] emdi
   ```

   # Create an eMDI instance named as **test**.

   ```
   [DeviceA-emdi] instance test
   ```

   # Specify the UDP data flow for the instance to monitor. Set the source and destination IP addresses to 192.168.1.2 and 10.1.1.2, respectively.

```
[DeviceA-emdi-instance-test] flow ipv4 udp source 192.168.1.2 destination 10.1.1.2
```

Set the monitoring period to 10 seconds.

```
[DeviceA-emdi-instance-test] monitor-period 10
```

Set the eMDI instance lifetime to 30 minutes.

```
[DeviceA-emdi-instance-test] lifetime minutes 30
```

Start the eMDI instance.

```
[DeviceA-emdi-instance-test] start
```

2. Configure Device B and Device C as Device A is configured. (Details not shown.)

Verifying the configuration

After the eMDI instance runs for a while, display brief monitoring statistics for the eMDI instance on each device to locate the fault.

Display brief monitoring statistics for eMDI instance **test** on Device A.

```
<DeviceA> display emdi statistics instance-name test
Instance name    : test
Instance ID      : 1
Monitoring period: 10 sec
Protocol         : UDP

Unit for RTP-LR, RTP-SER and RTP-ELF is 1/100000
Timestamp            Status   RTP-LR  RTP-SER  Jitter(us)  RTP-LP      RTP-ELF
2019/09/17 16:17:20  Normal   0       0        2560        0           0
2019/09/17 16:17:10  Abnormal 50000   0        2459        1           100000
2019/09/17 16:17:00  Abnormal 12634   33333    5236        3           23356
```

Contents

Configuring SQA 1
About SQA 1
SQA task at a glance 1
Configuring SIP-based SQA 1
Configuring H.323-based SQA 2
Display and maintenance commands for SQA 2
SQA configuration examples 2
Example: Configuring SIP-based SQA 2
Example: Configuring H.323-based SQA 4

Configuring SQA

About SQA

Service quality analysis (SQA) enables the device to identify SIP- or H.323-based multimedia traffic and preferentially forward this traffic to ensure multimedia service quality. Use this feature to decrease multimedia stuttering and improve user experience.

SQA task at a glance

To configure SQA, perform the following tasks:

1. Configure SQA

   Perform one of the following tasks according to the type of the VoIP protocol used.

   - Configuring SIP-based SQA
   - Configuring H.323-based SQA

Configuring SIP-based SQA

Restrictions and guidelines

Make sure SQA uses different port numbers to listen for SIP and H.323 packets.

Procedure

1. Enter system view.

   `system-view`
2. Enter SQA view.

   `sqa`
3. Enable SIP-based SQA.

   `sqa-sip enable`

   By default, SIP-based SQA is disabled.
4. Specify the SIP listening port number.

   `sqa-sip port` *`port-number`*

   By default, the SIP listening port number is 5060.

   Make sure the SIP listening port number on the device is the same as that on the SIP server.
5. (Optional.) Specify an IP address range for SIP-based SQA.

   `sqa-sip filter address` *`start-address end-address`*

   By default, no IP address range is specified for SIP-based SQA. The device performs SQA on all SIP packets.

   After this command is executed, the device performs SQA only on SIP calls in the specified IP address range.

   If you execute this command multiple times, the most recent configuration takes effect.

Configuring H.323-based SQA

Restrictions and guidelines

Make sure SQA uses different port numbers to listen for SIP and H.323 packets.

Procedure

1. Enter system view.
   system-view
2. Enter SQA view.
   sqa
3. Enable H.323-based SQA.
   sqa-h323 enable
   By default, H.323-based SQA is disabled.
4. Specify the H.255 listening port number.
   sqa-h225 port *port-number*
   By default, the H.255 listening port number is 1720.
   Make sure the H.255 listening port number on the device is the same as that on the H.323 server.
5. (Optional.) Specify an IP address range for H.323-based SQA.
   sqa-h323 filter address *start-address end-address*
   By default, no IP address range is specified for H.323-based SQA. The device performs SQA on all H.323 packets.
   After this command is executed, the device performs SQA only on H.323 calls in the specified IP address range.
   If you execute this command multiple times, the most recent configuration takes effect.

Display and maintenance commands for SQA

Execute **display** commands in any view .

Task	Command
Display SIP or H.323 call information.	**display sqa** { **h323** | **sip** } **call** [[**call-id** *call-id*] **verbose**]
Display SIP or H.323 call statistics.	**display sqa** { **h323** | **sip** } **call-statistics**

SQA configuration examples

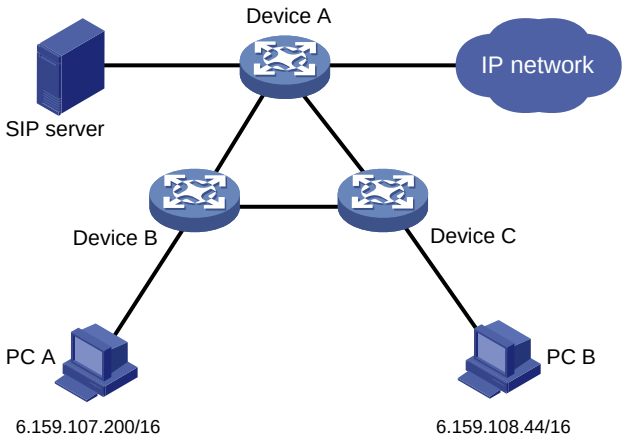
Example: Configuring SIP-based SQA

Network configuration

As shown in Figure 1, the SIP server installed with third-party VoIP server software acts as both the SIP proxy server and SIP registrar to manage SIP UA registration and SIP calls. Host A and Host B installed with third-party client software can place calls to each other.

Configure SIP-based SQA on Device A, Device B, and Device C to optimize multimedia traffic of SIP calls to provide high-quality multimedia services.

Figure 1 Network diagram



Prerequisites

Assign IP addresses to interfaces, and make sure the devices can reach each other.

Procedure

1. Configure the SIP server:

   Install third-party VoIP server software, register UAs, and specify the usernames (phone numbers) and passwords for the clients. (Details not shown.)

2. Configure PC A and PC B:

   Install third-party VoIP client software, specify the IP address of the SIP server, and configure the username (phone number), password, and other parameters on each PC. (Details not shown.)

3. Configure Device A:

   # Enable SIP-based SQA.

   ```
   <DeviceA> system-view
   [DeviceA] sqa
   [DeviceA-sqa] sqa-sip enable
   ```

   # Specify 5066 as the SIP listening port number. (Make sure this SIP listening port number is the same as that on the SIP server.)

   ```
   [DeviceA-sqa] sqa-sip port 5066
   ```

   # Specify an IP address range for SIP-based SQA. (Make sure Host A's IP address is included in the range.)

   ```
   [DeviceA-sqa] sqa-sip filter address 6.159.107.200 6.159.107.210
   [DeviceA-sqa] quit
   ```

4. Configure Device B and Device C:

   # Configure Device B and Device C in the same way as Device A is configured. (Details not shown.)

Verifying the configuration

Place a call from Host A to Host B. (Details not shown.)

Display brief information about all SIP calls on Device A.

```
[DeviceA] display sqa sip call
Caller                Callee                CallId
```

```
6.159.107.207:49172    6.159.108.51:52410     3101326658
6.159.107.203:49172    6.159.108.47:52410     4530332933
6.159.107.208:49172    6.159.108.52:52410     4445702693
6.159.107.206:49172    6.159.108.50:52410     8263542841
6.159.107.201:49172    6.159.108.45:52410     4752123310
6.159.107.200:49172    6.159.108.44:52410     99462146
```

The output shows that Device A has performed SQA on the call that Host A placed to Host B.

Display detailed information about the SIP call that Host A placed to Host B.

```
[DeviceA] display sqa sip call call-id 99462146 verbose
Call ID: 99462146
  Caller information:
    IP: 6.159.107.200      Port: 49172    MAC: a036-9fd4-b5bd
    Tag: 1111@6.159.1.10
  Callee information:
    IP: 6.159.108.44       Port: 52410    MAC: a036-9fd4-b5bc
    Tag: 2222@6.159.1.10
  Type: Audio     VLAN: 1               StartTime: 2019-7-14 15:40:39
  MOS(F): 4       MOS(R): 4
  Forward flow (octets): 4793344        Forward flow (packets): 4681
  Reverse flow (octets): 3810304        Reverse flow (packets): 3721
```

The output shows that the MOS value for both originating and returning traffic is high, which indicates the quality of the audio service is high.

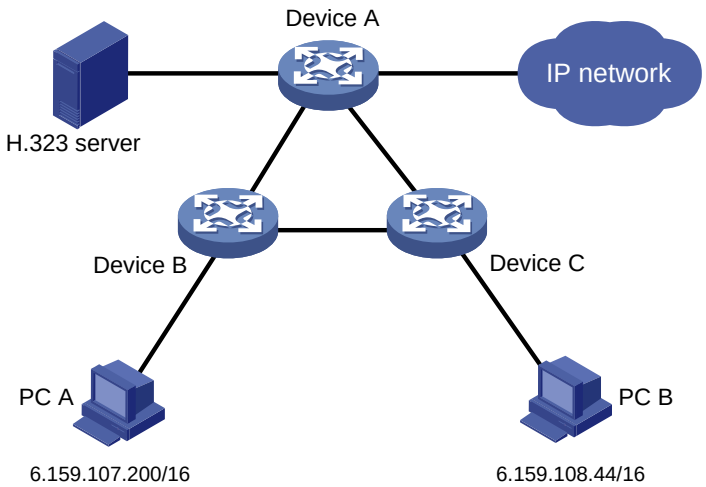
Example: Configuring H.323-based SQA

Network configuration

As shown in Figure 2, the H.323 server installed with third-party VoIP server software acts as both the H.323 proxy server and H.323 registrar to manage H.323 user registration and H.323 calls. Host A and Host B installed with third-party client software can place calls to each other.

Configure H.323-based SQA on Device A, Device B, and Device C to optimize multimedia traffic of H.323 calls to provide high-quality multimedia services.

Figure 2 Network diagram



Prerequisites

Assign IP addresses to interfaces, and make sure the devices can reach each other.

Procedure

1. Configure the H.323 proxy server:

   Install third-party VoIP server software, and specify the usernames (phone numbers) and passwords for the clients. (Details not shown.)

2. Configure PC A and PC B:

   Install third-party VoIP client software, specify the IP address of the H.323 server, and configure the username (phone number), password, and other parameters on each PC. (Details not shown.)

3. Configure Device A:

   # Enable H.323-based SQA.

   ```
   <DeviceA> system-view
   [DeviceA] sqa
   [DeviceA-sqa] sqa-h323 enable
   ```

   # Specify 1720 as the H.323 listening port number. (Make sure this H.323 listening port number is the same as that on the H.323 server.)

   ```
   [DeviceA-sqa] sqa-h225 port 1720
   ```

   # Specify an IP address range for SQA. (Make sure Host A's IP address is included in the range.)

   ```
   [DeviceA-sqa] sqa-h323 filter address 6.159.107.200 6.159.107.210
   [DeviceA-sqa] quit
   [DeviceA] quit
   ```

4. Configure Device B and Device C:

   # Configure Device B and Device C in the same way as Device A is configured. (Details not shown.)

Verifying the configuration

Place a call from Host A to Host B. (Details not shown.)

Display brief information about all H.323 calls on Device A.

```
[DeviceA] display sqa h323 call
Caller                Callee                CallId
6.159.107.207:49172   6.159.108.51:52410    42910c03-e31c-1910-9a63-000c29209aa9
6.159.107.203:49172   6.159.108.47:52410    3e33ecbd-6f1d-1910-8b52-6805ca5d1208
6.159.107.208:49172   6.159.108.52:52410    01a788fe-e21c-1910-8ee7-000c29209aa9
6.159.107.206:49172   6.159.108.50:52410    4e9516ff-e21c-1910-8f98-000c29209aa9
6.159.107.200:49172   6.159.108.44:52410    703b6fff-e21c-1910-9f51-000c29209aa9
```

The output shows that Device A has performed SQA on the call that Host A placed to Host B.

Display detailed information about the H.323 call that Host A placed to Host B.

```
[DeviceA] display sqa h323 call call-id 703b6fff-e21c-1910-9f51-000c29209aa9 verbose
Call ID: 703b6fff-e21c-1910-9f51-000c29209aa9
  Caller information:
    IP: 6.159.107.200      Port: 49172    MAC: a036-9fd4-b5bd
  Callee information:
    IP: 6.159.108.44       Port: 52410    MAC: a036-9fd4-b5bc
  GUID: 703b6fff-e21c-1910-9f61-000c29209aa9
  Protocol: TCP(6)
```

```
Session ID: 2
Type: Audio          Vlan: 1          StartTime: 2019-7-14 15:40:39
MOS(F): 4       MOS(R): 4
Forward flow (octets): 4793344        Forward flow (packets): 4681
Reverse flow (octets): 3810304        Reverse flow (packets): 3721
```

The output shows that the MOS value for both originating and returning traffic is high, which indicates the quality of the audio service is high.

Contents

Configuring TCP connection trace 1

About TCP connection trace 1
Operating mechanism 1
Common reasons for TCP disconnection 2
TCP connection trace tasks at a glance 3
Configuring TCP connection trace rules 3
Setting the maximum number of traced data packets per TCP connection 4
Setting a memory quota for recording TCP connection information 4
Enabling TCP connection trace 5
Specifying a cache mode for TCP connection trace 5
Enabling persistence for TCP connection trace configuration 5
Display and maintenance commands for TCP connection trace 6
TCP connection trace configuration examples 6
Example: Tracing IPv4 TCP connections 6
Example: Tracing IPv6 TCP connections 10

Configuring TCP connection trace

About TCP connection trace

Transmission Control Protocol (TCP) is a reliable connection-oriented network communication protocol. A TCP connection is established before data transmission, and it is released after data transmission is completed. TCP connections are complicated and prone to abnormal terminations due to various reasons.

The TCP connection trace feature traces TCP connections on the device, records information (including packet information) about existing connections, and caches information about abnormally closed connections. The recorded information helps you locate TCP disconnection reasons.

Operating mechanism

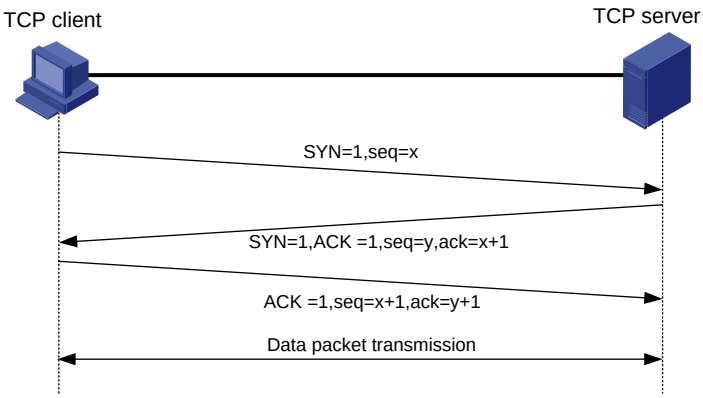
Tracing data packets after the connection establishment

As shown in Figure 1, the TCP client and the TCP server transmit data packets after they establish a TCP connection. The TCP connection trace feature traces data packets and records the connection information. If you enable this feature on both the TCP client and the TCP server, the two devices operate as follows:

- The TCP client records information about each sent data packet, including the send sequence number and the packet send time in the TCP header.
- The TCP server records packet information after it receives the packet.

To display packet information, execute the related **display** commands.

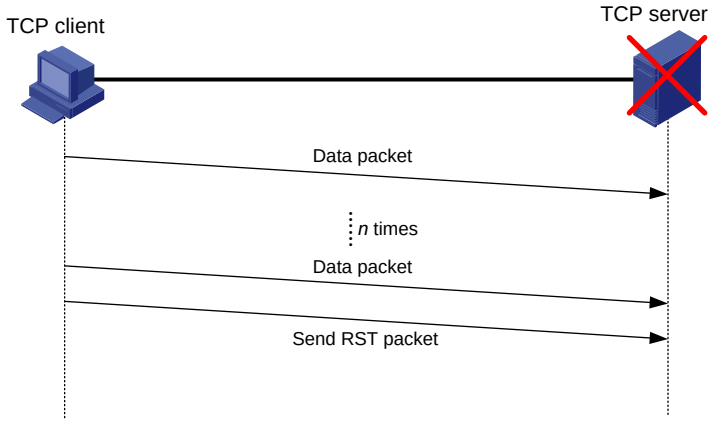
Figure 1 Packet trace after the TCP connection establishment



Caching abnormally closed TCP connections

By default, the TCP connection trace feature records information only about abnormally closed TCP connections. As shown in Figure 2, the TCP connection is unexpectedly closed on the TCP server when the server fails, but the TCP client is not aware of the server failure. With an expectation of an ACK packet from the server, the client sends data packets to the server and starts the retransmission timer. When the timer expires, the TCP client continues to retransmit packets if no ACK packet is received. Without receiving any ACK packet from the server after a number of retransmission attempts, the client sends an RST packet to terminate the connection. In this case, the TCP connection trace feature on the TCP client records the information about retransmitted packets and the disconnection reason in the memory. You can execute related **display** commands to view the information.

Figure 2 Abnormal disconnection of TCP connection



The device does not record the last RST packet sent before the disconnection, but it records the disconnection reason contained in the RST packet.

Common reasons for TCP disconnection

The following are typical reasons for abnormally disconnection of TCP connections:

- **Send reset/Received reset**—The device closes the TCP connection when it receives an RST packet from the peer or sends an RST packet to the peer in one of the following conditions:
  - A TCP server sends an RST packet to a TCP client when the server receives a TCP connection request on a non-open port from the client.
  - The device encounters an exception (for example, application crash) during TCP exchange and sends an RST packet to notify the peer to release TCP connection resources.
  - A TCP server replies an RST packet to a TCP client when the server receives a TCP packet that does not match any existing TCP connection.
  - A TCP client sends an RST packet to the server if the client does not receive any ACK packet after making the maximum number of retransmissions or the retransmission timeout.

  For this disconnection reason, use the following methods to resolve the issue:
  - Verify that the TCP port number is correct on the TCP client.
  - Verify that the TCP server operates normally.
  - Verity that the TCP-based application on the TCP client or TCP server operates correctly.
- **Application abnormal close**—The TCP connection is closed when the application on one end detects abnormal packets and notifies the TCP module to send a FIN packet to the other end.
- **Keepalive timeout**—After a TCP connection is established for a TCP-based application, the device detects the aliveness of the peer as follows:
  - If the application itself supports keepalive probing, the device uses keepalive probing of the application for aliveness detection.
  - If the application does not support keepalive probing, the device uses the keepalive probing of the TCP protocol for aliveness detection.

  When the idle time of the TCP connection exceeds the TCP keepalive duration, the TCP server sends a probe packet without any data to the peer. The server determines the client's status as follows:
  - If the server receives the ACK packet for the probe packet from the client, it determines that the TCP client is working normally.

    - If the server does not receive any ACK packet for multiple probe packets, it determines that the application on the TCP client has been crashed and terminates the connection.
- **Persist timeout**—The device closes the TCP connection if it does not receive any ACK packet from the peer within a period of time.

  If the TCP server does not have any free buffer space, it sends a notification to the TCP client. Upon receiving the notification, the TCP client stops sending packets to the TCP server and starts a timer. When the timer expires, the TCP client sends a space query packet to the TCP server.
    - If the TCP server answers that it has free space, the TCP client sends data packets to the TCP server.
    - If the TCP server answers that it has no free space, the TCP client resets the timer, and sends the query packet again after the timer expires. The client terminates the TCP connection if it does not receive any answer from the server for a long time.
- **Retransmission timeout**—The TCP client expects an ACK packet from the server after sending a data packet to the server. If the client does not receive an ACK packet in a period of time, it retransmits the data packet. If the client does not receive any ACK packet after making 12 retransmission attempts, the client terminates the TCP connection. To troubleshoot this failure, verify that the TCP server operates correctly and the link between the TCP server and client is connected.
- **Backup drop**—After TCP NSR is enabled, a backup TCP connection is established on the subordinate device. When a TCP connection exception occurs on the subordinate device, the TCP connection on the subordinate device is terminated.

TCP connection trace tasks at a glance

To configure TCP connection trace, perform the following tasks:

1. (Optional.) Configuring TCP connection trace rules
2. (Optional.) Setting the maximum number of traced data packets per TCP connection
3. (Optional.) Setting a memory quota for recording TCP connection information
4. Enabling TCP connection trace
5. (Optional.) Specifying a cache mode for TCP connection trace
6. Enabling persistence for TCP connection trace configuration

Configuring TCP connection trace rules

About this task

You can configure TCP connection trace rules to identify TCP connections of important services for monitoring.

Restrictions and guidelines

You can configure TCP connection trace rules only when the TCP connection trace feature is not enabled.

Procedure

To configure TCP connection trace rules, execute the following command in user view:

tcp trace filter { **acl** *acl-number* | **acl6** *ipv6-acl-number* | **only-new-connection** } *

By default, the feature traces all existing TCP connections.

Setting the maximum number of traced data packets per TCP connection

About this task

The device might have lots of TCP connections. If a traced TCP connection has a large number of packets, the device might not have enough memory to trace other TCP connections. To avoid this situation, set the maximum number of packets that can be traced per TCP connection.

This feature counts the number of incoming and outgoing data packets for each TCP connection. When the number of traced packets in a TCP connection exceeds the upper limit, new records of the connection overwrite its old records.

Restrictions and guidelines

You can set the maximum number of packets that can be traced per connection only when the TCP connection trace feature is not enabled. To disable TCP connection trace, execute the **undo tcp trace enable** command.

Procedure

To set the maximum number of data packets that can be traced per TCP connection, execute the following command in user view:

tcp trace max-packet-number *packet-number*

By default, the TCP connection trace feature can trace a maximum of 500 data packets per TCP connection.

Setting a memory quota for recording TCP connection information

About this task

After you enable the TCP connection trace feature, the device starts recording information about packets of specified TCP connections, including the packet sequence number and send time. If a large number of TCP connections exist on the device, it will consume a large amount of memory to record packet information, affecting the performance of other features. To avoid this situation, set a reasonable memory quota.

When the allocated memory is used up, the device does not record packet information about new TCP connections. For an existing connection, new packet records overwrite its old records.

Restrictions and guidelines

You can set the memory quota only when the TCP connection trace feature is not enabled. To disable TCP connection trace, execute the **undo tcp trace enable** command.

Procedure

To set the memory quota for recording TCP connection information, execute the following command in user view:

tcp trace memory-quota *memory-quota*

By default, a 100 MB memory is allocated for recording TCP connection information.

Enabling TCP connection trace

Restrictions and guidelines

You can enable TCP connection trace on either a TCP client or TCP server. Because a TCP server has more TCP connections than a TCP client, this feature on a TCP server consumes more memory resources than on the TCP client. If you want to trace specific TCP connections, enable this feature on the TCP client as a best practice.

Procedure

To enable TCP connection trace, execute the following command in user view:

tcp trace enable

By default, TCP connection trace is disabled.

Specifying a cache mode for TCP connection trace

About this task

By default, the TCP connection trace feature clears information about a connection when the connection is normally closed. To save information about normally closed connections, specify the **all** keyword in the **tcp trace cache-mode** command. The output from the following commands will display information about these connections:

- **display tcp trace cache-connection**
- **display tcp trace cache-packet ip**
- **display tcp trace cache-packet ipv6**

Procedure

To specify a cache mode for TCP connection trace, execute the following commands in user view:

tcp trace cache-mode { **abnormal-close** | **all** }

By default, the TCP connection trace feature records information only about the abnormally closed TCP connections.

Enabling persistence for TCP connection trace configuration

About this task

By default, the TCP connection trace configuration cannot survive a device reboot. To enable the existing TCP connection trace configuration to continue taking effect after the reboot, execute both the **tcp trace persist** and **save** commands.

- The **tcp trace persist** command enables the device to save the TCP connection trace configuration into the database file.
- The **save** command enables the device to save the TCP connection trace configuration in the database file into the .mdb binary configuration file.

When the device reboots, the TCP connection trace configuration is automatically restored from the binary file.

Procedure

To enable persistence for TCP connection trace configuration, execute the following command in user view:

tcp trace persist

By default, persistence for the TCP connection trace configuration is disabled.

Display and maintenance commands for TCP connection trace

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display information about all disconnected TCP connections that have been cached.	**display tcp trace cache-connection** { **ip** | **ipv6** } **slot** *slot-number*
Display packet information about a disconnected IPv4 TCP connection that is traced.	**display tcp trace cache-packet ip source-ip** *source-ipv4-address* **source-port** *source-port-number* **destination-ip** *destination-ipv4-address* **destination-port** *destination-port-number* [**vpn-instance** *vpn-instance-name*] [**verbose** | **wireshark**] [**last** *last-number*] **slot** *slot-number*
Display packet information about a disconnected IPv6 TCP connection that is traced.	**display tcp trace cache-packet ipv6 source-ip** *source-ipv6-address* **source-port** *source-port-number* **destination-ip** *destination-ipv6-address* **destination-port** *destination-port-number* [**vpn-instance** *vpn-instance-name*] [**verbose** | **wireshark**] [**last** *last-number*] **slot** *slot-number*
Display the TCP connection trace configuration and statistics.	**display tcp trace information slot** *slot-number*
Clear the connection and packet information about all disconnected TCP connections.	**reset tcp trace cache**

TCP connection trace configuration examples

Example: Tracing IPv4 TCP connections

Network configuration

As show in Figure 3, you can Telnet to Device B from Device A. Configure TCP connection trace on Device A to trace the TCP connection between Device A and Device B. When the connection is abnormally closed, you can use the trace information to locate disconnection reasons. Enable persistence for TCP connection trace configuration so that the configuration can survive Device A reboot.

Figure 3 Network diagram

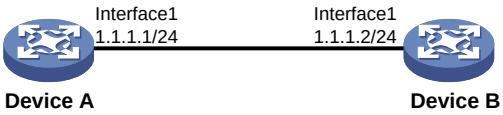


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

- Configure Device B as a Telnet server. For more information, see login management configuration in *Fundamentals Configuration Guide*.
- Assign an IP address to each interface as shown in Figure 3. (Details not shown.)

Procedure

Allow the TCP connection trace feature to trace a maximum of 600 data packets per TCP connection.

```
<DeviceA> tcp trace max-packet-number 600
```

Set the memory quota to 300 MB for recording TCP connection information.

```
<DeviceA> tcp trace memory-quota 300
```

Enable TCP connection trace.

```
<DeviceA> tcp trace enable
```

Enable persistence for the TCP connection trace configuration.

```
<DeviceA> tcp trace persist
```

Verifying the configuration

1. Telnet to Device B from Device A.

   ```
   <DeviceA> telnet 1.1.1.2
   ```

2. Display TCP connection information when you can correctly log in to Device B from Device A through Telnet.

   # Display information about all disconnected IPv4 TCP connections that have been recorded.

   ```
   <DeviceA> display tcp trace cache-connection ip slot 1
    R (Reason):
     PN: Peer normal close
     RR: Received reset
     SR: Sent reset
     AN: Application normal close
     AA: Application abnormal close
     KT: Keepalive timeout
     PT: Persist timeout
     RT: Retransmit timeout
     BD: Backup drop
     AD: Abnormal packet defend
    Local Addr:port      Foreign Addr:port    VPN name    R   Time
   ```

   # Display packet information about disconnected IPv4 TCP connections.

```
<DeviceA> display tcp trace cache-packet ip source-ip 1.1.1.1 source-port 23
destination-ip 1.1.1.2 destination-port 23 verbose slot 1
```

The TCP connection is normal and no output is displayed.

3. Shut down GigabitEthernet 1/0/1 on Device B.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] shutdown
```

After you shut down GigabitEthernet 1/0/1, the Telnet connection is terminated. Display TCP connection information on Device A after a period of time.

Display information about all disconnected IPv4 TCP connections that have been recorded.

```
<DeviceA> display tcp trace cache-connection ip slot 1
 R (Reason):
  PN: Peer normal close
  RR: Received reset
  SR: Sent reset
  AN: Application normal close
  AA: Application abnormal close
  KT: Keepalive timeout
  PT: Persist timeout
  RT: Retransmit timeout
  BD: Backup drop
  AD: Abnormal packet defend
 Local Addr:port      Foreign Addr:port    VPN name    R   Time
 1.1.1.1:9732         1.1.1.2:23           N/A         RT  20:29:08
                                                           July 1 2019
```

Displays information about the last three packets that were traced before the TCP connection was closed.

```
<DeviceA> display tcp trace cache-packet ip source-ip 1.1.1.1 source-port 9732
destination-ip 1.1.1.2 destination-port 23 verbose last 3 slot 1

 MDC ID: 1, Sep 18 15:38:58:503 2019, VPN name: N/A
 Connection: 1.1.1.1:9732 -> 1.1.1.2:23
 Reason: retransmit timeout
 sndmax/localackmax/scale/mss: 73/512/3/1448
 rcvmax/peerackmax/scale/mss: 512/67/3/0
 iss 2236101971, irs 3798739327, in 21 (drop 0), out 23, retrans 14
 connection time: Sep 18 15:29:39:003 2019

 (1) Sep 18 15:33:08:367 2019 RS
 seq 2236102038(67), ack 3798739839(512), data 6, NSR invalid (CLOSED)
 flag PSH ACK, status ESTABLISHED, win: 8145, csum: 0xeab6, ID: 200, ipcsum: 0x0
 iss/sndcc/unack/next/max/wnd: 2236101971/6/67/73/73/4344
 irs/rcvcc/undeliver/next/adv/wnd: 3798739327/0/0/512/65663/65160
 socket state: ISCONNECTED
 socket options: SO_OOBINLINE
 inpcb flags: INP_ANONPORT INP_PROTOCOL_PACKET INP_SYNCPCB
 inpcb extflag: N/A
```

```
 TCP options: TF_ACKNOW TF_REQ_SCALE TF_RCVD_SCALE TF_REQ_TSTMP TF_RCVD_TSTMP
TF_SACK_PERMIT
 recv delayack: 0
 time start/offset/nsroffset start/lastrcv/tsrecent tsrecentage
 4310545007/0/0
 4310754371/15578830/15690748
 4310546126
 rexmt shift/current/seq rtt/srtt/var low/min/best updatetimes
 10/15560/2236102038 0/0/25
 2/30/68 9

 (2) Sep 18 15:33:23:928 2019 RS
 seq 2236102038(67), ack 3798739839(512), data 6, NSR invalid (CLOSED)
 flag PSH ACK, status ESTABLISHED, win: 8145, csum: 0xadee, ID: 201, ipcsum: 0x0
 iss/sndcc/unack/next/max/wnd: 2236101971/6/67/73/73/4344
 irs/rcvcc/undeliver/next/adv/wnd: 3798739327/0/0/512/65663/65160
 socket state: ISCONNECTED
 socket options: SO_OOBINLINE
 inpcb flags: INP_ANONPORT INP_PROTOCOL_PACKET INP_SYNCPCB
 inpcb extflag: N/A
 TCP options: TF_ACKNOW TF_REQ_SCALE TF_RCVD_SCALE TF_REQ_TSTMP TF_RCVD_TSTMP
TF_SACK_PERMIT
 recv delayack: 0
 time start/offset/nsroffset start/lastrcv/tsrecent tsrecentage
 4310545007/0/0
 4310769931/15578830/15690748
 4310546126
 rexmt shift/current/seq rtt/srtt/var low/min/best updatetimes
 11/15560/2236102038 0/0/25
 2/30/68 9

 (3) Sep 18 15:33:39:488 2019 RS
 seq 2236102038(67), ack 3798739839(512), data 6, NSR invalid (CLOSED)
 flag PSH ACK, status ESTABLISHED, win: 8145, csum: 0x7126, ID: 202, ipcsum: 0x0
 iss/sndcc/unack/next/max/wnd: 2236101971/6/67/73/73/4344
 irs/rcvcc/undeliver/next/adv/wnd: 3798739327/0/0/512/65663/65160
 socket state: ISCONNECTED
 socket options: SO_OOBINLINE
 inpcb flags: INP_ANONPORT INP_PROTOCOL_PACKET INP_SYNCPCB
 inpcb extflag: N/A
 TCP options: TF_ACKNOW TF_REQ_SCALE TF_RCVD_SCALE TF_REQ_TSTMP TF_RCVD_TSTMP
TF_SACK_PERMIT
 recv delayack: 0
 time start/offset/nsroffset start/lastrcv/tsrecent tsrecentage
 4310545007/0/0
 4310785491/15578830/15690748
 4310546126
 rexmt shift/current/seq rtt/srtt/var low/min/best updatetimes
 12/15560/2236102038 0/0/25
```

```
2/30/68 9
```

The output shows that:

- The TCP connection between Device A and Device B has been closed.
- The disconnection reason is retransmission timeout. The client determines retransmission timeout because it does not receive any ACK packet from the server.

To resolve the issue, verify that the TCP server operates correctly and the link between the TCP client and server is connected.

Example: Tracing IPv6 TCP connections

Network configuration

As show in Figure 4, you can Telnet to Device B from Device A. Configure TCP connection trace on Device A to trace the IPv6 TCP connection between Device A and Device B. When the connection is abnormally closed, you can use the trace information to locate disconnection reasons. Enable persistence for TCP connection trace configuration so that the configuration can survive Device A reboot.

Figure 4 Network diagram

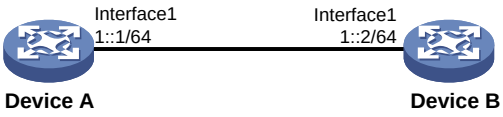


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Prerequisites

- Configure Device B as a Telnet server. For more information, see login management configuration in *Fundamentals Configuration Guide*.
- Assign an IPv6 address to each interface as shown in Figure 4. (Details not shown.)

Procedure

Allow the TCP connection trace feature to trace a maximum of 600 data packets per TCP connection.

```
<DeviceA> tcp trace max-packet-number 600
```

Set the memory quota to 300 MB for recording TCP connection information.

```
<DeviceA> tcp trace memory-quota 300
```

Enable TCP connection trace.

```
<DeviceA> tcp trace enable
```

Enable persistence for the TCP connection trace configuration.

```
<DeviceA> tcp trace persist
```

Verifying the configuration

1. Telnet to Device B from Device A.

   ```
   <DeviceA> telnet ipv6 1::2
   ```

2. Display TCP connection information when you can correctly log in to Device B from Device A through Telnet.

Display information about all disconnected IPv6 TCP connections that have been recorded.

```
<DeviceA> display tcp trace cache-connection ipv6 slot 1
R (Reason):
  PN: Peer normal close
  RR: Received reset
  SR: Sent reset
  AN: Application normal close
  AA: Application abnormal close
  KT: Keepalive timeout
  PT: Persist timeout
  RT: Retransmit timeout
  BD: Backup drop
  AD: Abnormal packet defend
 LAddr->port           FAddr->port           VPN name   R   Time
```

Display packet information about disconnected IPv6 TCP connections.

```
<DeviceA> display tcp trace cache-packet ipv6 source-ip 1::1 source-port 47745
destination-ip 1::2 destination-port 23 verbose slot 1
```

The TCP connection is normal and no output is displayed.

3. Shut down GigabitEthernet 1/0/1 on Device B.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] shutdown
```

After you shut down GigabitEthernet 1/0/1, the Telnet connection is terminated. Display TCP connection information on Device A after a period of time.

Display information about all disconnected IPv6 TCP connections that have been recorded.

```
<DeviceA> display tcp trace cache-connection ipv6 slot 1
 R (Reason):
  PN: Peer normal close
  RR: Received reset
  SR: Sent reset
  AN: Application normal close
  AA: Application abnormal close
  KT: Keepalive timeout
  PT: Persist timeout
  RT: Retransmit timeout
  BD: Backup drop
  AD: Abnormal packet defend
 LAddr->port           FAddr->port           VPN name   R   Time
 1::1->47745           1::2->23              N/A        RT  20:29:08
                                                           July 1 2019
```

Displays information about the last three packets that were traced before the IPv6 TCP connection was closed.

```
<DeviceA> display tcp trace cache-packet ipv6 source-ip 1::1 source-port 47745
destination-ip 1::2 destination-port 23 verbose last 3 slot 1

 MDC ID: 1, Sep 18 16:20:20:091 2019, VPN name: N/A
 Connection: 1::1->47745 --> 1::2->23
```

```
 Reason: retransmit timeout
 sndmax/localackmax/scale/mss: 77/460/3/1428
 rcvmax/peerackmax/scale/mss: 460/59/3/0
 iss 2284338403, irs 1251751781, in 9 (drop 0), out 11, retrans 20
 connection time: Sep 18 16:15:49:827 2019
 (1) Sep 18 16:18:50:817 2019 RS
 seq 2284338462(59), ack 1251752241(460), data 18, NSR invalid (CLOSED)
 flag PSH ACK, status ESTABLISHED, win: 8211, csum: 0x6edb
 iss/sndcc/unack/next/max/wnd: 2284338403/18/59/77/77/4280
 irs/rcvcc/undeliver/next/adv/wnd: 1251751781/0/0/460/66139/65688
 socket state: ISCONNECTED
 socket options: SO_OOBINLINE
 inpcb flags: INP_ANONPORT INP_PROTOCOL_PACKET IN6P_IPV6_V6ONLY INP_SYNCPCB
 inpcb extflag: N/A
 TCP options: TF_ACKNOW TF_REQ_SCALE TF_RCVD_SCALE TF_REQ_TSTMP TF_RCVD_TSTMP
TF_SACK_PERMIT TF_APP_SEND
 recv delayack: 0
 time start/offset/nsroffset start/lastrcv/tsrecent tsrecentage
 4313315830/0/0
 4313496820/18348628/18460545
 4313315924
 rexmt shift/current/seq rtt/srtt/var low/min/best updatetimes
 10/15560/2284338462 18529253/0/15
 1/30/64 5

 (2) Sep 18 16:18:58:881 2019 RS
 seq 2284338462(59), ack 1251752241(460), data 18, NSR invalid (CLOSED)
 flag PSH ACK, status ESTABLISHED, win: 8211, csum: 0x4f5a
 iss/sndcc/unack/next/max/wnd: 2284338403/18/59/77/77/4280
 irs/rcvcc/undeliver/next/adv/wnd: 1251751781/0/0/460/66139/65688
 socket state: ISCONNECTED
 socket options: SO_OOBINLINE
 inpcb flags: INP_ANONPORT INP_PROTOCOL_PACKET IN6P_IPV6_V6ONLY INP_SYNCPCB
 inpcb extflag: N/A
 TCP options: TF_ACKNOW TF_REQ_SCALE TF_RCVD_SCALE TF_REQ_TSTMP TF_RCVD_TSTMP
TF_SACK_PERMIT
 recv delayack: 0
 time start/offset/nsroffset start/lastrcv/tsrecent tsrecentage
 4313315830/0/0
 4313504885/18348628/18460545
 4313315924
 rexmt shift/current/seq rtt/srtt/var low/min/best updatetimes
 11/15560/2284338462 0/0/15
 1/30/64 5

 (3) Sep 18 16:19:14:441 2019 RS
 seq 2284338462(59), ack 1251752241(460), data 18, NSR invalid (CLOSED)
 flag PSH ACK, status ESTABLISHED, win: 8211, csum: 0x1292
 iss/sndcc/unack/next/max/wnd: 2284338403/18/59/77/77/4280
```

```
 irs/rcvcc/undeliver/next/adv/wnd: 1251751781/0/0/460/66139/65688
 socket state: ISCONNECTED
 socket options: SO_OOBINLINE
 inpcb flags: INP_ANONPORT INP_PROTOCOL_PACKET IN6P_IPV6_V6ONLY INP_SYNCPCB
 inpcb extflag: N/A
 TCP options: TF_ACKNOW TF_REQ_SCALE TF_RCVD_SCALE TF_REQ_TSTMP TF_RCVD_TSTMP
TF_SACK_PERMIT
 recv delayack: 0
 time start/offset/nsroffset start/lastrcv/tsrecent tsrecentage
 4313315830/0/0
 4313520445/18348628/18460545
 4313315924
 rexmt shift/current/seq rtt/srtt/var low/min/best updatetimes
 12/15560/2284338462 0/0/15
 1/30/64 5
```

The output shows that:

- The TCP connection between Device A and Device B has been closed.
- The disconnection reason is retransmission timeout. The client determines retransmission timeout because it does not receive any ACK packet from the server.

To resolve the issue, verify that the TCP server operates correctly and the link between the TCP client and server is connected.

Contents

Configuring SmartMC 1

About SmartMC 1
SmartMC network framework 1
SmartMC network establishment 1
SmartMC features 2
Restrictions and guidelines: SmartMC configuration 5
SmartMC tasks at a glance 5
Prerequisites for SmartMC 6
Enabling SmartMC 7
Specifying a file server 7
Configuring an outgoing interface for the SmartMC network 8
Enabling automatic Ethernet link aggregation 8
Modifying the password of the default user for members 9
Creating a SmartMC group 9
Creating a VLAN for members 10
Deploying a batch file to members 10
Configuring a batch file for ports connecting APs or IP phones 11
Backing up configuration files 11
Configuring resource monitoring 12
Upgrading the startup software and configuration file on members 13
About upgrading the startup software and configuration file on members 13
Restrictions and guidelines for startup software and configuration file upgrade 13
Prerequisites 13
Upgrading the startup software and configuration file on members 13
Upgrading the startup software and configuration file on all members in SmartMC groups 14
Managing the network topology 16
Refreshing the network topology 16
Saving the network topology 16
Replacing faulty members 16
Display and maintenance commands for SmartMC 17
SmartMC configuration examples 18
Example: Configuring SmartMC 18

Configuring SmartMC

About SmartMC

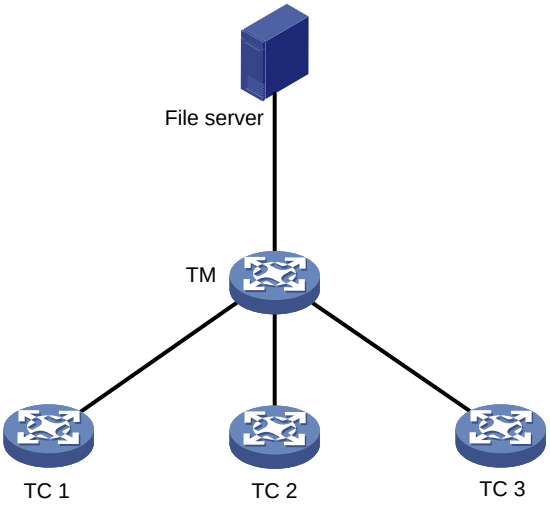
Smart Management Center (SmartMC) centrally manages and maintains dispersed network devices at network edges. In a SmartMC network, only one device acts as the commander and the remaining devices all act as members. SmartMC provides the following features for you to manage the members from the commander:

- Configuration file backup and download.
- Software upgrade.
- Configuration deployment.
- Faulty member replacement.

SmartMC network framework

Figure 1 shows the basic framework of a SmartMC network.

Figure 1 SmartMC network framework



The SmartMC network contains the following elements:

- **Commander**—Also called topology master (TM), which manages all members in the SmartMC network.
- **Member**—Also called topology client (TC), which is managed by the commander.
- **File server**—Stores startup software images and configuration files for the commander and members.

SmartMC network establishment

A SmartMC network can be established automatically. During automatic establishment, the commander obtains member information through NETCONF sessions to form the network topology. The member information includes port information, LLDP neighbor information, STP information, device type, and software version.

Automatic SmartMC network establishment

The commander and members use the following procedure to establish a SmartMC network:

1. After SmartMC is enabled, the commander broadcasts a SmartMC packet at an interval of 15 seconds to detect members in the network. The SmartMC packet contains information of the commander, such as its bridge MAC address and the IP address of VLAN-interface 1.
2. When a member receives the packet, it records the commander information, and returns a response packet to the commander. The response packet contains information of the member, such as its bridge MAC address and the IP address of VLAN-interface 1.
3. When the commander receives the response packet, it initiates a NETCONF session to the member with the default username **admin** and the default password **admin** and sets up TCP keepalive. The commander then obtains detailed information about the member through the session, including port information, LLDP neighbor information, STP information, device type, and software version.
4. The commander notifies the member of online success, and adds the member to the SmartMC network.
5. Based on the LLDP neighbor information obtained from all members, the commander forms a SmartMC topology.

After the SmartMC network is established, the commander and members check for the existence of each other by exchanging SmartMC packets.

After the successful establishment of the SmartMC network, the commander becomes aware of members through NETCONF session status, while members become aware of each other through SmartMC broadcast messages sent by the commander:

- When a member receives a SmartMC broadcast packet from the commander, it compares the bridge MAC address in the packet with the recorded bridge MAC address. If the two bridge MAC addresses are the same and the device with the MAC address is offline, the member returns a response packet to the commander. If the member does not receive a broadcast packet from the commander within the time limit, the member determines that the commander does not exist in the network anymore. Then, the member clears the commander information. The time limit is a random value in the range of 60 to 120 seconds.
- When the commander receives a response message from a member, it establishes a NETCONF session with the member and sets up TCP keepalive. If no data packet is exchanged within 40 seconds, TCP keepalive detection will be triggered. The device initiates a detection every 5 seconds. If the detection fails 4 times in a row, the member is considered offline, and the commander sets it to offline status. This means that if the commander does not receive any packets from the member within 60 seconds, it sets the member to offline status.

SmartMC features

Bulk configuration deployment for members

This feature allows you to deploy multiple command lines to members from the commander, eliminating the need to log in to members and configure the command one by one.

The procedure for bulk configuration deployment is as follows:

1. The commander acts as a Telnet client and establishes Telnet connections to the members.
2. The commander deploys a batch file to the members through NETCONF sessions. The batch file is created on the commander and contains command lines to be deployed.
3. The members run the command lines in the file.

Bulk configuration deployment for ports connecting APs and IP phones

With batch file deployment enabled, the commander automatically deploys configurations in the specified batch file to a port connecting an AP or IP phone, simplifying access port configuration.

When the commander first detects the association of an AP or IP phone on a port through LLDP, it deploys the command lines in the specified batch file to the port. If no batch file is specified for the device type, the configurations on the port remain unchanged.

If the AP or IP phone disconnects from the port, the configurations on the port remain. When a new device comes online from the port, configurations used by the port depend on the new device type.

- If the new device is an AP or IP phone and has the same type as the disconnected device, the configurations on the port remain unchanged.
- If the new device is an AP or IP phone but has a different type as the disconnected device, the commander deploys the command lines in the specified batch file to the port. If no batch file is specified for the device type, the configurations on the port remain unchanged.
- If the new device is neither an AP nor an IP phone, the configurations on the port remain unchanged.

To disable the commander from deploying a batch file to ports, remove the specified batch file or execute the **`undo smartmc batch-file-apply enable`** command to disable batch file deployment.

Configuration file backup

You can use the following methods to back up the next-startup configuration file on the commander and members:

- **Automatic backup**—Enable this feature for the commander and all members in the network to immediately perform a backup. After that, the commander and members back up the configuration file at a user-specified interval.
- **Manual backup**—Manually trigger a backup on the commander or the specified members or SmartMC groups.

To back up the configuration file on a member, the commander instructs the member by unicasting a SmartMC packet. When a member receives the packet, it saves the running configuration to the next-startup configuration file and uploads the file to the file server.

Startup software and configuration file upgrade

This feature enables users to upgrade startup software and the configuration file of member devices from the commander.

Before upgrade, you must upload the upgrade files from the commander to the file server and specify the upgrade files on the file server for the members to download.

The procedure for startup software and configuration file upgrade is as follows:

1. The commander instructs the members (or SmartMC group) to download the upgrade files from the file server.
2. The members download the upgrade files from the file server.
3. The members upgrade the startup software and configuration file as follows:
   - **Startup software upgrade**—Performs an ISSU with the upgrade startup software files. The members might be restarted during the upgrade process.
   - **Configuration file upgrade**—Replaces the current configuration file with the upgrade configuration file. The members will not be restarted during the upgrade process.

Faulty member replacement

You can use the following methods to replace a faulty member:

- **Automatic replacement**—Enables the commander to record the positions of all members in the topology for replacement. When the commander discovers that the new member has physically replaced the faulty member, it compares the new member with the faulty one. The commander performs a replacement if the following requirements are met:
  - The new member is deployed at the same topological position as the faulty one.
  - The models of the new member and faulty member are the same.

The commander then instructs the new member to download the configuration file of the faulty member from the file server. After downloading the configuration file, the new member runs the configuration file to complete the replacement.

- **Manual replacement**—After the faulty member is physically replaced, you manually trigger a configuration replacement. The new member will download the configuration file of the faulty member from the file server and run the file to complete the replacement.

Outgoing interface for a SmartMC network

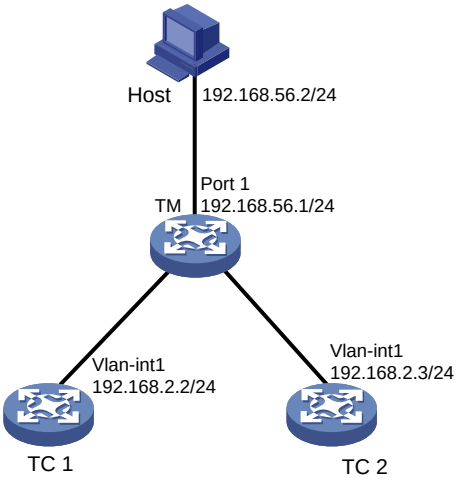
The outgoing interface feature allows hosts connecting to an outgoing interface to access all the members in a SmartMC network. You can configure multiple outgoing interfaces for a SmartMC network.

As shown in Figure 2, the host is connected to port 1 on the TM and TC 1 and TC 2 are in a different network segment than the host. The host can access the Web interface of the TM but cannot access the Web interface of any member.

If port 1 on the TM is configured as the outgoing interface, the system mirrors the IP address of each member to a new address. The new address contains the IP address of the outgoing interface and the port number assigned by the commander to the member in the format of *IP address*:*Port number*. This enables the host to access the Web interfaces of members from the Web interface of the TM.

To access the Web interface of a member, enter the Web interface of the commander, and click **Visibility** from the navigation pane. Then, click the **Topology** tab, select the target member, and click **Login to Web interface**.

Figure 2 SmartMC network



NOTE:

The S5590-HI, S5590-EI, and S5500V3-HI switch series do not support management from the Web interface.

Automatic link aggregation

Automatic link aggregation automatically bundles multiple physical Ethernet links between two members into one logical link, called an aggregate link. This feature provides increased link bandwidth and improved link reliability.

NOTE:

- Automatic link aggregation cannot be performed between the commander and a member, or between a member and a device outside the SmartMC network. You can aggregate the links between the commander and a member manually. For more information about manual link

aggregation, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*.

- If a member enabled with automatic link aggregation joins a SmartMC network whose commander is disabled with the aggregation feature, the feature will be disabled for the member as well. This might affect service traffic forwarding on the member.

VLAN creation for members

To simplify configuration and management, you can create a VLAN for members. Then, all access ports on a member that are not connected to other members or the commander are assigned to the VLAN.

If a member has access ports that are connected to offline devices, you must remove the offline devices before creating a VLAN for the member.

The VLAN creation fails for a member if one or more access ports cannot be assigned to the VLAN. If the VLAN creation fails, the VLAN memberships for the access ports are restored to the state before the VLAN was created.

The failure to create a VLAN for a member does not affect the VLAN creation for other members.

Resource monitoring

Resource monitoring allows you to view CPU usage, memory usage, temperature information, and packet dropping information of commanders and members on the commander. Packet dropping monitoring monitors packet dropping on members and on interfaces.

You can view the usage and temperature information on the commander.

Restrictions and guidelines: SmartMC configuration

Only users with the network-admin role can configure SmartMC. For more information about user roles, see RBAC in *Fundamentals Configuration Guide*.

Devices exchange SmartMC protocol packets and establish and maintain the SmartMC network only in VLAN 1. To ensure the proper functioning of SmartMC, make sure the interfaces connecting member devices within the SmartMC network allow VLAN 1 packets to pass through.

You need to enable SmartMC on both the commander and members and perform all the other SmartMC tasks only on the commander.

SmartMC tasks at a glance

To configure SmartMC, perform the following tasks:

1. Enabling SmartMC
2. Specifying a file server

   This task is required for configuring automatic configuration file backup, replacing faulty members, and upgrading the startup software and configuration file on members.
3. (Optional.) Configuring an outgoing interface for the SmartMC network
4. (Optional.) Enabling automatic Ethernet link aggregation
5. (Optional.) Modifying the password of the default user for members
6. Creating a SmartMC group

   This task is required for upgrading the startup software and configuration file on members and deploying a batch file to a SmartMC group.
7. (Optional.) Deploying and managing configuration

- Creating a VLAN for members
- Deploying a batch file to members
- Configuring a batch file for ports connecting APs or IP phones
- Backing up configuration files

8. (Optional.) Monitoring and maintaining the SmartMC network
   - Configuring resource monitoring
   - Upgrading the startup software and configuration file on members
   - Managing the network topology
   - Replacing faulty members

Prerequisites for SmartMC

Before you configure SmartMC, perform the following tasks on the commander and members:

- Enable the Telnet service, and configure scheme authentication for VTY user lines. For information about Telnet service and VTY user lines, see CLI login configuration in *Fundamentals Configuration Guide*.
- Configure a local user.
  - Specify the username and password.
    - On the commander, make sure the username and password are the same as the username and password configured by using the **smartmc tm username** *username* **password** { **cipher** | **simple** } *string* **enable** command.
    - On a member, set both the username and password to **admin**, and execute the **password-control length 4**, **password-control composition type-number 1 type-length 1**, and **undo password-control complexity user-name check** commands to lower the password complexity requirements.

      This is because SmartMC requires that the commander use username **admin** and password **admin** to communicate with members, which does not meet the default password complexity requirements. For more information about these commands, see password control commands in *Security Command Reference*.

      After the SmartMC network is established, you can increase the password complexity requirements and use the **smartmc tc password** command to modify the username and password.
  - Specify the Telnet, SSH, HTTP, and HTTPS services for the user.
  - Set the RBAC role of the local user to network-admin.

  For information about local users, see AAA configuration in *Security Configuration Guide*. For information about user roles, see RBAC configuration in *Fundamentals Configuration Guide*.
- Enable NETCONF over SOAP over HTTP. For information about NETCONF over SOAP, see NETCONF configuration in *Network Management and Monitoring Configuration Guide*.
- Enable NETCONF over SSH. For more information about NETCONF over SSH, see "Configuring NETCONF."
- Configure public key management for SSH authentication. Execute the **ssh user** *admin* **service-type netconf authentication-type publickey assign publickey** *-smartmc-publickey-* command on the member to configure the authentication method for SSH users. Digital signature is supported. The username must be **admin** for an SSH user and the public key name must be **-smartmc-publickey-** for the SSH client. For more information about public key management, see *Security Configuration Guide*.
- Enable LLDP globally. For information about LLDP, see *Layer 2—LAN Switching Configuration Guide*.

Enabling SmartMC

About this task

Enable this feature on both the commander and members to enable management of members from the commander.

Restrictions and guidelines

A SmartMC network must have one and only one commander.

If you change the role of the commander to member or disable SmartMC on the commander, all SmartMC settings in its running configuration will be cleared.

SmartMC fails to be enabled if ACL resources are insufficient. If ACL resources are insufficient, use the **undo acl** command to delete unnecessary ACLs and then enable SmartMC. You can execute the **display acl** command to view ACL configuration and match statistics. For more information about ACLs, see *ACL and QoS Configuration Guide*.

SmartMC fails to be enabled if ports 80 and 443 have been used.

If you execute the **smartmc enable** command multiple times, the most recent configuration takes effect. You can execute the command to change the device role or the password.

Procedure

1. Enter system view.

   system-view

2. Enable SmartMC and set the device role.

   smartmc { **tc** | **tm username** *username* **password** { **cipher** | **simple** } *string* } **enable**

   By default, SmartMC is disabled.

Specifying a file server

About this task

In a SmartMC network, a file server is used to store the following files:

- Upgrade startup software files and upgrade configuration file for members.
- Backup configuration files of the commander and members.

The system supports using an FTP or SFTP server as the file server. For information about FTP servers, see configuring FTP in *Fundamentals Configuration Guide*. For information about SFTP servers, see configuring SSH in *Security Configuration Guide*.

Restrictions and guidelines

You can use the following methods to specify a file server:

- Specify the IP address of a file server.
- Specify the IP address of the commander. The commander will act as a file server.

To configure the commander to act as a file server, make sure the commander has enough storage space for storing the files required by members.

To use an independent file server, connect the file server to the commander instead of the members as a best practice. The file server uses VLAN 1 to communicate with the SmartMC network. If you connect the file server to members, creating a VLAN for members will assign member interfaces connecting to the file server to the created VLAN, causing file server disconnection. For more information about member VLAN creation, see "Creating a VLAN for members."

Procedure

1. Enter system view.
   system-view
2. Specify a file server.
   smartmc { **ftp-server** | **sftp-server** } { *ipv4-address* | **ipv6** *ipv6-address* } [**port** *port*] [**vpn-instance** *vpn-instance-name*] [**directory** *directory*] **username** *username* **password** { **cipher** | **simple** } *string*

   By default, no file server is specified.

Configuring an outgoing interface for the SmartMC network

Restrictions and guidelines

VLAN-interface 1 cannot be used as an outgoing interface, because the SmartMC network is established in VLAN 1.

A member port of a Layer 3 aggregation group cannot be used as an outgoing interface for SmartMC, and vice versa. If a configured outgoing interface is a member port of a Layer 3 aggregation group, the outgoing interface does not take effect. For more information about Layer 3 aggregation groups, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*.

If you specify a Layer 3 Ethernet interface as the outgoing interface, the system issues the command only to VLAN-interface 1. If you remove the outgoing interface configuration, the **service slot** *slot-number* configuration remains, and you can modify the configuration manually as needed.

Procedure

1. Enter system view.
   system-view
2. Enter VLAN interface view or Layer 3 Ethernet interface view.
   - Enter VLAN interface view.
     interface vlan *interface-number*
   - Enter Layer 3 Ethernet interface view.
     interface *interface-type interface-number*
3. Configure the interface as an outgoing interface.
   smartmc outbound

   By default, no interface is used as an outgoing interface.

Enabling automatic Ethernet link aggregation

Restrictions and guidelines

Enabling or disabling automatic link aggregation might cause network flapping, and the members might go offline for a short period of time.

Procedure

1. Enter system view.
   system-view
2. Enable automatic Ethernet link aggregation.
   smartmc auto-link-aggregation enable

By default, automatic Ethernet link aggregation is disabled.

Modifying the password of the default user for members

About this task

During SmartMC network establishment, the commander uses the default username and password to establish NETCONF sessions to members automatically added to the network. The default username and password of the members for NETCONF session establishment are **admin** and **admin**.

To enhance security, you can perform this task to change the password for the default user **admin** of the members after the commander adds the members to the network.

Restrictions and guidelines

Do not modify the password for members that are manually added to the SmartMC network. If you modify the password for a manually added member, you will not be able to manage that member from the commander.

You can use the **display smartmc tc verbose** command to identify the method used to add the members.

Procedure

1. Enter system view.
   system-view
2. Modify the password of the default user for members.
   smartmc tc password [**cipher**] *string*

Creating a SmartMC group

About this task

This feature allows you to create a SmartMC group on the commander and add members to the group. When you perform the following operations, you can specify a SmartMC group to apply the operations to all members in the group:

- Startup software upgrade.
- Configuration file upgrade.
- Configuration deployment.

Procedure

1. Enter system view.
   system-view
2. Create a SmartMC group and enter its view.
   smartmc group *group-name*
3. (Optional.) Display predefined device types.
   match device-type ?
4. Set a match criterion.
   match { **device-type** *device-type* | **ip-address** *ip-address* { *ip-mask-length* | *ip-mask* } | **mac-address** *mac-address mac-mask-length* }
   By default, no match criterion is set.

5. If the device type of the members is not predefined on the commander, perform the following tasks to manually define the device type on the commander:
   a. Return to system view.
      quit
   b. Define a device type on the commander.
      smartmc tc sysoid *sysoid* **device-type** *device-type*
      To obtain the SYSOID of a member, execute the **display smartmc tc verbose** command.
      You cannot define a predefined member type as another type.

Creating a VLAN for members

Restrictions and guidelines

If you perform this task multiple times to create a VLAN for members, the most recent configuration takes effect.

Procedure

1. Enter system view.
   system-view
2. Creating a VLAN for members and assign access ports on the members to the VLAN.
   smartmc vlan *vlan-id* { **group** *group-name-list* | **tc** *tc-id-list* }

Deploying a batch file to members

About this task

Perform this task to distribute the configurations from a specified configuration file to multiple member devices at once, in order to reduce the workload of configuring each device individually.

Restrictions and guidelines

Make sure the configuration file does not contain the **undo smartmc enable** and **undo telnet server enable** commands as these two commands might cause the configuration link to disconnect, leading to the failure of the remaining configuration distribution.

Procedure

1. Execute the following command in user view to create a batch file and edit the command lines to be deployed to members.
   create batch-file *cmd-filename*
   Each command occupies a line in the batch file. When you finish editing, enter a percent sign (%) to return to user view.
   Make sure the command lines that you enter are correct because the system does not verify whether the command lines are correct.
2. Enter system view.
   system-view
3. Deploy the batch file to a list of members or SmartMC groups.
   smartmc batch-file *cmd-filename* **deploy** { **group** *group-name-list* | **tc** *tc-id-list* }

Configuring a batch file for ports connecting APs or IP phones

Restrictions and guidelines

All commands in the batch file must be commands used in interface view.

The size of the batch file cannot exceed 8190 characters.

Make sure the file name is correct when specifying the batch file because the system does not verify whether the file name is correct. After specifying the batch file, do not delete the file or rename the file.

Procedure

1. Execute the following command in user view to create a batch file and edit the command lines to be deployed to members.

   create batch-file *cmd-filename*

   Each command occupies a line in the batch file. When you finish editing, enter a percent sign (%) to return to user view.

   Make sure the command lines that you enter are correct because the system does not verify whether the command lines are correct.
2. Enter system view.

   system-view
3. Specify the batch file for ports connecting APs or IP phones.

   smartmc batch-file *batch-file-name* **apply** { **ap** | **phone** }
4. (Optional.) Disable batch file deployment.

   undo smartmc batch-file-apply enable

   By default, batch file deployment is enabled.

Backing up configuration files

About this task

Perform this task to back up the configuration file of the commander or the specified members. Configuration files automatically backed up to the file server are named in the format of *device_bridge_MAC_address*_**backup.cfg**.

Restrictions and guidelines

When you change the commander in the SmartMC network, make sure the backup configuration file of the original commander on the file server is deleted. If the file still exists, the new commander might download the file and run the settings. This will cause a conflict in the network.

The maximum number of members that can perform automatic configuration at the same time is limited by the performance of the file server. If automatic configuration backup fails, set the maximum number of members to a smaller value.

Prerequisites

Before performing this task, you must specify a file server (see "Specifying a file server").

Procedure

1. Enter system view.

   system-view

2. Set the maximum number of members that can perform configuration file backup at the same time.

   smartmc backup configuration max-number *max-number*

   By default, a maximum of five members can perform automatic configuration backup at the same time.

3. Back up configuration files.

   Choose one option as needed:

   - Enable automatic configuration file backup and set the backup interval.

     smartmc backup startup-configuration interval *interval-time*

     By default, automatic configuration file backup is disabled.

   - Manually back up the configuration file on the commander or the specified members.

     smartmc backup configuration { **group** *group-name-list* | **tc** [*tc-id-list*] }

     TC ID 0 represents the commander.

Configuring resource monitoring

Restrictions and guidelines

Enable and disable this feature as needed. Enabling resource monitoring when the number of online members reaches a certain threshold may cause the commander's CPU to become busy and memory space to be insufficient.

Procedure

1. Enter system view.

   system-view

2. Set the interval for the commander to obtain resource monitoring information.

   smartmc resource-monitor interval *interval*

   The default setting is 1 minute.

3. Set the aging time for resource monitoring information.

   smartmc resource-monitor max-age *max-age*

   The default setting is 24 hours.

4. Enable resource monitoring.

   smartmc resource-monitor [**cpu** | **memory** | **packet-drop** | **temperature**]* [**group** *group-name-list* | **tc** { *tc-id-list* | **mac-address** *mac-address* } | **tm**]

   By default, resource monitoring is disabled.

   If you do not specify a resource type, this command enables resource monitoring for all resource types.

   If you do not specify a device to monitor (member or commander), this command enables resource monitoring on the commander and all members.

Upgrading the startup software and configuration file on members

About upgrading the startup software and configuration file on members

You can use the following methods to upgrade the startup software and configuration file on members:

- Schedule an upgrade by specifying an upgrade time or upgrade delay.
- Upgrade immediately by not specifying an upgrade time or upgrade delay.

Restrictions and guidelines for startup software and configuration file upgrade

A member can perform only one upgrade task at a time.

An immediate upgrade cannot be cancelled. If you specify a delay time or upgrade time to perform a scheduled upgrade, the upgrade operation can be cancelled by using the **undo smartmc upgrade** command before it starts.

Prerequisites

Before performing this task, you must specify a file server (see "Specifying a file server").

Upgrading the startup software and configuration file on members

Upgrading the startup software and configuration file in one step

1. Enter system view.

   system-view

2. Upgrade the startup software on members in one step.

   smartmc upgrade boot-loader tc { *tc-id-list* { **boot** *boot-filename* **system** *system-filename* | **file** *ipe-filename* }&<1-40> [**delay** *delay-time* | **time** *in-time*]

   CAUTION:

   Upgrading the startup software might interrupt services. Before upgrading the startup software, make sure no services will be interrupted.

3. Upgrade the configuration file on members in one step.

   smartmc upgrade startup-configuration tc { *tc-id-list* *cfg-filename* }&<1-40> [**delay** *delay-time* | **time** *in-time*]

CAUTION:

After you update the configuration file, the configuration in the new configuration file will become the running configuration. Before upgrading the configuration file, make sure the contents of the new configuration file are correct.

Configuring startup software and configuration file upgrade step by step

1. Enter system view.

   system-view

2. Configure startup software upgrade for members step by step:

   a. Specify the upgrade startup software files.

      smartmc tc *tc-id* **boot-loader** { *ipe-filename* | **boot** *boot-filename* **system** *system-filename* }

   b. Upgrade the startup software on members.

   smartmc upgrade boot-loader tc *tc-id-list*

CAUTION:

Upgrading the startup software might interrupt services. Before upgrading the startup software, make sure no services will be interrupted.

3. Configure configuration file upgrade for members step by step:

   a. Specify the upgrade configuration file.

      smartmc tc *tc-id* **startup-configuration** *cfg-filename*

   b. Upgrade the configuration file on members.

   smartmc upgrade startup-configuration tc *tc-id-list*

CAUTION:

After you update the configuration file, the configuration in the new configuration file will become the running configuration. Before upgrading the configuration file, make sure the contents of the new configuration file are correct.

Upgrading the startup software and configuration file on all members in SmartMC groups

Upgrading the startup software and configuration file in one step

1. Enter system view.

   system-view

2. Upgrade the startup software on all members in SmartMC groups in one step.

   smartmc upgrade boot-loader group { *group-name-list* [**boot** *boot-filename* **system** *system-filename* | **file** *ipe-filename*] }&<1-40> [**delay** *minutes* | **time** *in-time*]

CAUTION:

Upgrading the startup software might interrupt services. Before upgrading the startup software, make sure no services will be interrupted.

3. Upgrade the configuration file on all members in SmartMC groups in one step.

smartmc upgrade startup-configuration group { *group-name-list* **file** *cfg-filename* }&<1-40> [**delay** *minutes* | **time** *in-time*]

CAUTION:

After you update the configuration file, the configuration in the new configuration file will become the running configuration. Before upgrading the configuration file, make sure the contents of the new configuration file are correct.

Configuring startup software and configuration file upgrade step by step

1. Enter system view.

   system-view

2. Enter SmartMC group view.

   smartmc group *group-name*

3. Specify the upgrade startup software files for the SmartMC group.

   boot-loader file { *ipe-filename* | **boot** *boot-filename* **system** *system-filename* }

   By default, no upgrade startup software files are specified for a SmartMC group.

4. Specify the upgrade configuration file for the SmartMC group.

   startup-configuration *cfgfile*

   By default, no upgrade configuration file is specified for a SmartMC group.

5. Return to system view.

   quit

6. Upgrade the startup software and configuration file on all members in the SmartMC group.

   Choose one option as needed:

   - Upgrade the startup software.

     smartmc upgrade boot-loader group *group-name-list* [**delay** *minutes* | **time** *in-time*]

     CAUTION:

     Upgrading the startup software might interrupt services. Before upgrading the startup software, make sure no services will be interrupted.

   - Upgrade the configuration file.

     smartmc upgrade startup-configuration group *group-name-list* [**delay** *minutes* | **time** *in-time*]

     CAUTION:

     After you update the configuration file, the configuration in the new configuration file will become the running configuration. Before upgrading the configuration file, make sure the contents of the new configuration file are correct.

Managing the network topology

Refreshing the network topology

About this task

You can use the following methods to refresh the network topology:

- **Automatic topology refresh**—Specify the refresh interval to allow the commander to refresh the network topology periodically.
- **Manual topology refresh**—Execute the **smartmc topology-refresh** command to manually refresh the network topology.

Restrictions and guidelines

The topology refresh time depends on the number of members in the network.

When the number of online members in the network is greater than 64, the commander calculates the refresh interval of the network topology based on the number of online members. The calculation formula is (*N*+64)/64 multiplied by the set refresh interval, where *N* represents the number of online members.

Procedure

Choose one option as needed:

- Manually refresh the network topology in any view.
  smartmc topology-refresh
- Configure automatic network topology refresh.
  a. Enter system view.
  system-view
  b. Set the automatic topology refresh interval.
  smartmc topology-refresh interval *interval*
  By default, the automatic topology refresh interval is 60 seconds.

Saving the network topology

About this task

This task allows you to save the current network topology to the **topology.dba** file in the flash memory. After the commander reboots, it uses the **topology.dba** file to restore the network topology.

Procedure

1. Enter system view.
   system-view
2. Save the network topology.
   smartmc topology-save

Replacing faulty members

Restrictions and guidelines

Make sure the new member for replacement and the faulty member have the same neighbor relationship and device model.

Make sure the new member has a different member ID than all the members in the SmartMC network, including offline members. Faulty members are considered offline.

To automatically replace a faulty member, first enable automatic replacement, and then install the new member at the location where the faulty member was installed and connect all cables.

To manually replace a faulty member, first install the new member at the location where the faulty member was installed, connect all cables, and then execute the manual replacement command.

Prerequisites

Before you replace a faulty member, specify a file server (see "Specifying a file server").

Procedure

1. Enter system view.
   system-view
2. Replace faulty members.
   Choose one option as needed:
   - Enable automatic faulty member replacement.
     smartmc auto-replace enable
     By default, automatic faulty member replacement is disabled.
   - Manually replace a faulty member.
     smartmc replace tc *tc-id1* **faulty-tc** *tc-id2*

Display and maintenance commands for SmartMC

Execute **display** commands in any view.

Task	Command
Display the backup status on members.	**display smartmc backup configuration status**
Display the batch file execution results.	**display smartmc batch-file status** [**ap** | **last** *number* | **phone**]
Display SmartMC configuration.	**display smartmc configuration**
Display connections between the devices in the SmartMC network.	**display smartmc device-link**
Display SmartMC group information.	**display smartmc group** [*group-name*] [**verbose**]
Display the faulty member replacement status.	**display smartmc replace status**
Display resource monitoring information.	**display smartmc resource-monitor** [**cpu** | **memory** | **temperature**] * [**tc** *tc-id* | **tm**]
Display resource monitoring configuration.	**display smartmc resource-monitor configuration**
Display member information.	**display smartmc tc** [*tc-id*] [**verbose**]
Display log information in the log buffer on a member.	**display smartmc tc** *tc-id* **log buffer** [**module** *module-name* [**mnemonic**

Task	Command
	mnemonic-value]]
Display restart log information for a member.	**display smartmc tc** *tc-id* **log restart**
Display VLAN creation results for members.	**display smartmc vlan**
Display member upgrade status.	**display smartmc upgrade status**

SmartMC configuration examples

Example: Configuring SmartMC

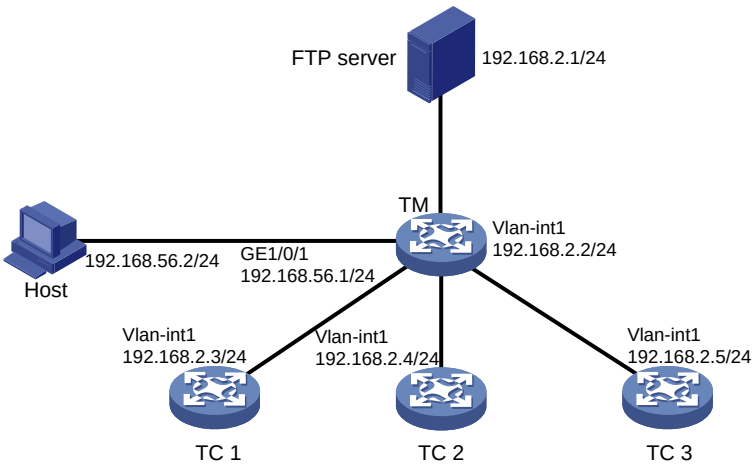
Network configuration

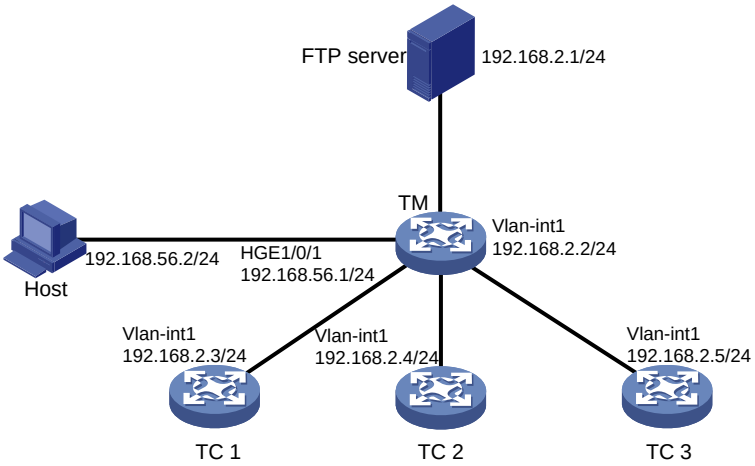
As shown in Figure 3, member 1, member 2, and member 3 belong to the same device type: S5590-EI series. The IP address of the file server is 192.168.2.1. The username and password for accessing the file server are **admin** and **hello12345**, respectively.

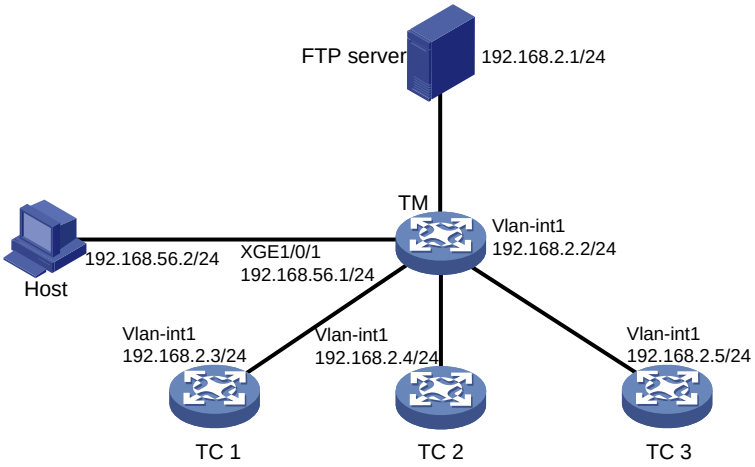
Perform the following tasks to establish a SmartMC network and upgrade the configuration file on the members:

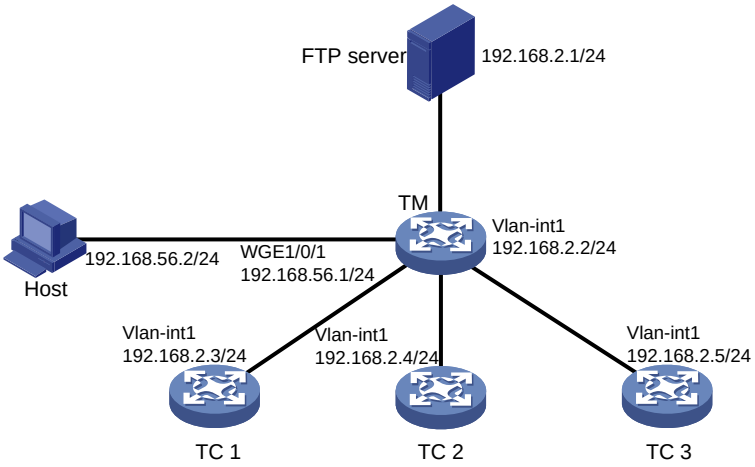
1. Configure the commander and members to automatically establish a SmartMC network.
2. Configure interface GigabitEthernet 1/0/1 as the outgoing interface for the SmartMC network.
3. Create a SmartMC group and add the members to the group.
4. Upgrade the configuration file on all members in the SmartMC group.
5. Save configuration file **startup.cfg** on the file server.

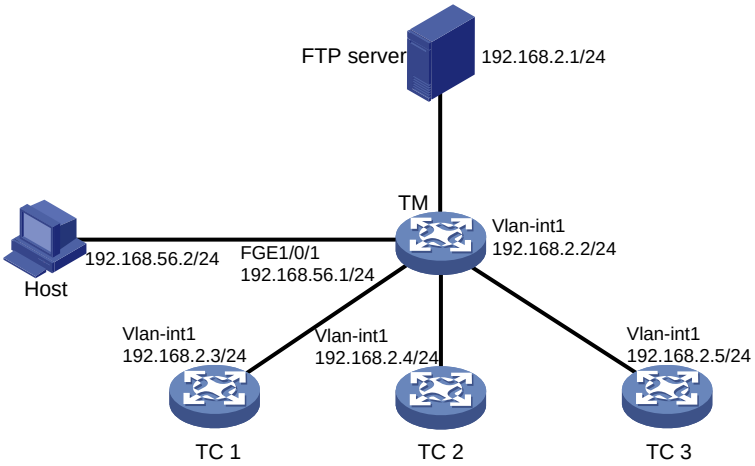
Figure 3 Network diagram

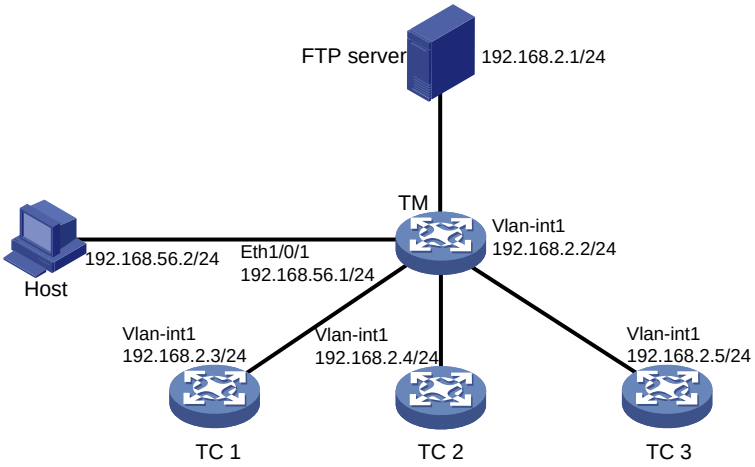


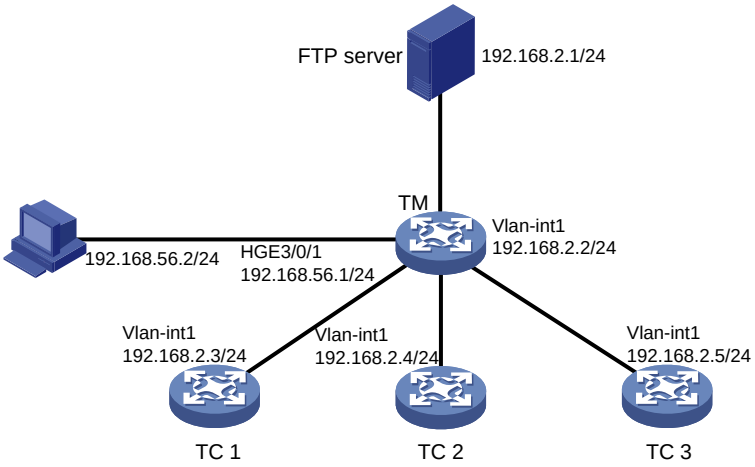

FTP server
192.168.2.1/24
TM
Vlan-int1
192.168.2.2/24
192.168.56.2/24
HGE1/0/1
192.168.56.1/24
Host
Vlan-int1
192.168.2.3/24
Vlan-int1
192.168.2.4/24
Vlan-int1
192.168.2.5/24
TC 1
TC 2
TC 3

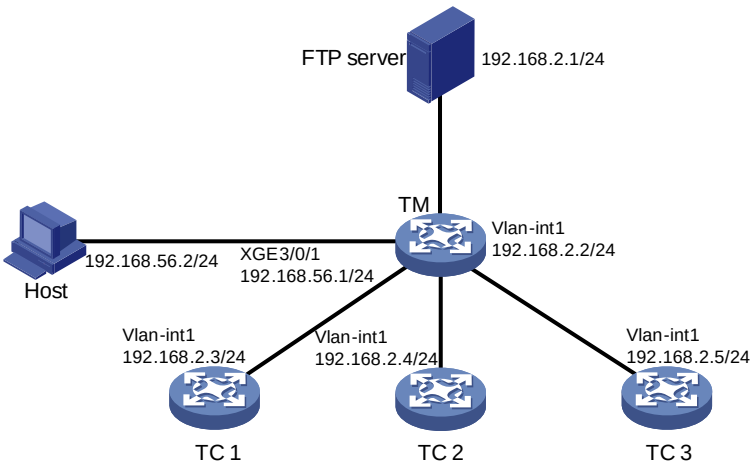

FTP server
192.168.2.1/24
TM
Vlan-int1
192.168.2.2/24
192.168.56.2/24
XGE1/0/1
192.168.56.1/24
Host
Vlan-int1
192.168.2.3/24
Vlan-int1
192.168.2.4/24
Vlan-int1
192.168.2.5/24
TC 1
TC 2
TC 3

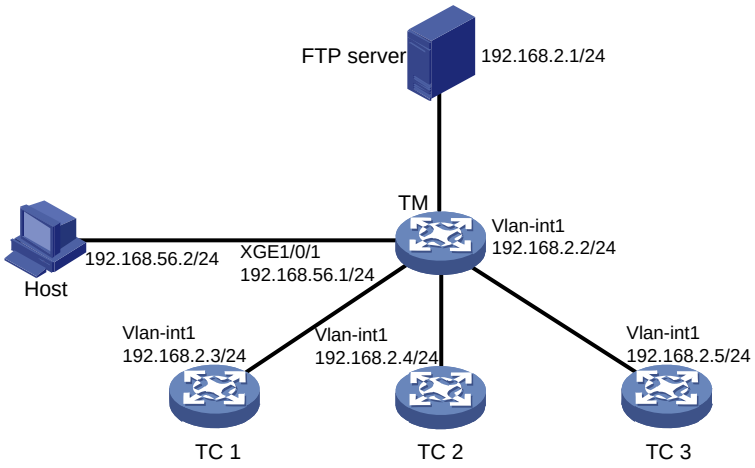

FTP server
192.168.2.1/24
TM
Vlan-int1
192.168.2.2/24
192.168.56.2/24
WGE1/0/1
192.168.56.1/24
Host
Vlan-int1
192.168.2.3/24
Vlan-int1
192.168.2.4/24
Vlan-int1
192.168.2.5/24
TC 1
TC 2
TC 3

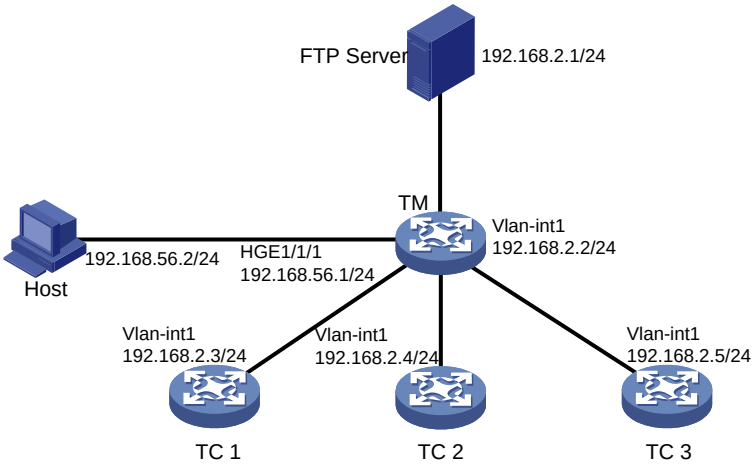

FTP server
192.168.2.1/24
TM
Vlan-int1
192.168.2.2/24
192.168.56.2/24
FGE1/0/1
192.168.56.1/24
Host
Vlan-int1
192.168.2.3/24
Vlan-int1
192.168.2.4/24
Vlan-int1
192.168.2.5/24
TC 1
TC 2
TC 3


FTP server
192.168.2.1/24
TM
Vlan-int1
192.168.2.2/24
192.168.56.2/24
Eth1/0/1
192.168.56.1/24
Host
Vlan-int1
192.168.2.3/24
Vlan-int1
192.168.2.4/24
Vlan-int1
192.168.2.5/24
TC 1
TC 2
TC 3


FTP server
192.168.2.1/24
TM
Vlan-int1
192.168.2.2/24
192.168.56.2/24
HGE3/0/1
192.168.56.1/24
Vlan-int1
192.168.2.3/24
Vlan-int1
192.168.2.4/24
Vlan-int1
192.168.2.5/24
TC 1
TC 2
TC 3







Procedure

1. Configure TC 1:

Configure VLAN-interface 1.

```
<TC1> system-view
[TC1] interface vlan-interface 1
[TC1-Vlan-interface1] ip address 192.168.2.3 24
[TC1-Vlan-interface1] quit
```

Enable the Telnet service.

```
[TC1] telnet server enable
```

Enable NETCONF over SOAP over HTTP.

```
[TC1] netconf soap http enable
```

Enable NETCONF over SSH.

```
[TC1] netconf ssh server enable
```

Enable LLDP globally.

```
[TC1] lldp global enable
```

Create a user.

```
[TC1] local-user admin
```

Lower password complexity requirements. For more information about these commands, see password control commands in *Security Command Reference*.

```
[TC1-luser-manage-admin] password-control length 4
[TC1-luser-manage-admin] password-control composition type-number 1 type-length 1
[TC1-luser-manage-admin] undo password-control complexity user-name check
```

Set the username and password to **admin**, add the **telnet**, **http**, and **https** service types, and authorize the user to use the **network-admin** user role.

```
[TC1-luser-manage-admin] password simple admin
[TC1-luser-manage-admin] service-type telnet http https ssh
[TC1-luser-manage-admin] authorization-attribute user-role network-admin
[TC1-luser-manage-admin] quit
```

Set the system to support public key authentication for the administrator to login through SSH. The public key name must be **-smartmc-publickey-**.

```
[TC1] ssh user admin service-type netconf authentication-type publickey assign
publickey -smartmc-publickey-
```

Set scheme authentication for VTY user lines 0 to 63.

```
[TC1] line vty 0 63
[TC1-line-vty0-63] authentication-mode scheme
[TC1-line-vty0-63] quit
```

Enable SmartMC and set the device role to **tc**.

```
[TC1] smartmc tc enable
```

2. Configure TC 2 and TC 3 in the same way TC 1 is configured. (Details not shown.)

3. Configure the TM:

Configure GigabitEthernet 1/0/1.

```
<TM> system-view
[TM] interface gigabitethernet 1/0/1
[TM-GigabitEthernet1/0/1] port link-mode route
[TM-GigabitEthernet1/0/1] ip address 192.168.52.2 24
[TM-GigabitEthernet1/0/1] quit
```

Configure VLAN-interface 1.

```
[TM] interface vlan-interface 1
[TM-Vlan-interface1] ip address 192.168.2.2 24
[TM-Vlan-interface1] quit
```

Enable the Telnet service.

```
[TM] telnet server enable
```

Enable NETCONF over SOAP over HTTP.

```
[TM] netconf soap http enable
```

Enable NETCONF over SSH.

```
[TM] netconf ssh server enable
```

Enable LLDP globally.

```
[TM] lldp global enable
```

Create a user. Set the username to **admin** and the password to **hello12345**, add the **telnet**, **http**, and **https** service types, and authorize the user to use the **network-admin** user role.

```
[TM] local-user admin
[TM-luser-manage-admin] password simple hello12345
[TM-luser-manage-admin] service-type telnet http https ssh
[TM-luser-manage-admin] authorization-attribute user-role network-admin
[TM-luser-manage-admin] quit
```

Set scheme authentication for VTY user lines 0 to 63.

```
[TM] line vty 0 63
[TM-line-vty0-63] authentication-mode scheme
[TM-line-vty0-63] quit
```

Enable SmartMC, set the device role to commander, and set the username to **admin** and the password (plaintext) to **hello12345**.

```
[TM] smartmc tm username admin password simple hello12345 enable
```

Specify GigabitEthernet 1/0/1 as the outgoing interface.

```
[TM] interface gigabitethernet 1/0/1
[TM-GigabitEthernet1/0/1] smartmc outbound
[TM-GigabitEthernet1/0/1] quit
```

Set the file server IP address, username, and plaintext password to 192.168.2.1, **admin**, and **hello12345**, respectively.

```
[TM] smartmc ftp-server 192.168.2.1 username admin password simple hello12345
```

Create SmartMC group **S1** and enter its view.

```
[TM] smartmc group S1
```

Create an IP address match criterion to add all members in the specified network segment to SmartMC group **S1**.

```
[TM-smartmc-group-S1] match ip-address 192.168.2.0 24
```

Specify the upgrade configuration file **startup.cfg** for SmartMC group **S1**.

```
[TM-smartmc-group-S1] startup-configuration startup.cfg
[TM-smartmc-group-S1] quit
```

Upgrade the configuration file on all members in SmartMC group **S1**.

```
[TM] smartmc upgrade startup-configuration group S1 file startup.cfg
```

Verifying the configuration

Display brief information about all members after the SmartMC network is established.

```
[TM] display smartmc tc
TCID DeviceType Sysname  IpAddress       MacAddress     Status   Version
1    S5590-EI   TC1      192.168.2.3     201c-e7c3-0300 Normal   COMWAREV700R001
2    S5590-EI   TC2      192.168.2.4     201c-e7c3-0301 Normal   COMWAREV700R001
3    S5590-EI   TC3      192.168.2.5     201c-e7c3-0302 Normal   COMWAREV700R001
```

Display the configuration file upgrade status on the members.

```
[TM] display smartmc upgrade status
ID    IpAddress        MacAddress      Status      UpdateTime        UpdateFile
1     192.168.2.3      201c-e7c3-0300  Finished    Immediately       startup.cfg
2     192.168.2.4      201c-e7c3-0301  Finished    Immediately       startup.cfg
```

```
3     192.168.2.5      201c-e7c3-0302  Finished    Immediately       startup.cfg
```

Contents

Quicknet commands 1

About Quicknet 1
Benefits 1
Quicknet network framework 1
Quicknet network management 2
Quicknet role selection 6
Quicknet network establishment and changing 7
Quicknet splitting 8
Quicknet merging 9
Quicknet configuration synchronization 10
Quicknet time synchronization (NTP) 11
Quicknet management system login with a fixed domain name 12
Restrictions and guidelines: Quicknet configuration 12
Quicknet tasks at a glance 12
Enabling Quicknet 12
Setting the network-wide Quicknet password 13
Display and maintenance commands for Quicknet 13
Quicknet configuration examples 14
Example: Configuring a wired Quicknet network 14
Example: Configuring a wired and wireless Quicknet network 17
Quicknet configuration examples 21
Example: Configuring a wired Quicknet network 21
Example: Configuring a wired and wireless Quicknet network 29

Quicknet commands

About Quicknet

Quicknet is a lightweight network management topology used to centrally manage and maintain dispersed network devices at network edges. In a Quicknet network, only one device acts as the TM and the remaining devices all act as TCs. The TM manages the entire Quicknet network in a unified way.

Benefits

Quicknet provides the following benefits:

- Simplified network setup. Quicknet supports plug-and-play for devices. Once initial settings are configured from the Web interface, devices can automatically form a Quicknet network based on the factory configuration.
- Simplified network maintenance. After the Quicknet network is built, the network administrator can log into the Quicknet management system and, through simple Web operations, can achieve unified management and maintenance of a large number of network devices. This effectively solves operation and maintenance issues for small and medium-sized campus networks.

Quicknet network framework

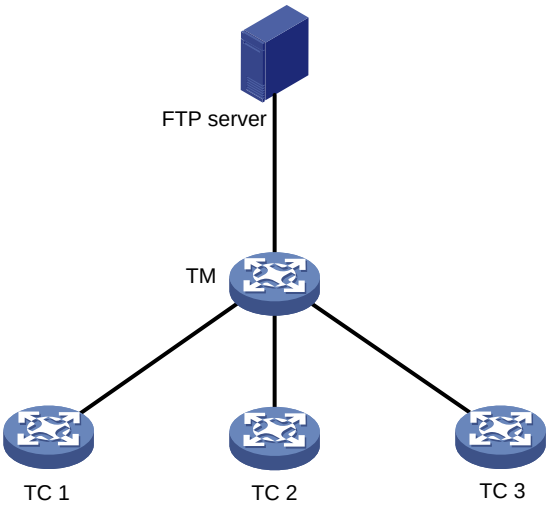
Figure 1 shows the basic framework of a Quicknet network. The Quicknet network contains the following elements:

- Topology master (TM)—Manages all members in the Quicknet network.
- **Topology client (TC)**—Managed by the TM.
- FTP server—Stores startup software images and configuration files for the TM and members.

Quicknet networks can include various models of network devices. The method for determining device roles varies by product positioning.

- Some device models support manually designating the device role as TC through the CLI.
- Some device models support manually designating the device role as TM through the CLI.
- Some device models support using the auto election mechanism to determine their roles (TC or TM) in a Quicknet network. For more information about auto election, see "Quicknet role selection."

Figure 1 Quicknet network framework



Quicknet network management

Introduction

Quicknet provides various flexible network device management solutions, allowing you to select any of the following methods as needed:

- Local unified management (mainly used for initial deployment)

  Through the Quicknet management system, all network devices can be centrally managed on-site, significantly improving operational efficiency, and is especially suitable for rapid deployment and management of small to medium-sized campus networks.

- Cloudnet remote management (recommended for O&M)

  With the Cloudnet management platform, you can remotely manage all devices in the Quicknet network anytime, anywhere.

- Single-device local configuration
  - Web interface: Access the graphical web page through the device IP to intuitively configure a single device.
  - CLI: Log in to the device command line interface and use commands for advanced configuration.

Quicknet management system

The Quicknet management system is a lightweight web-based network management platform and acts as the unified management entry for the Quicknet network running on TM. You can visit https://quicknet.h3c.com to access the Quicknet management system on the TM. It is mainly used for the deployment of Quicknet networks. After completing the initial deployment, you can perform simple operations and maintenance on all network devices through the Quicknet management system. To perform in-depth O&M, use the one-key onboarding function provided by the Quicknet management system as a best practice.

The Quicknet management system provides the following functions:

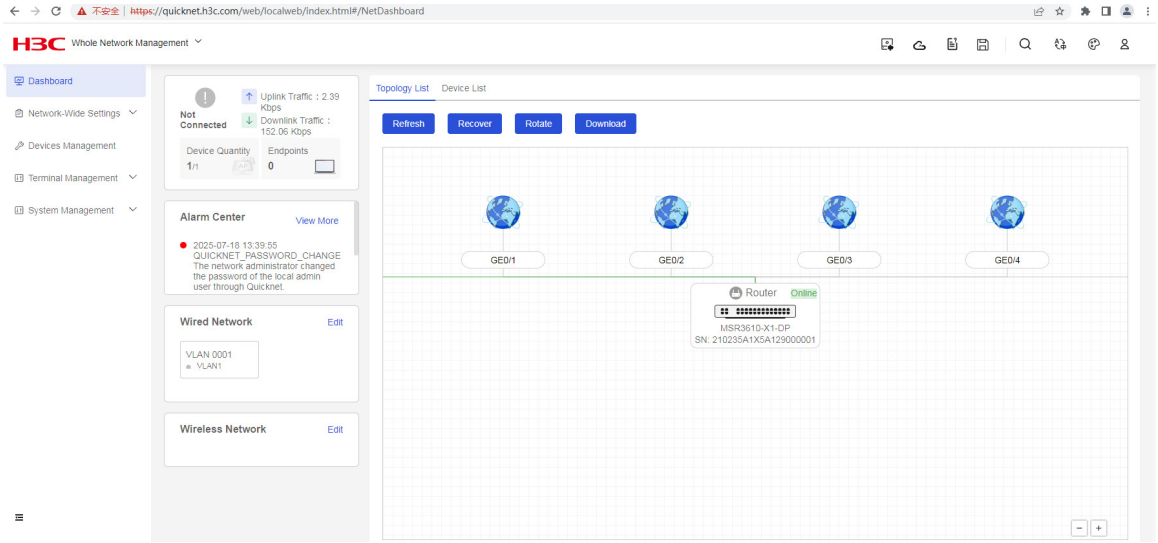
- Initialization—Configures the project name, network-wide password, Internet parameters, Wi-Fi settings, and time zone for the Quicknet network.
- Network Dashboard—Displays network connectivity information, alarm messages, wired network information, wireless service information, topology list, device list, and device details, and provides a portal for users to switch to local management.

- Network Configuration—Includes WAN and wired configuration. WAN settings include network access parameters and wired settings include the service VLAN.
- Device Management—Includes adding devices, removing offline devices, rebooting devices, upgrading devices, resetting devices, viewing device details, configuring device interfaces, creating VLANs, editing VLANs, and deleting VLANs.
- Terminal Management—Allows viewing and management of wired and wireless users.
- System Management—Includes upgrading, rebooting, and modifying the network-wide password.

By combining the aforementioned functions, you can achieve the following objectives:

- Management the Quicknet network topology.
- Manage member devices in bulk.

Figure 2 Quicknet management system



Quicknet network deployment

Quicknet provides an intelligent deployment solution, enabling network deployment in a few simple steps. The specific process is as follows:

1. Auto networking (plug and play)
   - Networking conditions:
     - Devices can reach each other at Layer 2 (allowing VLAN 1 packets to pass).
     - Quicknet is enabled (whether Quicknet is enabled by default depends on the device model).
   - Auto TM election: After the conditions are met, the devices automatically elect the TM and establish a Quicknet network.
2. Device role and information query
   - After completing physical connections, you can execute the **`display quicknet configuration`** command on the TM to verify the TM device role and execute the **`display quicknet configuration`** command on a TC to verify the TC device role and view TM information.
   - The network is not deployed yet. Executing the **`display quicknet tc`** command on the TM cannot display TC information.
3. Network deployment methods
   - Wireless access (recommended)

- Connect to the Wi-Fi whose name includes the Quicknet string.
- Use a browser to visit https://quicknet.h3c.com and perform initial deployment as instructed.

- Wired access (login to TM Web interface)
  - Connect your PC to VLAN-interface 1 on any device in the Quicknet network.
  - Use a browser to visit https://quicknet.h3c.com and perform initial deployment as instructed.

  In the Quicknet network, a DHCP server will be automatically deployed to assign IP addresses in the VLAN 1 segment to PCs, and provide translation from https://quicknet.h3c.com to the TM IP address. If a PC accesses the Quicknet management system across network segments, you must configure the mapping relationship between https://quicknet.h3c.com and the IP address of Quicknet on the local PC.

4. Initial deployment configuration
   - You can configure the network-wide password, Wi-Fi, and time zone settings.
   - The configuration will be automatically synchronized to all member devices and take effect.
5. Deployment process on the TM
   a. Generates a unique NetID for the current Quicknet network.
   b. Requires the logged-in administrator to set a management password (the Quicknet network-wide password) and synchronizes this password along with the NetID to all member devices in the Quicknet network.

Set the devices with the network-wide password and NetID to the deployed state. The state will affect the TM election later.

Quicknet network-wide password

The Quicknet network-wide password is used to ensure the security of the NETCONF communication channel between TM and TC devices.

After Layer 2 connections are established in the Quicknet network, the TM must establish Layer 3 communication channels with TCs to synchronize configurations and obtain device information.

To ensure communication security, the TM and TCs must perform identity authentication (the authentication credentials are the local user admin and the corresponding password). The device comes with default credentials from the factory. The username admin in the default credentials cannot be changed, but the password can be modified. This password is also known as the Quicknet network-wide password.

You can use either method to configure the network-wide password:

1. Modify during initial configuration

   Set the network-wide password in the Quicknet management system's deployment wizard.
2. Modify during device operation
   - Modify from the Quicknet management system's Web interface.
   - Modify from the TM's CLI.

The network-wide password synchronization mechanism is as follows:

- After the password is changed, the TM automatically synchronizes the new password to all network devices.
- The system will rebuild all communication channels based on the new password.

When you configure the network-wide password, follow these restrictions and guidelines:

- As a best practice for security purposes, change the default password upon initial deployment.
- Change password regularly to comply with corporate security policies.
- The password must contain a minimum of eight characters and contain uppercase and lowercase letters, digits, and special characters.

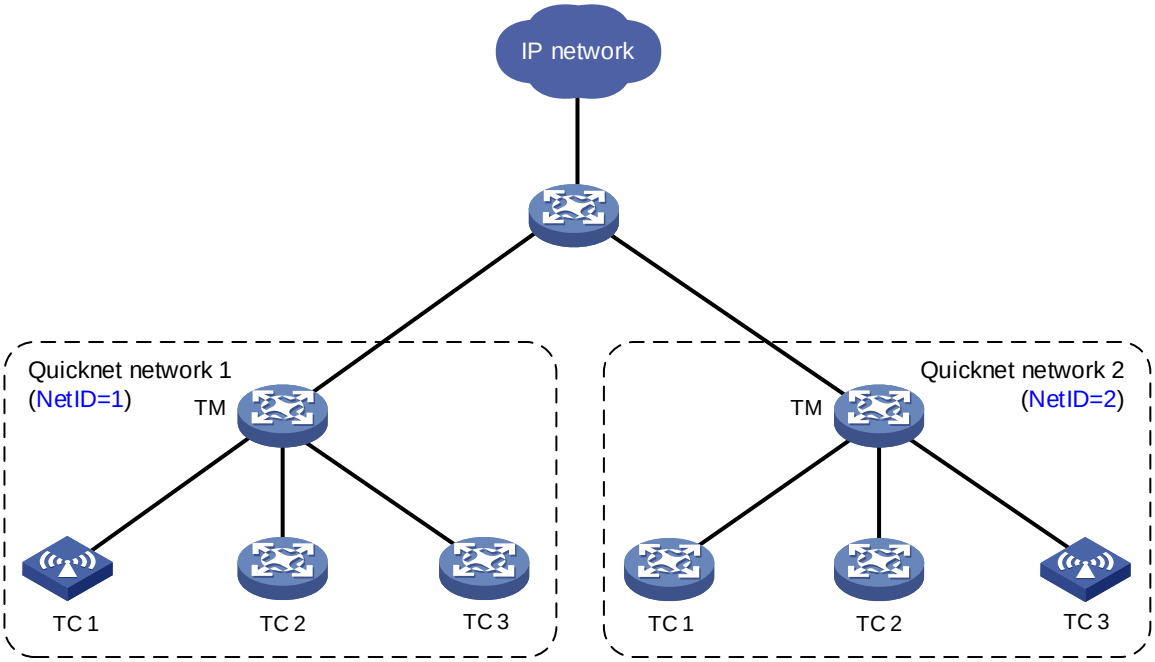
Quicknet NetID

A NetID is the identifier of a Quicknet network in the Quicknet system, and one Quicknet network corresponds to only one NetID.

To adapt to various networking applications, multiple Quicknet networks can be deployed in the same network, and NetID is used to distinguish between different Quicknet networks. One device can only join one Quicknet network. As shown in Figure 3, the TM broadcasts Quicknet protocol packets to discover TCs in the respective Quicknet network. To avoid interference between Quicknet protocol packets of two Quicknet networks, Quicknet uses NetID to distinguish between different Quicknet networks. When a member device in a Quicknet network receives a Quicknet packet with a different NetID, it directly discards it.

The NetID is automatically generated by the Quicknet module at the initial deployment phase.

Figure 3 Multi-Quicknet networking

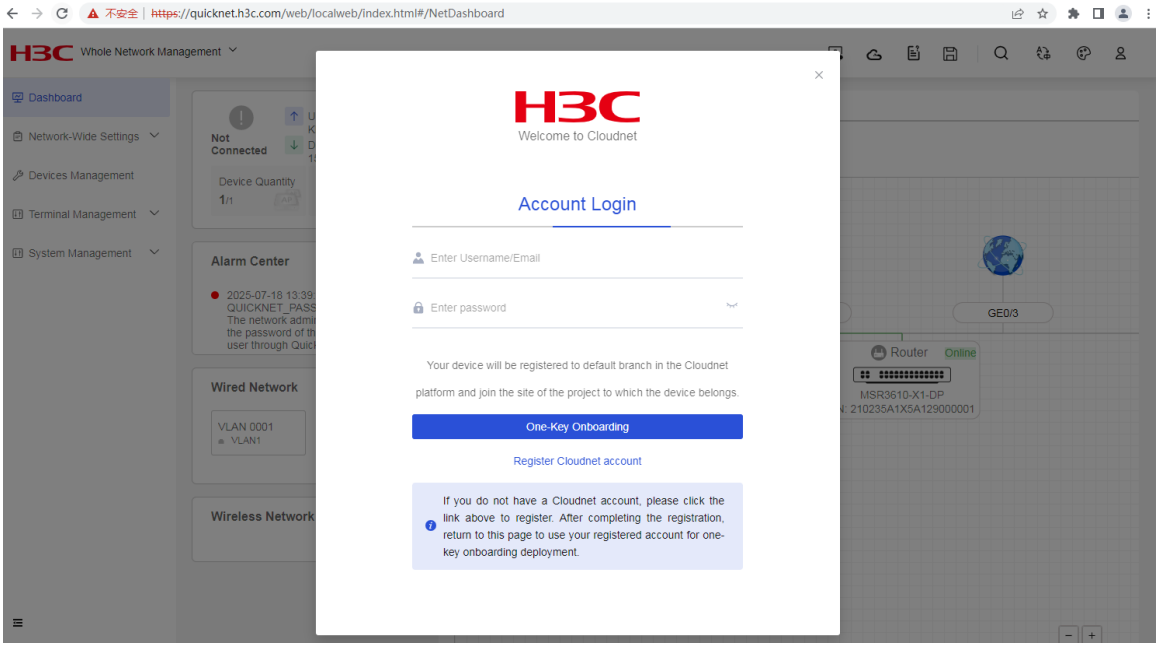


One-key onboarding

After completing the Quicknet deployment, to manage devices both locally and from the cloud, click the one-key onboarding icon in the top navigation bar of the Quicknet management system to register all network devices to Cloudnet.

After successfully executing one-key onboarding, the administrator can log in to the H3C Cloudnet management platform (https://cloudnet.h3c.com) to implement remote management.

Figure 4 One-key onboarding



Quicknet role selection

Quicknet networks can include various models of network devices. The method for determining device roles varies by product positioning.

- Some device models support manually designating the device role as TC through the CLI.
- Some device models support manually designating the device role as TM through the CLI.
- Some device models support using the auto election mechanism to determine their roles (TC or TM) in a Quicknet network.

For devices that support Quicknet, the role is not determined in the factory settings, and the TM is automatically elected through role election. A Quicknet network conducts role election in the following situations:

- The devices are first interconnected, which triggers Quicknet network formation.
- The TM leaves the network or malfunctions.
- The Quicknet network splits.
- Two (or more) independent Quicknet networks merge into one.

The following priority order is used to select the TM in the role election:

1. The TM specified by command execution is prioritized.
2. The device with a higher Quicknet election priority is prioritized.

   Each device comes with a priority in factory settings, which is determined by the device model and network location and cannot be modified from the CLI. The Quicknet election priorities of the following devices are in descending order: AC, router, security device, AP. The role of a switch requires manual configuration.

   - If none of the devices participating in the role election have finished startup deployment, the device with the highest priority becomes the TM.
   - When a device joins a deployed Quicknet network, if the priority of the new device is higher than the priority of the existing TM, the new device synchronizes configuration and then becomes the TM. If the priority of the new device is the same as or is lower than the priority of the existing TM, the new device joins the network as a TC.

3. Devices that have obtained Quicknet configuration and have been deployed are prioritized.

   If two devices have the same priority, the device that has obtained Quicknet configuration and has been deployed becomes the TM.

4. The TM in the larger Quicknet network is prioritized.

   When two Quicknet networks merge, the TM in the network with more members become the new TM.

5. Devices that have higher heath scores are prioritized.
6. Devices that have lower bridge MAC addresses are prioritized.

During the election process, the devices broadcast election packets at Layer 2 and follow the above rules until the only TM is selected.

Quicknet network establishment and changing

Quicknet network establishment

The network establishment process is as follows:

1. If a device is manually designated as TM or TC in the network, it will not participate in the role election process but will directly enter the TM or TC state. The TM will broadcast discovery packets to announce its identity.
2. For a device that supports automatic election, it will enter the TC(auto) state by default after startup.
   - If the device receives a TM broadcast packet within 5 seconds, its role is determined as TC, and the role election process will not be initiated.
   - If the device does not receive a TM broadcast packet within 5 seconds, it enters the role election phase and broadcasts an election packet (containing information such as bridge MAC, NetID, precedence, network scale, and VLAN-interface1 IP) at 5-second intervals. The optimal device is elected as the TM and other devices become TCs.
     - If the device receives an election packet from a device with better conditions within 20 seconds, it switches to TC. If the local device has better conditions, it becomes the TM.
     - If the device does not receive any election packets from other devices within 20 seconds, it switches to TM.
3. The TM broadcasts discovery packets at intervals of 5 seconds to announce its TM identity.
4. Each TC records the device information of the TM and sends a Quicknet discovery ACK packet back to the TM. At this time, you can use the **`display quicknet configuration`** command on the device to view the Quicknet configuration.
5. The network administrator can use one of the following methods to perform initial deployment:
   - Connect to the TM directly.
   - Connect to the Wi-Fi whose name includes the Quicknet string.

   Visit https://quicknet.h3c.com to log in to the management system. The system will automatically issue the LLDP and NETCONF configuration.
6. During the deployment process, the TM attempts to establish NETCONF sessions with all member devices and authenticates each TC using the username admin and the default password.
   - If the authentication is successful, the NETCONF session is established successfully.
   - If the authentication fails, the NETCONF session fails to be established.

   The TM sends commands to the TC through this NETCONF session to obtain information from the TC.
7. The TM obtains detailed information of the TC (including port, LLDP neighbors, and STP status) through a NETCONF session.
8. The TC sends a login request packet (containing device ID information).

9. The TM confirms the information and adds the TC to the network. You can use the **`display quicknet tc`** command to view TC information.
10. The TM builds the network topology based on LLDP information (the network topology can be viewed in Quicknet).
11. The administrator sets the network-wide password and Internet parameters according to the system guide to complete the deployment.
12. The TM synchronizes the password and rebuilds all NETCONF connections.

The Quicknet network has been officially established and can be uniformly operated and maintained through the management system.

New device joining

The mechanism for automatically electing the TM when a new device joins is as follows:

1. After the Quicknet network is established, the TM broadcast discovery packets at a fixed interval (default to 5 seconds), inquiring if member devices exist in the network.
2. When a new device is connected and powered on:
   - If the configuration of a newly added device specifies the role as TC, the device directly joins the Quicknet network as a TC.
   - If the configuration of a newly added device specifies the role as TM, the device cannot join the Quicknet network.
   - If the configuration of a newly added device uses the self-election function, the new device will enter the role election state and determine whether it acts as a TM or TC according to the role election rules.

Member device leaving

After the Quicknet network is successfully established, the TM and TCs perceive each other's existence through Quicknet broadcast packets and response packets:

- After the TM leaves, TCs will enter the role election state if they fail to receive Quicknet discovery packets from the TM within 20 seconds. TCs then re-elect an TM according to the role election rules.
- After a TC leaves, the TM considers a TC as offline if it fails to receive any Quicknet discovery ACK packet from the TC within 20 seconds.

Quicknet splitting

If link failures occur between member devices in a Quicknet network, TCs and TM may fail to reach each other, causing the Quicknet network to split into multiple Quicknet networks. After the splitting, one Quicknet network contains the original TM, while other Quicknet networks do not have a TM.

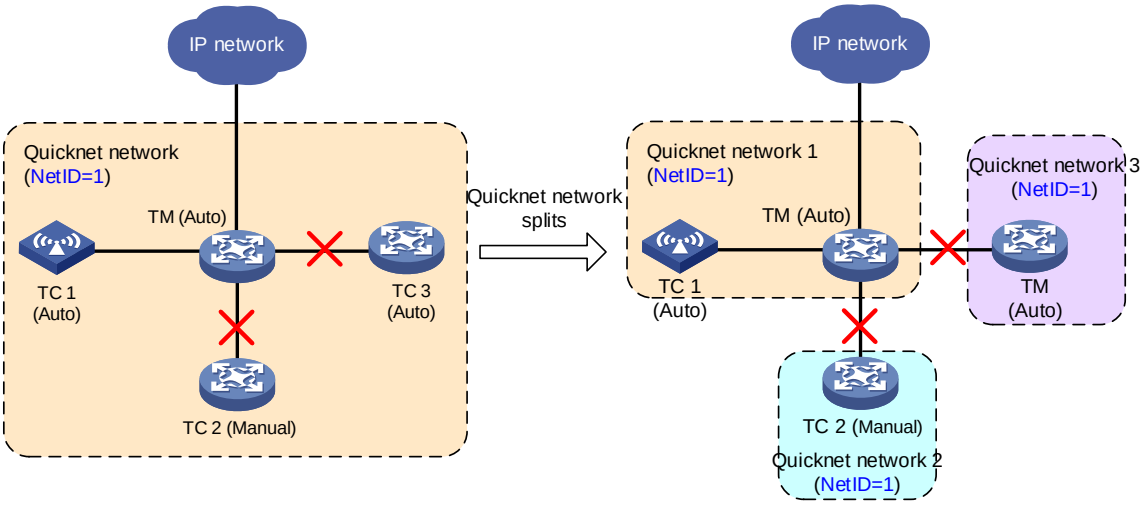
During the splitting, the Quicknet software module handles the situation as follows:

- For the Quicknet network that contains the original TM, the Quicknet network continues to operate normally, and the departing member devices become offline.
- For a Quicknet network that does not contain a TM, if some devices in the network support self-election, they consider the TM as faulty and trigger role election after 20 seconds without receiving any Quicknet broadcast discovery packets from a TM.

As shown in Figure 5, the roles of TM, TC 1, and TC 3 in the Quicknet network are decided through election, and the role of TC 2 is manually specified. Assuming that this Quicknet network has been established with a NetID of 1, the following scenarios might occur:

- When the link fails between TC 2 and the TM, TC 2 leaves the Quicknet network, but its role remains as TC and its NetID remains as 1.
- When the link fails between TC 3 and the TM, TC 3 leaves the Quicknet network. After 20 seconds without receiving any Quicknet broadcast discovery packets from the TM, TC 3 re-elects itself as the TM with a NetID of 1.

Figure 5 Quicknet splitting



Quicknet merging

The process of multiple stable Quicknet networks interconnecting to form a Quicknet network is called Quicknet merging.

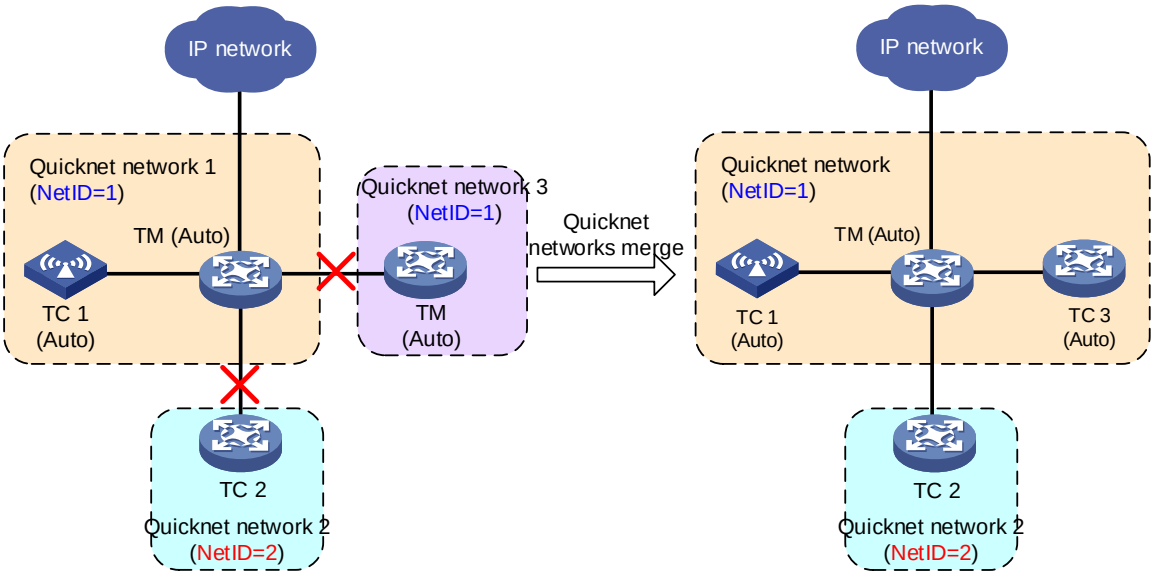
The following rules are followed during the Quicknet merging:

- For Quicknet networks that have already been established, only Quicknet networks with the same NetID can be merged. Quicknet networks with different NetIDs cannot be merged. Restoring factory settings or disabling and then enabling Quicknet can delete the NetID on a device.
- Established Quicknet networks can be merged with non-established Quicknet networks.
- When merging, the two Quicknet networks will compete for the TM role. For more information about the TM election rules, see "Quicknet role selection."

As shown in Figure 6, three Quicknet networks exist in the current network. Quicknet network 1 and Quicknet network 3 have the same NetID, and Quicknet network 2 has a different NetID. When the link failure is recovered, the Quicknet networks will try to merge, as follows:

- Quicknet network 1 has two member devices, which is greater than the number of member devices in Quicknet network 2. Therefore, Quicknet network 1 wins the election and its TM will become the TM of the merged Quicknet network.
- Quicknet network 2 cannot participate in the merge due to its NetID being inconsistent with other Quicknet networks. You can restore the devices to factory settings for them to join the merged Quicknet network.

Figure 6 Quicknet merging



Quicknet configuration synchronization

After the administrator completes the deployment through Quicknet, member devices form a Quicknet network through the Layer 2 network. The TM automatically establishes NETCONF sessions with all TCs through VLAN interface 1 for configuration deployment and parameter retrieval, achieving unified network management.

Quicknet supports the following configuration methods:

- Global configuration

  Administrators can configure settings on the Quicknet management system's web page or from the TM's CLI. These configurations are distributed to and take effect on all member devices. Global synchronized configurations include wireless, wired, Internet, Quicknet network-wide password, and FTP server settings.

- Batch configuration

  Administrators can configure multiple devices by selecting them on the Quicknet management system's web page. These configurations are distributed to and take effect on the selected member devices. Batch synchronization operations include batch reboot and factory reset.

- Single device configuration

  In addition to Quicknet global configuration and batch configuration, administrators can configure parameters for a single member device in the following methods:

  - Configure the settings for the specified device on the web page. These configurations will be distributed to and take effect on that device.
  - For configurations not supported by the web page, such as disabling/enabling interfaces or adding interfaces to VLANs, administrators must log in to the member device and use the CLI to configure the settings. To log in to a member device, use either of the following methods:
    - Use the eWEB or CLI in the Quicknet management system to access the local Web interface or CLI.
    - Use the IP address of the device to log in to the local Web interface or CLI of the device.

Quicknet time synchronization (NTP)

After the Quicknet network is deployed, the TM synchronizes the time to all TCs using Network Time Protocol (NTP). The time synchronization precision of NTP is in milliseconds, which can meet the time synchronization requirements of general computer networks. For more information about NTP, see "Configuring NTP."

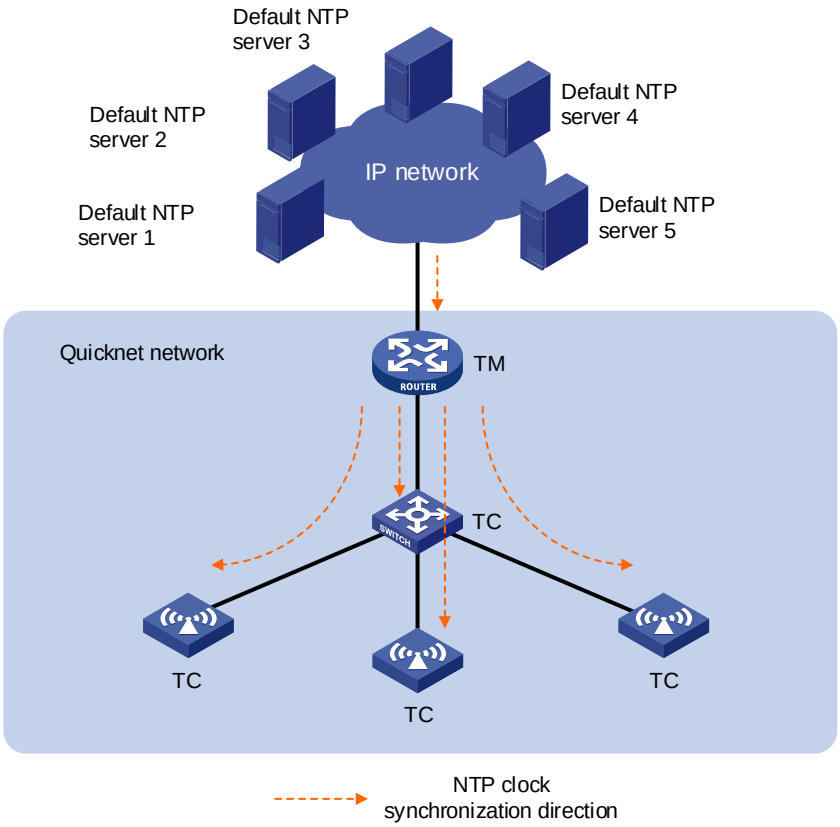
Quicknet synchronizes time settings across the network as follows:

1. When the Quicknet network is deployed, the TM automatically deletes the existing NTP configuration on the device and generates a new NTP configuration specific to the Quicknet network.
2. The TM acts as an NTP client and synchronizes time settings with external NTP servers.

   The device supports a default of four external NTP servers. When the device receives clock signals from multiple default NTP servers, it selects the optimal clock signal for synchronization. The addresses of the four clock servers are as follows:

   - s2f.time.edu.cn (Northeast network center)
   - cn.pool.ntp.org (NTP project domestic time server)
   - 1.cn.pool.ntp.org (NTP project domestic time server)
   - asia.pool.ntp.cn (NTP project Asian time server)
3. If all the four external NTP servers are unavailable, the TM modifies the level of the local clock to six and uses the local clock as the clock source for the Quicknet network.
4. The TM acts as an NTP server, and TCs act as NTP clients and synchronize time with the TM.

Figure 7 Quicknet time synchronization



Quicknet management system login with a fixed domain name

Typically, accessing the device's web interface requires its IP address or domain name. If you use a domain name, a DNS server must be deployed in the network and the domain-to-IP mapping must be configured. To simplify network deployment and user operations, Quicknet supports accessing the Quicknet management system using fixed domain name quicknet.h3c.com, even without a DNS server.

The TM automatically establishes a mapping between quicknet.h3c.com and the IP address of the TM's VLAN-interface 1 and synchronizes it with all member devices. Users can log into the Quicknet management system by connecting to any member device and entering the domain name.

Restrictions and guidelines: Quicknet configuration

The device starts up with factory settings. After the administrator completes initial deployment configuration from the Quicknet management system, the device can automatically establish a Quicknet network.

The devices exchange Quicknet protocol packets and establish and maintain the Quicknet network within VLAN 1. For Quicknet to operate normally, make sure the interfaces connecting member devices allow VLAN 1 packets to pass through (by default, the interfaces connecting member devices allow VLAN 1 packets to pass through).

Quicknet tasks at a glance

To configure Quicknet, perform the following tasks:

1. Enabling Quicknet
2. Setting the network-wide Quicknet password

   As a best practice to ensure inter-device communication, configure this feature.

Enabling Quicknet

Restrictions and guidelines

One Quicknet network can have only one TM.

If you switch the device role from TM to TC or disable Quicknet, the system clears all the running configuration related to Quicknet on the device.

Quicknet requires specific ACL resources. If ACL resources are insufficient, Quicknet might fail to be enabled. You can use the **`display acl`** command to view the ACL configuration and operation status and use the **`undo acl`** command to delete useless ACLs based on the actual conditions. After releasing ACL resources, you can try again to enable Quicknet. For more information about ACL, see ACL and QoS Configuration Guide.

When you enable Quicknet, the device checks whether port 80 and port 443 are occupied, as HTTP and HTTPS services require these ports. If port 80 or port 443 is occupied, Quicknet cannot be enabled.

Procedure

1. Enter system view.

system-view

2. Enable Quicknet. Perform either of the following tasks:
   - Enable Quicknet and specify the device role as TM.
     quicknet enable auto tm
   - Enable Quicknet and specify the device role as TC.
     quicknet tc enable

   For the S5500V3-HI switch series:

   If the device starts up with the initial configuration, Quicknet is disabled.

   If the device starts up with the factory defaults, Quicknet is enabled and the device role is TC.

   For the S5590-HI and S5590-EI switch series:

   By default, Quicknet is disabled and the device role is not specified.

Setting the network-wide Quicknet password

About this task

The TM manages and maintains each TC through NETCONF channels, and uses the local username admin and the default password to perform TC authentication. You can perform this task to edit the default password on the TM. The TM then issues the changed password to all TCs and re-establishes NETCONF channels.

Procedure

1. Enter system view.
   system-view
2. Set the network-wide Quicknet password.
   quicknet password [**cipher**] *string*

   For the S5500V3-HI switch series:

   If the device starts up with the initial configuration, the network-wide Quicknet password is not set.

   If the device starts up with the factory defaults, the default username is admin, the default network-wide Quicknet password is h3capadmin for APs, and the default network-wide Quicknet password is admin for other devices.

   For the S5590-HI and S5590-EI switch series:

   By default, the network-wide Quicknet password is not set.

Display and maintenance commands for Quicknet

Execute **display** commands in any view.

Table 1 Display and maintenance commands for Quicknet

Step	Command
Display Quicknet configuration.	**display quicknet configuration**
Display member device information.	**display quicknet tc** [*tc-id*] [**verbose**]

Quicknet configuration examples

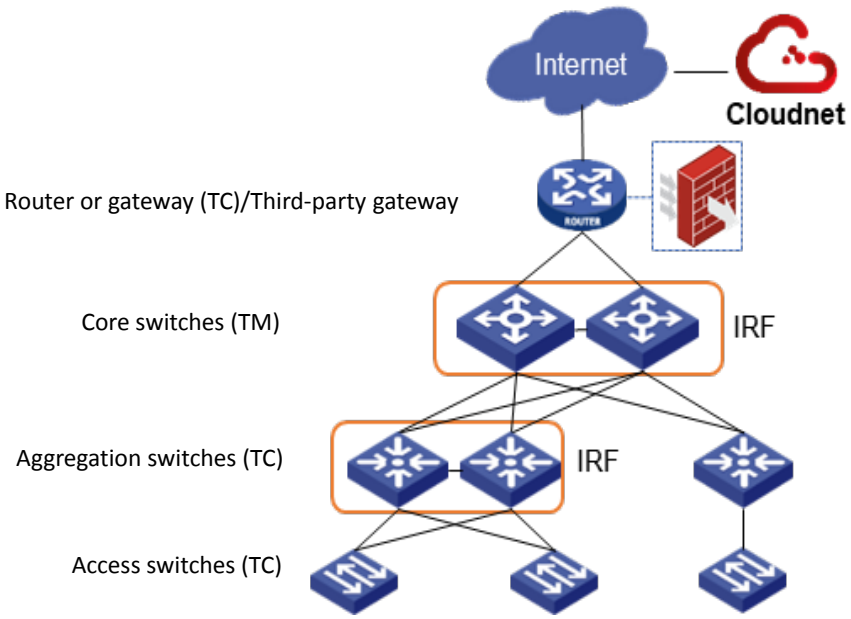
Example: Configuring a wired Quicknet network

Network configuration

As shown in Figure 8, the network contains devices of the following types: router, firewall, core switch, aggregation switch, and access switch. The gateway also acts as the DHCP server to assign an IP address to VLAN-interface 1 of each LAN device. The gateway comes with the DHCP server feature enabled, with the IP address as 192.168.0.1/24 and IP address range for allocation as 192.168.0.0/24 (excluding 192.168.0.1).

Configure Quicknet for devices to automatically form a network after startup and use the Quicknet management system to perform unified management.

Figure 8 Network diagram



Configuration requirements

In the wired network:

- Use the core switch as the TM and other devices as TCs.
- Use the router, gateway, or a third-party device as the gateway.
- If the firewall acts as the gateway, configure the firewall to operate in routing mode and communicate with the core switch in Layer 2 transparent mode.
- Before the initial deployment, manually specify the core switch as the TM and complete the initial settings.

As a best practice, use one-key onboarding to register devices on Cloudnet after the initial deployment is complete, thus remote O&M can be performed from the Cloudnet platform.

Prerequisites

1. Connect the TCs.

   Connect the TM and TCs through Layer 2 interfaces. Make sure the interfaces permit VLAN 1 traffic. By default, all Layer 2 interfaces on the devices permit VLAN 1 traffic.

2. Connect the PC.

Connect your PC to the Quicknet network to log in to the Quicknet management system. Use a network cable to connect the PC to a LAN port on any member device and add the PC to VLAN 1.

Configuring the PC

1. Configure the IP address of the PC:

   As a best practice, use the dynamic obtaining method. Configure the PC to automatically obtain an IP address and the DNS server address. To specify a static IP address, make sure the IP address is on the same subnet as the IP address of VLAN-interface 1 on the connected device.

2. Configure DNS:

   The PC stores a DNS mapping file. Add a mapping between the TM IP address and quicknet.h3c.com to this file. In Windows, the file name is hosts. You can find it at C:\Windows\System32\drivers\etc. Open the document in a text editor and add the following line:

   192.168.1.1 quicknet.h3c.com

   Here, 192.168.1.1 is the IP address of VLAN interface 1 for the TM in the network.

Configure the Quicknet management system

1. Use the Quicknet management system to perform initial deployment.

   Enter quicknet.h3c.com in the address bar of a browser to log in to the Quicknet management system. Complete deployment configuration as instructed to set up a Quicknet network.

NOTE:

Upon the first login to Quicknet, it takes 10 seconds or longer for the system to collect topology information. This time period increases with the number of devices in the network.

Figure 9 Login page of the Quicknet management system

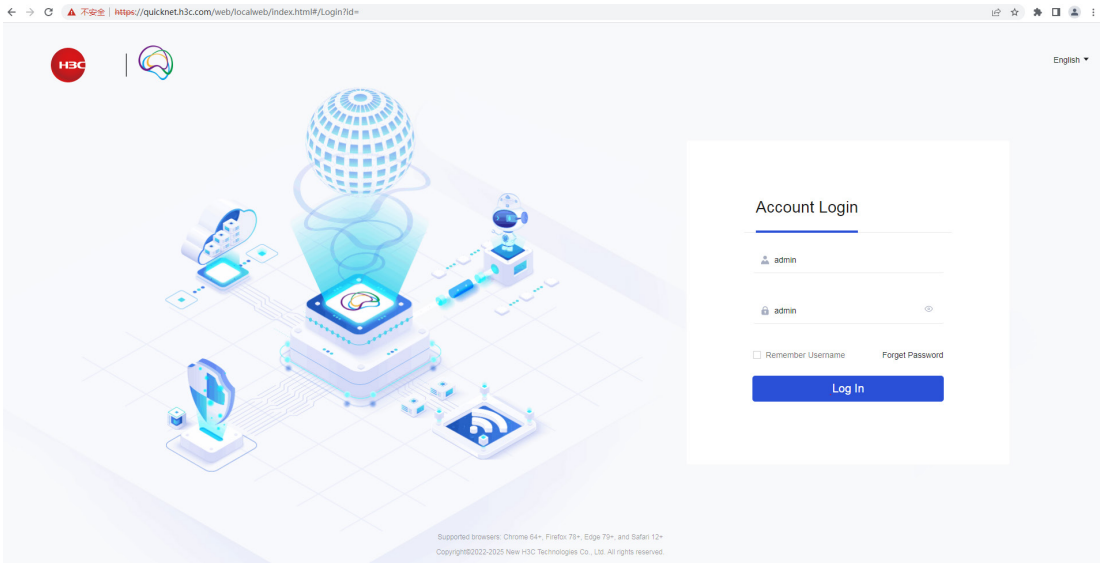
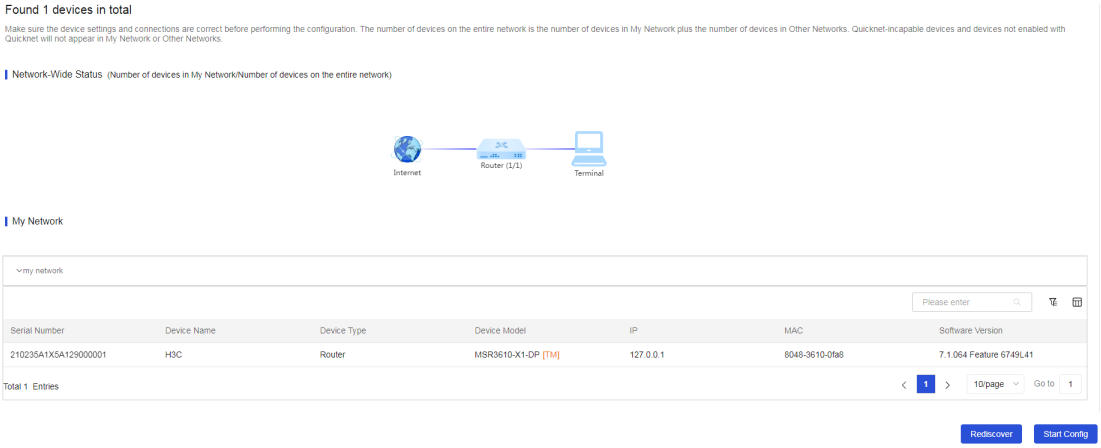
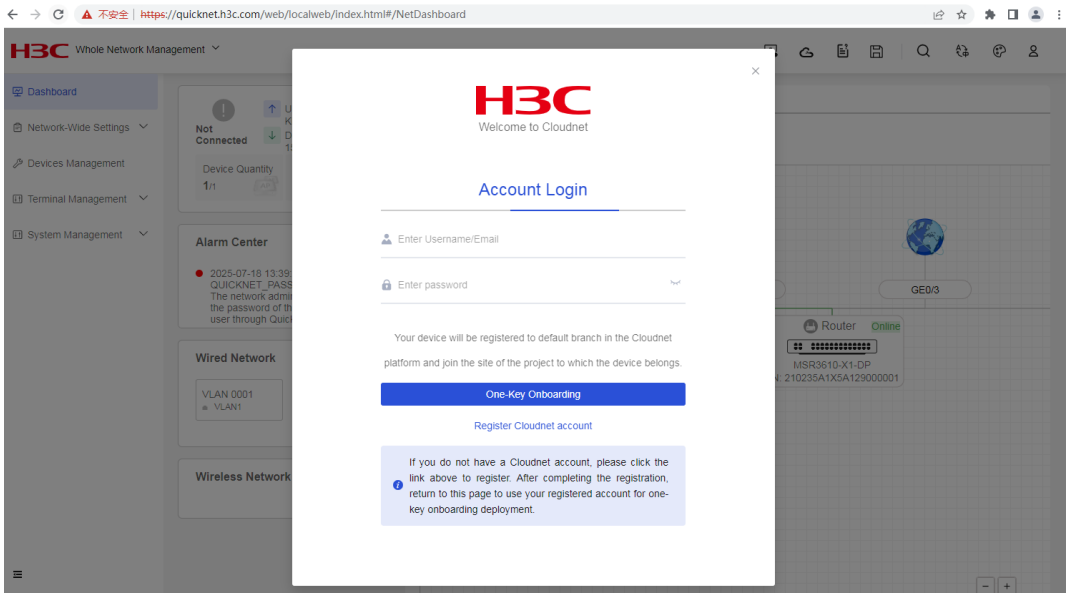


Figure 10 Homepage of the Quicknet management system



2. (Optional.) Perform one-key onboarding for all devices in the Quicknet network.
   a. Click the One-Key Onboarding icon at the top right corner of the page. The Cloudnet login window opens.

      Figure 11 One-key onboarding

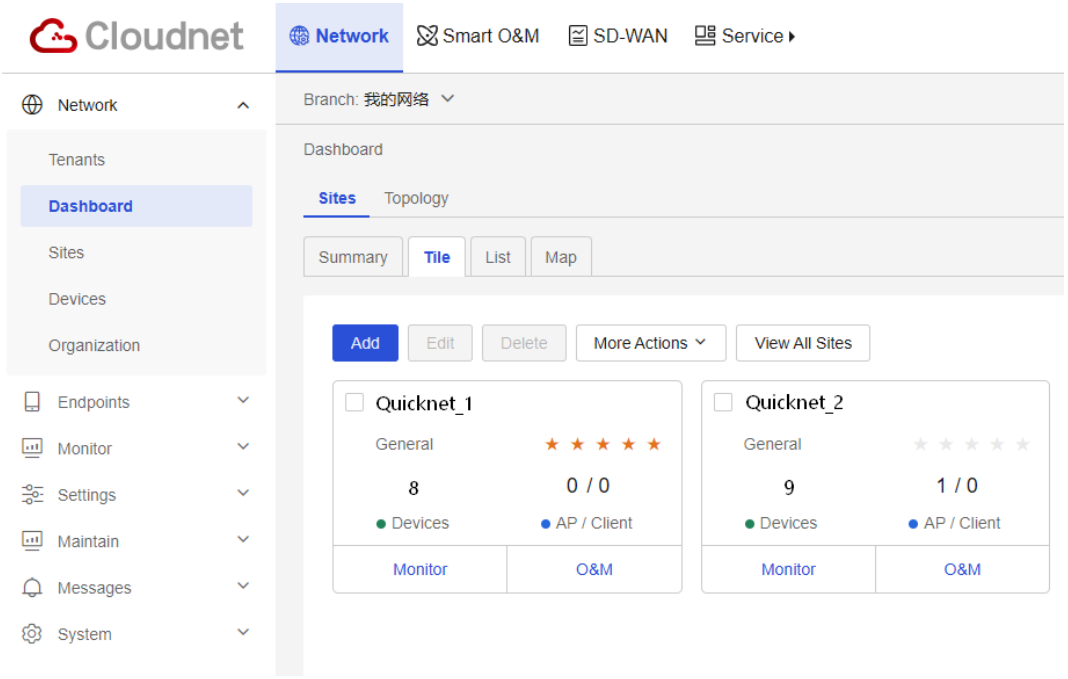


   b. Enter your Cloudnet account, and click One-Key Onboarding to access the Cloudnet dashboard page.

      If you do not have a Cloudnet account, register one first and then try to log in.

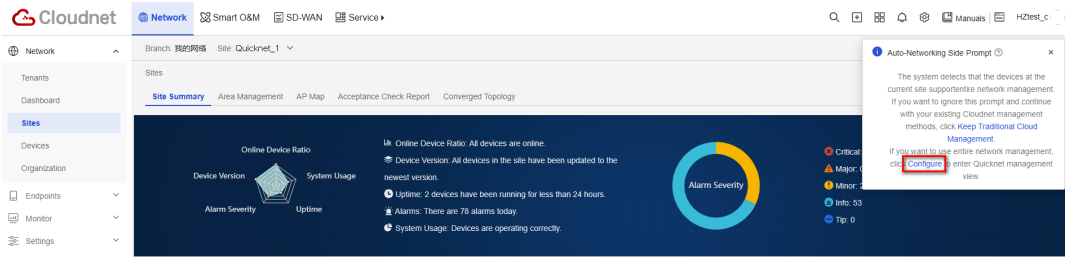
   c. The Cloudnet dashboard displays all sites created by the login user. Click the site to which the Quicknet network belongs to access the site management page. The site name is the same as the project name specified during Quicknet deployment.

Figure 12 Cloudnet dashboard



d. Click to access the configuration page as prompted at the upper right corner of the page. Manage all devices in the Quicknet network as needed.

Figure 13 Cloudnet site page



Verifying the configuration

Verify that you can view the network topology and information from the management system after the initial deployment is completed.

Configuration files

None (use the factory settings).

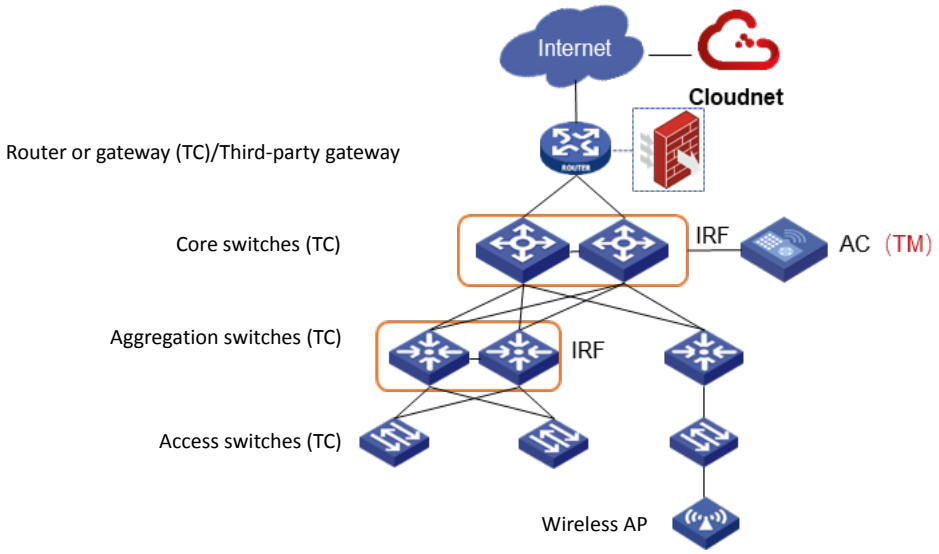
Example: Configuring a wired and wireless Quicknet network

Network configuration

As shown in Figure 14, the network contains devices of the following types: router, firewall, core switch, aggregation switch, access switch, AC, and AP. The gateway also acts as the DHCP server to assign an IP address to VLAN-interface 1 of each LAN device. The gateway comes with the DHCP server feature enabled, with the IP address as 192.168.0.1/24 and IP address range for allocation as 192.168.0.0/24 (excluding 192.168.0.1).

Configure Quicknet for devices to automatically form a network after startup and to achieve unified management.

Figure 14 Network diagram



Configuration requirements

In the wired + wireless network:

- The AC acts as the TM and other Quicknet-capable devices act as TCs.
- Use the router, gateway, or a third-party device as the gateway.
- If the firewall acts as the gateway, configure the firewall to operate in routing mode and communicate with the core switch in Layer 2 transparent mode.

As a best practice, use one-key onboarding to register devices on Cloudnet after the initial deployment is complete, thus remote O&M can be performed from the Cloudnet platform.

Prerequisites

1. Connect the TC.

   Connect the TM and TCs through Layer 2 interfaces. Make sure the interfaces permit VLAN 1 traffic. By default, all Layer 2 interfaces on the devices permit VLAN 1 traffic.

2. Connect the PC.

   Connect your PC to the Quicknet network to log in to the Quicknet management system:

   - Wired connection: Use a network cable to connect the PC to a LAN port on any member device and add the PC to VLAN 1.
   - Wireless connection: Connect the PC to the Wi-Fi whose name includes the Quicknet string, and specify a static IP address for the PC or configure the PC to obtain an IP address automatically to add the PC to VLAN 1.

Configuring the PC

1. Configure the IP address of the PC:

   As a best practice, use the dynamic obtaining method. Configure the PC to automatically obtain an IP address and the DNS server address. To specify a static IP address, make sure the IP address is on the same subnet as the IP address of VLAN-interface 1 on the connected device.

2. Configuring the DNS

   The PC stores a DNS mapping file. Add a mapping between the TM IP address and quicknet.h3c.com to this file. In Windows, the file name is hosts. You can find it at C:\Windows\System32\drivers\etc. Open the document in a text editor and add the following line:

192.168.1.1 quicknet.h3c.com

Here, 192.168.1.1 is the IP address of VLAN interface 1 for the TM in the network.

Configure the Quicknet management system

1. Use the Quicknet management system to perform initial deployment.

   Enter quicknet.h3c.com in the address bar of a browser to log in to the Quicknet management system. Complete deployment configuration as instructed to set up a Quicknet network.

NOTE:

Upon the first login to Quicknet, it takes 10 seconds or longer for the system to collect topology information. This time period increases with the number of devices in the network.

Figure 15 Login page of the Quicknet management system

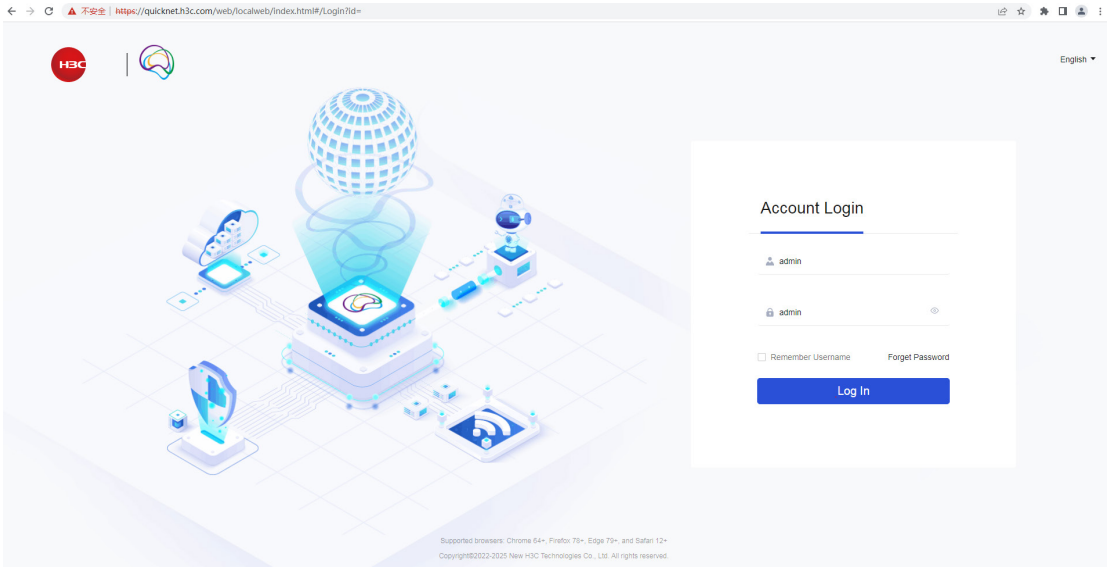


Figure 16 Homepage of the Quicknet management system



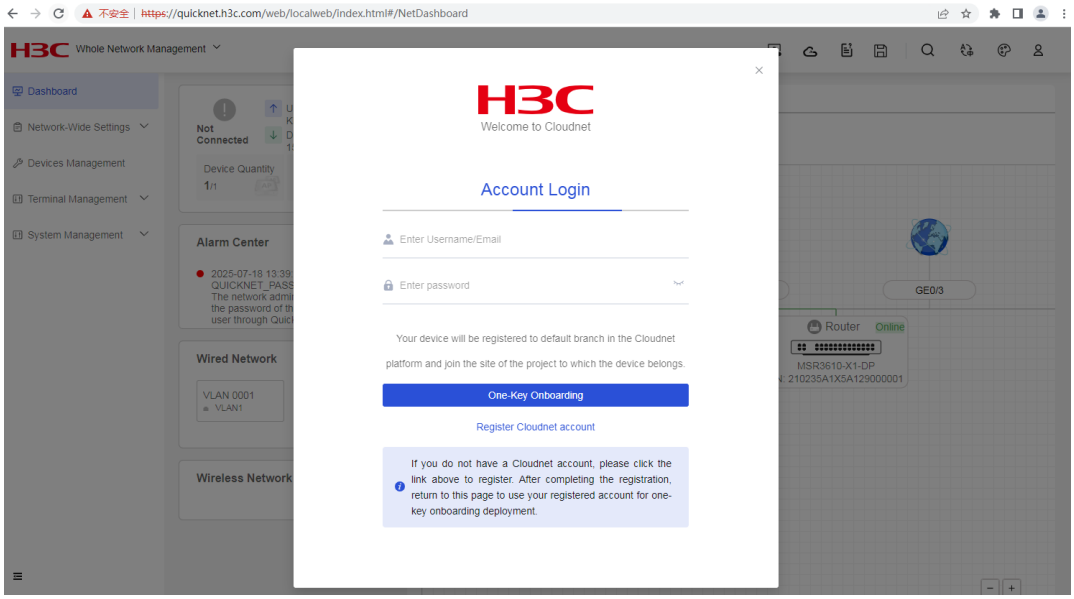
2. Manage the AP license and software version.

   If the AC does not have sufficient license seats, fit APs cannot register on the AC. As a best practice to ensure that the AC has sufficient license seats, click Continue Config after completing the deployment configuration, and go to the AP license management page to install AP licenses. Then, enter the AP software management page and upgrade the AP software as needed.

3. (Optional.) Perform one-key onboarding for all devices in the Quicknet network.

   a. Click the One-Key Onboarding icon at the top right corner of the page. The Cloudnet login window opens.

   Figure 17 One-key onboarding

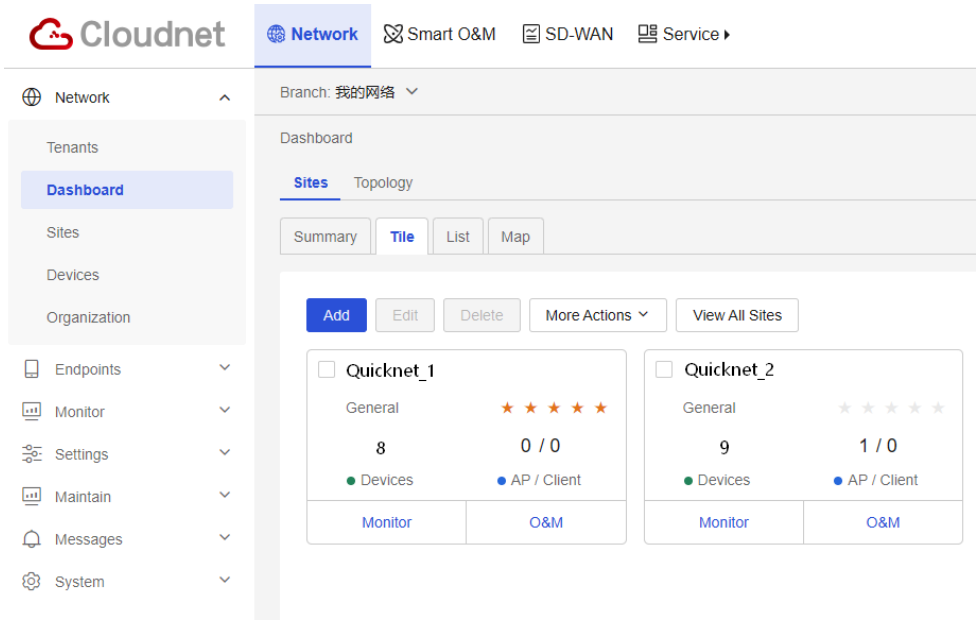


   b. Enter your Cloudnet account, and click One-Key Onboarding to access the Cloudnet dashboard page.

      If you do not have a Cloudnet account, register one first and then try to log in.

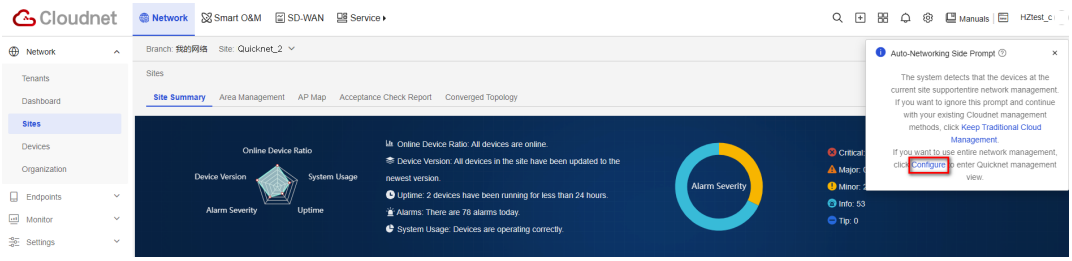
   c. The Cloudnet dashboard displays all sites created by the login user. Click the site to which the Quicknet network belongs to access the site management page. The site name is the same as the project name specified during Quicknet deployment.

   Figure 18 Cloudnet dashboard



   d. Click to access the configuration page as prompted at the upper right corner of the page. Manage all devices in the Quicknet network as needed.

Figure 19 Cloudnet site page



Verifying the configuration

On the system summary page of the Quicknet management system, verify that you can view the topology of all member devices and the related information.

Configuration files

None (use the factory settings).

Quicknet configuration examples

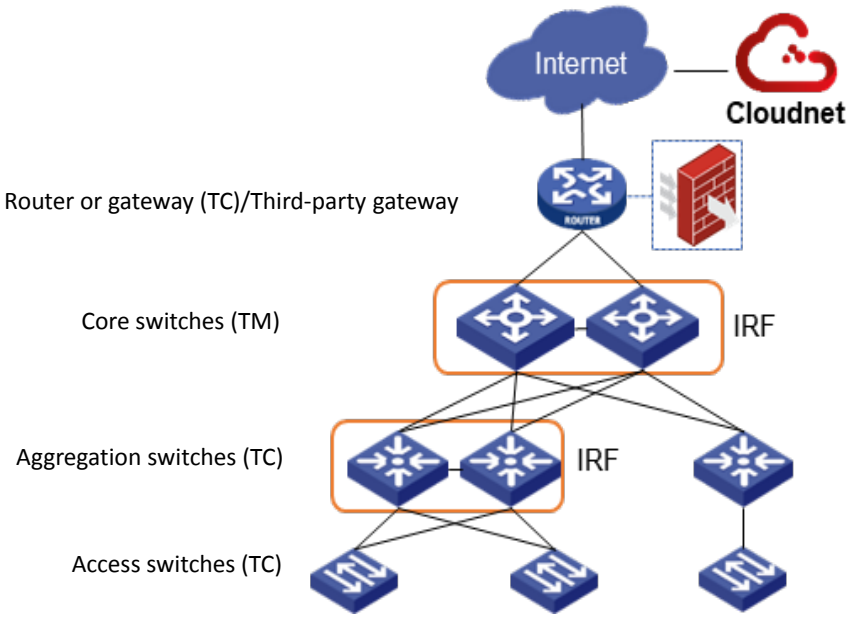
Example: Configuring a wired Quicknet network

Network configuration

As shown in Figure 20, the network contains devices of the following types: router, firewall, core switch, aggregation switch, and access switch. The gateway also acts as the DHCP server to assign an IP address to VLAN-interface 1 of each LAN device. The gateway comes with the DHCP server feature enabled, with the IP address as 192.168.0.1/24 and IP address range for allocation as 192.168.0.0/24 (excluding 192.168.0.1).

Configure Quicknet for devices to automatically form a network after startup and use the Quicknet management system to perform unified management.

Figure 20 Network diagram



Configuration requirements

In the wired network:

- Use the core switch as the TM and other devices as TCs.
- Use the router, gateway, or a third-party device as the gateway.
- If the firewall acts as the gateway, configure the firewall to operate in routing mode and communicate with the core switch in Layer 2 transparent mode.
- Before the initial deployment, manually specify the core switch as the TM and complete the initial settings.

As a best practice, use one-key onboarding to register devices on Cloudnet after the initial deployment is complete, thus remote O&M can be performed from the Cloudnet platform.

Prerequisites

1. Connect the TC.

   Connect the TM and TCs through Layer 2 interfaces. Make sure the interfaces permit VLAN 1 traffic. By default, all Layer 2 interfaces on the devices permit VLAN 1 traffic.

2. Connect the PC.

   Connect your PC to the Quicknet network to log in to the Quicknet management system. Use a network cable to connect the PC to a LAN port on any member device and add the PC to VLAN 1.

Configuring the TM

Enable the Web function so users can log in to the Quicknet management system.

```
<Sysname> system-view
[Sysname] sysname TM
[TM] ip http enable
[TM] ip https enable
```

Enable the LLDP function so that the TM can autodiscover TCs through Layer 2 protocols.

```
[TM] lldp global enable
```

Enable Telnet and NETCONF functions so that users can then send configuration to TCs from the console on the TM.

```
[TM] telnet server enable
Telnet is insecure because it transmits data in plaintext form.
[TM] netconf soap http enable
[TM] netconf ssh server enable
```

Specify the local username as admin and the password as admin. The TM, the local device, and TCs use the username admin and password admin for authentication when establishing NETCONF sessions. Devices that fail authentication cannot be managed by the TM. During deployment, the TM will replace this password with a more complex administrative password.

```
[TM] local-user admin
[TM-luser-manage-admin] password-control length 4
[TM-luser-manage-admin] password-control composition type-number 1 type-length 1
[TM-luser-manage-admin] undo password-control complexity user-name check
[TM-luser-manage-admin] password simple admin
[TM-luser-manage-admin] service-type telnet http https ssh
[TM-luser-manage-admin] authorization-attribute user-role network-admin
[TM-luser-manage-admin] quit
[TM] ssh user admin service-type netconf authentication-type publickey assign pub
lickey -smartmc-publickey-
[TM] line vty 0 63
```

```
[TM-line-vty0-63] authentication-mode scheme
[TM-line-vty0-63] quit
```

For security purposes, the password control function is enabled by default. This function requires strong passwords with sufficient length and complexity. Configure password control parameters before you set the password to admin. For more information about password control, see Security Command Reference.

Configure a DHCP address pool and apply it to VLAN-interface 1. (Quicknet uses VLAN 1 to communicate with member devices.)

```
[TM] dhcp enable
[TM] dhcp server ip-pool 1
[TM-dhcp-pool-1] network 192.168.1.1 24
[TM-dhcp-pool-1] gateway-list 192.168.1.1
[TM-dhcp-pool-1] forbidden-ip 192.168.1.1
[TM-dhcp-pool-1] quit
[TM] interface Vlan-interface 1
[TM-Vlan-interface1] ip address  192.168.1.1 24
[TM-Vlan-interface1] dhcp server apply ip-pool 1
[TM-Vlan-interface1] quit
```

Enable Quicknet and specify the device role as TM.

```
[TM] quicknet enable auto tm
```

Configuring the TC

NOTE:

If your TC model comes with Quicknet configuration in factory settings, you only need to power on the TCs and connect the cables. No extra configuration is required. (The model of TC comes with the Quicknet function enabled, along with the DHCP configuration required for Quicknet networking.)

Enable the Web function so users can log in to the TC Web management platform through the eWEB function of the Quicknet management system.

```
<Sysname> system-view
[Sysname] sysname TC1
[TC1] ip http enable
[TC1] ip https enable
```

Enable the LLDP function so that the TM can autodiscover TCs through Layer 2 protocols.

```
[TC1] lldp global enable
```

Enable Telnet and NETCONF functions so that users can then send configuration to TCs from the console on the TM.

```
[TC1] telnet server enable
Telnet is insecure because it transmits data in plaintext form.
[TC1] netconf soap http enable
[TC1] netconf ssh server enable
```

Specify the local username as admin and the password as admin. The TM, the local device, and TCs use the username admin and password admin for authentication when establishing NETCONF sessions. Devices that fail authentication cannot be managed by the TM. During deployment, the TM will replace this password with a more complex administrative password.

```
[TC1] local-user admin
[TC1-luser-manage-admin] password-control length 4
[TC1-luser-manage-admin] password-control composition type-number 1 type-length 1
```

```
[TC1-luser-manage-admin] undo password-control complexity user-name check
[TC1-luser-manage-admin] password simple admin
[TC1-luser-manage-admin] service-type telnet http https ssh
[TC1-luser-manage-admin] authorization-attribute user-role network-admin
[TC1-luser-manage-admin] quit
[TC1] ssh user admin service-type netconf authentication-type publickey assign pub
lickey -smartmc-publickey-
[TC1] line vty 0 63
[TC1-line-vty0-63] authentication-mode scheme
[TC1-line-vty0-63] quit
```

For security purposes, the password control function is enabled by default. This function requires strong passwords with sufficient length and complexity. Configure password control parameters before you set the password to admin. For more information about password control, see Security Command Reference.

Configure VLAN-interface 1 to automatically obtain an IP address through DHCP. (Quicknet uses VLAN 1 to communicate with member devices.)

```
[TC1] interface vlan-interface 1
[TC1-Vlan-interface1] ip address dhcp-alloc
[TC1-Vlan-interface1] quit
```

Enable Quicknet and specify the device role as TC.

```
[TC1] quicknet tc enable
```

Configuring the PC

1. Configure the IP address of the PC:

   As a best practice, use the dynamic obtaining method. Configure the PC to automatically obtain an IP address and the DNS server address. To specify a static IP address, make sure the IP address is on the same subnet as the IP address of VLAN-interface 1 on the connected device.

2. Configuring the DNS

   The PC stores a DNS mapping file. Add a mapping between the TM IP address and quicknet.h3c.com to this file. In Windows, the file name is hosts. You can find it at C:\Windows\System32\drivers\etc. Open the document in a text editor and add the following line:

   192.168.1.1 quicknet.h3c.com

   Here, 192.168.1.1 is the IP address of VLAN interface 1 for the TM in the network.

Configure the Quicknet management system

1. Use the Quicknet management system to perform initial deployment.

   Enter quicknet.h3c.com in the address bar of a browser to log in to the Quicknet management system. Complete deployment configuration as instructed to set up a Quicknet network.

NOTE:

Upon the first login to Quicknet, it takes 10 seconds or longer for the system to collect topology information. This time period increases with the number of devices in the network.

Figure 21 Login page of the Quicknet management system

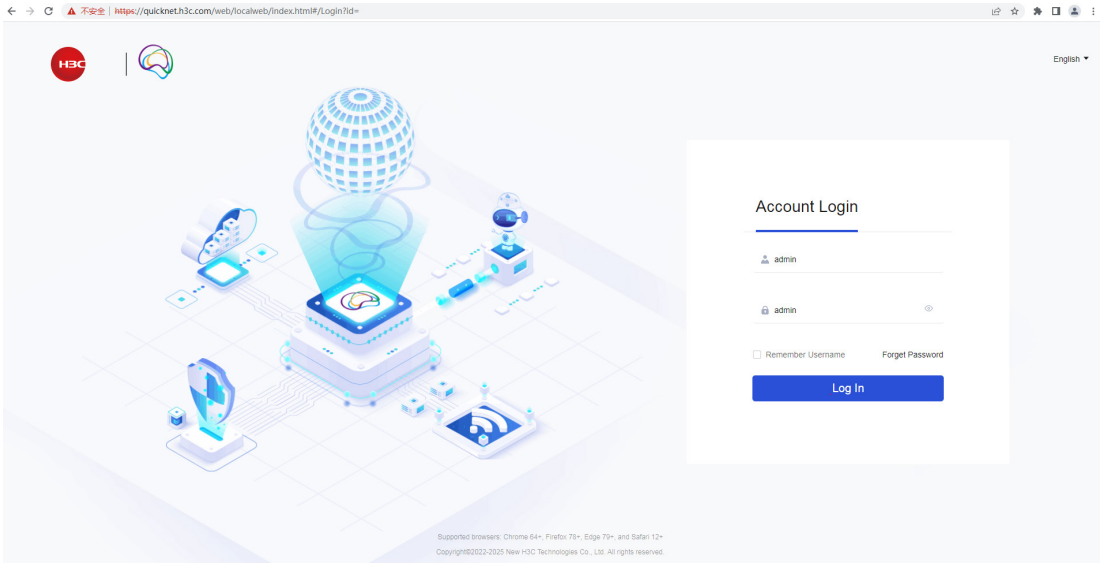


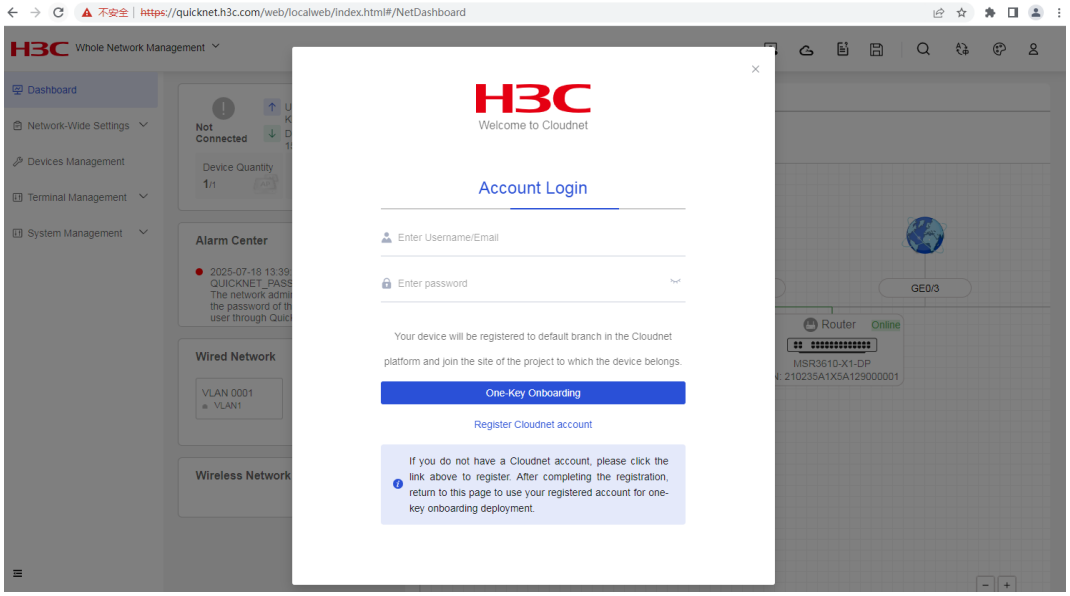
Figure 22 Homepage of the Quicknet management system



2. (Optional.) Perform one-key onboarding for all devices in the Quicknet network.

a. Click the One-Key Onboarding icon at the top right corner of the page. The Cloudnet login window opens.

Figure 23 One-key onboarding

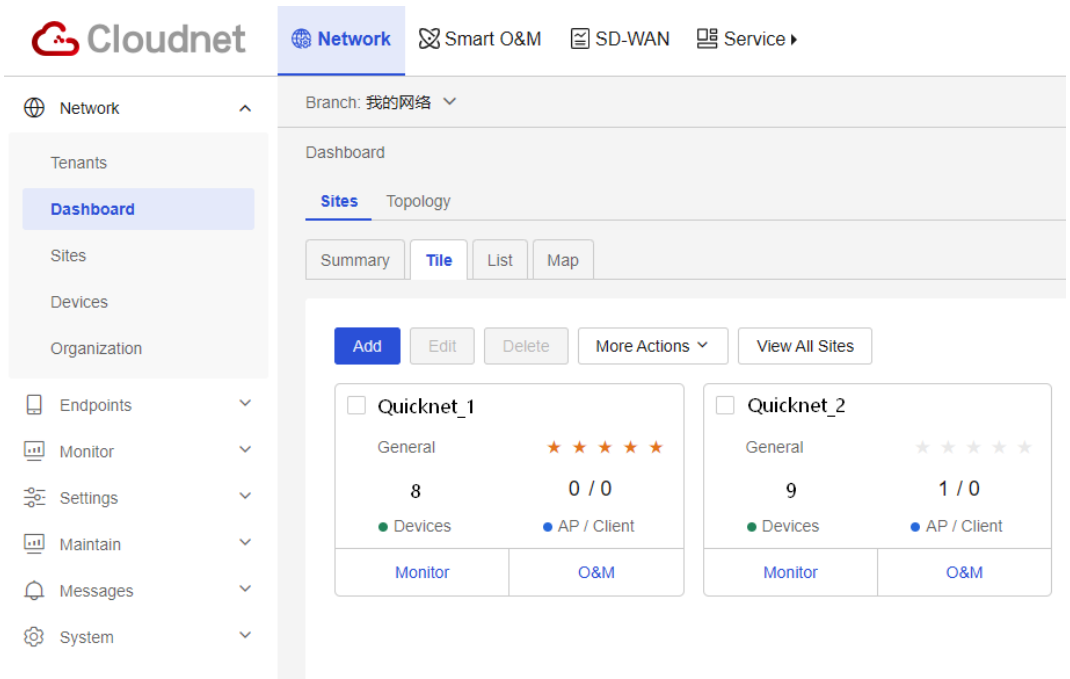


b. Enter your Cloudnet account, and click One-Key Onboarding to access the Cloudnet dashboard page.

If you do not have a Cloudnet account, register one first and then try to log in.

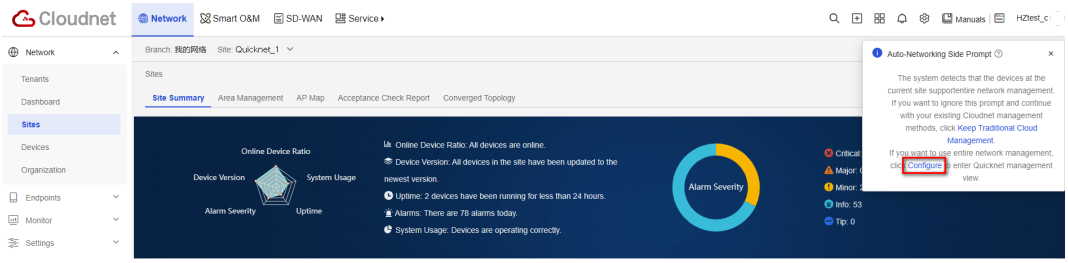
c. The Cloudnet dashboard displays all sites created by the login user. Click the site to which the Quicknet network belongs to access the site management page. The site name is the same as the project name specified during Quicknet deployment.

Figure 24 Cloudnet dashboard



d. Click to access the configuration page as prompted at the upper right corner of the page. Manage all devices in the Quicknet network as needed.

Figure 25 Cloudnet site page



Verifying the configuration

Verify that you can view the network topology and information from the management system after the initial deployment is completed.

Configuration files

- TM configuration file (includes some settings that the device automatically generates to support the Quicknet function after you enable it)

```
#
 sysname TM
#
 telnet server enable
#
 dhcp enable
#
 dns fast-reply enable
 ip host quicknet.h3c.com 192.168.1.1
#
 lldp global enable
#
dhcp server ip-pool 1
 gateway-list 192.168.1.1
 network 192.168.1.0 mask 255.255.255.0
 forbidden-ip 192.168.1.1
#
interface Vlan-interface1
 ip address 192.168.1.1 255.255.255.0
 dhcp server apply ip-pool 1
#
line vty 0 63
 authentication-mode scheme
 user-role network-operator
#
 ssh user admin service-type netconf authentication-type publickey assign publickey
-smartmc-publickey-
#
local-user admin class manage
 password hash
$h$6$13g0MKsz6iBwZC77$MT90PviiKQpRXDjseCSlbg1wjE0/lgR10jU9g2LjCRUGhCbWO8bbvjb6EfD
11kqBaqA77NBn+A/4awgfccbw7g==
 service-type telnet http https ssh
```

```
 authorization-attribute user-role network-admin
 authorization-attribute user-role network-operator
 password-control length 4
 password-control composition type-number 1 type-length 1
 undo password-control complexity user-name check
#
 netconf soap http enable
 netconf soap https enable
 netconf ssh server enable
#
 ip http enable
 ip https enable
#
 quicknet enable auto tm
 quicknet password cipher $c$3$DvAxSmoVCgX9H4fD39DN9/HKvZxALMTLMzo20m4A
#
return
```

- TC configuration file (includes some settings that the device automatically generates to support the Quicknet function after you enable it)

```
#
 sysname TC
#
 dns fast-reply enable
 ip host quicknet.h3c.com 192.168.1.1
#
 lldp global enable
#
interface Vlan-interface1
 ip address 192.168.1.2 255.255.255.0
#
line vty 0 63
 authentication-mode scheme
#
 ssh user admin service-type netconf authentication-type publickey assign publickey
-smartmc-publickey-
#
local-user admin class manage
 password hash
$h$6$PyB2gjZbpLLL5euH$5eznb/M9pKoAqDHdcv3MscydHojnwphdLlxG6fEj9T1Tp/+MAuuycTT8HfE
T/FRCEzZSu5EHH8Lb1EpMwzIUGA==
 service-type telnet http https ssh
 authorization-attribute user-role network-admin
 authorization-attribute user-role network-operator
 password-control length 4
 password-control composition type-number 1 type-length 1
 undo password-control complexity user-name check
#
public-key peer -smartmc-publickey-
 public-key-code begin
```

```
    30819F300D06092A864886F70D010101050003818D0030818902818100C69D96A6F187859B
    CFA359C257041A0AD5AF7F29F5AC06D3A0DFE22286DF0439BC610D062D493A2A73BF6BCED0
    9A31E234331B1CABE94AF0D346054230EC8E7E17BC341FEBE0D87498E148D521254ED0880B
    F5C03D472146FBE531B54E7F1A4A88A91FA011E3101358BEF957F9EF467955837761AC3D9E
    5841196EB0FC1070BF0203010001
  public-key-code end
  peer-public-key end
#
  netconf soap http enable
  netconf ssh server enable
#
  ip http enable
  ip https enable
#
  quicknet tc enable
#
return
```

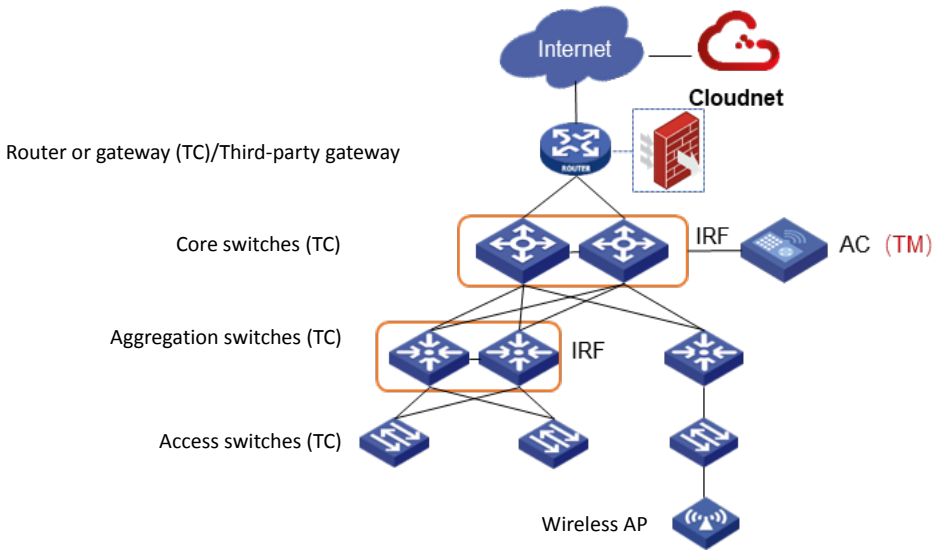
Example: Configuring a wired and wireless Quicknet network

Network configuration

As shown in Figure 26, the network contains devices of the following types: router, firewall, core switch, aggregation switch, access switch, AC, and AP. The gateway also acts as the DHCP server to assign an IP address to VLAN-interface 1 of each LAN device. The gateway comes with the DHCP server feature enabled, with the IP address as 192.168.0.1/24 and IP address range for allocation as 192.168.0.0/24 (excluding 192.168.0.1).

Configure Quicknet for devices to automatically form a network after startup and to achieve unified management.

Figure 26 Network diagram



Configuration requirements

In the wired + wireless network:

- The AC acts as the TM and other Quicknet-capable devices act as TCs.
- Use the router, gateway, or a third-party device as the gateway.
- If the firewall acts as the gateway, configure the firewall to operate in routing mode and communicate with the core switch in Layer 2 transparent mode.

As a best practice, use one-key onboarding to register devices on Cloudnet after the initial deployment is complete, thus remote O&M can be performed from the Cloudnet platform.

Prerequisites

1. Connect the TC.

Connect the TM and TCs through Layer 2 interfaces. Make sure the interfaces permit VLAN 1 traffic. By default, all Layer 2 interfaces on the devices permit VLAN 1 traffic.

2. Connect the PC.

Connect your PC to the Quicknet network to log in to the Quicknet management system:

- Wired connection: Use a network cable to connect the PC to a LAN port on any member device and add the PC to VLAN 1.
- Wireless connection: Connect the PC to the Wi-Fi whose name includes the Quicknet string, and specify a static IP address for the PC or configure the PC to obtain an IP address automatically to add the PC to VLAN 1.

Configuring the TM

Enable the Web function so users can log in to the Quicknet management system.

```
<Sysname> system-view
[Sysname] sysname TM
[TM] ip http enable
[TM] ip https enable
```

Enable the LLDP function so that the TM can autodiscover TCs through Layer 2 protocols.

```
[TM] lldp global enable
```

Enable Telnet and NETCONF functions so that users can then send configuration to TCs from the console on the TM.

```
[TM] telnet server enable
Telnet is insecure because it transmits data in plaintext form.
[TM] netconf soap http enable
[TM] netconf ssh server enable
```

Specify the local username as admin and the password as admin. The TM, the local device, and TCs use the username admin and password admin for authentication when establishing NETCONF sessions. Devices that fail authentication cannot be managed by the TM. During deployment, the TM will replace this password with a more complex administrative password.

```
[TM] local-user admin
[TM-luser-manage-admin] password-control length 4
[TM-luser-manage-admin] password-control composition type-number 1 type-length 1
[TM-luser-manage-admin] undo password-control complexity user-name check
[TM-luser-manage-admin] password simple admin
[TM-luser-manage-admin] service-type telnet http https ssh
[TM-luser-manage-admin] authorization-attribute user-role network-admin
[TM-luser-manage-admin] quit
[TM] ssh user admin service-type netconf authentication-type publickey assign pub
lickey -smartmc-publickey-
[TM] line vty 0 63
[TM-line-vty0-63] authentication-mode scheme
```

```
[TM-line-vty0-63] quit
```

For security purposes, the password control function is enabled by default. This function requires strong passwords with sufficient length and complexity. Configure password control parameters before you set the password to admin. For more information about password control, see Security Command Reference.

Configure a DHCP address pool and apply it to VLAN-interface 1. (Quicknet uses VLAN 1 to communicate with member devices.)

```
[TM] dhcp enable
[TM] dhcp server ip-pool 1
[TM-dhcp-pool-1] network 192.168.1.1 24
[TM-dhcp-pool-1] gateway-list 192.168.1.1
[TM-dhcp-pool-1] forbidden-ip 192.168.1.1
[TM-dhcp-pool-1] quit
[TM] interface Vlan-interface 1
[TM-Vlan-interface1] ip address  192.168.1.1 24
[TM-Vlan-interface1] dhcp server apply ip-pool 1
[TM-Vlan-interface1] quit
```

Enable Quicknet and specify the device role as TM.

```
[TM] quicknet enable auto tm
```

Configuring the TC

NOTE:

If your TC model comes with Quicknet configuration in factory settings, you only need to power on the TCs and connect the cables. No extra configuration is required. (The model of TC comes with the Quicknet function enabled, along with the DHCP configuration required for Quicknet networking.)

Enable the Web function so users can log in to the TC Web management platform through the eWEB function of the Quicknet management system.

```
<Sysname> system-view
[Sysname] sysname TC1
[TC1] ip http enable
[TC1] ip https enable
```

Enable the LLDP function so that the TM can autodiscover TCs through Layer 2 protocols.

```
[TC1] lldp global enable
```

Enable Telnet and NETCONF functions so that users can then send configuration to TCs from the console on the TM.

```
[TC1] telnet server enable
Telnet is insecure because it transmits data in plaintext form.
[TC1] netconf soap http enable
[TC1] netconf ssh server enable
```

Specify the local username as admin and the password as admin. The TM, the local device, and TCs use the username admin and password admin for authentication when establishing NETCONF sessions. Devices that fail authentication cannot be managed by the TM. During deployment, the TM will replace this password with a more complex administrative password.

```
[TC1] local-user admin
[TC1-luser-manage-admin] password-control length 4
[TC1-luser-manage-admin] password-control composition type-number 1 type-length 1
[TC1-luser-manage-admin] undo password-control complexity user-name check
```

```
[TC1-luser-manage-admin] password simple admin
[TC1-luser-manage-admin] service-type telnet http https ssh
[TC1-luser-manage-admin] authorization-attribute user-role network-admin
[TC1-luser-manage-admin] quit
[TC1] ssh user admin service-type netconf authentication-type publickey assign pub
lickey -smartmc-publickey-
[TC1] line vty 0 63
[TC1-line-vty0-63] authentication-mode scheme
[TC1-line-vty0-63] quit
```

For security purposes, the password control function is enabled by default. This function requires strong passwords with sufficient length and complexity. Configure password control parameters before you set the password to admin. For more information about password control, see Security Command Reference.

Configure VLAN-interface 1 to automatically obtain an IP address through DHCP. (Quicknet uses VLAN 1 to communicate with member devices.)

```
[TC1] interface vlan-interface 1
[TC1-Vlan-interface1] ip address dhcp-alloc
[TC1-Vlan-interface1] quit
```

Enable Quicknet and specify the device role as TC.

```
[TC1] quicknet tc enable
```

Configuring the PC

1. Configure the IP address of the PC:

   As a best practice, use the dynamic obtaining method. Configure the PC to automatically obtain an IP address and the DNS server address. To specify a static IP address, make sure the IP address is on the same subnet as the IP address of VLAN-interface 1 on the connected device.

2. Configuring the DNS

   The PC stores a DNS mapping file. Add a mapping between the TM IP address and quicknet.h3c.com to this file. In Windows, the file name is hosts. You can find it at C:\Windows\System32\drivers\etc. Open the document in a text editor and add the following line:

   192.168.1.1 quicknet.h3c.com

   Here, 192.168.1.1 is the IP address of VLAN interface 1 for the TM in the network.

Configure the Quicknet management system

1. Use the Quicknet management system to perform initial deployment.

   Enter quicknet.h3c.com in the address bar of a browser to log in to the Quicknet management system. Complete deployment configuration as instructed to set up a Quicknet network.

   NOTE:

   Upon the first login to Quicknet management system, it takes 10 seconds or longer for the system to collect topology information. This waiting time increases with the number of devices in the network.

Figure 27 Login page of the Quicknet management system

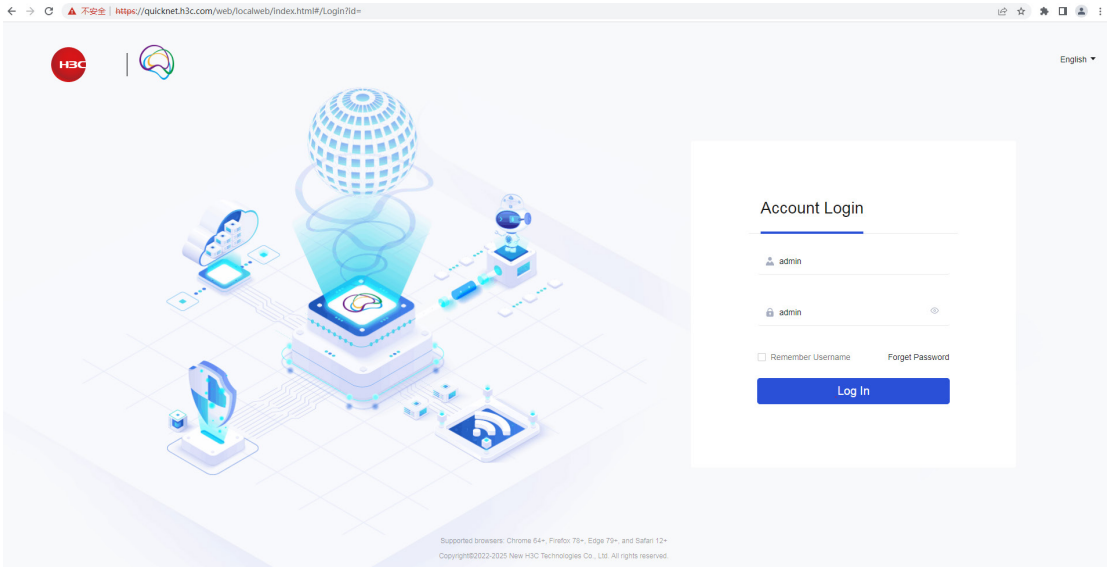


Figure 28 Homepage of the Quicknet management system



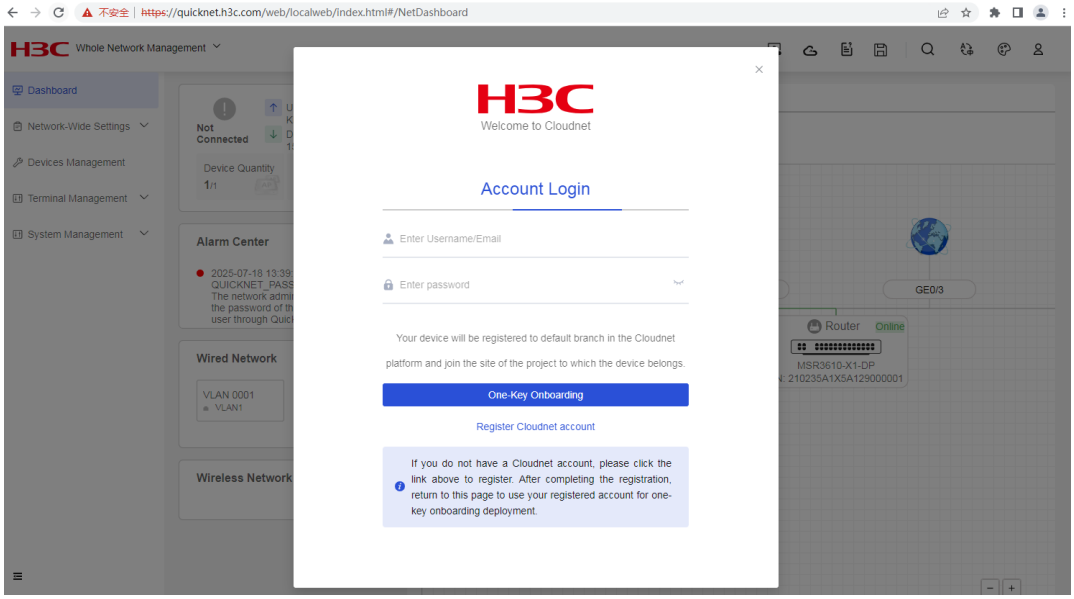
2. (Optional) Manage the AP license and software version

   If the AC does not have sufficient license seats, fit APs cannot register on the AC. As a best practice to ensure that the AC has sufficient license seats, click Continue Config after completing the deployment configuration, and go to the AP license management page to install AP licenses. Then, enter the AP software management page and upgrade the AP software as needed.

3. (Optional.) Perform one-key onboarding for all devices in the Quicknet network.

   a. Click the One-Key Onboarding icon at the top right corner of the page. The Cloudnet login window opens.

Figure 29 One-key onboarding

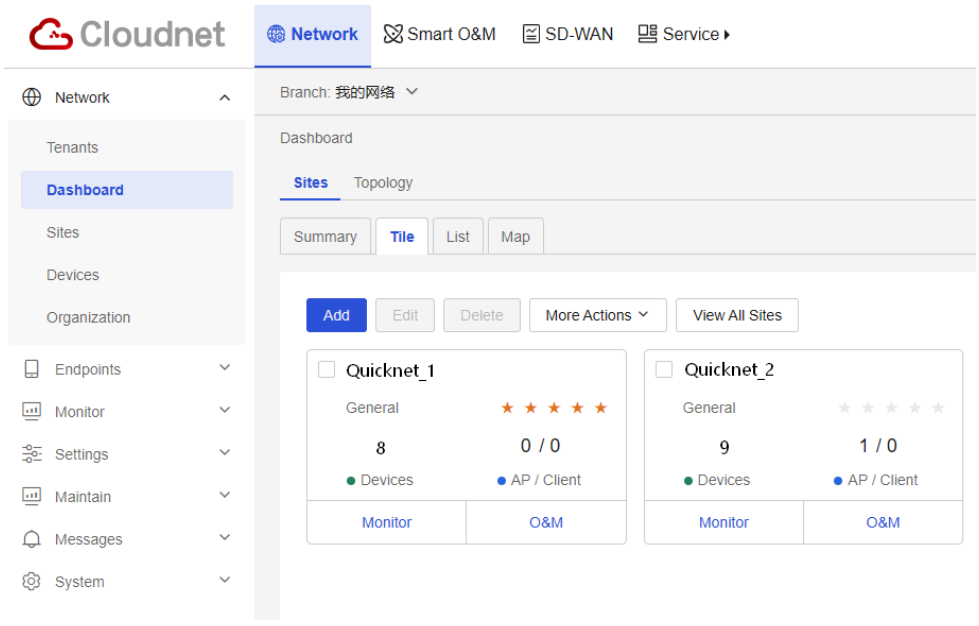


b. Enter your Cloudnet account, and click One-Key Onboarding to access the Cloudnet dashboard page.

If you do not have a Cloudnet account, register one first and then try to log in.

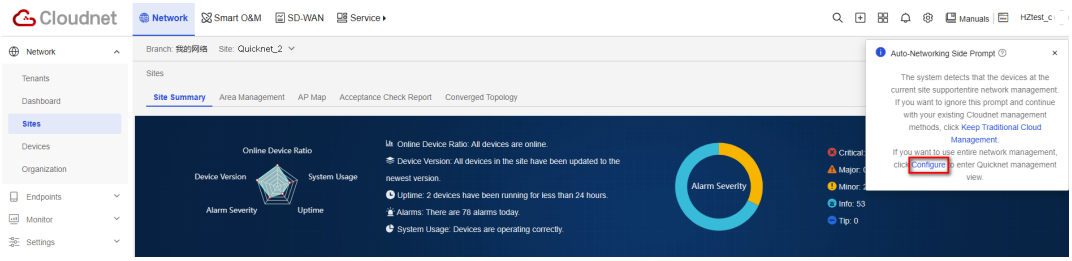
c. The Cloudnet dashboard displays all sites created by the login user. Click the site to which the Quicknet network belongs to access the site management page. The site name is the same as the project name specified during Quicknet deployment.

Figure 30 Cloudnet dashboard



d. Click to access the configuration page as prompted at the upper right corner of the page. Manage all devices in the Quicknet network as needed.

Figure 31 Cloudnet site page



Verifying the configuration

On the system summary page of the Quicknet management system, verify that you can view the topology of all member devices and the related information.

Configuration files

- TM configuration file (includes some settings that the device automatically generates to support the Quicknet function after you enable it)

```
#
 sysname TM
#
 telnet server enable
#
 dhcp enable
#
 dns fast-reply enable
 ip host quicknet.h3c.com 192.168.1.1
#
 lldp global enable
#
dhcp server ip-pool 1
 gateway-list 192.168.1.1
 network 192.168.1.0 mask 255.255.255.0
 forbidden-ip 192.168.1.1
#
interface Vlan-interface1
 ip address 192.168.1.1 255.255.255.0
 dhcp server apply ip-pool 1
#
line vty 0 63
 authentication-mode scheme
 user-role network-operator
#
 ssh user admin service-type netconf authentication-type publickey assign publickey
-smartmc-publickey-
#
local-user admin class manage
 password hash
$h$6$13g0MKsz6iBwZC77$MT90PviiKQpRXDjseCSlbg1wjE0/lgR10jU9g2LjCRUGhCbWO8bbvjb6EfD
11kqBaqA77NBn+A/4awgfccbw7g==
 service-type telnet http https ssh
```

```
 authorization-attribute user-role network-admin
 authorization-attribute user-role network-operator
 password-control length 4
 password-control composition type-number 1 type-length 1
 undo password-control complexity user-name check
#
 netconf soap http enable
 netconf soap https enable
 netconf ssh server enable
#
 ip http enable
 ip https enable
#
 quicknet enable auto tm
 quicknet password cipher $c$3$DvAxSmoVCgX9H4fD39DN9/HKvZxALMTLMzo20m4A
#
return
```

- TC configuration file (includes some settings that the device automatically generates to support the Quicknet function after you enable it)

```
#
 sysname TC
#
 dns fast-reply enable
 ip host quicknet.h3c.com 192.168.1.1
#
 lldp global enable
#
interface Vlan-interface1
 ip address 192.168.1.2 255.255.255.0
#
line vty 0 63
 authentication-mode scheme
#
 ssh user admin service-type netconf authentication-type publickey assign publickey
-smartmc-publickey-
#
local-user admin class manage
 password hash
$h$6$PyB2gjZbpLLL5euH$5eznb/M9pKoAqDHdcv3MscydHojnwphdLlxG6fEj9T1Tp/+MAuuycTT8HfE
T/FRCEzZSu5EHH8Lb1EpMwzIUGA==
 service-type telnet http https ssh
 authorization-attribute user-role network-admin
 authorization-attribute user-role network-operator
 password-control length 4
 password-control composition type-number 1 type-length 1
 undo password-control complexity user-name check
#
public-key peer -smartmc-publickey-
 public-key-code begin
```

```
   30819F300D06092A864886F70D010101050003818D0030818902818100C69D96A6F187859B
   CFA359C257041A0AD5AF7F29F5AC06D3A0DFE22286DF0439BC610D062D493A2A73BF6BCED0
   9A31E234331B1CABE94AF0D346054230EC8E7E17BC341FEBE0D87498E148D521254ED0880B
   F5C03D472146FBE531B54E7F1A4A88A91FA011E3101358BEF957F9EF467955837761AC3D9E
   5841196EB0FC1070BF0203010001
 public-key-code end
 peer-public-key end
#
 netconf soap http enable
 netconf ssh server enable
#
 ip http enable
 ip https enable
#
 quicknet tc enable
#
return
```

Contents

Configuring the AON Ethernet system ····· 1

About the AON Ethernet system ····· 1
AON Ethernet system framework ····· 1
Benefits ····· 2
AON node IDs ····· 2
AON node management methods ····· 2
Hardware compatibility with AON Ethernet system ····· 3
Restrictions and guidelines: AON Ethernet system configuration ····· 3
AON Ethernet system tasks at a glance ····· 3
Prerequisites ····· 4
Configuring the SmartMC management VLAN ····· 5
Enabling automatic upgrade for AON nodes ····· 6
Enabling SmartMC for the AON aggregation ····· 7
Managing an AON node in AON node view ····· 8
Configuring a description and a name for an AON node ····· 8
Configuring VLANs for an AON node ····· 8
Enabling interface isolation for an AON node ····· 10
Configuring PoE for an AON node ····· 10
Performing an internal loopback test for an interface of an AON node ····· 11
Shutting down interfaces on an AON node ····· 11
Rebooting an AON node ····· 12
Managing AON nodes in bulk by using an AON node configuration template ····· 12
Replacing a faulty AON node ····· 14
Display and maintenance commands for the AON Ethernet system ····· 14

Configuring the AON Ethernet system

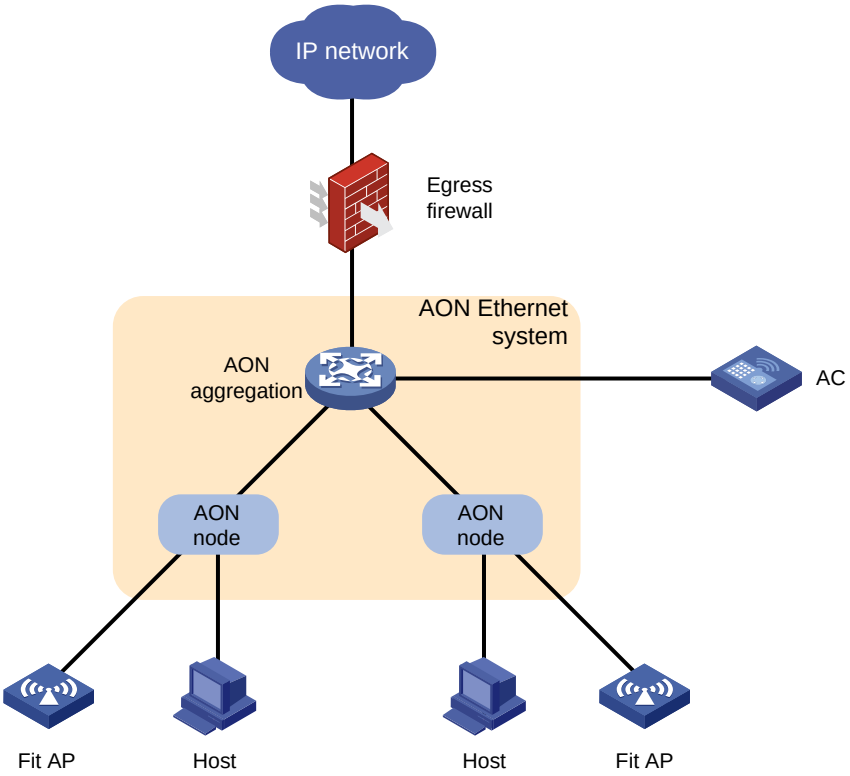
About the AON Ethernet system

The AON Ethernet system is a network management and maintenance solution based on a distributed deployment architecture. In the system, the AON aggregation (also called iOptic host) manages and maintains AON nodes (also called iOptic nodes), similar to how an AC manages APs in an AC+fit AP network. The AON aggregation functions like an AC, and the AON nodes function like APs. You can perform centralized control and management over campus access network devices by configuring and managing the AON aggregation.

AON Ethernet system framework

Figure 1 shows the basic framework of the AON Ethernet system.

Figure 1 AON Ethernet system framework



The AON Ethernet system contains the following types of physical devices:

- **AON aggregation**—The AON aggregation, core of the AON Ethernet system, automatically discovers and manages the AON nodes, forwards packets from the AON nodes, and performs centralized O&M on the AON nodes. On the AON aggregation, you can configure the AON nodes and obtain the interface status, interface statistics, PoE power, and transceiver module data of the AON nodes.
- **AON node**—An AON node acts as an expansion module of the AON aggregation, connecting to the external network through the AON aggregation. It is lightweight firmware that uses the manufacturer's built-in Microcontroller Unit (MCU) and achieves fast startup (within 10 seconds). Some AON node models support the Comware system, and others do not have an

operating system. In the AON Ethernet system, the AON nodes require zero configuration and are plug and play, functioning similarly as hubs.

Benefits

The AON Ethernet system offers the following benefits:

- Streamlined architecture. The traditional aggregation + access architecture has been transformed into an AON aggregation + AON nodes architecture, where an AON node operates as an expansion module to extend the capacity of the AON aggregation. This streamlined approach not only makes the network simpler but also allows for flexible expansion of the interface capacity for the AON aggregation.
- Plug-and-play AON nodes, simplifying network planning and deployment, and routine Ops.

AON node IDs

In the AON Ethernet system, an AON node is uniquely identified by its ID. The AON aggregation automatically allocates IDs to AON nodes as follows:

1. Enabled with SmartMC, the AON aggregation automatically broadcasts a discovery packet to discover AON nodes.
2. Upon receiving the discovery packets, the AON nodes automatically register with the AON aggregation.
3. The AON aggregation automatically allocates an ID to each AON node. The ID is bound to the SN and MAC address of an AON node. To view the mappings between AON nodes and their IDs, execute the **`display smartmc light-unit`** command on the AON aggregation. After an AON node goes offline, the AON aggregation still reserves the ID for the AON node until you execute the **`undo smartmc light-unit`** command to delete the AON node. Then, the device reclaims the ID of the AON node and allocates it to another AON node.

The AON aggregation uses AON node IDs to manage AON nodes. You can use the ID of an AON node to enter AON node view, where you can configure the name, description, interface VLANs, and PoE settings for the AON node.

AON node management methods

Managing a single AON node

When an AON node registers with the AON aggregation, the AON aggregation automatically allocates an ID to that AON node. The AON node is uniquely identified by the ID. The AON aggregation can manage AON nodes. After an AON node goes offline, the AON aggregation still reserves the ID for the AON node until you execute the **`undo smartmc light-unit`** command to delete the AON node. You can use the ID to enter AON node view and then configure parameters for the AON node.

Managing AON nodes in bulk

You can use an AON node configuration template to configure multiple AON nodes in bulk.

After you create an AON node configuration template, you can bind it to AON nodes and configure parameters for the AON nodes. When an AON node comes online, the configuration in AON node view, if any, applies. If the AON node does not have any configuration in AON node view and it matches the AON node configuration template, the AON aggregation automatically deploys configuration in AON node configuration template view to the AON node. If the deployment succeeds, the configuration in the template is converted to the configuration in AON node view. If the deployment fails, the configuration in the template is not converted to the configuration in AON node view.

You can bind an AON node configuration template to AON nodes in the following methods:

- Bind the template to the interface connected to the AON nodes on the AON aggregation. In this method, all AON nodes that come online via the interface can use the configuration in the template.

  When the uplink interfaces of an AON node supports aggregation and are connected to the AON aggregation, you must also configure an aggregate interface on the AON aggregation and bind the AON node configuration template to this aggregate interface.

- Bind the template to the device type of the AON nodes. In this method, all AON nodes of that type can use the configuration in the template when they come online.

Bind the template to the serial numbers of the AON nodes. In this method, only AON nodes of the specified serial numbers can use the configuration in the template when they come online. This method allows for both bulk and differentiated configuration.

Hardware compatibility with AON Ethernet system

The S5500V3-HI switch series does not support this feature.

Restrictions and guidelines: AON Ethernet system configuration

The AON Ethernet system uses VLAN 1 to broadcast discovery packets and uses the MAC address of VLAN-interface 1 to construct the discovery packets. Therefore, Vlan-interface 1 must exist on the AON aggregation and cannot be deleted.

If the AON aggregation connects to an AON node through a non-aggregate interface or an aggregate interface with incorrect configuration, it can detect the aggregate link connection error when the AON node comes online. If the aggregate interface configuration or aggregate link topology changes, the AON aggregation cannot immediately detect the aggregate link connection error when the AON node is online. The AON aggregation can detect the error only after the AON node comes online again.

AON Ethernet system tasks at a glance

To configure the AON Ethernet system, perform the following tasks:

1. (Optional.) Configuring the SmartMC management VLAN

   To modify the SmartMC management VLAN, make sure SmartMC is disabled. If you modify the SmartMC management VLAN when SmartMC is enabled, the modification does not take effect. For the modification to take effect, first disable SmartMC and enable SmartMC again.
2. (Optional.) Enabling automatic upgrade for AON nodes
3. Enabling SmartMC for the AON aggregation
4. (Optional.) Managing an AON node in AON node view
   - Configuring a description and a name for an AON node
   - Configuring VLANs for an AON node
   - Enabling interface isolation for an AON node
   - Configuring PoE for an AON node
   - Performing an internal loopback test for an interface of an AON node
   - Shutting down interfaces on an AON node
   - Rebooting an AON node
5. (Optional.) Managing AON nodes in bulk by using an AON node configuration template

Use an AON node configuration template for deployment configuration of AON nodes that come online for the first time. The AON node configuration template does not take effect on an AON node if the following conditions are met:

- The AON node has been online.
- Configuration exists in the corresponding AON node view.

6. (Optional.) Replacing a faulty AON node

Prerequisites

IP address planning

1. The IP address of the management VLAN interface on the AON aggregation and the IP addresses of AON nodes must reside in the same network segment. The default management VLAN is VLAN 1.
2. You can configure IP addresses by using either of the following methods:
   - Manual configuration.
   - DHCP (recommended).

     This method makes sure the IP address of the AON aggregation and the AON nodes reside in the same network segment, allowing for Layer 3 management channel establishment.

DHCP deployment recommendations

- As a best practice, deploy the DHCP server on the AON aggregation.
- Configure the following address pools on the AON aggregation:
  - Address pool for VLAN 1.
  - Address pool for the management VLAN of the AON Ethernet system.

For information about configuring the DHCP server, see DHCP configuration in *Layer 3—IP Services Configuration Guide*.

Configuring a local user

1. Enter system view.

   system-view
2. Create a local user named **admin**. (The name must be **admin**.)

   local-user admin
3. Configure a password for the local user.

   password { **hash** | **simple** } *string*

   The specified password must be the same as that configured by using the **smartmc tm username** *username* **password** { **cipher** | **simple** } *string* **enable** command.
4. Specify the service type for the local user.

   service-type { **http** | **https** | **ssh** | **telnet** } *

   By default, no service type is specified for a local user.
5. Assign the network-admin user role to the user.

   authorization-attribute user-role network-admin
6. Return to system view.

   quit
7. Enable scheme authentication for user lines VTY 0 through VTY 63.

   line vty 0 63

   authentication-mode scheme

By default, password authentication is enabled for VTY lines.

Configuring the SmartMC management VLAN

About this task

Application scenarios

Use the SmartMC management VLAN to separate the management packets from other service packets in the AON Ethernet system.

Operating mechanism

An AON Ethernet system is set up as follows:

1. Enabled with SmartMC, the AON aggregation broadcasts a SmartMC scanning packet in VLAN 1 to discover AON nodes. Upon receiving the scanning packet, an AON node sends a response to register with the AON aggregation.
2. The AON aggregation and AON node establish a CoAP session within the management VLAN. This CoAP session acts as the Layer 3 management channel between them.

   By default, the management VLAN is VLAN 1. In VLAN 1, the AON aggregation and AON node automatically request IP addresses and the AON aggregation actively establishes a CoAP session with the AON node.
3. If the SmartMC management VLAN is configured, the AON aggregation and AON node work as follows:
   a. The downlink interface on the AON aggregation and the uplink interfaces of the AON node automatically permit management VLAN packets.
   b. If you use the AON aggregation as the DHCP server, you must manually configure the IP address of the management VLAN interface, enable DHCP, and configure address pools on the AON aggregation. If you use another device as the DHCP server, the management VLAN interface of the AON aggregation automatically requests an IP address.
   c. The AON node automatically requests an IP address within the management VLAN.
   d. The AON aggregation and AON node re-establish a COAP session by using the IP addresses obtained within the management VLAN.
4. The AON aggregation sends a successful registration packet to the AON node and the AON node joins the AON Ethernet system.

The SmartMC management VLAN is dedicated to transmitting protocol packets for the AON Ethernet system management plane. Broadcast packets are restricted within the management VLAN, effectively separating protocol packets on the AON Ethernet system management plane from the service packets and other protocol packets. This facilitates the construction and maintenance of the AON Ethernet and enhances its security.

Restrictions and guidelines

When the device runs as a SmartMC TC, the system automatically deletes the **smartmc management-vlan** command.

To modify the SmartMC management VLAN, make sure SmartMC is disabled. If you modify the SmartMC management VLAN when SmartMC is enabled, the modification does not take effect. For the modification to take effect, first disable SmartMC and enable SmartMC again.

This feature is supported only when the device acts as an AON aggregation. When the device also acts as the TM in a SmartMC network, if you change the SmartMC management VLAN to a VLAN other than VLAN 1, TCs cannot come online.

Procedure

1. Enter system view.

   system-view

2. Create a management VLAN.

   vlan *vlan-id*

   By default, only VLAN 1 exists on the device.

   To successfully configure the SmartMC management VLAN, first create the management VLAN.

3. Return to system view.

   quit

4. Create a management VLAN interface.

   interface vlan-interface *vlan-id*

   By default, no VLAN interface exists.

5. Return to system view.

   quit

6. Configure the SmartMC management VLAN.

   smartmc management-vlan *vlan-id*

   By default, the SmartMC management VLAN is VLAN 1.

Enabling automatic upgrade for AON nodes

About this task

Configured with this feature, the AON aggregation automatically identifies whether the software version of an onboarding AON node is compatible with its own. If they are not compatible, the AON node software is automatically upgraded.

Restrictions and guidelines

In a distributed deployment scenario, you must also upload the target AON node version to the AON aggregation.

If you do not enable this feature and the AON node software version is incompatible with that of the AON aggregation, an AON node can still come online and forward packets. To avoid affecting AON node management on the AON aggregation, upgrade the AON node immediately.

Prerequisites

1. Execute the **display version** command on the AON aggregation to obtain its running software version.
2. Obtain the most recent software version of the AON nodes that is compatible with the AON aggregation software version based on the release notes of the AON aggregation.
3. Upload the startup software package of the AON nodes to the AON aggregation. For the automatic upgrade feature to take effect, make sure the AON aggregation has the startup software package of the AON nodes.

Procedure

1. Enter system view.

   system-view

2. Enable automatic upgrade for AON nodes.

   smartmc auto-upgrade enable

   By default, automatic upgrade is disabled for AON nodes.

Enabling SmartMC for the AON aggregation

About this task

An AON aggregation and its connected AON nodes form an AON Ethernet system. The members in the AON Ethernet system discover each other at Layer 2 through SmartMC. In this way, the AON Ethernet system is also a SmartMC network. With this feature configured, the AON aggregation acts as the topology master (TM), and the AON nodes act as topology clients (TCs). Enabled with the AON node mode, an AON node is automatically enabled with SmartMC and operates as a TC. In the AON Ethernet system, you only need to enable the AON node mode on AON nodes and no other settings are required for them. For information about the factory default mode of an AON node, see its product documentation. For more information about SmartMC, see *Network Management and Monitoring Configuration Guide*.

An AON aggregation supports both the SmartMC and AON Ethernet system features. For information about configuring a pure SmartMC network, see *Network Management and Monitoring Configuration Guide*. For information about configuring an AON Ethernet, see the configuration guide. Do not deploy both networks on an AON aggregation. An AON aggregation broadcasts scanning packets to discover TCs and determines the network type based on the type of the first TC that responds to the scanning packet.

Restrictions and guidelines

A SmartMC network can have only one AON aggregation.

Disabling SmartMC for the AON aggregation clears the running configuration related to SmartMC.

The SmartMC feature consumes certain ACL resources. If the ACL resources are insufficient, SmartMC will fail to be enabled. To resolve such an issue, perform the following tasks:

1. Obtain the ACL configuration and running status by using the **`display acl`** command.
2. Delete unnecessary ACLs by using the **`undo acl`** command as required.
3. Enable SmartMC again.

For more information about ACL, see *ACL and QoS Configuration Guide*.

When you enable SmartMC, the device identifies whether HTTP port 80, HTTPS port 443, and CoAP port 5683 are in use. If any of these ports is occupied, SmartMC will fail to be enabled.

If you execute this command multiple times, the most recent configuration takes effect. You can use the **`smartmc enable`** command to enable SmartMC on an AON Ethernet or SmartMC network. The SmartMC network requires a username and password, but the AON Ethernet does not.

Procedure

1. Enter system view.

   `system-view`
2. Enable SmartMC for the AON aggregation.

   `smartmc tm username` *`username`* **`password`** { **`cipher`** | **`simple`** } *`string`* **`enable`**

   By default, SmartMC is disabled for the AON aggregation.

Managing an AON node in AON node view

Configuring a description and a name for an AON node

About this task

In the AON Ethernet system, each AON node is uniquely identified by its ID. However, the AON node ID, a string of digits, is difficult to remember. To resolve such an issue, you can configure a name for each AON node for easier management.

To enter the configuration view of an AON node, configure a name for that AON node and then execute the **smartmc light-unit name** *name* command.

The system name of an AON node can contain only 1 to 31 characters. To manage AON nodes efficiently, you can also configure the IDs, geographic locations, and IP addresses as descriptions for them. The AON node description is a string of 1 to 255 characters.

Restrictions and guidelines

Do not configure the name of an AON node as the MAC address of another AON node.

Procedure

1. Enter system view.
   system-view
2. Enter AON node view.
   smartmc light-unit { *id* | **name** *name* }
3. Configure a name for the AON node.
   name *sysname*
   By default, the AON aggregation uses the MAC address of an AON node as its name.
4. Configure a description for the AON node.
   description *text*
   By default, an AON node does not have a description.

Configuring VLANs for an AON node

VLAN modes

An AON node supports the following VLAN modes:

- **Transparent mode**—In this mode, the AON node forwards packets based on the destination MAC address no matter whether service packets carry the VLAN tag or their VLAN IDs or not. In transparent mode, you cannot edit the VLAN settings of interfaces on the AON node. To configure VLANs based on MAC addresses or subnets for service packets, perform the configuration on the interface that connects the AON aggregation to the AON node.
- **Manual mode**—In this mode, an interface operates in hybrid mode and VLANs are configured based on interfaces. After you switch the VLAN mode to manual, the interface uses PVID VLAN 1, which is added to the permitted VLAN list of the interface as an untagged member. To change the PVID of the interface, use the **port hybrid pvid** command. To assign the interface to specified VLANs, use the **port hybrid vlan** command.

VLAN configuration guidelines for the uplink interface on an AON node

An AON node uses a physical interface as the uplink interface to connect to an AON aggregation. For AON nodes that have two uplink interfaces, they can aggregate the two interfaces automatically. Support for auto uplink interface aggregation depends on the model and software version of AON

nodes. An AON aggregation can automatically identify whether an AON node supports uplink interface aggregation. Use the **uplink** keyword or *port-id-list* argument as prompted.

- For AON nodes that aggregate uplink interfaces automatically, the two uplink interfaces operate as follows:
  - In versions earlier than a specific version, the two interfaces back up each other, but only one interface is active for packet forwarding at the same time. In this case, the **uplink** keyword is not supported. Use the *port-id-list* argument when you configure VLANs for AON nodes.
  - In a specific version and later versions, the two interfaces back up each other, and can forward traffic at the same time for load balancing. In this case, use the **uplink** keyword when you configure VLANs for AON nodes. To avoid configuration failures, do not use the *port-id-list* argument.
    - If you upgrade a device from a version that does not support the **uplink** keyword to a version that supports the **uplink** keyword, the *port-id-list* configuration in the configuration file still takes effect for backward compatibility.
    - If you downgrade a device from a version that supports the **uplink** keyword to a version that does not support the **uplink** keyword, and then upgrade the device back to a version that supports the keyword, the configuration file of the device contains both the **uplink** and *port-id-list* configurations. In this case, the **uplink** configuration takes effect.

  For information about the version from which an AON node supports uplink interface aggregation, see the documentation for the AON node.
- For AON nodes that do not support uplink interface aggregation, the two interfaces back up each other, but only one interface is active for packet forwarding at the same time. Use the *port-id-list* argument when you configure VLANs for AON nodes.

Restrictions and guidelines

In transparent mode, user traffic cannot be isolated based on VLANs on an AON node. To isolate user traffic, configure the interface isolation feature for the AON node or isolate user traffic on the AON aggregation.

Procedure

1. Enter system view.

   system-view
2. Enter AON node view.

   smartmc light-unit { *id* | **name** *name* }
3. Specify the VLAN mode for the AON node.

   vlan mode { **manual** | **transparent** }

   By default, the VLAN mode for an AON node is transparent.
4. Set the PVID of hybrid interfaces on the AON node.

   port { **uplink** | *port-id-list* } * **hybrid pvid vlan** *vlan-id*

   By default, the interfaces on an AON node do not check VLAN information for packets and transparently transmit all the packets.
5. Assign hybrid interfaces on the AON node to the specified VLANs.

   port { **uplink** | *port-id-list* } * **hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }

   By default, the interfaces on an AON node do not check VLAN information for packets and transparently transmit all the packets.

Enabling interface isolation for an AON node

About this task

Application scenarios

This feature prevents the endpoints connected to an AON node from accessing each other at Layer 2 to enhance LAN security.

Operating mechanism

By default, all interfaces on an AON node belong to the same VLAN and the endpoints connected to the interfaces can access each other at Layer 2.

This feature isolates the interfaces on an AON node and enables the endpoints connected to the AON node to forward packets through the AON aggregation. This allows you to manage the traffic of AON nodes on the AON aggregation, which enhances LAN security.

Procedure

1. Enter system view.

   system-view

2. Enter AON node view.

   smartmc light-unit { *id* | **name** *name* }

3. Enable interface isolation for the AON node.

   isolate enable

   By default, interface isolation is disabled for an AON node.

Configuring PoE for an AON node

About this task

Enabled with PoE, the PIs on an AON node can provide PoE power for PDs if they support PoE output.

Forced PoE power supply

Before supplying power to a PD, an AON node performs a detection of the PD by default. It supplies power to the PD only after the PD passes the detection. If the PD fails the detection but the power provided by the AON node meets the PD specifications, you can enable forced power supply to the PD.

PoE power supply standards

IEEE released 802.3af (PoE) and 802.3at (PoE+) standards in chronological sequence. The standards are backward compatible but forwards incompatible. Before supplying power to a PD, a PI will detect the PD, determine the power class for the PD based on its power requirement, and supply power to the PD based on the power class.

- **802.3af**—Defines four power classes 0 to 3 and provides a maximum power of 12.95 W.
- **802.3at**—Adds class 4 in addition to the four power classes defined by 802.3af and provides a maximum power of 25.5 W.

Specify a PoE standard based on the power requirement of the PD. If the maximum power allowed by the standard is less than that required by the PD, the PI will not supply power to the PD.

Restrictions and guidelines

The forced PoE power supply feature enables an AON node to supply power to a PD directly without performing a detection of the PD. To avoid damaging the PD, make sure the power provided by the AON node meets the PD specifications before configuring this feature.

Disabling PoE for PIs on an AON node restores the default settings of the **poe force-power** and **standard** commands on the PIs.

Procedure

1. Enter system view.
   system-view
2. Enter AON node view.
   smartmc light-unit { *id* | **name** *name* }
3. Enable PoE for PIs on the AON node.
   poe enable port *port-id-list*
   By default, PoE is enabled for PIs on an AON node.
4. Specify a PoE standard for PIs on the AON node.
   poe standard { **af** | **at** } **port** *port-id-list*
   By default, the IEEE 802.3af standard is used.
5. Enable forced PoE power supply for PIs on the AON node.
   poe force-power port *port-id-list*
   By default, forced PoE power supply is disabled for PIs on an AON node.
6. Restart PoE on PIs of the AON node.
   restart poe port *port-id-list*

Performing an internal loopback test for an interface of an AON node

About this task

Use this feature to identify whether the interfaces on an AON node can forward packets correctly.

After you configure this feature on an interface of an AON node, the interface returns the packets to be forwarded to the device, allowing the packets to loop through the internal line. Then, you can identify whether the Ethernet forwarding path within the device can operate correctly.

Restrictions and guidelines

If you enable loop guard on an interface of an AON node, you cannot perform an internal loopback test for that interface on the AON aggregation.

Procedure

1. Enter system view.
   system-view
2. Enter AON node view.
   smartmc light-unit { *id* | **name** *name* }
3. Perform an internal loopback test for an interface of the AON node.
   loopback-test internal port *port-id*

Shutting down interfaces on an AON node

About this task

Some interface settings might require an interface restart before they can take effect.

Restrictions and guidelines

CAUTION:

Executing the **shutdown** command disconnects the links of the specified interfaces and interrupts communication. Use this command with caution.

The AON Ethernet system automatically discovers and records the uplink interfaces of AON nodes. Do not use this command to shut down the uplink interfaces of AON nodes.

Procedure

1. Enter system view.
   system-view
2. Enter AON node view.
   smartmc light-unit { *id* | **name** *name* }
3. Shut down interfaces on the AON node.
   shutdown port *port-id-list*
   By default, the interfaces on an AON node are up.

Rebooting an AON node

About this task

Use this feature to reboot an AON node when it is faulty.

Procedure

1. Enter system view.
   system-view
2. Enter AON node view.
   smartmc light-unit { *id* | **name** *name* }
3. Reboot the AON node.
   Reboot

Managing AON nodes in bulk by using an AON node configuration template

Restrictions and guidelines

Use an AON node configuration template for deployment configuration of AON nodes that come online for the first time. The AON node configuration template does not take effect on an AON node if the following conditions are met:

- The AON node has been online.
- Configuration exists in the corresponding AON node view.

You can repeat this command to bind a template to different AON nodes. However, the same serial number, model, or connection interface (the interface connecting to AON nodes on the AON aggregation) can be bound to only one template. An AON node can match configuration templates by its serial number, model, and connection interface. When an AON node matches multiple templates, the following AON node configurations are in descending order: configuration in AON node view, configuration of the template bound to the serial number, configuration of the template bound to the device type, and configuration of the template bound to the connection interface.

Procedure

1. Enter system view.
   system-view
2. Create an AON node configuration template and enter its view.
   smartmc light-unit template *template-name*
3. (Optional.) Configure a description for the template.
   description *text*
   By default, an AON node configuration template does not have a description.
4. (Optional.) Configure VLAN parameters.
   - Specify the VLAN mode for the AON node.
     vlan mode { **manual** | **transparent** }
     By default, the VLAN mode for an AON node is transparent.
   - Set the PVID of hybrid interfaces on the AON node.
     port *port-id-list* **hybrid pvid vlan** *vlan-id*
     By default, the interfaces on an AON node do not check VLAN information for packets and transparently transmit all the packets.
   - Assign hybrid interfaces on the AON node to the specified VLANs.
     port *port-id-list* **hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }
     By default, the interfaces on an AON node do not check VLAN information for packets and transparently transmit all the packets.
5. (Optional.) Enable interface isolation for the AON node.
   isolate enable
   By default, interface isolation is disabled for an AON node.
6. (Optional.) Configure PoE.
   - Enable PoE for PIs on the AON node.
     poe enable port *port-id-list*
     By default, PoE is enabled for PIs on an AON node.
   - Specify a PoE standard for PIs on the AON node.
     poe standard { **af** | **at** } **port** *port-id-list*
     By default, IEEE 802.3af is used.
   - Enable forced PoE power supply for PIs on the AON node.
     poe force-power port *port-id-list*
     By default, forced PoE power supply is disabled for PIs on an AON node.
7. Return to system view.
   quit
8. Bind the AON node configuration template to AON nodes.
   smartmc light-unit bind template *template-name* { **with-interface** *interface-type interface-number* | **with-light-unit-type** *light-unit-type* | **with-sn** *sn* }
   By default, an AON node configuration template is not bound to any AON nodes.

Replacing a faulty AON node

About this task

When an AON node in the AON Ethernet system fails, you can use this feature to replace the faulty AON node.

1. Connect a new AON node to the AON aggregation. Make sure the new AON node and the faulty AON node are the same model.
2. The AON aggregation detects that the new AON node comes online. Execute the **display smartmc light-unit** command to view the ID and status of the new AON node. If the new AON node is in **Normal** status, the new AON node successfully comes online.
3. Execute the **smartmc replace** command.
4. The AON Ethernet system replaces the ID of the new AON node with that of the faulty AON node, and runs configuration of the faulty AON node on the new one.

Restrictions and guidelines

Make sure the new member and the faulty member have the same neighbor relationship and model.

Prerequisites

Connect the new AON node to the AON Ethernet system and execute the **display smartmc light-unit** command to obtain the ID and status of the new AON node. You can configure this feature only when the new AON node is in **Normal** status.

Procedure

1. Enter system view.
   system-view
2. Manually replace a faulty AON node.
   smartmc replace tc *tc-id1* **faulty-tc** *tc-id2*

Display and maintenance commands for the AON Ethernet system

Execute **display** commands in any view and the **reset** command in user view.

Task	Command
Display connections between the AON aggregation and AON nodes in the AON Ethernet system.	**display smartmc device-link**
Display AON node information.	**display smartmc light-unit** [*id* | **interface** *interface-type interface-number* | **name** *name*] [**verbose**]
Display the configuration deployed from the AON aggregation to AON nodes and successfully run on the AON nodes.	**display smartmc light-unit configuration history**
Display information about the PoE power supply status of an AON node.	**display smartmc light-unit** { *id* | **name** *name* } **poe**
Display power supply information for PIs on an AON node.	**display smartmc light-unit** { *id* | **name** *name* } **poe port** [*port-id*]

Task	Command
Display information about interfaces on an AON node.	**display smartmc light-unit** { *id* | **name** *name* } **port** [*port-id* | **brief** | **uplink**]
Display key parameters of transceiver modules on an AON node.	**display smartmc light-unit** { *id* | **name** *name* } **transceiver port** [*port-id*]
Display upgrade information for an AON node.	**display smartmc light-unit** { *id* | **name** *name* } **upgrade-info**
Display AON node configuration template information.	**display smartmc light-unit template** [**name** *template-name*] [**match-info**]
Clear the configuration deployed from the AON aggregation to AON nodes.	**reset smartmc light-unit configuration history**

Contents

Configuring NETCONF 1

About NETCONF 1
NETCONF structure 1
Basic NETCONF concepts 1
NETCONF message format 3
How to use NETCONF 5
Protocols and standards 5
FIPS compliance 5
NETCONF tasks at a glance 5
Configuring NETCONF to use module-specific namespaces 6
Establishing a NETCONF session 7
Restrictions and guidelines for NETCONF session establishment 7
Setting NETCONF session attributes 7
Establishing NETCONF over SOAP sessions 8
Establishing NETCONF over SSH sessions 10
Establishing NETCONF over Telnet or NETCONF over console sessions 10
Exchanging capabilities 11
Retrieving device configuration information 11
Restrictions and guidelines for device configuration retrieval 11
Retrieving device configuration and state information 12
Retrieving non-default settings 14
Retrieving NETCONF information 15
Retrieving YANG file content 15
Retrieving NETCONF session information 16
Example: Retrieving a data entry for the interface table 16
Example: Retrieving non-default configuration data 18
Example: Retrieving syslog configuration data 19
Example: Retrieving NETCONF session information 20
Filtering data 21
About data filtering 21
Restrictions and guidelines for data filtering 21
Table-based filtering 21
Column-based filtering 22
Example: Filtering data with regular expression match 24
Example: Filtering data by conditional match 26
Locking or unlocking the running configuration 27
About configuration locking and unlocking 27
Restrictions and guidelines for configuration locking and unlocking 27
Locking the running configuration 27
Unlocking the running configuration 28
Example: Locking the running configuration 28
Modifying the configuration 29
About the <edit-config> operation 29
Procedure 29
Example: Modifying the configuration 31
Saving the running configuration 32
About the <save> operation 32
Restrictions and guidelines 32
Procedure 32
Example: Saving the running configuration 33
Loading the configuration 34
About the <load> operation 34
Restrictions and guidelines 34
Procedure 34
Rolling back the configuration 35
Restrictions and guidelines 35
Rolling back the configuration based on a configuration file 35

Rolling back the configuration based on a rollback point 35
Managing the candidate configuration 39
About this task 39
Candidate configuration management tasks at a glance 39
Creating a candidate configuration datastore 40
Editing the candidate configuration datastore 40
Committing the candidate configuration immediately 41
Committing the candidate configuration only after confirmation 41
Deleting the candidate configuration datastore 42
Performing CLI operations through NETCONF 43
About CLI operations through NETCONF 43
Restrictions and guidelines 43
Procedure 43
Example: Performing CLI operations 43
Subscribing to events 44
About event subscription 44
Restrictions and guidelines 44
Subscribing to syslog events 45
Subscribing to events monitored by NETCONF 46
Subscribing to events reported by modules 47
Canceling an event subscription 48
Canceling an event subscription 49
Example: Subscribing to syslog events 49
Terminating NETCONF sessions 50
About NETCONF session termination 50
Procedure 50
Example: Terminating another NETCONF session 51
Returning to the CLI 51
Enabling SNMP notifications for NETCONF 52
Display and maintenance commands for NETCONF 53
Supported NETCONF operations 53
Candidate configuration datastore operations 53
cancel-commit 53
commit (immediate commit) 53
commit (confirmation-required commit) 54
commit (commit confirmation) 54
copy-config 55
delete-config 55
discard-changes 55
CLI operations 56
CLI 56
Edit operations 57
edit-config: create 57
edit-config: delete 57
edit-config: merge 58
edit-config: remove 58
edit-config: replace 58
edit-config: *test-option* 58
edit-config: default-operation 59
edit-config: error-option 60
edit-config: incremental 61
Get operations 62
get 62
get-bulk 62
get-bulk-config 62
get-config 63
get-sessions 63
Other operations 64
action 64
close-session 64
kill-session 64

load ····65
lock····65
rollback····65
save····66
unlock····66
xml2cli ····66

Configuring NETCONF

About NETCONF

Network Configuration Protocol (NETCONF) is an XML-based network management protocol. It provides programmable mechanisms to manage and configure network devices. Through NETCONF, you can configure device parameters, retrieve parameter values, and collect statistics. For a network that has devices from vendors, you can develop a NETCONF-based NMS system to configure and manage devices in a simple and effective way.

NETCONF structure

NETCONF has the following layers: content layer, operations layer, RPC layer, and transport protocol layer.

Table 1 NETCONF layers and XML layers

NETCONF layer	XML layer	Description
Content	Configuration data, status data, and statistics	Contains a set of managed objects, which can be configuration data, status data, and statistics. For information about the operable data, see the NETCONF XML API reference for the device.
Operations	<get>, <get-config>, <edit-config>…	Defines a set of base operations invoked as RPC methods with XML-encoded parameters. NETCONF base operations include data retrieval operations, configuration operations, lock operations, and session operations. For information about operations supported on the device, see "Supported NETCONF operations."
RPC	<rpc> and <rpc-reply>	Provides a simple, transport-independent framing mechanism for encoding RPCs. The <rpc> and <rpc-reply> elements are used to enclose NETCONF requests and responses (data at the operations layer and the content layer).
Transport protocol	In non-FIPS mode: Console, Telnet, SSH, HTTP, HTTPS, and TLS In FIPS mode: Console, SSH, HTTPS, and TLS	Provides reliable, connection-oriented, serial data links. The following transport layer sessions are available in non-FIPS mode: • CLI sessions, including NETCONF over Telnet sessions, NETCONF over SSH sessions, and NETCONF over console sessions. • NETCONF over SOAP sessions, including NETCONF over SOAP over HTTP sessions and NETCONF over SOAP over HTTPS sessions. The following transport layer sessions are available in FIPS mode: • CLI sessions, including NETCONF over SSH sessions and NETCONF over console sessions. • NETCONF over SOAP over HTTPS sessions.

Basic NETCONF concepts

XML

eXtensible Markup Language (XML) is a simple, text-based data storage language. It describes the content and structure of data exchanged between entities on the Internet.

XML is case sensitive.

NETCONF uses XML-based data encoding for the configuration data and protocol messages. NETCONF messages are XML-based and comply with RFC 4741.

Element

Elements are the basic building blocks in an XML document. They act as containers to store various types of content, including text, child elements, and attributes.

An element encloses content in a pair of start and end tags, in <*element-name*>*content*</*element-name*> format. For example, the **capability** element is `<capability>`*content*`</capability>`.

If an element does not have content, that element is called an empty element. You can present an empty element in a self-closing empty tag in <*element-name*/> format, for example, `<ok/>`.

Attribute

Attributes provide meta-level information about an element. They are added in *name=value* format after the element name, within the start tag. For examples, the `xmlns` and `operation` attributes in the following element specify the namespace and operation type, respectively.

```
<Syslog xmlns="http://www.h3c.com/netconf/config:1.0" operation="create">
```

Namespace

A namespace is an abstract container or environment created to hold a logical group of unique identifiers or names. An identifier defined in a namespace is associated only with that namespace. Each data module must have a world-unique namespace to avoid element and attribute naming conflicts.

The device supports the following types of namespaces:

- **urn:ietf:params:xml:ns:netconf:base:1.0 namespace**—This namespace defines all NETCONF elements. All NETCONF messages must declare this namespace in the RPC layer.
- **ietf-netconf namespace**—The **urn:ietf:params:xml:ns:yang:ietf-netconf-monitoring** namespace defined in a NETCONF protocol. NETCONF messages for the <get-schema> and <get><netconf-state/></get> operations must declare this namespace.
- **Common namespace**—A namespace that defines elements across service modules. In a packet that uses a common namespace, the namespace is indicated in the <top> element, and the modules are listed under the <top> element.
- **H3C common namespaces**—The http://www.h3c.com/netconf/*ns*:1.0 namespaces. The *ns* argument represents data, config, and action, which represent the data, configuration, and action namespaces, respectively. The common namespaces are shared by all modules. If used, a common namespace is typically specified for the xmlns attribute in the <top> element, and the modules are listed under the <top> element.

  Example:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <get-bulk>
      <filter type="subtree">
         <top xmlns="http://www.h3c.com/netconf/data:1.0">
            <Ifmgr>
               <Interfaces>
               </Interfaces>
            </Ifmgr>
         </top>
      </filter>
   </get-bulk>
</rpc>
```

- **Module-specific namespaces**—A module has its own namespace to define module-specific elements. A packet that uses a module-specific namespace does not have the <top> element. The namespace follows the module name.

  Example:

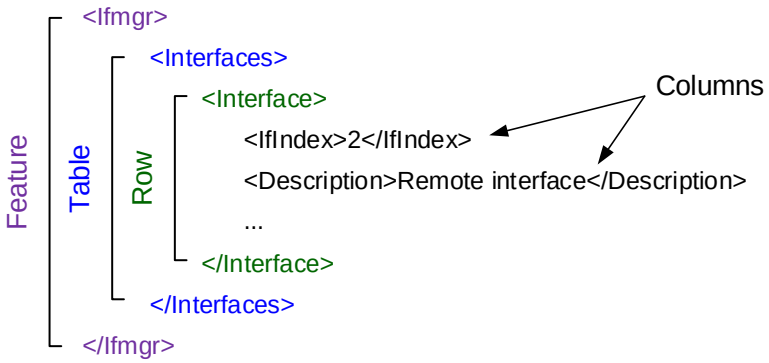
```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <get-bulk>
      <filter type="subtree">
            <Ifmgr xmlns="http://www.h3c.com/netconf/data:1.0-Ifmgr">
               <Interfaces>
               </Interfaces>
            </Ifmgr>
      </filter>
   </get-bulk>
</rpc>
```

Node, module, table, row, and column

In H3C NETCONF XML AI references, elements are also called nodes. An element might be a module (or feature), table, row, or column node, as shown in Figure 1.

- **Module**—A feature module defines a hierarchy of related nodes. The element name for a module node is the module name.
- **Table**—A table contains rows and columns, with each row representing an object instance and each column containing a parameter for the instance. A table might also contains subtables.
- **Row**—Each row represents an object instance in the table.
- **Column**—Each column contains a parameter for the instance.

Figure 1 Node, module, table, row, and column in NETCONF messages



In Figure 1, the `Ifmgr` element represents the interface management module (Ifmgr). The `Interfaces` element is a table node in the Ifmgr module and it contains a table of interfaces. The `Interface` element is a row node and it contains data about an interface in the interface table. The `IfIndex` and `Description` elements are column nodes that contain the interface index and description of the interface.

NETCONF message format

NETCONF

An incoming NETCONF message must pass XML schema check before it can be processed. If a NETCONF message fails XML schema check, the device sends an error message to the client.

For information about the NETCONF operations supported by the device and the operable data, see the NETCONF XML API reference for the device.

The following example shows a NETCONF message for getting all parameters of all interfaces on the device:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-bulk>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Ifmgr>
          <Interfaces>
                <Interface/>
          </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get-bulk>
</rpc>
```

NETCONF over SOAP

All NETCONF over SOAP messages are XML-based and comply with RFC 4741. NETCONF messages are contained in the <Body> element of SOAP messages. NETCONF over SOAP messages also comply with the following rules:

- SOAP messages must use the SOAP Envelope namespaces.
- SOAP messages cannot contain the following information:
  - DTD reference.
  - XML processing instructions.

The following example shows a NETCONF over SOAP message for getting all parameters of all interfaces on the device:

```
<env:Envelope xmlns:env="http://www.w3.org/2003/05/soap-envelope">
  <env:Header>
    <auth:Authentication env:mustUnderstand="1"
xmlns:auth="http://www.h3c.com/netconf/base:1.0">
      <auth:AuthInfo>800207F0120020C</auth:AuthInfo>
    </auth:Authentication>
  </env:Header>
  <env:Body>
    <rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <get-bulk>
        <filter type="subtree">
          <top xmlns="http://www.h3c.com/netconf/data:1.0">
            <Ifmgr>
              <Interfaces>
                <Interface/>
              </Interfaces>
            </Ifmgr>
          </top>
        </filter>
      </get-bulk>
    </rpc>
  </env:Body>
```

```
</env:Envelope>
```

How to use NETCONF

You can use NETCONF to manage and configure the device by using the methods in Table 2.

Table 2 NETCONF methods for configuring the device

Configuration tool	Login method	Remarks
CLI	• Console port • SSH • Telnet	To perform NETCONF operations, copy valid NETCONF messages to the CLI in XML view.
Custom user interface	N/A	This method requires a NETCONF over SOAP session or NETCONF over SSH session with the device. To use a NETCONF over SOAP session, you must enable NETCONF over SOAP. Then, NETCONF messages will be encapsulated in SOAP messages for transmission.

Protocols and standards

- RFC 3339, *Date and Time on the Internet: Timestamps*
- RFC 4741, *NETCONF Configuration Protocol*
- RFC 4742, *Using the NETCONF Configuration Protocol over Secure SHell (SSH)*
- RFC 4743, *Using NETCONF over the Simple Object Access Protocol (SOAP)*
- RFC 5277, *NETCONF Event Notifications*
- RFC 5381, *Experience of Implementing NETCONF over SOAP*
- RFC 5539, *NETCONF over Transport Layer Security (TLS)*
- RFC 6241, *Network Configuration Protocol*

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode (see *Security Configuration Guide*) and non-FIPS mode.

NETCONF tasks at a glance

To configure NETCONF, perform the following tasks:

1. Configuring NETCONF to use module-specific namespaces
2. Establishing a NETCONF session
   a. (Optional.) Setting NETCONF session attributes
   b. Establishing NETCONF over SOAP sessions
   c. Establishing NETCONF over SSH sessions
   d. Establishing NETCONF over Telnet or NETCONF over console sessions
   e. Exchanging capabilities
3. (Optional.) Retrieving device configuration information

- Retrieving device configuration and state information
- Retrieving non-default settings
- Retrieving NETCONF information
- Retrieving YANG file content
- Retrieving NETCONF session information

4. (Optional.) Filtering data
   - Table-based filtering
   - Column-based filtering
5. (Optional.) Locking or unlocking the running configuration
   a. Locking the running configuration
   b. Unlocking the running configuration
6. (Optional.) Modifying the configuration
7. (Optional.) Managing configuration files
   - Saving the running configuration
   - Loading the configuration
   - Rolling back the configuration
8. (Optional.) Managing the candidate configuration
9. (Optional.) Performing CLI operations through NETCONF
10. (Optional.) Subscribing to events
11. (Optional.) Terminating NETCONF sessions
12. (Optional.) Returning to the CLI
13. (Optional.) Enabling SNMP notifications for NETCONF

Configuring NETCONF to use module-specific namespaces

About this task

By default, the device uses the common namespace in the NETCONF responses to the requests for YANG file information. Perform this task to configure the device to use module-specific namespaces in the response messages.

The client performs the <get-schema> and <get><netconf-state/></get> operations to obtain information about YANG files on the device. For more information about the operations, see "Retrieving YANG file content" and "Retrieving NETCONF information."

Restrictions and guidelines

You do not need to perform this task if the client sends a hello message that includes the following capability during capability exchange:

```
<capability>urn:h3c:params:netconf:capability:module-specified-namespace:1.0</capability>
```

Procedure

1. Enter system view.

   `system-view`

2. Configure NETCONF to use module-specific namespaces in messages responding to the client request of retrieving YANG file information.

   `netconf capability specific-namespace`

By default, the common namespace is used in messages responding to the client request of retrieving YANG file information.

Establishing a NETCONF session

Restrictions and guidelines for NETCONF session establishment

After a NETCONF session is established, the device automatically sends its capabilities to the client. You must send the capabilities of the client to the device before you can perform any other NETCONF operations.

Before performing a NETCONF operation, make sure no other users are configuring or managing the device. If multiple users simultaneously configure or manage the device, the result might be different from what you expect.

You can use the **aaa session-limit** command to set the maximum number of NETCONF sessions that the device can support. If the upper limit is reached, new NETCONF users cannot access the device. For information about this command, see AAA in *Security Configuration Guide*.

Setting NETCONF session attributes

About this task

NETCONF supports the following types of namespaces:

- **Common namespace**—The common namespace is shared by all modules. In a packet that uses the common namespace, the namespace is indicated in the <top> element, and the modules are listed under the <top> element.

  Example:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <get-bulk>
      <filter type="subtree">
         <top xmlns="http://www.h3c.com/netconf/data:1.0">
            <Ifmgr>
               <Interfaces>
               </Interfaces>
            </Ifmgr>
         </top>
      </filter>
   </get-bulk>
</rpc>
```

- **Module-specific namespace**—Each module has its own namespace. A packet that uses a module-specific namespace does not have the <top> element. The namespace follows the module name.

  Example:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <get-bulk>
      <filter type="subtree">
            <Ifmgr xmlns="http://www.h3c.com/netconf/data:1.0-Ifmgr">
               <Interfaces>
               </Interfaces>
```

```
            </Ifmgr>
        </filter>
    </get-bulk>
</rpc>
```

The common namespace is incompatible with module-specific namespaces. To set up a NETCONF session, the device and the client must use the same type of namespaces. By default, the common namespace is used. If the client does not support the common namespace, use this feature to configure the device to use module-specific namespaces.

Procedure

1. Enter system view.

   system-view

2. Set the NETCONF session idle timeout time.

   netconf { **agent** | **soap** } **idle-timeout** *minute*

Keyword	Description
agent	Specifies the following sessions: • NETCONF over SSH sessions. • NETCONF over Telnet sessions. • NETCONF over console sessions. By default, the idle timeout time is 0, and the sessions never time out.
soap	Specifies the following sessions: • NETCONF over SOAP over HTTP sessions. • NETCONF over SOAP over HTTPS sessions. The default setting is 10 minutes.

3. Enable NETCONF logging.

   netconf log source { **all** | { **agent** | **soap** } * } { **protocol-operation** { **all** | { **action** | **config** | **get** | **set** | **session** | **syntax** | **others** } * } | **row-operation** | **verbose** }

   By default, NETCONF logging is disabled.

4. Enable the XML-to-CLI feature for NETCONF logging.

   netconf log xml2cli enable

   By default, the XML-to-CLI feature for NETCONF logging is disabled.

5. Configure NETCONF to use module-specific namespaces.

   netconf capability specific-namespace

   By default, the common namespace is used.

   For the setting to take effect, you must reestablish the NETCONF session.

Establishing NETCONF over SOAP sessions

About this task

You can use a custom user interface to establish a NETCONF over SOAP session to the device and perform NETCONF operations. NETCONF over SOAP encapsulates NETCONF messages into SOAP messages and transmits the SOAP messages over HTTP or HTTPS.

Restrictions and guidelines

You can add an authentication domain to the <UserName> parameter of a SOAP request. The authentication domain takes effect only on the current request.

The mandatory authentication domain configured by using the **netconf soap domain** command takes precedence over the authentication domain specified in the <UserName> parameter of a SOAP request.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Apply an SSL server policy to the NETCONF over SOAP over HTTPS service.

   netconf soap https ssl-server-policy *policy-name*

   By default, no SSL server policy is applied to the NETCONF over SOAP over HTTPS service.

   The NETCONF over SOAP over HTTPS service will use the SSL server policy to enhance service security. For more information about SSL server policies, see SSL configuration in *Security Configuration Guide*.

   If you modify the applied SSL server policy after NETCONF over SOAP over HTTPS is enabled, the changes apply only to the NETCONF over SOAP over HTTPS sessions established subsequently. They do not take effect on the NETCONF over SOAP over HTTPS sessions that have been established before you modify the policy.

3. Enable NETCONF over SOAP.

   In non-FIPS mode:

   netconf soap { **http** | **https** } **enable**

   In FIPS mode:

   netconf soap https enable

   By default, the NETCONF over SOAP feature is disabled.

4. (Optional.) Set the DSCP value for NETCONF over SOAP packets.

   In non-FIPS mode:

   netconf soap { **http** | **https** } **dscp** *dscp-value*

   In FIPS mode:

   netconf soap https dscp *dscp-value*

   By default, the DSCP value is 0 for NETCONF over SOAP packets.

5. (Optional.) Use an IPv4 ACL to control NETCONF over SOAP access. Only clients permitted by the IPv4 ACL can establish NETCONF over SOAP sessions.

   - Apply an ACL to control TCP connections from IPv4 HTTP clients.

     http acl { *advanced-acl-number* | *basic-acl-number* }

     By default, the device permits TCP connections from all IPv4 HTTP clients.

   - Apply an ACL to control TCP connections from IPv4 HTTPS clients.

     https acl { *advanced-acl-number* | *basic-acl-number* }

     By default, the device permits TCP connections from all IPv4 HTTPS clients.

   - Apply an IPv4 ACL to control NETCONF over SOAP access.

     In non-FIPS mode:

     netconf soap { **http** | **https** } **acl** { *ipv4-acl-number* | **name** *ipv4-acl-name* }

     In FIPS mode:

     netconf soap https acl { *ipv4-acl-number* | **name** *ipv4-acl-name* }

     By default, no IPv4 ACL is applied to control NETCONF over SOAP access.

6. (Optional.) Specify a mandatory authentication domain for NETCONF users.

   netconf soap domain *domain-name*

By default, no mandatory authentication domain is specified for NETCONF users. For information about authentication domains, see *Security Configuration Guide*.

7. Use the custom user interface to establish a NETCONF over SOAP session with the device.

   For information about the custom user interface, see the user guide for the interface.

Establishing NETCONF over SSH sessions

Prerequisites

Before establishing a NETCONF over SSH session, make sure the custom user interface can access the device through SSH.

Procedure

1. Enter system view.

   system-view

2. Enable NETCONF over SSH.

   netconf ssh server enable

   By default, NETCONF over SSH is disabled.

3. Specify the listening port for NETCONF over SSH packets.

   netconf ssh server port *port-number*

   By default, the listening port number is 830.

4. Use the custom user interface to establish a NETCONF over SSH session with the device. For information about the custom user interface, see the user guide for the interface.

Establishing NETCONF over Telnet or NETCONF over console sessions

Restrictions and guidelines

To ensure the format correctness of a NETCONF message, do not enter the message manually. Copy and paste the message.

While the device is performing a NETCONF operation, do not perform any other operations, such as pasting a NETCONF message or pressing **Enter**.

For the device to identify NETCONF messages, you must add end mark **]]>]]>** at the end of each NETCONF message. Examples in this document do not necessarily have this end mark. Do add the end mark in actual operations.

Prerequisites

To establish a NETCONF over Telnet session or a NETCONF over console session, first log in to the device through Telnet or the console port.

Procedure

To enter XML view, execute the following command in user view:

xml

If the XML view prompt appears, the NETCONF over Telnet session or NETCONF over console session is established successfully.

Exchanging capabilities

About this task

After a NETCONF session is established, the device sends its capabilities to the client. You must use a hello message to send the capabilities of the client to the device before you can perform any other NETCONF operations.

Hello message from the device to the client

```
<?xml version="1.0" encoding="UTF-8"?><hello
xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"><capabilities>...</capabilities><sess
ion-id>1</session-id></hello>]]>]]>
```

The <capabilities> element carries the capabilities supported by the device. The supported capabilities vary by device model.

The <session-id> element carries the unique ID assigned to the NETCONF session.

Hello message from the client to the device

After receiving the hello message from the device, copy the following hello message to notify the device of the capabilities supported by the client:

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
     capability-set
   </capability>
 </capabilities>
</hello>
```

Item	Description
capability-set	Specifies a set of capabilities supported by the client. Use the <capability> and </capability> tags to enclose each user-defined capability set.

Retrieving device configuration information

Restrictions and guidelines for device configuration retrieval

As a best practice to ensure retrieval performance, limit the number of data entries for each <get-bulk> or <get> operation to 1000. If the number of data entries to retrieve is higher than 1000, use the <get-bulk> operation as a best practice.

During a <get>, <get-bulk>, <get-config>, or <get-bulk-config> operation, NETCONF replaces unidentifiable characters in the retrieved data with question marks (?) before sending the data to the client. If the process for a relevant module is not started yet, the operation returns the following message:

```
<?xml version="1.0"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <data/>
</rpc-reply>
```

The <get><netconf-state/></get> operation does not support data filtering.

For more information about the NETCONF operations, see the NETCONF XML API references for the device.

Retrieving device configuration and state information

You can use the following NETCONF operations to retrieve device configuration and state information:

- **<get> operation**—Retrieves all device configuration and state information that matches the specified conditions.
- **<get-bulk> operation**—Retrieves data entries starting from the data entry next to the one with the specified index. One data entry contains a device configuration entry and a state information entry. The returned output does not include the index information.

The <get> message and <get-bulk> message share the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <getoperation>
    <filter>
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
          Specify the module, submodule, table name, and column name
      </top>
    </filter>
  </getoperation>
</rpc>
```

Item	Description
getoperation	Operation name, **get** or **get-bulk**.
filter	Specifies the filtering conditions, such as the module name, submodule name, table name, and column name. • If you specify a module name, the operation retrieves the data for the specified module. If you do not specify a module name, the operation retrieves the data for all modules. • If you specify a submodule name, the operation retrieves the data for the specified submodule. If you do not specify a submodule name, the operation retrieves the data for all submodules. • If you specify a table name, the operation retrieves the data for the specified table. If you do not specify a table name, the operation retrieves the data for all tables. • If you specify only the index column, the operation retrieves the data for all columns. If you specify the index column and any other columns, the operation retrieves the data for the index column and the specified columns.

A <get-bulk> message can carry the **count** and **index** attributes.

Item	Description
index	Specifies the index. If you do not specify this item, the index value starts with 1 by default.
count	Specifies the data entry quantity. The **count** attribute complies with the following rules: • The **count** attribute can be placed in the module node and table node. In other nodes, it cannot be resolved.

Item	Description
	• When the **count** attribute is placed in the module node, a descendant node inherits this count attribute if the descendant node does not contain the count attribute. • The <get-bulk> operation retrieves all the rest data entries starting from the data entry next to the one with the specified index if either of the following conditions occurs: ○ You do not specify the **count** attribute. ○ The number of matching data entries is less than the value of the **count** attribute.

The following <get-bulk> message example specifies the **count** and **index** attributes:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:xc="http://www.h3c.com/netconf/base:1.0">
  <get-bulk>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0"
xmlns:base="http://www.h3c.com/netconf/base:1.0">
        <Syslog>
          <Logs xc:count="5">
            <Log>
              <Index>10</Index>
           </Log>
            </Logs>
        </Syslog>
      </top>
    </filter>
  </get-bulk>
</rpc>
```

When retrieving interface information, the device cannot identify whether an integer value for the <IfIndex> element represents an interface name or index. When retrieving VPN instance information, the device cannot identify whether an integer value for the <vrfindex> element represents a VPN name or index. To resolve the issue, you can use the **valuetype** attribute to specify the value type.

The **valuetype** attribute has the following values:

Value	Description
name	The element carries a name.
index	The element carried an index.
auto	Default value. The device uses the value of the element as an index for information matching. If no match is found, the device uses the value as a name for information matching.

The following example specifies an index-type value for the <IfIndex> element:

```
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <getoperation>
    <filter>
      <top xmlns="http://www.h3c.com/netconf/config:1.0"
xmlns:base="http://www.h3c.com/netconf/base:1.0">
       <VLAN>
```

```
        <TrunkInterfaces>
          <Interface>
            <IfIndex base:valuetype="index">1</IfIndex>
          </Interface>
        </TrunkInterfaces>
       </VLAN>
      </top>
    </filter >
  </getoperation>
</rpc>
```

If the <get> or < get-bulk> operation succeeds, the device returns the retrieved data in the following format:

```
<?xml version="1.0"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <data>
     Device state and configuration data
  </data>
</rpc-reply>
```

Retrieving non-default settings

The <get-config> and <get-bulk-config> operations are used to retrieve all non-default settings. The <get-config> and <get-bulk-config> messages can contain the <filter> element for filtering data.

The <get-config> and <get-bulk-config> messages are similar. The following is a <get-config> message example:

```
<?xml version="1.0"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-config>
    <source>
      <datastore/>
    </source>
    <filter>
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
          Specify the module name, submodule name, table name, and column name
      </top>
    </filter>
  </get-config>
</rpc>
```

The value for *datastore* can be one of the following:

Value	Description
running	Specifies the running configuration datastore.
candidate	Specifies the candidate configuration datastore.

If the <get-config> or <get-bulk-config> operation succeeds, the device returns the retrieved data in the following format:

```
<?xml version="1.0"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
```

```
  <data>
    Data matching the specified filter
  </data>
</rpc-reply>
```

Retrieving NETCONF information

Use the <get><netconf-state/></get> message to retrieve NETCONF information.

Copy the following text to the client to retrieve NETCONF information:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="m-641" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get>
    <filter type='subtree'>
      <netconf-state xmlns='urn:ietf:params:xml:ns:yang:ietf-netconf-monitoring'>
         <getType/>
      </netconf-state>
    </filter>
  </get>
</rpc>
```

If you do not specify a value for *getType*, the retrieval operation retrieves all NETCONF information.

The value for *getType* can be one of the following operations:

Operation	Description
capabilities	Retrieves device capabilities.
datastores	Retrieves databases from the device.
schemas	Retrieves the list of the YANG file names from the device.
sessions	Retrieves session information from the device.
statistics	Retrieves NETCONF statistics.

If the <get><netconf-state/></get> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <data>
     Retrieved NETCONF information
  </data>
</rpc-reply>
```

Retrieving YANG file content

YANG files save the NETCONF operations supported by the device. You can identify the supported operations by retrieving and analyzing the content of YANG files.

YANG files are integrated in the device software and are named in the format of *yang_identifier@yang_version.yang*. You cannot view the YANG file names by executing the **dir** command. For information about retrieving the YANG file names, see "Retrieving NETCONF information."

Copy the following text to the client to retrieve the YANG file named **H3C-syslog-data@2019-01-01.yang**:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-schema xmlns='urn:ietf:params:xml:ns:yang:ietf-netconf-monitoring'>
    <identifier>H3C-syslog-data</identifier>
    <version>2019-01-01</version>
    <format>yang</format>
  </get-schema>
</rpc>
```

If the <get-schema> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <data>
     Content of the specified YANG file
  </data>
</rpc-reply>
```

Retrieving NETCONF session information

Use the <get-sessions> operation to retrieve NETCONF session information of the device.

Copy the following message to the client to retrieve NETCONF session information from the device:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-sessions/>
</rpc>
```

If the <get-sessions> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-sessions>
    <Session>
      <SessionID>Configuration session ID</SessionID>
      <Line>Line information</Line>
      <UserName>Name of the user creating the session</UserName>
      <Since>Time when the session was created</Since>
      <LockHeld>Whether the session holds a lock</LockHeld>
    </Session>
  </get-sessions>
</rpc-reply>
```

Example: Retrieving a data entry for the interface table

Network configuration

Retrieve a data entry for the interface table.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>urn:ietf:params:netconf:base:1.0</capability>
  </capabilities>
</hello>
```

Retrieve a data entry for the interface table.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:web="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-bulk>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0"
xmlns:web="http://www.h3c.com/netconf/base:1.0">
        <Ifmgr>
          <Interfaces web:count="1">
          </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get-bulk>
</rpc>
```

Verifying the configuration

If the client receives the following text, the <get-bulk> operation is successful:

```
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:web="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="100">
  <data>
    <top xmlns="http://www.h3c.com/netconf/data:1.0">
      <Ifmgr>
        <Interfaces>
          <Interface>
            <IfIndex>3</IfIndex>
            <Name>GigabitEthernet1/0/2</Name>
            <AbbreviatedName>GE1/0/2</AbbreviatedName>
            <PortIndex>3</PortIndex>
            <ifTypeExt>22</ifTypeExt>
            <ifType>6</ifType>
            <Description>GigabitEthernet1/0/2 Interface</Description>
            <AdminStatus>2</AdminStatus>
            <OperStatus>2</OperStatus>
            <ConfigSpeed>0</ConfigSpeed>
            <ActualSpeed>100000</ActualSpeed>
            <ConfigDuplex>3</ConfigDuplex>
            <ActualDuplex>1</ActualDuplex>
          </Interface>
        </Interfaces>
      </Ifmgr>
    </top>
  </data>
```

```
</rpc-reply>
```

Example: Retrieving non-default configuration data

Network configuration

Retrieve all non-default configuration data.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
            urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Retrieve all non-default configuration data.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-config>
    <source>
      <running/>
    </source>
  </get-config>
</rpc>
```

Verifying the configuration

If the client receives the following text, the <get-config> operation is successful:

```
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:web="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="100">
    <data>
        <top xmlns="http://www.h3c.com/netconf/config:1.0">
            <Ifmgr>
                <Interfaces>
                    <Interface>
                        <IfIndex>1307</IfIndex>
                        <Shutdown>1</Shutdown>
                    </Interface>
                    <Interface>
                        <IfIndex>1308</IfIndex>
                        <Shutdown>1</Shutdown>
                    </Interface>
                    <Interface>
                        <IfIndex>1309</IfIndex>
                        <Shutdown>1</Shutdown>
                    </Interface>
                    <Interface>
                        <IfIndex>1311</IfIndex>
```

```
                            <VlanType>2</VlanType>
                    </Interface>
                    <Interface>
                        <IfIndex>1313</IfIndex>
                            <VlanType>2</VlanType>
                    </Interface>
                </Interfaces>
            </Ifmgr>
            <Syslog>
                <LogBuffer>
                    <BufferSize>120</BufferSize>
                </LogBuffer>
            </Syslog>
            <System>
                <Device>
                    <SysName>Sysname</SysName>
                    <TimeZone>
                        <Zone>+11:44</Zone>
                        <ZoneName>beijing</ZoneName>
                    </TimeZone>
                </Device>
            </System>
        </top>
    </data>
</rpc-reply>
```

Example: Retrieving syslog configuration data

Network configuration

Retrieve configuration data for the Syslog module.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
            urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Retrieve configuration data for the Syslog module.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-config>
    <source>
      <running/>
    </source>
```

```
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Syslog/>
      </top>
    </filter>
  </get-config>
</rpc>
```

Verifying the configuration

If the client receives the following text, the <get-config> operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="100">
    <data>
        <top xmlns="http://www.h3c.com/netconf/config:1.0">
            <Syslog>
                <LogBuffer>
                    <BufferSize>120</BufferSize>
                </LogBuffer>
            </Syslog>
        </top>
    </data>
</rpc-reply>
```

Example: Retrieving NETCONF session information

Network configuration

Get NETCONF session information.

Procedure

Enter XML view.

```
<Sysname> xml
```

Copy the following message to the client to exchange capabilities with the device:

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
           urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Copy the following message to the client to get the current NETCONF session information on the device:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-sessions/>
</rpc>
```

Verifying the configuration

If the client receives a message as follows, the operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="100">
    <get-sessions>
        <Session>
            <SessionID>1</SessionID>
            <Line>vty0</Line>
            <UserName></UserName>
            <Since>2019-01-07T00:24:57</Since>
            <LockHeld>false</LockHeld>
        </Session>
    </get-sessions>
</rpc-reply>
```

The output shows the following information:

- The session ID of an existing NETCONF session is 1.
- The login user type is vty0.
- The login time is 2019-01-07T00:24:57.
- The user does not hold the lock of the configuration.

Filtering data

About data filtering

You can define a filter to filter information when you perform a <get>, <get-bulk>, <get-config>, or <get-bulk-config> operation. Data filtering includes the following types:

- **Table-based filtering**—Filters table information.
- **Column-based filtering**—Filters information for a single column.

Restrictions and guidelines for data filtering

For table-based filtering to take effect, you must configure table-based filtering before column-based filtering.

Table-based filtering

About this task

The namespace is **http://www.h3c.com/netconf/base:1.0**. The attribute name is **filter**. For information about the support for table-based match, see the NETCONF XML API references.

Copy the following text to the client to retrieve the longest data with IP address **1.1.1.0** and mask length **24** from the IPv4 routing table:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0">
  <get>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Route>
         <Ipv4Routes>
           <RouteEntry h3c:filter="IP 1.1.1.0 MaskLen 24 longer"/>
         </Ipv4Routes>
```

```
        </Route>
      </top>
    </filter>
  </get>
</rpc>
```

Restrictions and guidelines

To use table-based filtering, specify a match criterion for the **filter** row attribute.

Column-based filtering

About this task

Column-based filtering includes full match filtering, regular expression match filtering, and conditional match filtering. Full match filtering has the highest priority and conditional match filtering has the lowest priority. When more than one filtering criterion is specified, the one with the highest priority takes effect.

Full match filtering

You can specify an element value in an XML message to implement full match filtering. If multiple element values are provided, the system returns the data that matches all the specified values.

Copy the following text to the client to retrieve configuration data of all interfaces in UP state:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Ifmgr>
          <Interfaces>
            <Interface>
              <AdminStatus>1</AdminStatus>
            </Interface>
          </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get>
</rpc>
```

You can also specify an attribute name that is the same as a column name of the current table at the row to implement full match filtering. The system returns only configuration data that matches this attribute name. The XML message equivalent to the above element-value-based full match filtering is as follows:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get>
    <filter type="subtree">
      <top
xmlns="http://www.h3c.com/netconf/data:1.0"xmlns:data="http://www.h3c.com/netconf/dat
a:1.0">
        <Ifmgr>
          <Interfaces>
            <Interface data:AdminStatus="1"/>
          </Interfaces>
```

```
        </Ifmgr>
      </top>
    </filter>
  </get>
</rpc>
```

The above examples show that both element-value-based full match filtering and attribute-name-based full match filtering can retrieve the same index and column information for all interfaces in up state.

Regular expression match filtering

To implement a complex data filtering with characters, you can add a **regExp** attribute for a specific element.

The supported data types include integer, date and time, character string, IPv4 address, IPv4 mask, IPv6 address, MAC address, OID, and time zone.

Copy the following text to the client to retrieve the descriptions of interfaces, of which all the characters must be upper-case letters from A to Z:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0">
  <get-config>
    <source>
      <running/>
    </source>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Ifmgr>
          <Interfaces>
            <Interface>
              <Description h3c:regExp="^[A-Z]*$"/>
            </Interface>
          </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get-config>
</rpc>
```

Conditional match filtering

To implement a complex data filtering with digits and character strings, you can add a **match** attribute for a specific element. Table 3 lists the conditional match operators.

Table 3 Conditional match operators

Operation	Operator	Remarks
More than	match="more:*value*"	More than the specified value. The supported data types include date, digit, and character string.
Less than	match="less:*value*"	Less than the specified value. The supported data types include date, digit, and character string.
Not less than	match="notLess:*value*"	Not less than the specified *value*. The supported data types include date, digit, and character string.
Not more than	match="notMore:*value*"	Not more than the specified *value*. The supported data types

Operation	Operator	Remarks
		include date, digit, and character string.
Equal	match="equal:*value*"	Equal to the specified *value*. The supported data types include date, digit, character string, OID, and BOOL.
Not equal	match="notEqual:*value*"	Not equal to the specified *value*. The supported data types include date, digit, character string, OID, and BOOL.
Include	match="include:*string*"	Includes the specified *string*. The supported data types include only character string.
Not include	match="exclude:*string*"	Excludes the specified *string*. The supported data types include only character string.
Start with	match="startWith:*string*"	Starts with the specified *string*. The supported data types include character string and OID.
End with	match="endWith:*string*"	Ends with the specified *string.* The supported data types include only character string.

Copy the following text to the client to retrieve extension information about the entity whose CPU usage is more than 50%:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0">
  <get>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Device>
          <ExtPhysicalEntities>
            <Entity>
              <CpuUsage h3c:match="more:50"></CpuUsage>
            </Entity>
          </ExtPhysicalEntities>
        </Device>
      </top>
    </filter>
  </get>
</rpc>
```

Example: Filtering data with regular expression match

Network configuration

Retrieve all data including **Gig** in the **Description** column of the Interfaces table under the Ifmgr module.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
           urn:ietf:params:netconf:base:1.0
```

```
    </capability>
  </capabilities>
</hello>
```

Retrieve all data including **Gig** in the **Description** column of the Interfaces table under the Ifmgr module.

```
<?xml version="1.0"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0">
  <get>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Ifmgr>
          <Interfaces>
            <Interface>
              <Description h3c:regExp="(Gig)+"/>
            </Interface>
          </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get>
</rpc>
```

Verifying the configuration

If the client receives the following text, the operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0" message-id="100">
    <data>
        <top xmlns="http://www.h3c.com/netconf/data:1.0">
            <Ifmgr>
                <Interfaces>
                    <Interface>
                        <IfIndex>2681</IfIndex>
                        <Description>GigabitEthernet1/0/1 Interface</Description>
                    </Interface>
                    <Interface>
                        <IfIndex>2685</IfIndex>
                        <Description>GigabitEthernet1/0/2 Interface</Description>
                    </Interface>
                    <Interface>
                        <IfIndex>2689</IfIndex>
                        <Description>GigabitEthernet1/0/3 Interface</Description>
                    </Interface>
                 <Interface>
            </Ifmgr>
        </top>
    </data>
</rpc-reply>
```

Example: Filtering data by conditional match

Network configuration

Retrieve data in the **Name** column with the ifindex value not less than 5000 in the Interfaces table under the Ifmgr module.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
           urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Retrieve data in the **Name** column with the ifindex value not less than 5000 in the Interfaces table under the Ifmgr module.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0">
  <get>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Ifmgr>
          <Interfaces>
            <Interface>
              <IfIndex h3c:match="notLess:5000"/>
              <Name/>
            </Interface>
          </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get>
</rpc>
```

Verifying the configuration

If the client receives the following text, the operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0" message-id="100">
    <data>
        <top xmlns="http://www.h3c.com/netconf/data:1.0">
            <Ifmgr>
                <Interfaces>
                    <Interface>
                        <IfIndex>7241</IfIndex>
                        <Name>NULL0</Name>
```

```
                    </Interface>
                </Interfaces>
            </Ifmgr>
        </top>
    </data>
</rpc-reply>
```

Locking or unlocking the running configuration

About configuration locking and unlocking

Multiple methods are available for configuring the device, such as CLI, NETCONF, and SNMP. Before configuring, managing, or troubleshooting the device, you can lock the configuration to prevent other users from changing the device configuration. After you lock the configuration, only you can perform <edit-config> operations to change the configuration or unlock the configuration. Other users can only read the configuration.

If you close your NETCONF session, the system unlocks the configuration. You can also manually unlock the configuration.

Restrictions and guidelines for configuration locking and unlocking

The <lock> operation locks the running configuration of the device. You cannot use it to lock the configuration for a specific module.

Locking the running configuration

Copy the following text to the client to lock the running configuration:

```
<?xml version="1.0" encoding="UTF-8"?>
  <rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
    <lock>
      <target>
        <datastore/>
      </target>
    </lock>
  </rpc>
```

The value for *datastore* can be one of the following:

Value	Description
running	Specifies the running configuration datastore.
candidate	Specifies the candidate configuration datastore.

If the <lock> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
  <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
     <ok/>
</rpc-reply>
```

Unlocking the running configuration

Copy the following text to the client to unlock the running configuration:

```
<?xml version="1.0" encoding="UTF-8"?>
  <rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
    <unlock>
      <target>
        <datastore/>
      </target>
    </unlock>
  </rpc>
```

The value for *datastore* can be one of the following:

Value	Description
running	Specifies the running configuration datastore.
candidate	Specifies the candidate configuration datastore.

If the <unlock> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
    <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
</rpc-reply>
```

Example: Locking the running configuration

Network configuration

Lock the device configuration so other users cannot change the device configuration.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
    <capabilities>
        <capability>
            urn:ietf:params:netconf:base:1.0
        </capability>
    </capabilities>
</hello>
```

Lock the configuration.

```
<?xml version="1.0" encoding="UTF-8"?>
  <rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
    <lock>
      <target>
        <running/>
      </target>
    </lock>
```

```
  </rpc>
```

Verifying the configuration

If the client receives the following response, the <lock> operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
    <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
</rpc-reply>
```

If another client sends a lock request, the device returns the following response:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<rpc-error>
  <error-type>protocol</error-type>
  <error-tag>lock-denied</error-tag>
  <error-severity>error</error-severity>
  <error-message xml:lang="en"> Lock failed because the NETCONF lock is held by another
session.</error-message>
  <error-info>
    <session-id>1</session-id>
  </error-info>
  </rpc-error>
</rpc-reply>
```

The output shows that the <lock> operation failed. The client with session ID 1 is holding the lock,

Modifying the configuration

About the <edit-config> operation

The <edit-config> operation includes the following operations: merge, create, replace, remove, delete, default-operation, error-option, test-option, and incremental. For more information about the operations, see "Supported NETCONF operations."

Procedure

Copy the following text to perform the <edit-config> operation:

```
<?xml version="1.0"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <edit-config>
    <target><datastore/></target>
    <error-option>
       error-option
    </error-option>
    <config>
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        Specify the module name, submodule name, table name, and column name
      </top>
    </config>
  </edit-config>
```

```
</rpc>
```

The value for *datastore* can be one of the following:

Value	Description
running	Specifies the running configuration datastore.
candidate	Specifies the candidate configuration datastore.

When you modify the running configuration datastore, you can also specify the <error-option> element. This element indicates the action to take in response to an error that occurs during the operation. The following are available error options:

Value	Description
stop-on-error	Stops the <edit-config> operation.
continue-on-error	Continues the <edit-config> operation.
rollback-on-error	Rolls back the configuration to the configuration before the <edit-config> operation was performed. By default, an <edit-config> operation cannot be performed while the device is rolling back the configuration. If the rollback time exceeds the maximum time that the client can wait, the client determines that the <edit-config> operation has failed and performs the operation again. Because the previous rollback is not completed, the operation triggers another rollback. If this process repeats itself, CPU and memory resources will be exhausted and the device will reboot. To allow an <edit-config> operation to be performed during a configuration rollback, perform an <action> operation to change the value of the **DisableEditConfigWhenRollback** attribute to **false**.

If the <edit-config> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0">
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <ok/>
</rpc-reply>
```

If the device supports the **netconf log xml2cli enable** command, it includes <capability>urn:h3c:params:netconf:capability:h3c-xml2cli:1.0</capability> in the capability set exchanged with NETCONF clients. If a client also supports this capability, the device will return the commands for the <edit-config> operations it has successfully executed for that client. The following is the response message format:

```
<?xml version="1.0">
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <ok/>
  <xml2cli>
    Specified module name, submodule name, <cli> column, and target column name and table name in the <edit-config> request
  </xml2cli>
</rpc-reply>
```

The <cli> column contains the commands for the operations on the requested columns in the <edit-config> request. The following is a sample response:

```
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
   <xml2cli>
```

```
      <GRPC>
         <Configuration xc:operation="merge"
xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
            <cli>grpc enable
grpc port 50052
grpc idle-timeout 25</cli>
            <State>enable</State>
            <Port>50052</Port>
            <IdleTime>25</IdleTime>
         </Configuration>
      </GRPC>
   </xml2cli>
</rpc-reply>
```

You can also perform the <get> operation to verify that the current element value is the same as the value specified through the <edit-config> operation.

Example: Modifying the configuration

Network configuration

Change the log buffer size for the Syslog module to 512.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>urn:ietf:params:netconf:base:1.0</capability>
  </capabilities>
</hello>
```

Change the log buffer size for the Syslog module to 512.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:web="urn:ietf:params:xml:ns:netconf:base:1.0">
  <edit-config>
    <target>
      <running/>
    </target>
    <config>
      <top xmlns="http://www.h3c.com/netconf/config:1.0" web:operation="merge">
        <Syslog>
          <LogBuffer>
            <BufferSize>512</BufferSize>
          </LogBuffer>
        </Syslog>
      </top>
    </config>
  </edit-config>
</rpc>
```

Verifying the configuration

If the client receives the following text, the <edit-config> operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
    <ok/>
</rpc-reply>
```

Saving the running configuration

About the <save> operation

A <save> operation saves the running configuration to a configuration file and specifies the file as the main next-startup configuration file.

Restrictions and guidelines

The <save> operation is resource intensive. Do not perform this operation when system resources are heavily occupied.

Procedure

Copy the following text to the client:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save OverWrite="false" Binary-only="false" Safely="false">
    <file>Configuration file name</file>
  </save>
</rpc>
```

Item	Description
file	Specifies a .cfg configuration file by its name. The name must start with the storage medium name. If you specify the file column, a file name is required. If the **Binary-only** attribute is **false**, the device saves the running configuration to both the text and binary configuration files. • If the specified .cfg file does not exist, the device creates the binary and text configuration files to save the running configuration. • If you do not specify the file column, the device saves the running configuration to the text and binary next-startup configuration files.
OverWrite	Determines whether to overwrite the specified file if the file already exists. The following values are available: • **true**—Overwrite the file. • **false**—Do not overwrite the file. The running configuration cannot be saved, and the system displays an error message. The default value is **true**.
Binary-only	Determines whether to save the running configuration only to the binary configuration file. The following values are available: • **true**—Save the running configuration only to the binary configuration file. ○ If the file column specifies a nonexistent file, the <save> operation fails.

Item	Description
	○ If you do not specify the file column, the device attempts to save the running configuration to the binary file for the .cfg main next-startup configuration file. If the .cfg main next-startup configuration file does not exist, the <save> operation fails. • **false**—Save the running configuration to both the .cfg and binary configuration files. For more information, see the description for the file column in this table. Saving the running configuration only to the binary configuration file reduces the amount of time required to save the running configuration. The default value is **false**. The current software version does not support setting this column to **true**.
Safely	Determines whether to save the configuration file in safe mode or fast mode. Available values: • **true**—Saves the configuration in safe mode. In this mode, the system saves the configuration in a temporary file and starts overwriting the target next-startup configuration file after the save operation is complete. If a disruptive event such as a reboot or power failure occurs during the save operation, the device can still start up with the original next-startup configuration file. • **false**—Saves the configuration in fast mode. In this mode, the system directly overwrites the target next-startup configuration file. If a disruptive event such as a reboot, power failure, out-of-memory, out-of-storage event occurs during this process, the next-startup configuration file is lost. You must specify a new startup configuration file after the device reboots. By default, the Safely attribute is set to **false** if you do not specify a value for it. Safe mode is slower than fast mode, but more secure. As a best practice, use the safe mode if the power source is not reliable or you are remotely configuring the device.

If the <save> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
    <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
</rpc-reply>
```

Example: Saving the running configuration

Network configuration

Save the running configuration to the **config.cfg** file.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
            urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Save the running configuration of the device to the **config.cfg** file.

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save>
    <file>config.cfg</file>
  </save>
</rpc>
```

Verifying the configuration

If the client receives the following response, the <save> operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
    <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
</rpc-reply>
```

Loading the configuration

About the <load> operation

The <load> operation merges the configuration from a configuration file into the running configuration as follows:

- Loads settings that do not exist in the running configuration.
- Overwrites settings that already exist in the running configuration.

Restrictions and guidelines

When you perform a <load> operation, follow these restrictions and guidelines:

- The <load> operation is resource intensive. Do not perform this operation when the system resources are heavily occupied.
- Some settings in a configuration file might conflict with the existing settings. For the settings in the file to take effect, delete the existing conflicting settings, and then load the configuration file.

Procedure

Copy the following text to the client to load a configuration file for the device:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <load>
   <file>Configuration file name</file>
  </load>
</rpc>
```

The configuration file name must start with the storage media name and end with the **.cfg** extension.

If the <load> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
    <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
</rpc-reply>
```

Rolling back the configuration

Restrictions and guidelines

The <rollback> operation is resource intensive. Do not perform this operation when the system resources are heavily occupied.

By default, an <edit-config> operation cannot be performed while the device is rolling back the configuration. To allow an <edit-config> operation to be performed during a configuration rollback, perform an <action> operation to change the value of the **DisableEditConfigWhenRollback** attribute to **false**.

Rolling back the configuration based on a configuration file

Copy the following text to the client to roll back the running configuration to the configuration in a configuration file:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <rollback>
    <file>Specify the configuration file name</file>
  </rollback>
</rpc>
```

If the <rollback> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
    <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
</rpc-reply>
```

Rolling back the configuration based on a rollback point

About this task

You can roll back the running configuration based on a rollback point when one of the following situations occurs:

- A NETCONF client sends a rollback request.
- The NETCONF session idle time is longer than the rollback idle timeout time.
- A NETCONF client is unexpectedly disconnected from the device.

Restrictions and guidelines

Multiple users might simultaneously configure the device. As a best practice, lock the system before rolling back the configuration to prevent other users from modifying the running configuration.

Procedure

1. Lock the running configuration. For more information, see "Locking or unlocking the running configuration."
2. Enable configuration rollback based on a rollback point.

   # Copy the following text to the client to perform a <save-point>/<begin> operation:

   ```
   <rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
         <save-point>
           <begin>
             <confirm-timeout>100</confirm-timeout>
   ```

```
      </begin>
     </save-point>
</rpc>
```

Item	Description
confirm-timeout	Specifies the rollback idle timeout time in the range of 1 to 65535 seconds. The default is 600 seconds. This item is optional.

If the <save-point/begin> operation succeeds, the device returns a response in the following format:

```
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <data>
    <save-point>
       <commit>
          <commit-id>1</commit-id>
       </commit>
    </save-point>
  </data>
</rpc-reply>
```

3. Modify the running configuration. For more information, see "Modifying the configuration."
4. Mark the rollback point.

   The system supports a maximum of 50 rollback points. If the limit is reached, specify the **force** attribute for the <save-point>/<commit> operation to overwrite the earliest rollback point.

   # Copy the following text to the client to perform a <save-point>/<commit> operation:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save-point>
    <commit>
      <label>SUPPORT VLAN<label>
      <comment>vlan 1 to 100 and interfaces.</comment>
     </commit>
  </save-point>
</rpc>
```

   The <label> and <comment> elements are optional.

   If the <save-point>/<commit> operation succeeds, the device returns a response in the following format:

```
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <data>
    <save-point>
       <commit>
          <commit-id>2</commit-id>
       </commit>
    </save-point>
  </data>
</rpc-reply>
```

5. Retrieve the rollback point configuration records.

   The following text shows the message format for a <save-point/get-commits> request:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save-point>
    <get-commits>
      <commit-id/>
```

```
      <commit-index/>
      <commit-label/>
    </get-commits>
  </save-point>
</rpc>
```

Specify the <commit-id/>, <commit-index/>, or <commit-label/> element to retrieve the specified rollback point configuration records. If no element is specified, the operation retrieves records for all rollback point settings.

Copy the following text to the client to perform a <save-point>/<get-commits> operation:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save-point>
    <get-commits>
      <commit-label>SUPPORT VLAN</commit-label>
    </get-commits>
  </save-point>
</rpc>
```

If the <save-point/get-commits> operation succeeds, the device returns a response in the following format:

```
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <data>
      <save-point>
         <commit-information>
           <CommitID>2</CommitID>
           <TimeStamp>Sun Jan 1 11:30:28 2019</TimeStamp>
           <UserName>test</UserName>
           <Label>SUPPORT VLAN</Label>
         </commit-information>
    </save-point>
  </data>
</rpc-reply>
```

6. Retrieve the configuration data corresponding to a rollback point.

The following text shows the message format for a <save-point>/<get-commit-information> request:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save-point>
    <get-commit-information>
       <commit-information>
         <commit-id/>
         <commit-index/>
         <commit-label/>
       </commit-information>
       <compare-information>
         <commit-id/>
         <commit-index/>
         <commit-label/>
       </compare-information>
    </get-commit-information>
  </save-point>
</rpc>
```

Specify one of the following elements: <commit-id/>, <commit-index/>, and <commit-label/>. The <compare-information> element is optional.

Item	Description
commit-id	Uniquely identifies a rollback point.
commit-index	Specifies 50 most recently configured rollback points. The value of 0 indicates the most recently configured one and 49 indicates the earliest configured one.
commit-label	Specifies a unique label for a rollback point.
get-commit-information	Retrieves the configuration data corresponding to the most recently configured rollback point.

Copy the following text to the client to perform a <save-point>/<get-commit-information> operation:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save-point>
    <get-commit-information>
               <commit-information>
                  <commit-label>SUPPORT VLAN</commit-label>
               </commit-information>
    </get-commit-information>
  </save-point>
</rpc>
```

If the <save-point/get-commit-information> operation succeeds, the device returns a response in the following format:

```
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <data>
     <save-point>
        <commit-information>
           <content>
              …
              interface vlan 1
              …
           </content>
        </commit-information>
     </save-point>
   </data>
</rpc-reply>
```

7. Roll back the configuration based on a rollback point.

The configuration can also be automatically rolled back based on the most recently configured rollback point when the NETCONF session idle timer expires.

Copy the following text to the client to perform a <save-point>/<rollback> operation:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save-point>
    <rollback>
      <commit-id/>
      <commit-index/>
      <commit-label/>
    </rollback>
```

```
  </save-point>
</rpc>
```

Specify one of the following elements: <commit-id/>, <commit-index/>, and <commit-label/>. If no element is specified, the operation rolls back configuration based on the most recently configured rollback point.

Item	Description
commit-id	Uniquely identifies a rollback point.
commit-index	Specifies 50 most recently configured rollback points. The value of 0 indicates the most recently configured one and 49 indicates the earliest configured one.
commit-label	Specifies the unique label of a rollback point.

If the <save-point/rollback> operation succeeds, the device returns a response in the following format:

```
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <ok></ok>
</rpc-reply>
```

8. End the rollback configuration.

   # Copy the following text to the client to perform a <save-point>/<end> operation:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save-point>
    <end/>
  </save-point>
</rpc>
```

   If the <save-point/end> operation succeeds, the device returns a response in the following format:

```
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
</rpc-reply>
```

9. Unlock the configuration. For more information, see "Locking or unlocking the running configuration."

Managing the candidate configuration

About this task

The candidate configuration datastore can be used to store or edit configuration data to be committed without impacting the device's running configuration. A commit operation commits the candidate configuration to the running configuration. After a successful commit operation, the system creates a configuration rollback point.

The candidate configuration datastore is globally unique.

Candidate configuration management tasks at a glance

To manage the candidate configuration, perform the following tasks:

1. Creating a candidate configuration datastore
2. Editing the candidate configuration datastore

3. Commit the candidate configuration

   Choose one of the following tasks:

   - Committing the candidate configuration immediately
   - Committing the candidate configuration only after confirmation

4. (Optional.) Deleting the candidate configuration datastore

   Perform this task to release system resources.

Creating a candidate configuration datastore

About this task

By default, the candidate configuration datastore does not exist. To create a candidate configuration datastore, perform the <copy-config> operation. This operation will create the candidate configuration datastore and copy the running configuration to the candidate configuration datastore. If the candidate configuration datastore already exists, the <copy-config> operation will overwrite existing candidate configuration in the datastore.

Procedure

Copy the following text to the client to create a candidate configuration datastore and copy the running configuration to the candidate configuration datastore:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <copy-config>
    <source>
      <running/>
    </source>
    <target>
     <candidate/>
    </target>
  </copy-config>
</rpc>
```

If the <copy-config> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
</rpc-reply>
```

Editing the candidate configuration datastore

Restrictions and guidelines

As a best practice to prevent other users from changing the candidate configuration and to avoid configuration conflicts, lock the candidate configuration before editing it.

Procedure

1. Lock the candidate configuration. For more information, see "Locking or unlocking the running configuration." Make sure you specify the candidate configuration datastore.
2. (Optional.) Retrieve non-default settings through a <get-config> or <get-bulk-config> operation. For more information, see "Retrieving non-default settings."

   Skip this step if you know the non-default settings.

3. Edit the candidate configuration datastore through an <edit-config> operation. For more information, see "Modifying the configuration." Make sure you specify the candidate configuration datastore.
4. (Optional.) Restore the candidate configuration to the running configuration through a <discard-changes> operation. To perform a <discard-changes> operation, copy the following text to the client:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <discard-changes/>
</rpc>
```

Committing the candidate configuration immediately

Copy the following text to the client to commit the candidate configuration to the running configuration immediately:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <commit/>
</rpc>
```

If the <commit> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
</rpc-reply>
```

Committing the candidate configuration only after confirmation

About this task

A confirmation-required commit supports automatic rollback. If you require confirmation in the commit operation, the system will start a confirmation timeout timer.

- Before the timer expires, you can choose to commit the candidate configuration to the running configuration immediately or roll back the running configuration to the previously committed configuration.
- After the timer expires, the device automatically rolls back the running configuration to the previously committed configuration.

Procedure

1. Copy the following text to the client to commit the candidate configuration to the running configuration only after confirmation:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <commit>
    <confirmed/>
    <confirm-timeout>confirm-timeout</confirm-timeout>
    <persist>config-name</persist>
  </commit>
</rpc>
```

Element	Description
confirm-timeout	Specifies the confirmation timeout timer, in seconds. The default is 600 seconds.

Element	Description
config-name	Specifies the name of the candidate configuration to be committed. The <persist> element enables the candidate configuration to survive after the NETCONF session terminates. • If you include the <persist> element in the commit operation, you can confirm the commit in any NETCONF session. Make sure you specify the <persist-id> element with the same candidate configuration name in the commit confirmation operation. • If you do not include the <persist> element in this commit operation, you can confirm the commit only in the current NETCONF session.

2. Choose one of the following tasks before the confirmation timeout timer expires:
   - Commit the candidate configuration to the running configuration immediately by copying the following text to the client:

     ```
     <rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
       <commit>
         <persist-id>config-name</persist-id>
       </commit>
     </rpc>
     ```

Item	Description
config-name	Specifies the name of the candidate configuration to be committed. The name must be the same as that specified in the <persist> element. If you do not include the <persist> element in the commit operation, you do not need to include the <persist-id> element in the confirmation.

   - Cancel the commit and roll back the running configuration to the previously committed configuration by copying the following text to the client:

     ```
     <rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
       <cancel-commit/>
     </rpc>
     ```

   If the operation succeeds, the device returns a response in the following format:

   ```
   <?xml version="1.0" encoding="UTF-8"?>
   <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
   </rpc-reply>
   ```

Deleting the candidate configuration datastore

Copy the following text to the client to delete the candidate configuration datastore:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <delete-config>
    <target>
      <candidate/>
    </target>
  </delete-config>
```

```
</rpc>
```

If the <delete-config> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
</rpc-reply>
```

Performing CLI operations through NETCONF

About CLI operations through NETCONF

You can enclose command lines in XML messages to configure the device.

Restrictions and guidelines

Performing CLI operations through NETCONF is resource intensive. As a best practice, do not perform the following tasks:

- Enclose multiple command lines in one XML message.
- Use NETCONF to perform a CLI operation when other users are performing NETCONF CLI operations.

Procedure

Copy the following text to the client to execute the commands:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <CLI>
    <Execution>
      Commands
    </Execution>
  </CLI>
</rpc>
```

The <Execution> element can contain multiple commands, with one command on one line.

If the CLI operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <CLI>
    <Execution>
      <![CDATA[Responses to the commands]]>
    </Execution>
  </CLI>
</rpc-reply>
```

Example: Performing CLI operations

Network configuration

Send the **`display vlan`** command to the device.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
            urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Copy the following text to the client to execute the **`display vlan`** command:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <CLI>
    <Execution>
          display vlan
    </Execution>
  </CLI>
</rpc>
```

Verifying the configuration

If the client receives the following text, the operation is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <CLI>
    <Execution><![CDATA[
<Sysname>display vlan
 Total VLANs: Problem 12 1
 The VLANs include:
 1(default)
   ]]>
   </Execution>
  </CLI>
</rpc-reply>
```

Subscribing to events

About event subscription

When an event takes place on the device, the device sends information about the event to NETCONF clients that have subscribed to the event.

Restrictions and guidelines

Event subscription is not supported for NETCONF over SOAP sessions.

A subscription takes effect only on the current session. It is canceled when the session is terminated.

If you do not specify the event stream to be subscribed to, the device sends syslog event notifications to the NETCONF client.

Subscribing to syslog events

Copy the following message to the client to complete the subscription:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <create-subscription xmlns="urn:ietf:params:xml:ns:netconf:notification:1.0">
    <stream>NETCONF</stream>
      <filter>
        <event xmlns="http://www.h3c.com/netconf/event:1.0">
          <Code>code</Code>
            <Group>group</Group>
              <Severity>severity</Severity>
        </event>
      </filter>
      <startTime>start-time</startTime>
      <stopTime>stop-time</stopTime>
   </create-subscription>
</rpc>
```

Item	Description
stream	Specifies the event stream. The name for the syslog event stream is **NETCONF**.
event	Specifies the event. For information about the events to which you can subscribe, see the system log message references for the device.
code	Specifies the mnemonic symbol of the log message.
group	Specifies the module name of the log message.
severity	Specifies the severity level of the log message.
start-time	Specifies the start time of the subscription.
stop-time	Specifies the end time of the subscription.

If the subscription succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns:netconf="urn:ietf:params:xml:ns:netconf:base:1.0">
    <ok/>
</rpc-reply>
```

If the subscription fails, the device returns an error message in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<rpc-error>
   <error-type>error-type</error-type>
   <error-tag>error-tag</error-tag>
   <error-severity>error-severity</error-severity>
   <error-message xml:lang="en">error-message</error-message>
```

```
</rpc-error>
</rpc-reply>
```

For more information about error messages, see RFC 4741.

Subscribing to events monitored by NETCONF

After you subscribe to events as described in this section, NETCONF regularly polls the subscribed events and sends the events that match the subscription condition to the NETCONF client.

Copy the following message to the client to complete the subscription:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<create-subscription xmlns='urn:ietf:params:xml:ns:netconf:notification:1.0'>
  <stream>NETCONF_MONITOR_EXTENSION</stream>
  <filter>
    <NetconfMonitor xmlns='http://www.h3c.com/netconf/monitor:1.0'>
      <XPath>XPath</XPath>
      <Interval>interval</Interval>
      <ColumnConditions>
        <ColumnCondition>
          <ColumnName>ColumnName</ColumnName>
          <ColumnValue>ColumnValue</ColumnValue>
          <ColumnCondition>ColumnCondition</ColumnCondition>
        </ColumnCondition>
      </ColumnConditions>
      <MustIncludeResultColumns>
        <ColumnName>columnName</ColumnName>
      </MustIncludeResultColumns>
    </NetconfMonitor>
  </filter>
<startTime>start-time</startTime>
<stopTime>stop-time</stopTime>
</create-subscription>
</rpc>
```

Item	Description
stream	Specifies the event stream. The name for the event stream is **NETCONF_MONITOR_EXTENSION**.
NetconfMonitor	Specifies the filtering information for the event.
XPath	Specifies the path of the event in the format of *ModuleName*[/*SubmoduleName*]/*TableName*.
interval	Specifies the interval for NETCONF to obtain events that matches the subscription condition. The value range is 1 to 4294967 seconds. The default value is 300 seconds.
ColumnName	Specifies the name of a column in the format of [*GroupName*.]*ColumnName.*
ColumnValue	Specifies the baseline value.
ColumnCondition	Specifies the operator:

Item	Description
	• **more**. • **less**. • **notLess**. • **notMore**. • **equal**. • **notEqual**. • **include**. • **exclude**. • **startWith**. • **endWith**. Choose an operator according to the type of the baseline value.
start-time	Specifies the start time of the subscription.
stop-time	Specifies the end time of the subscription.

If the subscription succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns:netconf="urn:ietf:params:xml:ns:netconf:base:1.0">
    <ok/>
</rpc-reply>
```

Subscribing to events reported by modules

After you subscribe to events as described in this section, the specified modules report subscribed events to NETCONF. NETCONF sends the events to the NETCONF client.

Copy the following message to the client to complete the subscription:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:xs="http://www.h3c.com/netconf/base:1.0">
<create-subscription xmlns="urn:ietf:params:xml:ns:netconf:notification:1.0">
<stream>XXX_STREAM</stream>
   <filter type="subtree">
<event xmlns="http://www.h3c.com/netconf/event:1.0/xxx-features-list-name:1.0">
       <ColumnName xs:condition="Condition">value</ColumnName>
</event>
</filter>
<startTime>start-time</startTime>
<stopTime>stop-time</stopTime>
</create-subscription>
</rpc>
```

Item	Description
stream	Specifies the event stream. Supported event streams vary by device model.
event	Specifies the event name. An event stream includes multiple events. The events use the same namespaces as the event stream.
ColumnName	Specifies the name of a column.

Item	Description
Condition	Specifies the operator: • **more**. • **less**. • **notLess**. • **notMore**. • **equal**. • **notEqual**. • **include**. • **exclude**. • **startWith**. • **endWith**. Choose an operator according to the type of the baseline value.
value	Specifies the baseline value for the column.
start-time	Specifies the start time of the subscription.
stop-time	Specifies the end time of the subscription.

If the subscription succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns:netconf="urn:ietf:params:xml:ns:netconf:base:1.0">
    <ok/>
</rpc-reply>
```

Canceling an event subscription

Copy the following message to the client to cancel a subscription:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <cancel-subscription xmlns="urn:ietf:params:xml:ns:netconf:notification:1.0">
    <stream>XXX_STREAM</stream>
  </cancel-subscription>
</rpc>
```

Item	Description
stream	Specifies the event stream.

If the cancelation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="100">
    <ok/>
</rpc-reply>
```

If the subscription to be canceled does not exist, the device returns an error message in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<rpc-error>
   <error-type>error-type</error-type>
   <error-tag>error-tag</error-tag>
```

```
   <error-severity>error-severity</error-severity>
<error-message xml:lang="en">The subscription stream to be canceled doesn't exist: Stream name=XXX_STREAM.</error-message>
</rpc-error>
</rpc-reply>
```

Canceling an event subscription

Copy the following message to the client to cancel a subscription:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <cancel-subscription xmlns="urn:ietf:params:xml:ns:netconf:notification:1.0">
    <stream>XXX_STREAM</stream>
  </cancel-subscription>
</rpc>
```

Item	Description
stream	Specifies the event stream.

If the cancelation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="100">
    <ok/>
</rpc-reply>
```

If the subscription to be canceled does not exist, the device returns an error message in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<rpc-error>
   <error-type>error-type</error-type>
   <error-tag>error-tag</error-tag>
   <error-severity>error-severity</error-severity>
<error-message xml:lang="en">The subscription stream to be canceled doesn't exist: Stream name=XXX_STREAM.</error-message>
</rpc-error>
</rpc-reply>
```

Example: Subscribing to syslog events

Network configuration

Configure a client to subscribe to syslog events with no time limitation. After the subscription, all events on the device are sent to the client before the session between the device and client is terminated.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
```

```
  <capabilities>
    <capability>
            urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Subscribe to syslog events with no time limitation.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <create-subscription xmlns="urn:ietf:params:xml:ns:netconf:notification:1.0">
    <stream>NETCONF</stream>
  </create-subscription>
</rpc>
```

Verifying the configuration

If the client receives the following response, the subscription is successful:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="100">
    <ok/>
</rpc-reply>
```

When another client (192.168.100.130) logs in to the device, the device sends a notification to the client that has subscribed to all events:

```
<?xml version="1.0" encoding="UTF-8"?>
<notification xmlns="urn:ietf:params:xml:ns:netconf:notification:1.0">
    <eventTime>2019-01-04T12:30:52</eventTime>
    <event xmlns="http://www.h3c.com/netconf/event:1.0">
        <Group>SHELL</Group>
        <Code>SHELL_LOGIN</Code>
        <Slot>1</Slot>
        <Severity>Notification</Severity>
        <context>VTY logged in from 192.168.100.130.</context>
    </event>
</notification>
```

Terminating NETCONF sessions

About NETCONF session termination

NETCONF allows one client to terminate the NETCONF sessions of other clients. A client whose session is terminated returns to user view.

Procedure

Copy the following message to the client to terminate a NETCONF session:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <kill-session>
    <session-id>
      Specified session-ID
    </session-id>
```

```
  </kill-session>
</rpc>
```

If the <kill-session> operation succeeds, the device returns a response in the following format:

```
<?xml version="1.0" encoding="UTF-8"?>
    <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
      <ok/>
</rpc-reply>
```

Example: Terminating another NETCONF session

Network configuration

The user whose session's ID is 1 terminates the session with session ID 2.

Procedure

Enter XML view.

```
<Sysname> xml
```

Notify the device of the NETCONF capabilities supported on the client.

```
<hello xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <capabilities>
    <capability>
            urn:ietf:params:netconf:base:1.0
    </capability>
  </capabilities>
</hello>
```

Terminate the session with session ID 2.

```
<rpc message-id="100"xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <kill-session>
    <session-id>2</session-id>
  </kill-session>
</rpc>
```

Verifying the configuration

If the client receives the following text, the NETCONF session with session ID 2 has been terminated, and the client with session ID 2 has returned from XML view to user view:

```
<?xml version="1.0" encoding="UTF-8"?>
  <rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
    <ok/>
</rpc-reply>
```

Returning to the CLI

Restrictions and guidelines

Before returning from XML view to the CLI, you must first complete capability exchange between the device and the client.

Procedure

Copy the following text to the client to return from XML view to the CLI:

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
```

```
   <close-session/>
</rpc>
```

When the device receives the close-session request, it sends the following response and returns to CLI's user view:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
</rpc-reply>
```

Enabling SNMP notifications for NETCONF

About this task

To report critical NETCONF events to an SNMP NMS, enable SNMP notifications for NETCONF. For NETCONF SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about configuring SNMP, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   `system-view`

2. Enable SNMP notifications for NETCONF.

   `snmp-agent trap enable netconf`

   By default, SNMP notifications are enabled for NETCONF.

Display and maintenance commands for NETCONF

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display current NETCONF service status and global NETCONF service statistics.	**display netconf service**
Display NETCONF session status and statistics.	**display netconf session**
Clear NETCONF service statistics.	**reset netconf service statistics**
Clear NETCONF session statistics.	**reset netconf session statistics**

Supported NETCONF operations

This chapter describes NETCONF operations available with Comware 7.

NETCONF packet strings are case sensitive.

Candidate configuration datastore operations

This section describes operations specific to the candidate configuration datastore. Besides these operations, you can also perform the following operations on the candidate configuration datastore:

- edit-config: create, edit-config: delete, edit-config: merge, edit-config: remove, and edit-config: replace.
- get-bulk-config and get-config.
- lock and unlock.

cancel-commit

Usage guidelines

This operation cancels an ongoing confirmed commit and roll back the running configuration to the previously committed configuration. Perform this task before the confirmation timeout timer expires.

XML example

Cancel an ongoing commit and roll back the running configuration to the previously committed configuration.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <cancel-commit/>
</rpc>
```

commit (immediate commit)

Usage guidelines

This operation commits the candidate configuration to the running configuration immediately.

XML example

Commit the candidate configuration to the running configuration immediately.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <commit/>
</rpc>
```

commit (confirmation-required commit)

Usage guidelines

This operation commits the candidate configuration to the running configuration only after the user confirms the commit. If the user does not confirm the commit, the device automatically rolls back the running configuration to the previously committed configuration.

When you perform this operation, you must specify the <confirmed> node and can specify the following columns:

- **<confirm-timeout>**—Specifies the confirmation timeout timer, in seconds. The default is 600 seconds.
- **<persist>**—Specifies the name of the candidate configuration and enable the candidate configuration to survive after the NETCONF session terminates. If you do not include the <persist> element in this commit operation, you can confirm the commit only in the current NETCONF session.

XML example

Configure name **test** for the candidate configuration and commit the candidate configuration to the running configuration only after a confirmation. Set the confirmation timeout time to 600 seconds.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <commit>
    <confirmed/>
    <confirm-timeout>600</confirm-timeout>
    <persist>test</persist>
  </commit>
</rpc>
```

commit (commit confirmation)

Usage guidelines

This operation confirms to commit the candidate configuration to the running configuration.

You must confirm the commit operation before the confirmation timeout timer expires.

If you include the <persist> element in the confirming commit operation, you must include the <persist-id> element in the confirmation operation. Make sure the value for the <persist-id> element is the same as that specified in the <persist> element.

If you do not include the <persist> element, the commit confirmation operation is the same as the immediate commit operation.

XML example

Confirm to commit the candidate configuration with name **test** to the running configuration.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <commit>
    <persist-id>test</persist-id>
  </commit>
```

```
</rpc>
```

copy-config

Usage guidelines

This operation copies the running configuration to a target configuration datastore. If the target configuration datastore does not exist, this operation creates the datastore.

In the current software version, this operation can only copy the running configuration to the candidate configuration store.

XML example

Copy the running configuration to the candidate configuration datastore.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <copy-config>
    <source>
      <running/>
    </source>
    <target>
     <candidate/>
    </target>
  </copy-config>
</rpc>
```

delete-config

Usage guidelines

This operation deletes a configuration datastore.

In the current software version, this operation can only delete the candidate configuration store.

XML example

Delete the candidate configuration datastore.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <delete-config>
    <target>
      <candidate/>
    </target>
  </delete-config>
</rpc>
```

discard-changes

Usage guidelines

This operation restores the candidate configuration to the previously committed configuration.

XML example

Restore the candidate configuration to the previously committed configuration.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <discard-changes/>
</rpc>
```

CLI operations

CLI

Usage guidelines

This operation executes CLI commands.

A request message encloses commands in the <CLI> element. A response message encloses the command output in the <CLI> element.

You can use the following elements to execute commands:

- **Execution**—Executes commands in user view.
- **Configuration**—Executes commands in system view.

  To use this element, include the **exec-use-channel** attribute and specify a value for the attribute:

  - **false**—Executes commands without using a channel.
  - **true**—Executes commands by using a temporary channel. The channel is automatically closed after the execution.
  - **persist**—Executes commands by using the persistent channel for the session.

    To use the persistent channel, first perform an <Open-channel> operation to open the persistent channel. If you do not do so, the system will automatically open the persistent channel.

    After using the persistent channel, perform a <Close-channel> operation to close the channel and return to system view. If you do not perform an <Open-channel> operation, the system stays in the view and will execute subsequent commands in the view.

  You can also specify the **error-when-rollback** attribute in the <Configuration> element to indicate whether CLI operations are allowed during a configuration error-triggered configuration rollback. This attribute takes effect only if the value of the <error-option> element in <edit-config> operations is set to **rollback-on-error**. It has the following values:

  - **true**—Rejects CLI operation requests and returns error messages.
  - **false (the default)**—Allows CLI operations.

  For CLI operations to be correctly performed, set the value of the **error-when-rollback** attribute to **true**.

NOTE:

To execute commands in a lower-level view of the system view, use the <Configuration> element to enter the view first.

A NETCONF session can have multiple temporary channels, but it can have only one persistent channel.

This operation does not support executing interactive CLI commands.

You cannot execute the **quit** command by using a channel to exit user view.

XML example

Execute the **vlan 3** command in system view without using a channel.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <CLI>
     <Configuration exec-use-channel="false" error-when-rollback="true">vlan
3</Configuration>
  </CLI>
```

```
</rpc>
```

Edit operations

edit-config: create

Usage guidelines

This operation creates target configuration items.

To use the **create** attribute in an <edit-config> operation, you must specify the target configuration item.

- If the table supports creating a target configuration item and the item does not exist, the operation creates the item and configures the item.
- If the specified item already exists, a data-exist error message is returned.

XML example

Set the buffer size to 120.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
  <edit-config>
    <target>
      <running/>
    </target>
    <config>
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Syslog xmlns="http://www.h3c.com/netconf/config:1.0" xc:operation="create">
           <LogBuffer>
             <BufferSize>120</BufferSize>
          </LogBuffer>
        </Syslog>
      </top>
    </config>
  </edit-config>
</rpc>
```

edit-config: delete

Usage guidelines

This operation deletes the specified configuration.

- If the specified target has only the table index, the operation removes all configuration of the specified target, and the target itself.
- If the specified target has the table index and configuration data, the operation removes the specified configuration data of this target.
- If the specified target does not exist, an error message is returned, showing that the target does not exist.

XML example

The syntax is the same as the edit-config message with the **create** attribute. Change the operation attribute from **create** to **delete**.

edit-config: merge

Usage guidelines

This operation commits target configuration items to the running configuration.

To use the **merge** attribute in an <edit-config> operation, you must specify the target configuration item (on a specific level):

- If the specified item exists, the operation directly updates the setting for the item.
- If the specified item does not exist, the operation creates the item and configures the item.
- If the specified item does not exist and it cannot be created, an error message is returned.

XML example

The XML data format is the same as the edit-config message with the **create** attribute. Change the operation attribute from **create** to **merge**.

edit-config: remove

Usage guidelines

This operation removes the specified configuration.

- If the specified target has only the table index, the operation removes all configuration of the specified target, and the target itself.
- If the specified target has the table index and configuration data, the operation removes the specified configuration data of this target.
- If the specified target does not exist, or the XML message does not specify any targets, a success message is returned.

XML example

The syntax is the same as the edit-config message with the **create** attribute. Change the operation attribute from **create** to **remove**.

edit-config: replace

Usage guidelines

This operation replaces the specified configuration.

- If the specified target exists, the operation replaces the configuration of the target with the configuration carried in the message.
- If the specified target does not exist but is allowed to be created, the operation creates the target and then applies the configuration.
- If the specified target does not exist and is not allowed to be created, the operation is not conducted and an invalid-value error message is returned.

XML example

The syntax is the same as the edit-config message with the **create** attribute. Change the operation attribute from **create** to **replace**.

edit-config: *`test-option`*

Usage guidelines

This operation determines whether to commit a configuration item in an <edit-configure> operation. The <test-option> element has one of the following values:

- **test-then-set**—Performs a syntax check, and commits an item if the item passes the check. If the item fails the check, the item is not committed. This is the default test-option value.
- **set**—Commits the item without performing a syntax check.
- **test-only**—Performs only a syntax check. If the item passes the check, a success message is returned. Otherwise, an error message is returned.

XML example

Test the configuration for an interface.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <edit-config>
    <target>
      <running/>
    </target>
    <test-option>test-only</test-option>
    <config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Ifmgr xc:operation="merge">
          <Interfaces>
            <Interface>
              <IfIndex>262</IfIndex>
              <Description>222</Description>
                <ConfigSpeed>2</ConfigSpeed>
                <ConfigDuplex>1</ConfigDuplex>
               </Interface>
          </Interfaces>
        </Ifmgr>
      </top>
    </config>
  </edit-config>
</rpc>
```

edit-config: default-operation

Usage guidelines

This operation modifies the running configuration of the device by using the default operation method.

NETCONF uses one of the following operation attributes to modify configuration: **merge**, **create**, **delete**, and **replace**. If you do not specify an operation attribute for an edit-config message, NETCONF uses the default operation method. The value specified for the <default-operation> element takes effect only once. If you do not specify an operation attribute or the default operation method for an <edit-config> message, **merge** always applies.

The <default-operation> element has the following values:

- **merge**—Default value for the <default-operation> element.
- **replace**—Value used when the operation attribute is not specified and the default operation method is specified as **replace**.
- **none**—Value used when the operation attribute is not specified and the default operation method is specified as **none**. If this value is specified, the <edit-config> operation is used only for schema verification rather than issuing a configuration. If the schema verification is passed, a successful message is returned. Otherwise, an error message is returned.

XML example

Issue an empty operation for schema verification purposes.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <edit-config>
    <target>
      <running/>
    </target>
    <default-operation>none</default-operation>
    <config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Ifmgr>
          <Interfaces>
            <Interface>
              <IfIndex>262</IfIndex>
              <Description>222222</Description>
            </Interface>
          </Interfaces>
        </Ifmgr>
      </top>
    </config>
  </edit-config>
</rpc>
```

edit-config: error-option

Usage guidelines

This operation determines the action to take in case of a configuration error.

The <error-option> element has the following values:

- **stop-on-error**—Stops the operation and returns an error message. This is the default error-option value.
- **continue-on-error**—Continues the operation and returns an error message.
- **rollback-on-error**—Rolls back the configuration.

XML example

Issue the configuration for two interfaces, setting the <error-option> element to **continue-on-error**.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <edit-config>
    <target>
      <running/>
    </target>   <error-option>continue-on-error</error-option>
    <config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Ifmgr xc:operation="merge">
          <Interfaces>
            <Interface>
              <IfIndex>262</IfIndex>
              <Description>222</Description>
                <ConfigSpeed>1024</ConfigSpeed>
```

```
            <ConfigDuplex>1</ConfigDuplex>
          </Interface>
          <Interface>
            <IfIndex>263</IfIndex>
            <Description>333</Description>
              <ConfigSpeed>1024</ConfigSpeed>
              <ConfigDuplex>1</ConfigDuplex>
          </Interface>
        </Interfaces>
      </Ifmgr>
    </top>
  </config>
 </edit-config>
</rpc>
```

edit-config: incremental

Usage guidelines

This operation adds configuration data to a column without affecting the original data.

The **incremental** attribute applies to a list column such as the vlan permitlist column.

You can use the **incremental** attribute for <edit-config> operations with an attribute other than **replace**.

Support for the **incremental** attribute varies by module. For more information, see NETCONF XML API references.

XML example

Add VLANs 1 through 10 to an untagged VLAN list that has untagged VLANs 12 through 15.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:h3c="http://www.h3c.com/netconf/base:1.0">
  <edit-config>
    <target>
      <running/>
    </target>
    <config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <VLAN xc:operation="merge">
          <HybridInterfaces>
            <Interface>
              <IfIndex>262</IfIndex>
              <UntaggedVlanList  h3c:incremental="true">1-10</UntaggedVlanList>
               </Interface>
          </HybridInterfaces>
        </VLAN>
      </top>
    </config>
  </edit-config>
</rpc>
```

Get operations

get

Usage guidelines

This operation retrieves device configuration and state information.

XML example

Retrieve device configuration and state information for the Syslog module.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:xc="http://www.h3c.com/netconf/base:1.0">
  <get>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Syslog>
        </Syslog>
      </top>
    </filter>
  </get>
</rpc>
```

get-bulk

Usage guidelines

This operation retrieves a number of data entries (including device configuration and state information) starting from the data entry next to the one with the specified index.

XML example

Retrieve device configuration and state information for all interfaces.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-bulk>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/data:1.0">
        <Ifmgr>
          <Interfaces xc:count="5" xmlns:xc="http://www.h3c.com/netconf/base:1.0">
            <Interface/>
          </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get-bulk>
</rpc>
```

get-bulk-config

Usage guidelines

This operation retrieves a number of non-default configuration data entries starting from the data entry next to the one with the specified index.

XML example

Retrieve non-default configuration for all interfaces.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <get-bulk-config>
    <source>
      <running/>
    </source>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Ifmgr>
        </Ifmgr>
      </top>
    </filter>
  </get-bulk-config>
</rpc>
```

get-config

Usage guidelines

This operation retrieves non-default configuration data. If no non-default configuration data exists, the device returns a response with empty data.

XML example

Retrieve non-default configuration data for the interface table.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"
xmlns:xc="http://www.h3c.com/netconf/base:1.0">
  <get-config>
    <source>
      <running/>
    </source>
    <filter type="subtree">
      <top xmlns="http://www.h3c.com/netconf/config:1.0">
        <Ifmgr>
           <Interfaces>
              <Interface/>
            </Interfaces>
        </Ifmgr>
      </top>
    </filter>
  </get-config>
</rpc>
```

get-sessions

Usage guidelines

This operation retrieves information about all NETCONF sessions in the system. You cannot specify a session ID to retrieve information about a specific NETCONF session.

XML example

Retrieve information about all NETCONF sessions in the system.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<get-sessions/>
</rpc>
```

Other operations

action

Usage guidelines

This operation sends a non-configuration operation instruction to the device, for example, a reset instruction.

XML example

Clear statistics information for all interfaces.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <action>
    <top xmlns="http://www.h3c.com/netconf/action:1.0">
      <Ifmgr>
            <ClearAllIfStatistics>
                <Clear>
                </Clear>
        </ClearAllIfStatistics>
      </Ifmgr>
    </top>
  </action>
</rpc>
```

close-session

Usage guidelines

This operation terminates the current NETCONF session, unlocks the configuration, and releases the resources (for example, memory) used by the session. After this operation, you exit the XML view.

XML example

Terminate the current NETCONF session.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<close-session/>
</rpc>
```

kill-session

Usage guidelines

This operation terminates the NETCONF session for another user. This operation cannot terminate the NETCONF session for the current user.

XML example

Terminate the NETCONF session with session ID 1.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <kill-session>
    <session-id>1</session-id>
  </kill-session>
</rpc>
```

load

Usage guidelines

This operation loads the configuration. After the device finishes a <load> operation, the configuration in the specified file is merged into the running configuration of the device.

XML example

Merge the configuration in file **a1.cfg** to the running configuration of the device.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <load>
    <file>a1.cfg</file>
  </load>
</rpc>
```

lock

Usage guidelines

This operation locks the configuration. After you lock the configuration, other users cannot perform <edit-config> operations. Other NETCONF operations are allowed.

After a user locks the configuration, other users cannot use NETCONF or any other configuration methods such as CLI and SNMP to configure the device.

XML example

Lock the configuration.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
 <lock>
    <target>
        <running/>
    </target>
</lock>
</rpc>
```

rollback

Usage guidelines

This operation rolls back the configuration. To do so, you must specify the configuration file in the <file> element. After the device finishes the <rollback> operation, the current device configuration is totally replaced with the configuration in the specified configuration file.

XML example

Roll back the running configuration to the configuration in file **1A.cfg**.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <rollback>
    <file>1A.cfg</file>
  </rollback>
</rpc>
```

save

Usage guidelines

This operation saves the running configuration. You can use the <file> element to specify a file for saving the configuration. If the text does not include the file column, the running configuration is automatically saved to the main next-startup configuration file.

The **OverWrite** attribute determines whether the running configuration overwrites the original configuration file when the specified file already exists.

The **Binary-only** attribute determines whether to save the running configuration only to the binary configuration file.

The **Safely** attribute determines whether to save the configuration file in safe mode or fast mode.

XML example

Save the running configuration to file **test.cfg**.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <save OverWrite="false">
    <file>test.cfg</file>
  </save>
</rpc>
```

unlock

Usage guidelines

This operation unlocks the configuration, so other users can configure the device.

Terminating a NETCONF session automatically unlocks the configuration.

XML example

Unlock the configuration.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<unlock>
    <target>
        <running/>
    </target>
</unlock>
</rpc>
```

xml2cli

Usage guidelines

This operation retrieves the CLI command for an operation issued in XML format.

The returned commands are enclosed in a pair of <cli> and </cli> tags.

XML example

Retrieve the commands for an operation on the GRPC/Configuration table.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<xml2cli>
    <GRPC>
        <Configuration>
            <State>enable</State>
            <Port>50052</Port>
            <IdleTime>25</IdleTime>
        </Configuration>
    </GRPC>
 </xml2cli>
</rpc>
```

The device returns the commands for the operation, as shown below:

```
<?xml version="1.0">
<rpc-reply message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<xml2cli>
    <GRPC>
        <Configuration>
            <cli>grpc enable
            grpc port 50052
            grpc idle-timeout 25</cli>
            <State>enable</State>
            <Port>50052</Port>
            <IdleTime>25</IdleTime>
        </Configuration>
    </GRPC>
</xml2cli>
</rpc-reply>
```

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Telemetry Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes gRPC features and configuration tasks.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring gRPC 1

About gRPC 1
gRPC protocol stack layers 1
Network architecture 1
Telemetry technology based on gRPC 1
Telemetry modes 2
Telemetry data model architectures 2
gNMI 2
Protocols 3
FIPS compliance 3
Restrictions and guidelines: gRPC configuration 3
gRPC configuration tasks at a glance 3
Configuring the gRPC dial-in mode 3
Configuring the gRPC service 3
Configuring a gRPC user 4
Enabling gRPC logging in dial-in mode 5
Configuring the gRPC dial-out mode 5
Enabling the gRPC service 5
Configuring sensors 6
Configuring collectors 7
Configuring a subscription 8
Enabling gRPC logging in dial-out mode 10
Specifying the PKI domain for secure communication with collectors 10
Setting the maximum CPU usage for gRPC 11
Enabling event-triggered data to be pushed in a prioritized manner 11
Display and maintenance commands for gRPC 12
gRPC configuration examples 12
Example: Configuring the gRPC dial-in mode 12
Example: Configuring event-triggered telemetry in gRPC dial-out mode 13
Example: Configuring the gRPC dial-out mode 14

Protocol buffer code 15

Protocol buffer code format 15
Proto definition files 16
Proto definition files in dial-in mode 16
Proto definition file in dial-out mode 18
Obtaining proto definition files 20
Example: Developing a gRPC collector-side application 20
Prerequisites 20
Generating the C++ code for the proto definition files 21
Developing the collector-side application (dial-in mode) 21
Developing the collector-side application (dial-out mode) 34
Appendix: Methods and attributes for gNMI operations 35

Configuring gRPC

About gRPC

gRPC is an open source remote procedure call (RPC) system initially developed at Google. It uses HTTP 2.0 and provides network device configuration and management methods that support multiple programming languages.

gRPC protocol stack layers

Table 1 describes the gRPC protocol stack layers.

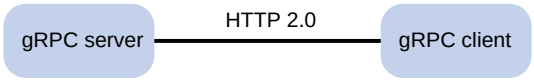
Table 1 gRPC protocol stack layers

Layer	Description
Content layer	Defines the data of the service module. Two peers must notify each other of the data models that they are using.
Protocol buffer encoding layer	Encodes data by using the protocol buffer code format.
gRPC layer	Defines the protocol interaction format for remote procedure calls.
HTTP 2.0 layer	Carries gRPC.
TCP layer	Provides connection-oriented reliable data links.

Network architecture

As shown in Figure 1, the gRPC network uses the client/server model. It uses HTTP 2.0 for packet transport.

Figure 1 gRPC network architecture



The gRPC network uses the following mechanism:

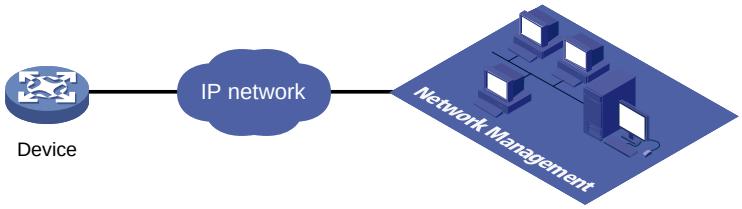
1. The gRPC server listens to connection requests from clients at the gRPC service port.
2. A user runs the gRPC client application to log in to the gRPC server, and uses methods provided in the .proto file to send requests.
3. The gRPC server responds to requests from the gRPC client.

The device can act as the gRPC server or client.

Telemetry technology based on gRPC

Telemetry is a remote data collection technology for monitoring device performance and operating status. H3C telemetry technology uses gRPC to push data from the device to the collectors on the NMSs. As shown in Figure 2, after a gRPC connection is established between the device and NMSs, the NMSs can subscribe to data of modules on the device.

Figure 2 Telemetry technology based on gRPC



Telemetry modes

The device supports the following telemetry modes:

- **Dial-in mode**—The device acts as a gRPC server and the collectors act as gRPC clients. A collector initiates a gRPC connection to the device to subscribe to device data.

  Dial-in mode supports the following operations:

  - **Get**—Obtains device status and settings.
  - **CLI**—Executes commands on the device.
  - **gNMI**—gRPC Network Management Interface operations, which include the following subtypes of operations:
    - **gNMI Capabilities**—Obtains the capacities of the device.
    - **gNMI Get**—Obtains the status and settings of the device.
    - **gNMI Set**—Deploys settings to the device.
    - **gNMI Subscribe**—Subscribes to data push services provided by the device. The data might be generated by periodical data collection, event-triggered data collection, or condition-triggered data collection.
- **Dial-out mode**—The device acts as a gRPC client and the collectors act as gRPC servers. The device initiates gRPC connections to the collectors and pushes device data to the collectors as configured.

Telemetry data model architectures

The device supports the following telemetry data model architectures:

- **Three-layer data model architecture**—Contains the following three layers:
  - **RPC layer**—Defined in public proto definition files **grpc_dialout.proto** and **grpc_dialout_v3.proto**. This layer provides public RPC methods such as the message format.
  - **Telemetry layer**—Defined in the public proto definition file **telemetry.proto**. This layer provides data collection-related services.
  - **Content layer**—Carries GPB or JSON encoded service data.

  The RPC and telemetry layers are identical to the gRPC layer in the gRPC protocol stack.
- **Two-layer data model architecture**—Contains the RPC and content layers. Two-layer data models support only JSON for service data encoding.

gNMI

gRPC Network Management Interface (gNMI) is a gRPC-based protocol for network device management. It defines a series of RPC methods to obtain or configure the states of devices.

The device supports gNMI and non-gNMI subscriptions for pushing data to a collector.

Protocols

RFC 7540, *Hypertext Transfer Protocol version 2 (HTTP/2)*

FIPS compliance

The device supports the FIPS mode that complies with NIST FIPS 140-2 requirements. Support for features, commands, and parameters might differ in FIPS mode and non-FIPS mode. For more information about FIPS mode, see *Security Configuration Guide*.

gRPC is not supported in FIPS mode.

Restrictions and guidelines: gRPC configuration

Disabling the gRPC service deletes all gRPC-related settings.

If you do not specify the source IP address for packets sent to collectors on the device, the **deviceIpAddr** field displays **not-config** in data pushed to the collectors in dial-out mode when the two-layer telemetry data model is used. To specify the source IP address for packets sent to collectors, use the **`source-address`** command.

gRPC configuration tasks at a glance

Configure gRPC dial-in and dial-out modes as needed:

1. Configuring the gRPC dial-in mode

   Configure gRPC dial-in when the device acts as a gRPC server to communicate with collectors that act as gRPC clients.

   a. Configuring the gRPC service
   b. Configuring a gRPC user
   c. (Optional.) Enabling gRPC logging in dial-in mode
2. Configuring the gRPC dial-out mode

   Configure gRPC dial-out when the device acts as a gRPC client to communicate with collectors that act as gRPC servers.

   a. Enabling the gRPC service
   b. Configuring sensors
   c. Configuring collectors
   d. Configuring a subscription
   e. (Optional.) Enabling gRPC logging in dial-out mode
3. (Optional.) Specifying the PKI domain for secure communication with collectors
4. (Optional.) Setting the maximum CPU usage for gRPC
5. (Optional.) Enabling event-triggered data to be pushed in a prioritized manner

Configuring the gRPC dial-in mode

Configuring the gRPC service

1. Enter system view.

system-view

2. Enable the gRPC service.

   grpc enable

   By default, the gRPC service is disabled.

3. (Optional.) Specify an ACL to identify the gRPC clients allowed to establish gRPC sessions.

   grpc acl [**ipv6**] { *acl-number* | **name** *acl-name* }

   By default, no ACL is used for identifying the gRPC clients allowed to establish gRPC sessions.

   Associating an ACL with gRPC causes the gRPC service to restart, disconnecting the ongoing sessions. The clients must re-establish gRPC sessions.

   You can use this command to specify only one IPv4 ACL and one IPv6 ACL. If you execute this command multiple times, the most recent specified IPv4 ACL and IPv6 ACL take effect.

4. (Optional.) Set the gRPC service port number.

   grpc port *port-number*

   By default, the gRPC service port number is 50051.

5. (Optional.) Set the gRPC session idle timeout timer.

   grpc idle-timeout *minutes*

   By default, the gRPC session idle timeout timer is 5 minutes.

6. (Optional.) Specify a data compression method for the gRPC server.

   grpc compression gzip

   By default, the gRPC server does not compress the transmitted data.

   Changing the data compression method for the gRPC server reboots the gRPC service and disconnects all sessions established between the gRPC server and its gRPC clients. The gRPC clients must reinitiate the sessions.

7. (Optional.) Disable the gRPC dial-in mode.

   undo grpc server enable

   By default, the gRPC dial-in mode is enabled.

Configuring a gRPC user

About this task

For gRPC clients to establish gRPC sessions with the device, you must configure local users for the gRPC clients.

Procedure

1. Enter system view.

   system-view

2. Add a local user with the device management right.

   local-user *user-name* [**class manage**]

3. Configure a password for the user.

   password [{ **hash** | **simple** } *password*]

   By default, no password is configured for a local user. A non-password-protected user can pass authentication after providing the correct username and passing attribute checks.

4. Assign user role network-admin to the user.

   authorization-attribute user-role *user-role*

   By default, a local user is assigned the network-operator role.

5. Authorize the user to use the HTTPS service.

service-type https

By default, no service types are authorized to a local user.

For more information about the **local-user**, **password**, **authorization-attribute**, and **service-type** commands, see AAA configuration in *Security Command Reference*.

Enabling gRPC logging in dial-in mode

About this task

To identify issues with gRPC in dial-in mode, enable gRPC operations logging in dial-in mode.

This feature generates gRPC operation logs in dial-in mode and sends them to the information center.

With the information center, you can configure log destinations and output rules. For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

gRPC operation logging might degrade device performance if gRPC operations are frequent. As a best practice, use gRPC operation logging only when necessary and log only gRPC operations of interest if gRPC operation logging is enabled.

Procedure

1. Enter system view.

   system-view
2. Enable gRPC logging in dial-in mode. Choose the options to configure as needed:
   - Enable gRPC logging for RPC operations in dial-in mode.

     grpc log dial-in rpc { **all** | { **cli** | **get** }* }

     By default, gRPC logging is disabled for RPC operations in dial-in mode.
   - Enable gRPC logging for gNMI operations in dial-in mode.

     grpc log dial-in gnmi { **all** | { **capabilities** | **get** | **set** | **subscribe** }* }

     By default, gRPC logging is enabled for gNMI Set operations and disabled for other gNMI operations in dial-in mode.

Configuring the gRPC dial-out mode

Enabling the gRPC service

1. Enter system view.

   system-view
2. Enable the gRPC service.

   grpc enable

   By default, the gRPC service is disabled.
3. (Optional.) Specify the architecture of telemetry data models.

   grpc data-model { **2-layer** | **3-layer** }

   By default, the device uses two-layer telemetry data models to push data.

   With the two-layer telemetry data model architecture, the device can only encode pushed data in JSON format.

Configuring sensors

About this task

A sensor is a data path from which the device collects data and then pushes the data to a collector. You can create a sensor group and then specify sensor paths in the group. Sensor groups include non-gNMI sensor groups and gNMI sensor groups. If the collector uses gNMI, you must create gNMI sensor groups.

Sensor paths have the following types:

- **Event-triggered**—The device pushes data from the sensor path to the collector when certain events occur. For information about event-triggered sensor paths, see *NETCONF XML API Event Reference* for the module.
- **Periodic**—The device pushes data from the sensor path to the collector at intervals. For information about periodic sensor paths, see the NETCONF XML API references for the module except for *NETCONF XML API Configuration Reference* and *NETCONF XML API Data Reference*.
- **Condition-triggered**—The device checks the data of the sensor path periodically and pushes the data from the sensor path to the collector when certain conditions are met. For information about condition-triggered sensor paths, the sensor path check interval, and data push conditions, contact H3C Support.

  Condition-triggered sensor paths are supported only by gNMI sensor groups.

Restrictions and guidelines

As a best practice for the device to push data from sensor paths, specify sensor paths of the same type in a sensor group.

Configuring a non-gNMI sensor group

1. Enter system view.

   system-view
2. Enter telemetry view.

   telemetry
3. Create a non-gNMI sensor group and enter sensor group view.

   sensor-group *group-name*
4. Specify a sensor path.

   sensor path *path* [**selection-nodes** *node-list*]

   To specify multiple sensor paths, execute this command multiple times. If you execute this command multiple times with the same sensor path specified, the most recent configuration takes effect.

Configuring a gNMI sensor group

1. Enter system view.

   system-view
2. Enter telemetry view.

   telemetry
3. Create a gNMI sensor group and enter sensor group view.

   sensor-group *group-name* **gnmi**

   To enter the view of an existing gNMI sensor group, you do not need to specify the **gnmi** keyword.
4. Specify a sensor path.

   sensor path *path* [**selection-nodes** *node-list*]

To specify multiple sensor paths, execute this command multiple times.

Configuring collectors

About this task

Collectors receive sampled data from network devices. For the device to communicate with collectors, you must create a destination group and add collectors to the destination group.

You can add collectors to a destination group by their IP addresses or domain names. When you specify collectors by their domain names, use the following restrictions and guidelines:

- You must configure DNS to make sure the device can translate the domain names of the collectors to IP addresses. For more information about DNS, see *Layer 3—IP Services Configuration Guide*.
- To view domain name and IP address mappings, use the **display dns host** command. If a domain name maps to multiple IP addresses, the device will push data to the first reachable IP address.

To encrypt the gRPC connections between the device and the target collector, use one of the following methods:

- Enable TLS encryption globally for all gRPC connections by executing the **grpc pki domain** command in system view. The device uses the TLS certificates in the specified PKI domain for encryption.
- Enable collector-specific TLS encryption by specifying the **tls** keyword when you add a collector to a destination group. The device uses one root TLS certificate that came with it to encrypt the gRPC connection to the specified collector.

For a collector, the setting in system view has higher priority than the collector-specific setting.

Restrictions and guidelines

As a best practice, configure a maximum of five destination groups. If you configure too many destination groups, system performance might degrade.

Prerequisites

To configure a collector that resides in a VPN instance, configure the VPN instance first. For more information about VPN instance configuration, see MCE configuration in *MCE Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enter telemetry view.

   telemetry

3. Create a destination group and enter destination group view.

   destination-group *group-name*

4. Add a collector to the destination group by its IP address.

   IPv4:

   ipv4-address *ipv4-address* [**port** *port-number*] [**vpn-instance** *vpn-instance-name*] [**tls**]

   IPv6:

   ipv6-address *ipv6-address* [**port** *port-number*] [**vpn-instance** *vpn-instance-name*] [**tls**]

   The IPv6 address cannot be a link-local address. For more information about link-local addresses, see IPv6 basics configuration in *Layer 3—IP Services Configuration Guide*.

To add multiple collectors, repeat this command. A collector is uniquely identified by a combination of the IP address, VPN instance name, and port number elements. One collector must have a different element than the other collectors in the destination group.

5. Add a collector to the destination group by its domain name.

   IPv4:

   domain-name *domain-name* [**port** *port-number*] [**vpn-instance** *vpn-instance-name*] [**tls**]

   IPv6:

   ipv6 domain-name *domain-name* [**port** *port-number*] [**vpn-instance** *vpn-instance-name*] [**tls**]

   To add multiple collectors, repeat this command. A collector is uniquely identified by a combination of the domain name, VPN instance name, and port number elements. One collector must have a different element than the other collectors in the destination group.

Configuring a subscription

About this task

A subscription binds sensor groups to destination groups. Then, the device pushes data from the specified sensors to the collectors in the destination groups.

Subscriptions include non-gNMI subscriptions and gNMI subscriptions. To push data to a collector that uses gNMI, you must configure a gNMI subscription.

Restrictions and guidelines

For a non-gNMI subscription, specify non-gNMI sensor groups.

For a gNMI subscription, specify gNMI sensor groups.

You cannot use a destination group in both a gRPC subscription and a gNMI subscription.

For the device to correctly push data from sensor paths in a sensor group, make sure the subscription settings for the sensor paths are compliant with Table 2.

Table 2 Compatibility of sensor path types and subscription settings

Sensor path type	Push mode	Data push interval	Data push suppression interval
Periodical	Default	Configured	Not configured
Event-triggered	Default	Not configured	Not configured
Condition-triggered	Condition-triggered	Not configured	Configured

Configuring a non-gNMI subscription

1. Enter system view.

   system-view

2. Enter telemetry view.

   telemetry

3. Create a non-gNMI subscription and enter subscription view.

   subscription *subscription-name*

4. (Optional.) Specify a transport protocol and a data compression method for the device to send packets to collectors.

   protocol grpc compression gzip

By default, the device uses gRPC (TCP) as the transport protocol to send packets to collectors and does not compress the packet data.

5. (Optional.) Set the DSCP value for packets sent to collectors.

   dscp *dscp-value*

   By default, the DSCP value for packets sent to collectors is 0.

   A greater DSCP value represents a higher priority.

6. (Optional.) Specify the source IP address for packets sent to collectors.

   source-address { *ipv4-address* | **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }

   By default, the device uses the primary IPv4 address of the output interface for the route to the collectors as the source address.

   Changing the source IP address for packets sent to collectors causes the device to re-establish the connection to the gRPC server.

7. (Optional.) Specify the encoding format for pushed data.

   encoding { **gpb** | **json** }

   By default, the device uses the JSON format to encode pushed data.

   The device can use the GPB format to encode pushed data only when it uses three-layer telemetry data models to push data.

8. (Optional.) Enable per-row time-stamping for JSON-encoded subscription data.

   json row-timestamp enable

   By default, the device time-stamps JSON-encoded subscription data on a per-message basis.

   This command is not supported when the device uses the three-layer telemetry data model architecture.

9. Specify a sensor group.

   sensor-group *group-name* [**sample-interval** [**msec**] *interval* | **suppress-time** *suppress-time*]

   By default, no sensor groups are specified.

10. Specify a destination group.

    destination-group *group-name*

    By default, no destination groups are specified.

Configuring a gNMI subscription

1. Enter system view.

   system-view

2. Enter telemetry view.

   telemetry

3. *Create a gNMI subscription* and enter subscription view.

   subscription *subscription-name* **gnmi**

   To enter the view of an existing gNMI *subscription*, you do not need to specify the **gnmi** keyword.

4. (Optional.) Set the data push mode.

   push-mode condition-triggered

   Perform this step only if a sensor group specified for a subscription contains condition-triggered sensor paths.

   By default, the data push mode for a sensor path in a subscription is periodical or event-triggered.

5. (Optional.) Set the DSCP value for packets sent to collectors.

dscp *dscp-value*

By default, the DSCP value for packets sent to collectors is 0.

A greater DSCP value represents a higher priority.

6. (Optional.) Specify the source IP address for packets sent to collectors.

   source-address { *ipv4-address* | **interface** *interface-type interface-number* | **ipv6** *ipv6-address* }

   By default, the device uses the primary IPv4 address of the output interface for the route to the collectors as the source address.

   Changing the source IP address for packets sent to collectors causes the device to re-establish the connection to the gRPC server.

7. Specify a sensor group.

   sensor-group *group-name* [**sample-interval** [**msec**] *interval* | **suppress-time** *suppress-time*]

   By default, no sensor groups are specified.

8. Specify a destination group.

   destination-group *group-name*

   By default, no destination groups are specified.

Enabling gRPC logging in dial-out mode

About this task

To identify issues with gRPC in dial-out mode, enable gRPC data collection logging in dial-out mode.

This feature generates gRPC data collection logs in dial-out mode and sends them to the information center.

With the information center, you can configure log destinations and output rules. For more information about the information center, see *Network Management and Monitoring Configuration Guide*.

Restrictions and guidelines

gRPC logging in dial-out mode is unavailable for gNMI subscriptions.

Procedure

1. Enter system view.

   system-view

2. Enable gRPC logging in dial-out mode.

   grpc log dial-out { **all** | { **event** | **sample** }* }

   By default, gRPC logging is disabled in dial-out mode.

Specifying the PKI domain for secure communication with collectors

About this task

By default, the gRPC connection between the device and a collector does not provide data encryption service or require authentication. After you specify a PKI domain, the device and the collector will use TLS for data encryption and bidirectional certificate-based authentication to improve communication security.

Restrictions and guidelines

The specified PKI domain must already exist and has correct certificate and key settings. For more information about PKI domains, see PKI configuration in *Security Configuration Guide*.

After you specify the PKI domain, the gRPC function will reboot, closing the connections to collectors. In dial-in mode, the collectors must re-initiate the connections. In dial-out mode, the device will automatically re-initiate the connections.

Procedure

1. Enter system view.
   system-view
2. Specify the PKI domain for secure communication with collectors.
   grpc pki domain *domain-name*
   By default, no PKI domain is specified for secure communication with collectors.

Setting the maximum CPU usage for gRPC

About this task

gRPC data sampling is CPU intensive. To prevent gRPC from negatively affecting other services, limit its maximum CPU usage. If the specified maximum CPU usage is exceeded during data sampling, gRPC suspends data sampling until after its CPU usage drops below the specified limit.

Procedure

1. Enter system view.
   system-view
2. Specify the PKI domain for secure communication with collectors.
   grpc cpu-usage max-percent *percentage*
   By default, no CPU usage limit is set for gRPC.

Enabling event-triggered data to be pushed in a prioritized manner

About this task

The gRPC feature might consume a large number of CPU resources during sampling and data pushing. To avoid affecting other services, configure this feature to control the amount of data pushed to the collector.

After you enable this feature, if the CPU usage reaches the threshold and triggers an alarm, only some event-triggered data (such as CPU usage and memory threshold) is pushed to the collector. Reporting of other event-triggered data is delayed or stopped to reduce CPU usage.

Procedure

1. Enter system view.
   system-view
2. Enable event-triggered data to be pushed in a prioritized manner.
   grpc event-trigger priority enable
   By default, event-triggered data is disabled from being pushed in a prioritized manner.

Display and maintenance commands for gRPC

Execute **display** commands in any view.

Task	Command
Display gRPC information.	**display grpc** [**verbose**]
Display periodic sensor paths that have the minimum sampling interval.	**display telemetry sensor-path**

gRPC configuration examples

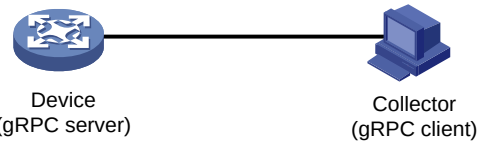
These configuration examples describe only CLI configuration tasks on the device. The collectors need to run an extra application. For information about collector-side application development, see "Developing the collector-side application (dial-in mode)."

Example: Configuring the gRPC dial-in mode

Network configuration

As shown in Figure 3, configure the gRPC dial-in mode on the device so the device acts as the gRPC server and the gRPC client can subscribe to LLDP events on the device.

Figure 3 Network diagram



Procedure

1. Assign IP addresses to interfaces on the gRPC server and client and configure routes. Make sure the server and client can reach each other.
2. Configure the device as the gRPC server:

   # Enable the gRPC service.

   ```
   <Device> system-view
   [Device] grpc enable
   ```

   # Create a local user named **test**. Set the password of the user, and assign user role network-admin and the HTTPS service to the user.

   ```
   [Device] local-user test
   [Device-luser-manage-test] password simple 123456TESTplat&!
   [Device-luser-manage-test] authorization-attribute user-role network-admin
   [Device-luser-manage-test] service-type https
   [Device-luser-manage-test] quit
   ```

3. Configure the gRPC client.
   a. Prepare a PC and install the gRPC environment on the PC. For more information, see the user guide for the gRPC environment.
   b. Obtain the H3C proto definition file and uses the protocol buffer compiler to generate code of a specific language, for example, Java, Python, C/C++, or Go.
   c. Create a client application to call the generated code.

d. Start the application to log in to the gRPC server.

Verifying the configuration

When an LLDP event occurs on the gRPC server, verify that the gRPC client receives the event.

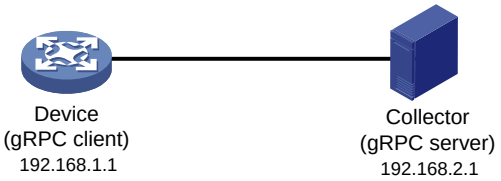
Example: Configuring event-triggered telemetry in gRPC dial-out mode

Network configuration

As shown in Figure 4, the device acts as a gRPC client and connects to the telemetry data collector (a gRPC server). The collector receives data on port 50050.

Configure the device to operate in dial-out mode to push interface event data to the collector.

Figure 4 Network diagram



Restrictions and guidelines

When you execute the **sensor-group** command in subscription view to specify a sensor group that contains event-triggered sensor paths for a subscription, do not use the **sample-interval** to set a data push interval.

NOTE:

An event-triggered sensor path typically contains a string of **event** in its path name.

Procedure

Configure IP addresses as required so the device and the collector can reach each other. (Details not shown.)

Enable gRPC.

```
<Device> system-view
[Device] grpc enable
```

Create a sensor group named **Test**, and add sensor path **ifmgr/interfaceevent** to it.

```
[Device] telemetry
[Device-telemetry] sensor-group Test
[Device-telemetry-sensor-group-Test] sensor path ifmgr/interfaceevent
[Device-telemetry-sensor-group-Test] quit
```

Create a destination group named **collector1** and add a collector to the group. In this example, the collector receives data on port 50050 at IPv4 address 192.168.2.1.

```
[Device-telemetry] destination-group collector1
[Device-telemetry-destination-group-collector1] ipv4-address 192.168.2.1 port 50050
[Device-telemetry-destination-group-collector1] quit
```

Configure a subscription named **B** to bind sensor group **Test** with destination group **collector1**.

```
[Device-telemetry] subscription B
[Device-telemetry-subscription-B] sensor-group Test
```

```
[Device-telemetry-subscription-B] destination-group collector1
[Device-telemetry-subscription-B] quit
```

Verifying the configuration

Access the collector and verify that the collector can receive the interface event data pushed by the device. (Details not shown.)

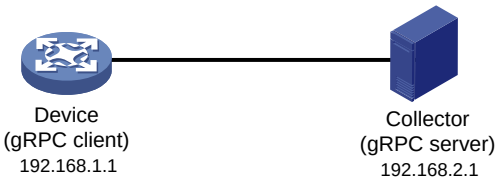
Example: Configuring the gRPC dial-out mode

Network configuration

As shown in Figure 5, the device is connected to a collector. The collector uses port 50050.

Configure gRPC dial-out mode on the device so the device pushes the device capability information of its interface module to the collector at 10-second intervals.

Figure 5 Network diagram



Procedure

Configure IP addresses as required so the device and the collector can reach each other. (Details not shown.)

Enable the gRPC service.

```
<Device> system-view
[Device] grpc enable
```

Create a sensor group named **test**, and add sensor path **ifmgr/devicecapabilities/**.

```
[Device] telemetry
[Device-telemetry] sensor-group test
[Device-telemetry-sensor-group-test] sensor path ifmgr/devicecapabilities/
[Device-telemetry-sensor-group-test] quit
```

Create a destination group named **collector1**. Specify a collector that uses IPv4 address 192.168.2.1 and port number 50050.

```
[Device-telemetry] destination-group collector1
[Device-telemetry-destination-group-collector1] ipv4-address 192.168.2.1 port 50050
[Device-telemetry-destination-group-collector1] quit
```

Configure a subscription named **A** to bind sensor group **test** with destination group **collector1**. Set the data push interval to 10 seconds.

```
[Device-telemetry] subscription A
[Device-telemetry-subscription-A] sensor-group test sample-interval 10
[Device-telemetry-subscription-A] destination-group collector1
[Device-telemetry-subscription-A] quit
```

Verifying the configuration

Verify that the collector receives the device capability information of the interface module from the device at 10-second intervals. (Details not shown.)

Protocol buffer code

Protocol buffer code format

Google Protocol Buffers provide a flexible mechanism for serializing structured data. Different from XML code and JSON code, the protocol buffer code is binary and provides higher performance.

Table 3 compares a protocol buffer code format example and the corresponding JSON code format example.

Table 3 Protocol buffer and JSON code format examples

Protocol buffer code format example	Corresponding JSON code format example

{
1:“H3C”
2:“H3C”
3:“H3C device_test”
4:“Syslog/LogBuffer”
5:"notification": {
"Syslog": {
"LogBuffer": {
"BufferSize": 512,
"BufferSizeLimit": 1024,
"DroppedLogsCount": 0,
"LogsCount": 100,
"LogsCountPerSeverity": {
"Alert": 0,
"Critical": 1,
"Debug": 0,
"Emergency": 0,
"Error": 3,
"Informational": 80,
"Notice": 15,
"Warning": 1
},
"OverwrittenLogsCount": 0,
"State": "enable"
}
},
"Timestamp": "1527206160022"
}
}
	{
	"producerName": "H3C",
	"deviceName": "H3C",
	"deviceModel": "H3C device_test",
	"sensorPath": "Syslog/LogBuffer",
	"jsonData": {
	"notification": {
	"Syslog": {
	"LogBuffer": {
	"BufferSize": 512,
	"BufferSizeLimit": 1024,
	"DroppedLogsCount": 0,
	"LogsCount": 100,
	"LogsCountPerSeverity": {
	"Alert": 0,
	"Critical": 1,
	"Debug": 0,
	"Emergency": 0,
	"Error": 3,
	"Informational": 80,
	"Notice": 15,
	"Warning": 1
	},
	"OverwrittenLogsCount": 0,
	"State": "enable"
	}
	},
	"Timestamp": "1527206160022"
	}
	}
	}

Proto definition files

You can define data structures in a proto definition file. Then, you can compile the file with utility protoc to generate code in a programing language such as Java and C++. Using the generated code, you can develop an application for a collector to communicate with the device.

H3C provides proto definition files for both dial-in mode and dial-out mode.

Proto definition files in dial-in mode

Public proto definition files

Dial-in mode supports the following public proto definition files:

- **grpc_service.proto**—Defines the public RPC methods in dial-in mode.
- **gnmi.proto**—Defines the public RPC methods for gNMI operations.
- **gnmi_ext.proto**—Defines the extended message structures required by file **gnmi.proto**.

Files **gnmi.proto** and **gnmi_ext.proto** are from Google. For information about the download paths, see "Obtaining proto definition files."

The **grpc_service.proto** file is provided by H3C. The following are the contents of the file:

```
syntax = "proto2";
package grpc_service;
message GetJsonReply {   // Reply to the Get method
    required string result = 1;
}
message SubscribeReply { // Subscription result
    required string result = 1;
}
message ConfigReply {    // Configuration result
    required string result = 1;
}
message ReportEvent {    // Subscribed event
    required string token_id = 1;    // Login token_id
    required string stream_name = 2; // Event stream name
    required string event_name = 3;  // Event name
    required string json_text = 4;   // Subscription result, a JSON string
}
message GetReportRequest{ // Obtains the event subscription result
    required string token_id = 1; // Returns the token_id upon a successful login
}
message LoginRequest {  // Login request parameters
    required string user_name = 1; // Username
    required string password = 2; // Password
}
message LoginReply {  // Reply to a login request
    required string token_id = 1; // Returns the token_id upon a successful login
}
message LogoutRequest { // Logout parameter
    required string token_id = 1; // token_id
}
```

```
message LogoutReply { // Reply to a logout request
    required string result = 1; // Logout result
}
message SubscribeRequest { // Event stream name
    required string stream_name = 1;
}
message CliConfigArgs { // Sends a configuration command and the parameters to the device
    required int64 ReqId = 1; // Command request ID
    required string cli = 2;  // Command string
}
message CliConfigReply { // Reply to a configuration command execution request
    required int64 ResReqId = 1; // Request ID, which corresponds to that in CliConfigArgs
  required string output = 2;    // Output from the command
    required string errors = 3;  // Command execution result
}
message DisplayCmdArgs { // Sends a display command and the parameters to the device
    required int64 ReqId = 1; // Command request ID
    required string cli = 2;  // Command string
}
message DisplayCmdReply { // Reply to a display command execution request
    required int64 ResReqId =1; // Request ID, which corresponds to that in DisplayCmdArgs
    required string output = 2;  // Output from the command
    required string errors = 3;  // Command execution result
}
service GrpcService {    // gRPC methods
    rpc Login (LoginRequest) returns (LoginReply) {}      // Login method
    rpc Logout (LogoutRequest) returns (LogoutReply) {}   // Logout method
    rpc SubscribeByStreamName (SubscribeRequest) returns (SubscribeReply) {} // Event
subscription method
    rpc GetEventReport (GetReportRequest) returns (stream ReportEvent) {} // Method for
obtaining the subscribed event
    rpc CliConfig (CliConfigArgs)  returns (stream CliConfigReply) {} // Method for
executing a configuration command and returning the execution result
    rpc DisplayCmdTextOutput(DisplayCmdArgs)  returns(stream DisplayCmdReply) {} //
Method for executing a display command and returning the execution result
}
```

Proto definition files for service modules

The dial-in mode supports proto definition files for the following service modules: Device, Ifmgr, IPFW, LLDP, and Syslog.

The following are the contents of the **Device.proto** file, which defines the RPC methods for the Device module:

```
syntax = "proto2";
import "grpc_service.proto";
package  device;
message DeviceBase { // Structure for obtaining basic device information
    optional string HostName = 1;    // Device name
    optional string HostOid = 2;     // sysoid
    optional uint32 MaxChassisNum = 3;    // Maximum number of chassis
```

```
    optional uint32 MaxSlotNum = 4;  // Maximum number of slots
    optional string HostDescription = 5; // Device description
}
message DevicePhysicalEntities { // Structure for obtaining physical entity information
of the device
    message Entity {
        optional uint32 PhysicalIndex = 1;    // Entity index
        optional string VendorType = 2;       // Vendor type
        optional uint32 EntityClass = 3;      // Entity class
        optional string SoftwareRev = 4;      // Software version
        optional string SerialNumber = 5;     // Serial number
        optional string Model = 6;            // Model
    }
    repeated Entity entity = 1;
}
service DeviceService { // RPC methods
    rpc GetJsonDeviceBase(DeviceBase) returns (grpc_service.GetJsonReply) {} // Method
for obtaining basic device information
    rpc GetJsonDevicePhysicalEntities(DevicePhysicalEntities) returns
(grpc_service.GetJsonReply) {}   // Method for obtaining physical entity information of
the device
}
```

Proto definition file in dial-out mode

Public proto definition files

Dial-out mode supports the following public proto definition files:

- **grpc_dialout.proto**—Defines the public RPC methods in dial-out mode.
- **grpc_dialout_v3.proto**—Defines the public RPC methods for three-layer telemetry data models in dial-out mode.
- **telemetry.proto**—Defines data sampling parameters for three-layer telemetry data models in dial-out mode.
- **dial_out.proto**—Defines the subscription message formats for gNMI subscriptions in dial-out mode.
- **gnmi.proto**—Defines the public RPC methods in gNMI subscriptions.
- **gnmi_ext.proto**—Defines the extended message structures required by file **gnmi.proto**.

File **dial_out.proto** is from SONIC, and files **gnmi.proto** and **gnmi_ext.proto** are from Google. For information about the download paths, see "Obtaining proto definition files."

The **grpc_dialout.proto** file is provided by H3C. The following are the contents of the file:

```
syntax = "proto2";
package grpc_dialout;
message DeviceInfo{  // Pushed device information
    required string producerName = 1; // Vendor name
    required string deviceName = 2;   // Device name
    required string deviceModel = 3;  // Device model
}
message DialoutMsg{   // Format of the pushed data
    required DeviceInfo deviceMsg = 1; // Device information described by DeviceInfo
```

```
    required string sensorPath = 2; // Sensor path, which corresponds to xpath in NETCONF
    required string jsonData = 3;  // Sampled data, a JSON string
}
message DialoutResponse{  // Response from the collector. Reserved. The value is not
processed.
    required string response = 1;
}
service GRPCDialout {  // Data push method
    rpc Dialout(stream DialoutMsg) returns (DialoutResponse);
}
```

The following is the content of the **grpc_dialout_v3.proto** file:

```
syntax = "proto3";
package grpc_dialout_v3;

message DialoutV3Args{
    int64 ReqId = 1;// Request ID
    bytes data = 2;// Carried data
    string errors = 3;// Error description.
    int32 totalSize = 4;// Total size of the message when fragmented. Value 0 indicates
that the message is not fragmented.
}
service gRPCDialoutV3{
    rpc DialoutV3(stream DialoutV3Args) returns (stream DialoutV3Args) {};
}
```

The following is the content of the **telemetry.proto** file:

```
syntax = "proto3";
package telemetry;
message Telemetry {
    string producer_name = 1;// Vendor name
    string node_id_str = 2;// Device name
    string product_name = 3;// Product name
    string subscription_id_str = 15;// Subscription name
    string sensor_path = 16;// Sensor path
    uint64 collection_id = 17;// Sampling ID
    uint64 collection_start_time = 18;// Start time of the sampling period
    uint64 msg_timestamp = 19;// Timestramp when this message was generated
    uint64 collection_end_time = 20;// End time of the sampling period
    uint32 current_period = 21;// Sampling accuracy
    string except_desc = 22;// Error description
    enum Encoding {
        Encoding_JSON = 0;// GPB encoding format
        Encoding_GPB = 1;// JSON encoding format
        };
    Encoding encoding = 23;// Data encoding format
    string data_str = 24;// This field is valid only when a non-GPB encoding is used. If
GPB encoding is used, this field is null.
    TelemetryGPBTable data_gpb = 25;// This filed indicates that the data carried is
defined in TelemetryGPBTable.
}
```

```
message TelemetryGPBTable {
    repeated TelemetryRowGPB row = 1;// Array definition, which indicates that the data
contains multiple TelemetryRowGPB structures.
}
message TelemetryRowGPB {
    uint64 timestamp = 1;// Timestramp of the current sampling instance
    bytes keys = 10;// Reserved field
    bytes content = 11;// Data carried by the sampling instance
}
```

Proto definition files for service modules

If service data is in GPB format, the device must use the proto definition file for the service module to decode the data.

The dial-out mode supports proto definition files for the Device and Ifmgr service modules.

Obtaining proto definition files

To obtain proto files, contact H3C Support.

You can also download files **gnmi.proto**, **gnmi_ext.proto**, and **dial-out.proto** from the following websites respectively:

- https://github.com/openconfig/gnmi/tree/master/proto/gnmi/gnmi.proto
- https://github.com/openconfig/gnmi/tree/master/proto/gnmi_ext/gnmi_ext.proto
- https://github.com/Azure/sonic-telemetry/blob/master/proto/dial_out.proto

Example: Developing a gRPC collector-side application

Use a language (for example, C++) to develop a gRPC collector-side application on Linux to achieve the following goals:

- Collect device data by using Get, gNMI Capabilities, gNMI Get, or gNMI Subscribe operations in dial-in mode or by using dial-out mode.
- Send settings to the device by using gNMI Set or CLI operations in dial-in mode.

Prerequisites

1. Obtain proto definition files.
   - For Get operations in dial-in mode, obtain the **grpc_service.proto** file and proto definition files for service modules.
   - For gNMI operations in dial-in mode, obtain files **grpc_service.proto**, **gnmi.proto**, and **gnmi_ext.proto**.
   - For CLI operations in dial-in mode, obtain the **grpc_service.proto** file.
   - For non-gNMI subscriptions in dial-out mode, obtain the **grpc_dialout.proto** file.
   - For gNMI subscriptions in dial-out mode, obtain the **dial_out.proto**, **gnmi.proto**, and **gnmi_ext.proto** files.
2. Obtain utility protoc from https://github.com/google/protobuf/releases.
3. Obtain the protobuf plug-in for C++ (**protobuf-cpp**) from https://github.com/google/protobuf/releases.

Generating the C++ code for the proto definition files

Dial-in mode

Copy the required proto definition files to the current directory, for example, **grpc_service.proto** and **BufferMonitor.proto**.

```
$protoc --plugin=./grpc_cpp_plugin  --grpc_out=. --cpp_out=. *.proto
```

Dial-out mode

Copy proto definition file **grpc_dialout.proto** to the current directory.

```
$ protoc --plugin=./grpc_cpp_plugin  --grpc_out=.  --cpp_out=. *.proto
```

Developing the collector-side application (dial-in mode)

In dial-in mode, the application needs to provide the code to be run on the gRPC client.

The C++ code generated from the proto definition files already encapsulates the service classes. For the gRPC client to initiate RPC requests, you only need to call the RPC method in the application.

The application performs the following operations:

- Log in to obtain the token_id.
- Prepare parameters for the RPC method, use the service classes generated from the proto definition files to call the RPC method, and resolve the returned result.
- Log out.

Using Get operations

Service classes GrpcService and BufferMonitorService are used in this example.

To develop the collector-side application in dial-in mode:

1. Create a GrpcServiceTest class.

   # In the class, use the GrpcService::Stub class generated from grpc_service.proto. Implement login and logout with the Login and Logout methods generated from grpc_service.proto.

```
class GrpcServiceTest
{
public:
    /* Constructor functions */
   GrpcServiceTest(std::shared_ptr<Channel> channel):
GrpcServiceStub(GrpcService::NewStub(channel))  {}

    /* Member functions */
    int Login(const std::string& username, const std::string& password);
   void Logout();
   void listen();
Status listen(const std::string& command);

    /* Member variable */
    std::string token;

private:
    std::unique_ptr<GrpcService::Stub> GrpcServiceStub;  // Use the
GrpcService::Stub class generated from grpc_service.proto.
};
```

2. Customize the Login method.

   # Call the Login method of the GrpcService::Stub class to allow a user who provides the correct the username and password to log in.

```
int GrpcServiceTest::Login(const std::string& username, const std::string& password)
{
    LoginRequest request;   // Username and password.
    request.set_user_name(username);
    request.set_password(password);

LoginReply reply;
ClientContext context;

   // Call the Login method.
   Status status = GrpcServiceStub->Login(&context, request, &reply);
   if (status.ok())
   {
       std::cout << "login ok!" << std::endl;
       std::cout <<"token id is :" << reply.token_id() << std::endl;
       token = reply.token_id();  // The login succeeds. The token is obtained.
       return 0;
   }
   else{
       std::cout << status.error_code() << ": " << status.error_message()
                 << ". Login failed!" << std::endl;
       return -1;
   }
}
```

3. Initiate an RPC request to the device. In this example, the application subscribes to interface packet drop events.

```
rpc SubscribePortQueDropEvent(PortQueDropEvent) returns
(grpc_service.SubscribeReply) {}
```

4. Create the BufMon_GrpcClient class to encapsulate the RPC method.

   # Use the BufferMonitorService::Stub class generated from BufferMonitor.proto to call the RPC method.

```
class BufMon_GrpcClient
{
public:
    BufMon_GrpcClient(std::shared_ptr<Channel> channel):
mStub(BufferMonitorService::NewStub(channel))
    {}

    std::string BufMon_Sub_AllEvent(std::string token);
    std::string BufMon_Sub_BoardEvent(std::string token);
    std::string BufMon_Sub_PortOverrunEvent(std::string token);
    std::string BufMon_Sub_PortDropEvent(std::string token);

    /* Get entries */
    std::string BufMon_Sub_GetStatistics(std::string token);
    std::string BufMon_Sub_GetGlobalCfg(std::string token);
```

```
    std::string BufMon_Sub_GetBoardCfg(std::string token);
    std::string BufMon_Sub_GetNodeQueCfg(std::string token);
    std::string BufMon_Sub_GetPortQueCfg(std::string token);

private:
    std::unique_ptr<BufferMonitorService::Stub> mStub; // Use the class generated from BufferMonitor.proto.
};
```

5. Use std::string BufMon_Sub_PortDropEvent(std::string token) to implement interface packet drop event subscription.

```
std::string BufMon_GrpcClient::BufMon_Sub_PortDropEvent(std::string token)
{
    std::cout << "-------BufMon_Sub_PortDropEvent-------- " << std::endl;

    PortQueDropEvent stNodeEvent;
    PortQueDropEvent_PortQueDrop* pstParam = stNodeEvent.add_portquedrop();

    UINT uiIfIndex = 0;
    UINT uiQueIdx = 0;
    UINT uiAlarmType = 0;

    std::cout<<"Please input interface queue info : ifIndex queIdx alarmtype " << std::endl;
    cout<<"alarmtype : 1 for ingress; 2 for egress; 3 for port headroom"<<endl;

    std::cin>>uiIfIndex>>uiQueIdx>>uiAlarmType; // Set the subscription parameters and interface index.
    pstParam->set_ifindex(uiIfIndex);
    pstParam->set_queindex(uiQueIdx);
    pstParam->set_alarmtype(uiAlarmType);

    ClientContext context;

    /* Token needs to be added to context */ // Set the token_id to be returned after a successful login
    std::string key = "token_id";
    std::string value = token;
    context.AddMetadata(key, value);

    SubscribeReply reply;
    Status status = mStub->SubscribePortQueDropEvent(&context,stNodeEvent,&reply); // Call the RPC method.

    return reply.result();
}
```

6. Use a loop to listen for event reports.

Implement this method in the GrpcServiceTest class.

```
void GrpcServiceTest::listen()
{
```

```
    GetReportRequest reportRequest;
    ClientContext context;
    ReportEvent reportedEvent;

   /* Add the token to the request */
   reportRequest.set_token_id(token);

    std::unique_ptr< ClientReader< ReportEvent>>
reader(GrpcServiceStub->GetEventReport(&context, reportRequest)); // Use
GetEventReport (which is generated from grpc_service.proto) to obtain event
information.

    std::string streamName;
    std::string eventName;
    std::string jsonText;
   std::string token;

    JsonFormatTool jsonTool;

   std::cout << "Listen to server for Event" << std::endl;
    while(reader->Read(&reportedEvent) ) // Read the received event report.
    {
        streamName = reportedEvent.stream_name();
        eventName = reportedEvent.event_name();
        jsonText = reportedEvent.json_text();
        token = reportedEvent.token_id();

          std::cout << "/***********EVENT COME**************/" << std::endl;
        std::cout << "TOKEN: " << token << std::endl;
        std::cout << "StreamName: "<< streamName << std::endl;
        std::cout << "EventName: " << eventName << std::endl;
        std::cout << "JsonText without format: " << std::endl << jsonText << std::endl;
        std::cout << std::endl;
        std::cout << "JsonText Formated: " << jsonTool.formatJson(jsonText) <<
std::endl;
        std::cout << std::endl;
    }

    Status status = reader->Finish();
    std::cout << "Status Message:" << status.error_message() << "ERROR code :" <<
status.error_code();
}  // Login and RPC request finished.
```

7. Call the Logout method to log out.

```
void GrpcServiceTest:: Logout ()
{
    LogoutRequest request;
   request.set_token_id(token);
   LogoutReply reply;
   ClientContext context;
```

```
    Status status = mStub->Logout(&context, request, &reply);
std::cout << "Logout! :" << reply.result() << std::endl;
}
```

Using gNMI Capabilities operations

1. Create a GrpcServiceTest class.
   See step 1 in "Using Get operations."
2. Customize the Login method.
   See step 2 in "Using Get operations."
3. Initiate an RPC request to the device.
   ```
   rpc Capabilities(CapabilityRequest) returns (CapabilityResponse);
   ```
4. Create the gnmi_client class to encapsulate the RPC method.
   ```
   class gnmi_client
   {
   public:
       explicit gnmi_client(const std::string &address,const std::string &tokenId);

       bool TestCapabilities();
       bool TestGet();
       bool TestSet();
       bool TestSubscribePool();
       bool TestSubscribeOnce();
       bool TestSubscribeStream();
       bool TestSubscribeStreamWithAlias();


   private:
       void PrintCapabilityResponse(const gnmi::CapabilityResponse &response);
       void PrintGetResponse(const gnmi::GetResponse &response);
       void PrintSubscribeResponse(const gnmi::SubscribeResponse &response);
       void PrintSubscribeRequest(const gnmi::SubscribeRequest &request);
       void PrintGetRequest(const gnmi::GetRequest &request);
       void PrintSetRequest(const gnmi::SetRequest &request);

       void FillGetRequest(gnmi::GetRequest &request);
       void FillSetRequest(gnmi::SetRequest &request);
       void FillSubscribeRequestByOnce(gnmi::SubscribeRequest &request);
       void FillSubscribeRequestByPool(gnmi::SubscribeRequest &request);
       void FillSubscribeRequestByStream(gnmi::SubscribeRequest &request);
       void FillSubscribePool(gnmi::SubscribeRequest &request);
       void FillSubscribeAlias(gnmi::SubscribeRequest &request);

   private:
       std::unique_ptr<gnmi::gNMI::Stub> mStubGnmiService;
       std::string mTokenID;
   };
   ```
5. Use the TestCapabilities() method to obtain the capabilities.
   ```
   bool gnmi_client::TestCapabilities()
   ```

```
{
    CapabilityRequest request;
    CapabilityResponse response;
    ClientContext context;

  context.AddMetadata("token_id", mTokenID);

    std::cout << std::endl << "CapabilitiesRequest => " << std::endl;
    Status ret = mStubGnmiService->Capabilities(&context,request,&response);

    if( StatusCode::OK != ret.error_code() )
    {
        cout << "TestCapabilities ErrorMessage:" << ret.error_message() << endl;
        return false;
    }
    else
    {
        PrintCapabilityResponse(response);
        return true;
    }
}
```

6. Call the Logout method to log out.
   See step 7 in "Using Get operations."

Using gNMI Get operations

1. Create a GrpcServiceTest class.
   See step 1 in "Using Get operations."
2. Customize the Login method.
   See step 2 in "Using Get operations."
3. Initiate an RPC request to the device.
   ```
   rpc Get(GetRequest) returns (GetResponse);
   ```
4. Create the gnmi_client class to encapsulate the RPC method.
   See step 4 in "Using gNMI Capabilities operations."
5. Use the TestGet() method to obtain the Device/Base/HostName.
   ```
   bool gnmi_client::TestGet()
   {
     GetRequest request;
     FillGetRequest(request);
     PrintGetRequest(request);
     GetResponse response;
     ClientContext context;

     context.AddMetadata("token_id", mTokenID);

     Status ret = mStubGnmiService->Get(&context,request,&response);

     if( StatusCode::OK != ret.error_code() )
     {
   ```

```
    cout << "TestGet ErrorMessage:" << ret.error_message() << endl;
    return false;
  }
  else
  {
    PrintGetResponse(response);
    return true;
  }
}

void gnmi_client::FillGetRequest(gnmi::GetRequest &request)
{
    auto prefix = request.mutable_prefix();
    auto pathelem01 = prefix->add_elem();
    pathelem01->set_name("Device");

    auto path1 = request.add_path();
    auto pathelem02 = path1->add_elem();
    pathelem02->set_name("Base");
    auto pathelem03 = path1->add_elem();
    pathelem03->set_name("HostName");
}
```

6. Call the Logout method to log out.
   See step 7 in "Using Get operations."

Using gNMI Set operations

1. Create a GrpcServiceTest class.
   See step 1 in "Using Get operations."
2. Customize the Login method.
   See step 2 in "Using Get operations."
3. Initiate an RPC request to the device.
   This example uses the Device module.
   ```
   rpc Set(SetRequest) returns (SetResponse)
   ```
4. Create a gNMITest class to encapsulate the RPC method.
   Use the gNMI::Stub class that is automatically created by gnmi.proto to call the RPC method.
   ```
   class gNMITest
   {
   public:
   gNMITest(std::shared_ptr<Channel> channel, const std::string tokenId ):

   mStubGrpcService(GrpcService::NewStub(channel)),

   mStubgNMIService(gNMI::NewStub(channel)),
                                                mTokenID(tokenId ){}
        SetResponse TestSetResponseInformation(SetRequest &request, const std::string
   tokenId);
   /*---delete: Device/Base/HostName. Restore the default for HostName -----------*/
   SetResponse DeleteDeviceBaseHostName();      /* delete: Device Base/HostName*/
   ```

```
/* update: Device/Base/HostName, string_val("string_hostname") */
SetResponse UpdateDeviceBaseHostNameStringVal();

/* replace: Device/Base/HostName, string_val("string_hostname") */
REPL_001 SetResponse ReplaceDeviceBaseHostNameStringVal();

private:
std::unique_ptr<GrpcService::Stub> mStubGrpcService;
std::unique_ptr<gNMI::Stub> mStubgNMIService;
std::string mTokenID;
};
```

5. Use customized methods to perform gNMI Set operations on service data, for example, on the Device module.

```
// Call the Set method to implement communication between client and server and get
the response.
SetResponse gNMITest::TestSetResponseInformation(SetRequest &request, const
std::string tokenId)
{
SetResponse reply;
ClientContext context;
context.AddMetadata("token_id", tokenId);

/* Call the Set method */
Status ret = mStubgNMIService->Set(&context,request,&reply);
if( StatusCode::OK != ret.error_code())
{
std::cout<<"error: "<<ret.error_message()<<std::endl;
}
return reply;
}
// Delete operation
SetResponse gNMITest:: DeleteDeviceBaseHostName ()       /* prefix == Device/Base
*/
{
SetRequest  request;
/* SetRequest->prefix */
Path        *path01 = request.mutable_prefix();
PathElem    *pathelem01 = path01->add_elem();
pathelem01->set_name("Device");
PathElem    *pathelem02 = path01->add_elem();
pathelem02->set_name("Base");

/* SetRequest->delete */
Path        *path02 = request.add_delete_();
PathElem    *pathelem03 = path02->add_elem();
pathelem03->set_name("HostName");
/* Gather response info. */
return TestSetResponseInformation(request, mTokenID);
```

```
}
// Update operation
SetResponse gNMITest::UpdateDeviceBaseHostNameStringVal()
{
SetRequest  request;

/* SetRequest->prefix */
Path        *path01 = request.mutable_prefix();
PathElem    *pathelem01 = path01->add_elem();
pathelem01->set_name("Device");
PathElem    *pathelem02 = path01->add_elem();
pathelem02->set_name("Base");

/* SetRequest->update */
Update      *update01 = request.add_update();
Path        *path02 = update01->mutable_path();
PathElem    *pathelem03 = path02->add_elem();
pathelem03->set_name("HostName");

TypedValue  *typevalue01 = update01->mutable_val();
typevalue01->set_string_val("string_hostname");

/* Gather response info. */
return TestSetResponseInformation(request, mTokenID);
}
// Replace operation
SetResponse gNMITest::ReplaceDeviceBaseHostNameStringVal()
{
SetRequest  request;

/* SetRequest->prefix */
Path        *path01 = request.mutable_prefix();
PathElem    *pathelem01 = path01->add_elem();
pathelem01->set_name("Device");
PathElem    *pathelem02 = path01->add_elem();
pathelem02->set_name("Base");

/* SetRequest->replace */
Update      *replace01 = request.add_replace();
Path        *path02 = replace01->mutable_path();
PathElem    *pathelem03 = path02->add_elem();
pathelem03->set_name("HostName");

TypedValue  *typevalue01 = replace01->mutable_val();
typevalue01->set_string_val("string_hostname");

/* Gather response info. */
return TestSetResponseInformation(request, mTokenID);
```

```
}
```

6. Call the Logout method to log out.
   See step 7 in "Using Get operations."

Using gNMI Subscribe operations

1. Create a GrpcServiceTest class.
   See step 1 in "Using Get operations."
2. Customize the Login method.
   See step 2 in "Using Get operations."
3. Initiate an RPC request to the device.

```
rpc Subscribe(stream SubscribeRequest) returns (stream SubscribeResponse);
```

4. Create the gnmi_client class to encapsulate the RPC method.
   See step 4 in "Using gNMI Capabilities operations."
5. Create a subscribe method.

```
void gnmi_client::FillSubscribeRequestByOnce(gnmi::SubscribeRequest &request)
{
  auto subscribeList = request.mutable_subscribe();

  auto prefix = subscribeList->mutable_prefix();
  auto pathelem01 = prefix->add_elem();
  pathelem01->set_name("LLDP");

  auto subscribe = subscribeList->add_subscription();

  auto path = subscribe->mutable_path();
  auto pathelem02 = path->add_elem();
  pathelem02->set_name("NeighborEvent");

  auto pathelem03 = path->add_elem();
  pathelem03->set_name("Neighbor");
  (*pathelem03->mutable_key())["IfName"] = "xxx";

    subscribeList->set_mode(::gnmi::SubscriptionList_Mode_ONCE);
    subscribeList->set_encoding(::gnmi::JSON);
}

void gnmi_client::FillSubscribeRequestByPool(gnmi::SubscribeRequest &request)
{
    auto subscribeList =  request.mutable_subscribe();

    auto prefix = subscribeList->mutable_prefix();
    auto pathelem01 = prefix->add_elem();
    pathelem01->set_name("Device");

    auto subscribe = subscribeList->add_subscription();

    auto path = subscribe->mutable_path();
    auto pathelem02 = path->add_elem();
```

```
    pathelem02->set_name("CPUs");
    auto pathelem03 = path->add_elem();
    pathelem03->set_name("CPU");
    auto pathelem04 = path->add_elem();
    pathelem04->set_name("CPUUsage");

    subscribeList->set_mode(::gnmi::SubscriptionList_Mode_POLL);
    subscribeList->set_encoding(::gnmi::JSON);
}

void gnmi_client::FillSubscribeRequestByStream(gnmi::SubscribeRequest &request)
{
     auto subscribeList =  request.mutable_subscribe();

     auto prefix = subscribeList->mutable_prefix();
     auto pathelem01 = prefix->add_elem();
     pathelem01->set_name("Diagnostic");

     auto subscribe = subscribeList->add_subscription();

     auto path = subscribe->mutable_path();
     auto pathelem02 = path->add_elem();
     pathelem02->set_name("CPUEvent");
     auto pathelem03 = path->add_elem();
     pathelem03->set_name("CPU");
     (*pathelem03->mutable_key())["Chassis#condition"] = "equal:1";
     subscribe->set_mode(::gnmi::ON_CHANGE);
     subscribe->set_sample_interval(1000);
     subscribe->set_suppress_redundant(false);
     subscribe->set_heartbeat_interval(1000);

     subscribeList->set_mode(::gnmi::SubscriptionList_Mode_STREAM);
     subscribeList->set_encoding(::gnmi::JSON);
}

void gnmi_client::FillSubscribeAlias(gnmi::SubscribeRequest &request)
{
    auto aliases = request.mutable_aliases();
    auto alias = aliases->add_alias();

    auto path = alias->mutable_path();
    auto pathelem01 = path->add_elem();
    pathelem01->set_name("Device");
    auto pathelem02 = path->add_elem();
    pathelem02->set_name("CPUs");
    auto pathelem03 = path->add_elem();
    pathelem03->set_name("CPU");
    auto pathelem04 = path->add_elem();
```

```
    pathelem04->set_name("CPUUsage");

    alias->set_alias("#cpu_usage");
}
```

6. Call the Logout method to log out.

   See step 7 in "Using Get operations."

Using CLI operations

1. Create a GrpcServiceTest class.

   See step 1 in "Using Get operations."

2. Customize the Login method.

   See step 2 in "Using Get operations."

3. Initiate an RPC request to the device.

   This example uses the CliConfig method in file **grpc_service.proto**.

   ```
   rpc CliConfig (CliConfigArgs)  returns (stream CliConfigReply) {}
   ```

4. Use the GrpcServiceTest class to encapsulate the RPC method in the same way you do for Get operations.

5. Use a customized method to support CliConfig operations.

```
// Make a thread to listen to the sever and get messages
Status GrpcServiceTest::listen(const std::string& command)
{
CliConfigArgs reportRequest;
ClientContext context;
CliConfigReply reportedEvent;
std::string key = "token_id";
std::string value = token;
context.AddMetadata(key, value);
/* Add token to request */
reportRequest.set_reqid(12345678);
reportRequest.set_cli(command);

std::unique_ptr< ClientReader< CliConfigReply>> reader(mStub->CliConfig(&context,
reportRequest));

std::string streamName;
std::string output;
int64 resreqid;

std::cout << "Command result" << std::endl;
while( reader->Read(&reportedEvent) )
{
streamName = reportedEvent.errors();
output = reportedEvent.output();
resreqid = reportedEvent.resreqid();
std::cout << "resreqid: "<< resreqid << std::endl;
std::cout << "errors: "<< streamName << std::endl;
std::cout << "output: \n"<< output << "\n"<< std::endl;
}
```

```
Status status = reader->Finish();
return status;
}
```

6. Use a loop in the main function to wait for commands.

   Use the GrpcServiceTest class.

```
int main(int argc, char *argv[])
{
const char *cmd;
unsigned int i = 0;
unsigned int cycle = 0;

if (4 == argc)
{
g_server_address = argv[1];
g_username = argv[2];
g_password = argv[3];
std::cout << "server_address: " << g_server_address <<std::endl;
std::cout << "username: " << g_username << " " << "password: " << g_password <<
std::endl;
auto channel =
grpc::CreateChannel(g_server_address,grpc::InsecureChannelCredentials());
 // 1. Log in
GrpcServiceTest reporter(channel);
if(0 != reporter.Login(g_username, g_password))
        {
            return 0;
        }
        while(1)
        {// 2. Read a command and execute the command
                std::cout<<"\n\nPlease Input Command:\n";
                getline(std::cin,g_command);          // Read a command
                Status status = reporter.listen(g_command);
                if (!status.ok())
               {
                   std::cout << status.error_code() << ": " << status.error_message()
<< std::endl;
                    break;
               }
        }
    }
    std::cout<<"Complete exec Command."<<std::endl;
    return 0;
}
```

7. Call the Logout method to log out.

   See step 7 in "Using Get operations."

Developing the collector-side application (dial-out mode)

In dial-out mode, the application needs to provide the gRPC server code so the collector can receive and resolve data obtained from the device.

The application performs the following operations:

- Inherit the automatically generated GRPCDialout::Service class, reload the automatically generated RPC Dialout service, and resolve the fields.
- Register the RPC service with the specified listening port.

To develop the collector-side application in dial-out mode:

1. Inherit and overload RPC service Dialout.

Create class DialoutTest and inherit GRPCDialout::Service.

```
class DialoutTest final : public GRPCDialout::Service { // Overload the automatically generated abstract class.
    Status Dialout(ServerContext* context, ServerReader< DialoutMsg>* reader, DialoutResponse* response) override; // Implement RPC method Dialout.
};
```

2. Register the DialoutTest service as a gRPC service and specify the listening port.

```
using grpc::Server;
using grpc::ServerBuilder;
std::string server_address("0.0.0.0:60057"); // Specify the address and port to listen to.
DialoutTest dialout_test; // Define the object declared in step 1.
ServerBuilder builder;
builder.AddListeningPort(server_address, grpc::InsecureServerCredentials());// Add the listening port.
builder.RegisterService(&dialout_test); // Register the service.
std::unique_ptr<Server> server(builder.BuildAndStart()); // Start the service.
server->Wait();
```

3. Implement the Dialout method and data resolution.

```
Status DialoutTest::Dialout(ServerContext* context, ServerReader< DialoutMsg>* reader, DialoutResponse* response)
{
        DialoutMsg msg;

        while( reader->Read(&msg))
        {
            const DeviceInfo &device_msg = msg.devicemsg();
            std::cout<< "Producer-Name: " << device_msg.producername() << std::endl;
            std::cout<< "Device-Name: " << device_msg.devicename() << std::endl;
            std::cout<< "Device-Model: " << device_msg.devicemodel() << std::endl;
            std::cout<<"Sensor-Path: " << msg.sensorpath()<<std::endl;
            std::cout<<"Json-Data: " << msg.jsondata()<<std::endl;
            std::cout<<std::endl;
        }
        response->set_response("test");

        return Status::OK;
}
```

4. After obtaining the DialoutMsg object (generated from the proto definition file) through the Read method, call the method to obtain the field values.

Appendix: Methods and attributes for gNMI operations

gNMI operations are intended for data tables and event tables.

- Some data tables are periodically reported and some data tables are reported only when triggered. The latter data tables are not supported in the current software version.
- All event tables are reported only when triggered.

Methods for adding columns in gNMI operations

Table 4 describes the methods for adding columns in gNMI operations.

Table 4 Methods for adding columns in gNMI operations

Request description	Sample request	Sample client-side code

Set a column of the key.
For example:
ifTypeExt=*colvalue*
	path:
	<
	elem:
	<
	name:"Interface"
	key:
	<
	name:"ifTypeExt"
	value:"*colvalue*"
	>
	>
	>
		auto pathelem01 = path->add_elem();
		
		pathelem01->set_name("Interface");
		
		(*pathelem01->mutable_key())["ifTypeExt"] = "*colvalue*";

Set an attribute for a column of the key.
For example, set the attr attribute of ifTypeExt:
attr=*attrvalue*
	path:
	<
	elem:
	<
	name:"Interface"
	key:
	<
	name:"ifTypeExt#attr"
	value:"*attrvalue*"
	>
	>
	>
		auto pathelem01 = path->add_elem();
		
		pathelem01->set_name("Interface");
		
		(*pathelem01->mutable_key())["ifTypeExt#attr"] = "*attrvalue*";

Set a column in a group of the key.
For example, set the Upper column of the Broadcast group:
Upper=*colvalue*
	path:
	<
	elem:
	<
	name:"Interface"
	key:
	<
	name:"Broadcast:Upper"
	value:"*colvalue*"
	>
	>
	>
		auto pathelem01 = path->add_elem();
		
		pathelem01->set_name("Interface");
		
		(*pathelem01->mutable_key())["Broadcast:Upper"] = "*colvalue*";

Set an attribute for a column in a group of the key.
For example, set the attr attribute of the Upper column of the Broadcast group:
attr=*attrvalue*
	path:
	<
	elem:
	<
	name:"Interface"
	key:
	<
		auto pathelem01 = path->add_elem();
		
		pathelem01->set_name("Interface");
		
		(*pathelem01->mutable_key())["Br

Request description	Sample request	Sample client-side code
	name:"Broadcast:Upper#attr" value:"*attrvalue*" > > >	oadcast:Upper#attr"] = "*attrvalue*";

Set an attribute of a column through the key:
For example, set the attr attribute of Interface: attr=*attrvalue*
	path:
	
	elem:
	
	name:"Interface"
	key:
	
	name:"#attr"
	value:"*attrvalue*"
	>
	>
	>
		auto pathelem01 = path->add_elem();
		
		pathelem01->set_name("Interface");
		
		(*pathelem01->mutable_key())["#attr"] = "*attrvalue*";

Set a column or column attribute in JSON:
For example:
Set Upper to *colvalue* and set the attr attribute of Upper to *attrvalue*.
	{
	"Upper": {
	"#text": "*colvalue*",
	"@#attr": "*attrvalue*"
	}
	}
		{
		"Upper": {
		"#text": "*colvalue*",
		"@#attr": "*attrvalue*"
		}
		}

The YANG files for different YANG models might have a top-level node with the same name. To specify the top-level node in the YANG file for a particular YANG model, add the YANG model name before the top-level node name. If you do not add the YANG model name, the top-level node in the first YANG file is specified.
For example, add oc-if before node interfaces to specify the interfaces node in the YANG file for OpenConfig.
	path:
	
	elem:
	
	name:"oc-if:interfaces"
	>
	>
		auto pathelem01 = path->add_elem();
		
		pathelem01->set_name("oc-if:interfaces");

gNMI Get operation attributes

gNMI Get operations support the following attributes:

- **regExp**—Regular expression.

  To implement a complex filtering with characters, you can add a **regExp** attribute for a specific column.

  The following example obtains the interface descriptions that contain only characters in upper case:

```
auto prefix = request.mutable_prefix();
 auto pathelem01 = prefix->add_elem();
 pathelem01->set_name("Ifmgr");

 auto path1 = request.add_path();
 auto pathelem02 = path1->add_elem();
 pathelem02->set_name("Interfaces");
 auto pathelem03 = path1->add_elem();
 pathelem03->set_name("Interface");
```

```
auto pathelem04 = path1->add_elem();
pathelem04->set_name("Description");
(*pathelem04->mutable_key())["#regExp"] = "^[A-Z]*$";
```

- **match**—Conditional match.

  Table 5 describes the conditional match operators.

  Table 5 Conditional match operators

Operation	Operator	Description
More than	match="more:*value*"	More than the specified value. The supported data types include date, digit, and character string.
Less than	match="less:*value*"	Less than the specified value. The supported data types include date, digit, and character string.
Not less than	match="notLess:*value*"	Not less than the specified *value*. The supported data types include date, digit, and character string.
Not more than	match="notMore:*value*"	Not more than the specified *value*. The supported data types include date, digit, and character string.
Equal to	match="equal:*value*"	Equal to the specified *value*. The supported data types include date, digit, character string, OID, and BOOL.
Not equal to	match="notEqual:*value*"	Not equal to the specified *value*. The supported data types include date, digit, character string, OID, and BOOL.
Include	match="include:*string*"	Includes the specified *string*. The supported data types include only character string.
Not include	match="exclude:*string*"	Excludes the specified *string*. The supported data types include only character string.
Start with	match="startWith:*string*"	Starts with the specified *string*. The supported data types include character string and OID.
End with	match="endWith:*string*"	Ends with the specified *string*. The supported data types include only character string.

  The following example obtains extension information about the entity whose CPU usage is more than 50%:

```
auto prefix = request.mutable_prefix();
 auto pathelem01 = prefix->add_elem();
 pathelem01->set_name("Device");

 auto path1 = request.add_path();
 auto pathelem02 = path1->add_elem();
 pathelem02->set_name("ExtPhysicalEntities");
 auto pathelem03 = path1->add_elem();
 pathelem03->set_name("Entity");
 auto pathelem04 = path1->add_elem();
 pathelem04->set_name("CpuUsage");
 (*pathelem04->mutable_key())["#match"] = "more:50";
```

- **valuetype**—Attribute value type.

  When obtaining interface information, the device cannot identify whether an integer value for the IfIndex or vrfindex column represents an interface name or index. To resolve the issue, you can use the **valuetype** attribute to specify the value type.

  The **valuetype** attribute has the following values:

Value	Description
name	The column is carrying a name.
index	The column is carrying an index.
auto	Default value. The device uses the value of the column as a name for interface matching. If no match is found, the device uses the value as an index for interface matching.

The following example sets the valuetype attribute to index and set the index to 1 for IfIndex:

```
auto prefix = request.mutable_prefix();
 auto pathelem01 = prefix->add_elem();
 pathelem01->set_name("VLAN");

 auto path1 = request.add_path();
 auto pathelem02 = path1->add_elem();
 pathelem02->set_name("TrunkInterfaces");
 auto pathelem03 = path1->add_elem();
 pathelem03->set_name("Interface");
(*pathelem03->mutable_key())["IfIndex"] = "1";
(*pathelem03->mutable_key())["IfIndex#valuetype"] = "index";
```

- **count**—Specifies the data entry quantity.

  The example sets the count attribute:

  ```
  auto prefix = request.mutable_prefix();
   auto pathelem01 = prefix->add_elem();
   pathelem01->set_name("Syslog");

   auto path1 = request.add_path();
   auto pathelem02 = path1->add_elem();
   pathelem02->set_name("Logs");
  (*pathelem03->mutable_key())["count"] = "5";
   auto pathelem03 = path1->add_elem();
   pathelem03->set_name("Log");
  (*pathelem03->mutable_key())["Index"] = "10";
  ```

- Value filtering

  You can specify a value for a column in a table to implement full match filtering. If you specify values for multiple columns, the system returns the data that matches all the specified values.

  The following example obtains information about interface whose IfIndex is 1.

  ```
  auto prefix = request.mutable_prefix();
   auto pathelem01 = prefix->add_elem();
   pathelem01->set_name("VLAN");

   auto path1 = request.add_path();
   auto pathelem02 = path1->add_elem();
   pathelem02->set_name("TrunkInterfaces");
   auto pathelem03 = path1->add_elem();
   pathelem03->set_name("Interface");
  (*pathelem03->mutable_key())["IfIndex"] = "1";
  ```

gNMI Set operation attributes

gNMI Set operations support the following attributes:

- **incremental**—Adds configuration data to a column without affecting the original data. This attribute applies to a list column such as the vlan permitlist column.

  This example adds VLANs 1 through 10 to an untagged VLAN list that has untagged VLANs 12 through 15.

```
Path        *path01 = request.mutable_prefix();
PathElem    *pathelem01 = path01->add_elem();
pathelem01->set_name("VLAN");
PathElem    *pathelem02 = path01->add_elem();
pathelem02->set_name("HybridInterfaces");

Update  *Update01 = request.add_update();
Path        *path02 = Update01->mutable_path();
PathElem    *pathelem03 = path02->add_elem();
pathelem03->set_name("Interface");
(*pathelem03->mutable_key())["IfName"] = "262";
(*pathelem03->mutable_key())["UntaggedVlanList#incremental"] = "true";
(*pathelem03->mutable_key())["UntaggedVlanList"] = "1-10";
```

- **valuetype**—Attribute value type.

  When configuring an interface, the device cannot identify whether an integer value for the IfIndex or vrfindex column represents an interface name or index. To resolve the issue, you can use the **valuetype** attribute to specify the value type.

  The **valuetype** attribute has the following values:

Value	Description
name	The column is carrying a name.
index	The column is carrying an index.
auto	Default value. The device uses the value of the column as a name for interface matching. If no match is found, the device uses the value as an index for interface matching.

  The following example sets the valuetype attribute to index and set the index to 1 for IfIndex:

```
    Path    *path01 = request.mutable_prefix();
    PathElem*pathelem01 = path01->add_elem();
    pathelem01->set_name("VLAN");
    PathElem*pathelem02 = path01->add_elem();
    pathelem02->set_name("TrunkInterfaces");

    /* SetRequest->delete */
    Path    *path02 = request.add_delete_();
    PathElem*pathelem04 = path02->add_elem();
    pathelem04->set_name("Interface ");
(*pathelem04->mutable_key())["IfIndex"] = "1";
(*pathelem04->mutable_key())["IfIndex#valuetype"] = "index";
```

gNMI Subscribe operation attributes

gNMI Subscribe operations support the following attributes:

- All gNMI Get operation attributes. See "gNMI Get operation attributes."

These attributes are used for subscription to periodically reported data tables.

- **condition**—Data pushing condition.

  This attribute is used for event table subscription. Values for this attribute are as follows:

  - **more**—More than.
  - **less**—Less than.
  - **notLess**—Not less than.
  - **notMore**—Not more than.
  - **equal**—Equal to.
  - **notEqual**—Not equal to.
  - **include**—Include.
  - **exclude**—Not include.
  - **startWith**—Start with.
  - **endWith**—End with.

  The following example monitors LLDP events:

```
auto prefix = subscribeList->mutable_prefix();
auto pathelem01 = prefix->add_elem();
pathelem01->set_name("LLDP");

auto subscribe = subscribeList->add_subscription();
auto path = subscribe->mutable_path();
auto pathelem02 = path->add_elem();
pathelem02->set_name("NeighborEvent");
auto pathelem03 = path->add_elem();
pathelem03->set_name("Neighbor");
(*pathelem03->mutable_key())["IfName"] = "GigabitEthernet1/0/1";
(*pathelem03->mutable_key())["IfName#condition"] = "equal";
```

- **match**—Conditional match.

  This attribute is used for data table subscription.

  The matching object can be a fixed value or another column. The latter supports only data tables that are reported only when triggered.

  The conditional match operators table in "gNMI Get operation attributes" describes the conditional match operators for matching a fixed value. Attribute values for matching another column are as follows:

  - **refMore**—More than.
  - **refLess**—Less than.
  - **refNotLess**—Not less than.
  - **refNotMore**—Not more than.
  - **refEqual**—Equal to.
  - **refNotEqual**—Not equal to.

  In the following example, a report is triggered only if the current-usage column has a greater value than the alarm-threshold column:

```
auto subscribe = subscribeList->add_subscription();
auto path = subscribe->mutable_path();
auto pathelem02 = path->add_elem();
pathelem02->set_name("queue");
auto pathelem03 = path->add_elem();
pathelem03->set_name("state");
```

```
(*pathelem03->mutable_key())["current-usage#match"] = "refMore:alarm-threshold";
```

- **select**—Reports certain columns in a data table that is reported only when triggered

  The following example reports only the current-usage column and the available-usage column:

```
auto subscribe = subscribeList->add_subscription();
auto path = subscribe->mutable_path();
auto pathelem02 = path->add_elem();
pathelem02->set_name("queue");
auto pathelem03 = path->add_elem();
pathelem03->set_name("state");
(*pathelem03->mutable_key())["current-usage#select"] = "";
(*pathelem03->mutable_key())["available-usage#select"] = "";
```

Contents

Configuring flow groups 1
About flow groups 1
Restrictions and guidelines: Flow group configuration 1
Procedure 2
Display and maintenance commands for flow group 2

Configuring flow groups

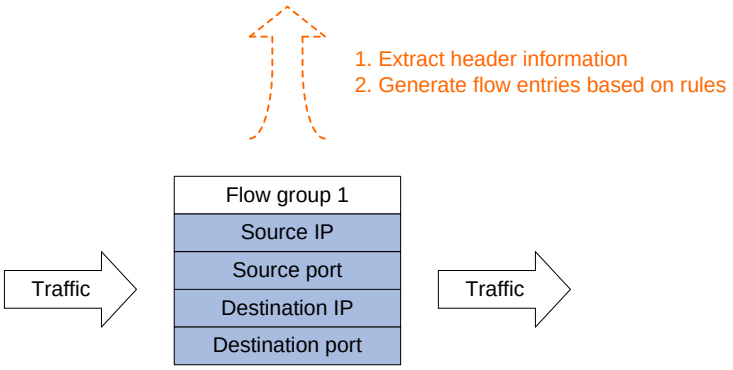
About flow groups

A flow group allows you to identify flows based on flow generation rules. The device extracts traffic characteristics (for example, 5-tuples in the packet header) and generates flow entries according to the header fields specified in a flow generation rule.

A flow group can use an ACL to limit the traffic for which flow entries are generated. A flow entry is aged out if no matching packets are received before the aging timer expires.

Figure 1 Flow entry generation

Flow table					
Src IP	Src Port	Dst IP	Dst Port	Aging	Count
10.0.0.2	2000	10.0.0.20	3000	20	5120
10.0.0.2	2002	10.0.0.20	3002	18	25
10.0.0.2	2020	10.0.1.20	3000	15	813



The flow entries generated by a flow group can be used by other features. A flow group can be in one of the following modes:

- **Common MOD mode**—Used by MOD. This mode occupies hardware resources but has a lower burden on the CPU. For more information about MOD, see *Telemetry Configuration Guide*.
- **Simple MOD mode**—Used by MOD. This mode has a higher burden on the CPU but saves hardware resources.

Restrictions and guidelines: Flow group configuration

A flow group can reference only one ACL.

Because a flow can belong to only one flow group, make sure the same flow is not assigned to more than one flow group when specifying ACLs. For information about ACLs, see *ACL and QoS Configuration Guide*.

To delete an applied flow group, first remove the application and then delete the flow group.

Only one flow group can be applied.

Do not use a flow group to monitor VXLAN packets with the outer destination port number as 4789 or 6081. The generated flow entries do not contain the correct source and destination port numbers.

The traffic statistics for a Layer 3 Ethernet interface or Layer 3 aggregate interface in the flow table include the traffic statistics for its subinterfaces.

Procedure

1. Enter system view.
   system-view
2. Create a flow group and enter its view.
   telemetry flow-group *group-id* [**name** *group-name*] [**mode simple-mod**]
3. Specify an ACL.
   if-match acl [**ipv6** | **mac**] { *acl-number* | **name** *acl-name* }
   By default, no ACL is specified.
4. Configure the header fields used for generating flow entries.
   template { **destination-ip** | **destination-port** | **protocol** | **source-ip** | **source-port** | **vxlan** { **inner-destination-ip** | **inner-destination-port** | **inner-protocol** | **inner-source-ip** | **inner-source-port** | **vxlan-id** } * } *
   By default, no header fields are used for generating flow entries.
   If an IPv4 ACL has been specified for the flow group, only the **destination-ip**, **source-ip**, **inner-destination-ip**, **inner-source-ip**, and **vxlan-id** keywords in this command can be configured.
5. Return to system view.
   quit
6. (Optional.) Set the aging time for flow entries.
   telemetry flow-group aging-time *aging-time*
   The default setting is 15 minutes.
7. (Optional.) Set the maximum number of flow entries generated.
   telemetry flow-group max-entry *max-entries*
   By default, the number of flow entries is not limited.
8. Apply the flow group.
   telemetry apply flow-group { *group-id* | **name** *group-name* }
   By default, no flow group is applied.

Display and maintenance commands for flow group

Execute **display** commands in any view.

Task	Command
Display the configuration and application status of flow groups.	**display telemetry flow-group** [*group-id*	**name** *group-name*] [**slot** *slot-number*]
Display flow entries.	**display telemetry flow-group flow-table** [[*group-id*	**name** *group-name*]	**mod**]

Task	Command
	[**destination-ip** { *dst-ipv4* | *dst-ipv6* } | **destination-port** *dst-port* | **protocol** *protocol* | **source-ip** { *src-ipv4* | *src-ipv6* } | **source-port** *src-port* | **vxlan** [**inner-destination-ip** { *dst-ipv4* | *dst-ipv6* } | **inner-destination-port** *dst-port* | **inner-protocol** *protocol* | **inner-source-ip** { *src-ipv4* | *src-ipv6* } | **inner-source-port** *src-port* | **vxlan-id** *vxlan-id*] *****] ***** { **slot** *slot-number* }

Contents

Configuring MOD 2
Overview 2
Restrictions and guidelines: MOD configuration 2
Procedure 2
Display and maintenance commands for MOD 3
MOD configuration examples 3
Example: Configuring MOD to send packets through UDP 3
Example: Configuring MOD to send packets through gRPC 5

Configuring MOD

Overview

Mirror On Drop (MOD) can detect packet drops during the forwarding process on the device. When a packet is dropped, MOD can send the packet drop reason and the characteristics of the dropped packet to the collector.

MOD uses the flow entries generated based on a flow group to detect dropped packets. For more information about flow groups, see flow group configuration in Telemetry Configuration Guide.

MOD works as follows:

1. The device generates flow entries based on a flow group.
2. Based on the packet drop reason list of MOD, the device monitors packet drops for packets matching the flow entries.
3. If a packet is dropped for a reason on the packet drop reason list, the device sends the packet drop reason and the characteristics of the dropped packet (flow entry matching the packet) to the collector.

Restrictions and guidelines: MOD configuration

MOD takes effect only on packets matching ACLs referenced by flow groups.

MOD monitors only packets dropped for reasons on the configured packet drop reason list. MOD does not monitor packets dropped for other reasons.

Procedure

1. Enter system view.

   system-view
2. Enter MOD view.

   telemetry mod
3. Configure the device ID for MOD.

   device-id *address*

   By default, no device ID is configured for MOD.
4. Enable sampling for MOD.

   sampler *sampler-name*

   By default, sampling is disabled for MOD.
5. Specify a protocol that MOD uses to send packets to the collector.

   transport-protocol { **grpc** | **udp** }

   By default, MOD uses UDP to send packets to the collector.
6. (Optional.) Configure the encapsulation information for packets sent to the collector by MOD.

   collector source-ip *source-address* **destination-ip** *destination-address* **source-port** *source-port* **destination-port** *destination-port* [**vlan** *vlan-id*]

   By default, the encapsulation information for packets sent to the collector by MOD is not configured.

   This command takes effect only if MOD uses UDP to send packets to the collector.

7. Configure the packet drop reason list monitored by MOD.

   reason-list { *reason-list* | **all** }

   By default, no packet drop reasons is configured, and the device does not monitor packet drops.

   For more information about the packet drop reason list, see the **reason-list** command in Telemetry Command Reference.

Display and maintenance commands for MOD

Execute **display** commands in any view.

Task	Command
Display MOD configuration information.	**display telemetry mod** [**slot** *slot-number*]

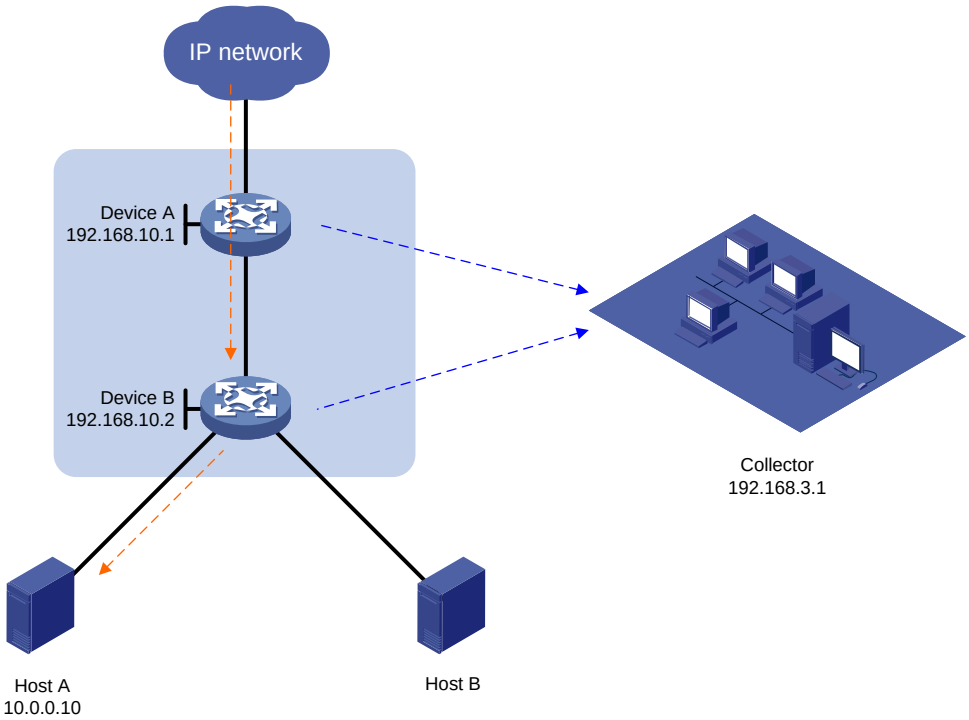
MOD configuration examples

Example: Configuring MOD to send packets through UDP

Network configuration

As shown in Figure 1, Host A and Host B access the external network through Device A and Device B. Some packets from the external network are missing on Host A. Configure MOD on the access device and aggregation device to identify whether packets are dropped during the forwarding process on the two devices.

Figure 1 Network diagram



Prerequisites

Configure IP addresses for devices, and make sure they can reach each other at Layer 3.

Procedures

1. Configure Device A:
   a. Configure a flow group:

      # Create advanced IPv4 ACL 3000, and configure a rule to match packets with destination IP address 10.0.0.10 for the ACL.

      ```
      <DeviceA> system-view
      [DeviceA] acl advanced 3000
      [DeviceA-acl-ipv4-adv-3000] rule permit ip destination 10.0.0.10 0
      [DeviceA-acl-ipv4-adv-3000] quit
      ```

      # Create flow group 1 and configure it to reference ACL 3000.

      ```
      [DeviceA] telemetry flow-group 1
      [DeviceA-telemetry-flow-group-1] if-match acl 3000
      ```

      # Configure the flow group to generate flow entries based on the destination IP address.

      ```
      [DeviceA-telemetry-flow-group-1] template destination-ip
      [DeviceA-telemetry-flow-group-1] quit
      ```

      # Set the flow entry aging time to 10 minutes.

      ```
      [DeviceA] telemetry flow-group aging-time 10
      ```

      # Apply flow group 1.

      ```
      [DeviceA] telemetry apply flow-group 1
      ```

   b. Configure MOD:

      # Configure the device ID for MOD as 192.168.10.1.

      ```
      [DeviceA] telemetry mod
      [DeviceA-telemetry-mod] device-id 192.168.10.1
      [DeviceA-telemetry-mod] quit
      ```

      # Create sampler samp, set the sampling mode to random, and set the sampling rate to 2 to the 8th power.

      ```
      [DeviceA] sampler samp mode random packet-interval n-power 8
      ```

      # Enable sampling for MOD, and reference sampler samp.

      ```
      [DeviceA] telemetry mod
      [DeviceA-telemetry-mod] sampler samp
      ```

      # Encapsulate the packets sent to the collector by MOD with the following information: source IP address 192.168.10.1, destination IP address 192.168.3.1, source port number 1000, and destination port number 2333.

      ```
      [DeviceA-telemetry-mod] collector source-ip 192.168.10.1 destination-ip
      192.168.3.1 source-port 1000 destination-port 2333
      ```

      # Configure MOD to monitor all packet drop reasons.

      ```
      [DeviceA-telemetry-mod] reason-list all
      [DeviceA-telemetry-mod] quit
      ```

2. Configure Device B.

   Configure Device B in the same way Device A is configured except the device ID for MOD and the source IP address encapsulated in the packets sent to the collector.

Verifying the configuration

1. Verify the configuration on Device A:

   # Display the ACL configuration.

```
<DeviceA> display acl 3000
Advanced IPv4 ACL 3000, 1 rule,
ACL's step is 5, start ID is 0
 rule 0 permit ip destination 10.0.0.10 0
```

Display the flow group configuration.

```
<DeviceA> display telemetry flow-group 1
Flow group 1 (Successful)
  ACL       : 3000
  Template  :
    destination-ip
Aging time: 10 minutes
Rate limit: -
Max  entry: -
```

Display the MOD configuration.

```
<DeviceA> display telemetry mod
Status          : Successful
Drop reason list:
...
Sampler            : samp
DeviceID           : 192.168.10.1
Transport protocol : UDP
Collector
  Source IP        : 192.168.10.1
  Destination IP   : 192.168.3.1
  Source port      : 1000
  Destination port : 2333
```

2. Verify the configuration on Device B.

   The configuration on Device B is the same as the configuration on Device A except the device ID for MOD and the source IP address encapsulated in the packets sent to the collector. (Details not shown.)

3. Verify the configuration on the collector.

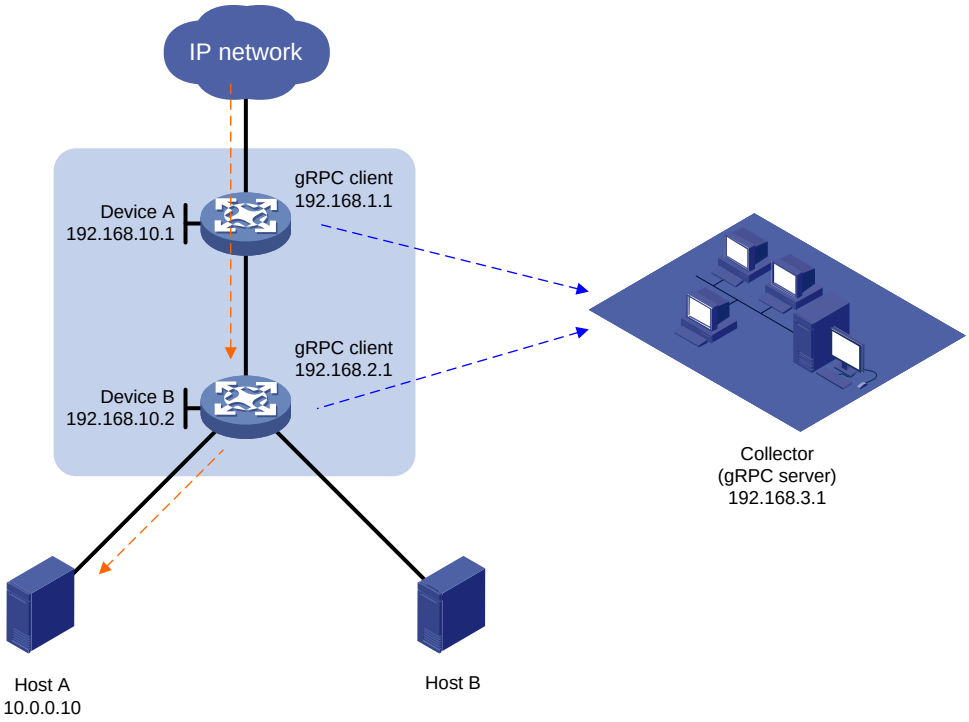
   When a packet destined for IP address 10.0.0.10 is dropped during the forwarding process on Device A and Device B, the collector can receive the packet drop reason and the characteristics of the dropped packet from the devices.

Example: Configuring MOD to send packets through gRPC

Network configuration

As shown in Figure 2, Host A and Host B access the external network through Device A and Device B. Some packets from the external network are missing on Host A. Configure MOD on the access device and aggregation device to identify whether packets are dropped during the forwarding process on the two devices.

Figure 2 Network diagram



Prerequisites

Configure IP addresses for devices, and make sure they can reach each other at Layer 3.

Procedures

1. Configure Device A:
   a. Configure a flow group:

      # Create advanced IPv4 ACL 3000, and configure a rule to match packets with destination IP address 10.0.0.10 for the ACL.

      ```
      <DeviceA> system-view
      [DeviceA] acl advanced 3000
      [DeviceA-acl-ipv4-adv-3000] rule permit ip destination 10.0.0.10 0
      [DeviceA-acl-ipv4-adv-3000] quit
      ```

      # Create flow group 1 and configure it to reference ACL 3000.

      ```
      [DeviceA] telemetry flow-group 1
      [DeviceA-telemetry-flow-group-1] if-match acl 3000
      ```

      # Configure the flow group to generate flow entries based on the destination IP address.

      ```
      [DeviceA-telemetry-flow-group-1] template destination-ip
      [DeviceA-telemetry-flow-group-1] quit
      ```

      # Set the flow entry aging time to 50 minutes.

      ```
      [DeviceA] telemetry flow-group aging-time 50
      ```

      # Apply flow group 1.

      ```
      [DeviceA] telemetry apply flow-group 1
      ```

   b. Configure MOD:

      # Configure the device ID for MOD as 192.168.10.1.

      ```
      [DeviceA] telemetry mod
      [DeviceA-telemetry-mod] device-id 192.168.10.1
      ```

```
[DeviceA-telemetry-mod] quit
```

Create sampler **samp**, set the sampling mode to random, and set the sampling rate to 2 to the 8th power.

```
[DeviceA] sampler samp mode random packet-interval n-power 8
```

Enable sampling for MOD, and reference sampler **samp**.

```
[DeviceA] telemetry mod
[DeviceA-telemetry-mod] sampler samp
```

Configure MOD to use gRPC to send packets to the collector.

```
[DeviceA-telemetry-mod] transport-protocol grpc
```

Configure MOD to monitor all packet drop reasons.

```
[DeviceA-telemetry-mod] reason-list all
[DeviceA-telemetry-mod] quit
```

c. Configure gRPC:

Enable the gRPC service.

```
[DeviceA] grpc enable
```

Create a sensor group named **test**, and add sensor path **telemetryftrace/Genevent**.

```
[DeviceA] telemetry
[DeviceA-telemetry] sensor-group test
[DeviceA-telemetry-sensor-group-test] sensor path telemetryftrace/Genevent
[DeviceA-telemetry-sensor-group-test] quit
```

Create a destination group named **collector1**. Specify a collector that uses IPv4 address 192.168.3.1 and port number 50050.

```
[DeviceA-telemetry] destination-group collector1
[DeviceA-telemetry-destination-group-collector1] ipv4-address 192.168.3.1 port 50050
[DeviceA-telemetry-destination-group-collector1] quit
```

Configure a subscription named **A** to bind sensor group **test** with destination group **collector1**.

```
[DeviceA-telemetry] subscription A
[DeviceA-telemetry-subscription-A] sensor-group test
[DeviceA-telemetry-subscription-A] destination-group collector1
[DeviceA-telemetry-subscription-A] quit
[DeviceA-telemetry] quit
```

2. Configure Device B.

Configure Device B in the same way Device A is configured except the device ID for MOD.

Verifying the configuration

1. Verify the configuration on Device A:

Display the ACL configuration.

```
<DeviceA> display acl 3000
Advanced IPv4 ACL 3000, 1 rule,
ACL's step is 5, start ID is 0
 rule 0 permit ip destination 10.0.0.10 0
```

Display the flow group configuration.

```
<DeviceA> display telemetry flow-group 1
Flow group 1 (Successful)
  ACL       : 3000
  Template  :
```

```
    destination-ip
Aging time: 50 minutes
Rate limit: -
Max  entry: -
```

Display the MOD configuration.

```
<DeviceA> display telemetry mod
Status          : Successful
Drop reason list:
...
Sampler            : samp
DeviceID           : 192.168.10.1
Transport protocol : gRPC
```

2. Verify the configuration on Device B.

The configuration on Device B is the same as the configuration on Device A except the device ID for MOD. (Details not shown.)

3. Verify the configuration on the collector.

When a packet destined for IP address 10.0.0.10 is dropped during the forwarding process on Device A and Device B, the collector (gRPC server) can receive the packet drop reason and the characteristics of the dropped packet from the devices (gRPC clients).

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

OpenFlow Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the OpenFlow fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring OpenFlow 1

About OpenFlow 1
OpenFlow network components 1
OpenFlow switch 1
OpenFlow port 1
OpenFlow instance 2
OpenFlow flow table 3
Group table 5
Meter table 5
OpenFlow channel 6
OpenFlow controller 7
Protocols and standards 8
Restrictions: Licensing requirements for OpenFlow 9
Restrictions and guidelines: OpenFlow configuration 9
OpenFlow tasks at a glance 10
Configuring OpenFlow instances 10
Creating an OpenFlow instance 10
Configuring the OpenFlow instance mode 11
Configuring inband management VLANs 11
Configuring flow tables and flow entries for an OpenFlow instance 12
Setting the controller connection mode 13
Preventing an OpenFlow instance from reporting the specified types of ports to controllers 13
Enabling support for smart group entries 13
Activating or reactivating an OpenFlow instance 14
Configuring OpenFlow instance attributes 15
Configuring controllers for an OpenFlow switch 16
Refreshing all Layer 3 flow entries in the MAC-IP flow tables for an OpenFlow instance 17
Ignoring the permit flag added by OpenFlow 17
Shutting down an interface by OpenFlow 17
Enabling SNMP notifications for OpenFlow 18
Configuring OpenFlow flow monitor parameters 18
Display and maintenance commands for OpenFlow 19
OpenFlow configuration examples 20
Example: Configuring OpenFlow in VLAN mode 20
Appendix A Application restrictions 0
Flow entry restrictions 0
Restrictions for merging the action list into the action set 1
Packet-out messages restrictions 1
Packet-in messages restrictions 2
LLDP frame matching 2
Flow table modification messages restrictions 3
Appendix B MAC-IP flow table 3
Capabilities supported by the MAC-IP flow table 3
MAC-IP flow table restrictions 4
Table-miss flow entry of MAC-IP flow tables 5
Dynamic aware 5
MAC-IP flow table cooperating with extensibility flow table 5
Capabilities supported by the VLAN tagging flow table 6
Capabilities supported by the VLAN untagging flow table 6
Multi-level table-based TTP pipeline 7
Table-miss flow entry of TTP flow tables 7
Capabilities supported by the TTP-VLAN flow table 8
Capabilities supported by the TTP-termination-MAC flow table 8
Capabilities supported by the TTP-unicast-routing flow table 9
Capabilities supported by the TTP-bridging flow table 10
Capabilities supported by the TTP-ingress-ACL flow table 10
Capabilities supported by the TTP-egress-ACL flow table 11

Capabilities supported by the TTP-L2-interface-group flow table ····· 12
Capabilities supported by the TTP-L3-unicast-group flow table ····· 12
Capabilities supported by the TTP-L3-ECMP-group table ····· 13

Configuring OpenFlow

About OpenFlow

OpenFlow is the communications interface defined between the control and forwarding layers of a Software-Defined Networking architecture. With OpenFlow, you can perform centralized data forwarding management for physical and virtual devices through controllers.

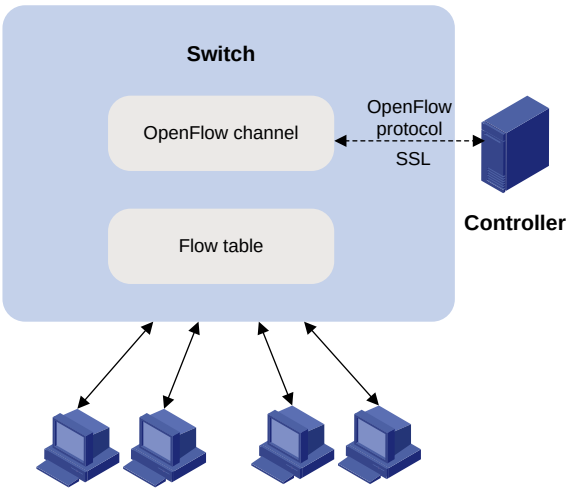
OpenFlow network components

OpenFlow separates the data forwarding and routing decision functions. It keeps the flow-based forwarding function and employs a separate controller to make routing decisions. An OpenFlow switch communicates with the controller through an OpenFlow channel. An OpenFlow channel can be encrypted by using TLS or run directly over TCP. An OpenFlow switch exchanges control messages with the controller through an OpenFlow channel to perform the following operations:

- Receive flow table entries or data from the controller.
- Report information to the controller.

Unless otherwise stated, a switch refers to an OpenFlow switch throughout this document.

Figure 1 OpenFlow network diagram



OpenFlow switch

OpenFlow switches include the following types:

- **OpenFlow-only**—Supports only OpenFlow operation.
- **OpenFlow-hybrid**—Supports both OpenFlow operation and traditional Ethernet switching operation.

OpenFlow port

OpenFlow supports the following types of ports:

- **Physical port**—Corresponds to a hardware interface, such as an Ethernet interface. A physical port can be either an ingress port or an output port.

- **Logical port**—Does not correspond to a hardware interface and might be defined by non-OpenFlow methods. For example, aggregate interfaces are logical ports. An aggregate interface can only be an ingress port.
- **Reserved port**—Defined by OpenFlow to specify forwarding actions. Reserved ports include the following types:
  - **Controller**—OpenFlow controller.
  - **Local**—Local CPU.
  - **Normal**—Normal forwarding process.

OpenFlow instance

Unless otherwise stated, an OpenFlow switch refers to an OpenFlow instance throughout this document.

You can configure one or more OpenFlow instances on the same device. A controller considers each OpenFlow instance as a separate OpenFlow switch and deploys forwarding instructions to it.

OpenFlow instance mode

An OpenFlow instance operates in one of the following modes:

- **Global mode**—When the global mode is enabled for an OpenFlow instance, the flow entries take effect on packets within the network. All service interfaces on the device belong to the OpenFlow instance.
- **VLAN mode**—When the VLAN mode is enabled for an OpenFlow instance, the flow entries take effect only on packets within VLANs associated with the OpenFlow instance. You must associate the OpenFlow instance with VLANs. All interfaces in the associated VLANs belong to the OpenFlow instance.

Activation and reactivation

The configurations for an OpenFlow instance take effect only after the OpenFlow instance is activated.

The controller can deploy flow entries to an OpenFlow instance only after the OpenFlow instance reports the following device information to the controller:

- Capabilities supported by OpenFlow.
- Information about ports that belong to the OpenFlow instance.

An activated OpenFlow instance must be reactivated when any of the OpenFlow instance configurations are changed.

After reactivation, the OpenFlow instance is disconnected from all controllers and then reconnected to them.

OpenFlow instance port

An OpenFlow switch sends information about the following ports to the controller:

- Physical ports.
- Logical ports.
- Reserved ports of the **Local** type.

In loosen mode, a port belongs to the OpenFlow instance when VLANs associated with the OpenFlow instance overlap with the port's allowed VLANs. Otherwise, a port belongs to an OpenFlow instance only when VLANs associated with the OpenFlow instance are within the port's allowed VLAN list.

OpenFlow flow table

An OpenFlow switch matches packets with one or more flow tables. A flow table contains flow entries, and packets are matched based on the matching precedence of flow entries.

Flow table types

OpenFlow flow tables include the following types:

- **MAC-IP**—Combines the MAC address table and FIB table.

  A MAC-IP flow table provides the following match fields:

  - Destination MAC address.
  - VLAN.
  - Destination IP address.

  A MAC-IP flow table provides the following actions:

  - Modifying the destination MAC address.
  - Modifying the source MAC address.
  - Modifying the VLAN.
  - Modifying the tunnel ID.
  - Specifying the output port.
- **Extensibility**—Uses ACLs to match packets.

Flow entry

A flow entry contains the following fields:

- **Match fields**—Matching rules of the flow entry. These contain the ingress port, packet headers, and metadata specified by the previous table.
- **Priority**—Matching precedence of the flow entry. When a packet is matched with the flow table, only the highest priority flow entry that matches the packet is selected.
- **Counters**—Counts of the packets and bytes that match the flow entry.
- **Instructions**—Used to modify the action set or pipeline processing. Instructions include the following types:
  - **Meter**—Directs the packets to the specified meter to rate limit the packets.
  - **Apply-Actions**—Applies the specified actions in the action list immediately.
  - **Clear-Actions**—Clears all actions in the action set immediately.
  - **Write-Actions**—Modifies all actions in the action set immediately.
  - **Write-Metadata**—Modifies packets between two flow tables if multiple flow tables exist.
  - **Goto-Table**—Indicates the next flow table in the processing pipeline.

  Actions are executed in one of the following ways:

  - **Action Set**—When the instruction set of a flow entry does not contain a **Goto-Table** instruction, pipeline processing stops. Then, the actions in the action set are executed in the order specified by the instruction list. An action set contains a maximum of one action of each type.
  - **Action List**—The actions in the action list are executed immediately in the order specified by the action list. The effect of those actions is cumulative.

  Actions include the following types:

  - **(Required.) Output**—The Output action forwards a packet to the specified OpenFlow port. OpenFlow switches must support forwarding packets to physical ports, logical ports, and reserved ports.

- **(Required.) Drop**—No explicit action exists to represent drops. Packets whose action sets have no output actions are dropped. Typically, packets are dropped due to empty instruction sets, empty action sets, or the executing a Clear-Actions instruction.
- **(Required.) Group**—Process the packet through the specified group. The exact interpretation depends on group type.
- **(Optional.) Set-Queue**—The Set-Queue action sets the queue ID for a packet. When the packet is forwarded to a port by the output action, the packet is assigned to the queue attached to this port for scheduling and forwarding. The forwarding behavior is dictated by the configuration of the queue and provides basic QoS support.
- **(Optional.) Push-Tag/Pop-Tag**—Switches support the ability to push or pop tags, such as VLAN tags.
- **(Optional.) Set-Field**—The Set-Field actions are identified by their field type and modify the values of corresponding header fields in the packet. Set-Field actions are always applied to the outermost header. For example, a Set VLAN ID action always sets the ID of the outermost VLAN tag.
- **(Optional.) Change-TTL**—The Change-TTL actions modify the values of the IPv4 TTL or IPv6 Hop Limit in the packet. Change-TTL actions are always applied to the outermost header. The Change-TTL actions include the following the actions: Set TTL, Decrement TTL, and Copy TTL (outwards or inwards).

- **Timeouts**—Maximum amount of idle time or hard time for the flow entry.
  - **idle time**—The flow entry is removed if it has matched no packets within the idle time.
  - **hard time**—The flow entry is removed when the hard time timeout is exceeded, regardless of whether or not it has matched packets.
- **Cookie**—Custom data specified by the controller. The data is not used for processing packets, and might be used by the controller for matching flow entries.

Figure 2 Flow entry components



OpenFlow pipeline

The OpenFlow pipeline processing defines how packets interact with flow tables contained by a switch.

The flow tables of an OpenFlow switch are sequentially numbered, starting at 0. The packet is first matched with flow entries of the first flow table, which is flow table 0. A flow entry can only direct a packet to a flow table number that is greater than its own flow table number.

When a packet matches a flow entry, the OpenFlow switch updates the action set for the packet and passes the packet to the next flow table. In the last flow table, the OpenFlow switch executes all actions to modify packet contents and specify the output port for packet forwarding. If the instruction set of a flow table contains an action list, the OpenFlow switch immediately executes the actions for a copy of the packet in this table.

Figure 3 OpenFlow forwarding workflow

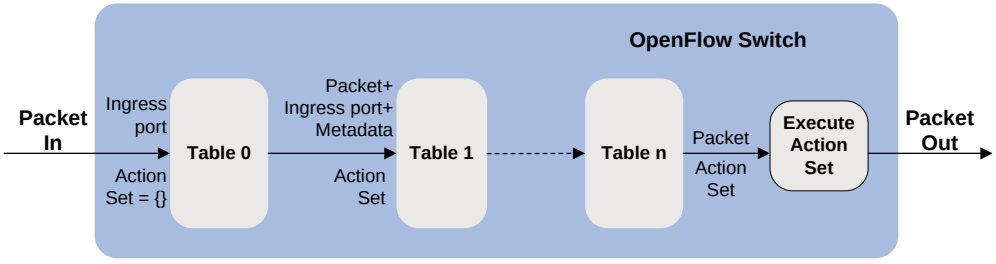


Table-miss flow entry

Every flow table must support a table-miss flow entry to process table misses. The table-miss flow entry specifies how to process packets that were not matched by other flow entries in the flow table.

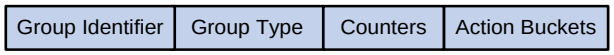
The table-miss flow entry wildcards all match fields (all fields omitted) and has the lowest priority 0.

The table-miss flow entry behaves in most ways like any other flow entry.

Group table

The ability for a flow entry to point to a group enables OpenFlow to represent additional methods of forwarding. A group table contains group entries.

Figure 4 Group entry components



A group entry contains the following fields:

- **Group Identifier**—A 32 bit unsigned integer uniquely identifying the group.
- **Group Type**—Type of the group:
  - **All**—Execute all buckets in the group. This group is used for multicast or broadcast forwarding.
  - **Select**—Execute one bucket in the group.
  - **Indirect**—Execute the one defined bucket in the group.
  - **Fast failover**—Execute the first live bucket.
- **Counters**—Updated when packets are processed by a group.
- **Action Buckets**—An ordered list of action buckets, where each action bucket contains a set of actions to execute and associated parameters. Table 1 shows the actions supported by action buckets.

  Table 1 Actions supported by action buckets

Action	Type	Description
Output	Required	Forwards a packet to a specific OpenFlow port, for example, physical port, logical port, and reserved port.

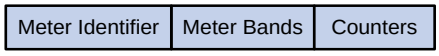
Meter table

Meters enable OpenFlow to implement various simple QoS operations, such as rate-limiting. A meter table contains meter entries.

A meter entry contains the following fields:

- **Meter Identifier**—A 32 bit unsigned integer uniquely identifying the meter.
- **Meter Bands**—Each meter can have one meter band, which specifies the rate at which the band applies and the way packets should be processed. If the current rate of packets exceeds the rate of the band, packets are processed according to the method defined in the band.
- **Counters**—Updated when packets are processed by a meter.

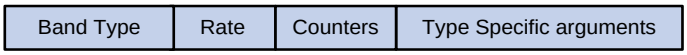
Figure 5 Meter entry components



A meter band contains the following fields:

- **Band Type**—(Optional.) Packet processing methods. Options are:
  - **Drop**—Discards the packet when the rate of the packet exceeds the band rate.
  - **DSCP Remark**—Remarks the DSCP field in the IP header of the packet.
- **Rate**—Defines the lowest rate at which the band can apply.
- **Counters**—Updated when packets are processed by a band.
- **Type Specific Arguments**—Some band types have specific arguments.

Figure 6 Band components



OpenFlow channel

The OpenFlow channel is the interface that connects each OpenFlow switch to a controller. The controller uses the OpenFlow channel to exchange control messages with the switch to perform the following operations:

- Configure and manage the switch.
- Receive events from the switch.
- Send packets out the switch.

The OpenFlow channel is usually encrypted by using TLS. Also, an OpenFlow channel can be run directly over TCP.

The OpenFlow protocol supports the following message types: controller-to-switch, asynchronous, and symmetric. Each message type has its own subtypes.

Controller-to-switch messages

Controller-to-switch messages are initiated by the controller and used to directly manage or inspect the state of the switch. Controller-to-switch messages might or might not require a response from the switch.

The controller-to-switch messages include the following subtypes:

- **Features**—The controller requests the basic capabilities of a switch by sending a features request. The switch must respond with a features reply that specifies the basic capabilities of the switch.
- **Configuration**—The controller sets and queries configuration parameters in the switch. The switch only responds to a query from the controller.
- **Modify-State**—The controller sends Modify-State messages to manage state on the switches. Their primary purpose is to add, delete, and modify flow or group entries in the OpenFlow tables and to set switch port properties.
- **Read-State**—The controller sends Read-State messages to collect various information from the switch, such as current configuration and statistics.
- **Packet-out**—These are used by the controller to send packets out of the specified port on the switch, or to forward packets received through packet-in messages. Packet-out messages must contain a full packet or a buffer ID representing a packet stored in the switch. The message must also contain a list of actions to be applied in the order they are specified. An empty action list drops the packet.
- **Barrier**—Barrier messages are used to confirm the completion of the previous operations. The controller send s Barrier request. The switch must send a Barrier reply when all the previous operations are complete.

- **Role-Request**—Role-Request messages are used by the controller to set the role of its OpenFlow channel, or query that role. It is typically used when the switch connects to multiple controllers.
- **Asynchronous-Configuration**—These are used by the controller to set an additional filter on the asynchronous messages that it wants to receive, or to query that filter. It is typically used when the switch connects to multiple controllers.

Asynchronous messages

Switches send asynchronous messages to controllers to inform a packet arrival or switch state change. For example, when a flow entry is removed due to timeout, the switch sends a flow-removed message to inform the controller.

The asynchronous messages include the following subtypes:

- **Packet-In**—Transfer the control of a packet to the controller. For all packets forwarded to the Controller reserved port using a flow entry or the table-miss flow entry, a packet-in event is always sent to controllers. Other processing, such as TTL checking, can also generate packet-in events to send packets to the controller. The packet-in events can include the full packet or can be configured to buffer packets in the switch. If the packet-in event is configured to buffer packets, the packet-in events contain only some fraction of the packet header and a buffer ID. The controller processes the full packet or the combination of the packet header and the buffer ID. Then, the controller sends a packet-out message to direct the switch to process the packet.
- **Flow-Removed**—Inform the controller about the removal of a flow entry from a flow table. These are generated due to a controller flow delete request or the switch flow expiry process when one of the flow timeouts is exceeded.
- **Port-status**—Inform the controller of a state or setting change on a port.
- **Error**—Inform the controller of a problem or error.

Symmetric messages

Symmetric messages are sent without solicitation, in either direction.

The symmetric messages contain the following subtypes:

- **Hello**—Hello messages are exchanged between the switch and controller upon connection startup.
- **Echo**—Echo request or reply messages can be sent from either the switch or the controller, and must return an echo reply. They are mainly used to verify the liveness of a controller-switch connection, and might also be used to measure its latency or bandwidth.
- **Experimenter**—This is a staging area for features meant for future OpenFlow revisions.

OpenFlow timers

An OpenFlow switch supports the following timers:

- **Connection detection interval**—Interval at which the OpenFlow switch sends an Echo Request message to a controller. When the OpenFlow switch receives no Echo Reply message within three intervals, the OpenFlow switch is disconnected from the controller.
- **Reconnection interval**—Interval for the OpenFlow switch to wait before it attempts to reconnect to a controller.

OpenFlow controller

Controller roles

A switch can establish connections with multiple controllers. When OpenFlow operation is initiated, a switch is simultaneously connected to multiple controllers in Equal state. A controller can request its role to be changed at any time. The controller role contains the following types:

- **Equal**—In this role, the controller has full access to the switch and is equal to other controllers in the same role. By default, the controller receives all switch asynchronous messages such as packet-in and flow-removed messages. The controller can send controller-to-switch messages to modify the state of the switch.
- **Master**—This role is similar to the Equal role and has full access to the switch. The difference is that up to one controller in this role is allowed for a switch.
- **Slave**—In this role, the controller has read-only access to the switch.

  The controller cannot send controller-to-switch messages to perform the following operations:

  - Deploy flow entries, group entries, and meter entries.
  - Modify the port and switch configurations.
  - Send packet-out messages.

  By default, the controller does not receive switch asynchronous messages except Port-status messages. The controller can send Asynchronous-Configuration messages to set the asynchronous message types it wants to receive.

Controller connection modes

An OpenFlow instance can connect to one or more controllers, depending on the controller connection mode the OpenFlow instance uses:

- **Single**—The OpenFlow instance connects to only one controller at a time. When communication with the current controller fails, the OpenFlow instance uses another controller.
- **Multiple**—The OpenFlow instance can simultaneously connect to multiple controllers. When communication with any controller fails, the OpenFlow instance attempts to reconnect to the controller after a reconnection interval.

Main and auxiliary connections

The OpenFlow channel can have one main connection and multiple auxiliary connections.

- **Main connection**—Processes control messages to complete operations such as deploying entries, obtaining data, and sending information. The main connection must be a reliable TCP or SSL connection.
- **Auxiliary connection**—Improves the communication performance between the controller and OpenFlow switches. An auxiliary connection can have the different destination IP address and port number than the main connection.

Connection interruption mode

When an OpenFlow switch is disconnected from all controllers, the OpenFlow switch is set to either of the following modes:

- **Secure**—The OpenFlow switch forwards traffic based on flow tables and does not remove unexpired flow entries. If the output action in a matching flow entry is to forward traffic to a controller, the traffic is discarded. This is the default forwarding mode.
- **Smart**—The OpenFlow switch forwards traffic based on flow tables and does not remove unexpired flow entries. If the output action in a matching flow entry is to forward traffic to a controller, the traffic is forwarded in normal process.
- **Standalone**—The OpenFlow switch uses the normal forwarding process.

The OpenFlow switch forwards traffic based on flow tables when it reconnects to a controller successfully.

Protocols and standards

OpenFlow Switch Specification Version 1.3.3

Restrictions: Licensing requirements for OpenFlow

OpenFlow requires an MC-NAT license to run on the device for receiving group entries that contain Set-Field actions. For more information about the Set-Field actions, see Group table. For more information about licenses, see license management in *Fundamentals Configuration Guide*.

Restrictions and guidelines: OpenFlow configuration

Layer 3 Ethernet interfaces, Layer 3 aggregate interfaces, Layer 3 Ethernet subinterfaces, and Layer 3 aggregate subinterfaces do not support OpenFlow in the current software version.

The device does not support group entries with the output ports as Layer 3 Ethernet interfaces, Layer 3 aggregate interfaces, Layer 3 Ethernet subinterfaces, or Layer 3 aggregate subinterfaces.

The output port for the output action of a group entry cannot be a Layer 3 Ethernet interface.

The device does not support extensibility flow entries with the Set-Field action of setting the ip_ecn (the ECN field in IP packets). For more information about ECN, see *ACL and QoS Configuration Guide*.

When configuring a global instance, follow these restrictions and guidelines:

- When the action of a flow entry is output, the output port cannot be Flood or All.
- When the action of a flow entry is to set the destination MAC or source MAC, you must specify the output action and the output port must be an Ethernet interface or aggregate interface.
- When local first load sharing is enabled for Ethernet link aggregation, the OpenFlow controller cannot create group entries on the device. The two features are mutually exclusive. For more information about load sharing for Ethernet link aggregation, see Ethernet link aggregation in *Layer 2—LAN Switching Configuration Guide*.
- The OpenFlow controller supports creating only group entries with the output port as physical ports or aggregate interfaces.
- The OpenFlow controller does not support creating group entries with the action of setting the IPv6 address.
- A group entry with the action of setting the UDP port number takes effect only on IPv4 UDP packets.
- When the OpenFlow controller creates a group entry on the device, the packets sent out of the output port might carry VLAN tags. Modify the VLAN configuration of the output port as needed. For more information about port VLAN configuration, see VLAN in *Layer 2—LAN Switching Configuration Guide*.

If the output port is also the input port for a packet matching an OpenFlow flow entry, you must execute the **`port bridge enable`** command on the output port to enable bridging. Otherwise, traffic cannot be forwarded properly. For more information about bridging, see Ethernet interfaces in *Layer 2—LAN Switching Configuration Guide*.

Group entries are mutually exclusive with the **`link-aggregation load-sharing mode local-first`** command. For the group table to operate normally, do not execute the **`link-aggregation load-sharing mode local-first`** command on the device.

For an OpenFlow flow entry deployed by the controller to match IPv6 packets, you must specify an IPv6 packet field in the match fields in the flow entry (for example, set the Ethernet type to 0x86DD). To match MPLS packets, you must specify an MPLS packet field in the match fields in the flow entry (for example, set the Ethernet type to 0x8847 or 0x8848).

OpenFlow tasks at a glance

To configure OpenFlow, perform the following tasks:

1. Configuring OpenFlow instances
   a. Creating an OpenFlow instance
   b. Configuring the OpenFlow instance mode
   c. (Optional.) Configuring inband management VLANs
   d. (Optional.) Configuring flow tables and flow entries for an OpenFlow instance

      This task is applicable if the OpenFlow instance does not use TTP flow tables.
   e. (Optional.) Setting the controller connection mode
   f. (Optional.) Preventing an OpenFlow instance from reporting the specified types of ports to controllers
   g. (Optional.) Enabling support for smart group entries

      Perform this task to optimize the IPTV service. A controller supporting smart group entries is required to cooperate with the OpenFlow instance.
   h. Activating or reactivating an OpenFlow instance
   i. (Optional.) Configuring OpenFlow instance attributes
2. Configuring controllers for an OpenFlow switch
3. (Optional.) Refreshing all Layer 3 flow entries in the MAC-IP flow tables for an OpenFlow instance
4. (Optional.) Ignoring the permit flag added by OpenFlow
5. (Optional.) Shutting down an interface by OpenFlow
6. (Optional.) Enabling SNMP notifications for OpenFlow
7. (Optional.) Configuring OpenFlow flow monitor parameters

Configuring OpenFlow instances

Creating an OpenFlow instance

1. Enter system view.

   system-view
2. Create an OpenFlow instance and enter its view.

   openflow instance *instance-id*
3. (Optional) Set the datapath ID.

   datapath-id *id*

   By default, the datapath ID of an OpenFlow instance contains the instance ID and the bridge MAC address of the device. The upper 16 bits are the instance ID and the lower 48 bits are the bridge MAC address of the device.

   The datapath ID uniquely identifies an OpenFlow switch (OpenFlow instance). Do not set the same datapath ID for different OpenFlow switches.
4. Set a DSCP value for OpenFlow packets.

   tcp dscp *dscp-value*

   By default, no DSCP value is set for OpenFlow packets.

Configuring the OpenFlow instance mode

Restrictions and guidelines

When you associate an OpenFlow instance with VLANs, follow these guidelines:

- For VLAN traffic to be processed correctly, associate different OpenFlow instances with different VLANs.
- When you activate an OpenFlow instance that is associated with nonexistent VLANs, the system automatically creates the VLANs.
- Do not associate VLANs permitted on a port with different OpenFlow instances. Otherwise, port modification messages of different OpenFlow instances deployed from different controllers overwrite each other.
- Do not configure BFD MAD on the VLAN interface for a VLAN that is associated with an OpenFlow instance. For more information about BFD MAD, see *Virtual Technologies Configuration Guide*.

Enabling the global mode for an OpenFlow instance

1. Enter system view.
   system-view
2. Enter OpenFlow instance view.
   openflow instance *instance-id*
3. Enable the global mode for the OpenFlow instance.
   Classification global
   By default, the OpenFlow instance mode is not configured.

Enabling the VLAN mode for an OpenFlow instance

1. Enter system view.
   system-view
2. Enter OpenFlow instance view.
   openflow instance *instance-id*
3. Configure the OpenFlow instance mode.
   classification vlan *vlan-id* [**mask** *vlan-mask*] [**loosen**]
   By default, the OpenFlow instance mode is not configured.

Configuring inband management VLANs

About this task

Traffic in the inband management VLANs are forwarded in the normal forwarding process instead of the OpenFlow forwarding process. Inband management VLANs are used by an OpenFlow instance to establish OpenFlow channels to controllers.

Restrictions and guidelines

The ports that are assigned only to inband management VLANs are not OpenFlow ports.

Inband management VLANs configured for an OpenFlow instance in VLAN mode must be within the list of the VLANs associated with the OpenFlow instance.

If an OpenFlow instance in global or VLAN mode connects to a controller through a non-management Ethernet interface, configure the VLAN to which the interface belongs as an inband management VLAN.

Procedure

1. Enter system view.
   system-view
2. Enter OpenFlow instance view.
   openflow instance *instance-id*
3. Configure inband management VLANs for the OpenFlow instance.
   in-band management vlan *vlan-id-list*
   By default, no inband management VLANs are configured for an OpenFlow instance.

Configuring flow tables and flow entries for an OpenFlow instance

Restrictions and guidelines

The extensibility flow table ID must be greater than the MAC-IP flow table ID.

Procedure

1. Enter system view.
   system-view
2. Enter OpenFlow instance view.
   openflow instance *instance-id*
3. Configure flow tables for the OpenFlow instance.
   flow-table { **mac-ip** *mac-ip-table-id* | **extensibility** *extensibility-table-id* }*
   By default, an OpenFlow instance contains one extensibility flow table with an ID of 0.
4. Set the maximum number of flow entries that each extensibility flow table supports.
   flow-entry max-limit *limit-value*
   By default, each extensibility flow table supports a maximum of 65535 flow entries.
   When the maximum number is reached, the OpenFlow instance does not accept new flow entries for that table and sends a deployment failure notification to the controller.
5. Configure the OpenFlow instance to allow dynamic ARP entries to overwrite OpenFlow ARP entries.
   precedence dynamic arp
   By default, an OpenFlow instance does not allow dynamic ARP entries to overwrite OpenFlow ARP entries.
   Only MAC-IP flow tables support this feature.
6. Allow the deployed flow tables to include link aggregation member ports.
   permit-port-type member-port
   By default, the deployed flow tables cannot include link aggregation member ports.
7. Configure the default action of table-miss flow entries to forward packets to the normal pipeline.
   default table-miss permit
   By default, the default action of table-miss flow entries is to drop packets.

Setting the controller connection mode

Procedure

1. Enter system view.
   system-view
2. Enter OpenFlow instance view.
   openflow instance *instance-id*
3. Set the controller connection mode.
   controller mode { **multiple** | **single** }
   By default, the **multiple** mode is used.

Preventing an OpenFlow instance from reporting the specified types of ports to controllers

About this task

Perform this task to prevent an OpenFlow instance from reporting controllers information about the following types of interfaces that belong to the OpenFlow instance:

- Layer 3 Ethernet interface.
- Layer 3 aggregate interface.
- VLAN interface.
- Virtual switch instance (VSI) interface.

Procedure

1. Enter system view.
   system-view
2. Enter OpenFlow instance view
   openflow instance *instance-id*
3. Prevent an OpenFlow instance from reporting the specified types of ports to controllers.
   forbidden port { **l3-physical-interface** | **vlan-interface** | **vsi-interface** } *
   By default, no port types are prevented from being reported to the controllers. All ports that belong to an OpenFlow instance are reported to the controllers.

Enabling support for smart group entries

About this task

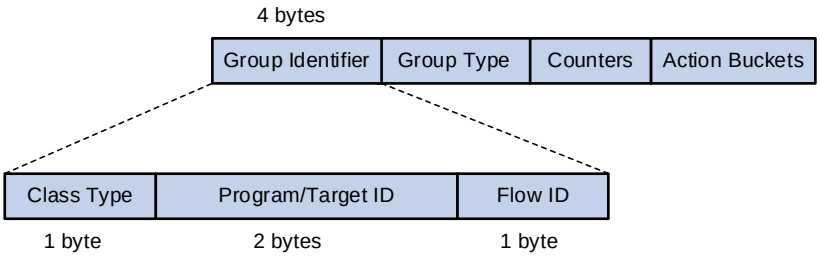
On an IPTV service network, the traffic of an IPTV program might contain multiple types of traffic, such as flows of videos, audios in different languages, and subtitles in different languages. The controller needs to issue multiple flow entries for the OpenFlow instance to forward the traffic flows of an IPTV program to target users. The speed of program switchover might be affected due to the controller performance, which will further decrease the user experience.

To address this issue, support for smart group entries is introduced. The controller can issue smart group entries that define all possible forwarding actions for the IPTV service in advance. Once a program switchover occurs, the controller sends only a controller-to-switch message instead of multiple flow entries to the OpenFlow switch. This message will instruct the switch to combine smart group entries to forward the traffic flows of the new program to the target user.

Smart group entries include the following types:

- **Program group entry**—Identifies forwarding actions for flows of an IPTV program. A program group entry can be used by an OpenFlow flow entry to provide extra forwarding functions. By default, a program group entry does not contain any buckets. To include buckets in the program group entry, bind the program group entry to target group entries.
- **Target group entry**—Identifies a target user for an IPTV program. A target group entry contains only a bucket, which defines the output port. Target group entries cannot be used by OpenFlow flow entries but they can be bound to a program group entry. A program group entry can be bound to multiple target group entries to forward flows of an IPTV program to multiple target users.

Figure 7 Smart group entry components



A smart group entry contains the following fields:

- **Class Type**—Identifies the type of a smart group entry. This field contains one byte. 0x00 indicates a program group entry and 0x01 indicates a target group entry.
- **Program/Target ID**—Identifies the ID of a program or target user. This field contains two bytes.
- **Flow ID**—Identifies the ID of a flow for a program or target user. This field contains one byte.

Restrictions and guidelines

To modify a target group entry issued to an OpenFlow instance, the controller must first delete the target group entry and then re-issue the target group entry.

Procedure

1. Enter system view.
   `system-view`
2. Enter OpenFlow instance view.
   `openflow instance` *`instance-id`*
3. Enable the OpenFlow instance to support smart group entries.
   `smart-group enable`
   By default, support for smart group entries is disabled for an OpenFlow instance.

Activating or reactivating an OpenFlow instance

Restrictions and guidelines

To avoid failure to activate an OpenFlow instance containing a MAC-IP flow table, first execute the **`default table-miss permit`** command.

Procedure

1. Enter system view.
   `system-view`
2. Enter OpenFlow instance view.

openflow instance *instance-id*

3. Activate or reactivate the OpenFlow instance.

   active instance

   By default, an OpenFlow instance is not activated.

Configuring OpenFlow instance attributes

Restrictions and guidelines

The OpenFlow instance attribute configurations of an OpenFlow instance can take effect without activation for the OpenFlow instance.

Procedure

1. Enter system view.

   system-view

2. Enter OpenFlow instance view.

   openflow instance { *instance-id* }

3. Configure a description for the OpenFlow instance.

   description *text*

   By default, an OpenFlow instance does not have a description.

4. Set OpenFlow timers.
   - Set the connection detection interval.

     controller echo-request interval *interval*

     The default setting is 5 seconds.
   - Set the reconnection interval.

     controller connect interval *interval*

     The default setting is 60 seconds.

5. Configure MAC address-related features.
   - Forbid MAC address learning for VLANs associated with the OpenFlow instance.

     mac-learning forbidden

     By default, MAC address learning is allowed for VLANs associated with an OpenFlow instance.

     The configuration does not take effect on inband management VLANs.
   - Configure the OpenFlow instance to support matching the dynamic MAC addresses in the query and deletion flow entry instructions sent from controllers.

     mac-ip dynamic-mac aware

     By default, an OpenFlow instance ignores dynamic MAC addresses in the query and deletion flow entry instructions sent from controllers.

     Only MAC-IP flow tables support this feature.

6. Prevent the OpenFlow instance from reporting ARP packets to the specified controllers.

   forbidden packet-in arp controller *controller-id-list*

   By default, no controllers to which ARP packets are forbidden to be reported are configured.

   This feature forbids an OpenFlow instance to report ARP packets to the specified controllers to prevent the controllers from being affected by a large number of packets.

7. Disable logging for successful flow table modifications.

   flow-log disable

   By default, logging for successful flow table modifications is enabled.

8. Enable loop guard for the OpenFlow instance.

   loop-protection enable

   By default, loop guard is disabled for an OpenFlow instance.

   After an OpenFlow instance is deactivated, you can enable loop guard for the OpenFlow instance to avoid loops. This feature enables the deactivated OpenFlow instance to create a flow entry for dropping all traffic in these VLANs.

Configuring controllers for an OpenFlow switch

Restrictions and guidelines

Make sure the configuration of an auxiliary connection does not conflict with the configuration of the main connection. Otherwise, the auxiliary connection cannot be established.

Procedure

1. Enter system view.

   system-view

2. Enter OpenFlow instance view.

   openflow instance *instance-id*

3. Specify a controller and configure the main connection to the controller.

   controller *controller-id* **address** { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**port** *port-number*] [**local address** { **ip** *local-ipv4-address* | **ipv6** *local-ipv6-address* } [**port** *local-port- number*]] [**ssl** *ssl-policy-name* [**access-control-policy** *acp-policy-name*]] [**vrf** *vrf-name*]

   By default, an OpenFlow instance does not have a main connection to a controller.

   The source IP address must be the IP address of a port belonging to an OpenFlow instance. Otherwise, the OpenFlow switch might fail to establish a connection with the controller.

4. (Optional) Specify a controller and configure an auxiliary connection to the controller.

   controller *id* **auxiliary** *auxiliary-id* **transport** { **tcp** | **udp** | **ssl** *ssl-policy-name* } [**address** { **ip** *ipv4-address* | **ipv6** *ipv6-address* }] [**port** *port-number*]

   By default, an OpenFlow instance does not have auxiliary connections to a controller.

   If no destination IP address and port number are specified, the auxiliary connection uses the destination IP address and port number configured for the main connection.

5. (Optional) Set the connection interruption mode.

   fail-open mode { **secure** | **smart** | **standalone** }

   By default, the **secure** mode is used.

6. (Optional) Enable OpenFlow connection backup.

   tcp-connection backup

   By default, OpenFlow connection backup is enabled.

   This feature prevents connection interruption when an active/standby switchover occurs.

   This feature is available only on an IRF fabric with two member devices.

   This feature is available for only OpenFlow connections established over TCP.

Refreshing all Layer 3 flow entries in the MAC-IP flow tables for an OpenFlow instance

About this task

Perform this task to obtain all Layer 3 flow entries in the MAC-IP flow tables from the controller again if the Layer 3 flow entries have been overwritten.

Restrictions and guidelines

1. Enter system view.
   system-view
2. Enter OpenFlow instance view.
   openflow instance *instance-id*
3. Refresh all Layer 3 flow entries in the MAC-IP flow tables.
   refresh ip-flow

Ignoring the permit flag added by OpenFlow

About this task

Packets matching an OpenFlow flow entry with the output action as Normal are permitted to pass through and marked with a permit flag. Packets with a permit flag are not dropped by a discard action in a QoS policy (for example, a CAR discard action).

This feature allows the device to drop packets with a permit flag by using a discard action in a QoS policy.

Procedure

1. Enter system view.
   system-view
2. Ignore the permit flag added by OpenFlow.
   openflow permit-flag ignore
   By default, the permit flag added by OpenFlow is not ignored.

Shutting down an interface by OpenFlow

About this task

After an interface is shut down by OpenFlow, the **Current state** field displays **OFP DOWN** in the **display interface** command output.

You can use the **undo openflow shutdown** command to bring up an interface shut down by OpenFlow. The interface can also be brought up by port modification messages from controllers.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Shut down an interface by OpenFlow.

openflow shutdown

By default, an interface is not shut down by OpenFlow.

Enabling SNMP notifications for OpenFlow

About this task

This feature enables OpenFlow to generate SNMP notifications for important OpenFlow events. For the notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see SNMP configuration in *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for OpenFlow.

   snmp-agent trap enable openflow [**connect-state**]

   By default, SNMP notifications are enabled for OpenFlow.

Configuring OpenFlow flow monitor parameters

About this task

The controller can use NETCONF to enable the flow monitor feature for flow entries on the device. On the device, you can configure flow monitor parameters at the CLI.

For an OpenFlow flow entry with flow monitor enabled, the device monitors the changes in the numbers of bytes and packets matching the flow entry at the flow monitor interval. If these counters do not change consecutively for the specified times, the device uses NETCONF to report to the controller the event that the traffic matching the flow entry does not increase. If these counters increase later, the device uses NETCONF to report to the controller the event that the traffic matching the flow entry increases.

NOTE:

Use the **display openflow flow-table** command to view flow entries. In the command output, the byte count and packet count fields for a flow entry represent the number of bytes matching the flow entry and the number of packets matching the flow entry.

The controller deploys the <action> operation to enable flow monitor for a flow entry on the device. The following NETCONF message enables flow monitor for all flow entries with CookieID 1 in flow table 0 for the OpenFlow instance corresponding to the specified datapath ID.

```
<rpc message-id="100" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <action>
    <error-option>stop-on-error</error-option>
    <top xmlns="http://www.h3c.com/netconf/action:1.0">
       <OFP>
         <FlowMonitor>
            <Application>
               <DatapathID>410026760491999</DatapathID>
               <TableID>0</TableID>
               <CookieID>1</CookieID>
               <Enable>1</Enable>
```

```
            </Application>
          </FlowMonitor>
        </OFP>
      </top>
    </action>
</rpc>
```

Restrictions and guidelines

With OpenFlow flow monitor enabled for a large number of flow entries, the CPU load of the device might be too heavy if the configured flow monitor interval is small.

The OpenFlow flow monitor feature monitors only the traffic matching the flow entries deployed by the controller to an OpenFlow instance, and does not monitor flow entries with the default and local attributes.

After the OpenFlow flow monitor feature is enabled for a flow entry, the idle time attribute (the flow entry is removed if the flow entry does not match any data stream during the idle time) that the controller deploys to the flow entry does not take effect.

Procedure

1. Enter system view.
   `system-view`
2. Enter OpenFlow instance view.
   `openflow instance` *`instance-id`*
3. Configure the flow monitor interval.
   `flow-monitor interval` *`interval`*

   By default, the flow monitor interval is not configured.

   If you do not configure the flow monitor interval, the OpenFlow flow monitor feature does not take effect.
4. (Optional.) Set the monitor threshold for consecutive traffic interruptions.
   `flow-monitor threshold` *`threshold`*

   By default, the monitor threshold for consecutive traffic interruptions is 2.

Display and maintenance commands for OpenFlow

Execute **`display`** commands in any view.

Task	Command
Display the detailed information for an OpenFlow instance.	**`display openflow instance`** [*`instance-id`*]
Display controller information for an OpenFlow instance.	**`display openflow instance`** *`instance-id`* { **`controller`** [*`controller-id`*] | **`listened`** }
Display auxiliary connection information.	**`display openflow instance`** *`instance-id`* **`auxiliary`** [*`controller-id`* [**`auxiliary`** *`auxiliary-id`*]]
Display flow table entries for an OpenFlow instance.	**`display openflow instance`** *`instance-id`* **`flow-table`** [*`table-id`*]

Task	Command
Display group table information for an OpenFlow instance.	**display openflow instance** *instance-id* **group** [*group-id*]
Display meter table information for an OpenFlow instance.	**display openflow instance** *instance-id* **meter** [*meter-id*]
Display smart group entry information for an OpenFlow instance.	**display openflow instance** *instance-id* **smart-group** { **program-group** [*program-id*] | **target-group** [*target-id*] }
Display summary OpenFlow instance information.	**display openflow summary**
Clear statistics on packets that a controller sends and receives for an OpenFlow instance.	**reset openflow instance** *instance-id* { **controller** { **controller** [*controller-id*] | **listened** }

OpenFlow configuration examples

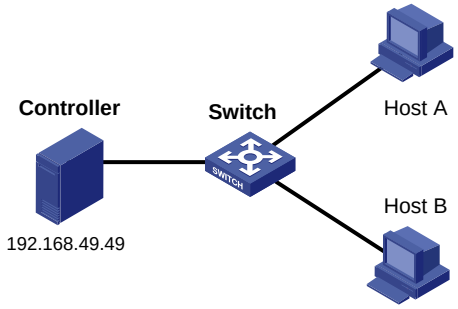
Example: Configuring OpenFlow in VLAN mode

Network configuration

As shown in Figure 8, an OpenFlow switch communicates with the controller. Perform the following tasks on the OpenFlow switch:

- Create OpenFlow instance 1, associate VLANs 4092 and 4094 with the OpenFlow instance, and activate the OpenFlow instance.
- Configure the IP address for controller 1 to have the controller manage the OpenFlow switch.

Figure 8 Network diagram



Procedure

Create VLANs 4092 and 4094.

```
<Switch> system-view
[Switch] vlan 4092
[Switch-vlan4092] quit
[Switch] vlan 4094
[Switch-vlan4094] quit
```

Create OpenFlow instance 1 and associate VLANs with it.

```
[Switch] openflow instance 1
```

```
[Switch-of-inst-1] classification vlan 4092 mask 4093
```

Specify controller 1 for OpenFlow instance 1 and activate the instance.

```
[Switch-of-inst-1] controller 1 address ip 192.168.49.49
[Switch-of-inst-1] active instance
[Switch-of-inst-1] quit
```

Verifying the configuration

View detailed information about the OpenFlow instance.

```
[Switch] display openflow instance 1
Instance 1 information:

Configuration information:
 Description   : --
 Active status : Active
 Inactive configuration:
  None
 Active configuration:
  Classification VLAN, total VLANs(2)
   4092, 4094
...
Port information:
 GigabitEthernet1/0/3
Active channel information:
 Controller 1 IP address: 192.168.49.49 port: 6633
```

Appendix A Application restrictionsFlow entry restrictions

Matching restrictions

VLAN matching

Table 2 describes the VLAN matching restrictions when an OpenFlow instance is associated with VLANs.

Table 2 VLAN matching

VLAN	Mask	Matching packets
-	-	All packets in the VLANs that are associated with the OpenFlow instance.
0	-	Packets without a VLAN tag. The PVID of the ingress port must be associated with the OpenFlow instance.
0	Value	Unsupported.
Valid VLAN	-/value	Unsupported.
0x1000	-/value (except 0x1000)	Unsupported.
0x1000	0x1000	Packets with a VLAN tag. The VLAN ID of the VLAN tag must be associated with the OpenFlow instance.
Valid VLAN | 0x1000	-/value	Matching packets by the combination of the VLAN ID and VLAN mask. The VLANs obtained through the combination of the VLAN ID and VLAN mask must be associated with the OpenFlow instance.
Other	Other	Unsupported.

Protocol packet matching

If protocols are enabled, protocol packets (except LLDP frames) are processed by the corresponding protocols instead of the OpenFlow protocol.

For more information about LLDP frame matching, see "LLDP frame matching."

Metadata matching

Metadata passes matching information between flow tables. The controller deploys metadata matching entries only to non-first flow tables. If the controller deploys a metadata matching entry to the first flow table, the switch returns an unsupported flow error.

Instruction restrictions

Table 3 Instruction restrictions

Instruction type	Restrictions
Clear-Actions	The Clear-Actions instruction has the following restrictions: • For the single flow table, the flow entries of the table cannot include this instruction and other instructions at the same time. • For multiple flow tables of the pipeline, only the flow entries of the first flow table can include this instruction and other instructions at the same time.
Apply-Actions	The action list of the Apply-Actions instruction cannot include multiple Output actions.

Instruction type	Restrictions
	When the action list includes only one Output action, the switch processes the action list as described in "Restrictions for merging the action list into the action set."
Write-Metadata/mask	The flow entries of the last table of the pipeline cannot include this instruction. Otherwise, the switch returns an unsupported flow error.
Goto-Table	

Restrictions for merging the action list into the action set

The switch follows the following restrictions to merge the action list into the action set.

Non-output actions

When the action set and the action list do not contain the Output or Group action, the following rules apply:

- If actions in the action set do not conflict with actions in the action list, the switch merges the action list into the action set.
- If actions in the action set conflict with actions in the action list, actions in the action list are replaced with actions in the action set.

Output actions

When the action set and the action list contain the Output action or the Group action, the following rules apply:

- If both the action list and the action set contain an Output action, the Output action in the action list takes precedence. The Output action in the action list does not modify the packet. The Output action in the action set is executed at the last step of the pipeline processing to modify the packet.
- If either the action list or the action set contains an Output action, the port specified by the Output action is treated as the output port. The actions are executed in the order defined by the action set rules.
- If the action list contains an Output action and the action set contains a Group action, the following rules apply:
  - The Output action does not modify the packet.
  - The Group action is executed.

Packet-out messages restrictions

Ingress port

The ingress port must be a physical or logical port when one of the following reserved ports is the output port in a packet-out message:

- Normal.
- Local.
- In Port.
- Controller.

Buffer ID co-existing with packet

If a packet-out message contains both the packet and the buffer ID representing the packet stored in the switch, the switch processes only the buffered packet. The switch ignores the packet in the message.

Packets without a VLAN tag

If the packet contained in a packet-out message has no VLAN tag, the switch performs the following operations:

- Tags the packet with the PVID of the ingress port.
- Forwards the packet within the VLAN.

The switch processes the packet as follows when the ingress port is a reserved port:

- If the output port is a physical or logical port, the switch tags the packet with the PVID of the output port and forwards the packet within the VLAN.
- If the output port is the Flood or All reserved port, the switch processes the packet as described in "Output port."

Output port

If the output port in a packet-out message is the Flood or All reserved port, the switch processes the packet contained in the packet-out message as follows:

- When the output port is the Flood reserved port:
  - If the packet has a VLAN tag, the switch broadcasts the packet within the VLAN.
  - If the packet has no VLAN tag and the ingress port is a physical or logical port, the switch tags the packet with the PVID of the ingress port. The switch then forwards the packet within the VLAN.
  - If the packet has no VLAN tag and the ingress port is the Controller reserved port, the switch forwards the packet out all OpenFlow ports.
- When the output port is the All reserved port:
  - If the packet has a VLAN tag, the switch broadcasts the packet within the VLAN.
  - If the packet has no VLAN tag, the switch forwards the packet out of all OpenFlow ports regardless of the ingress port type.

Packet-in messages restrictions

Processing VLAN tags

When sending a packet-in message to the controller, the switch processes the VLAN tag of the packet contained in the packet-out message as follows:

- If the VLAN tag of the packet is the same as the PVID of the ingress port, the switch removes the VLAN tag.
- If the VLAN tag of the packet is different from the PVID of the ingress port, the switch does not remove the VLAN tag.

Packet buffer

If a packet-in message is sent to controller due to no matching flow entry, the switch supports buffering the packet contained in the packet-in message. The buffer size is 1K packets.

If a packet-in message is sent to controller for other reasons, the switch does not support buffering the packet contained in the packet-in message. The switch must send the full packet to the controller, and the cookie field of the packet is set to 0xFFFFFFFFFFFFFFFF.

LLDP frame matching

LLDP is used to perform topology discovery in an OpenFlow network. LLDP must be enabled globally on a device. A switch sends a LLDP frame to the controller through the packet-in message when the following conditions exist:

- The port that receives the LLDP frame from the controller belongs to OpenFlow instances.

- The flow tables in the OpenFlow instance have a flow entry that matches the LLDP frame (the output port is the Controller reserved port).

Flow table modification messages restrictions

The flow table modification messages have the following restrictions for the table-miss flow entry and common flow entries.

Table-miss flow entry

The controller deploys the table-miss flow entry (the action is Drop) to an OpenFlow instance after the OpenFlow instance is activated.

The controller cannot query the table-miss flow entry through Multipart messages.

The controller cannot modify the table-miss flow entry through the Modify request. The controller can only modify the table-miss flow entry through the Add request.

The controller can modify or delete the table-miss flow entry only through the strict version of the Modify or Delete request. The controller cannot modify or remove the table-miss flow entry through the non-strict version of the Modify or Delete request despite that the match fields are wildcarded.

The controller deploys a table-miss flow entry (the action is Drop) to an OpenFlow instance after the current table-miss flow entry is deleted.

Common flow entries

The controller cannot modify or remove all common flow entries through the non-strict version of the Modify or Delete request despite that the match fields are wildcarded.

Appendix B MAC-IP flow tableCapabilities supported by the MAC-IP flow table

The controller must include the required match fields and actions and can include the optional match fields and actions in the flow entries deployed to the MAC-IP flow table. If the controller does not include the optional match fields and actions in the flow entries, the switch adds them to the flow entries by default.

The Layer 2 flow entries are implemented by using MAC address entries. Table 4 describes the capabilities supported by Layer 2 flow entries.

Table 4 Capabilities supported by Layer 2 flow entries

Item	Capabilities
Required match fields	The MAC-IP flow table must support the following match fields: • VLAN ID. • Unicast destination MAC address.
Optional match fields	N/A
Required actions	Specifying the output port.
Optional actions	The MAC-IP flow table can optionally support the following instructions: • **Goto-Table**—When the switch has multiple tables, the switch adds this instruction by default if the controller does not deploy it. • **Write-Metadata**—When the switch has multiple tables, the switch adds this instruction by default if the controller does not deploy it.

The Layer 3 flow entries are implemented by using routing entries. Table 5 describes the capabilities supported by Layer 3 flow entries.

Table 5 Capabilities supported by Layer 3 flow entries

Item	Capabilities
Required match fields	The MAC-IP flow table must support the following match fields: • VLAN ID. • Unicast destination IP address. • Unicast destination MAC address, which must be the MAC address of the VLAN interface for the VLAN that is matched.
Optional match fields	N/A
Required actions	Specifying the output port.
Optional actions	The MAC-IP flow table can optionally support the following actions: • **Modify source MAC address**—The switch modifies the source MAC address to the MAC address of the VLAN interface for the VLAN to which the output port belongs. • **Decrement TTL by one**. • **Goto-Table**—When the switch has multiple tables, the switch adds this instruction by default if the controller does not deploy it. • **Write-Metadata**—When the switch has multiple tables, the switch adds this instruction by default if the controller does not deploy it.

MAC-IP flow table restrictions

Controller must follow the restrictions in the following tables to deploy flow entries for MAC-IP flow table. Otherwise, forwarding failure might occur.

Table 6 Restrictions for deploying Layer 2 flow entries for the MAC-IP flow table

Item	Restrictions
Match fields	The destination MAC address cannot be the MAC address of the switch to which the flow entry is deployed.
Actions	The output port must belong to the VLAN that is matched.

Table 7 Restrictions for deploying Layer 3 flow entries for the MAC-IP flow table

Item	Restrictions
Match fields	The VLAN interface of the VLAN that is matched is in up state. The destination MAC address is the MAC address of the VLAN interface for the VLAN that is matched. The destination IP address cannot be the IP address of the switch to which the flow entry is deployed.
Actions	The specified output port must belong to the destination VLAN. The destination MAC address cannot be the MAC address of the switch to which the flow entry is deployed. If the switch modifies the source MAC address, the source MAC address must be the MAC address of the VLAN interface for the VLAN to which the output port belongs.

To deploy a Layer 3 flow entry, make sure the following requirements are met:

- The VLAN interface of the matched VLAN is in up state.

- The switch sends the controller a packet that indicates the VLAN interface acts as an OpenFlow port. The link state and the MAC address of the VLAN interface are also included in the packet.

The switch reports the VLAN interface deletion to the controller and the controller removes the corresponding Layer 3 flow entry.

The controller ensures the correctness of Layer 3 flow entries. The switch does not check for the restrictions for Lay 3 flow entries.

Table-miss flow entry of MAC-IP flow tables

The table-miss flow entry of a MAC-IP flow table supports the following output actions:

- **Goto-Table**—Direct the packet to the next table.
- **Drop**—Drop the packet.
- **Controller**—Send the packet to the controller.
- **Normal**—Forward the packet to the normal pipeline.

Dynamic aware

On an OpenFlow switch that supports MAC-IP flow tables, you can configure OpenFlow to support querying and deleting dynamic MAC address flow entries.

The controller can query and delete dynamic MAC address flow entries by specifying a VLAN, a MAC address, or the combination of a MAC address and a VLAN.

MAC-IP flow table cooperating with extensibility flow table

Metadata/mask

The MAC-IP flow table supports the Write Metadata/mask instruction and the extensibility flow table supports metadata/mask matching. The MAC-IP flow table can cooperate with an extensibility flow table to perform the pipeline process of multiple tables by using metadata/mask.

Each metadata mask bit has a different meaning. The corresponding metadata bit being set indicates that the metadata mask bit is matched. When the corresponding metadata bit is not set, the metadata mask bit is wildcarded.

Table 8 Metadata mask meanings

Metadata mask bit	Meaning	Metadata
Bit 0	Destination MAC address	• **1**—Set. Matches the destination MAC address. • **0**—Not set. Does not match the destination MAC address.
Bit 1	Source MAC address	• **1**—Set. Matches the source MAC address. • **0**—Not set. Does not match the source MAC address.
Bit 2	Destination IP address	• **1**—Set. Matches the destination IP address. • **0**—Not set. Does not match the destination IP address.
Others	Reserved	Reserved.

Matching restrictions

When the output action in an extensibility flow table is not Normal, the following rules apply:

- The MAC-IP flow table does not take effect.
- All actions are executed according to the extensibility flow table.

When the output action in an extensibility flow table is Normal, the following rules apply:

- The output action is executed according to the MAC-IP flow table.
- The other actions are executed according to the extensibility flow table.

Capabilities supported by the VLAN tagging flow table

The controller must include the required match fields and actions and can include the optional match fields and actions in the flow entries deployed to the VLAN tagging flow table. If the controller does not include the optional match fields and actions in the flow entries, the switch adds them to the flow entries by default.

Table 9 describes the capabilities supported by the flow entries in the VLAN tagging flow table.

Table 9 Capabilities supported by flow entries in the VLAN tagging flow table

Item	Capabilities
Required match fields	The VLAN tagging flow table must support the following match fields: • **input-port**. • **vlan**.
Optional match fields	N/A
Required actions	The following actions in the action list of the Apply-Actions instruction must be applied immediately: • **Push-Tag**. • **Set-Field (VLAN)**.
Optional actions	The VLAN tagging flow table can optionally support the following actions: • **Output (normal)**. • **Goto-Table**.

The Push-Tag and Set-Field (VLAN) actions must be in the action list of the Apply-Actions instruction. The Push-Tag and Set-Field (VLAN) actions can be used as follows:

- **Push-Tag + Set-Field (value1)**—Adds a VLAN tag **value1**.
- **Set-Field (value1) + Push-Tag + Set-Field (value2)**—Modifies the VLAN tag of the packet to **value1** and adds a VLAN tag **value2**.
- **Push-Tag + Set-Field (value1) + Push-Tag + Set-Field (value2)**—Adds inner VLAN tag **value2** and outer VLAN tag **value1**.

The **Goto-Table** instruction is optional and does not take effect. The flow table specified by this instruction can only be the next of the VLAN tagging flow table.

Capabilities supported by the VLAN untagging flow table

Table 10 Capabilities supported by flow entries in the VLAN untagging flow table

Item	Capabilities
Required match fields	The VLAN untagging flow table must support the following match fields: • **egress port**—Matches the egress port of packets.

Item	Capabilities
	• **vlan**—Matches the outer VLAN tag of packets.
Optional match fields	The VLAN untagging flow table can optionally support the **inner vlan** match field that matches the inner VLAN tag of double-tagged packets.
Required actions	The following actions in the action list of the Apply-Actions instruction must be applied immediately: • **Pop-Tag**. • **Set-Field (VLAN)**.
Optional actions	The VLAN untagging flow table can optionally support the **Output (normal)** action.

The VLAN untagging flow table applies only to double-tagged packets.

The **egress port** and **inner vlan** are extended match fields that use the Experimenter ID H3C (0xFE2) and take the private match field values 47 and 48, respectively. To deploy flow entries that contain the extended match fields, make sure the controllers are developed to be compatible with the Experimenter ID and the extended match fields.

The Pop-Tag and Set-Field (VLAN) actions must be in the action list of the Apply-Actions instruction. The Pop-Tag and Set-Field (VLAN) actions can be used as follows:

- **Pop-Tag**—Removes the outer VLAN tag.
- **Pop-Tag + Pop-Tag**—Removes the inner and outer VLAN tags.
- **Pop-Tag + Set-Field (value)**—Removes the outer VLAN tag and modifies the inner VLAN tag to **value**.

Multi-level table-based TTP pipeline

A TTP pipeline can include multiple types of flow tables and the tables can be cascaded. Figure 9 shows the forwarding workflow of a TTP pipeline.

Figure 9 Multi-level table-based TTP pipeline

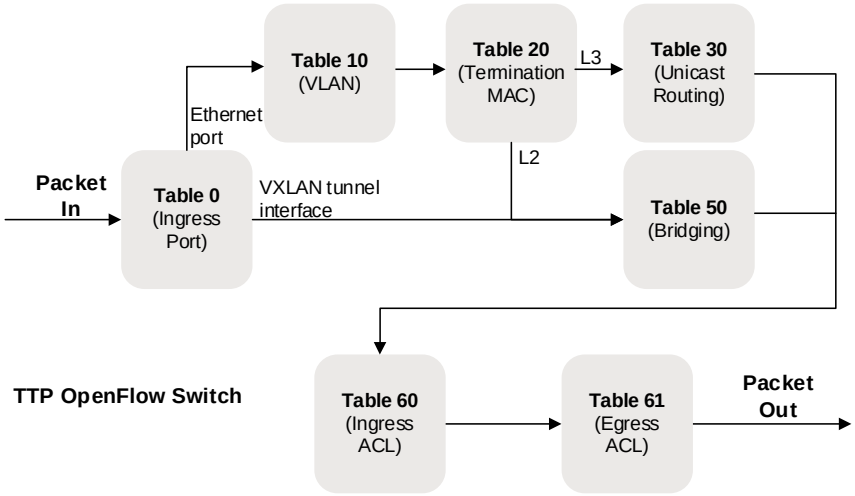


Table-miss flow entry of TTP flow tables

The table-miss flow entry of a TTP flow table supports the following output actions:

- **Normal**—Forward the packet to the normal pipeline.
- **Goto-table-50**—Direct the packet to table 50 (the TTP-Bridging flow table).

- **Controller**—Send the packet to the controller.

Table 11 describes the instruction supported by the table-miss flow entry of different types in TTP flow tables. The instruction in the table-miss entry of TTP flow tables 0 through 50 cannot be modified.

Table 11 Instruction supported by the table-miss flow entry of TTP flow tables

Table ID	TTP flow table	Instruction in the table-miss entry
0	TTP-ingress-port	Normal
10	TTP-VLAN	Normal
20	TTP-termination-MAC	Goto-table-50
30	TTP-unicast-routing	Normal
50	TTP-bridging	Controller
60	TTP-ingress-ACL	N/A
61	TTP-egress-ACL	N/A

Capabilities supported by the TTP-VLAN flow table

A TTP-VLAN flow entry is used to match untagged or tagged packets arriving at a physical interface, direct the packets to the TTP-termination-MAC flow table, and transmit VXLAN ID information.

Table 12 Capabilities supported by TTP-VLAN flow entries (for untagged packets)

Item	Capabilities
Required match fields	N/A
Optional match fields	**In-Port**—Type of the input port (physical interface).
Required actions	• **Write-metadata**—Writes the VNI information. • **Goto-table-20**—Directs the packet to table 20 (TTP-termination-MAC flow table).
Optional actions	N/A

Table 13 Capabilities supported by TTP-VLAN flow entries (for tagged packets)

Item	Capabilities
Required match fields	N/A
Optional match fields	• **In-Port**—Type of the input port (physical interface). • **VLAN ID**—VLAN ID.
Required actions	• **Write-metadata**—Writes the VNI information. • **Goto-table-20**—Directs the packet to table 20 (TTP-termination-MAC flow table).
Optional actions	N/A

Capabilities supported by the TTP-termination-MAC flow table

A TTP-termination-MAC flow entry is used to select Layer 2 or 3 forwarding based on the destination MAC address of a packet, direct the packet to a flow table, and transmit the VRF information.

Table 14 Capabilities supported by TTP-termination-MAC flow entries

Item	Capabilities
Required match fields	N/A
Optional match fields	• **ETH_TYPE**—Ethernet type. Value 0x800 indicates Ethernet. • **Metadata**—VNI. • **ETH_DST**—Destination MAC address (the MAC address of the local gateway).
Required actions	• **Write-metadata**—Writes VRF index information. Value 0 indicates the public network. • **Goto-table**—Directs the packet to a table. ○ If the packet matches the value of the ETH_DST field, the switch directs the packet to table 30 (TTP-unicast-routing flow table) for Layer 3 forwarding. ○ If the packet does not match the value of the ETH_DST field, the switch directs the packet to table 50 (TTP-bridging flow table) for Layer 2 forwarding.
Optional actions	N/A

Capabilities supported by the TTP-unicast-routing flow table

A TTP-unicast-routing flow entry is used to forward Layer 3 packets.

Table 15 Capabilities supported by TTP-unicast-routing flow entries (for normal IPv4 unicast packets)

Item	Capabilities
Required match fields	N/A
Optional match fields	• **ETH_TYPE**—Ethernet type. Value 0x800 indicates Ethernet. • **Metadata**—VNI. • **IP_DST**—Destination IP address.
Required actions	The TTP-unicast-routing flow table must support the following actions: • **Write-actions**—Writes actions in to the action set. ○ **Set L3_Unicast_Group_ID**—Specifies actions in a Layer 3 unicast group entry ID. For more information about Layer 3 unicast group entries, see "Capabilities supported by the TTP-L3-unicast-group flow table." ○ **Set L3_ECMP_Group_ID**—Specifies actions in a Layer 3 ECMP group entry ID. For more information about Layer 3 ECMP group entries, see "Capabilities supported by the TTP-L3-ECMP-group table." • **Set TTL**—Decreases the TTL by 1. • **Goto-table-60**—Directs the packet to table 60 (TTP-ingress-ACL flow table).
Optional actions	N/A

Table 16 Capabilities supported by TTP-unicast-routing flow entries (for IPv4 unicast packets with special destination IP addresses)

Item	Capabilities
Required match fields	N/A
Optional match fields	• **ETH_TYPE**—Ethernet type. Value 0x800 indicates Ethernet. • **Metadata**—VNI. • **IP_DST**—Destination IP address (a subnet or an IP address of the switch).

Required actions	**Controller**—Sends the packet to the controller.
Optional actions	N/A

Capabilities supported by the TTP-bridging flow table

A TTP-bridging flow entry is used to forward a Layer 2 packet, determine the type of the output port based on the destination MAC address of the packet, and perform one of the following actions:

- If the output port is a local physical interface, the switch tags or untags the packet before forwarding.
- If the output port is a VXLAN tunnel interface, the switch sets the VNI for the packet and untags the packet before forwarding.

Table 17 Capabilities supported by TTP-bridging flow entries

Item	Capabilities
Required match fields	N/A
Optional match fields	• **ETH_DST**—Destination MAC address. • **Metadata**—VNI.
Required actions	• **Write-actions**—Writes actions in to the action set by using the Set L2_Interface_Group_ID field. For more information about the action set, see "Capabilities supported by the TTP-L2-interface-group flow table." • **Goto-table-60**—Directs the packet to table 60 (TTP-ingress-ACL flow table).
Optional actions	N/A

Capabilities supported by the TTP-ingress-ACL flow table

The TTP-ingress-ACL flow entry is used to rate limit, redirect, mirror, or discard matching packets in the inbound direction or send the packets to the controller.

The TTP-ingress-ACL flow table does not contain a table-miss entry and executes all actions in the action set on a packet that does not match any flow entry.

Table 18 Capabilities supported by TTP-ingress-ACL flow entries (for LLDP and ARP packets)

Item	Capabilities
Required match fields	N/A
Optional match fields	**ETH_TYPE**—Ethernet type.
Required actions	**Controller**—Sends the packet to the controller.
Optional actions	N/A

Table 19 Capabilities supported by TTP-ingress-ACL flow entries (for DHCP packets)

Item	Capabilities
Required match fields	N/A
Optional match fields	• **ETH_TYPE**—Ethernet type. • **IP_PROTO**—IP protocol type. • **UDP_SRC**—Source UDP port number. • **UDP_DST**—Destination UDP port number.

Required actions	**Controller**—Sends the packet to the controller.
Optional actions	N/A

Table 20 Capabilities supported by TTP-ingress-ACL flow entries (for all packets)

Item	Capabilities
Required match fields	N/A
Optional match fields	• **In-Port**—Input port. • **Metadata**—VNI or VRF. • **IPv4SRC**—Source IPv4 address. • **IPv4DST**—Destination IPv4 address. • **IP_PROTO**—IP protocol type. • **TCP_SRC/UDP_SRC**—Source TCP or UDP port number. • **TCP_DST/UDP_DST**—Destination TCP or UDP port number.
Required actions	**A TTP-ingress-ACL flow entry must contain one of the following actions:** • **Write meter-ID**—Sets a meter entry for rate limit. • **Clear-actions + Output port number**—Clears the action set and specifies an output port for packet redirection. • **Output port number**—Specifies an output port to mirror the packet. • **Clear-actions**—Clears the action set to discard the packet.
Optional actions	N/A

Capabilities supported by the TTP-egress-ACL flow table

A TTP-egress-ACL flow entry is used to rate limit, mirror, or discard matching packets in the outbound direction.

The TTP-egress-ACL flow table does not contain a table-miss entry and uses the normal forwarding process to forward a packet that does not match any flow entry.

All packets processed by the TTP-ingress-ACL flow table will be directed to the TTP-egress-ACL flow table for processing by default.

Table 21 Capabilities supported by TTP-egress-ACL flow entries

Item	Capabilities
Required match fields	N/A
Optional match fields	• **Metadata**—VNI. • **IPv4SRC**—Source IPv4 address. • **IPv4DST**—Destination IPv4 address. • **IP_PROTO**—IP protocol type. • **TCP_SRC/UDP_SRC**—Source TCP or UDP port number. • **TCP_DST/UDP_DST**—Destination TCP or UDP port number.
Required actions	**A TTP-egress-ACL flow entry must contain one of the following actions:** • **Write meter-ID**—Sets a meter table for rate limit. • **Output port number**—Specifies an output port to mirror the packet. • **Clear-actions**—Clears the action set to discard the packet.
Optional actions	N/A

Capabilities supported by the TTP-L2-interface-group flow table

A TTP-L2-interface-group flow entry is used to tag or untag a packet and set the VXLAN ID of the packet.

For a TTP-L2-interface-group flow entry, the group type is indirect and the group entry contains only one action bucket.

Table 22 Capabilities supported by TTP-L2-interface-group flow entries (for VXLAN interfaces)

Item	Capabilities
Required actions	• **Set Tunnel_ID**—Sets the VNI. • **POP VLAN**—Untags the packet. Do not execute this action on an untagged packet. • **Output-port-number**—Sets the output port to the VXLAN tunnel interface.
Optional actions	N/A

Table 23 Capabilities supported by TTP-L2-interface-group flow entries (for untagging physical interfaces)

Item	Capabilities
Required actions	• **POP VLAN**—Untags the packet. Do not execute this action on an untagged packet. • **Output-port-number**—Sets the output port.
Optional actions	N/A

Table 24 Capabilities supported by TTP-L2-interface-group flow entries (for tagging physical interfaces)

Item	Capabilities
Required actions	• **PUSH VLAN**—Tags the packet with a VLAN ID. • **POP VLAN**—Untags the packet. Do not execute this action on an untagged packet. • **Output-port-number**—Sets the output port.
Optional actions	N/A

Capabilities supported by the TTP-L3-unicast-group flow table

A TTP-L3-unicast-group flow table is used to perform the following actions on a packet:

- Set the destination MAC address and source MAC address.
- Specify a TTP-L2-interface-group entry.

In a TTP-L3-unicast-group flow entry, the group type is indirect and the group entry contains only one action bucket.

Table 25 Capabilities supported by TTP-L3-unicast-group flow entries

Item	Capabilities
Required actions	• **Set ETH_DST**—Sets the destination MAC address. • **Set ETH_SRC**—Sets the source MAC address. • **Set L2_Interface_Group_ID**—Specifies a TTP-L2-interface-group flow table.
Optional actions	N/A

Capabilities supported by the TTP-L3-ECMP-group table

A TTP-L3-ECMP-group table is used to direct a packet to multiple TTP-L3-unicast-group entries for processing.

For a TTP-L3-ECMP-group entry, the group type is select, and the group entry contains multiple action buckets.

Table 26 Capabilities supported by TTP-L3-ECMP-group flow entries

Item	Capabilities
Required actions	• Bucket 1:{Set L3_Unicast_Group_ID} • Bucket 2:{Set L3_Unicast_Group_ID} • … • Bucket n:{Set L3_Unicast_Group_ID}
Optional actions	N/A

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

VXLAN Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the VXLAN fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

VXLAN overview 1
VXLAN benefits 1
VXLAN network model 1
VXLAN packet format 2
VXLAN working mechanisms 3
Generic VXLAN network establishment and forwarding process 3
VXLAN tunnel establishment and assignment 3
Assignment of traffic to VXLANs 3
MAC learning 4
Unicast forwarding 5
Flood 6
Access modes of VSIs 8
ARP and ND flood suppression 9
VXLAN IP gateways 10
VXLAN M-LAG 10
Support of multicast-mode VXLAN for M-LAG 11
Protocols and standards 13
Configuring basic VXLAN features 14
Restrictions: System operating mode 14
VXLAN tasks at a glance 14
Prerequisites for VXLAN 14
Creating a VXLAN on a VSI 15
Configuring a VXLAN tunnel 15
Manually creating a VXLAN tunnel 15
Enabling BFD on a VXLAN tunnel 17
Manually assigning VXLAN tunnels to a VXLAN 17
Assigning customer frames to a VSI 18
Restrictions and guidelines for configuring traffic assignment methods 18
Mapping a static Ethernet service instance to a VSI 18
Mapping dynamic Ethernet service instances to VSIs 19
Configuring VLAN-based VXLAN assignment 21
Managing MAC address entries 21
About MAC address entry management 21
Configuring static MAC address entries 22
Disabling remote-MAC address learning 22
Enabling local-MAC logging 22
Enabling software-based MAC learning on an interface 23
Configuring a multicast-mode VXLAN 24
About multicast methods for multicast-mode VXLANs 24
Prerequisites for multicast-mode VXLANs 24
Configuring a multicast-mode VXLAN that uses the PIM method 24
Configuring a multicast-mode VXLAN that uses the IGMP host method 25
Configuring private VSI 25
About private VSI 25
Restrictions and guidelines 27
Configuring a primary VSI 27
Configuring a secondary VSI 27
Setting the destination UDP port number of VXLAN packets 27
Configuring VXLAN packet check 28
Configuring the DSCP values in the outer headers of VXLAN tunnel packets 28
Disabling flooding for a VSI 29
Enabling ARP flood suppression 29
Enabling ND flood suppression 30
Enabling VXLAN packet statistics 30
Enabling packet statistics for a VSI 30
Enabling packet statistics for an AC 31

Setting the L2VPN statistics collection interval ····31
Enabling overlay OAM ····32
Pinging an IPv4 VXLAN tunnel destination ····32
Pinging an IPv6 VXLAN tunnel destination ····33
Tracing the path to an IPv4 VXLAN tunnel destination ····34
Tracing the path to an IPv6 VXLAN tunnel destination ····35
Display and maintenance commands for VXLANs ····36
VXLAN configuration examples ····37
Example: Configuring a unicast-mode VXLAN ····37
Example: Configuring a multicast-mode VXLAN ····41
Example: Configuring a multicast-mode VXLAN to support M-LAG on a public network in M-LAG access mode ····48
Configuring VXLAN IP gateways ····59
About VXLAN IP gateways ····59
VXLAN IP gateways separated from VTEPs ····59
Centralized VXLAN IP gateway deployment ····60
Distributed VXLAN IP gateway deployment ····61
VXLAN IP gateway tasks at a glance ····66
Prerequisites for VXLAN IP gateway configuration ····66
Configuring a centralized VXLAN IP gateway ····67
Restrictions and guidelines ····67
Configuring a gateway interface on a centralized VXLAN IP gateway ····67
Assigning a subnet to a VSI ····67
Configuring a distributed VXLAN IP gateway ····68
Restrictions and guidelines for distributed VXLAN IP gateway configuration ····68
Configuring a gateway interface on a distributed VXLAN IP gateway ····68
Enabling dynamic ARP or ND entry synchronization for distributed VXLAN IP gateways ····69
Assigning a subnet to a VSI ····70
Managing ARP entries and ND entries ····70
Adding a static ARP entry ····70
Disabling remote ARP or ND learning for VXLANs ····71
Configuring a VSI interface ····71
Configuring optional parameters for a VSI interface ····71
Restoring the default settings of the VSI interface ····72
Configuring VXLAN M-LAG ····72
Configuring VXLAN M-LAG ····72
Enabling an interface to join a multicast group ····74
Display and maintenance commands for VXLAN IP gateways ····75
VXLAN IP gateway configuration examples ····76
Example: Configuring a centralized VXLAN IP gateway ····76
Example: Configuring distributed VXLAN IP gateways (IPv4 underlay) ····81
Example: Configuring distributed VXLAN IP gateways (IPv6 underlay) ····92
Example: Configuring VXLAN M-LAG with a direct peer link ····101
Configuring the VTEP as an OVSDB VTEP ····113
About OVSDB VTEP ····113
Working mechanisms ····113
Protocols and standards ····113
Restrictions and guidelines: OVSDB VTEP configuration ····113
OVSDB VTEP tasks at a glance ····113
Prerequisites for OVSDB VTEP configuration ····114
Setting up an OVSDB connection to a controller ····114
About OVSDB connection types ····114
Restrictions and guidelines for OVSDB controller connection setup ····114
Prerequisites for OVSDB controller connection setup ····114
Configuring active SSL connection settings ····115
Configuring passive SSL connection settings ····115
Configuring active TCP connection settings ····115
Configuring passive TCP connection settings ····116
Enabling the OVSDB server ····116
Enabling the OVSDB VTEP service ····116

Specifying a global source address for VXLAN tunnels ····· 116
Specifying a VTEP access port ····· 117
Disabling the ACLs issued by the OVSDB controller ····· 117
OVSDB VTEP configuration examples ····· 117
Example: Configuring a unicast-mode VXLAN ····· 117

VXLAN overview

Virtual eXtensible LAN (VXLAN) is a MAC-in-UDP technology that provides Layer 2 connectivity between distant network sites across an IP network. VXLAN is typically used in data centers and the access layer of campus networks for multitenant services.

VXLAN benefits

VXLAN provides the following benefits:

- **Support for more virtual switched domains than VLANs**—Each VXLAN is uniquely identified by a 24-bit VXLAN ID. The total number of VXLANs can reach 16777216 (2^{24}). This specification makes VXLAN a better choice than 802.1Q VLAN to isolate traffic for user terminals.
- **Easy deployment and maintenance**—VXLAN requires deployment only on the edge devices of the transport network. Devices in the transport network perform typical Layer 3 forwarding.

VXLAN network model

As shown in Figure 1, a VXLAN is a virtual Layer 2 network (known as the overlay network) built on top of an existing physical Layer 3 network (known as the underlay network). The overlay network encapsulates inter-site Layer 2 frames into VXLAN packets and forwards the packets to the destination along the Layer 3 forwarding paths provided by the underlay network. The underlay network is transparent to tenants, and geographically dispersed sites of a tenant are merged into a Layer 2 network.

The site networks and the undelay network can be IPv4 or IPv6 networks.

The transport edge devices assign user terminals to different VXLANs, and then forward traffic between sites for user terminals by using VXLAN tunnels. Supported user terminals include PCs, wireless terminals, and VMs on servers.

NOTE:

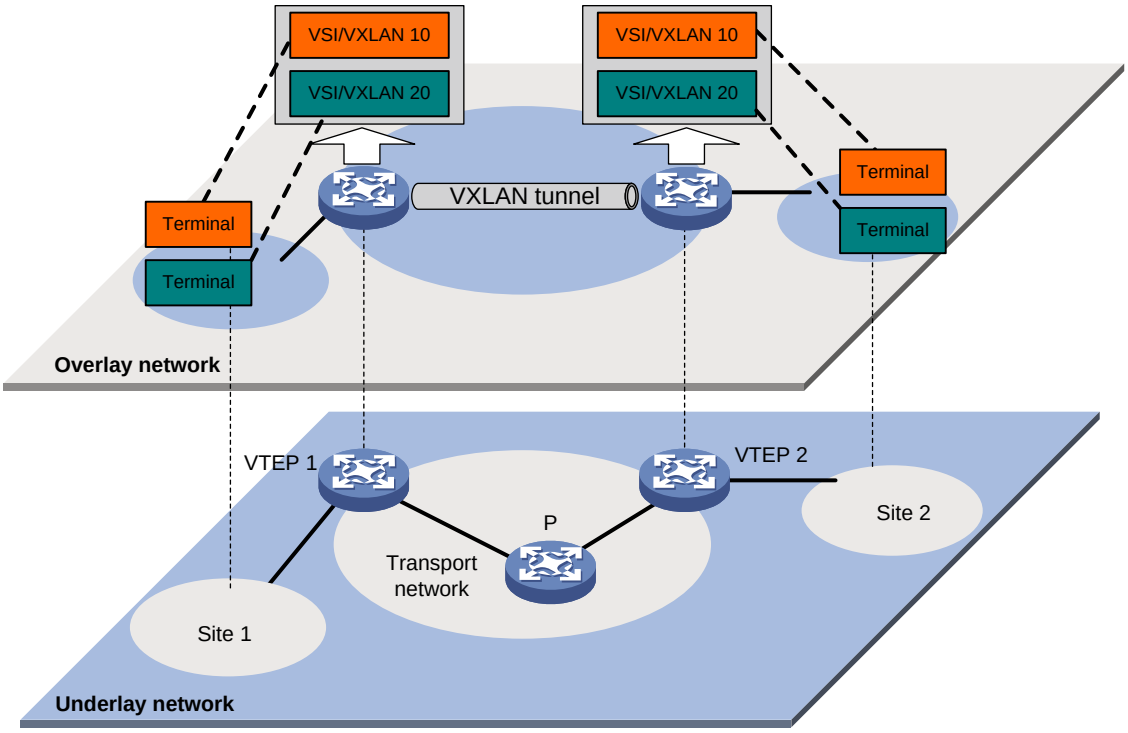
This document uses VMs as examples to describe the mechanisms of VXLAN. The mechanisms do not differ between different kinds of user terminals.

The transport edge devices are VXLAN tunnel endpoints (VTEP). The VTEP implementation of the device uses ACs, VSIs, and VXLAN tunnels to provide VXLAN services.

- **VSI**—A virtual switch instance is a virtual Layer 2 switched domain. Each VSI provides switching services only for one VXLAN. VSIs learn MAC addresses and forward frames independently of one another. VMs in different sites have Layer 2 connectivity if they are in the same VXLAN.
- **Attachment circuit (AC)**—An AC is a physical or virtual link that connects a VTEP to a local site. Typically, ACs are Ethernet service instances that are associated with the VSI of a VXLAN. Traffic received from an AC is assigned to the VSI associated with the AC. Ethernet service instances are created on site-facing Layer 2 interfaces. An Ethernet service instance matches a list of custom VLANs by using a frame match criterion.
- **VXLAN tunnel**—Logical point-to-point tunnels between VTEPs over the transport network. Each VXLAN tunnel can trunk multiple VXLANs.

VTEPs encapsulate VXLAN traffic in the VXLAN, outer UDP, and outer IP headers. The devices in the transport network forward VXLAN traffic only based on the outer IP header.

Figure 1 VXLAN network model

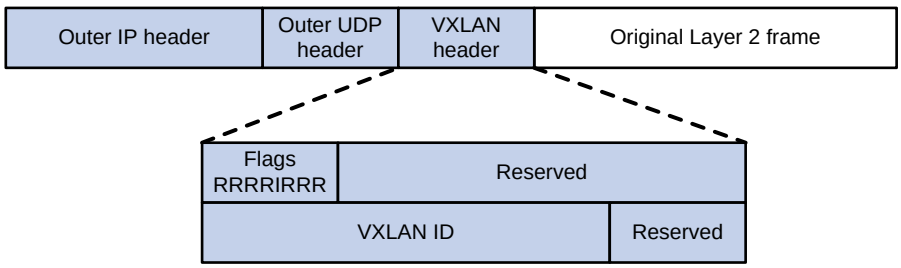


VXLAN packet format

As shown in Figure 2, a VTEP encapsulates a frame in the following headers:

- **8-byte VXLAN header**—VXLAN information for the frame.
  - **Flags**—If the I bit is 1, the VXLAN ID is valid. If the I bit is 0, the VXLAN ID is invalid. All other bits are reserved and set to 0.
  - **24-bit VXLAN ID**—Identifies the VXLAN of the frame. It is also called the virtual network identifier (VNI).
- **8-byte outer UDP header for VXLAN**—The default VXLAN destination UDP port number is 4789.
- **20-byte outer IP header**—Valid addresses of VTEPs or VXLAN multicast groups on the transport network. Devices in the transport network forward VXLAN packets based on the outer IP header.

Figure 2 VXLAN packet format



VXLAN working mechanisms

Generic VXLAN network establishment and forwarding process

The VTEP uses the following process to establish the VXLAN network and forward an inter-site frame:

1. Discovers remote VTEPs, establishes VXLAN tunnels, and assigns the VXLAN tunnels to VXLANs.
2. Assigns the frame to its matching VXLAN if the frame is sent between sites.
3. Performs MAC learning on the VXLAN's VSI.
4. Forwards the frame through VXLAN tunnels.

This section describes this process in detail. For intra-site frames in a VSI, the system performs typical Layer 2 forwarding, and it processes 802.1Q VLAN tags as described in "Access modes of VSIs."

VXLAN tunnel establishment and assignment

To provide Layer 2 connectivity for a VXLAN between two sites, you must create a VXLAN tunnel between the sites and assign the tunnel to the VXLAN.

VXLAN tunnel establishment

VXLAN supports manual and automatic VXLAN tunnel establishment.

- **Manual creation**—Manually create a VXLAN tunnel interface, and specify the tunnel source and destination IP addresses on the peer VTEPs.
- **Automatic creation**—Configure Ethernet Virtual Private Network (EVPN) to automatically discover VTEPs and set up VXLAN tunnels. For more information about EVPN, see *EVPN Configuration Guide*.

VXLAN tunnel assignment

VXLAN supports manual and automatic VXLAN tunnel assignment.

- **Manual assignment**—Manually assign VXLAN tunnels to VXLANs.
- **Automatic assignment**—Run EVPN to automatically assign VXLAN tunnels to VXLANs. For more information about EVPN, see *EVPN Configuration Guide*.

Assignment of traffic to VXLANs

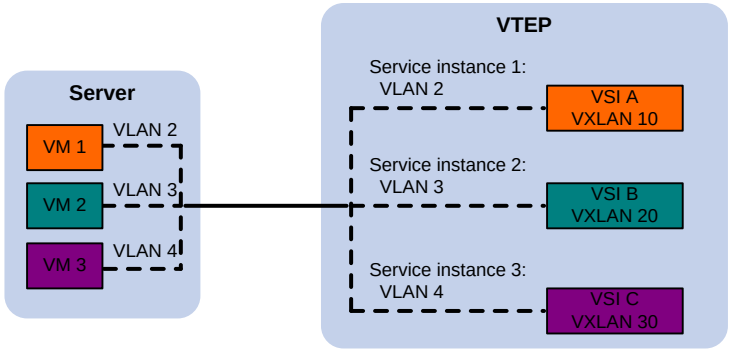
Traffic from the local site to a remote site

The VTEP uses the following methods to assign customer frames to a VXLAN:

- **Ethernet service instance-to-VSI mapping**—This method uses the frame match criterion of an Ethernet service instance to match a list of VLANs on a site-facing Layer 2 interface. The frame match criterion specifies the characteristics of traffic from the VLANs, such as tagging status and VLAN IDs. The VTEP assigns customer traffic to a VXLAN by mapping the Ethernet service instance to a VSI.
- **VLAN-based VXLAN assignment**—This method maps a VLAN to a VXLAN. The VTEP assigns all frames of the VLAN to the VXLAN.

As shown in Figure 3, Ethernet service instance 1 matches VLAN 2 and is mapped to VSI A (VXLAN 10). When a frame from VLAN 2 arrives, the VTEP assigns the frame to VXLAN 10, and looks up VSI A's MAC address table for the outgoing interface.

Figure 3 Identifying traffic from the local site



Traffic from a remote site to the local site

When a frame arrives at a VXLAN tunnel, the VTEP uses the VXLAN ID in the frame to identify its VXLAN.

MAC learning

The VTEP performs source MAC learning on the VSI as a Layer 2 switch.

- For traffic from the local site to the remote site, the VTEP learns the source MAC address before VXLAN encapsulation.
- For traffic from the remote site to the local site, the VTEP learns the source MAC address after removing the VXLAN header.

A VSI's MAC address table includes the following types of MAC address entries:

- **Local MAC**—MAC entries learned from the local site. The outgoing interfaces for the MAC address entries are site-facing interfaces.
  - **Static**—Manually added MAC entries.
  - **Dynamic**—Dynamically learned MAC entries.
- **Remote MAC**—MAC entries learned from a remote site, including static and dynamic MAC entries. The outgoing interfaces for the MAC addresses are VXLAN tunnel interfaces.
  - **Static**—Manually added MAC entries.
  - **Dynamic**—MAC entries learned in the data plane from incoming traffic on VXLAN tunnels. The learned MAC addresses are contained in the inner Ethernet header.
  - **BGP EVPN**—MAC entries advertised through BGP EVPN. For more information, see *EVPN Configuration Guide*.
  - **OpenFlow**—MAC entries issued by a remote controller through OpenFlow. For more information, see *OpenFlow Configuration Guide*.
  - **OVSDB**—MAC entries issued by a remote controller through OVSDB.

  The following shows the priority order of different types of remote MAC address entries:

  a. Static MAC address entries, and MAC address entries issued by a remote controller through OpenFlow or OVSDB. These types of entries have the same priority and overwrite each other.

  b. MAC address entries advertised through BGP EVPN.

  c. Dynamic MAC address entries.

Unicast forwarding

Intra-site unicast forwarding

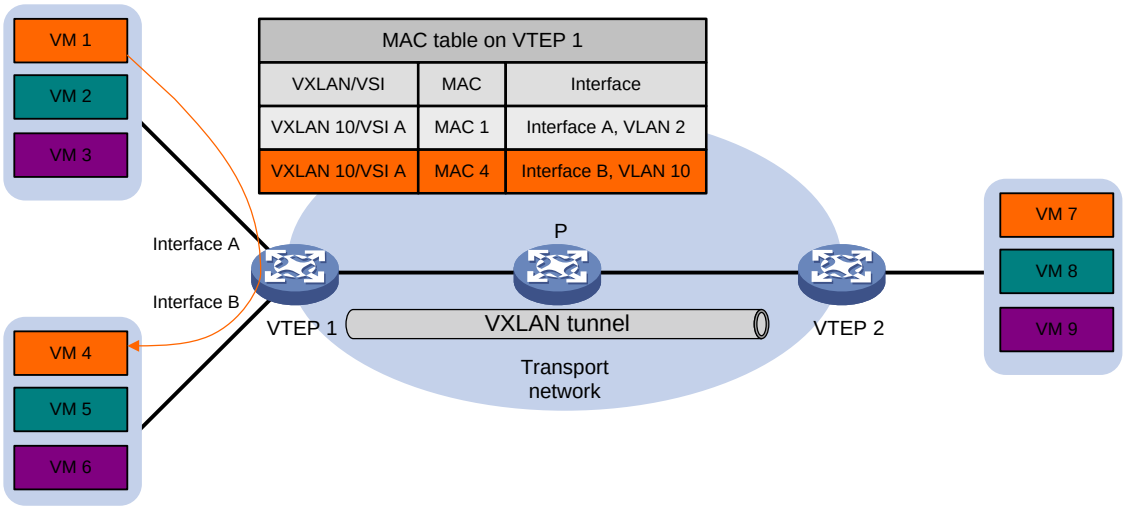
The VTEP uses the following process to forward a known unicast frame within a site:

1. Identifies the VSI of the frame.
2. Looks up the destination MAC address in the VSI's MAC address table for the outgoing interface.
3. Sends the frame out of the matching outgoing interface.

As shown in Figure 4, VTEP 1 forwards a frame from VM 1 to VM 4 within the local site in VLAN 10 as follows:

4. Identifies that the frame belongs to VSI A when the frame arrives at Interface A.
5. Looks up the destination MAC address (MAC 4) in the MAC address table of VSI A for the outgoing interface.
6. Sends the frame out of the matching outgoing interface (Interface B) to VM 4 in VLAN 10.

Figure 4 Intra-site unicast



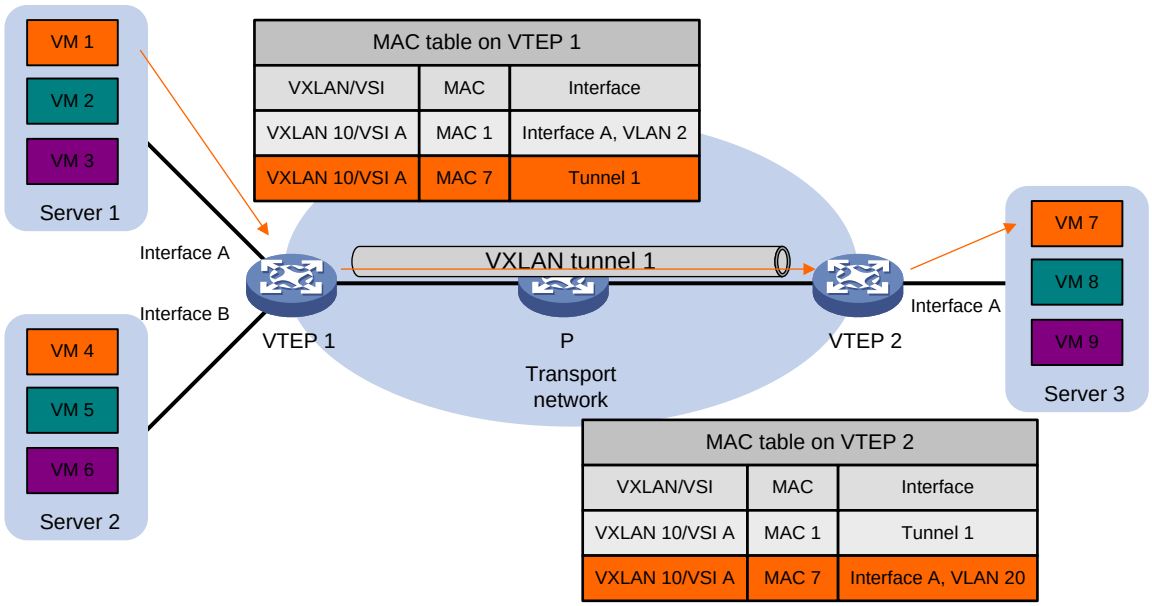
Inter-site unicast forwarding

The following process (see Figure 5) applies to a known unicast frame between sites:

1. The source VTEP encapsulates the Ethernet frame in the VXLAN/UDP/IP header.

   In the outer IP header, the source IP address is the source VTEP's VXLAN tunnel source IP address. The destination IP address is the VXLAN tunnel destination IP address.
2. The source VTEP forwards the encapsulated packet out of the outgoing VXLAN tunnel interface found in the VSI's MAC address table.
3. The intermediate transport devices (P devices) forward the frame to the destination VTEP by using the outer IP header.
4. The destination VTEP removes the headers on top of the inner Ethernet frame. It then performs MAC address table lookup in the VXLAN's VSI to forward the frame out of the matching outgoing interface.

Figure 5 Inter-site unicast



Flood

IMPORTANT:

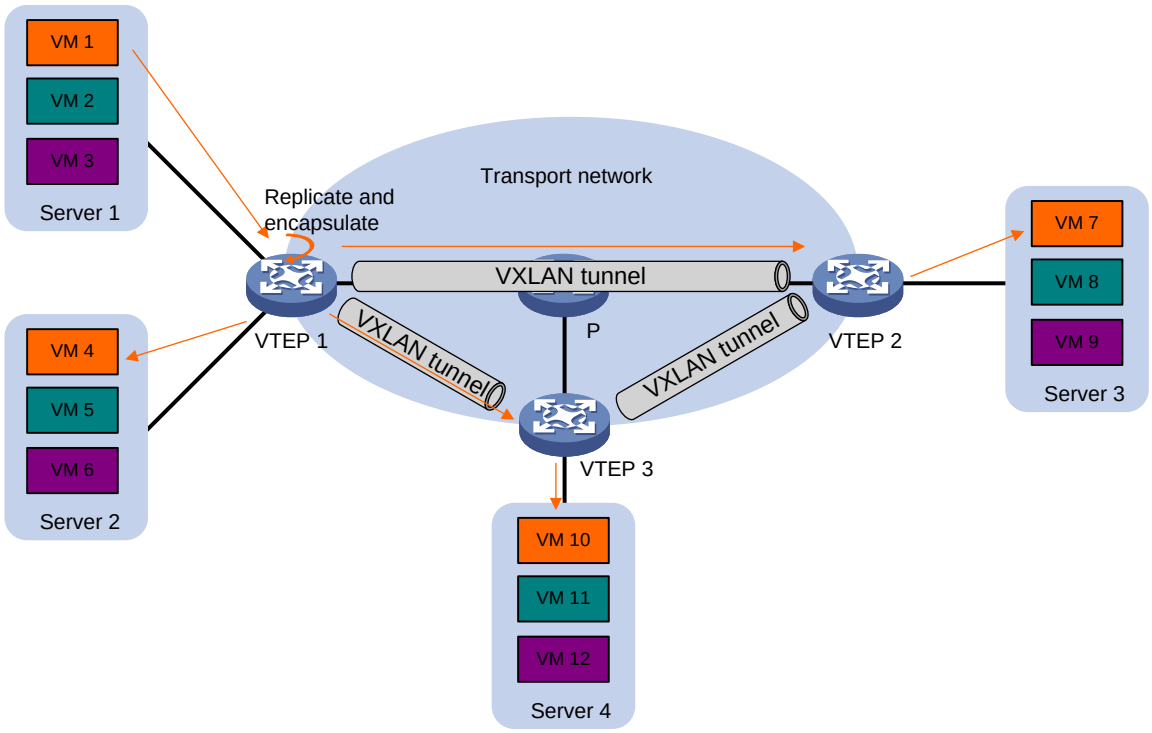
Execute the **`display device`** command with the **`verbose`** keyword to view detailed hardware information of the device. If the **Switch chip version** field displays **000**, the device does not support multicast mode (tandem replication). If the field shows **001**, the device supports multicast mode (tandem replication).

The source VTEP floods a broadcast, multicast, or unknown unicast frame to all site-facing interfaces and VXLAN tunnels in the VXLAN, except for the incoming interface. Each destination VTEP floods the inner Ethernet frame to all site-facing interfaces in the VXLAN. To avoid loops, the destination VTEPs do not flood the frame back to VXLAN tunnels.

Unicast mode (head-end replication)

As shown in Figure 6, the source VTEP replicates the flood frame, and then sends one replica to the destination IP address of each VXLAN tunnel in the VXLAN.

Figure 6 Unicast mode



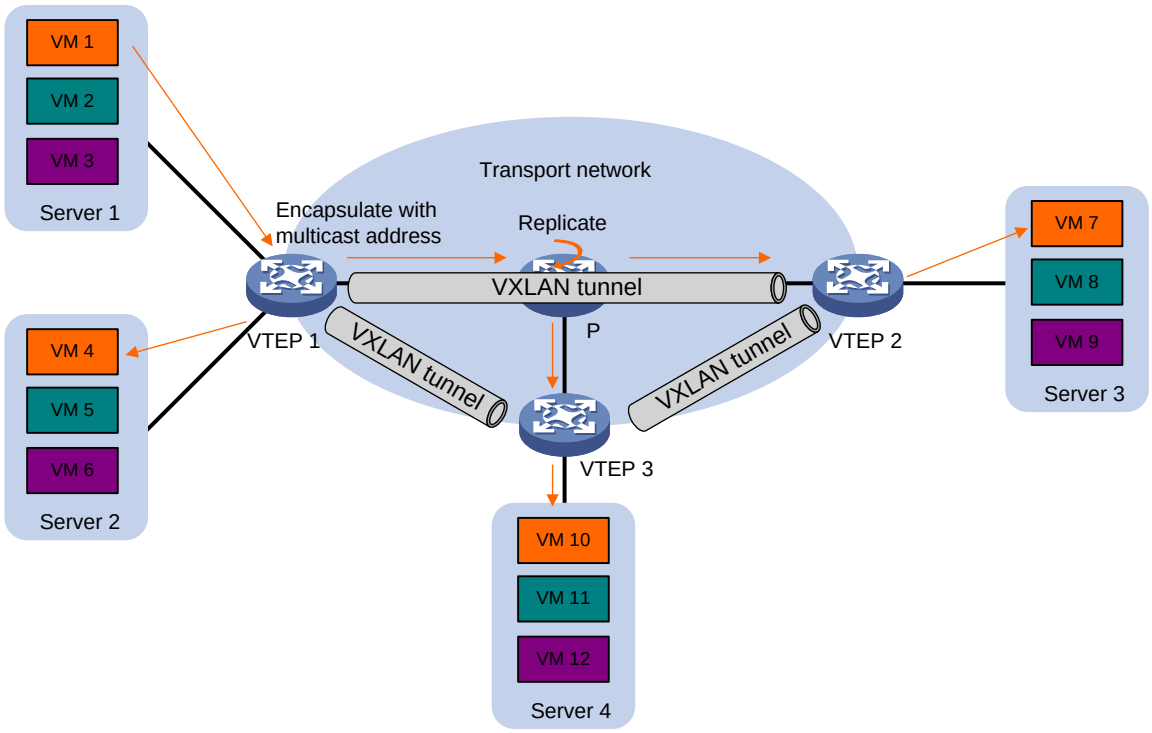
Multicast mode (tandem replication)

IMPORTANT:

Multicast mode is not supported if the underlay network is an IPv6 network.

As shown in Figure 7, the source VTEP sends the flood frame in a multicast VXLAN packet destined for a multicast group address. Transport network devices replicate and forward the packet to remote VTEPs based on their multicast forwarding entries.

Figure 7 Multicast mode



Access modes of VSIs

The access mode of a VSI determines how the VTEP processes the 802.1Q VLAN tags in the Ethernet frames.

VLAN access mode

In this mode, Ethernet frames received from or sent to the local site must contain 802.1Q VLAN tags.

- For an Ethernet frame received from the local site, the VTEP removes all its 802.1Q VLAN tags before forwarding the frame.
- For an Ethernet frame destined for the local site, the VTEP adds 802.1Q VLAN tags to the frame before forwarding the frame.

In VLAN access mode, VXLAN packets sent between sites do not contain 802.1Q VLAN tags. You can use different 802.1Q VLANs to provide the same service in different sites.

Ethernet access mode

The VTEP does not process the 802.1Q VLAN tags of Ethernet frames received from or sent to the local site.

- For an Ethernet frame received from the local site, the VTEP forwards the frame with the 802.1Q VLAN tags intact.
- For an Ethernet frame destined for the local site, the VTEP forwards the frame without adding 802.1Q VLAN tags.

In Ethernet access mode, VXLAN packets sent between VXLAN sites contain 802.1Q VLAN tags. You must use the same VLAN to provide the same service between sites.

ARP and ND flood suppression

ARP or ND flood suppression reduces ARP request broadcasts or ND request multicasts by enabling the VTEP to reply to ARP or ND requests on behalf of VMs.

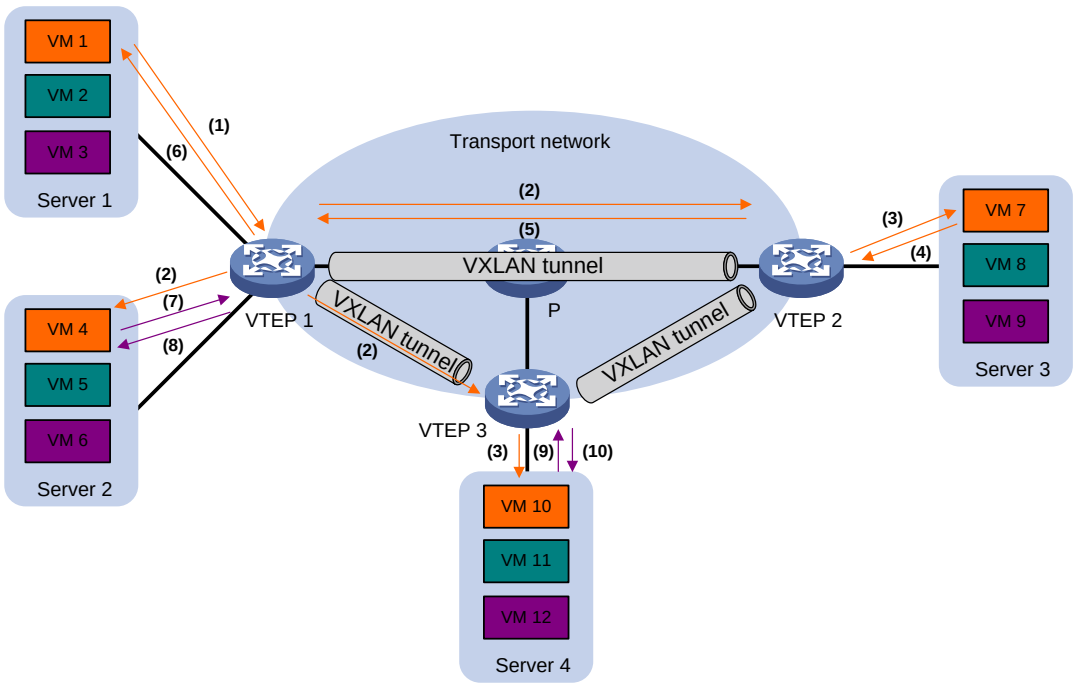
As shown in Figure 8, this feature works as follows:

1. The VTEP snoops ARP or ND packets to populate the ARP or ND flood suppression table with local and remote MAC addresses.
2. If an ARP or ND request from the local site does not have a matching entry, the VTEP drops the request or floods it over the transport network. If an ARP or ND request has a matching entry, the VTEP acts as follows:
   - In proxy reply mode, the VTEP replies to the ARP or ND request on behalf of the target IP address. In the ARP reply, the sender IP address is the target IP address in the ARP request, and the sender and source MAC addresses are the MAC address in the matching ARP flood suppression entry. In the ND reply, the sender IP address is the target IP address in the ND request, and the source MAC address is the MAC address in the matching ND flood suppression entry.
   - In unicast forwarding mode, the VTEP replaces the target MAC address and destination MAC address in the ARP request with the MAC address in the entry and unicasts the ARP request. For an ND request, the VTEP replaces the destination MAC address with the MAC address in the entry and unicasts the ND request.

Proxy reply mode decreases system resource consumption. Unicast forwarding mode reduces the ARP or ND entry learning errors caused by the delay in refreshing ARP or ND flood suppression entries at the expense of high system resource usage.

To avoid incorrect packet discards, do not enable the VTEP to discard ARP or ND requests that do not match any flood suppression entries before the VTEP creates ARP or ND flood suppression entries.

Figure 8 ARP flood suppression



The following uses ARP flood suppression that operates in proxy reply mode and floods ARP requests that do not match any ARP flood suppression entries as an example to explain the flood suppression workflow:

1. VM 1 sends an ARP request to obtain the MAC address of VM 7.
2. VTEP 1 creates a suppression entry for VM 1, and floods the ARP request in the VXLAN.
3. VTEP 2 and VTEP 3 de-encapsulate the ARP request. The VTEPs create a suppression entry for VM 1, and broadcast the request in the local site.
4. VM 7 sends an ARP reply.
5. VTEP 2 creates a suppression entry for VM 7 and forwards the ARP reply to VTEP 1.
6. VTEP 1 de-encapsulates the ARP reply, creates a suppression entry for VM 7, and forwards the ARP reply to VM 1.
7. VM 4 sends an ARP request to obtain the MAC address of VM 1 or VM 7.
8. VTEP 1 creates a suppression entry for VM 4 and replies to the ARP request.
9. VM 10 sends an ARP request to obtain the MAC address of VM 1.
10. VTEP 3 creates a suppression entry for VM 10 and replies to the ARP request.

VXLAN IP gateways

A VXLAN IP gateway provides Layer 3 forwarding services for VMs in VXLANs. A VXLAN IP gateway can be an independent device or be collocated with a VTEP. For more information about VXLAN IP gateway placement, see "Configuring VXLAN IP gateways."

VXLAN M-LAG

IMPORTANT:

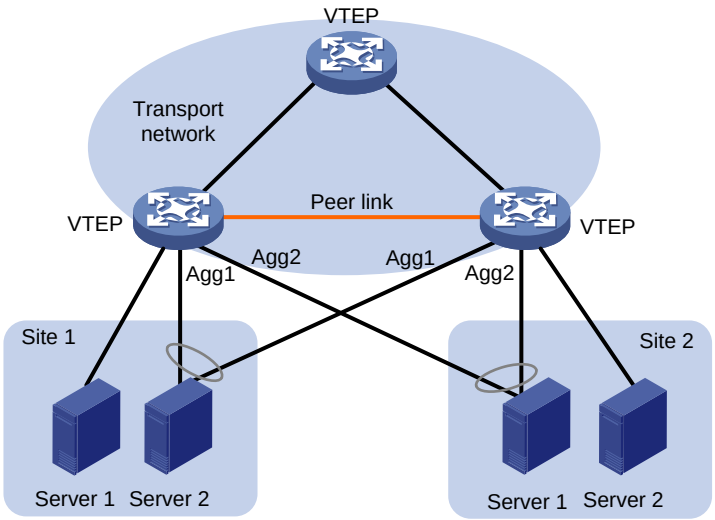
VXLAN M-LAG supports only IPv4 sites and IPv4 underlay networks.

In the current software version, tunnel peer links are not supported.

About VXLAN M-LAG

As shown in Figure 9, VXLAN multichassis link aggregation (M-LAG) virtualizes two VTEPs or VXLAN IP gateways into one M-LAG system through M-LAG to avoid single points of failure. The VTEPs or VXLAN IP gateways are called M-LAG member devices. For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*.

Figure 9 VXLAN M-LAG



VM reachability information synchronization

To ensure VM reachability information consistency in the M-LAG system, the M-LAG member devices synchronize MAC address entries and ARP information with each other through a peer link. The peer link can be an Ethernet aggregate link or a VXLAN tunnel, which are referred to as direct peer link and tunnel peer link, respectively.

IMPORTANT:

The tunnel peer link is automatically associated with all VXLANs on each M-LAG member device.

Virtual VTEP address

The M-LAG member devices use the same IP address to set up VXLAN tunnels with remote VTEPs or VXLAN IP gateways.

Dual-homed ACs

As shown in Figure 9, a VM is dual-homed to the VTEPs in an M-LAG system through Ethernet links. The Ethernet links connected to the same VM are aggregated into one Layer 2 aggregation group across VTEPs. On the corresponding Layer 2 aggregate interface, Ethernet service instances are configured as ACs of VXLANs to match customer traffic. These ACs are dual-homed ACs.

- If a direct peer link is used, the site-facing link backup mechanism is as follows:

  When a site-facing AC is configured on a VTEP, the VTEP automatically creates an AC on the peer link with the same traffic match criterion as the site-facing AC. Then, it maps the automatically created AC to the VSI of the site-facing AC. When the site-facing AC is down, traffic sent to the AC is forwarded to the other VTEP through the peer link. This mechanism ensures service continuity in case of AC failure.

- If a tunnel peer link is used, the site-facing link backup mechanism is as follows:

  If a site-facing AC on a VTEP is down, traffic sent to the AC will be encapsulated into VXLAN packets. The VXLAN ID belongs to the VXLAN that is associated with the VSI of the site-facing AC. The VTEP forwards the VXLAN packets through the tunnel peer link to the peer VTEP. The peer VTEP assigns the traffic to the correct VSI based on the VXLAN ID in the received packets.

Communication between single-homed ACs

An AC that is attached to only one of the VTEPs in an M-LAG system is called a single-homed AC. Two single-homed ACs attached to different VTEPs communicate through the peer link.

- If a direct peer link is used, the traffic forwarding mechanism is as follows:

  When a single-homed AC is configured on a VTEP, the VTEP automatically creates an AC on the peer link with the same traffic match criterion as the single-homed AC. Then, it maps the automatically created AC to the VSI of the single-homed AC. When receiving traffic from the single-homed AC, the VTEP sends the traffic to the other VTEP through the peer link. Then, the other VTEP identifies the VSI of the traffic and forwards it.

- If a tunnel peer link is used, the traffic forwarding mechanism is as follows:

  When receiving traffic from a single-homed AC, a VTEP encapsulates the traffic into VXLAN packets and sends them to the other VTEP through the peer link. The VXLAN ID in the VXLAN packets belongs to the VSI to which the single-homed AC is mapped. Then, the other VTEP identifies the VSI of the traffic and forwards it.

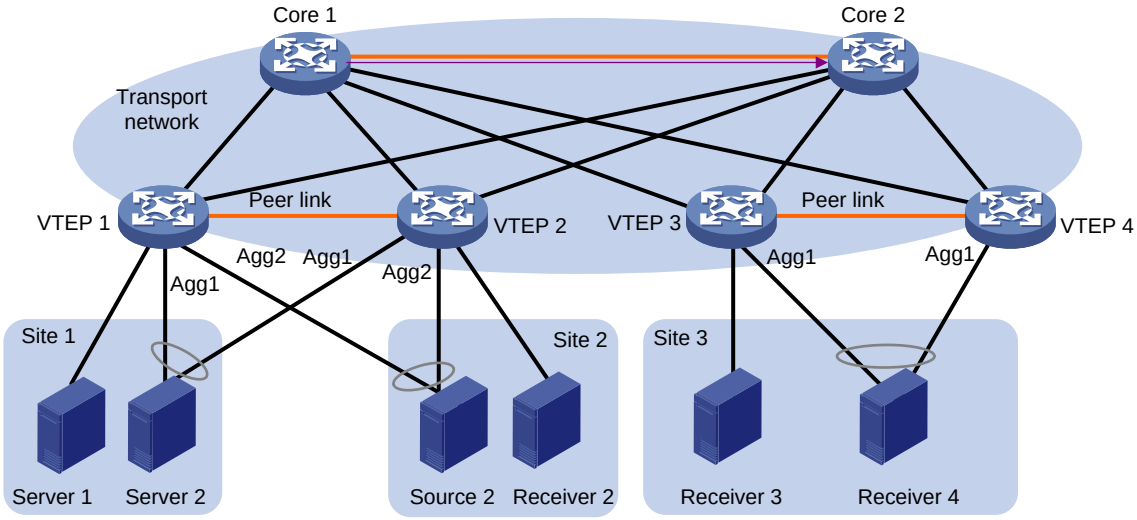
Support of multicast-mode VXLAN for M-LAG

As shown in Figure 10, when a multicast-mode VXLAN forwards flood traffic, it not only allows the VTEPs to form M-LAG systems but also enables core devices in the underlay to form M-LAG systems, enhancing network reliability.

TIP:

- In the current software version, in a network where multicast-mode VXLAN supports M-LAG, the underlay can achieve load sharing either by using M-LAG or by IP routing ECMP.
- In a network where multicast-mode VXLAN supports M-LAG, when the underlay uses M-LAG for load sharing, the VTEPs must use the virtual IP addresses of their M-LAG systems to establish VXLAN multicast tunnels.

Figure 10 Support of multicast-mode VXLAN for M-LAG



M-LAG member devices might forward the same multicast traffic simultaneously, which consumes network resources. To avoid this issue, M-LAG member devices use an odd/even address-based forwarding principle for load-sharing VXLAN packets encapsulated with multicast addresses. More specifically, one M-LAG member device forwards packets with odd-numbered multicast addresses (the destination multicast addresses encapsulated in VXLAN packets), and the other M-LAG member device forwards packets with even-numbered multicast addresses. As shown in Figure 11, VTEP 1 and Core 2 forward the traffic for odd multicast addresses, and VTEP 2 and Core 1 forward the traffic for even multicast addresses.

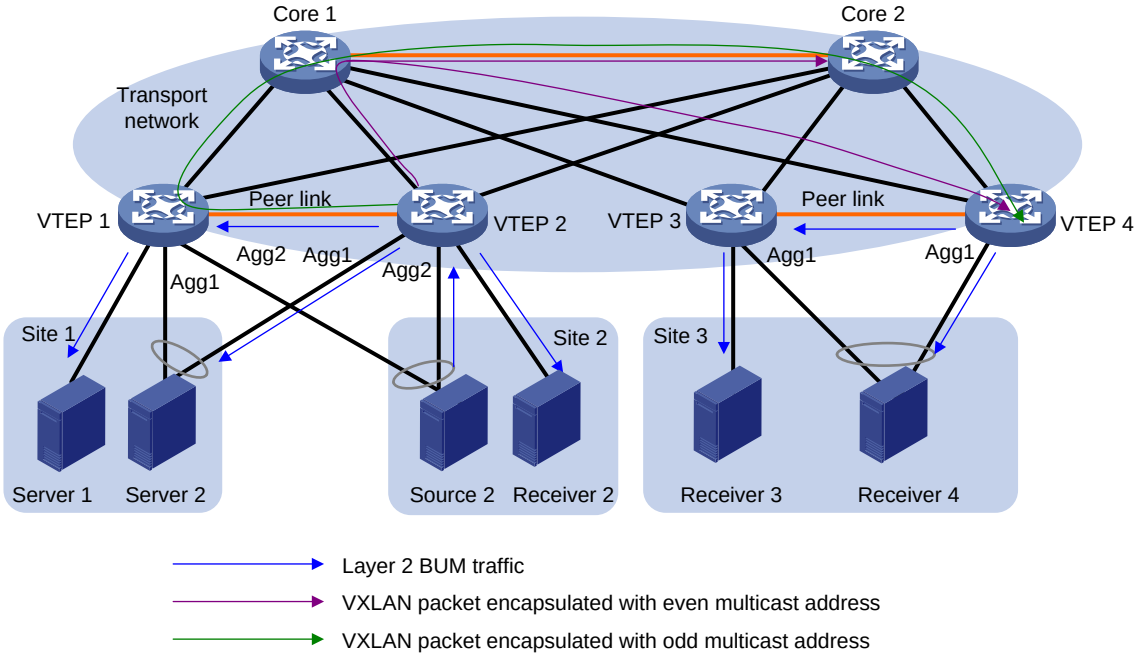
As shown in Figure 11, in the network where multicast-mode VXLAN supports M-LAG, the traffic forwarding process for flood traffic sent by devices accessing in M-LAG mode is as follows:

- **Local forwarding**—After VTEP 2 receives the flood traffic sent by a device accessing in M-LAG mode, VTEP 2 forwards the traffic to the other local devices accessing in M-LAG mode and local single-homed devices.
- **Forwarding between M-LAG member devices**—After VTEP 2 receives the flood traffic sent by a device accessing in M-LAG mode, VTEP 2 synchronizes the flood traffic to the peer M-LAG member device (VTEP 1) through the peer link. VTEP 1 forwards the flood traffic to the local single-homed devices. Traffic received from the peer link will not be forwarded to devices accessing in M-LAG mode or to VXLAN tunnels.
- **VXLAN tunnel forwarding**—After VTEP 2 receives the flood traffic sent by the device accessing in M-LAG mode, VTEP 2 adds VXLAN encapsulation to the flood traffic. The encapsulated destination address is the VXLAN flood multicast address configured by using the **group** command. Based on the odd/even address-based multicast traffic forwarding principle, the traffic is forwarded as follows:
  - If the encapsulated destination multicast address of a packet is even, VTEP 2 directly forwards the packet to Core 1 through the VXLAN tunnel, and then Core 1 forwards the packet to the remote VTEP. The remote VTEP decapsulates the packet and sends the original flood packet to the local site.

- If the encapsulated destination multicast address of a packet is odd, VTEP 2 synchronizes the encapsulated multicast VXLAN packet to VTEP 1 through the peer link. VTEP 1 forwards the packet to Core 2 through the VXLAN tunnel, and then Core 2 forwards the packet to the remote VTEP. The remote VTEP decapsulates the packet and sends the original flood packet to the local site.

The forwarding process for flood traffic sent by a single-homed access device is similar and will not be described here.

Figure 11 Traffic forwarding schematic diagram when multicast-mode VXLAN supports M-LAG



In a network where multicast-mode VXLAN supports M-LAG, follow these restrictions:

- On a leaf device, the virtual IP address of an M-LAG system cannot act as the RP address. If the RP is selected dynamically, the configured C-RP address range cannot include the virtual IP address of the M-LAG system. If you manually configure the RP, you cannot configure the virtual IP address of the M-LAG system as the static RP address.
- When you use IP routing ECMP for load sharing in an underlay, the M-LAG virtual IP address on a core device cannot act as the RP address.
- In a leaf-spine architecture, each leaf node separately selects an RPF interface. For the same multicast tunnel, it's impossible to ensure that all egress leaf nodes select the same spine as their intermediate replication node.

Protocols and standards

RFC 7348, *Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks*

Configuring basic VXLAN features

Restrictions: System operating mode

To use VXLAN, first set the system operating mode to expert by using the **system-working-mode expert** command in system view, save the running configuration, and then reboot the device. For more information about the system operating mode, see device management in *Fundamentals Configuration Guide*.

Do not configure an interface both as an AC interface and for carrying VXLAN tunnel traffic.

VXLAN tasks at a glance

To configure basic VXLAN settings, perform the following tasks on VTEPs:

1. Creating a VXLAN on a VSI
2. Configuring a VXLAN tunnel
3. Manually assigning VXLAN tunnels to a VXLAN
4. Assigning customer frames to a VSI
5. (Optional.) Managing MAC address entries
   - Configuring static MAC address entries
   - Disabling remote-MAC address learning
   - Enabling local-MAC logging
   - Enabling software-based MAC learning on an interface
6. Configuring a multicast-mode VXLAN

   If the network is multicast dense, configure the VTEP to flood VXLAN traffic in multicast mode.
7. (Optional.) Configuring private VSI

   Configure private VSI on the VTEP and private VLAN on an attached server to isolate traffic of different VMs or apply security policies.
8. (Optional.) Configuring VXLAN packet parameters
   - Setting the destination UDP port number of VXLAN packets
   - Configuring VXLAN packet check
   - Configuring the DSCP values in the outer headers of VXLAN tunnel packets
9. (Optional.) Reducing flood traffic in the transport network
   - Disabling flooding for a VSI
   - Enabling ARP flood suppression
   - Enabling ND flood suppression
10. Maintaining VXLAN networks
    - Enabling VXLAN packet statistics
    - Testing the connectivity of a VXLAN tunnel

Prerequisites for VXLAN

Configure a routing protocol on the devices in the transport network to make sure the VTEPs can reach one another.

Creating a VXLAN on a VSI

Procedure

1. Enter system view.
   system-view
2. Enable L2VPN.
   l2vpn enable
   By default, L2VPN is disabled.
3. Create a VSI and enter VSI view.
   vsi *vsi-name*
4. Enable the VSI.
   undo shutdown
   By default, a VSI is enabled.
5. Create a VXLAN and enter VXLAN view.
   vxlan *vxlan-id*
   You can create only one VXLAN on a VSI.
   The VXLAN ID must be unique for each VSI.
6. (Optional.) Configure VSI parameters:
   a. Return to VSI view.
      quit
   b. Configure a VSI description.
      description *text*
      By default, a VSI does not have a description.
   c. Set the MTU for the VSI.
      mtu *size*
      The default MTU for a VSI is 1500 bytes.
   d. Set the broadcast, multicast, or unknown unicast restraint bandwidth for the VSI.
      restrain { **broadcast** | **multicast** | **unknown-unicast** } *bandwidth*
      By default, a VSI's broadcast restraint bandwidth, unknown multicast restraint bandwidth, and unknown unicast restraint bandwidth are not set.
   e. Enable MAC address learning for the VSI.
      mac-learning enable
      By default, MAC address learning is enabled for a VSI.

Configuring a VXLAN tunnel

Manually creating a VXLAN tunnel

About this task

When you manually create a VXLAN tunnel, specify addresses on the local VTEP and the remote VTEP as the tunnel source and destination addresses, respectively.

Restrictions and guidelines

As a best practice, do not configure multiple VXLAN tunnels to use the same source and destination IP addresses.

Make sure the following VXLAN tunnels are not associated with the same VXLAN when they have the same tunnel destination IP address:

- A VXLAN tunnel automatically created by EVPN.
- A manually created VXLAN tunnel.

For more information about EVPN, see *EVPN Configuration Guide*.

This task provides basic VXLAN tunnel configuration. For more information about tunnel configuration and commands, see *Layer 3—IP Services Configuration Guide* and *Layer 3—IP Services Command Reference*.

In versions earlier than R8307P53, to ensure correct VXLAN traffic forwarding, do not use the following interfaces as the transport-facing outgoing interfaces for VXLAN tunnels:

- Layer 3 Ethernet subinterfaces.
- Layer 3 Ethernet interfaces that have subinterfaces.

In versions earlier than R8336P05, if an interface has a subinterface configured but neither the subinterface nor its main interface acts as a tunnel outgoing interface, the interface cannot decapsulate tunnel-encapsulated packets it receives.

If ECMP routes exist between the two ends of a VXLAN tunnel, the VXLAN tunnel performs load sharing only on known unicast traffic.

Procedure

1. Enter system view.

   system-view

2. (Optional.) Specify a global source IP address for VXLAN tunnels.

   tunnel global source-address { *ipv4-address* | **ipv6** *ipv6-address* }

   By default, no global source IP address is specified for VXLAN tunnels.

   A VXLAN tunnel uses the global source address if you do not specify a source interface or source address for the tunnel.

3. Create a VXLAN tunnel interface and enter tunnel interface view.

   interface tunnel *tunnel-number* **mode vxlan** [**ipv6**]

   The endpoints of a tunnel must use the same tunnel mode.

4. Specify a source address for the tunnel. Choose one of the following methods:
   - Specify a source IP address for the tunnel.

     source { *ipv4-address* | *ipv6-address* }

     The specified IP address is used in the outer IP header of tunneled VXLAN packets.
   - Specify a source interface for the tunnel.

     source *interface-type interface-number*

     The IP address of the specified interface is used in the outer IP header of tunneled VXLAN packets.

     By default, no source IP address or source interface is specified for a tunnel.

     Do not perform this step if you are using OVSDB for VXLAN tunnel management.

     Do not use a Layer 3 Ethernet subinterface to provide the tunnel source IP address. If you do so, the VXLAN tunnel cannot transmit VXLAN packets.

5. Specify a destination IP address for the tunnel.

   destination { *ipv4-address* | *ipv6-address* }

By default, no destination IP address is specified for a tunnel.

Specify the remote VTEP's IP address. This IP address will be the destination IP address in the outer IP header of tunneled VXLAN packets.

Enabling BFD on a VXLAN tunnel

About this task

Enable BFD on both ends of a VXLAN tunnel for quick link connectivity detection. The VTEPs periodically send BFD single-hop control packets to each other through the VXLAN tunnel. A VTEP sets the tunnel state to Defect if it has not received control packets from the remote end for 5 seconds. In this situation, the tunnel interface state is still Up. The tunnel state will change from Defect to Up if the VTEP can receive BFD control packets again.

Restrictions and guidelines

You must enable BFD on both ends of a VXLAN tunnel.

This feature supports only IPv4 VXLAN tunnels and does not support IPv6 VXLAN tunnels or VXLAN DCI tunnels.

Procedure

1. Enter system view.

   system-view
2. Specify the reserved VXLAN.

   reserved vxlan *vxlan-id*

   By default, no VXLAN has been reserved.

   For BFD sessions to come up, you must reserve a VXLAN.

   You can specify only one reserved VXLAN on the VTEP. The reserved VXLAN cannot be the VXLAN created on any VSI.
3. Enter VXLAN tunnel interface view.

   interface tunnel *tunnel-number*
4. Enable BFD on the tunnel.

   tunnel bfd enable destination-mac *mac-address*

   By default, BFD is disabled on a tunnel.

Manually assigning VXLAN tunnels to a VXLAN

About this task

To provide Layer 2 connectivity for a VXLAN between two sites, you must assign the VXLAN tunnel between the sites to the VXLAN.

You can assign multiple VXLAN tunnels to a VXLAN, and configure a VXLAN tunnel to trunk multiple VXLANs. For a unicast-mode VXLAN, the system floods unknown unicast, multicast, and broadcast traffic to each tunnel associated with the VXLAN.

Restrictions and guidelines

For full Layer 2 connectivity in the VXLAN, make sure the VXLAN contains the VXLAN tunnel between each pair of sites in the VXLAN.

Procedure

1. Enter system view.

   system-view

2. Enter VSI view.
   vsi *vsi-name*
3. Enter VXLAN view.
   vxlan *vxlan-id*
4. Assign VXLAN tunnels to the VXLAN.
   tunnel { *tunnel-number* [**backup-tunnel** *tunnel-number*] | **all** }
   By default, a VXLAN does not contain any VXLAN tunnels.

Parameter	Description
backup-tunnel *tunnel-number*	Specifies a backup tunnel. When the primary VXLAN tunnel is operating correctly, the backup VXLAN tunnel does not forward traffic. When the primary VXLAN tunnel goes down, traffic is switched to the backup VXLAN tunnel.

Assigning customer frames to a VSI

Restrictions and guidelines for configuring traffic assignment methods

VLAN-based VXLAN assignment is mutually exclusive with the manually created Ethernet service instances. To create these Ethernet service instances, you must first disable VLAN-based VXLAN assignment by using the **undo vxlan vlan-based** command. To enable VLAN-based VXLAN assignment, you must first delete all Ethernet service instances.

Mapping a static Ethernet service instance to a VSI

About this task

A static Ethernet service instance matches a list of VLANs on a site-facing interface. The VTEP assigns customer traffic from the VLANs to a VXLAN by mapping the Ethernet service instance to a VSI.

Restrictions and guidelines

Link aggregation group membership is mutually exclusive with Ethernet service instance-to-VSI mappings on a Layer 2 interface. Do not map a VSI to an Ethernet service instance on a Layer 2 interface if the interface is in a Layer 2 aggregation group.

If an Ethernet service instance matches both inner and outer VLAN IDs on an interface, do not configure any other Ethernet service instances to match the same outer VLAN ID on that interface.

For information about the frame match criterion configuration restrictions and guidelines of Ethernet service instances, see *VXLAN Command Reference*.

Do not configure both a PVID AC (with the PVID specified in the **encapsulation s-vid** command) and an untagged AC (**encapsulation untagged**) for Ethernet service instances on the same interface. If you do that, untagged packets might match an unexpected AC, disrupting traffic forwarding.

When you use the VLAN access mode, do not set the frame match criterion of an Ethernet service instance to **encapsulation default**.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Create an Ethernet service instance and enter Ethernet service instance view.
   service-instance *instance-id*
4. Configure a frame match criterion. Choose one of the following options:
   - Match frames tagged with the specified inner 802.1Q VLAN ID.
     encapsulation c-vid *vlan-id*
   - Match frames tagged with the specified outer 802.1Q VLAN ID.
     encapsulation s-vid *vlan-id* [**only-tagged**]
   - Match frames tagged with the specified outer and inner 802.1Q VLAN IDs.
     encapsulation s-vid *vlan-id* **c-vid** { *vlan-id* | **all** }
   - Match any untagged frames.
     encapsulation untagged
   - Match frames that do not match any other service instance on the interface.
     encapsulation default

     An interface can contain only one Ethernet service instance that uses the **encapsulation default** criterion.

     An Ethernet service instance that uses the **encapsulation default** criterion matches any frames if it is the only instance on the interface.

   By default, an Ethernet service instance does not contain a frame match criterion.
5. Map the Ethernet service instance to a VSI.
   xconnect vsi *vsi-name* [**access-mode** { **ethernet** | **vlan** }] [**track** *track-entry-number*&<1-3>]

   By default, an Ethernet service instance is not mapped to any VSI.

Mapping dynamic Ethernet service instances to VSIs

About this task

The 802.1X or MAC authentication feature can use the authorization VSI, the guest VSI, the Auth-Fail VSI, and the critical VSI to control the access of users to network resources. When assigning a user to a VSI, 802.1X or MAC authentication sends the VXLAN feature the VSI information and the user's access information, including access interface, VLAN, and MAC address. Then the VXLAN feature creates a dynamic Ethernet service instance for the user and maps it to the VSI. For more information about 802.1X authentication and MAC authentication, see *Security Configuration Guide*.

A dynamic Ethernet service instance supports the following traffic match modes:

- **VLAN-based mode**—Matches frames by VLAN ID.
- **MAC-based mode**—Matches frames by VLAN ID and source MAC address.

By default, dynamic Ethernet service instances use VLAN-based traffic match mode. To use MAC-based traffic match mode for dynamic Ethernet service instances, you must enable MAC authentication or 802.1X authentication that uses MAC-based access control.

Restrictions and guidelines for dynamic Ethernet service instance mappings

Dynamic Ethernet service instances cannot be created on member ports of a Layer 2 aggregation group.

Configuring the VLAN-based traffic match mode

To use the VLAN-based traffic match mode, configure 802.1X authentication or MAC authentication and perform one of the following tasks:

- Configure the guest VSI, Auth-Fail VSI, or critical VSI on the 802.1X- or MAC authentication-enabled interface.
- Issue an authorization VSI to an 802.1X or MAC authentication user from a remote AAA server.

Then, the device will automatically create a dynamic Ethernet service instance for the 802.1X or MAC authentication user and map the Ethernet service instance to a VSI.

For more information about configuring 802.1X authentication and MAC authentication, see *Security Configuration Guide*.

Configuring the MAC-based traffic match mode

1. Enter system view.

   `system-view`
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view**.**

     `interface` *`interface-type interface-number`*
   - Enter Layer 2 aggregate interface view.

     `interface bridge-aggregation` *`interface-number`*
3. Enable MAC-based traffic match mode for dynamic Ethernet service instances on the interface.

   `mac-based ac`

   By default, MAC-based traffic match mode is disabled for dynamic Ethernet service instances. Dynamic Ethernet service instances use VLAN-based traffic match mode.

   For more information about this command, see *VXLAN Command Reference*.
4. Enable MAC authentication or 802.1X authentication that uses MAC-based access control.

   To use the MAC-based traffic match mode, configure MAC authentication or 802.1X authentication that uses MAC-based access control and perform one of the following tasks:
   - Configure the guest VSI, Auth-Fail VSI, or critical VSI on the 802.1X- or MAC authentication-enabled interface.
   - Issue an authorization VSI to an 802.1X or MAC authentication user from a remote AAA server.

   Then, the device will automatically create a dynamic Ethernet service instance for the 802.1X or MAC authentication user and map the Ethernet service instance to a VSI.

   For more information about configuring 802.1X authentication and MAC authentication, see *Security Configuration Guide*.

Configuring VLAN-based VXLAN assignment

About this task

VLAN-based VXLAN assignment enables the device to assign all traffic of a VLAN to a VXLAN. If you enable this feature and map a VLAN to a VXLAN, the device automatically performs the following operations:

1. Creates an Ethernet service instance that uses the VLAN ID as its instance ID on each interface in the VLAN. The matching outer VLAN ID of the Ethernet service instances is the VLAN ID.
2. Maps the Ethernet service instances to the VSI of the VXLAN.

Restrictions and guidelines

Do not configure this feature together with EVPN M-LAG. For information about EVPN M-LAG, see *EVPN Configuration Guide*.

If you map a VLAN to a VXLAN, the VTEP cannot perform non-VXLAN Layer 2 forwarding in the VLAN. Also, the VLAN interface of the VLAN cannot perform Layer 3 forwarding.

The Ethernet service instance creation or deletion time is affected by the number of VLANs mapped to a VXLAN and the number of trunk ports assigned to the VLANs. The larger the numbers, the longer the time. During AC creation or deletion, other operations are queued.

Prerequisites

Use the **vxlan** command to create the VXLAN to which a VLAN is mapped.

Procedure

1. Enter system view.
   system-view
2. Enable VLAN-based VXLAN assignment.
   vxlan vlan-based
   By default, VLAN-based VXLAN assignment is disabled.
3. Create a VLAN and enter VLAN view.
   vlan *vlan-id*
   Do not specify VLAN 1 for VLAN-based VXLAN assignment.
4. Map the VLAN to a VXLAN.
   vxlan vni *vxlan-id*
   By default, a VLAN is not mapped to a VXLAN.
   Do not map a VLAN to the L3 VXLAN ID of EVPN.

Managing MAC address entries

About MAC address entry management

Local-MAC address entries are dynamically learned. You can log local MAC addresses and local-MAC changes.

Remote-MAC address entries have a variety of types, including manually added entries and dynamically learned entries.

Configuring static MAC address entries

Restrictions and guidelines

Do not configure static remote-MAC entries for VXLAN tunnels that are automatically established by using EVPN.

- ENDP or EVPN re-establishes VXLAN tunnels if the transport-facing interface goes down and then comes up. If you have configured static remote-MAC entries, the entries are deleted when the tunnels are re-established.
- ENDP or EVPN re-establishes VXLAN tunnels if you perform configuration rollback. If the tunnel IDs change during tunnel re-establishment, configuration rollback fails, and static remote-MAC entries on the tunnels cannot be restored.

  For more information about EVPN, see *EVPN Configuration Guide*.

Procedure

1. Enter system view.

   system-view
2. Add a static remote-MAC address entry.

   mac-address static *mac-address* **interface tunnel** *tunnel-number* **vsi** *vsi-name*

   For the setting to take effect, make sure the VSI's VXLAN has been specified on the VXLAN tunnel.

Disabling remote-MAC address learning

About this task

When network attacks occur, disable remote-MAC address learning to prevent the device from learning incorrect remote MAC addresses. You can manually add static remote-MAC address entries.

Restrictions and guidelines

This feature takes effect only on manually created VXLAN tunnels. Remote-MAC address learning is always disabled for automatically created tunnels, and it is not controlled by this command.

Procedure

1. Enter system view.

   system-view
2. Disable remote-MAC address learning.

   vxlan tunnel mac-learning disable

   By default, remote-MAC address learning is enabled.

Enabling local-MAC logging

About this task

When the local-MAC logging feature is enabled, the VXLAN module immediately sends a log message with its local MAC addresses to the information center. When a local MAC address is added or removed, a log message is also sent to the information center to report the local-MAC change.

With the information center, you can set log message filtering and output rules, including output destinations. For more information about configuring the information center, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.
   system-view
2. Enable local-MAC logging.
   vxlan local-mac report
   By default, local-MAC logging is disabled.

Enabling software-based MAC learning on an interface

About this task

This feature is applicable to SDN networks.

To reduce broadcast traffic in an SDN network, the controller synchronizes the MAC addresses that each VTEP learns among all VTEPs. On a VTEP, an interface can learn MAC addresses in hardware or software.

- In hardware-based learning mode, the software periodically obtains new MAC addresses from the hardware and advertises the MAC addresses to the controller.
- In software-based learning mode, the software instantly issues new MAC addresses to the hardware and the controller as soon as they are learned.

Software-based MAC learning shortens the interval at which the VTEP advertises MAC address reachability information to the controller. However, this mode is resource intensive. When you use this mode, you must fully understand its impact on the device performance.

Restrictions and guidelines

Software-based MAC learning consumes more resources than the hardware learning method. As a best practice to ensure device performance, do not enable software-based MAC learning if MAC addresses change frequently in the network.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Enable software-based MAC learning on the interface.
   l2vpn mac-address software-learning enable
   By default, hardware-based MAC learning is used.
4. (Optional.) Set the MAC learning limit on the interface.
   mac-address max-mac-count *count*
   By default, the MAC learning limit is not set on an interface.
   For more information about this command, see MAC address table commands in *Layer 2—LAN Switching Command Reference*.

Configuring a multicast-mode VXLAN

About multicast methods for multicast-mode VXLANs

A multicast-mode VXLAN supports the following multicast methods:

- **PIM**—VTEPs and transport network devices run PIM to generate multicast forwarding entries. To forward multicast traffic correctly, you must use the source IP address of an up VXLAN tunnel as the source IP address of multicast VXLAN packets. As a best practice, use the source IP address of a VXLAN tunnel that uses the IP address of a loopback interface. If the VTEP has multiple transport-facing interfaces, PIM dynamically selects the outgoing interfaces for multicast VXLAN packets.
- **IGMP host**—VTEPs and transport network devices run PIM and IGMP to generate multicast forwarding entries.
  - Transport-facing interfaces of VTEPs act as IGMP hosts.
  - Transport network devices connected to a VTEP run IGMP.
  - All transport network devices run PIM.

  On a VTEP, you must use the IP address of the transport-facing interface as the source IP address for multicast VXLAN packets. If the VTEP has multiple transport-facing interfaces, multicast VXLAN packets are sent to the transport network through the interface that provides the source IP address for multicast VXLAN packets.

VTEPs in a multicast-mode VXLAN can use different multicast methods.

Prerequisites for multicast-mode VXLANs

For a multicast-mode VXLAN to flood traffic, you must perform the following tasks in addition to multicast-mode configuration:

- Enable IP multicast routing on all VTEPs and transport network devices.
- Configure a multicast routing protocol on transport network devices. A VTEP can be both a multicast source and multicast group member. As a best practice, use BIDIR-PIM.
- Enable IGMP on transport network devices that are connected to an IGMP host-enabled VTEP.

Configuring a multicast-mode VXLAN that uses the PIM method

1. Enter system view.

   system-view
2. Enter VSI view.

   vsi *vsi-name*
3. Enter VXLAN view.

   vxlan *vxlan-id*
4. Assign a multicast group address for flood traffic, and specify a source IP address for multicast VXLAN packets.

   group *group-address* **source** *source-address*

   By default, a VXLAN uses unicast mode for flood traffic. No multicast group address or source IP address is specified for multicast VXLAN packets.

   You must assign all VTEPs in a multicast-mode VXLAN to the same multicast group.
5. Enter interface view.

interface *interface-type interface-number*

Enable PIM on the loopback interface and all transport-facing interfaces.

6. Enable PIM. Choose one of the following modes:
   - Enable PIM-SM.

     pim sm
   - Enable PIM-DM.

     pim dm

   By default, PIM is disabled on an interface.

Configuring a multicast-mode VXLAN that uses the IGMP host method

Restrictions and guidelines

If you enable both the IGMP host feature and the IGMP protocol on an interface, the IGMP protocol will not take effect.

Procedure

1. Enter system view.

   system-view
2. Enter VSI view.

   vsi *vsi-name*
3. Enter VXLAN view.

   vxlan *vxlan-id*
4. Assign a multicast group address for flood traffic, and specify a source IP address for multicast VXLAN packets.

   group *group-address* **source** *source-address*

   By default, a VXLAN uses unicast mode for flood traffic. No multicast group address or source IP address is specified for multicast VXLAN packets.

   You must assign all VTEPs in a multicast-mode VXLAN to the same multicast group.
5. Enter the view of the transport-facing interface.

   interface *interface-type interface-number*
6. Enable the IGMP host feature.

   igmp host enable

   By default, the IGMP host feature is disabled on an interface.

   The IGMP host feature enables the interface to send IGMP reports in response to IGMP queries before it can receive traffic from the multicast group.

Configuring private VSI

About private VSI

In a data center network, typically VMs on a server are assigned to the same subnet. For security purposes, you can configure private VSI on the attached VTEP and private VLAN on the server to isolate the VMs based on VM attributes, operating system, and VM name. The server will send inter-VM traffic to the VTEP, and the VTEP will isolate traffic of different VMs or apply security policies as configured.

The private VSI feature uses a two-tier VSI structure, as shown in Figure 12. Each primary VSI is associated with one or multiple secondary VSIs, and the secondary VSIs communicate at Layer 3 through the primary VSI. A secondary VSI can be one of the following types:

- **Community VSI**—A secondary VSI whose ACs have Layer 2 connectivity to one another.
- **Isolated VSI**—A secondary VSI whose ACs are isolated at Layer 2.

Figure 12 Private VSI example

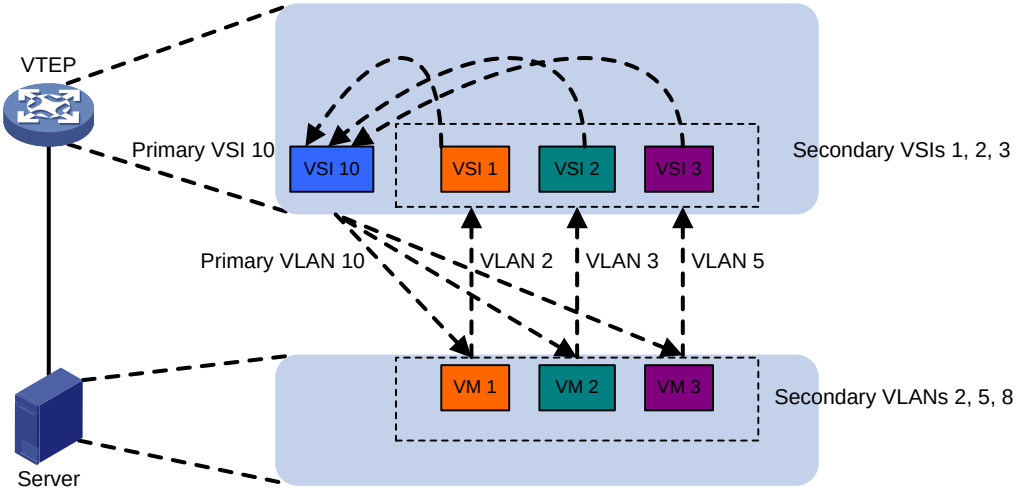


Figure 12 shows a private VSI application example. The private VLAN feature is configured on the server as follows:

- VLAN 10 is a primary VLAN.
- VLAN 2, VLAN 5, and VLAN 8 are secondary VLANs of VLAN 10.

The following VSIs are configured on the VTEP:

VSI	VSI type	VXLAN ID	Mapped VLAN
VSI 10	Primary VSI	10	VLAN 10
VSI 2	Community VSI	2	VLAN 2
VSI 5	Community VSI	5	VLAN 5
VSI 8	Community VSI	8	VLAN 8

To enable Layer 3 communication among the secondary VSIs, VSI-interface 1 is configured as the gateway interface for all VSIs, and local ARP and ND proxy is enabled on VSI-interface 1. The VTEP sends all packets destined for the VMs on the server only through the ACs of the primary VSI.

The following uses packet forwarding from VM 1 to VM 3 as an example to introduce the inter-VM communication procedure:

2. VM 1 sends a packet in VLAN 2. The destination IP address of the packet is the IP address of VM 3, and the destination MAC address is the MAC address of VSI-interface 1.
3. The server forwards the packet to the VTEP.
4. The VTEP receives the packet from an AC of VSI 2 and forwards the packet to VSI-interface 1.
5. VSI-interface 1 looks up the ARP table for the outgoing interface, encapsulates the packet, and forwards it through an AC of VSI 10 on the outgoing interface.
6. The server forwards the packet to VM 3.

Restrictions and guidelines

As a best practice, configure Ethernet service instances for VSIs as follows:

- Configure a primary VSI to match the packets of a primary VLAN on the peer server.
- Configure a secondary VSI to match the packets of a secondary VLAN on the peer server.

On a physical interface, you can map only one AC to a primary VSI.

You can configure only one-to-one mappings between private VLAN sets and private VSI sets.

You must use a VSI interface as the gateway interface for primary and secondary VSIs.

If a primary VSI is configured on an M-LAG system that performs Layer 3 forwarding, the traffic sent over the peer link will be flooded in the private VSIs.

Configuring a primary VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Configure the VSI as a primary VSI and enable the private VSI feature.
   private-vsi primary
   By default, the private VSI feature is disabled.
4. Specify secondary VSIs.
   private-vsi secondary *vsi-name*&<1-8>
   By default, a primary VSI is not associated with any secondary VSIs.

Configuring a secondary VSI

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Configure the VSI as a secondary VSI and enable the private VSI feature.
   private-vsi { **community** | **isolated** }
   By default, the private VSI feature is disabled.

Setting the destination UDP port number of VXLAN packets

1. Enter system view.
   system-view
2. Set a destination UDP port for VXLAN packets.
   vxlan udp-port *port-number*
   By default, the destination UDP port number is 4789 for VXLAN packets.
   You must configure the same destination UDP port number on all VTEPs in a VXLAN.

Configuring VXLAN packet check

About this task

The device can check the UDP checksum and 802.1Q VLAN tags of each received VXLAN packet.

- **VLAN tag check**—The device checks the inner Ethernet header of each VXLAN packet for 802.1Q VLAN tags. If the header contains 802.1Q VLAN tags, the device drops the packet.

Restrictions and guidelines

If a remote VTEP uses the Ethernet access mode, its VXLAN packets might contain 802.1Q VLAN tags. To prevent the local VTEP from dropping the VXLAN packets, do not execute the **`vxlan invalid-vlan-tag discard`** command on the local VTEP.

The access mode is configurable by using the **`xconnect vsi`** command.

Procedure

1. Enter system view.

   `system-view`
2. Enable the VTEP to drop VXLAN packets that have 802.1Q VLAN tags in the inner Ethernet header.

   `vxlan invalid-vlan-tag discard`

   By default, the VTEP does not check the inner Ethernet header for 802.1Q VLAN tags.

Configuring the DSCP values in the outer headers of VXLAN tunnel packets

About this task

VXLAN QoS forwards VXLAN tunnel packets based on the DSCP values in the outer headers. The DSCP values in the outer headers can be generated in the following ways:

- **copy-from-inner**—Directly copies the DSCP values from the inner headers of VXLAN tunnel packets to the DSCP values in the outer headers of the VXLAN tunnel packets.
- **local-preference-map**—Generates the DSCP values in the outer headers of VXLAN tunnel packets based on the colored mapping table that maps the QoS local precedence values to the VXLAN tunnel DSCP values.

Before configuring this feature, first configure the following QoS-related features:

- Execute the **`remark local-precedence`** command to set the local precedence values of the packets, which specify the local precedence values and packet colors for generating the DSCP values in the outer headers of VXLAN tunnel packets.
- Execute the **`qos map-table color`** command with the **`lp-tunnel-dscp`** keyword specified to enable mappings from local preference values to VXLAN tunnel DSCP values.

Restrictions and guidelines

With this feature configured, the DSCP values in the outer headers of all VXLAN tunnel packets will be affected.

Procedure

1. Enter system view.

   `system-view`
2. Configure the DSCP values in the outer headers of VXLAN tunnel packets.

vxlan-tunnel outer dscp { **copy-from-inner** | **local-precedence-map** }

By default, the DSCP values in the outer headers of VXLAN tunnel packets are copied from the DSCP values in the inner headers of the VXLAN tunnel packets.

Disabling flooding for a VSI

About this task

By default, the VTEP floods broadcast, unknown unicast, and unknown multicast frames received from the local site to the following interfaces in the frame's VXLAN:

- All site-facing interfaces except for the incoming interface.
- All VXLAN tunnel interfaces.

When receiving broadcast, unknown unicast, and unknown multicast frames on VXLAN tunnel interfaces, the device floods the frames to all site-facing interfaces in the frames' VXLAN.

To confine a kind of flood traffic, disable flooding for that kind of flood traffic on the VSI bound to the VXLAN.

To exclude a remote MAC address from the remote flood suppression done by using this feature, enable selective flood for the MAC address. The VTEP will flood the frames destined for the MAC address to remote sites.

Procedure

1. Enter system view.

   system-view
2. Enter VSI view.

   vsi *vsi-name*
3. Disable flooding for the VSI.

   flooding disable { **all** | { **broadcast** | **unknown-multicast** | **unknown-unicast** } * } [**all-direction**]

   By default, flooding is enabled for a VSI.

   You must specify the **unknown-multicast**, **broadcast**, and **unknown-unicast** keywords simultaneously when you execute this command. You cannot specify only one or two of them.
4. (Optional.) Enable selective flood for a MAC address.

   selective-flooding mac-address *mac-address*

Enabling ARP flood suppression

Restrictions and guidelines

The aging timer is fixed at 25 minutes for ARP flood suppression entries. If the suppression table is full, the VTEP stops learning new entries. For the VTEP to learn new entries, you must wait for old entries to age out, or use the **reset arp suppression vsi** command to clear the table.

If the **flooding disable** command is configured, set the MAC aging timer to a higher value than the aging timer for ARP flood suppression entries on all VTEPs. This setting prevents the traffic blackhole that occurs when a MAC address entry ages out before its ARP flood suppression entry ages out. To set the MAC aging timer, use the **mac-address timer** command.

When remote ARP learning is disabled for VXLANs, the device does not use ARP flood suppression entries to respond to ARP requests received on VXLAN tunnels.

Procedure

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. (Optional.) Set the ARP flood suppression mode.
   arp suppression mode { **proxy-reply** | **unicast-forward** } [**mismatch-discard**]
   If ARP flood suppression is enabled, the default ARP flood suppression mode is proxy reply. In proxy reply mode, the device broadcasts the ARP requests that are not targeted at the device and do not match any ARP flood suppression entries.
4. Enable ARP flood suppression.
   arp suppression enable [**no-broadcast**]
   By default, ARP flood suppression is disabled.

Enabling ND flood suppression

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. (Optional.) Set the ND flood suppression mode.
   ipv6 nd suppression mode { **proxy-reply** | **unicast-forward** } [**mismatch-discard**]
   If ND flood suppression is enabled, the default ND flood suppression mode is proxy reply. In proxy reply mode, the device broadcasts the NS packets that are not targeted at the device and do not match any ND flood suppression entries.
4. Enable ND flood suppression.
   ipv6 nd suppression enable [**no-broadcast**]
   By default, ND flood suppression is disabled.

Enabling VXLAN packet statistics

Enabling packet statistics for a VSI

Restrictions and guidelines

To display the packet statistics for a VSI, use the **display l2vpn vsi verbose** command in any view.

To clear the packet statistics for a VSI, use the **reset l2vpn statistics vsi** command in user view.

Procedure

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*

3. Enable packet statistics for the VSI.
   statistics enable
   By default, the packet statistics feature is disabled for all VSIs.

Enabling packet statistics for an AC

Restrictions and guidelines

For the **statistics enable** command to take effect on an Ethernet service instance, you must configure a frame match criterion for the Ethernet service instance and map it to a VSI. When you modify the frame match criterion or VSI mapping, the packet statistics of the instance are cleared.

Enabling packet statistics for an Ethernet service instance

1. Enter system view.
   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.
     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.
     interface bridge-aggregation *interface-number*
3. Enter Ethernet service instance view.
   service-instance *instance-id*
4. Enable packet statistics for the Ethernet service instance.
   statistics enable
   By default, the packet statistics feature is disabled for all Ethernet service instances.

Enabling packet statistics for Ethernet service instances of a VLAN

1. Enter system view.
   system-view
2. Enter VLAN view.
   vlan *vlan-id*
3. Enable packet statistics for Ethernet service instances of the VLAN.
   ac statistics enable
   By default, packet statistics are disabled for Ethernet service instances of a VLAN.
   This feature enables packet statistics for the Ethernet service instances that are automatically created for VLAN-based VXLAN assignment. Before you enable this feature, you must use the **vxlan vlan-based** command to enable VLAN-based VXLAN assignment.

Setting the L2VPN statistics collection interval

About this task

Perform this task to set the interval for collecting VSI, AC, and VXLAN tunnel packet statistics.

Procedure

1. Enter system view.
   system-view
2. Set the L2VPN statistics collection interval
   l2vpn statistics interval *interval*

The default L2VPN statistics collection interval is 15 minutes.

Testing the connectivity of a VXLAN tunnel

Enabling overlay OAM

About this task

You must enable overlay OAM on the tunnel destination device for a VXLAN tunnel before you can use the **ping vxlan** or **tracert vxlan** command to test reachability of the VXLAN tunnel on the tunnel source device.

Restrictions and guidelines

To specify the **-r** 3 parameter in the **ping vxlan**, **tracert vxlan**, **ping vxlan ipv6**, or **tracert vxlan ipv6** command on the tunnel source device, you must also enable overlay OAM on the tunnel source device.

Procedure

1. Enter system view.
   system-view
2. Enable overlay OAM.
   overlay oam enable
   By default, overlay OAM is disabled.

Pinging an IPv4 VXLAN tunnel destination

About this task

Perform this task to test the connectivity of a VXLAN tunnel in a VXLAN or EVPN VXLAN network when the tunnel has traffic loss or interruption issues. The process of a ping VXLAN operation is as follows:

1. The tunnel source VTEP sends VXLAN-encapsulated VXLAN echo requests to the tunnel destination VTEP.
2. The tunnel destination VTEP responds with VXLAN echo replies.
3. The tunnel source VTEP outputs packet statistics and the test result based on the received VXLAN echo replies.

Restrictions and guidelines

Before you perform this task on the tunnel source device, you must enable overlay OAM on the tunnel destination device by using the **overlay oam enable** command.

The VTEP can distribute VXLAN echo requests among multiple paths to the destination based on the source UDP port. When a VXLAN tunnel has multiple paths on the transport network, you can configure load sharing parameters to ensure accuracy of the test result. You can use one of the following methods to configure source UDP ports for VXLAN echo requests:

- Specify a source UDP port range. The device will send VXLAN echo requests sourced from each UDP port in the UDP port range. You need to execute the **ping vxlan** command only once.
- Specify load balancing parameters such as source and destination MAC addresses, source and destination IP addresses, and protocol for the VTEP to calculate a source UDP port number. You need to execute the **ping vxlan** command multiple times to test connectivity of all paths.

The load balancing parameters change only the source UDP port number of VXLAN echo requests. Other fields of the requests will not be changed.

If you specify the **vxlan-source-udpport** *vxlan-source-udpport* [*end-vxlan-src-udpport*] parameters, the number of VXLAN echo requests sourced from each UDP port in the UDP port range is determined by the **-c** *count* parameter.

Procedure

Execute the following command in any view:

ping vxlan [**-a** *inner-src-address* | **-c** *count* | **-m** *interval* | **-r** *reply-mode* | **-t** *timeout* | **-tos** *tos-value*] * **vxlan-id** *vxlan-id* **tunnel-source** *source-address* **tunnel-destination** *dest-address* [**destination-udpport** *dest-port*] [**vxlan-source-address** *vxlan-source-address*] [**load-balance** { **vxlan-source-udpport** *vxlan-source-udpport* [*end-vxlan-src-udpport*] | **source-address** *lb-src-address* **destination-address** *lb-dest-address* **protocol** { **udp** | *lb-protocol-id* } **source-port** *lb-src-port* **destination-port** *lb-dest-port* **source-mac** *lb-source-mac* **destination-mac** *lb-destination-mac* }]

Pinging an IPv6 VXLAN tunnel destination

About this task

Perform this task to test the connectivity of a VXLAN tunnel in a VXLAN or EVPN VXLAN network when the tunnel has traffic loss or interruption issues. The process of a ping VXLAN operation is as follows:

1. The tunnel source VTEP sends VXLAN-encapsulated VXLAN echo requests to the tunnel destination VTEP.
2. The tunnel destination VTEP responds with VXLAN echo replies.
3. The tunnel source VTEP outputs packet statistics and the test result based on the received VXLAN echo replies.

Restrictions and guidelines

Before you perform this task on the tunnel source device, you must enable overlay OAM on the tunnel destination device by using the **overlay oam enable** command.

The VTEP can distribute VXLAN echo requests among multiple paths to the destination based on the source UDP port. When a VXLAN tunnel has multiple paths on the transport network, you can configure load sharing parameters to ensure accuracy of the test result. You can use one of the following methods to configure source UDP ports for VXLAN echo requests:

- Specify a source UDP port range. The device will send VXLAN echo requests sourced from each UDP port in the UDP port range. You need to execute the **ping vxlan ipv6** command only once.
- Specify load balancing parameters such as source and destination MAC addresses, source and destination IP addresses, and protocol for the VTEP to calculate a source UDP port number. You need to execute the **ping vxlan ipv6** command multiple times to test connectivity of all paths.

The load balancing parameters change only the source UDP port number of VXLAN echo requests. Other fields of the requests will not be changed.

If you specify the **vxlan-source-udpport** *vxlan-source-udpport* [*end-vxlan-src-udpport*] parameters, the number of VXLAN echo requests sourced from each UDP port in the UDP port range is determined by the **-c** *count* parameter.

Procedure

Execute the following command in any view:

ping vxlan [**-a** *inner-src-ipv6-address* | **-c** *count* | **-m** *interval* | **-r** *reply-mode* | **-t** *timeout* | **-tos** *tos-value*] * **ipv6 vxlan-id** *vxlan-id* **tunnel-source** *source-ipv6-address* **tunnel-destination** *dest-ipv6-address* [**destination-udpport** *dest-port*] [**vxlan-source-address** *vxlan-source-ipv6-address*] [**load-balance** { **vxlan-source-udpport** *vxlan-source-udpport* [*end-vxlan-src-udpport*] | **source-address** *lb-src-address* **destination-address** *lb-dest-address* **protocol** { **udp** | *lb-protocol-id* } **source-port** *lb-src-port* **destination-port** *lb-dest-port* **source-mac** *lb-source-mac* **destination-mac** *lb-destination-mac* }]

Tracing the path to an IPv4 VXLAN tunnel destination

About this task

Perform this task to locate failed nodes on the path for a VXLAN tunnel that has traffic loss or interruption issues in a VXLAN or EVPN VXLAN network. The process of a tracert VXLAN operation is as follows:

1. The tunnel source VTEP sends VXLAN-encapsulated VXLAN echo requests to the tunnel destination VTEP. The TTL in the IP header of the requests is set to 1.
2. The first hop on the path responds to the tunnel source VTEP with a TTL-expired ICMP error message.
3. The tunnel source VTEP sends VXLAN echo requests with the TTL set to 2.
4. The second hop responds with a TTL-expired ICMP error message.
5. This process continues until a VXLAN echo request reaches the tunnel destination VTEP or the maximum TTL value is reached. If a VXLAN echo request reaches the tunnel destination VTEP, the tunnel destination VTEP sends a VXLAN echo reply to the tunnel source VTEP.
6. The tunnel source VTEP outputs packet statistics and the test result based on the received ICMP error messages and whether a VXLAN echo reply is received.

Restrictions and guidelines

Before you perform this task on the tunnel source device, you must enable overlay OAM on the tunnel destination device by using the **overlay oam enable** command.

The VTEP can distribute VXLAN echo requests among multiple paths to the destination based on the source UDP port. When a VXLAN tunnel has multiple paths on the transport network, you can configure load sharing parameters to ensure accuracy of the test result. You can use one of the following methods to configure source UDP ports for VXLAN echo requests:

- Specify a source UDP port range. The device will send VXLAN echo requests sourced from each UDP port in the UDP port range. You need to execute the **tracert vxlan** command only once.
- Specify load balancing parameters such as source and destination MAC addresses, source and destination IP addresses, and protocol for the VTEP to calculate a source UDP port number. You need to execute the **tracert vxlan** command multiple times to test connectivity of all paths.

The load balancing parameters change only the source UDP port number of VXLAN echo requests. Other fields of the requests will not be changed.

Procedure

Execute the following command in any view:

tracert vxlan [**-a** *inner-src-address* | **-h** *ttl-value* | **-r** *reply-mode* | **-t** *timeout*] * **vxlan-id** *vxlan-id* **tunnel-source** *source-address* **tunnel-destination** *dest-address* [**destination-udpport** *dest-port*] [**vxlan-source-address** *vxlan-source-address*] [**load-balance** { **vxlan-source-udpport** *vxlan-source-udpport* | **source-address**

lb-src-address **destination-address** *lb-dest-address* **protocol** { **udp** | *lb-protocol-id* } **source-port** *lb-src-port* **destination-port** *lb-dest-port* **source-mac** *lb-source-mac* **destination-mac** *lb-destination-mac* }]

Tracing the path to an IPv6 VXLAN tunnel destination

About this task

Perform this task to locate failed nodes on the path for a VXLAN tunnel that has traffic loss or interruption issues in a VXLAN or EVPN VXLAN network. The process of a tracert VXLAN operation is as follows:

1. The tunnel source VTEP sends VXLAN-encapsulated VXLAN echo requests to the tunnel destination VTEP. The Hop Limit in the IPv6 header of the requests is set to 1.
2. The first hop on the path responds to the tunnel source VTEP with a Hop Limit-expired ICMP error message.
3. The tunnel source VTEP sends VXLAN echo requests with the Hop Limit set to 2.
4. The second hop responds with a Hop Limit-expired ICMP error message.
5. This process continues until a VXLAN echo request reaches the tunnel destination VTEP or the maximum Hop Limit value is reached. If a VXLAN echo request reaches the tunnel destination VTEP, the tunnel destination VTEP sends a VXLAN echo reply to the tunnel source VTEP.
6. The tunnel source VTEP outputs packet statistics and the test result based on the received ICMP error messages and whether a VXLAN echo reply is received.

Restrictions and guidelines

Before you perform this task on the tunnel source device, you must enable overlay OAM on the tunnel destination device by using the **overlay oam enable** command.

The VTEP can distribute VXLAN echo requests among multiple paths to the destination based on the source UDP port. When a VXLAN tunnel has multiple paths on the transport network, you can configure load sharing parameters to ensure accuracy of the test result. You can use one of the following methods to configure source UDP ports for VXLAN echo requests:

- Specify a source UDP port range. The device will send VXLAN echo requests sourced from each UDP port in the UDP port range. You need to execute the **tracert vxlan ipv6** command only once.
- Specify load balancing parameters such as source and destination MAC addresses, source and destination IPv6 addresses, and protocol for the VTEP to calculate a source UDP port number. You need to execute the **tracert vxlan ipv6** command multiple times to test connectivity of all paths.

The load balancing parameters change only the source UDP port number of VXLAN echo requests. Other fields of the requests will not be changed.

Procedure

Execute the following command in any view.

tracert vxlan [**-a** *inner-src-ipv6-address* | **-h** *hop-limit-value* | **-r** *reply-mode* | **-t** *timeout*] * **ipv6 vxlan-id** *vxlan-id* **tunnel-source** *source-ipv6-address* **tunnel-destination** *dest-ipv6-address* [**destination-udpport** *dest-port*] [**vxlan-source-address** *vxlan-source-ipv6-address*] [**load-balance** { **vxlan-source-udpport** *vxlan-source-udpport* | **source-address** *lb-src-address* **destination-address** *lb-dest-address* **protocol** { **udp** | *lb-protocol-id* } **source-port** *lb-src-port* **destination-port** *lb-dest-port* **source-mac** *lb-source-mac* **destination-mac** *lb-destination-mac* }]

Display and maintenance commands for VXLANs

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display ARP flood suppression entries on VSIs.	**display arp suppression vsi** [**name** *vsi-name*] [**slot** *slot-number*] [**count**]
Display information about the multicast groups that contain IGMP host-enabled interfaces.	**display igmp** [**vpn-instance** *vpn-instance-name*] **host group** [*group-address* | **interface** *interface-type interface-number*] [**verbose**]
Display information about tunnel interfaces.	**display interface** [**tunnel** [*number*]] [**brief** [**description** | **down**]]
Display information about VXLAN tunnel interfaces.	**display vxlan tunnel-interface** [**tunnel** *number* [**verbose**]]
Display the IP addresses terminated by VXLAN tunnels.	**display vxlan tunnel termination**
Display ND flood suppression entries.	**display ipv6 nd suppression vsi** [**name** *vsi-name*] [**slot** *slot-number*] [**count**]
Display MAC address entries for VSIs.	**display l2vpn mac-address** [**vsi** *vsi-name*] [**dynamic**] [**count** | **verbose**]
Display MAC move records for VSIs.	**display l2vpn mac-address mac-move** [**slot** *slot-number*]
Display information about Ethernet service instances.	**display l2vpn service-instance** [**interface** *interface-type interface-number* [**service-instance** *instance-id*]] [**verbose**]
Display information about VSIs.	**display l2vpn vsi** [**name** *vsi-name*] [**verbose**]
Display VXLAN tunnel information for VXLANs.	**display vxlan tunnel** [**vxlan-id** *vxlan-id* [**tunnel** *tunnel-number*]]
Clear ARP flood suppression entries on VSIs.	**reset arp suppression vsi** [**name** *vsi-name*]
Clear ND flood suppression entries on VSIs.	**reset ipv6 nd suppression vsi** [**name** *vsi-name*]
Clear dynamic MAC address entries on VSIs.	**reset l2vpn mac-address** [**vsi** *vsi-name*]
Clear MAC move records for all VSIs.	**reset l2vpn mac-address mac-move**
Clear packet statistics on ACs.	**reset l2vpn statistics ac** [**interface** *interface-type interface-number* **service-instance** *instance-id*]
Clear packet statistics on VSIs.	**reset l2vpn statistics vsi** [**name** *vsi-name*]

NOTE:

For more information about the **`display interface tunnel`** command, see tunneling commands in *Layer 3—IP Services Command Reference*.

VXLAN configuration examples

Example: Configuring a unicast-mode VXLAN

Network configuration

As shown in Figure 13:

- Configure VXLAN 10 as a unicast-mode VXLAN on Switch A, Switch B, and Switch C to provide Layer 2 connectivity for the VMs across the network sites.
- Manually establish VXLAN tunnels and assign the tunnels to VXLAN 10.
- Enable remote-MAC address learning.

Figure 13 Network diagram

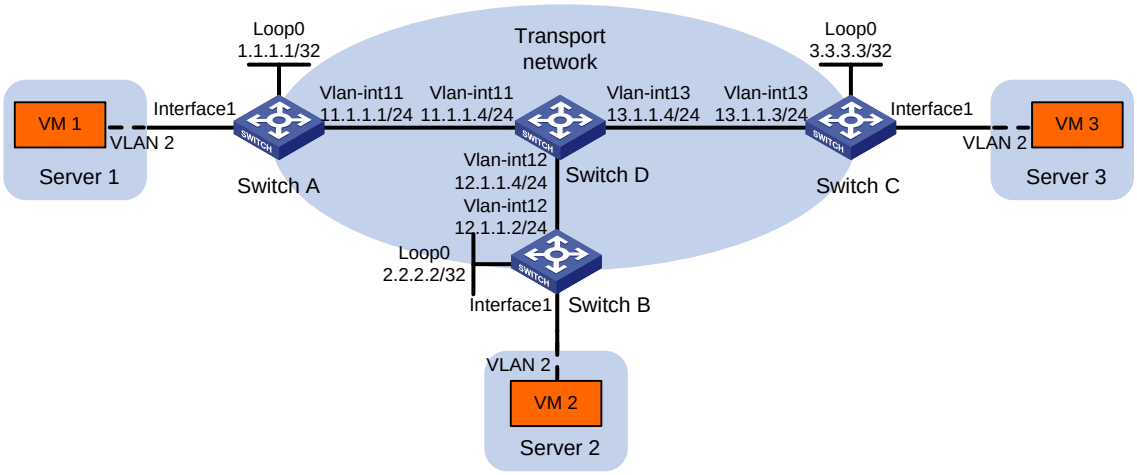


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Create VLANs and VLAN interfaces on all devices. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 13. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D). (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Create VSI **vpna** and VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch B and Switch C.

```
[SwitchA] interface loopback 0
[SwitchA-Loopback0] ip address 1.1.1.1 255.255.255.255
[SwitchA-Loopback0] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 1**.

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 1.1.1.1
[SwitchA-Tunnel1] destination 2.2.2.2
[SwitchA-Tunnel1] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 2**.

```
[SwitchA] interface tunnel 2 mode vxlan
[SwitchA-Tunnel2] source 1.1.1.1
[SwitchA-Tunnel2] destination 3.3.3.3
[SwitchA-Tunnel2] quit
```

Assign Tunnel 1 and Tunnel 2 to VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] tunnel 1
[SwitchA-vsi-vpna-vxlan-10] tunnel 2
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch C.

```
[SwitchB] interface loopback 0
[SwitchB-Loopback0] ip address 2.2.2.2 255.255.255.255
[SwitchB-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 2**.

```
[SwitchB] interface tunnel 2 mode vxlan
[SwitchB-Tunnel2] source 2.2.2.2
[SwitchB-Tunnel2] destination 1.1.1.1
[SwitchB-Tunnel2] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 3**.

```
[SwitchB] interface tunnel 3 mode vxlan
[SwitchB-Tunnel3] source 2.2.2.2
[SwitchB-Tunnel3] destination 3.3.3.3
[SwitchB-Tunnel3] quit
```

Assign Tunnel 2 and Tunnel 3 to VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] tunnel 2
[SwitchB-vsi-vpna-vxlan-10] tunnel 3
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch B.

```
[SwitchC] interface loopback 0
[SwitchC-Loopback0] ip address 3.3.3.3 255.255.255.255
[SwitchC-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 1**.

```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 3.3.3.3
[SwitchC-Tunnel1] destination 1.1.1.1
[SwitchC-Tunnel1] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 3**.

```
[SwitchC] interface tunnel 3 mode vxlan
[SwitchC-Tunnel3] source 3.3.3.3
[SwitchC-Tunnel3] destination 2.2.2.2
```

```
[SwitchC-Tunnel3] quit
```

Assign Tunnel 1 and Tunnel 3 to VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] tunnel 1
[SwitchC-vsi-vpna-vxlan-10] tunnel 3
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchC-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchC-GigabitEthernet1/0/1-srv1000] quit
[SwitchC-GigabitEthernet1/0/1] quit
```

Verifying the configuration

1. Verify the VXLAN settings on the VTEPs. This example uses Switch A.

Verify that the VXLAN tunnel interfaces on the VTEP are up.

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that the VXLAN tunnels have been assigned to the VXLAN.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
```

```
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name           Link ID    State  Type        Flood proxy
    Tunnel1               0x5000001  Up     Manual      Disabled
    Tunnel2               0x5000002  Up     Manual      Disabled
  ACs:
    AC                              Link ID    State    Type
    GE1/0/1 srv1000                 0          Up       Manual
```

Verify that the VTEP has learned the MAC addresses of remote VMs.

```
<SwitchA> display l2vpn mac-address
* - The output interface is issued to another VSI
MAC Address    State     VSI Name                        Link ID/Name   Aging
cc3e-5f9c-6cdb Dynamic   vpna                            Tunnel1        Aging
cc3e-5f9c-23dc Dynamic   vpna                            Tunnel2        Aging
--- 2 mac address(es) found  ---
```

2. Verify that VM 1, VM 2, and VM 3 can ping each other. (Details not shown.)

Example: Configuring a multicast-mode VXLAN

Network configuration

As shown in Figure 14:

- Configure VXLAN 10 as a multicast-mode VXLAN on Switch A, Switch B, and Switch C to provide Layer 2 connectivity for the VMs across the network sites.
- Manually establish VXLAN tunnels and assign the tunnels to VXLAN 10.
- Enable remote-MAC address learning.

Figure 14 Network diagram

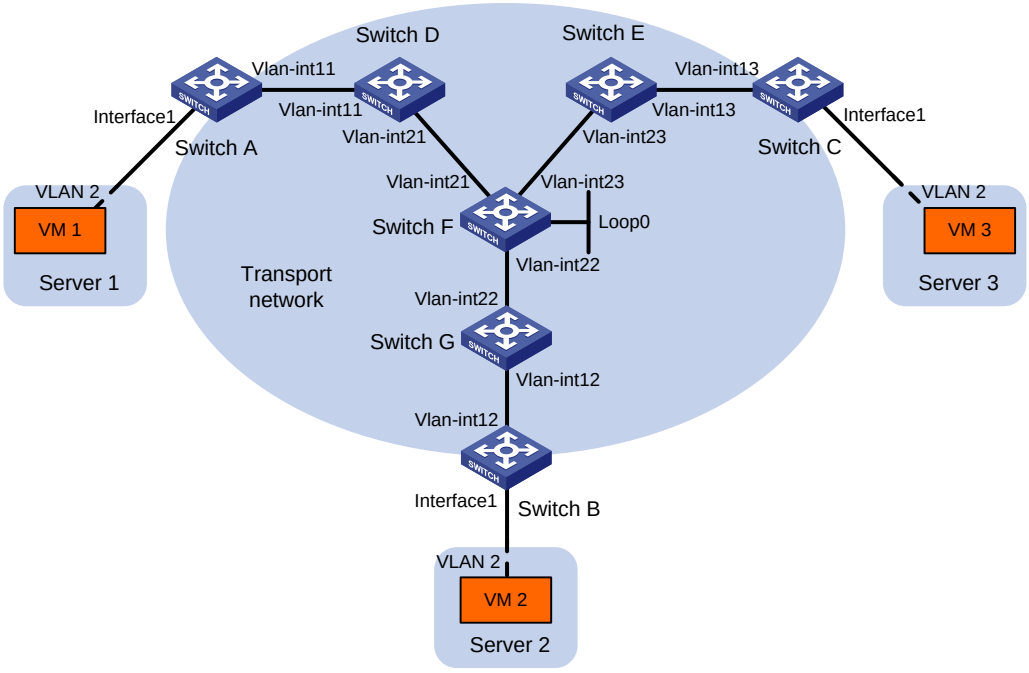


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Table 3 IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int11	11.1.1.1/24	Switch C	Vlan-int13	13.1.1.3/24
Switch D	Vlan-int11	11.1.1.4/24	Switch E	Vlan-int13	13.1.1.5/24
	Vlan-int21	21.1.1.4/24		Vlan-int23	23.1.1.5/24
Switch F	Vlan-int21	21.1.1.6/24	Switch G	Vlan-int12	12.1.1.7/24
	Vlan-int22	22.1.1.6/24		Vlan-int22	22.1.1.7/24
	Vlan-int23	23.1.1.6/24	Switch B	Vlan-int12	12.1.1.2/24
	Loop0	6.6.6.6/32			

Procedure

1.
2. Create VLANs and VLAN interfaces on all devices. (Details not shown.)
3. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 14. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through G). (Details not shown.)
4. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Enable IP multicast routing.

```
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Assign an IP address to VLAN-interface 11, and enable the IGMP host feature on the interface. This interface's IP address will be the source IP address of VXLAN packets sent by the VTEP.

```
[SwitchA] interface vlan-interface 11
[SwitchA-Vlan-interface11] ip address 11.1.1.1 24
[SwitchA-Vlan-interface11] igmp host enable
[SwitchA-Vlan-interface11] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 1**.

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 11.1.1.1
[SwitchA-Tunnel1] destination 12.1.1.2
[SwitchA-Tunnel1] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 2**.

```
[SwitchA] interface tunnel 2 mode vxlan
[SwitchA-Tunnel2] source 11.1.1.1
[SwitchA-Tunnel2] destination 13.1.1.3
[SwitchA-Tunnel2] quit
```

Assign Tunnel 1 and Tunnel 2 to VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] tunnel 1
[SwitchA-vsi-vpna-vxlan-10] tunnel 2
```

Configure the multicast group address and source IP address for multicast VXLAN packets.

```
[SwitchA-vsi-vpna-vxlan-10] group 225.1.1.1 source 11.1.1.1
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

5. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Enable IP multicast routing.

```
[SwitchB] multicast routing
[SwitchB-mrib] quit
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Assign an IP address to VLAN-interface 12, and enable the IGMP host feature on the interface. This interface's IP address will be the source IP address of VXLAN packets sent by the VTEP.

```
[SwitchB] interface vlan-interface 12
[SwitchB-Vlan-interface12] ip address 12.1.1.2 24
[SwitchB-Vlan-interface12] igmp host enable
[SwitchB-Vlan-interface12] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 2**.

```
[SwitchB] interface tunnel 2 mode vxlan
[SwitchB-Tunnel2] source 12.1.1.2
[SwitchB-Tunnel2] destination 11.1.1.1
[SwitchB-Tunnel2] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 3**.

```
[SwitchB] interface tunnel 3 mode vxlan
[SwitchB-Tunnel3] source 12.1.1.2
[SwitchB-Tunnel3] destination 13.1.1.3
[SwitchB-Tunnel3] quit
```

Assign Tunnel 2 and Tunnel 3 to VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] tunnel 2
[SwitchB-vsi-vpna-vxlan-10] tunnel 3
```

Configure the VXLAN multicast group address and the source IP address for VXLAN packets.

```
[SwitchB-vsi-vpna-vxlan-10] group 225.1.1.1 source 12.1.1.2
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

6. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Enable IP multicast routing.

```
[SwitchC] multicast routing
[SwitchC-mrib] quit
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchC] vsi vpna
```

```
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Assign an IP address to VLAN-interface 13, and enable the IGMP host feature on the interface. This interface's IP address will be the source IP address of VXLAN packets sent by the VTEP.

```
[SwitchC] interface vlan-interface 13
[SwitchC-Vlan-interface13] ip address 13.1.1.3 24
[SwitchC-Vlan-interface13] igmp host enable
[SwitchC-Vlan-interface13] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 1**.

```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 13.1.1.3
[SwitchC-Tunnel1] destination 11.1.1.1
[SwitchC-Tunnel1] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 3**.

```
[SwitchC] interface tunnel 3 mode vxlan
[SwitchC-Tunnel3] source 13.1.1.3
[SwitchC-Tunnel3] destination 12.1.1.2
[SwitchC-Tunnel3] quit
```

Assign Tunnel 1 and Tunnel 3 to VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] tunnel 1
[SwitchC-vsi-vpna-vxlan-10] tunnel 3
```

Configure the multicast group address and source IP address for VXLAN multicast packets.

```
[SwitchC-vsi-vpna-vxlan-10] group 225.1.1.1 source 13.1.1.3
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchC-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchC-GigabitEthernet1/0/1-srv1000] quit
[SwitchC-GigabitEthernet1/0/1] quit
```

7. Configure Switch D:

Enable IP multicast routing.

```
<SwitchD> system-view
[SwitchD] multicast routing
[SwitchD-mrib] quit
```

Enable IGMP and PIM-SM on VLAN-interface 11.

```
[SwitchD] interface vlan-interface 11
[SwitchD-Vlan-interface11] igmp enable
[SwitchD-Vlan-interface11] pim sm
[SwitchD-Vlan-interface11] quit
```

Enable PIM-SM on VLAN-interface 21.

```
[SwitchD] interface vlan-interface 21
[SwitchD-Vlan-interface21] pim sm
[SwitchD-Vlan-interface21] quit
```

Enable BIDIR-PIM.

```
[SwitchD] pim
[SwitchD-pim] bidir-pim enable
[SwitchD-pim] quit
```

8. Configure Switch E:

Enable IP multicast routing.

```
<SwitchE> system-view
[SwitchE] multicast routing
[SwitchE-mrib] quit
```

Enable IGMP and PIM-SM on VLAN-interface 13.

```
[SwitchE] interface vlan-interface 13
[SwitchE-Vlan-interface13] igmp enable
[SwitchE-Vlan-interface13] pim sm
[SwitchE-Vlan-interface13] quit
```

Enable PIM-SM on VLAN-interface 23.

```
[SwitchE] interface vlan-interface 23
[SwitchE-Vlan-interface23] pim sm
[SwitchE-Vlan-interface23] quit
```

Enable BIDIR-PIM.

```
[SwitchE] pim
[SwitchE-pim] bidir-pim enable
[SwitchE-pim] quit
```

9. Configure Switch F:

Enable IP multicast routing.

```
<SwitchF> system-view
[SwitchF] multicast routing
[SwitchF-mrib] quit
```

Enable PIM-SM on VLAN-interface 21, VLAN-interface 22, VLAN-interface 23, and Loopback 0.

```
[SwitchF] interface vlan-interface 21
[SwitchF-Vlan-interface21] pim sm
[SwitchF-Vlan-interface21] quit
[SwitchF] interface vlan-interface 22
[SwitchF-Vlan-interface22] pim sm
[SwitchF-Vlan-interface22] quit
[SwitchF] interface vlan-interface 23
[SwitchF-Vlan-interface23] pim sm
[SwitchF-Vlan-interface23] quit
[SwitchF] interface loopback 0
[SwitchF-LoopBack0] pim sm
[SwitchF-LoopBack0] quit
```

Enable BIDIR-PIM.

```
[SwitchF] pim
[SwitchF-pim] bidir-pim enable
```

Configure VLAN-interface 22 as a candidate-BSR, and configure Loopback 0 as a candidate-RP for BIDIR-PIM.

```
[SwitchF-pim] c-bsr 22.1.1.6
[SwitchF-pim] c-rp 6.6.6.6 bidir
[SwitchF-pim] quit
```

10. Configure Switch G:

Enable IP multicast routing.

```
<SwitchG> system-view
[SwitchG] multicast routing
[SwitchG-mrib] quit
```

Enable IGMP and PIM-SM on VLAN-interface 12.

```
[SwitchG] interface vlan-interface 12
[SwitchG-Vlan-interface12] igmp enable
[SwitchG-Vlan-interface12] pim sm
[SwitchG-Vlan-interface12] quit
```

Enable PIM-SM on VLAN-interface 22.

```
[SwitchG] interface vlan-interface 22
[SwitchG-Vlan-interface22] pim sm
[SwitchG-Vlan-interface22] quit
```

Enable BIDIR-PIM.

```
[SwitchG] pim
[SwitchG-pim] bidir-pim enable
[SwitchG-pim] quit
```

Verifying the configuration

1. Verify the VXLAN settings on the VTEPs. This example uses Switch A.

Verify that the VXLAN tunnel interfaces on the VTEP are up.

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Output queue - Urgent queuing: Size/Length/Discards 0/100/0
Output queue - Protocol queuing: Size/Length/Discards 0/500/0
Output queue - FIFO queuing: Size/Length/Discards 0/75/0
Last clearing of counters: Never
Tunnel source 11.1.1.1, destination 12.1.1.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that the VXLAN tunnels have been assigned to the VXLAN.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
```

```
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : Unlimited
  Broadcast Restrain      : Unlimited
  Multicast Restrain      : Unlimited
  Unknown Unicast Restrain: Unlimited
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name           Link ID    State  Type        Flood proxy
    Tunnel1               0x5000001  Up     Manual      Disabled
    Tunnel2               0x5000002  Up     Manual      Disabled
    Mtunnel10             0x6000000  Up     Auto        Disabled
  ACs:
    AC                              Link ID    State    Type
    GE1/0/1 srv1000                 0          Up       Manual
```

Verify that the VTEP has learned the MAC addresses of remote VMs.

```
<SwitchA> display l2vpn mac-address
* - The output interface is issued to another VSI
MAC Address    State     VSI Name                        Link ID/Name   Aging
cc3e-5f9c-6cdb Dynamic   vpna                            Tunnel1        Aging
cc3e-5f9c-23dc Dynamic   vpna                            Tunnel2        Aging
--- 2 mac address(es) found  ---
```

Verify that the VTEP has joined the VXLAN multicast group on VLAN-interface 11.

```
<SwitchA> display igmp host group
IGMP host groups in total: 1
 Vlan-interface11(11.1.1.1):
  IGMP host groups in total: 1
   Group address      Member state      Expires
   225.1.1.1          Idle              Off
```

2. Verify that VM 1, VM 2, and VM 3 can ping each other. (Details not shown.)

Example: Configuring a multicast-mode VXLAN to support M-LAG on a public network in M-LAG access mode

Network configuration

As shown in Figure 15:

- Configure a multicast-mode VXLAN on Switch A, Switch B, and Switch C to provide Layer 2 connectivity for the VMs across the network sites.
- Manually establish VXLAN tunnels and assign the tunnels to the VXLAN.
- Enable remote-MAC address learning.

Figure 15 Network diagram

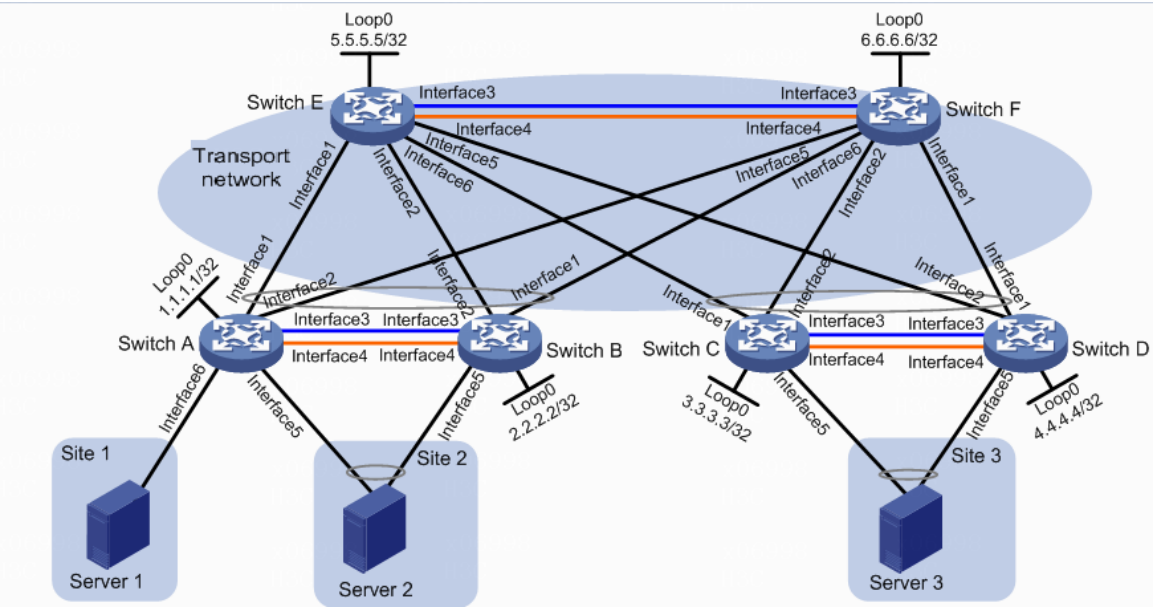


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5
Interface6	GigabitEthernet1/0/6

Table 5 IP address assignment

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int20	10.1.1.1/24	Switch C	Vlan-int30	16.1.1.5/24
	Vlan-int40	1.1.1.100/24		Vlan-int40	1.1.1.102/24
	GE1/0/3	60.1.2.1/24		GE1/0/3	60.4.5.4/24
	Loop1	1.1.2.2/32		Loop1	3.3.4.4/32
Switch B	Vlan-int20	10.1.1.2/24	Switch D	Vlan-int30	16.1.1.6/24
	Vlan-int40	1.1.1.101/24		Vlan-int40	1.1.1.103/24
	GE1/0/3	60.1.2.2/24		GE1/0/3	60.4.5.5/24
	Loop1	1.1.2.2/32		Loop1	3.3.4.4/32
Switch E	Vlan-int20	10.1.1.3/24	Switch F	Vlan-int20	10.1.1.4/24
	Vlan-int30	16.1.1.3/24		Vlan-int30	16.1.1.4/24
	GE1/0/3	60.1.4.1/24		GE1/0/3	60.1.4.2/24
	Loop1	5.5.6.6/32		Loop1	5.5.6.6/32

Procedure

1. Create VLANs and VLAN interfaces on all devices. (Details not shown.)

2. Configure IP addresses and unicast routing settings:

Assign IP addresses to interfaces, as shown in Figure xxx. (Details not shown.)

Configure OSPF on all transport network devices. (Details not shown.)

3. Configure Switch A:

Enable L2VPN.

```
<SwitchA> system-view
[SwitchA] l2vpn enable
```

Enable IP multicast routing.

```
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

Enable IGMP snooping.

```
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Specify a static RP.

```
[SwitchA] pim
[SwitchA-pim] static-rp 5.5.6.6
[SwitchA-pim] quit
```

Create VSI instance **vpna** and VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Configure M-LAG system settings.

```
[SwitchA] m-lag system-mac 1-1-1
[SwitchA] m-lag system-number 1
[SwitchA] m-lag system-priority 123
```

Configure the destination and source IP addresses of keepalive packets.

```
[SwitchA] m-lag keepalive ip destination 60.1.2.2 source 60.1.2.1
```

Set the link mode of GigabitEthernet 1/0/3 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] port link-mode route
[SwitchA-GigabitEthernet1/0/3] ip address 60.1.2.1 24
[SwitchA-GigabitEthernet1/0/3] quit
```

Exclude the interface used for M-LAG keepalive detection from the shutdown action by MAD.

```
[SwitchA] m-lag mad exclude interface
```

Create Layer 2 aggregate interface Bridge-Aggregation 12 and set its aggregation mode to dynamic.

```
[SwitchA] interface bridge-aggregation 12
[SwitchA-Bridge-Aggregation12] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation12] quit
```

Assign GigabitEthernet1/0/4 to aggregation group 12.

```
[SwitchA] interface gigabitethernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] port link-aggregation group 12
[SwitchA-GigabitEthernet1/0/4] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 12 as the peer-link interface.

```
[SwitchA] interface bridge-aggregation 12
```

```
[SwitchA-Bridge-Aggregation12] port m-lag peer-link 1
[SwitchA-Bridge-Aggregation12] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 1 and configure the interface to operate in dynamic mode.

```
[SwitchA] interface bridge-aggregation 1
[SwitchA-Bridge-Aggregation1] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation1] quit
```

Assign GigabitEthernet1/0/1 and GigabitEthernet1/0/2 to aggregation group 1.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-aggregation group 1
[SwitchA-GigabitEthernet1/0/2] quit
```

Assign Layer 2 aggregate interface Bridge-Aggregation 1 to M-LAG group 1.

```
[SwitchA] interface bridge-aggregation 1
[SwitchA-Bridge-Aggregation1] port m-lag group 1
[SwitchA-Bridge-Aggregation1] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 2 and configure the interface to operate in dynamic mode.

```
[SwitchA] interface bridge-aggregation 2
[SwitchA-Bridge-Aggregation2] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation2] quit
```

Assign GigabitEthernet1/0/5 to aggregation group 2.

```
[SwitchA] interface gigabitethernet 1/0/5
[SwitchA-GigabitEthernet1/0/5] port link-aggregation group 2
[SwitchA-GigabitEthernet1/0/5] quit
```

Assign Layer 2 aggregate interface Bridge-Aggregation 2 to M-LAG group 2.

```
[SwitchA] interface bridge-aggregation2
[SwitchA-Bridge-Aggregation2] port m-lag group 2
[SwitchA-Bridge-Aggregation2] quit
```

Assign aggregate interface 1 to VLAN 20. Create VLAN-interface 20. Assign an IP address to VLAN-interface 20, and enable PIM-SM on the interface.

```
[SwitchA] interface bridge-aggregation 1
[SwitchA-Bridge-Aggregation1] port trunk permit vlan 20
[SwitchA-Bridge-Aggregation1] quit
[SwitchA] vlan 20
[SwitchA] interface vlan-interface 20
[SwitchA-Vlan-interface20] ip address 10.1.1.1 24
[SwitchA-Vlan-interface20] pim sm
[SwitchA-Vlan-interface20] quit
```

Enable PIM SM on interface **LoopBack1**.

```
[SwitchA] interface loopback 1
[SwitchA-Loopback1] pim sm
[SwitchA-Loopback1] quit
```

Establish a VXLAN tunnel between Switch A and Switch C:

- Create tunnel interface **Tunnel1** in VXLAN mode.

- Specify the source address of the tunnel as 1.1.2.2, which is the address of local interface **LoopBack1**.
- Specify the destination address of the tunnel as 3.3.4.4, which is the address of interface **LoopBack1** on Switch C.

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 1.1.2.2
[SwitchA-Tunnel1] destination 3.3.4.4
[SwitchA-Tunnel1] quit
```

Assign tunnel interface **Tunnel1** to VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] tunnel 1
```

Assign multicast group address 239.0.0.1 for flood traffic, and specify source IP address 1.1.2.2 for multicast VXLAN packets.

```
[SwitchA-vsi-vpna-vxlan-10] group 239.0.0.1 source 1.1.2.2
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

On aggregate interface 2 that connects to servers, create Ethernet service instance 1000 and configure it to match data frames from VLAN 2.

```
[SwitchA] interface bridge-aggregation 2
[SwitchA-Bridge-Aggregation2] port trunk permit vlan 2
[SwitchA-Bridge-Aggregation2] service-instance 1000
[SwitchA-Bridge-Aggregation2-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-Bridge-Aggregation2-srv1000] xconnect vsi vpna
[SwitchA-Bridge-Aggregation2-srv1000] quit
[SwitchA-Bridge-Aggregation2] quit
```

Enable VLAN-interface 40 to join a multicast group.

```
[SwitchA] vlan 40
[SwitchA] interface vlan-interface 40
[SwitchA-Vlan-interface40] ip address 1.1.1.100 24
[SwitchA-Vlan-interface40] pim sm
[SwitchA-Vlan-interface40] vxlan multicast-group member enable
```

4. Configure Switch B:

Configure Switch B in the same way Switch A is configured. (Details not shown.)

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Enable IP multicast routing.

```
[SwitchC] multicast routing
[SwitchC-mrib] quit
```

Enable IGMP snooping.

```
[SwitchC] igmp-snooping
[SwitchC-igmp-snooping] quit
```

Specify a static RP.

```
[SwitchC] pim
[SwitchC-pim] static-rp 5.5.6.6
```

```
[SwitchC-pim] quit
```

Create VSI instance **vpna** and VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Configure M-LAG system settings.

```
[SwitchC] m-lag system-mac 2-2-2
[SwitchC] m-lag system-number 1
[SwitchC] m-lag system-priority 123
```

Configure the destination and source IP addresses of keepalive packets.

```
[SwitchC] m-lag keepalive ip destination 60.4.5.5 source 60.4.5.4
```

Set the link mode of GigabitEthernet1/0/3 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[SwitchC] interface gigabitethernet 1/0/3
[SwitchC-GigabitEthernet1/0/3] port link-mode route
[SwitchC-GigabitEthernet1/0/3] ip address 60.4.5.4 24
[SwitchC-GigabitEthernet1/0/3] quit
```

Exclude the interface used for M-LAG keepalive detection from the shutdown action by MAD.

```
[SwitchC] m-lag mad exclude interface gigabitethernet 1/0/3
```

Create Layer 2 aggregate interface Bridge-Aggregation 12 and set its aggregation mode to dynamic.

```
[SwitchC] interface bridge-aggregation 12
[SwitchC-Bridge-Aggregation12] link-aggregation mode dynamic
[SwitchC-Bridge-Aggregation12] quit
```

Assign GigabitEthernet1/0/4 to aggregation group 12.

```
[SwitchC] interface gigabitethernet 1/0/4
[SwitchC-GigabitEthernet1/0/4] port link-aggregation group 12
[SwitchC-GigabitEthernet1/0/4] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 12 as the peer-link interface.

```
[SwitchC] interface bridge-aggregation 12
[SwitchC-Bridge-Aggregation12] port m-lag peer-link 1
[SwitchC-Bridge-Aggregation12] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 1 and configure the interface to operate in dynamic mode.

```
[SwitchC] interface bridge-aggregation 1
[SwitchC-Bridge-Aggregation1] link-aggregation mode dynamic
[SwitchC-Bridge-Aggregation1] quit
```

Assign GigabitEthernet1/0/1 and GigabitEthernet1/0/2 to aggregation group 1.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchC-GigabitEthernet1/0/1] quit
[SwitchC] interface gigabitethernet 1/0/2
[SwitchC-GigabitEthernet1/0/2] port link-aggregation group 1
[SwitchC-GigabitEthernet1/0/2] quit
```

Assign Layer 2 aggregate interface Bridge-Aggregation 1 to M-LAG group 1.

```
[SwitchC] interface bridge-aggregation 1
[SwitchC-Bridge-Aggregation1] port m-lag group 1
```

```
[SwitchC-Bridge-Aggregation1] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 2 and configure the interface to operate in dynamic mode.

```
[SwitchC] interface bridge-aggregation 2
[SwitchC-Bridge-Aggregation2] link-aggregation mode dynamic
[SwitchC-Bridge-Aggregation2] quit
```

Assign GigabitEthernet1/0/5 to aggregation group 2.

```
[SwitchC] interface gigabitethernet 1/0/5
[SwitchC-GigabitEthernet1/0/5] port link-aggregation group 2
[SwitchC-GigabitEthernet1/0/5] quit
```

Assign Layer 2 aggregate interface Bridge-Aggregation 2 to M-LAG group 2.

```
[SwitchC] interface bridge-aggregation2
[SwitchC-Bridge-Aggregation2] port m-lag group 2
[SwitchC-Bridge-Aggregation2] quit
```

Assign aggregate interface 1 to VLAN 30. Create VLAN-interface 30. Assign an IP address to VLAN-interface 30, and enable PIM-SM on the interface.

```
[SwitchC] interface bridge-aggregation 1
[SwitchC-Bridge-Aggregation1] port trunk permit vlan 30
[SwitchC-Bridge-Aggregation1] quit
[SwitchC] vlan 30
[SwitchC] interface vlan-interface 30
[SwitchC-Vlan-interface30] ip address 16.1.1.5 24
[SwitchC-Vlan-interface30] pim sm
[SwitchC-Vlan-interface30] quit
```

Enable PIM SM on interface **LoopBack1**.

```
[SwitchC] interface loopback 1
[SwitchC-Loopback1] pim sm
[SwitchC-Loopback1] quit
```

Establish a VXLAN tunnel between Switch C and Switch A:

- Create tunnel interface **Tunnel1** in VXLAN mode.
- Specify the source address of the tunnel as 3.3.4.4, which is the address of local interface **LoopBack1**.
- Specify the destination address of the tunnel as 1.1.2.2, which is the address of interface **LoopBack1** on Switch A.

```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 3.3.4.4
[SwitchC-Tunnel1] destination 1.1.2.2
[SwitchC-Tunnel1] quit
```

Assign tunnel interface **Tunnel1** to VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] tunnel 1
```

Assign multicast group address 239.0.0.1 for flood traffic, and specify source IP address 3.3.4.4 for multicast VXLAN packets.

```
[SwitchC-vsi-vpna-vxlan-10] group 239.0.0.1 source 3.3.4.4
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

On aggregate interface 2 that connects to servers, create Ethernet service instance 1000 and configure it to match data frames from VLAN 2.

```
[SwitchC] interface bridge-aggregation 2
[SwitchC-Bridge-Aggregation2] port trunk permit vlan 2
[SwitchC-Bridge-Aggregation2] service-instance 1000
[SwitchC-Bridge-Aggregation2-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchC-Bridge-Aggregation2-srv1000] xconnect vsi vpna
[SwitchC-Bridge-Aggregation2-srv1000] quit
[SwitchC-Bridge-Aggregation2] quit
```

Enable VLAN-interface 40 to join a multicast group.

```
[SwitchC] vlan 40
[SwitchC] interface vlan-interface 40
[SwitchC-Vlan-interface40] ip address 1.1.1.103 24
[SwitchC-Vlan-interface40] pim sm
[SwitchC-Vlan-interface40] vxlan multicast-group member enable
```

6. Configure Switch D:

Configure Switch D in the same way Switch C is configured. (Details not shown.)

7. Configure Switch E:

Enable L2VPN.

```
<SwitchE> system-view
[SwitchE] l2vpn enable
```

Enable IP multicast routing.

```
[SwitchE] multicast routing
[SwitchE-mrib] quit
```

Enable IGMP snooping.

```
[SwitchE] igmp-snooping
[SwitchE-igmp-snooping] quit
```

Specify a static RP.

```
[SwitchE] pim
[SwitchE-pim] static-rp 5.5.6.6
[SwitchE-pim] quit
```

Configure M-LAG system settings.

```
[SwitchE] m-lag system-mac 34-34-34
[SwitchE] m-lag system-number 1
[SwitchE] m-lag system-priority 123
```

Configure the destination and source IP addresses of keepalive packets.

```
[SwitchE] m-lag keepalive ip destination 60.1.4.2 source 60.1.4.1
```

Set the link mode of GigabitEthernet1/0/3 to Layer 3, and assign the interface an IP address. The IP address will be used as the source IP address of keepalive packets.

```
[SwitchE] interface gigabitethernet 1/0/3
[SwitchE-GigabitEthernet1/0/3] port link-mode route
[SwitchE-GigabitEthernet1/0/3] ip address 60.1.4.1 24
[SwitchE-GigabitEthernet1/0/3] quit
```

Exclude the interface used for M-LAG keepalive detection from the shutdown action by MAD.

```
[SwitchE] m-lag mad exclude interface gigabitethernet 1/0/3
```

Create Layer 2 aggregate interface Bridge-Aggregation 12 and set its aggregation mode to dynamic.

```
[SwitchE] interface bridge-aggregation 12
[SwitchE-Bridge-Aggregation12] link-aggregation mode dynamic
[SwitchE-Bridge-Aggregation12] quit
```

Assign GigabitEthernet1/0/4 to aggregation group 12.

```
[SwitchE] interface gigabitethernet 1/0/4
[SwitchE-GigabitEthernet1/0/4] port link-aggregation group 12
[SwitchE-GigabitEthernet1/0/4] quit
```

Configure Layer 2 aggregate interface Bridge-Aggregation 12 as the peer-link interface.

```
[SwitchE] interface bridge-aggregation 12
[SwitchE-Bridge-Aggregation12] port m-lag peer-link 1
 [SwitchE-Bridge-Aggregation12] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 1 and configure the interface to operate in dynamic mode.

```
[SwitchE] interface bridge-aggregation 1
[SwitchE-Bridge-Aggregation1] link-aggregation mode dynamic
[SwitchE-Bridge-Aggregation1] quit
```

Enable PIM SM on interface **LoopBack1**.

```
[SwitchE] interface loopback 1
[SwitchE-Loopback1] pim sm
[SwitchE-Loopback1] quit
```

Assign GigabitEthernet1/0/1 and GigabitEthernet1/0/2 to aggregation group 1.

```
[SwitchE] interface gigabitethernet 1/0/1
[SwitchE-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchE-GigabitEthernet1/0/1] quit
[SwitchE] interface gigabitethernet 1/0/2
[SwitchE-GigabitEthernet1/0/2] port link-aggregation group 1
[SwitchE-GigabitEthernet1/0/2] quit
```

Assign Layer 2 aggregate interface Bridge-Aggregation 1 to M-LAG group 1.

```
[SwitchE] interface bridge-aggregation 1
[SwitchE-Bridge-Aggregation1] port m-lag group 1
[SwitchE-Bridge-Aggregation1] quit
```

Create Layer 2 aggregate interface Bridge-Aggregation 2 and configure the interface to operate in dynamic mode.

```
[SwitchE] interface bridge-aggregation 2
[SwitchE-Bridge-Aggregation2] link-aggregation mode dynamic
[SwitchE-Bridge-Aggregation2] quit
```

Assign GigabitEthernet1/0/5 and GigabitEthernet1/0/6 to aggregation group 2.

```
[SwitchE] interface gigabitethernet 1/0/5
[SwitchE-GigabitEthernet1/0/5] port link-aggregation group 2
[SwitchE-GigabitEthernet1/0/5] quit
[SwitchE] interface gigabitethernet 1/0/6
[SwitchE-GigabitEthernet1/0/6] port link-aggregation group 2
[SwitchE-GigabitEthernet1/0/6] quit
```

Assign Layer 2 aggregate interface Bridge-Aggregation 2 to M-LAG group 2.

```
[SwitchE] interface bridge-aggregation2
[SwitchE-Bridge-Aggregation2] port m-lag group 2
[SwitchE-Bridge-Aggregation2] quit
```

Assign aggregate interface 1 to VLAN 20. Create VLAN-interface 20. Assign an IP address to VLAN-interface 20, and enable PIM-SM on the interface.

```
[SwitchE] interface bridge-aggregation 1
[SwitchE-Bridge-Aggregation1] port trunk permit vlan 20
[SwitchE-Bridge-Aggregation1] quit
[SwitchE] vlan 20
[SwitchE] interface vlan-interface 20
[SwitchE-Vlan-interface20] ip address 10.1.1.3 24
[SwitchE-Vlan-interface20] pim sm
[SwitchE-Vlan-interface20] quit
```

Assign aggregate interface 2 to VLAN 30. Create VLAN-interface 30. Assign an IP address to VLAN-interface 30, and enable PIM-SM on the interface.

```
[SwitchE] interface bridge-aggregation 2
[SwitchE-Bridge-Aggregation2] port trunk permit vlan 30
[SwitchE-Bridge-Aggregation2] quit
[SwitchE] vlan 30
[SwitchE] interface vlan-interface 30
[SwitchE-Vlan-interface30] ip address 16.1.1.3 24
[SwitchE-Vlan-interface30] pim sm
[SwitchE-Vlan-interface30] quit
```

8. Configure Switch F:

Configure Switch F in the same way Switch E is configured. (Details not shown.)

Verifying the configuration

1. Verify the VXLAN settings on the VTEPs. This example uses Switch A.

Verify that the VXLAN tunnel interfaces on Switch A are up.

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Output queue - Urgent queuing: Size/Length/Discards 0/100/0
Output queue - Protocol queuing: Size/Length/Discards 0/500/0
Output queue - FIFO queuing: Size/Length/Discards 0/75/0
Last clearing of counters: Never
Tunnel source 1.1.2.2, destination 3.3.4.4
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Display the PIM routing entries on Switch A.

```
[SwitchA] display pim routing-table
Total 1 (*, G) entries; 1 (S, G) entries

(*, 239.0.0.1)
```

```
    RP: 5.5.6.6
    Protocol: pim-sm, Flag: WC
    UpTime: 00:55:36
    Upstream interface: Vlan-interface10
        Upstream neighbor: 10.1.1.4
        RPF prime neighbor: 10.1.1.4
    Downstream interface information:
    Total number of downstream interfaces: 1
        1: Mtunnel10 (public-instance)
            Protocol: VXLAN, UpTime: 00:55:36, Expires: -

(1.1.2.2, 239.0.0.1)
    RP: 5.5.6.6
    Protocol: pim-sm, Flag: RPT SPT LOC VXLAN
    UpTime: 00:55:36
    Upstream interface: Mtunnel10
        Upstream neighbor: NULL
        RPF prime neighbor: NULL
    Downstream interface information:
    Total number of downstream interfaces: 2
        1: Vlan-interface10
            Protocol: pim-sm, UpTime: 00:55:36, Expires: 00:02:45
        2: Vlan-interface40
            Protocol: VXLAN_MEMBER, UpTime: 00:00:41, Expires: -
```

Configuring VXLAN IP gateways

About VXLAN IP gateways

The following are available IP gateway placement designs for VXLANs:

- **VXLAN IP gateways separated from VTEPs**—Use a VXLAN-unaware device as a gateway to the external network for VXLANs. On the gateway, you do not need to configure VXLAN settings.
- **VXLAN IP gateways collocated with VTEPs**—Include the following placement designs:
  - **Centralized VXLAN IP gateway deployment**—Use one VTEP to provide Layer 3 forwarding for VXLANs. Typically, the gateway-collocated VTEP connects to other VTEPs and the external network. To use this design, make sure the IP gateway has sufficient bandwidth and processing capability.
  - **Distributed VXLAN IP gateway deployment**—Deploy one VXLAN IP gateway on each VTEP to provide Layer 3 forwarding for VXLANs at their respective sites. This design distributes the Layer 3 traffic load across VTEPs. However, its configuration is more complex than the centralized VXLAN IP gateway design.

In a collocation design, the VTEPs use virtual Layer 3 VSI interfaces as gateway interfaces to provide services for VXLANs.

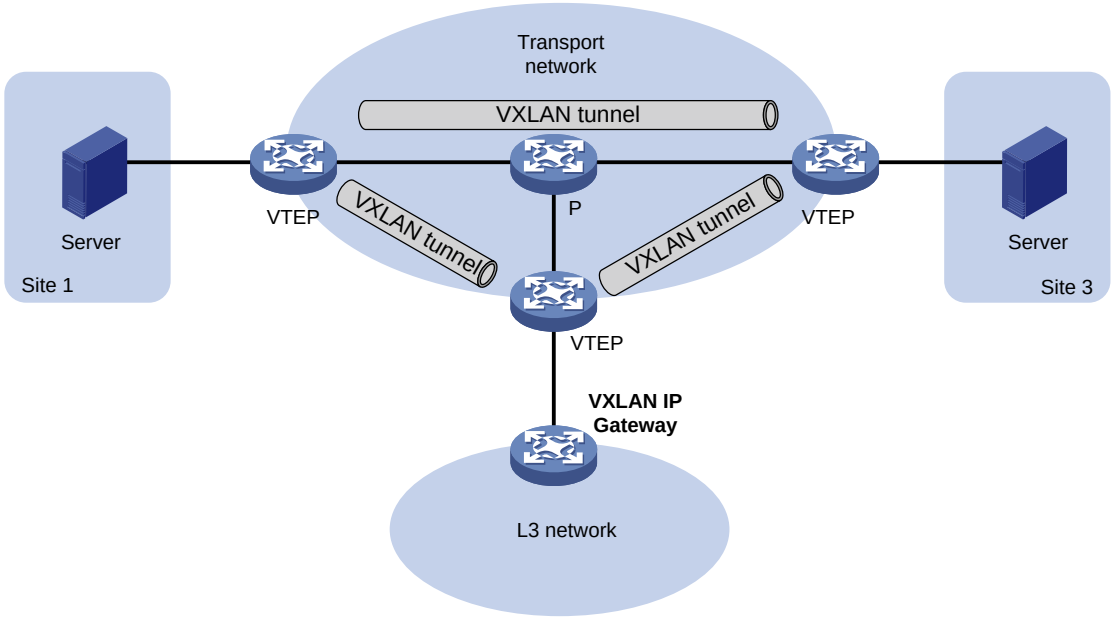
NOTE:

The following information describes traffic forwarding of VXLAN IP gateways in IPv4 networks. Traffic forwarding of VXLAN IP gateways in IPv6 networks is similar to that in IPv4 networks.

VXLAN IP gateways separated from VTEPs

As shown in Figure 16, an independent VXLAN IP gateway connects a Layer 3 network to a VTEP. VMs send Layer 3 traffic in Layer 2 frames to the gateway through VXLAN tunnels. When the tunneled VXLAN packets arrive, the VTEP terminates the VXLANs and forwards the inner frames to the gateway. In this gateway placement design, the VTEP does not perform Layer 3 forwarding for VXLANs.

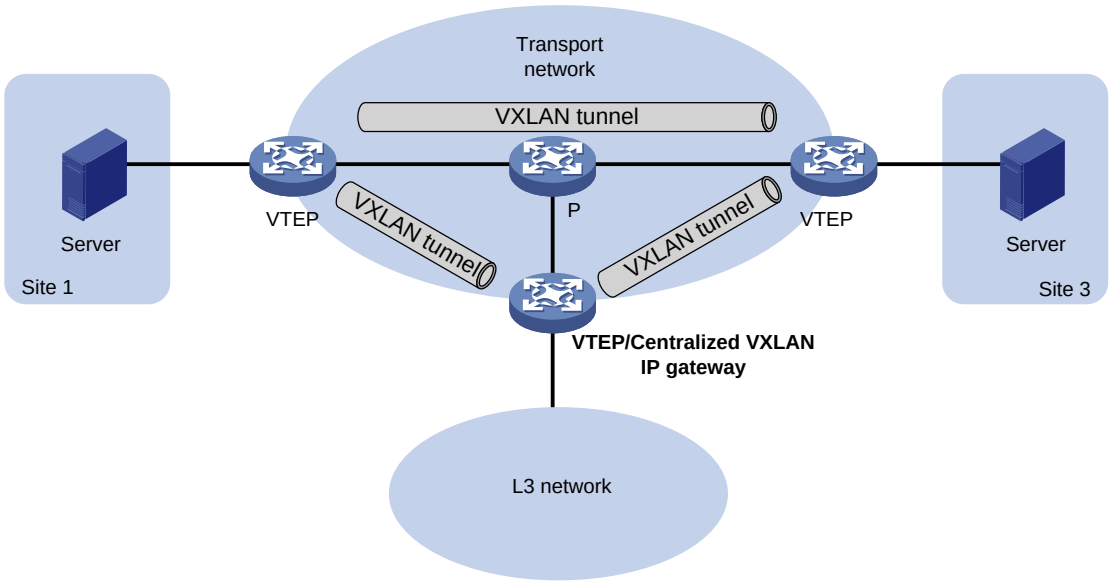
Figure 16 VXLAN IP gateway separated from VTEPs



Centralized VXLAN IP gateway deployment

As shown in Figure 17, a VTEP acts as a gateway for VMs in the VXLANs. The VTEP both terminates the VXLANs and performs Layer 3 forwarding for the VMs.

Figure 17 Centralized VXLAN IP gateway placement design



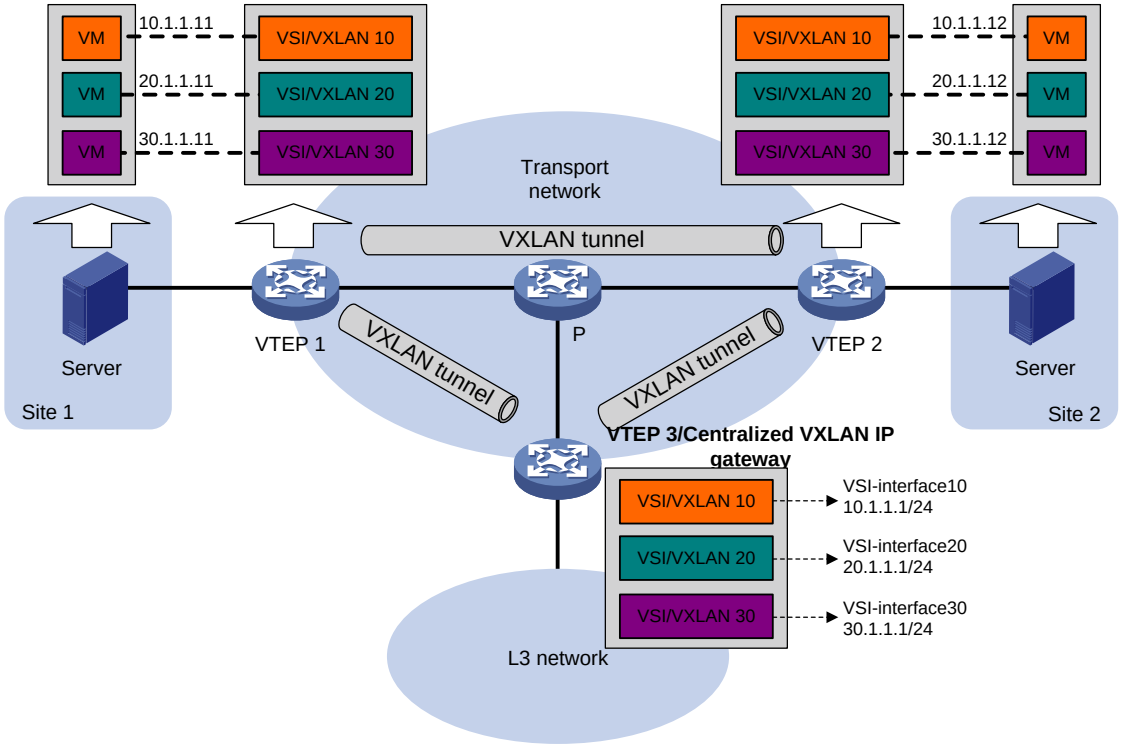
As shown in Figure 18, the network uses the following process to forward Layer 3 traffic from VM 10.1.1.11 to the Layer 3 network:

2. The VM sends an ARP request to obtain the MAC address of the gateway (VTEP 3) at 10.1.1.1.
3. VTEP 1 floods the ARP request to all remote VTEPs.
4. VTEP 3 de-encapsulates the ARP request, creates an ARP entry for the VM, and sends an ARP reply to the VM.

5. VTEP 1 forwards the ARP reply to the VM.
6. The VM learns the MAC address of the gateway, and sends the Layer 3 traffic to the gateway.
7. VTEP 3 removes the VXLAN encapsulation and inner Ethernet header for the traffic, and forwards the traffic to the destination node.

Inter-VXLAN forwarding is the same as this process except for the last step. At the last step of inter-VLAN forwarding, the gateway replaces the source-VXLAN encapsulation with the destination-VXLAN encapsulation, and then forwards the traffic.

Figure 18 Example of centralized VXLAN IP gateway deployment



Distributed VXLAN IP gateway deployment

About distributed VXLAN IP gateway deployment

As shown in Figure 19, each site's VTEP acts as a gateway to perform Layer 3 forwarding for the VXLANs of the local site. A VTEP acts as a border gateway to the Layer 3 network for the VXLANs.

Figure 19 Distributed VXLAN IP gateway placement design

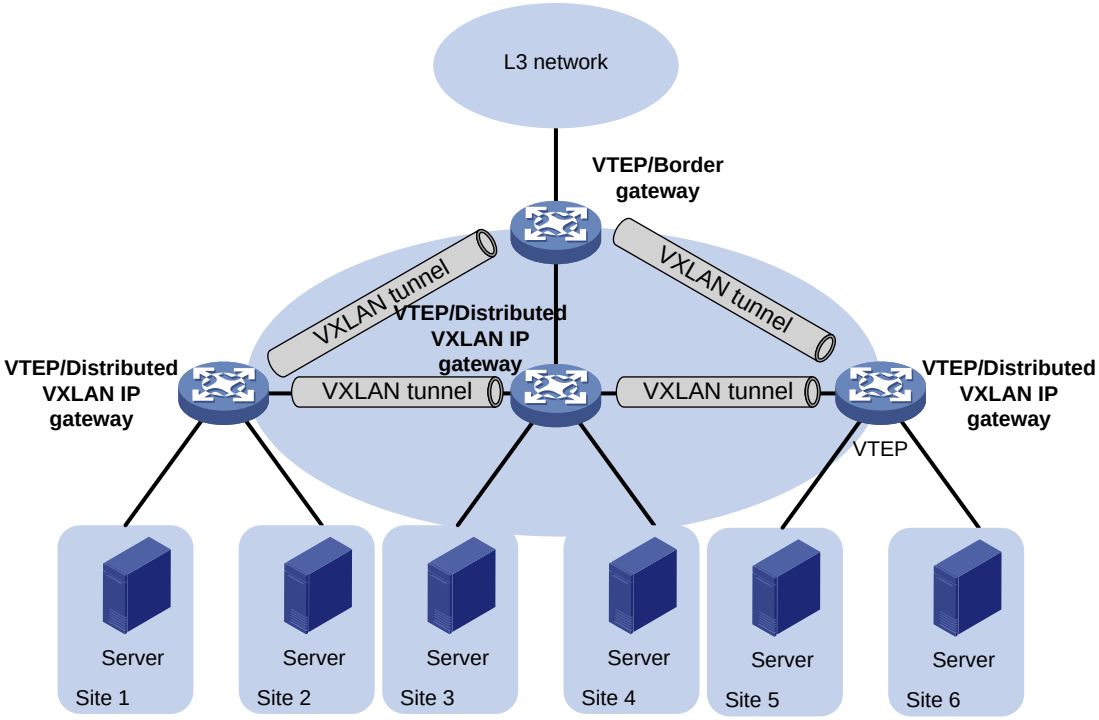
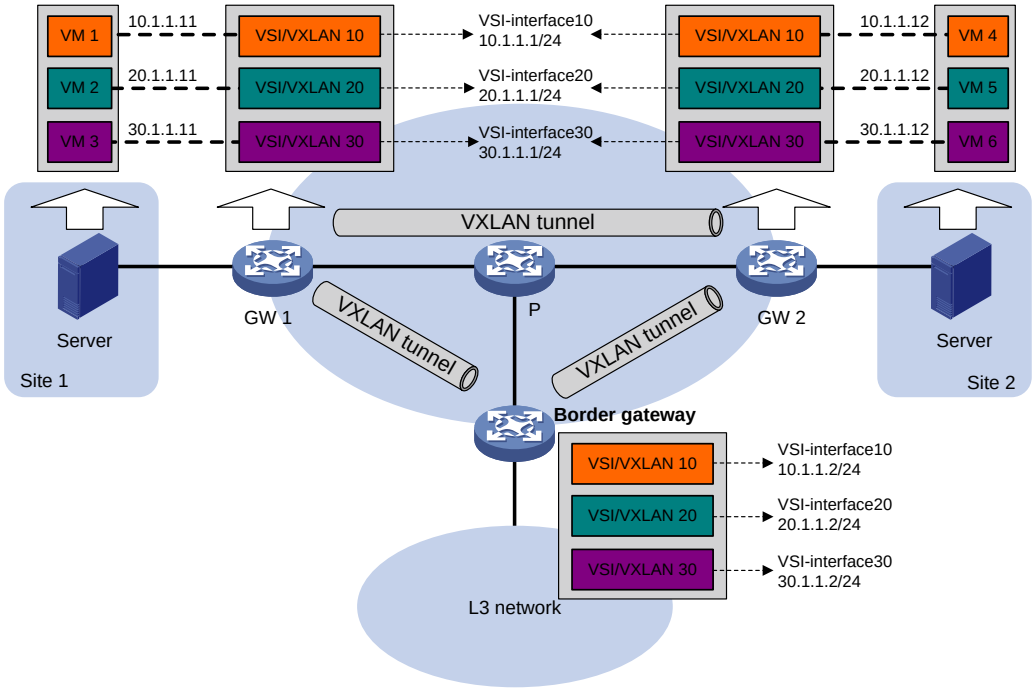


Figure 20 shows an example of distributed VXLAN IP gateway deployment. Create VSI interfaces on each distributed VXLAN IP gateway and the border gateway as gateway interfaces. Assign the same IP address to the same VSI interface on the distributed VXLAN IP gateways. Enable one of the following features on a distributed VXLAN IP gateway:

- ARP or ND flood suppression. The gateway performs Layer 2 forwarding based on MAC address entries and performs Layer 3 forwarding based on ARP or ND entries.
- Local proxy ARP or local ND proxy. The gateway performs Layer 3 forwarding based on ARP or ND entries. The following sections use distributed VXLAN IP gateways enabled with the local proxy ARP or local ND proxy feature to describe the forwarding processes for intra-VXLAN traffic, inter-VXLAN traffic, and traffic from a VXLAN to an external network.

A distributed VXLAN IP gateway can generate ARP or ND entries by a variety of methods. The following sections use dynamically learned ARP or ND entries to describe the forwarding processes.

Figure 20 Example of distributed VXLAN IP gateway deployment



Intra-VXLAN traffic forwarding between sites

As shown in Figure 21, the network uses the following process to forward traffic in a VXLAN between sites (for example, from VM 1 to VM 4 in VXLAN 10):

1. VM 1 sends an ARP request to obtain the MAC address of VM 4.
2. GW 1 performs the following operations:
   a. Creates an ARP entry for VM 1 and replies with the MAC address of VSI-interface 10 (the gateway interface for VXLAN 10).
   b. Replaces the sender MAC address of the ARP request with the MAC address of VSI-interface 10, and then floods the request to all sites in VXLAN 10.
3. VM 1 creates an ARP entry for VM 4. The MAC address in the entry is the MAC address of VSI-interface 10 on GW 1.
4. GW 2 (the VTEP for VM 4) performs the following operations:
   a. De-encapsulates the ARP request and creates an ARP entry for VM 1. The entry contains VM 1's IP address (10.1.1.11), the MAC address of VSI-interface 10 on GW 1, and the incoming tunnel interface.
   b. Replaces the sender MAC address of the request with the MAC address of VSI-interface 10 on GW 2, and then floods the request to the local site in VXLAN 10.
5. VM 4 creates an ARP entry for VM 1, and then sends a reply to GW 2. The MAC address in the ARP entry is the MAC address of VSI-interface 10 on GW 2.
6. GW 2 performs the following operations:
   a. Creates an ARP entry for VM 4.
   b. Replaces the sender MAC address of the reply with the MAC address of VSI-interface 10 on GW 2, and sends the reply to GW 1.
7. GW 1 de-encapsulates the ARP reply and creates an ARP entry for VM 4. The entry contains VM 4's IP address (10.1.1.12), the MAC address of VSI-interface 10 on GW 2, and the incoming tunnel interface.
8. For subsequent traffic between VM 1 and VM 4, GW 1 and GW 2 use their respective ARP tables to make the forwarding decision.

Figure 21 Example of distributed VXLAN IP gateway deployment

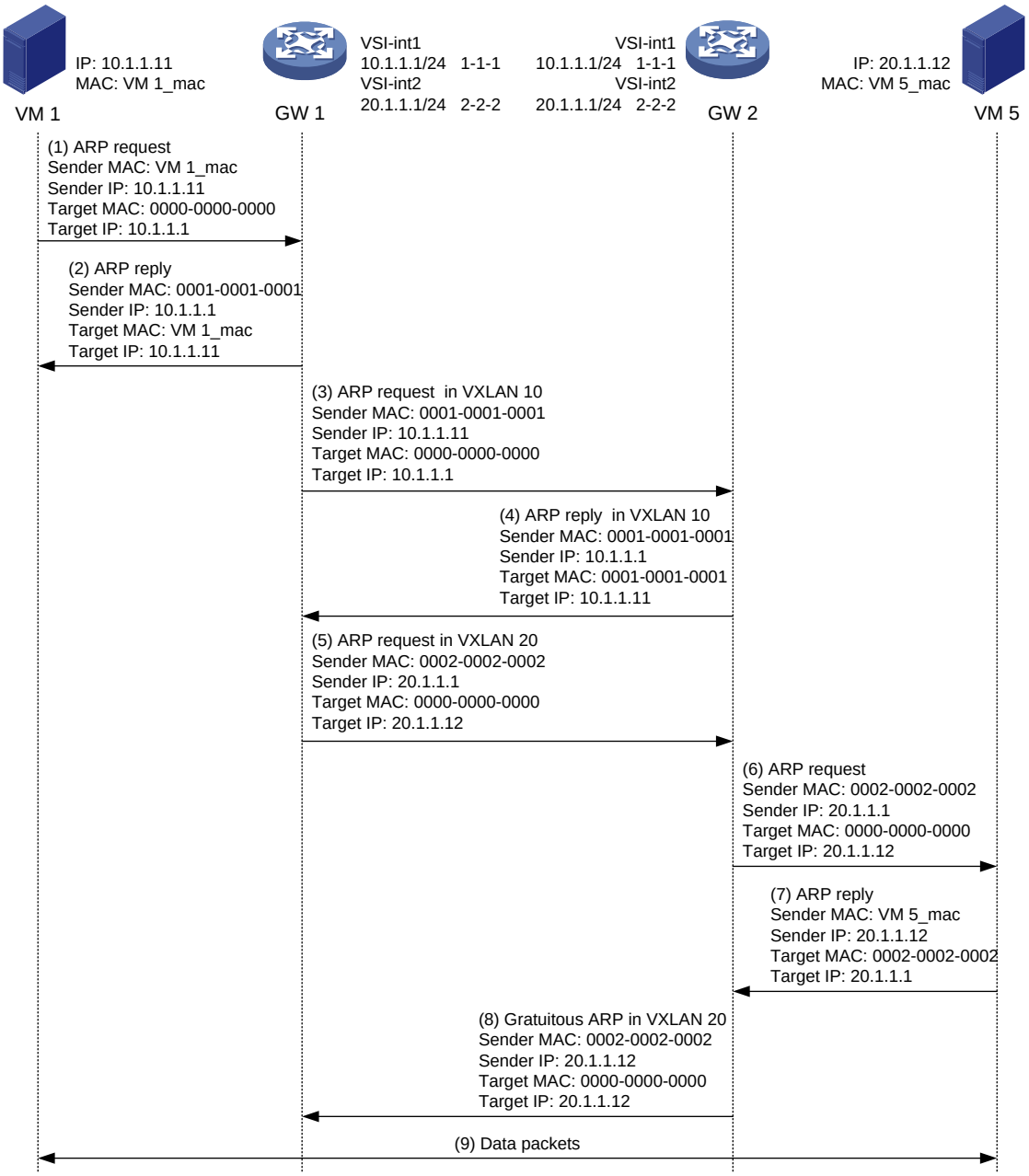
9.

Inter-VXLAN traffic forwarding between sites

As shown in Figure 22, the network uses the following process to forward traffic between VXLANs (for example, from VM 1 in VXLAN 10 to VM 5 in VXLAN 20):

1. VM 1 sends an ARP request to obtain the MAC address of the gateway at 10.1.1.1.
2. GW 1 creates an ARP entry for VM 1 and replies with the MAC address of VSI-interface 10 (the gateway interface for VXLAN 10) so VM 1 will send the packets destined for VM 5 to GW 1.
3. GW 1 sends an ARP request to the local and remote sites in VXLAN 10. In the ARP request, the sender IP address is 10.1.1.11, and the sender MAC address is the MAC address of VSI-interface 10 on GW 1.
4. GW 2 performs the following operations:
   a. De-encapsulates the ARP request and creates an ARP entry for VM 1. The entry contains IP address 10.1.1.11 and MAC address of VSI-interface 10 on GW 1, and the incoming tunnel interface.
   b. Replaces the sender MAC address of the request with the MAC address of VSI-interface 10 on GW 2, and then floods the request to the local site in VXLAN 10.
   c. Sends an ARP reply to GW 1. The reply contains IP address 10.1.1.1 and MAC address of VSI-interface 10 on GW 2).
5. When sending an ARP request in VXLAN 10, GW 1 also sends an ARP request to the local and remote sites in VXLAN 20 to obtain the MAC address of VM 5. In the ARP request, the sender IP address is 20.1.1.1, and the sender MAC address is the MAC address of VSI-interface 20 on GW 1.
6. GW 2 de-encapsulates the ARP request of VXLAN 20, replaces the sender MAC address of the request with the MAC address of VSI-interface 20 on GW 2, and then floods the request to the local site in VXLAN 20.
7. VM 5 creates an ARP entry for GW 2, and then sends a reply to GW 2. The entry contains IP address 20.1.1.1 and MAC address of VSI-interface 20 on GW 2.
8. GW 2 performs the following operations:
   a. Creates an ARP entry for VM 5.
   b. Sends a gratuitous ARP packet to the local and remote sites. In the packet, the sender IP address is 20.1.1.12, and the sender MAC address is the MAC address of VSI-interface 20 on GW 2.
9. GW 1 de-encapsulates the gratuitous ARP packet and creates an ARP entry for VM 5. The entry contains VM 5's IP address 20.1.1.12, the MAC address of VSI-interface 20 on GW 2, and the incoming tunnel interface.
10. For subsequent traffic between VM 1 and VM 5, GW 1 and GW 2 use their respective ARP tables to make the forwarding decision.

Figure 22 Inter-VXLAN traffic forwarding between sites



VXLAN-to-external network traffic forwarding

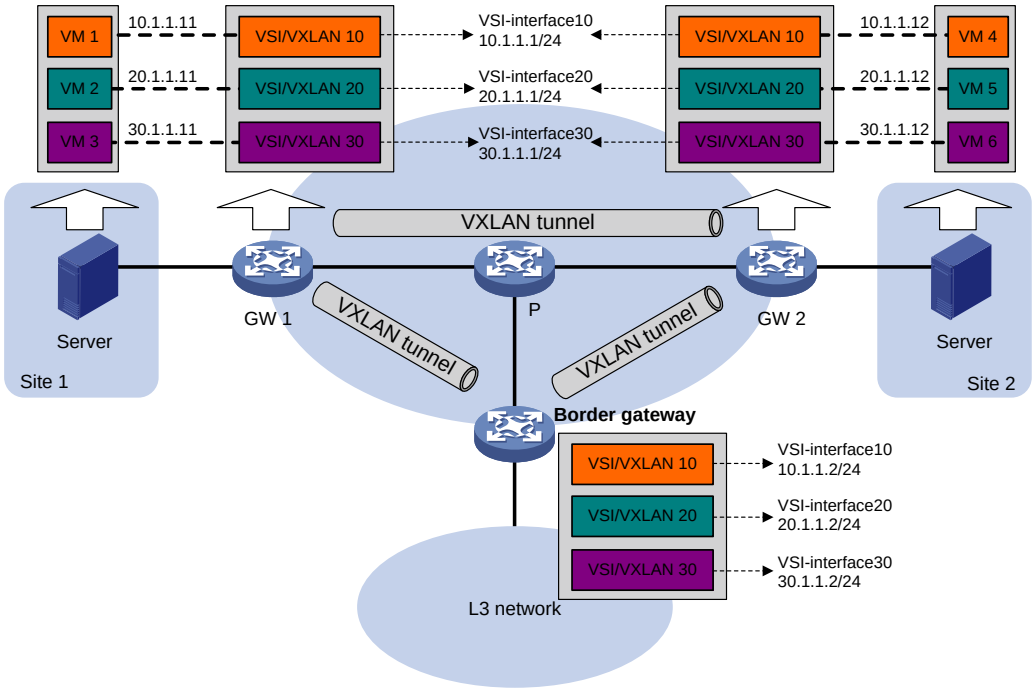
As shown in Figure 23, the network uses the following process to forward traffic from a VXLAN to the Layer 3 network (for example, from VM 1 to the host at 50.1.1.1):

1. VM 1 sends an ARP request to obtain the MAC address of the gateway at 10.1.1.1.
2. GW 1 creates an ARP entry for VM 1 and replies with the MAC address of VSI-interface 10 (the gateway interface for VXLAN 10).
3. VM 1 sends a packet destined for the host to GW 1.
4. GW 1 performs the following operations:
   a. Searches the IP routing policies or routing table for the next hop. In this example, the next hop for the packet is 10.1.1.2 (the border gateway).

b. Floods an ARP request to the local and remote sites in VXLAN 10 to obtain the MAC address of 10.1.1.2.

5. The border gateway de-encapsulates the ARP request, creates an ARP entry for GW 1, and tunnels a reply to GW 1.
6. GW 1 de-encapsulates the ARP reply and creates an ARP entry for 10.1.1.2.
7. GW 1 sends the packet destined for the host to the border gateway.
8. The border gateway de-encapsulates the packet and forwards it to the host.

Figure 23 Example of distributed VXLAN IP gateway deployment



VXLAN IP gateway tasks at a glance

To configure a VXLAN IP gateway, perform the following tasks:

1. Configure a VXLAN IP gateway

   Choose one of the following tasks:

   - Configuring a centralized VXLAN IP gateway
   - Configuring a distributed VXLAN IP gateway
2. (Optional.) Managing ARP entries and ND entries
3. (Optional.) Configuring a VSI interface
4. (Optional.) Configuring VXLAN M-LAG

Prerequisites for VXLAN IP gateway configuration

Before you configure a centralized or distributed VXLAN IP gateway, you must perform the following tasks on VTEPs:

- Create VSIs and VXLANs.
- Configure VXLAN tunnels and assign them to VXLANs.

Configuring a centralized VXLAN IP gateway

Restrictions and guidelines

Do not execute the **local-proxy-arp enable** command on a centralized VXLAN IP gateway.

Configuring a gateway interface on a centralized VXLAN IP gateway

1. Enter system view.
   system-view
2. Create a VSI interface and enter VSI interface view.
   interface vsi-interface *vsi-interface-id*
3. Assign an IP address to the VSI interface.
   IPv4:
   ip address *ipv4-address* { *mask* | *mask-length* }
   IPv6:
   For information about assigning an IPv6 address to an interface, see IPv6 basics in *Layer 3—IP Services Configuration Guide*.
   By default, no IP address is assigned to a VSI interface.
4. Return to system view.
   quit
5. Enter VSI view.
   vsi *vsi-name*
6. Specify a gateway interface for the VSI.
   gateway vsi-interface *vsi-interface-id*
   By default, no gateway interface is specified for a VSI.

Assigning a subnet to a VSI

About this task

Perform this task on VSIs that share a gateway interface. This task enables the VSI interface to identify the VSI of a packet.

You can assign a maximum of eight IPv4 and IPv6 subnets to a VSI. Make sure these subnets are on the same network as one of the IP addresses on the gateway interface.

For VSIs that share a gateway interface, the subnets must be unique.

If you remove the gateway interface from the VSI, the VSI's subnet settings are automatically deleted.

Restrictions and guidelines

This task does not prevent a VSI from learning ARP or ND entries for the subnets not assigned to the VSI. If the VSI shares a VSI interface with other VSIs, it can learn ARP or ND entries for the subnets assigned to those VSIs.

Procedure

1. Enter system view.

system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Assign a subnet to the VSI.

   gateway subnet { *ipv4-address wildcard-mask* | *ipv6-address prefix-length* }

   By default, no subnet exists on a VSI.

Configuring a distributed VXLAN IP gateway

Restrictions and guidelines for distributed VXLAN IP gateway configuration

For a VXLAN that requires access to the external network, specify the VXLAN's VSI interface on the border gateway as the next hop by using one of the following methods:

- Configure a static route.
- Configure a routing policy, and apply the policy by using the **apply default-next-hop** command. For more information about configuring routing policies, see routing policy configuration in *Layer 3—IP Routing Configuration Guide*.

If both ARP (or ND) flood suppression and local proxy ARP (or local ND proxy) are enabled on a distributed VXLAN IP gateway, only ARP (or ND) flood suppression takes effect. As a best practice, do not use these features together on distributed VXLAN IP gateways. For more information about ARP (or ND) flood suppression, see "Enabling ARP flood suppression" or "Enabling ND flood suppression."

Make sure a VSI interface uses the same MAC address to provide service on distributed VXLAN IP gateways connected to IPv4 sites. Make sure a VSI interface uses different link-local addresses to provide service on distributed VXLAN IP gateways connected to both IPv4 and IPv6 sites.

Configuring a gateway interface on a distributed VXLAN IP gateway

1. Enter system view.

   system-view

2. Create a VSI interface and enter VSI interface view.

   interface vsi-interface *vsi-interface-id*

3. Assign an IP address to the VSI interface.

   IPv4:

   ip address *ip-address* { *mask* | *mask-length* } [**sub**]

   IPv6:

   See IPv6 basics in *Layer 3—IP Services Configuration Guide*.

   By default, no IP address is assigned to a VSI interface.

4. Specify the VSI interface as a distributed gateway.

   distributed-gateway local

   By default, a VSI interface is not a distributed gateway.

5. Enable local proxy ARP or local ND proxy.

IPv4:

local-proxy-arp enable [**ip-range** *startIP* **to** *endIP*]

By default, local proxy ARP is disabled.

For more information about this command, see proxy ARP commands in *Layer 3—IP Services Command Reference*.

IPv6:

local-proxy-nd enable

By default, local ND proxy is disabled.

For more information about this command, see IPv6 basics commands in *Layer 3—IP Services Command Reference*.

6. Bring up the VSI interface.

   undo shutdown

   By default, a VSI interface is up.

7. Return to system view.

   quit

8. Enter VSI view.

   vsi *vsi-name*

9. Specify the VSI interface as the gateway interface for the VSI.

   gateway vsi-interface *vsi-interface-id*

   By default, no gateway interface is specified for a VSI.

Enabling dynamic ARP or ND entry synchronization for distributed VXLAN IP gateways

About this task

When local proxy ARP or local ND proxy is enabled on distributed VXLAN IP gateways, enable this feature for all gateways to have the same ARP or ND entries.

A controller or the EVPN feature can also synchronize ARP or ND entries among distributed VXLAN IP gateways. When you use a controller or the EVPN feature, do not enable dynamic ARP or ND entry synchronization.

Enabling dynamic ARP entry synchronization

1. Enter system view.

   system-view

2. Enable dynamic ARP entry synchronization for distributed VXLAN IP gateways.

   arp distributed-gateway dynamic-entry synchronize

   By default, dynamic ARP entry synchronization is disabled for distributed VXLAN IP gateways.

Enabling dynamic ND entry synchronization

1. Enter system view.

   system-view

2. Enable dynamic ND entry synchronization for distributed VXLAN IP gateways.

   ipv6 nd distributed-gateway dynamic-entry synchronize

   By default, dynamic ND entry synchronization is disabled for distributed VXLAN IP gateways.

Assigning a subnet to a VSI

About this task

Perform this task on VSIs that share a gateway interface. This task enables the VSI interface to identify the VSI of a packet.

You can assign a maximum of eight IPv4 and IPv6 subnets to a VSI. Make sure these subnets are on the same network as one of the IP addresses on the gateway interface.

For VSIs that share a gateway interface, the subnets must be unique.

If you remove the gateway interface from the VSI, the VSI's subnet settings are automatically deleted.

Restrictions and guidelines

This task does not prevent a VSI from learning ARP or ND entries for the subnets not assigned to the VSI. If the VSI shares a VSI interface with other VSIs, it can learn ARP or ND entries for the subnets assigned to those VSIs.

Procedure

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Assign a subnet to the VSI.

   gateway subnet { *ipv4-address wildcard-mask* | *ipv6-address prefix-length* }

   By default, no subnet exists on a VSI.

Managing ARP entries and ND entries

Adding a static ARP entry

About this task

A VXLAN IP gateway can dynamically learn ARP entries and use manually configured static ARP entries.

Procedure

1. Enter system view.

   system-view

2. Add a static local-ARP entry.

   arp static *ip-address mac-address* **vsi-interface** *vsi-interface-id interface-type interface-number* **service-instance** *instance-id* **vsi** *vsi-name* [**vpn-instance** *vpn-instance-name*]

   For more information about this command, see ARP commands in *Layer 3—IP Services Command Reference*.

3. Add a static remote-ARP entry.

   arp static *ip-address mac-address* **vsi-interface** *vsi-interface-id* **tunnel** *number* **vsi** *vsi-name* [**vpn-instance** *vpn-instance-name*]

   For more information about this command, see ARP commands in *Layer 3—IP Services Command Reference*.

Disabling remote ARP or ND learning for VXLANs

About this task

By default, the device learns ARP or ND information of remote user terminals from packets received on VXLAN tunnel interfaces. To save resources on VTEPs in an SDN transport network, you can temporarily disable remote ARP or ND learning when the controller and VTEPs are synchronizing entries. After the entry synchronization is completed, enable remote ARP or ND learning.

Restrictions and guidelines

This feature takes effect only on manually created VXLAN tunnels. Remote ARP/ND learning is always disabled for automatically created tunnels, and it is not controlled by this command.

As a best practice, disable remote ARP or ND learning for VXLANs only when the controller and VTEPs are synchronizing entries.

Procedure

1. Enter system view.
   system-view
2. Disable remote ARP learning.
   vxlan tunnel arp-learning disable
   By default, remote ARP learning is enabled for VXLANs.
3. Disable remote ND learning.
   vxlan tunnel nd-learning disable
   By default, remote ND learning is enabled for VXLANs.

Configuring a VSI interface

Configuring optional parameters for a VSI interface

1. Enter system view.
   system-view
2. Enter VSI interface view.
   interface vsi-interface *vsi-interface-id*
3. Assign a MAC address to the VSI interface.
   mac-address *mac-address*
   By default, VSI interfaces use the MAC address of VLAN-interface 4094 on the device.
4. Configure the description of the VSI interface.
   description *text*
   The default description of a VSI interface is *interface-name* plus **Interface** (for example, **Vsi-interface100 Interface**).
5. Set the MTU for the VSI interface.
   mtu *size*
   The default MTU is 1444 bytes.
6. Set the expected bandwidth for the VSI interface.
   bandwidth *bandwidth-value*
   The default expected bandwidth (in kbps) equals the interface baudrate divided by 1000.

The expected bandwidth is an informational parameter used only by higher-layer protocols for calculation. You cannot adjust the actual bandwidth of an interface by using this command.

7. Set an ARP packet sending rate limit for the VSI interface.

   arp send-rate *pps*

   By default, the ARP packet sending rate is not limited for a VSI interface.

   This limit applies only to the ARP request packets sourced from the VSI interface.

Restoring the default settings of the VSI interface

Restrictions and guidelines

CAUTION:

This operation might interrupt ongoing network services. Make sure you are fully aware of the impact of this operation when you perform it on a live network.

This operation might fail to restore the default settings for some commands for reasons such as command dependencies or system restrictions. Use the **display this** command in interface view to identify these commands. Use their **undo** forms or follow the command reference to restore their default settings. If your restoration attempt still fails, follow the error message instructions to resolve the problem.

Procedure

1. Enter system view.

   system-view

2. Enter VSI interface view.

   interface vsi-interface *vsi-interface-id*

3. Restore the default settings of the VSI interface.

   default

Configuring VXLAN M-LAG

Configuring VXLAN M-LAG

About this task

VXLAN M-LAG virtualizes two VTEPs or VXLAN IP gateways into one M-LAG system to avoid single points of failure.

Restrictions and guidelines

When you configure VXLAN M-LAG, follow these restrictions and guidelines:

- An M-LAG system does not support Layer 2 multicast.
- M-LAG member devices must use the same tunnel source address to establish VXLAN tunnels to remote devices.

In a VXLAN M-LAG network, when overlay Layer 3 packets encapsulated by VXLAN are forwarded through a peer link, VLAN-interface 1 must be created on the M-LAG member devices.

If a direct peer link is used, follow these restrictions:

- If the frame match criteria of dynamic ACs on the peer link are created based on site-facing Ethernet service instances, you can configure only the following criteria for site-facing Ethernet service instances:

- **encapsulation s-vid** { *vlan-id* | *vlan-id-list* }
- **encapsulation untagged**

In addition, you must set the access mode to VLAN for site-facing Ethernet service instances.

- If the frame match criteria of dynamic ACs on the peer link are created based on VXLAN IDs, you must set the access mode to VLAN for all ACs.
- Make sure the Ethernet service instances that use the same match criterion are mapped to the same VSI.
- Make sure the following settings are consistent on the M-LAG member devices:
  - Ethernet service instances and their match criterion on the M-LAG interfaces in the same M-LAG group.
  - VXLAN IDs of VSIs.

  In addition, the Ethernet service instances must be created manually.

By default, the up/down state of a VSI interface mapped to a VSI remains unaffected whether the VSI is mapped to an Ethernet service instance or not in an M-LAG network. If a VSI is not mapped to an Ethernet service instance, or the port of the mapped Ethernet service instance is down, the VSI interface has no actual traffic requirement. If it stays in up state, traffic forwarding mistakes might occur. To resolve this issue, do not create a VXLAN on the VSI mapped to a single-homed AC, and only map the VSI to a VSI interface and execute the **l2vpn m-lag peer-link ac-match-rule vxlan-mapping** command. In this way, the VSI interface mapped to the VSI can come up only when the VSI is mapped to an Ethernet service instance and the port hosting the Ethernet service instance is up. This mechanism applies only to single-homed AC scenarios.

The VTEPs in an M-LAG system synchronize local MAC address entries with each other over the peer link. However, they do not synchronize MAC address entry deletions. When you delete a MAC address entry from one VTEP, the other VTEP retains the entry that contains the same MAC address until the entry ages out.

Procedure

Configure M-LAG as described in *Layer 2—LAN Switching Configuration Guide*.

1. Exclude interfaces from the shutdown action by M-LAG MAD.

   m-lag mad exclude interface *interface-type interface-number*

   By default, M-LAG MAD shuts down all network interfaces when detecting a multi-active collision, except for the network interfaces set by the system to not shut down.

   Exclude all interfaces used by VXLAN from the MAD shutdown action by M-LAG. The interfaces include VSI interfaces, interfaces used for setting up the keepalive link, and transport-facing outgoing interfaces of VXLAN tunnels.

   For more information about this command, see M-LAG commands in *Layer 2—LAN Switching Command Reference*.

2. Set the data restoration interval.

   m-lag restore-delay *value*

   By default, the data restoration interval is 30 seconds.

   Set the data restoration interval to a value equal to or larger than 180 seconds on a VXLAN M-LAG system.

   For more information about this command, see M-LAG commands in *Layer 2—LAN Switching Command Reference*.

3. Configure VXLAN as described in "Configuring basic VXLAN features" and "Configuring VXLAN IP gateways."

4. (Optional.) Enable the device to create frame match criteria based on VXLAN IDs for the dynamic ACs on the direct peer link.

   l2vpn m-lag peer-link ac-match-rule vxlan-mapping

By default, on an M-LAG system with a direct peer link, dynamic ACs on the peer link use frame match criteria that are identical to those of site-facing ACs.

If you do not execute this command, do not configure overlapping outer VLAN IDs for Ethernet service instances of different VSIs.

If you execute this command, do not create VXLANs with IDs larger than 16000000.

For more information about this command, see *EVPN Command Reference*.

Enabling an interface to join a multicast group

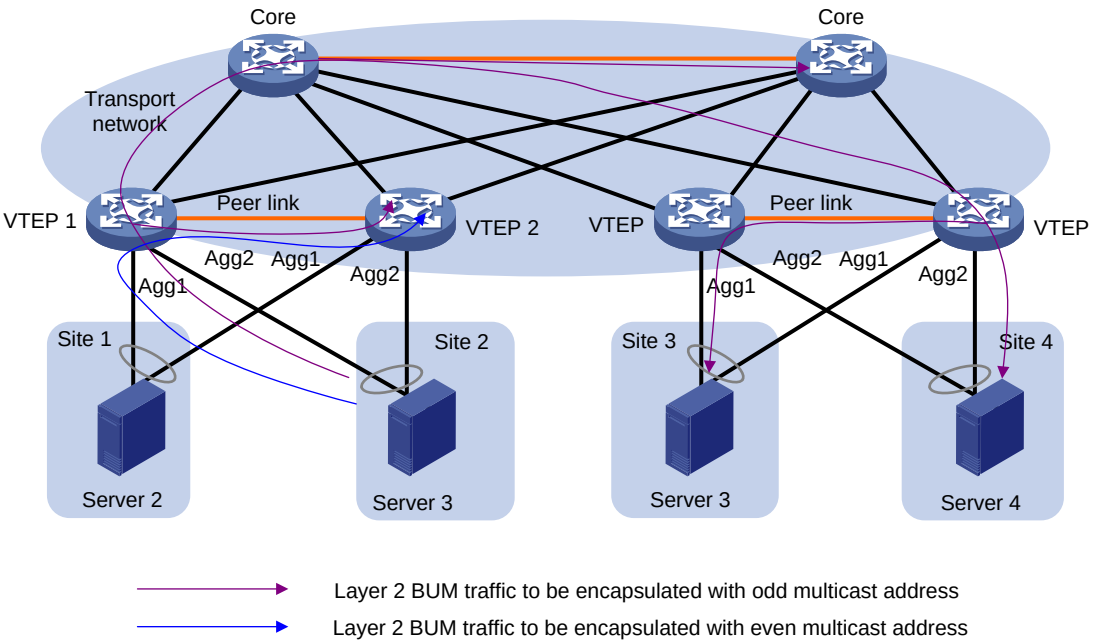
About this task

As shown in Figure 24, in the VXLAN+M-LAG scenario, VXLAN uses multicast routes to forward BUM traffic. An M-LAG member device VTEP 1 forwards packets with odd-numbered multicast addresses (the destination multicast addresses encapsulated in VXLAN packets), and the other M-LAG member device VTEP 2 forwards packets with even-numbered multicast addresses.

After VTEP 1 receives Layer 2 BUM traffic, VTEP 1 synchronizes it to VTEP 2 through the peer link. At the same time, VTEP 1 adds VXLAN encapsulation to the BUM traffic.

- If VTEP 1 assigns an odd multicast address as the destination address in the VXLAN encapsulation of the traffic, VTEP 1 can forward the traffic normally.
- If VTEP 1 assigns an even multicast address as the destination address in the VXLAN encapsulation of the traffic, VTEP 1 will not forward the traffic through the VXLAN tunnel, following the even/odd address-based forwarding principle. Although VTEP 2 receives the BUM traffic from the peer link, VTEP 2 does not forward the BUM traffic through the VXLAN tunnel because BUM traffic received from the peer link is not sent to the VXLAN tunnel. As a result, traffic forwarding fails.

Figure 24 Traffic forwarding issues with multicast-mode VXLAN support for M-LAG



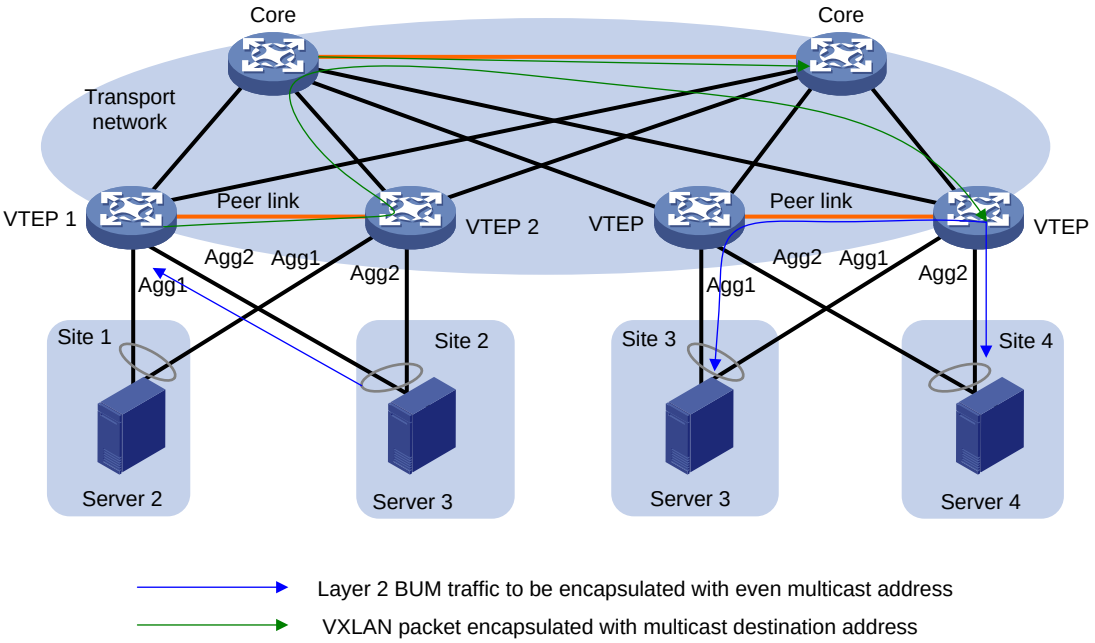
To address the traffic forwarding failure issue, use this feature to synchronize encapsulated multicast VXLAN packets between M-LAG member devices through the peer link.

After you assign the peer link to a VLAN and enable the VLAN interface for that VLAN to join a multicast group, the VLAN interface will automatically join the multicast group specified by using the **group** command in VXLAN view. The peer link will then allow forwarding traffic for that multicast group. In this way, M-LAG member devices can synchronize VXLAN packets encapsulated with

multicast addresses through the peer link. After you enable an interface to join a multicast group, if the local M-LAG member device receives a multicast VXLAN packet from the peer M-LAG member device through the peer link, it will not decapsulate the multicast VXLAN packet and send it to the local AC. Instead, it will send the packet to the remote VTEP through the transport network.

As shown in Figure 25, this configuration allows VTEP 1 to synchronize the encapsulated multicast VXLAN packets to VTEP 2 through the peer link. VTEP 2 then forwards these packets to the transport network, resolving the traffic forwarding failure issue.

Figure 25 Schematic diagram for an interface joining a multicast group



Restrictions and guidelines

On the same device, you can enable only one VLAN interface to join a multicast group. As a best practice to avoid traffic interference, do not deploy other services on that VLAN interface.

You cannot enable a VLAN interface bound to a VPN instance to join a multicast group.

Procedure

1. Enter system view.
   `system-view`
2. Enter VLAN interface view.
   `interface vlan-interface` *`vsi-interface-id`*
3. Enable the interface to join a multicast group.
   `vxlan multicast-group member enable`
   By default, an interface is disabled from joining a multicast group.

Display and maintenance commands for VXLAN IP gateways

Execute **`display`** commands in any view and **`reset`** commands in user view.

Task	Command
Display L2VPN ARP entries on an M-LAG member device.	**display l2vpn m-lag arp** [**local** | **remote**] [**count** | { **public-instance** | **vpn-instance** *vpn-instance-name* } [**count** | *ip-address*]]
Display L2VPN ARP flood suppression entries on an M-LAG member device.	**display l2vpn m-lag arp suppression** [**local** | **remote**] [**count** | **vsi** *vsi-name* [**count** | *ip-address*]]
Display L2VPN MAC address entries on an M-LAG member device.	**display l2vpn m-lag mac-address** [**local** | **remote**] [**count** | **vsi** *vsi-name* [**count** | *mac-address*]]
Display L2VPN ND entries on an M-LAG member device.	**display l2vpn m-lag nd** [**local** | **remote**] [**count** | { **public-instance** | **vpn-instance** *vpn-instance-name* } [**count** | *ipv6-address*]]
Display L2VPN ND flood suppression entries on an M-LAG member device.	**display l2vpn m-lag nd suppression** [**local** | **remote**] [**count** | **vsi** *vsi-name* [**count** | *ipv6-address*]]
Display information about VSI interfaces.	**display interface** [**vsi-interface** [*vsi-interface-id*]] [**brief** [**description** | **down**]]
Clear statistics on VSI interfaces.	**reset counters interface** [**vsi-interface** [*vsi-interface-id*]]
Clear MAC address entries synchronized from the peer in an M-LAG system.	**reset l2vpn m-lag mac-address** [**vsi** *vsi-name*]

VXLAN IP gateway configuration examples

Example: Configuring a centralized VXLAN IP gateway

Network configuration

As shown in Figure 26:

- Configure VXLAN 10 as a unicast-mode VXLAN on Switch A, Switch B, and Switch C to provide connectivity for the VMs across the network sites.
- Configure a centralized VXLAN IP gateway on Switch B to provide gateway services for VXLAN 10.
- Manually establish VXLAN tunnels and assign the tunnels to VXLAN 10.
- Enable remote-MAC address learning.

Figure 26 Network diagram

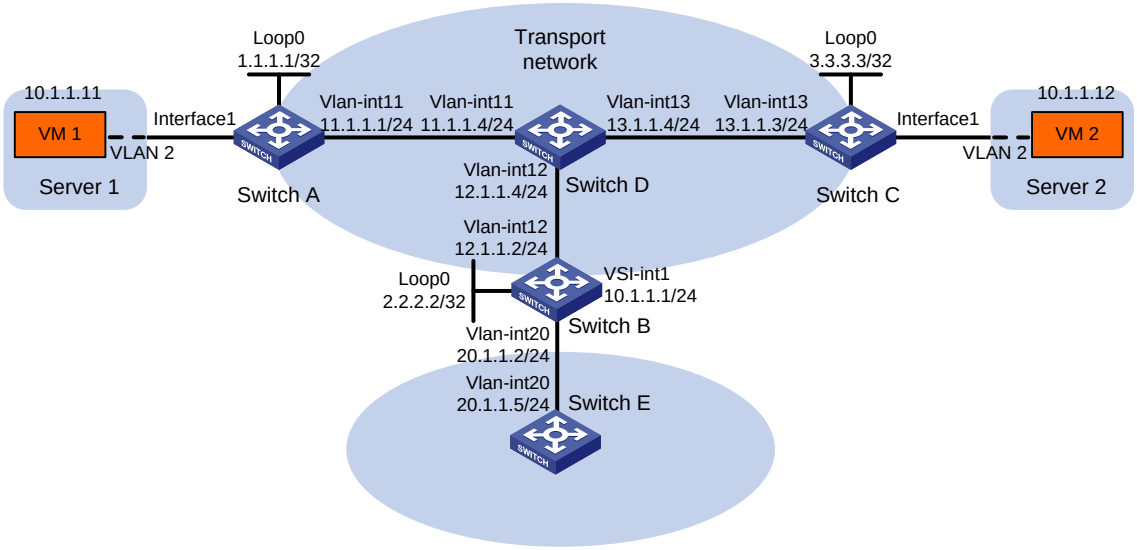


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1.
2. Create VLANs and VLAN interfaces on all devices. (Details not shown.)
3. On VM 1 and VM 2, specify 10.1.1.1 as the gateway address. (Details not shown.)
4. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 26. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D). (Details not shown.)

   # Configure OSPF to advertise routes to networks 10.1.1.0/24 and 20.1.1.0/24 on Switch B and Switch E. (Details not shown.)
5. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Create VSI **vpna** and VXLAN 10.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] vxlan 10
   [SwitchA-vsi-vpna-vxlan-10] quit
   [SwitchA-vsi-vpna] quit
   ```

   # Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch B and Switch C.

   ```
   [SwitchA] interface loopback 0
   [SwitchA-Loopback0] ip address 1.1.1.1 255.255.255.255
   [SwitchA-Loopback0] quit
   ```

   # Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 1**.

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 1.1.1.1
[SwitchA-Tunnel1] destination 2.2.2.2
[SwitchA-Tunnel1] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 2**.

```
[SwitchA] interface tunnel 2 mode vxlan
[SwitchA-Tunnel2] source 1.1.1.1
[SwitchA-Tunnel2] destination 3.3.3.3
[SwitchA-Tunnel2] quit
```

Assign Tunnel 1 and Tunnel 2 to VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] tunnel 1
[SwitchA-vsi-vpna-vxlan-10] tunnel 2
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

6. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch C.

```
[SwitchB] interface loopback 0
[SwitchB-Loopback0] ip address 2.2.2.2 255.255.255.255
[SwitchB-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 2**.

```
[SwitchB] interface tunnel 2 mode vxlan
[SwitchB-Tunnel2] source 2.2.2.2
[SwitchB-Tunnel2] destination 1.1.1.1
[SwitchB-Tunnel2] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 3**.

```
[SwitchB] interface tunnel 3 mode vxlan
[SwitchB-Tunnel3] source 2.2.2.2
[SwitchB-Tunnel3] destination 3.3.3.3
```

```
[SwitchB-Tunnel3] quit
```

Assign Tunnel 2 and Tunnel 3 to VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] tunnel 2
[SwitchB-vsi-vpna-vxlan-10] tunnel 3
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create VSI-interface 1 and assign the interface an IP address. The IP address will be used as the gateway address for VXLAN 10.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchB-Vsi-interface1] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
[SwitchB-vsi-vpna] quit
```

7. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch B.

```
[SwitchC] interface loopback 0
[SwitchC-Loopback0] ip address 3.3.3.3 255.255.255.255
[SwitchC-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 1**.

```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 3.3.3.3
[SwitchC-Tunnel1] destination 1.1.1.1
[SwitchC-Tunnel1] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 3**.

```
[SwitchC] interface tunnel 3 mode vxlan
[SwitchC-Tunnel3] source 3.3.3.3
[SwitchC-Tunnel3] destination 2.2.2.2
[SwitchC-Tunnel3] quit
```

Assign Tunnel 1 and Tunnel 3 to VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] tunnel 1
[SwitchC-vsi-vpna-vxlan-10] tunnel 3
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchC-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchC-GigabitEthernet1/0/1-srv1000] quit
[SwitchC-GigabitEthernet1/0/1] quit
```

Verifying the configuration

1. Verify the VXLAN IP gateway settings on Switch B:

Verify that the VXLAN tunnel interfaces are up on Switch B.

```
[SwitchB] display interface tunnel 2
Tunnel2
Current state: UP
Line protocol state: UP
Description: Tunnel2 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 2.2.2.2, destination 1.1.1.1
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that VSI-interface 1 is up.

```
[SwitchB] display interface vsi-interface brief
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP      Description
Vsi1                 UP   UP       10.1.1.1
```

Verify that the VXLAN tunnels have been assigned to the VXLAN, and VSI-interface 1 is the gateway interface of VSI **vpna**.

```
[SwitchB] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
```

```
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name           Link ID    State  Type        Flood proxy
    Tunnel2               0x5000002  Up     Manual      Disabled
    Tunnel3               0x5000003  Up     Manual      Disabled
```

Verify that Switch B has created ARP entries for the VMs. (Details not shown.)

Verify that Switch B has created FIB entries for the VMs.

```
[SwitchB] display fib 10.1.1.11
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination/Mask   Nexthop         Flag     OutInterface/Token       Label
10.1.1.11/32       10.1.1.11       UH       Vsi1                     Null
```

2. Verify that the VMs can access the WAN:

Verify that VM 1 and VM 2 can ping each other. (Details not shown.)

Verify that VM 1, VM 2, and VLAN-interface 20 (20.1.1.5) on Switch E can ping each other. (Details not shown.)

Example: Configuring distributed VXLAN IP gateways (IPv4 underlay)

Network configuration

As shown in Figure 27:

- Configure VXLAN 10, VXLAN 20, and VXLAN 30 as unicast-mode VXLANs on Switch A, Switch B, and Switch C to provide connectivity for the VMs across the network sites.
- Manually establish VXLAN tunnels and assign the tunnels to the VXLANs.
- Configure distributed VXLAN IP gateways on Switch A and Switch C to forward traffic between the VXLANs.
- Configure Switch B as a border gateway to forward traffic between the VXLANs and the WAN connected to Switch E.

NOTE:

This example provides configuration of IPv4 sites. The configuration procedure does not differ between IPv4 and IPv6 sites.

Figure 27 Network diagram

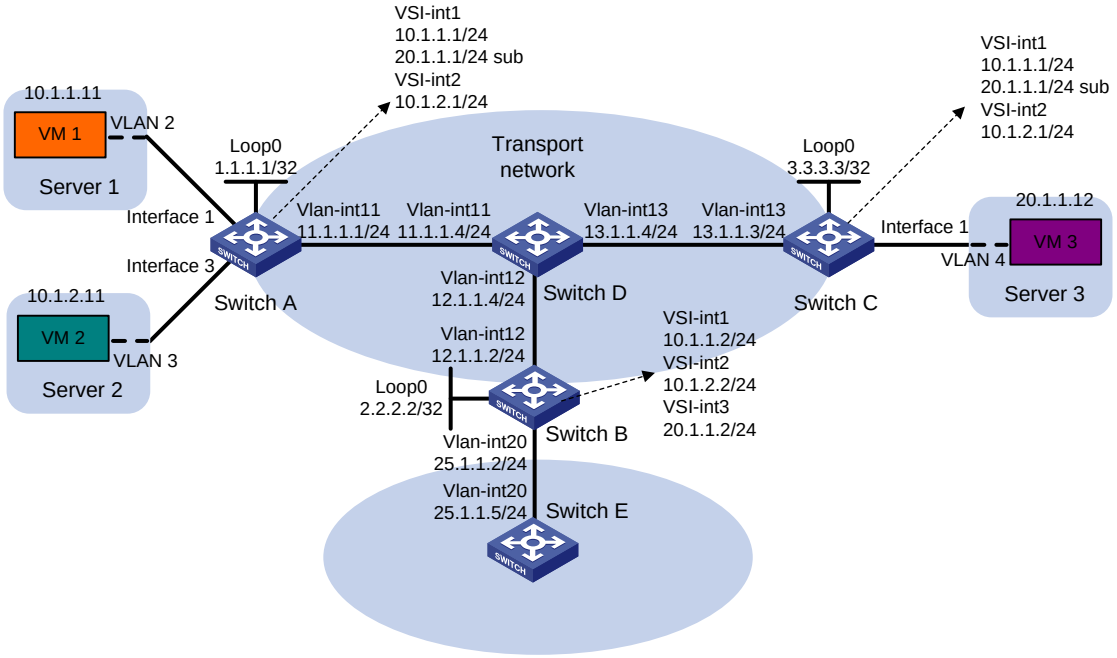


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface3	GigabitEthernet1/0/3

Procedure

1.
2. Create VLANs and VLAN interfaces on all devices. (Details not shown.)
3. On VM 1, VM 2, and VM 3, specify 10.1.1.1, 10.1.2.1, and 20.1.1.1 as the gateway address, respectively. (Details not shown.)
4. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 27. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D). (Details not shown.)

   # Configure OSPF to advertise routes to networks 10.1.1.0/24, 10.1.2.0/24, 20.1.1.0/24, and 25.1.1.0/24 on Switch B and Switch E. (Details not shown.)
5. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Create VSI **vpna** and VXLAN 10.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] vxlan 10
   [SwitchA-vsi-vpna-vxlan-10] quit
   [SwitchA-vsi-vpna] quit
   ```

   # Create VSI **vpnb** and VXLAN 20.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

Create VSI **vpnc** and VXLAN 30.

```
[SwitchA] vsi vpnc
[SwitchA-vsi-vpnc] vxlan 30
[SwitchA-vsi-vpnc-vxlan-30] quit
[SwitchA-vsi-vpnc] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch B and Switch C.

```
[SwitchA] interface loopback 0
[SwitchA-Loopback0] ip address 1.1.1.1 255.255.255.255
[SwitchA-Loopback0] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 1**.

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 1.1.1.1
[SwitchA-Tunnel1] destination 2.2.2.2
[SwitchA-Tunnel1] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 2**.

```
[SwitchA] interface tunnel 2 mode vxlan
[SwitchA-Tunnel2] source 1.1.1.1
[SwitchA-Tunnel2] destination 3.3.3.3
[SwitchA-Tunnel2] quit
```

Assign Tunnel 1 and Tunnel 2 to VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] tunnel 1
[SwitchA-vsi-vpna-vxlan-10] tunnel 2
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Assign Tunnel 1 and Tunnel 2 to VXLAN 20.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] tunnel 1
[SwitchA-vsi-vpnb-vxlan-20] tunnel 2
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

Assign Tunnel 2 to VXLAN 30.

```
[SwitchA] vsi vpnc
[SwitchA-vsi-vpnc] vxlan 30
[SwitchA-vsi-vpnc-vxlan-30] tunnel 2
[SwitchA-vsi-vpnc-vxlan-30] quit
[SwitchA-vsi-vpnc] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

On GigabitEthernet 1/0/3, create Ethernet service instance 1000 to match VLAN 3.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] service-instance 1000
[SwitchA-GigabitEthernet1/0/3-srv1000] encapsulation s-vid 3
```

Map Ethernet service instance 1000 to VSI **vpnb**.

```
[SwitchA-GigabitEthernet1/0/3-srv1000] xconnect vsi vpnb
[SwitchA-GigabitEthernet1/0/3-srv1000] quit
[SwitchA-GigabitEthernet1/0/3] quit
```

Create VSI-interface 1 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 10.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchA-Vsi-interface1] mac-address 1-1-1
```

Specify VSI-interface 1 as a distributed gateway and enable local proxy ARP on the interface.

```
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-arp enable
[SwitchA-Vsi-interface1] quit
```

Create VSI-interface 2 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 20.

```
[SwitchA] interface vsi-interface 2
[SwitchA-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchA-Vsi-interface2] mac-address 2-2-2
```

Specify VSI-interface 2 as a distributed gateway and enable local proxy ARP on the interface.

```
[SwitchA-Vsi-interface2] distributed-gateway local
[SwitchA-Vsi-interface2] local-proxy-arp enable
[SwitchA-Vsi-interface2] quit
```

Enable dynamic ARP entry synchronization for distributed VXLAN IP gateways.

```
[SwitchA] arp distributed-gateway dynamic-entry synchronize
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**. Assign subnet 10.1.1.0/24 to the VSI.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] gateway subnet 10.1.1.0 0.0.0.255
[SwitchA-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] gateway vsi-interface 2
[SwitchA-vsi-vpnb] quit
```

Assign a secondary IP address to VSI-interface 1. The IP address will be used as the gateway address for VXLAN 30.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip address 20.1.1.1 255.255.255.0 sub
[SwitchA-Vsi-interface1] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpnc**. Assign subnet 20.1.1.0/24 to the VSI.

```
[SwitchA] vsi vpnc
[SwitchA-vsi-vpnc] gateway vsi-interface 1
[SwitchA-vsi-vpnc] gateway subnet 20.1.1.0 0.0.0.255
[SwitchA-vsi-vpnc] quit
```

Configure a PBR policy for VXLAN 10. Set the policy name to **vxlan10**, and set the default next hop to 10.1.1.2 (VSI-interface 1 on Switch B).

```
[SwitchA] acl advanced 3000
[SwitchA-acl-ipv4-adv-3000] rule 0 permit ip
[SwitchA-acl-ipv4-adv-3000] quit
[SwitchA] policy-based-route vxlan10 permit node 5
[SwitchA-pbr-vxlan10-5] if-match acl 3000
[SwitchA-pbr-vxlan10-5] apply default-next-hop 10.1.1.2
[SwitchA-pbr-vxlan10-5] quit
```

Configure a PBR policy for VXLAN 20. Set the policy name to **vxlan20**, and set the default next hop to 10.1.2.2 (VSI-interface 2 on Switch B).

```
[SwitchA] policy-based-route vxlan20 permit node 5
[SwitchA-pbr-vxlan20-5] if-match acl 3000
[SwitchA-pbr-vxlan20-5] apply default-next-hop 10.1.2.2
[SwitchA-pbr-vxlan20-5] quit
```

Apply policies **vxlan10** and **vxlan20** to VSI-interface 1 and VSI-interface 2, respectively.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip policy-based-route vxlan10
[SwitchA-Vsi-interface1] quit
[SwitchA] interface vsi-interface 2
[SwitchA-Vsi-interface2] ip policy-based-route vxlan20
[SwitchA-Vsi-interface2] quit
```

6. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create VSI **vpnb** and VXLAN 20.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Create VSI **vpnc** and VXLAN 30.

```
[SwitchB] vsi vpnc
[SwitchB-vsi-vpnc] vxlan 30
[SwitchB-vsi-vpnc-vxlan-30] quit
[SwitchB-vsi-vpnc] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch C.

```
[SwitchB] interface loopback 0
[SwitchB-Loopback0] ip address 2.2.2.2 255.255.255.255
[SwitchB-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 2**.

```
[SwitchB] interface tunnel 2 mode vxlan
[SwitchB-Tunnel2] source 2.2.2.2
[SwitchB-Tunnel2] destination 1.1.1.1
[SwitchB-Tunnel2] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 3**.

```
[SwitchB] interface tunnel 3 mode vxlan
[SwitchB-Tunnel3] source 2.2.2.2
[SwitchB-Tunnel3] destination 3.3.3.3
[SwitchB-Tunnel3] quit
```

Assign Tunnel 2 to VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] tunnel 2
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Assign Tunnel 2 to VXLAN 20.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] tunnel 2
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Assign Tunnel 3 to VXLAN 30.

```
[SwitchB] vsi vpnc
[SwitchB-vsi-vpnc] vxlan 30
[SwitchB-vsi-vpnc-vxlan-30] tunnel 3
[SwitchB-vsi-vpnc-vxlan-30] quit
[SwitchB-vsi-vpnc] quit
```

Create VSI-interface 1 and assign the interface an IP address.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ip address 10.1.1.2 255.255.255.0
[SwitchB-Vsi-interface1] quit
```

Create VSI-interface 2 and assign the interface an IP address.

```
[SwitchB] interface vsi-interface 2
[SwitchB-Vsi-interface2] ip address 10.1.2.2 255.255.255.0
[SwitchB-Vsi-interface2] quit
```

Create VSI-interface 3 and assign the interface an IP address.

```
[SwitchB] interface vsi-interface 3
[SwitchB-Vsi-interface3] ip address 20.1.1.2 255.255.255.0
[SwitchB-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
```

```
[SwitchB-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] gateway vsi-interface 2
[SwitchB-vsi-vpnb] quit
```

Specify VSI-interface 3 as the gateway interface for VSI **vpnc**.

```
[SwitchB] vsi vpnc
[SwitchB-vsi-vpnc] gateway vsi-interface 3
[SwitchB-vsi-vpnc] quit
```

7. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Create VSI **vpnb** and VXLAN 20.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Create VSI **vpnc** and VXLAN 30.

```
[SwitchC] vsi vpnc
[SwitchC-vsi-vpnc] vxlan 30
[SwitchC-vsi-vpnc-vxlan-30] quit
[SwitchC-vsi-vpnc] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch B.

```
[SwitchC] interface loopback 0
[SwitchC-Loopback0] ip address 3.3.3.3 255.255.255.255
[SwitchC-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 1**.

```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 3.3.3.3
[SwitchC-Tunnel1] destination 1.1.1.1
[SwitchC-Tunnel1] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 3**.

```
[SwitchC] interface tunnel 3 mode vxlan
[SwitchC-Tunnel3] source 3.3.3.3
[SwitchC-Tunnel3] destination 2.2.2.2
[SwitchC-Tunnel3] quit
```

Assign Tunnel 1 to VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] tunnel 1
```

```
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Assign Tunnel 1 to VXLAN 20.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] tunnel 1
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Assign Tunnel 1 and Tunnel 3 to VXLAN 30.

```
[SwitchC] vsi vpnc
[SwitchC-vsi-vpnc] vxlan 30
[SwitchC-vsi-vpnc-vxlan-30] tunnel 1
[SwitchC-vsi-vpnc-vxlan-30] tunnel 3
[SwitchC-vsi-vpnc-vxlan-30] quit
[SwitchC-vsi-vpnc] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 4.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 4
```

Map Ethernet service instance 1000 to VSI **vpnc**.

```
[SwitchC-GigabitEthernet1/0/1-srv1000] xconnect vsi vpnc
[SwitchC-GigabitEthernet1/0/1-srv1000] quit
[SwitchC-GigabitEthernet1/0/1] quit
```

Create VSI-interface 1 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 10.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchC-Vsi-interface1] mac-address 1-1-1
```

Specify VSI-interface 1 as a distributed gateway and enable local proxy ARP on the interface.

```
[SwitchC-Vsi-interface1] distributed-gateway local
[SwitchC-Vsi-interface1] local-proxy-arp enable
[SwitchC-Vsi-interface1] quit
```

Create VSI-interface 2 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 20.

```
[SwitchC] interface vsi-interface 2
[SwitchC-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchC-Vsi-interface2] mac-address 2-2-2
```

Specify VSI-interface 2 as a distributed gateway and enable local proxy ARP on the interface.

```
[SwitchC-Vsi-interface2] distributed-gateway local
[SwitchC-Vsi-interface2] local-proxy-arp enable
[SwitchC-Vsi-interface2] quit
```

Enable dynamic ARP entry synchronization for distributed VXLAN IP gateways.

```
[SwitchC] arp distributed-gateway dynamic-entry synchronize
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**. Assign subnet 10.1.1.0/24 to the VSI.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] gateway vsi-interface 1
[SwitchC-vsi-vpna] gateway subnet 10.1.1.0 0.0.0.255
```

```
[SwitchC-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] gateway vsi-interface 2
[SwitchC-vsi-vpnb] quit
```

Assign a secondary IP address to VSI-interface 1. The IP address will be used as the gateway address for VXLAN 30.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip address 20.1.1.1 255.255.255.0 sub
[SwitchC-Vsi-interface1] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpnc**. Assign subnet 20.1.1.0/24 to the VSI.

```
[SwitchC] vsi vpnc
[SwitchC-vsi-vpnc] gateway vsi-interface 1
[SwitchC-vsi-vpnc] gateway subnet 20.1.1.0 0.0.0.255
[SwitchC-vsi-vpnc] quit
```

Configure a PBR policy for the VXLANs. Set the policy name to **vxlan**, and set the default next hop to 20.1.1.2 (VSI-interface 1 on Switch B).

```
[SwitchC] acl advanced 3000
[SwitchC-acl-ipv4-adv-3000] rule 0 permit ip
[SwitchC-acl-ipv4-adv-3000] quit
[SwitchC] policy-based-route vxlan permit node 5
[SwitchC-pbr-vxlan-5] if-match acl 3000
[SwitchC-pbr-vxlan-5] apply default-next-hop 20.1.1.2
[SwitchC-pbr-vxlan-5] quit
```

Apply policy **vxlan** to VSI-interface 1.

```
[SwitchC] interface vsi-interface1
[SwitchC-Vsi-interface1] ip policy-based-route vxlan
[SwitchC-Vsi-interface1] quit
```

Verifying the configuration

1. Verify the VXLAN IP gateway settings on Switch A:

Verify that the VXLAN tunnel interfaces are up on Switch A.

```
[SwitchA] display interface tunnel 2
Tunnel2
Current state: UP
Line protocol state: UP
Description: Tunnel2 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 3.3.3.3
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that VSI-interface 1 is up.

```
[SwitchA] display interface vsi-interface brief
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP      Description
Vsi1                 UP   UP       10.1.1.1
Vsi2                 UP   UP       10.1.2.1
```

Verify that the VXLAN tunnels have been assigned to VXLAN 10, and VSI-interface 1 is the gateway interface for VSI **vpna**.

```
[SwitchA] display l2vpn vsi name vpna verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type        Flood proxy
    Tunnel1              0x5000001  Up     Manual      Disabled
    Tunnel2              0x5000002  Up     Manual      Disabled
  ACs:
     AC                              Link ID    State    Type
     GE1/0/1 srv1000                 0          Up       Manual
```

Verify that Switch A has created ARP entries for the VMs. (Details not shown.)

2. Verify the configuration on the border gateway Switch B:

Verify that the VXLAN tunnel interfaces are up on Switch B.

```
[SwitchB] display interface tunnel 2
Tunnel2
Current state: UP
Line protocol state: UP
Description: Tunnel2 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 2.2.2.2, destination 1.1.1.1
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
```

```
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that VSI-interface 1 is up.

```
[SwitchB] display interface vsi-interface brief
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP      Description
Vsi1                 UP   UP       10.1.1.2
Vsi2                 UP   UP       10.1.2.2
Vsi3                 UP   UP       20.1.1.2
```

Verify that the VXLAN tunnels have been assigned to VXLAN 10, and VSI-interface 1 is the gateway interface for VSI **vpna**.

```
[SwitchB] display l2vpn vsi name vpna verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type        Flood proxy
    Tunnel1              0x5000001  Up     Manual      Disabled
    Tunnel2              0x5000002  Up     Manual      Disabled
```

Verify that Switch B has created ARP entries for the VMs. (Details not shown.)

Verify that Switch B has created FIB entries for the VMs.

```
[SwitchB] display fib 10.1.1.11
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination/Mask   Nexthop         Flag     OutInterface/Token       Label
10.1.1.11/32       10.1.1.11       UH       Vsi1                     Null
[SwitchB] display fib 20.1.1.12
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
```

```
  R:Relay     F:FRR
Destination/Mask    Nexthop         Flag     OutInterface/Token       Label
20.1.1.12/32        20.1.1.12       UH       Vsi3                     Null
```

3. Verify that the network connectivity for VMs meets the requirements:

   # Verify that VM 1, VM 2, and VM 3 can ping each other. (Details not shown.)

   # Verify that VM 1, VM 2, and VM 3 can ping VLAN-interface 20 (25.1.1.5) on Switch E for WAN access. (Details not shown.)

Example: Configuring distributed VXLAN IP gateways (IPv6 underlay)

Network configuration

As shown in Figure 28:

- Configure VXLAN 10 and VXLAN 20 as unicast-mode VXLANs on Switch A, Switch B, and Switch C to provide connectivity for the VMs across the network sites.
- Manually establish VXLAN tunnels and assign the tunnels to the VXLANs.
- Configure distributed VXLAN IP gateways on Switch A and Switch C to forward traffic between the VXLANs.
- Configure Switch B as a border gateway to forward traffic between the VXLANs and the WAN connected to Switch E.

NOTE:

This example provides configuration of IPv6 sites. The configuration procedure does not differ between IPv4 and IPv6 sites.

Figure 28 Network diagram

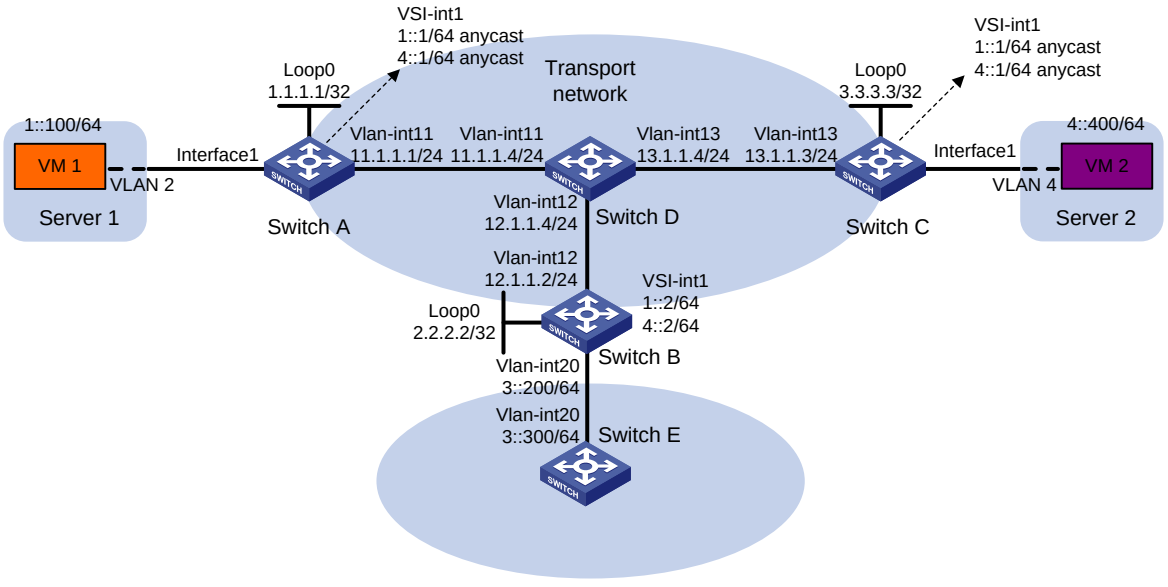


Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Create VLANs and VLAN interfaces on all devices. (Details not shown.)
2. On VM 1 and VM 2, specify 1::1 and 4::1 as the gateway address, respectively. (Details not shown.)
3. Configure IPv6 addresses and unicast routing settings:

   # Assign IPv6 addresses to interfaces, as shown in Figure 28. (Details not shown.)

   # Configure OSPFv3 on all transport network switches (Switches A through D). (Details not shown.)

   # Configure OSPFv3 to advertise routes to networks 1::/64, 4::/64, and 3::/64 on Switch B and Switch E. (Details not shown.)
4. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Create VSI **vpna** and VXLAN 10.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] vxlan 10
   [SwitchA-vsi-vpna-vxlan-10] quit
   [SwitchA-vsi-vpna] quit
   ```

   # Create VSI **vpnb** and VXLAN 20.

   ```
   [SwitchA] vsi vpnb
   [SwitchA-vsi-vpnb] vxlan 20
   [SwitchA-vsi-vpnb-vxlan-20] quit
   [SwitchA-vsi-vpnb] quit
   ```

   # Assign an IPv6 address to Loopback 0. The IPv6 address will be used as the source IP address of the VXLAN tunnels to Switch B and Switch C.

   ```
   [SwitchA] interface loopback 0
   [SwitchA-Loopback0] ipv6 address 7::7 128
   [SwitchA-Loopback0] quit
   ```

   # Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 1**.

   ```
   [SwitchA] interface tunnel 1 mode vxlan ipv6
   [SwitchA-Tunnel1] source 7::7
   [SwitchA-Tunnel1] destination 8::8
   [SwitchA-Tunnel1] quit
   ```

   # Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 2**.

   ```
   [SwitchA] interface tunnel 2 mode vxlan ipv6
   [SwitchA-Tunnel2] source 7::7
   [SwitchA-Tunnel2] destination 9::9
   [SwitchA-Tunnel2] quit
   ```

   # Assign Tunnel 1 and Tunnel 2 to VXLAN 10.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] vxlan 10
   [SwitchA-vsi-vpna-vxlan-10] tunnel 1
   [SwitchA-vsi-vpna-vxlan-10] tunnel 2
   [SwitchA-vsi-vpna-vxlan-10] quit
   [SwitchA-vsi-vpna] quit
   ```

   # Assign Tunnel 1 and Tunnel 2 to VXLAN 20.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] tunnel 1
[SwitchA-vsi-vpnb-vxlan-20] tunnel 2
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

Create VSI-interface 1 and assign the interface two IPv6 anycast addresses. The IP addresses will be used as gateway addresses for VXLAN 10 and VXLAN 20.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ipv6 address 1::1/64 anycast
[SwitchA-Vsi-interface1] ipv6 address 4::1/64 anycast
```

Specify VSI-interface 1 as a distributed gateway and enable local ND proxy on the interface.

```
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-nd enable
[SwitchA-Vsi-interface1] quit
```

Enable dynamic ND entry synchronization for distributed VXLAN IP gateways.

```
[SwitchA] ipv6 nd distributed-gateway dynamic-entry synchronize
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**. Assign subnet 1::1/64 to the VSI.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] gateway subnet 1::1 64
[SwitchA-vsi-vpna] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpnb**. Assign subnet 4::1/64 to the VSI.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] gateway vsi-interface 1
[SwitchA-vsi-vpnb] gateway subnet 4::1 64
[SwitchA-vsi-vpnb] quit
```

Configure an IPv6 static route. Set the destination address to 3::/64 and the next hop to 1::2.

```
[SwitchA] ipv6 route-static 3:: 64 1::2
```

5. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create VSI **vpnb** and VXLAN 20.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Assign an IPv6 address to Loopback 0. The IPv6 address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch C.

```
[SwitchB] interface loopback 0
[SwitchB-Loopback0] ipv6 address 8::8 128
[SwitchB-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 2**.

```
[SwitchB] interface tunnel 2 mode vxlan ipv6
[SwitchB-Tunnel2] source 8::8
[SwitchB-Tunnel2] destination 7::7
[SwitchB-Tunnel2] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 3**.

```
[SwitchB] interface tunnel 3 mode vxlan ipv6
[SwitchB-Tunnel3] source 8::8
[SwitchB-Tunnel3] destination 9::9
[SwitchB-Tunnel3] quit
```

Assign Tunnel 2 and Tunnel 3 to VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] tunnel 2
[SwitchB-vsi-vpna-vxlan-10] tunnel 3
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Assign Tunnel 2 and Tunnel 3 to VXLAN 20.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] tunnel 2
[SwitchB-vsi-vpnb-vxlan-20] tunnel 3
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Create VSI-interface 1 and assign the interface IPv6 addresses.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ipv6 address 1::2/64
[SwitchB-Vsi-interface1] ipv6 address 4::2/64
[SwitchB-Vsi-interface1] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
[SwitchB-vsi-vpna] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpnb**.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] gateway vsi-interface 1
[SwitchB-vsi-vpnb] quit
```

6. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Create VSI **vpnb** and VXLAN 20.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Assign an IPv6 address to Loopback 0. The IPv6 address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch B.

```
[SwitchC] interface loopback 0
[SwitchC-Loopback0] ipv6 address 9::9 128
[SwitchC-Loopback0] quit
```

Create a VXLAN tunnel to Switch A. The tunnel interface name is **Tunnel 1**.

```
[SwitchC] interface tunnel 1 mode vxlan ipv6
[SwitchC-Tunnel1] source 9::9
[SwitchC-Tunnel1] destination 7::7
[SwitchC-Tunnel1] quit
```

Create a VXLAN tunnel to Switch B. The tunnel interface name is **Tunnel 3**.

```
[SwitchC] interface tunnel 3 mode vxlan ipv6
[SwitchC-Tunnel3] source 9::9
[SwitchC-Tunnel3] destination 8::8
[SwitchC-Tunnel3] quit
```

Assign Tunnel 1 and Tunnel 3 to VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] tunnel 1
[SwitchC-vsi-vpna-vxlan-10] tunnel 3
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Assign Tunnel 1 and Tunnel 3 to VXLAN 20.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] tunnel 1
[SwitchC-vsi-vpnb-vxlan-20] tunnel 3
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 4.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 4
```

Map Ethernet service instance 1000 to VSI **vpnb**.

```
[SwitchC-GigabitEthernet1/0/1-srv1000] xconnect vsi vpnb
```

```
[SwitchC-GigabitEthernet1/0/1-srv1000] quit
[SwitchC-GigabitEthernet1/0/1] quit
```

Create VSI-interface 1 and assign the interface two IPv6 anycast addresses. The IP addresses will be used as gateway addresses for VXLAN 10 and VXLAN 20.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ipv6 address 1::1/64 anycast
[SwitchC-Vsi-interface1] ipv6 address 4::1/64 anycast
```

Specify VSI-interface 1 as a distributed gateway and enable local ND proxy on the interface.

```
[SwitchC-Vsi-interface1] distributed-gateway local
[SwitchC-Vsi-interface1] local-proxy-nd enable
[SwitchC-Vsi-interface1] quit
```

Enable dynamic ND entry synchronization for distributed VXLAN IP gateways.

```
[SwitchC] ipv6 nd distributed-gateway dynamic-entry synchronize
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**. Assign subnet 1::1/64 to the VSI.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] gateway vsi-interface 1
[SwitchC-vsi-vpna] gateway subnet 1::1 64
[SwitchC-vsi-vpna] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpnb**. Assign subnet 4::1/64 to the VSI.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] gateway vsi-interface 1
[SwitchC-vsi-vpnb] gateway subnet 4::1 64
[SwitchC-vsi-vpnb] quit
```

Configure an IPv6 static route. Set the destination address to 3::/64 and the next hop to 4::2.

```
[SwitchC] ipv6 route-static 3:: 64 4::2
```

Verifying the configuration

1. Verify the distributed VXLAN IP gateway settings on Switch A:

Verify that the VXLAN tunnel interfaces are up on Switch A.

```
[SwitchA] display interface tunnel 2
Tunnel2
Current state: UP
Line protocol state: UP
Description: Tunnel2 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 7::7, destination 9::9
Tunnel protocol/transport UDP_VXLAN/IPv6
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that VSI-interface 1 is up.

```
[SwitchA] display ipv6 interface vsi-interface brief
*down: administratively down
```

```
(s): spoofing
Interface                                Physical Protocol IPv6 Address
Vsi-interface1                           up       up       1::1
```

Verify that the VXLAN tunnels have been assigned to VXLAN 10 and VXLAN 20, and VSI-interface 1 is the gateway interface for VSIs **vpna** and **vpnb**.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type        Flood proxy
    Tunnel1              0x5000001  Up     Manual      Disabled
    Tunnel2              0x5000002  Up     Manual      Disabled
  ACs:
     AC                              Link ID    State    Type
     GE1/0/1 srv1000                 0          Up       Manual

VSI Name: vpnb
  VSI Index               : 1
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 20
  Tunnels:
    Tunnel Name          Link ID    State  Type        Flood proxy
    Tunnel1              0x5000001  Up     Manual      Disabled
```

```
    Tunnel2              0x5000002  Up     Manual      Disabled
```

Verify that Switch A has created neighbor entries for the VMs.

```
[SwitchA] display ipv6 neighbors all
Type: S-Static    D-Dynamic    O-Openflow     R-Rule    I-Invalid
IPv6 address                  Link layer     VID  Interface      State T  Age
1::2                          3c8c-400d-867a 0    Tunnel1        STALE D  7
1::100                        0001-0000-0047 0    0              STALE D  22
4::400                        0002-0000-0047 1    Tunnel2        REACH D  5
FE80::201:FF:FE00:47          0001-0000-0047 0    Tunnel1        REACH D  30
FE80::202:FF:FE00:0           0002-0000-0000 1    Tunnel1        REACH D  27
FE80::202:FF:FE00:47          0002-0000-0047 0    0              DELAY D  5
```

Verify that Switch A has created FIB entries for the VMs.

```
[SwitchA] display ipv6 fib 4::400
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination: 4::400                                    Prefix length: 128
Nexthop    : 4::400                                    Flags: UH
Time stamp : 0x2c                                      Label: Null
Interface  : Tunnel2                                   Token: Invalid
[SwitchA] display ipv6 fib 3::300
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination: 3::                                       Prefix length: 40
Nexthop    : 1::2                                      Flags: USGR
Time stamp : 0x23                                      Label: Null
Interface  : Tunnel1                                   Token: Invalid
```

2. Verify the configuration on the border gateway Switch B:

Verify that the VXLAN tunnel interfaces are up on Switch B.

```
[SwitchB] display interface tunnel 2
Tunnel2
Current state: UP
Line protocol state: UP
Description: Tunnel2 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 8::8, destination 7::7
Tunnel protocol/transport UDP_VXLAN/IPv6
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that VSI-interface 1 is up.

```
[SwitchB] display ipv6 interface vsi-interface brief
*down: administratively down
(s): spoofing
Interface                              Physical Protocol IPv6 Address
Vsi-interface1                         up       up       1::2
```

Verify that the VXLAN tunnels have been assigned to VXLAN 10, and VSI-interface 1 is the gateway interface for VSI **vpna**.

```
[SwitchB] display l2vpn vsi name vpna verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type        Flood proxy
    Tunnel1              0x5000001  Up     Manual      Disabled
    Tunnel2              0x5000002  Up     Manual      Disabled
```

Verify that Switch B has created neighbor entries for the VMs.

```
[SwitchB] display ipv6 neighbors all
Type: S-Static    D-Dynamic    O-Openflow     R-Rule    I-Invalid
IPv6 address                   Link layer     VID  Interface      State T  Age
3::300                         0003-0000-0047 20   Vlan20         DELAY D  3
FE80::203:FF:FE00:47           0003-0000-0047 20   Vlan20         STALE D  222
1::100                         0001-0000-0047 0    Tunnel2        STALE D  232
4::400                         0002-0000-0047 1    Tunnel3        REACH D  3
FE80::201:FF:FE00:0            0001-0000-0000 0    Tunnel2        STALE D  237
FE80::201:FF:FE00:47           0001-0000-0047 20   Vlan20         STALE D  222
FE80::202:FF:FE00:0            0002-0000-0000 1    Tunnel3        STALE D  345
```

Verify that Switch B has created FIB entries for the VMs.

```
[SwitchB] display ipv6 fib 1::100
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination: 1::100                                  Prefix length: 128
Nexthop    : 1::100                                  Flags: UH
Time stamp : 0x21                                    Label: Null
```

```
Interface  : Vsi1                                          Token: Invalid
[SwitchB] display ipv6 fib 4::400
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination: 4::                                           Prefix length: 64
Nexthop    : ::                                            Flags: U
Time stamp : 0x19                                          Label: Null
Interface  : Vsi1                                          Token: Invalid
```

3. Verify that the network connectivity for the VMs meets the network requirements:

   # Verify that VM 1 and VM 2 can ping each other. (Details not shown.)

   # Verify that VM 1, VM 2, and VLAN-interface 20 (3::300) on Switch E can ping each other. (Details not shown.)

Example: Configuring VXLAN M-LAG with a direct peer link

Network configuration

As shown in Figure 29, perform the following tasks to make sure the VMs can communicate with one another:

- Configure Switch A, Switch B, and Switch C as distributed VXLAN IP gateways.
- Configure VXLAN 10 and VXLAN 20 on Switch A, Switch B, and Switch C.
- Configure VXLAN M-LAG on Switch A and Switch B to virtualize them into one device. The switches use an Ethernet aggregate link as the peer link.

Figure 29 Network diagram

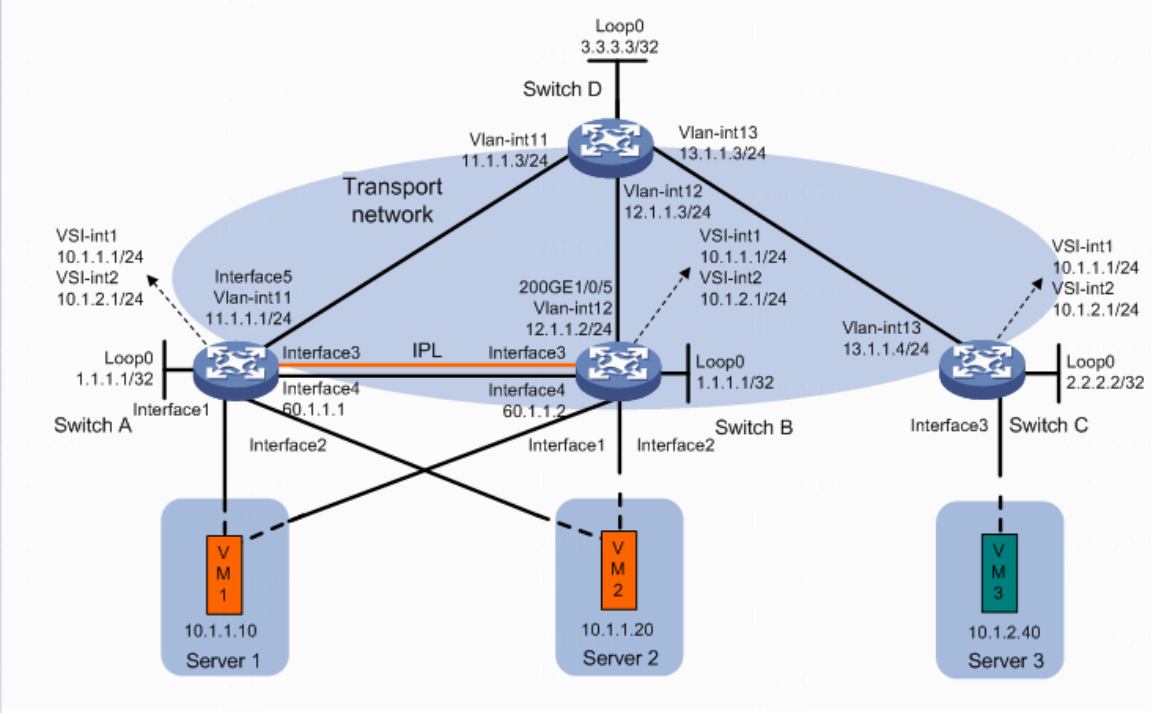


Table 9 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

1.
2. On VM 1 and VM 2, specify 10.1.1.1 as the gateway address. On VM 3, specify 10.1.2.1 as the gateway address. (Details not shown.)
3. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces (including loopback interfaces), as shown in Figure 29. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)

   # Execute the **`ospf peer hold-max-cost duration`** command on the interfaces used for setting up OSPF neighbor relationships. (Details not shown.)
4. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Choose one of the following frame match criterion creation methods for dynamic ACs:

   - Enable the device to create frame match criteria based on VXLAN IDs for the dynamic ACs on the peer link.

     ```
     [SwitchA] l2vpn m-lag peer-link ac-match-rule vxlan-mapping
     ```
   - Use the default setting for dynamic ACs on the peer link to use frame match criteria identical to those of site-facing ACs.

   # Configure M-LAG system settings.

   ```
   [SwitchA] m-lag system-mac 0001-0001-0001
   [SwitchA] m-lag system-number 1
   [SwitchA] m-lag system-priority 10
   [SwitchA] m-lag keepalive ip destination 60.1.1.2 source 60.1.1.1
   [SwitchA] m-lag restore-delay 180
   ```

   # Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3.

   ```
   [SwitchA] interface bridge-aggregation 3
   [SwitchA-Bridge-Aggregation3] link-aggregation mode dynamic
   [SwitchA-Bridge-Aggregation3] quit
   ```

   # Assign GigabitEthernet 1/0/3 to link aggregation group 3.

   ```
   [SwitchA] interface gigabitethernet 1/0/3
   [SwitchA-GigabitEthernet1/0/3] port link-aggregation group 3
   [SwitchA-GigabitEthernet1/0/3] quit
   ```

   # Specify Bridge-Aggregation 3 as the peer-link interface.

   ```
   [SwitchA] interface bridge-aggregation 3
   [SwitchA-Bridge-Aggregation3] port m-lag peer-link 1
   ```

```
[SwitchA-Bridge-Aggregation3] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 to link aggregation group 4.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 4
[SwitchA-GigabitEthernet1/0/1] quit
```

Assign Bridge-Aggregation 4 to M-LAG group 4.

```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] port m-lag group 4
[SwitchA-Bridge-Aggregation4] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5.

```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/2 to link aggregation group 5.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-aggregation group 5
[SwitchA-GigabitEthernet1/0/2] quit
```

Assign Bridge-Aggregation 5 to M-LAG group 5.

```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] port m-lag group 5
[SwitchA-Bridge-Aggregation5] quit
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Create VSI **vpnb** and VXLAN 20.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnel to Switch C.

```
[SwitchA] interface loopback 0
[SwitchA-Loopback0] ip address 1.1.1.1 255.255.255.255
[SwitchA-Loopback0] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 1**.

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 1.1.1.1
[SwitchA-Tunnel1] destination 2.2.2.2
[SwitchA-Tunnel1] quit
```

Assign Tunnel 1 to VXLAN 10.

```
[SwitchA] vsi vpna
```

```
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] tunnel 1
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Assign Tunnel 1 to VXLAN 20.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] tunnel 1
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

On Bridge-Aggregation 4, create Ethernet service instance 1000 to match VLAN 1.

```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] service-instance 1000
[SwitchA-Bridge-Aggregation4-srv1000] encapsulation s-vid 1
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-Bridge-Aggregation10-srv1000] xconnect vsi vpna
[SwitchA-Bridge-Aggregation10-srv1000] quit
```

On Bridge-Aggregation 5, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] service-instance 1000
[SwitchA-Bridge-Aggregation5-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-Bridge-Aggregation5-srv1000] xconnect vsi vpna
[SwitchA-Bridge-Aggregation5-srv1000] quit
```

Create VSI-interface 1 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 10.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchA-Vsi-interface1] mac-address 1-1-1
```

Specify VSI-interface 1 as a distributed gateway.

```
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-arp enable
[SwitchA-Vsi-interface1] quit
```

Create VSI-interface 2 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 20.

```
[SwitchA] interface vsi-interface 2
[SwitchA-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchA-Vsi-interface2] mac-address 2-2-2
```

Specify VSI-interface 2 as a distributed gateway.

```
[SwitchA-Vsi-interface2] distributed-gateway local
[SwitchA-Vsi-interface2] local-proxy-arp enable
[SwitchA-Vsi-interface2] quit
```

Enable dynamic ARP entry synchronization for distributed VXLAN IP gateways.

```
[SwitchA] arp distributed-gateway dynamic-entry synchronize
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] gateway vsi-interface 2
[SwitchA-vsi-vpnb] quit
```

Exclude the interfaces used by VXLAN from the shutdown action by M-LAG MAD.

```
[SwitchA] m-lag mad exclude interface loopback 0
[SwitchA] m-lag mad exclude interface gigabitethernet 1/0/4
[SwitchA] m-lag mad exclude interface gigabitethernet 1/0/5
[SwitchA] m-lag mad exclude interface vsi-interface 1
[SwitchA] m-lag mad exclude interface vsi-interface 2
[SwitchA] m-lag mad exclude interface vlan-interface 11
```

5. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Choose one of the following frame match criterion creation methods for dynamic ACs:

- Enable the device to create frame match criteria based on VXLAN IDs for the dynamic ACs on the peer link.

  ```
  [SwitchB] l2vpn m-lag peer-link ac-match-rule vxlan-mapping
  ```

- Use the default setting for dynamic ACs on the peer link to use frame match criteria identical to those of site-facing ACs.

Configure M-LAG system settings.

```
[SwitchB] m-lag system-mac 0001-0001-0001
[SwitchB] m-lag system-number 2
[SwitchB] m-lag system-priority 10
[SwitchB] m-lag keepalive ip destination 60.1.1.1 source 60.1.1.2
[SwitchB] m-lag restore-delay 180
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3.

```
[SwitchB] interface bridge-aggregation 3
[SwitchB-Bridge-Aggregation3] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/3 to link aggregation group 3.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port link-aggregation group 3
[SwitchB-GigabitEthernet1/0/3] quit
```

Specify Bridge-Aggregation 3 as the peer-link interface.

```
[SwitchB] interface bridge-aggregation 3
[SwitchB-Bridge-Aggregation3] port m-lag peer-link 1
[SwitchB-Bridge-Aggregation3] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 to link aggregation group 4.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-aggregation group 4
[SwitchB-GigabitEthernet1/0/1] quit
```

Assign Bridge-Aggregation 4 to M-LAG group 4.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] port m-lag group 4
[SwitchB-Bridge-Aggregation4] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/2 to link aggregation group 5.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-aggregation group 5
[SwitchB-GigabitEthernet1/0/2] quit
```

Assign Bridge-Aggregation 5 to M-LAG group 5.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] port m-lag group 5
[SwitchB-Bridge-Aggregation5] quit
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create VSI **vpnb** and VXLAN 20.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnel to Switch C.

```
[SwitchB] interface loopback 0
[SwitchB-Loopback0] ip address 1.1.1.1 255.255.255.255
[SwitchB-Loopback0] quit
```

Create a VXLAN tunnel to Switch C. The tunnel interface name is **Tunnel 1**.

```
[SwitchB] interface tunnel 1 mode vxlan
[SwitchB-Tunnel1] source 1.1.1.1
[SwitchB-Tunnel1] destination 2.2.2.2
[SwitchB-Tunnel1] quit
```

Assign Tunnel 1 to VXLAN 10.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] tunnel 1
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Assign Tunnel 1 to VXLAN 20.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] tunnel 1
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

On Bridge-Aggregation 4, create Ethernet service instance 1000 to match VLAN 1.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] service-instance 1000
[SwitchB-Bridge-Aggregation4-srv1000] encapsulation s-vid 1
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-Bridge-Aggregation10-srv1000] xconnect vsi vpna
[SwitchB-Bridge-Aggregation10-srv1000] quit
```

On Bridge-Aggregation 5, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] service-instance 1000
[SwitchB-Bridge-Aggregation5-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-Bridge-Aggregation5-srv1000] xconnect vsi vpna
[SwitchB-Bridge-Aggregation5-srv1000] quit
```

Create VSI-interface 1 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 10.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchB-Vsi-interface1] mac-address 1-1-1
```

Specify VSI-interface 1 as a distributed gateway.

```
[SwitchB-Vsi-interface1] distributed-gateway local
[SwitchB-Vsi-interface1] local-proxy-arp enable
[SwitchB-Vsi-interface1] quit
```

Create VSI-interface 2 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 20.

```
[SwitchB] interface vsi-interface 2
[SwitchB-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchB-Vsi-interface2] mac-address 2-2-2
```

Specify VSI-interface 2 as a distributed gateway.

```
[SwitchB-Vsi-interface2] distributed-gateway local
[SwitchB-Vsi-interface2] local-proxy-arp enable
[SwitchB-Vsi-interface2] quit
```

Enable dynamic ARP entry synchronization for distributed VXLAN IP gateways.

```
[SwitchB] arp distributed-gateway dynamic-entry synchronize
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
[SwitchB-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] gateway vsi-interface 2
[SwitchB-vsi-vpnb] quit
```

Exclude the interfaces used by VXLAN from the shutdown action by M-LAG MAD.

```
[SwitchB] m-lag mad exclude interface loopback 0
[SwitchB] m-lag mad exclude interface gigabitethernet 1/0/4
[SwitchB] m-lag mad exclude interface gigabitethernet 1/0/5
[SwitchB] m-lag mad exclude interface vsi-interface 1
[SwitchB] m-lag mad exclude interface vsi-interface 2
```

```
[SwitchB] m-lag mad exclude interface vlan-interface 12
```

6. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Create VSI **vpna** and VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Create VSI **vpnb** and VXLAN 20.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Assign an IP address to Loopback 0. The IP address will be used as the source IP address of the VXLAN tunnels to Switch A and Switch B.

```
[SwitchC] interface loopback 0
[SwitchC-Loopback0] ip address 2.2.2.2 255.255.255.255
[SwitchC-Loopback0] quit
```

Create a VXLAN tunnel to Switch A and Switch B. The tunnel interface name is **Tunnel 1**.

```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 2.2.2.2
[SwitchC-Tunnel1] destination 1.1.1.1
[SwitchC-Tunnel1] quit
```

Assign Tunnel 1 to VXLAN 10.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] tunnel 1
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Assign Tunnel 1 to VXLAN 20.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] tunnel 1
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

On GigabitEthernet 1/0/3, create Ethernet service instance 1000 to match VLAN 4.

```
[SwitchC] interface gigabitethernet 1/0/3
[SwitchC-GigabitEthernet1/0/3] service-instance 1000
[SwitchC-GigabitEthernet1/0/3] encapsulation s-vid 4
```

Map Ethernet service instance 1000 to VSI **vpnb**.

```
[SwitchC-GigabitEthernet1/0/3] xconnect vsi vpnb
[SwitchC-GigabitEthernet1/0/3] quit
```

Create VSI-interface 1 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 10.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
```

```
[SwitchC-Vsi-interface1] mac-address 1-1-1
```

Specify VSI-interface 1 as a distributed gateway and enable local proxy ARP on the interface.

```
[SwitchC-Vsi-interface1] distributed-gateway local
[SwitchC-Vsi-interface1] local-proxy-arp enable
[SwitchC-Vsi-interface1] quit
```

Create VSI-interface 2 and assign the interface an IP address and a MAC address. The IP address will be used as the gateway address for VXLAN 20.

```
[SwitchC] interface vsi-interface 2
[SwitchC-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchC-Vsi-interface2] mac-address 2-2-2
```

Specify VSI-interface 2 as a distributed gateway and enable local proxy ARP on the interface.

```
[SwitchC-Vsi-interface2] distributed-gateway local
[SwitchC-Vsi-interface2] local-proxy-arp enable
[SwitchC-Vsi-interface2] quit
```

Enable dynamic ARP entry synchronization for distributed VXLAN IP gateways.

```
[SwitchC] arp distributed-gateway dynamic-entry synchronize
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] gateway vsi-interface 1
[SwitchC-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] gateway vsi-interface 2
[SwitchC-vsi-vpnb] quit
```

Verifying the configuration

1. Verify the M-LAG configuration on Switch A:

Verify that the VXLAN tunnel to Switch C is up, and the tunnel source address is 1.1.1.1.

```
[SwitchA]display interface tunnel
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Output queue - Urgent queuing: Size/Length/Discards 0/100/0
Output queue - Protocol queuing: Size/Length/Discards 0/500/0
Output queue - FIFO queuing: Size/Length/Discards 0/75/0
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 50506 bytes/sec, 404048 bits/sec, 295 packets/sec
Input: 4 packets, 240 bytes, 0 drops
Output: 257011 packets, 43395523 bytes, 0 drops
```

Verify that ACs are automatically created on the peer link and assigned to VSIs.

- For dynamic ACs whose frame match criteria are generated based on VXLAN IDs:

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 1
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : Unlimited
  Broadcast Restrain      : Unlimited
  Multicast Restrain      : Unlimited
  Unknown Unicast Restrain: Unlimited
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel1              0x5000000  UP       Manual      Disabled
  ACs:
    AC                              Link ID  State       Type
    BAGG3 srv1                      0        Up          Dynamic (M-LAG)
    BAGG4 srv1000                   1        Up          Manual
    BAGG5 srv1000                   2        Up          Manual
```

- For dynamic ACs whose frame match criteria are identical to those of site-facing ACs:

```
[SwitchA] display l2vpn vsi name vpna verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : 5120 kbps
  Multicast Restrain      : 5120 kbps
  Unknown Unicast Restrain: 5120 kbps
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : Unlimited
  Drop Unknown            : Disabled
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel1              0x5000001  UP       Manual      Disabled
  ACs:
    AC                                 Link ID    State       Type
    BAGG4 srv1000                      0          Up          Manual
```

```
      BAGG3 srv1                      1          Up          Dynamic (M-LAG)
      BAGG5 srv1000                   2          Up          Manual
      BAGG3 srv2                      3          Up          Dynamic (M-LAG)
```

2. Verify the distributed VXLAN IP gateway settings on Switch C:

Verify that the VXLAN tunnel to Switch A and Switch B is up.

```
[SwitchC]display interface tunnel
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Output queue - Urgent queuing: Size/Length/Discards 0/100/0
Output queue - Protocol queuing: Size/Length/Discards 0/500/0
Output queue - FIFO queuing: Size/Length/Discards 0/75/0
Last clearing of counters: Never
Tunnel source 2.2.2.2, destination 1.1.1.1
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 8 packets, 480 bytes, 0 drops
Output: 4 packets, 240 bytes, 0 drops
```

Verify that the VXLAN tunnel has been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

```
[SwitchC] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : 5120 kbps
  Multicast Restrain      : 5120 kbps
  Unknown Unicast Restrain: 5120 kbps
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : Unlimited
  Drop Unknown            : Disabled
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel1              0x5000001  UP       Manual      Disabled

VSI Name: vpnb
  VSI Index               : 1
```

```
VSI State               : Up
MTU                     : 1500
Bandwidth               : -
Broadcast Restrain      : 5120 kbps
Multicast Restrain      : 5120 kbps
Unknown Unicast Restrain: 5120 kbps
MAC Learning            : Enabled
MAC Table Limit         : -
MAC Learning rate       : Unlimited
Drop Unknown            : Disabled
Flooding                : Enabled
Statistics              : Disabled
Gateway Interface       : VSI-interface 2
VXLAN ID                : 20
Tunnels:
  Tunnel Name          Link ID    State    Type        Flood proxy
  Tunnel1              0x5000001  UP       Manual      Disabled
```

3. Verify network connectivity for the VMs:

Verify that VM 1, VM 2, and VM 3 can communicate when both Switch A and Switch B are operating correctly. (Details not shown.)

Verify that VM 1, VM 2, and VM 3 can communicate when Switch A's or Switch B's links to the local site are disconnected. (Details not shown.)

Configuring the VTEP as an OVSDB VTEP

About OVSDB VTEP

An H3C network virtualization controller can use the Open vSwitch Database (OVSDB) management protocol to deploy and manage VXLANs on VTEPs. To work with a controller, you must configure the VTEP as an OVSDB VTEP.

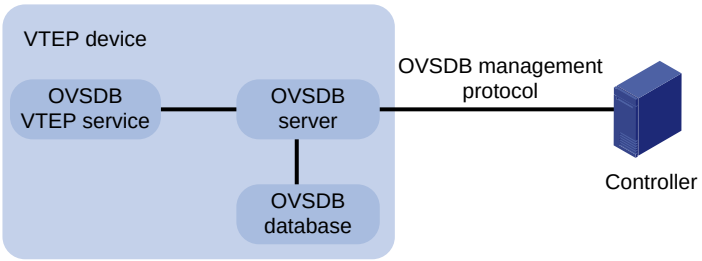
Working mechanisms

As shown in Figure 30, an OVSDB VTEP stores all of its VXLAN settings in the form of entries in an OVSDB database. The OVSDB database, OVSDB VTEP service, and the controller interact through the OVSDB server. The controller communicates with the OVSDB server through the OVSDB protocol to manage the OVSDB database. The OVSDB VTEP service reads and writes data in the OVSDB database through the OVSDB server.

The OVSDB VTEP service performs the following operations to manage the VXLAN settings on the VTEP:

- Converts data in the OVSDB database into VXLAN configuration and deploys the configuration to the VTEP. For example, create or remove a VXLAN or VXLAN tunnel.
- Adds site-facing interface information and the global source address of VXLAN tunnels to the OVSDB database. The information is reported to the controller by the OVSDB server.

Figure 30 OVSDB network model



Protocols and standards

RFC 7047, *The Open vSwitch Database Management Protocol*

Restrictions and guidelines: OVSDB VTEP configuration

You can configure a VTEP both at the CLI and through a controller. As a best practice, do not manually remove the VXLAN configuration issued by the controller.

OVSDB VTEP tasks at a glance

To configure OVSDB VTEPs, perform the following tasks:

1. Setting up an OVSDB connection to a controller

    - Configuring active SSL connection settings
    - Configuring passive SSL connection settings
    - Configuring active TCP connection settings
    - Configuring passive TCP connection settings
2. Enabling the OVSDB server
3. Enabling the OVSDB VTEP service
4. Specifying a global source address for VXLAN tunnels
5. Specifying a VTEP access port
6. (Optional.) Disabling the ACLs issued by the OVSDB controller

Prerequisites for OVSDB VTEP configuration

Before you configure the VTEP as an OVSDB VTEP, enable L2VPN by using the **`l2vpn enable`** command.

Before you set up SSL connections to controllers, you must configure SSL as described in *Security Configuration Guide*.

Setting up an OVSDB connection to a controller

About OVSDB connection types

The OVSDB server supports the following types of OVSDB connections:

- **Active SSL connection**—The OVSDB server initiates an SSL connection to the controller.
- **Passive SSL connection**—The OVSDB server accepts the SSL connection from the controller.
- **Active TCP connection**—The OVSDB server initiates a TCP connection to the controller.
- **Passive TCP connection**—The OVSDB server accepts the TCP connection from the controller.

Restrictions and guidelines for OVSDB controller connection setup

When you set up OVSDB connections, follow these restrictions and guidelines:

- You can set up multiple OVSDB connections. For the device to establish the connections, you must enable the OVSDB server. You must disable and then re-enable the OVSDB server if it has been enabled.
- You must specify the same PKI domain and CA certificate file for all active and passive SSL connections.

Prerequisites for OVSDB controller connection setup

Make sure you have configured a PKI domain before specifying it for SSL. For more information about configuring a PKI domain, see *Security Configuration Guide*.

Configuring active SSL connection settings

1. Enter system view.
   system-view
2. Specify a PKI domain for SSL.
   ovsdb server pki domain *domain-name*
   By default, no PKI domain is specified for SSL.
3. (Optional.) Specify a CA certificate file for SSL.
   ovsdb server bootstrap ca-certificate *ca-filename*
   By default, SSL uses the CA certificate file in the PKI domain.
   If the specified CA certificate file does not exist, the device obtains a self-signed certificate from the controller. The obtained file uses the name specified for the *ca-filename* argument.
4. Set up an active SSL connection.
   ovsdb server ssl ip *ip-address* **port** *port-number*
   By default, the device does not have active OVSDB SSL connections.
   You can set up a maximum of eight OVSDB SSL connections.

Configuring passive SSL connection settings

1. Enter system view.
   system-view
2. Specify a PKI domain for SSL.
   ovsdb server pki domain *domain-name*
   By default, no PKI domain is specified for SSL.
3. (Optional.) Specify a CA certificate file for SSL.
   ovsdb server bootstrap ca-certificate *ca-filename*
   By default, SSL uses the CA certificate file in the PKI domain.
   If the specified CA certificate file does not exist, the device obtains a self-signed certificate from the controller. The obtained file uses the name specified for the *ca-filename* argument.
4. Enable the device to listen for SSL connection requests.
   ovsdb server pssl [**port** *port-number*]
   By default, the device does not listen for SSL connection requests.
   You can specify only one port to listen for OVSDB SSL connection requests.

Configuring active TCP connection settings

1. Enter system view.
   system-view
2. Set up an active TCP connection.
   ovsdb server tcp ip *ip-address* **port** *port-number*
   By default, the device does not have active OVSDB TCP connections.
   You can set up a maximum of eight active OVSDB TCP connections.

Configuring passive TCP connection settings

1. Enter system view.
   system-view
2. Enable the device to listen for TCP connection requests.
   ovsdb server ptcp [**port** *port-number*] [**acl** *acl-number*]
   By default, the device does not listen for TCP connection requests.
   You can specify only one port to listen for OVSDB TCP connection requests.

Enabling the OVSDB server

Prerequisites

Make sure you have complete OVSDB connection setup before you enable the OVSDB server. If you change OVSDB connection settings after the OVSDB server is enabled, you must disable and then re-enable the OVSDB server for the change to take effect.

Procedure

1. Enter system view.
   system-view
2. Enable the OVSDB server.
   ovsdb server enable
   By default, the OVSDB server is disabled.

Enabling the OVSDB VTEP service

1. Enter system view.
   system-view
2. Enable the OVSDB VTEP service.
   vtep enable
   By default, the OVSDB VTEP service is disabled.

Specifying a global source address for VXLAN tunnels

About this task

The VTEP reports the global VXLAN tunnel source address to the controller for VXLAN tunnel setup.

Restrictions and guidelines

For correct VXLAN deployment and VTEP management, do not manually specify tunnel-specific source addresses for VXLAN tunnels if OVSDB is used.

Procedure

1. Enter system view.
   system-view
2. Specify a global source address for VXLAN tunnels.

tunnel global source-address { *ipv4-address* | **ipv6** *ipv6-address* }

By default, no global source address is specified for VXLAN tunnels.

Specifying a VTEP access port

About this task

For the controller to manage a site-facing interface, you must specify the interface as a VTEP access port.

Procedure

1. Enter system view.
   system-view
2. Enter interface view.
   interface *interface-type interface-number*
3. Specify the interface as a VTEP access port.
   vtep access port
   By default, an interface is not a VTEP access port.

Disabling the ACLs issued by the OVSDB controller

About this task

Perform this task on a VTEP to disable all the ACLs issued by the OVSDB controller in order to save ACL resources on the VTEP.

Prerequisites

Before you perform this task, you must enable the OVSDB VTEP service by using the **vtep enable** command.

Procedure

1. Enter system view.
   system-view
2. Disable the ACLs issued by the OVSDB controller.
   vtep acl disable
   By default, the ACLs issued by the OVSDB controller are enabled on a VTEP.

OVSDB VTEP configuration examples

Example: Configuring a unicast-mode VXLAN

Network configuration

As shown in Figure 31, configure the controller cluster to deploy unicast-mode VXLAN 10 to Switch A, Switch B, and Switch C to provide Layer 2 connectivity for the VMs across the network sites.

Figure 31 Network diagram

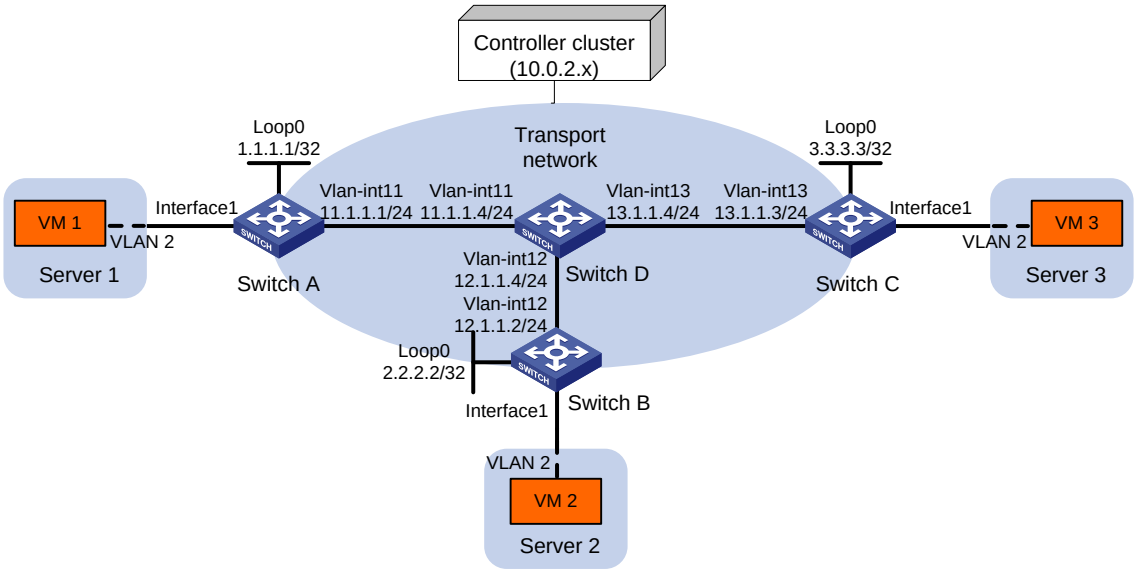


Table 10 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. Create VLANs and VLAN interfaces on all devices. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 31. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D). (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Configure active TCP connection settings.

   ```
   [SwitchA] ovsdb server tcp ip 10.0.2.15 port 6632
   ```

   # Enable the OVSDB server.

   ```
   [SwitchA] ovsdb server enable
   ```

   # Enable the OVSDB VTEP service.

   ```
   [SwitchA] vtep enable
   ```

   # Assign an IP address to Loopback 0. Specify the IP address as the global source address for VXLAN tunnels.

   ```
   [SwitchA] interface loopback 0
   [SwitchA-LoopBack0] ip address 1.1.1.1 255.255.255.255
   [SwitchA-LoopBack0] quit
   [SwitchA] tunnel global source-address 1.1.1.1
   ```

   # Specify site-facing interface GigabitEthernet 1/0/1 as a VTEP access port.

   ```
   [SwitchA] interface gigabitethernet 1/0/1
   [SwitchA-GigabitEthernet1/0/1] vtep access port
   ```

```
[SwitchA-GigabitEthernet1/0/1] quit
```

4. Configure Switch B:

   # Enable L2VPN.

   ```
   <SwitchB> system-view
   [SwitchB] l2vpn enable
   ```

   # Configure active TCP connection settings.

   ```
   [SwitchB] ovsdb server tcp 10.0.2.15 port 6632
   ```

   # Enable the OVSDB server.

   ```
   [SwitchB] ovsdb server enable
   ```

   # Enable the OVSDB VTEP service.

   ```
   [SwitchB] vtep enable
   ```

   # Assign an IP address to Loopback 0. Specify the IP address as the global source address for VXLAN tunnels.

   ```
   [SwitchB] interface loopback 0
   [SwitchB-LoopBack0] ip address 2.2.2.2 255.255.255.255
   [SwitchB-LoopBack0] quit
   [SwitchB] tunnel global source-address 2.2.2.2
   ```

   # Specify site-facing interface GigabitEthernet 1/0/1 as a VTEP access port.

   ```
   [SwitchB] interface gigabitethernet 1/0/1
   [SwitchB-GigabitEthernet1/0/1] vtep access port
   [SwitchB-GigabitEthernet1/0/1] quit
   ```

5. Configure Switch C:

   # Enable L2VPN.

   ```
   <SwitchC> system-view
   [SwitchC] l2vpn enable
   ```

   # Configure active TCP connection settings.

   ```
   [SwitchC] ovsdb server tcp ip 10.0.2.15 port 6632
   ```

   # Enable the OVSDB server.

   ```
   [SwitchC] ovsdb server enable
   ```

   # Enable the OVSDB VTEP service.

   ```
   [SwitchC] vtep enable
   ```

   # Assign an IP address to Loopback 0. Specify the IP address as the global source address for VXLAN tunnels.

   ```
   [SwitchC] interface loopback 0
   [SwitchC-LoopBack0] ip address 3.3.3.3 255.255.255.255
   [SwitchC-LoopBack0] quit
   [SwitchC] tunnel global source-address 3.3.3.3
   ```

   # Specify site-facing interface GigabitEthernet 1/0/1 as a VTEP access port.

   ```
   [SwitchC] interface gigabitethernet 1/0/1
   [SwitchC-GigabitEthernet1/0/1] vtep access port
   [SwitchC-GigabitEthernet1/0/1] quit
   ```

6. Configure VXLAN settings on the controller. (Details not shown.)

Verifying the configuration

1. Verify the VXLAN settings on the VTEPs. This example uses Switch A.

   # Verify that the VXLAN tunnel interfaces on the VTEP are up.

   ```
   [SwitchA] display interface tunnel 1
   Tunnel1
   ```

```
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that the VXLAN tunnels have been assigned to the VXLAN.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: evpn2014
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type        Flood proxy
    Tunnel1              0x5000001  Up     Manual      Disabled
    Tunnel2              0x5000002  Up     Manual      Disabled
  ACs:
    AC                              Link ID    State    Type
    GE1/0/1 srv2                    0          Up       Manual
```

Verify that the VTEP has learned the MAC addresses of remote VMs.

```
<SwitchA> display l2vpn mac-address
* - The output interface is issued to another VSI
MAC Address    State     VSI Name                       Link ID/Name   Aging
cc3e-5f9c-6cdb Dynamic   evpn2014                       Tunnel1        Aging
cc3e-5f9c-23dc Dynamic   evpn2014                       Tunnel2        Aging
--- 2 mac address(es) found  ---
```

2. Verify that VM 1, VM 2, and VM 3 can ping each other. (Details not shown.)

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

EVPN Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the EVPN fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

EVPN overview 1
EVPN solutions 1
EVPN VXLAN 1
EVPN benefits 1
Layered transport network 2
MP-BGP extension for EVPN 2
RD and route target selection of BGP EVPN routes 3

EVPN overview

Ethernet Virtual Private Network (EVPN) is a Layer 2 VPN technology that provides both Layer 2 and Layer 3 connectivity between distant network sites across an IP or MPLS network. EVPN uses MP-BGP in the control plane and Virtual eXtensible LAN (VXLAN) in the data plane. EVPN is typically used in data centers for multitenant services.

EVPN solutions

EVPN provides the EVPN VXLAN solution.

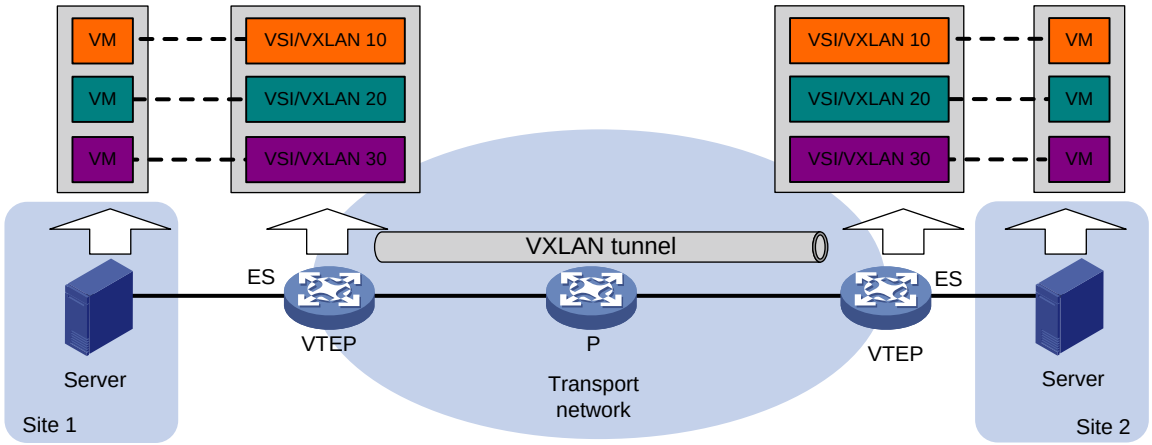
EVPN VXLAN

As shown in Figure 1, EVPN VXLAN uses the VXLAN technology for traffic forwarding in the data plane. The transport edge devices assign VMs to different VXLANs, and then forward traffic at Layer 2 between sites for VMs by using VXLAN tunnels. The transport edge devices are VXLAN tunnel endpoints (VTEPs). All EVPN VXLAN processing is performed on VTEPs

To provide Layer 3 connectivity between subnets of a tenant and between the EVPN VXLAN network and external networks, you can deploy EVPN gateways.

For more information about EVPN VXLAN, see "Configuring EVPN VXLAN."

Figure 1 EVPN VXLAN network model



EVPN benefits

EVPN provides the following benefits:

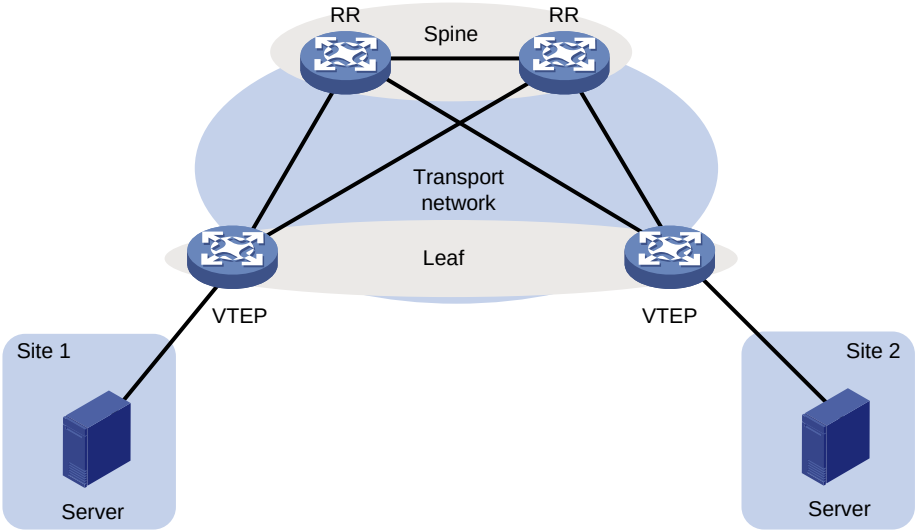
- **Configuration automation**—MP-BGP automates VTEP discovery, VXLAN tunnel establishment, and VXLAN tunnel assignment to ease deployment.
- **Separation of the control plane and the data plane**—EVPN uses MP-BGP to advertise host reachability information in the control plane and uses VXLAN to forward traffic in the data plane.
- **Integrated routing and bridging (IRB)**—MP-BGP advertises both Layer 2 and Layer 3 host reachability information to provide optimal forwarding paths and minimize flooding in an EVPN VXLAN network.

- **Point-to-point and point-to-multipoint connection**—Layer 2 frames are transmitted transparently across the IP transport network between sites after they are encapsulated into VXLAN packets packets.

Layered transport network

As shown in Figure 2, typically the EVPN transport network uses a layered structure. On the transport network, leaf nodes act as VTEPs or PEs to provide VXLAN services, and spine nodes perform forwarding for VXLAN traffic based on the outer IP header. If all VTEPs and transport network devices of an EVPN network belong to the same AS, the spine nodes can act as route reflectors (RRs) to reflect routes between the VTEPs. In this scenario, the spine nodes advertise and receive BGP EVPN routes, but do not perform VXLAN encapsulation and de-encapsulation.

Figure 2 Layered transport network



MP-BGP extension for EVPN

To support EVPN, MP-BGP introduces the EVPN subsequent address family under the L2VPN address family and the following network layer reachability information (BGP EVPN routes):

- **MAC/IP advertisement route**—Advertises MAC reachability information and host route information (host ARP or ND information).
- **Inclusive multicast Ethernet tag (IMET) route**—Advertises VTEP and VXLAN mappings for automating VTEP discovery, VXLAN tunnel establishment, and VXLAN tunnel assignment in an EVPN VXLAN network.
- **IP prefix advertisement route**—Advertises BGP IPv4 or IPv6 unicast routes as IP prefixes.

MP-BGP uses the route distinguisher (RD) field to differentiate BGP EVPN routes of different VSIs and uses route targets to control the advertisement and acceptance of BGP EVPN routes. MP-BGP supports the following types of route targets:

- **Export target**—A VTEP sets the export targets for BGP EVPN routes learned from the local site before advertising them to remote VTEPs.
- **Import target**—A VTEP checks the export targets of BGP EVPN routes received from remote VTEPs. The VTEP imports the BGP EVPN routes only when their export targets match the local import targets.

RD and route target selection of BGP EVPN routes

As shown in Table 1, you can configure RDs and route targets for BGP EVPN routes in multiple views.

Table 1 Supported views for RD and route target configuration

Item	Views
RD	• VSI EVPN instance view • VPN instance view • Public instance view
Route targets	• VSI EVPN instance view • VPN instance view • VPN instance IPv4 address family view • VPN instance IPv6 address family view • VPN instance EVPN view • Public instance view • Public instance IPv4 address family view • Public instance IPv6 address family view • Public instance EVPN view **NOTE:** Route targets configured in VPN instance view apply to IPv4 VPN, IPv6 VPN, and EVPN. Route targets configured in IPv4 address family view apply only to IPv4 VPN. Route targets configured in IPv6 address family view apply only to IPv6 VPN. Route targets configured in VPN instance EVPN view apply only to EVPN. Route targets configured in IPv4 address family view, IPv6 address family view, or VPN instance EVPN view take precedence over those in VPN instance view. The precedence order for different views of a VPN instance also applies to the views of the public instance.

The device selects RDs and route targets for BGP EVPN routes by using the following rules:

- **IMET routes and MAC/IP advertisement routes that contain only MAC addresses**—The device uses the RD and route targets configured in VSI EVPN instance view when advertising and accepting the routes.
- **MAC/IP advertisement routes that contain ARP or ND information**—The device uses the following settings when advertising the routes:
  - RD and export route targets configured in VSI EVPN instance view.
  - Export route targets configured for EVPN on a VPN instance or the public instance (VPN instance view, EVPN view of a VPN instance or the public instance, and public instance view).

  The device uses the import route targets configured for EVPN on a VPN instance or the public instance when accepting the routes.
- **IP prefix advertisement routes**—The device uses the route targets configured for the IPv4 or IPv6 address family on a VPN instance or the public instance when advertising and accepting the routes.

Contents

EVPN VXLAN overview 1
- Network model 1
- Configuration automation 2
- Assignment of traffic to VXLANs 2
  - Traffic from the local site to a remote site 2
  - Traffic from a remote site to the local site 3
- Layer 2 forwarding 3
  - MAC learning 3
  - Unicast 3
  - Flood 4
- Centralized EVPN gateway deployment 5
- Distributed EVPN gateway deployment 6
  - About distributed EVPN gateway deployment 6
  - Symmetric IRB 7
  - Asymmetric IRB 10
- EVPN VXLAN multihoming 11
  - About EVPN multihoming 11
  - DF election 12
  - Split horizon 14
  - Redundancy mode 14
  - IP aliasing 14
- ARP and ND flood suppression 15
- MAC mobility 16
- EVPN M-LAG 16
  - About EVPN M-LAG 16
  - VM reachability information synchronization 16
  - Virtual VTEP address 17
  - Independent BGP neighbor relationship establishment 17
  - Site-facing link redundancy 17

Configuring EVPN VXLAN 0
- EVPN VXLAN **tasks at a glance** 0
- Restrictions and guidelines: EVPN VXLAN configuration 1
- Restrictions: System operating mode 1
- Configuring a VXLAN on a VSI 1
  - Restrictions and guidelines for VXLAN configuration on a VSI 1
  - Creating a VXLAN on a VSI 1
  - Configuring VSI parameters 2
- Configuring an EVPN instance 2
  - About EVPN instance configuration 2
  - Restrictions and guidelines for EVPN instance configuration 2
  - Configuring an EVPN instance created in VSI view 2
- Configuring EVPN multihoming 3
  - Restrictions and guidelines for EVPN multihoming 3
  - Assigning an ESI to an interface 3
  - Configuring the DF election algorithm 4
  - Enabling AC state-based DF election 5
  - Setting the DF election delay 5
  - Enabling non-revertive mode for preference-based DF election 6
  - Configuring ES route advertisement delay 7
  - Disabling advertisement of EVPN multihoming routes 7
- Configuring BGP to advertise BGP EVPN routes 8
  - Restrictions and guidelines for BGP EVPN route advertisement 8
  - Enabling BGP to advertise BGP EVPN routes 8
  - Configuring BGP EVPN route settings 8
  - Enabling the device to ignore default routes in route recursion 12
  - Maintaining BGP sessions 12

Mapping ACs to a VSI 13
Mapping a static Ethernet service instance to a VSI 13
Configuring a centralized EVPN gateway 14
Restrictions and guidelines for centralized EVPN gateway configuration 14
Configuring a centralized gateway interface 14
Setting the static flag for the MAC addresses of centralized gateway interfaces 15
Configuring a distributed EVPN gateway 15
Restrictions and guidelines for distributed EVPN gateway configuration 15
Prerequisites for distributed EVPN gateway configuration 16
Configuring the traffic forwarding mode for EVPN VXLAN 16
Configuring a VSI interface 16
Configuring an L3 VXLAN ID for a VSI interface 17
Configuring IP prefix route advertisement 20
Configuring the EVPN global MAC address 21
Disabling generation of IP prefix advertisement routes for the subnets of a VSI interface 21
Enabling a distributed EVPN gateway to send RA messages over VXLAN tunnels 22
Enabling traffic statistics for the VSIs automatically created for L3 VXLAN IDs 23
Enabling the device to advertise ARP information for the distributed EVPN gateway interfaces through MAC/IP advertisement routes 23
Managing remote MAC address entries and remote ARP or ND learning 24
Disabling remote MAC address learning and remote ARP or ND learning 24
Disabling MAC address advertisement 24
Enabling MAC mobility event suppression 25
Disabling learning of MAC addresses from ARP or ND information 25
Disabling ARP information advertisement 26
Enabling ND information advertisement 26
Disabling the VSI interface on a centralized EVPN gateway from learning ARP or ND information across subnets 27
Enabling ARP mobility event suppression 27
Enabling ND mobility event suppression 28
Enabling ARP request proxy 29
Enabling ND request proxy 29
Configuring BGP EVPN route redistribution and advertisement 30
Redistributing MAC/IP advertisement routes into BGP unicast routing tables 30
Setting the metric of BGP EVPN routes added to a VPN instance's routing table 31
Setting the preference of MAC/IP advertisement routes redistributed to the BGP-VPN unicast address families 32
Enabling BGP EVPN route advertisement to the local site 33
Disabling flooding for a VSI 33
Enabling ARP or ND flood suppression 34
Configuring ARP/ND flood suppression on a VSI 34
Configuring ARP flood suppression for a VLAN 35
Enabling packet statistics for VXLAN tunnels 35
Testing the connectivity of a VXLAN tunnel 36
Enabling overlay OAM 36
Pinging a VXLAN tunnel destination 36
Tracing the path to a VXLAN tunnel destination 37
Enabling SNMP notifications for EVPN 38
Configuring EVPN M-LAG 38
About this task 38
Restrictions and guidelines 39
Prerequisites 40
Procedure (IPv4) 41
Procedure (IPv6) 41
Configuring EVPN VXLAN tunnels to learn dynamic entries 42
Display and maintenance commands for EVPN 42
EVPN VXLAN configuration examples 45
Example: Configuring a centralized EVPN gateway 45
Example: Configuring distributed EVPN gateways in symmetric IRB mode (IPv4 underlay network) 52
Example: Configuring distributed EVPN gateways in symmetric IRB mode (IPv6 underlay network) 61
Example: Configuring distributed IPv4 EVPN gateways in asymmetric IRB mode 70
Example: Configuring communication between EVPN networks and the public network 79

Example: Configuring IPv4 EVPN M-LAG with a direct peer link ····· 88
Example: Configuring IPv4 EVPN M-LAG with a tunnel peer link ····· 100
Example: Configuring IPv4 EVPN multihoming ····· 111

EVPN VXLAN overview

EVPN VXLAN uses EVPN routes for automatic VXLAN tunnel establishment and assignment and MAC reachability information advertisement in the control plane and uses VXLAN for forwarding in the data plane.

Network model

As shown in Figure 1, EVPN uses the VXLAN technology for traffic forwarding in the data plane. The transport edge devices assign user terminals to different VXLANs, and then forward traffic between sites for user terminals by using VXLAN tunnels. The transport edge devices are VXLAN tunnel endpoints (VTEPs).

The EVPN network sites and transport network can be IPv4 or IPv6 networks. Supported user terminals include PCs and VMs on servers.

NOTE:

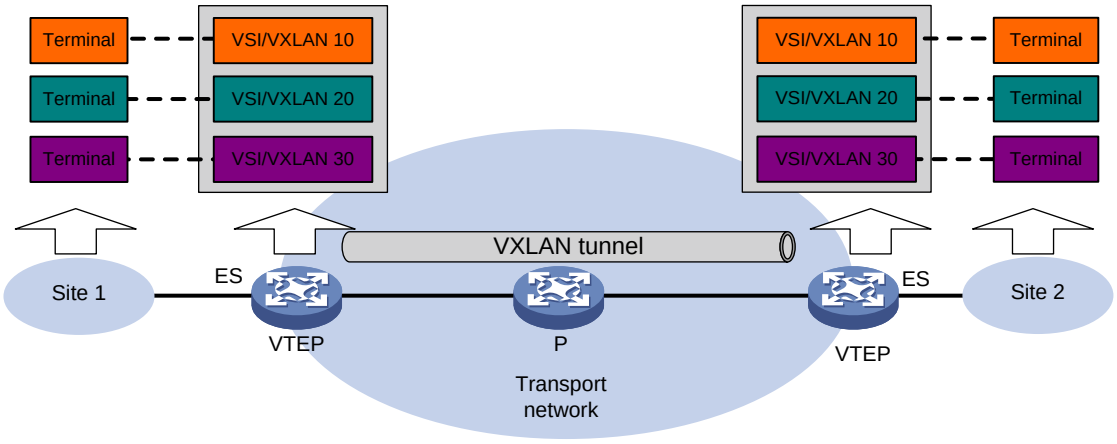
This document uses VMs as examples to describe the mechanisms of EVPN. The mechanisms do not differ between different kinds of user terminals.

A VTEP uses ESs, VSIs, and VXLAN tunnels to provide VXLAN services:

- **Ethernet segment (ES)**—An ES is a link that connects a site to a VTEP. Each ES is uniquely identified by an Ethernet segment identifier (ESI).
- **VSI**—A virtual switch instance is a virtual Layer 2 switched domain. Each VSI provides switching services only for one VXLAN. VSIs learn MAC addresses and forward frames independently of one another. User terminals in different sites have Layer 2 connectivity if they are in the same VXLAN. A VXLAN is identified by a 24-bit VXLAN ID which is also called the virtual network identifier (VNI). A VXLAN corresponds to an EVPN instance.
- **VXLAN tunnel**—Logical point-to-point tunnels between VTEPs over the transport network. Each VXLAN tunnel can trunk multiple VXLANs.

All VXLAN processing is performed on VTEPs. The ingress VTEP encapsulates VXLAN traffic in the VXLAN, outer UDP, and outer IP headers, and forwards the traffic through VXLAN tunnels. The egress VTEP removes the VXLAN encapsulation and forwards the traffic to the destination. Transport network devices (for example, the P device in Figure 1) forward VXLAN traffic only based on the outer IP header of VXLAN packets.

Figure 1 EVPN network model



Configuration automation

If EVPN is used for Layer 2 forwarding, VTEPs use the following BGP EVPN routes to discover VTEP neighbors, establish VXLAN tunnels, and assign the tunnels to VXLANs:

- **IMET route**—VTEPs advertise the VXLAN IDs they have through IMET routes. If two VTEPs have the same VXLAN ID, they automatically establish a VXLAN tunnel and assign the tunnel to the VXLAN.
- **MAC/IP advertisement route**—VTEPs advertise local MAC addresses and VXLAN IDs through MAC/IP advertisement routes. If two VTEPs have the same VXLAN ID, they automatically establish a VXLAN tunnel and assign the tunnel to the VXLAN.

If EVPN is used for Layer 3 forwarding, VTEPs use the following BGP EVPN routes to discover VTEP neighbors, establish VXLAN tunnels, and assign the tunnels to VXLANs:

- **IMET route**—VTEPs advertise the VXLAN IDs they have through IMET routes. If two VTEPs have the same VXLAN ID, they automatically establish a VXLAN tunnel and assign the tunnel to the VXLAN.
- **MAC/IP advertisement route and IP prefix advertisement route**—In the EVPN gateway deployment, VTEPs advertise MAC/IP advertisement routes or IP prefix advertisement routes with the export targets. When a VTEP receives a route, it compares the export targets of the route with the local import targets. If the route targets match, the VTEP establishes a VXLAN tunnel with the remote VTEP and associates the tunnel with the L3 VXLAN ID of the corresponding VPN instance. For more information about the L3 VXLAN ID, see "Distributed EVPN gateway deployment."

Assignment of traffic to VXLANs

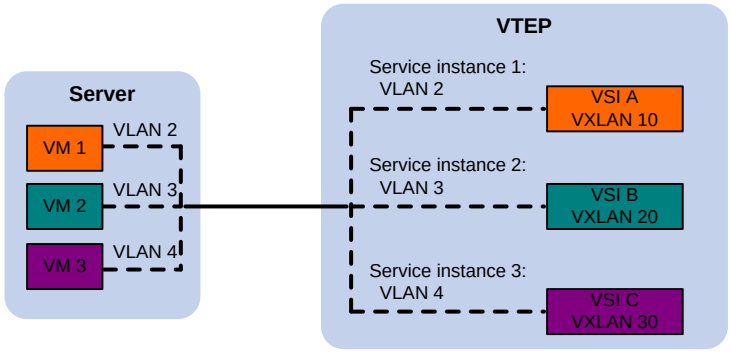
Traffic from the local site to a remote site

The VTEP uses an Ethernet service instance to match customer traffic on a site-facing interface. The VTEP assigns customer traffic to a VXLAN by mapping the Ethernet service instance to a VSI.

An Ethernet service instance is identical to an attachment circuit (AC) in L2VPN. An Ethernet service instance matches a list of VLANs on a Layer 2 Ethernet interface by using a frame match criterion. The frame match criterion specifies the characteristics of traffic from the VLANs, such as tagging status and VLAN IDs.

As shown in Figure 2, Ethernet service instance 1 matches VLAN 2 and is mapped to VSI A (VXLAN 10). When a frame from VLAN 2 arrives, the VTEP assigns the frame to VXLAN 10, and looks up VSI A's MAC address table for the outgoing interface.

Figure 2 Identifying traffic from the local site



Traffic from a remote site to the local site

When a VXLAN packet arrives at a VXLAN tunnel interface, the VTEP uses the VXLAN ID in the packet to identify its VXLAN.

Layer 2 forwarding

MAC learning

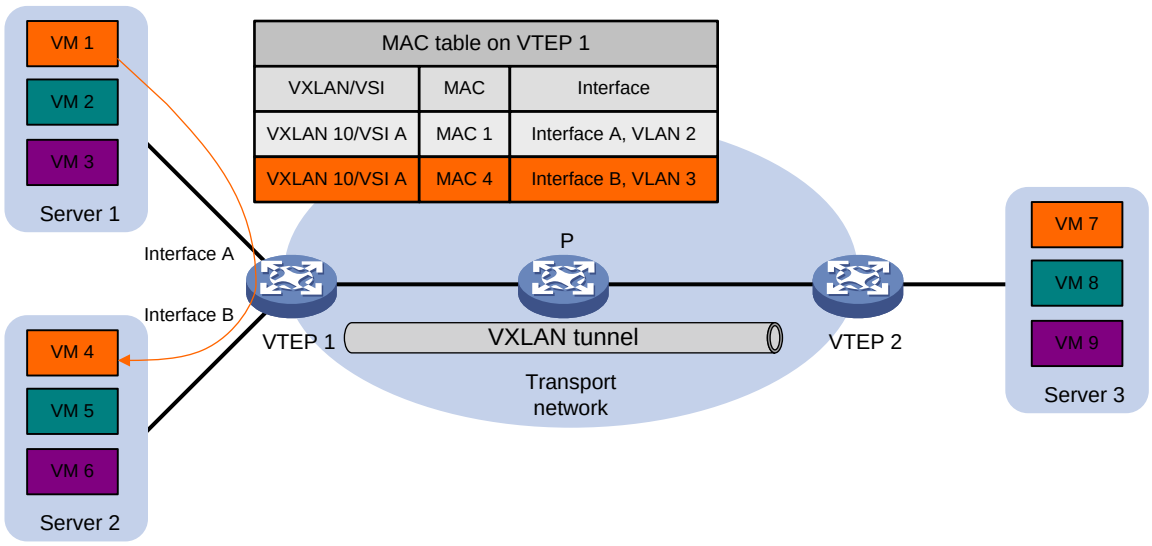
The VTEP performs Layer 2 forwarding based on a VSI's MAC address table. The VTEP learns MAC addresses by using the following methods:

- **Local MAC learning**—The VTEP automatically learns the source MAC addresses of frames sent from the local site. The outgoing interfaces of local MAC address entries are site-facing interfaces on which the MAC addresses are learned.
- **Remote MAC learning**—The VTEP uses MP-BGP to advertise local MAC reachability information to remote sites and learn MAC reachability information from remote sites. The outgoing interfaces of MAC address entries advertised from a remote site are VXLAN tunnel interfaces.

Unicast

As shown in Figure 3, the VTEP performs typical Layer 2 forwarding for known unicast traffic within the local site.

Figure 3 Intra-site unicast



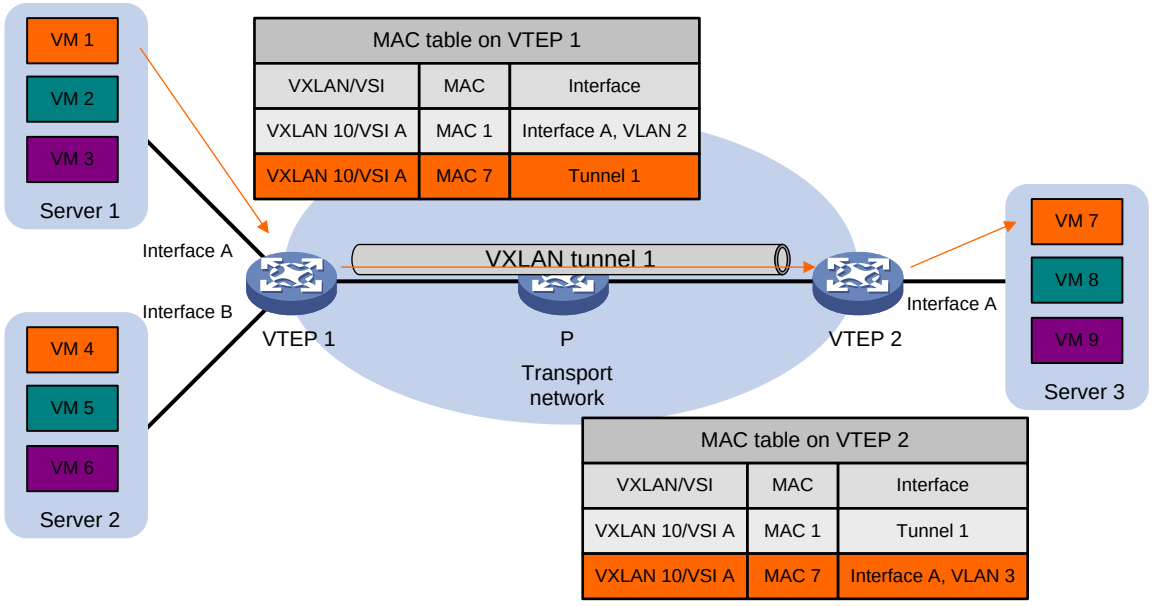
As shown in Figure 4, the following process applies to a known unicast frame between sites:

1. The source VTEP encapsulates the Ethernet frame in the VXLAN/UDP/IP header.

   In the outer IP header, the source IP address is the source VTEP's VXLAN tunnel source IP address. The destination IP address is the VXLAN tunnel destination IP address.
2. The source VTEP forwards the encapsulated packet out of the outgoing VXLAN tunnel interface found in the VSI's MAC address table.
3. The intermediate transport devices (P devices) forward the packet to the destination VTEP by using the outer IP header.

4. The destination VTEP removes the headers on top of the inner Ethernet frame. It then performs MAC address table lookup in the VXLAN's VSI to forward the frame out of the matching outgoing interface.

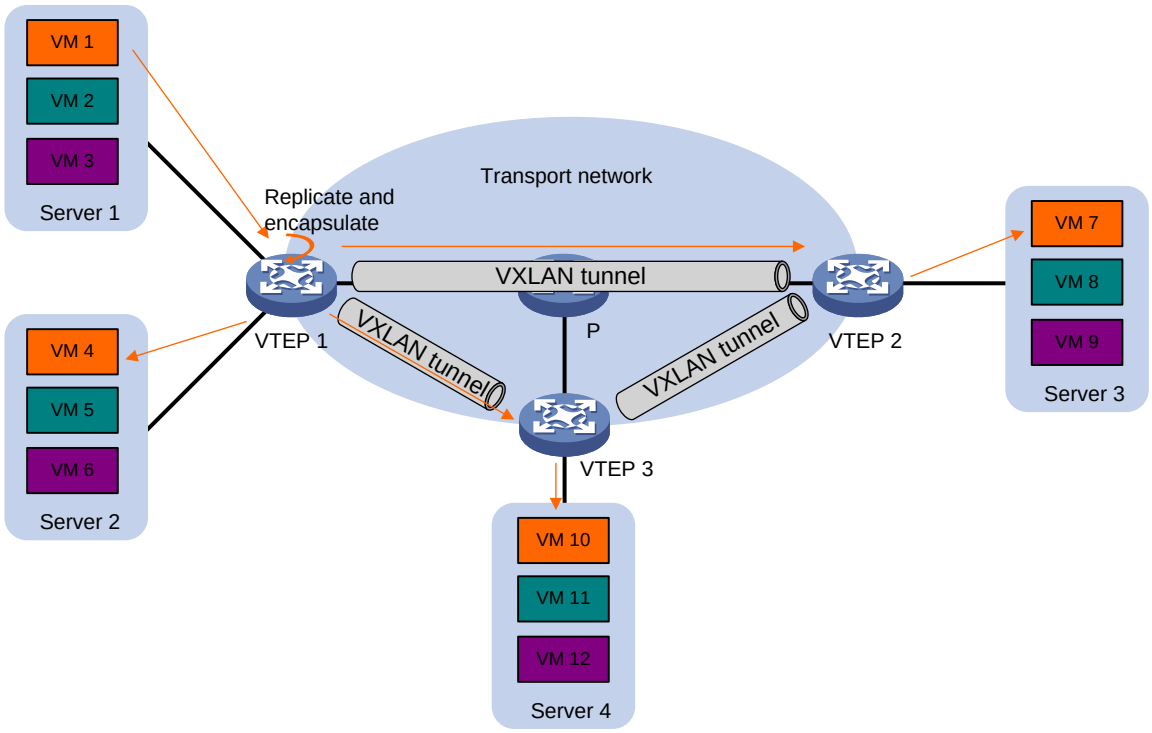
Figure 4 Inter-site unicast



Flood

As shown in Figure 5, a VTEP floods a broadcast, multicast, or unknown unicast frame to all site-facing interfaces and VXLAN tunnels in the VXLAN, except for the incoming interface. The source VTEP replicates the flood frame, and then sends one replica to the destination IP address of each VXLAN tunnel in the VXLAN. Each destination VTEP floods the inner Ethernet frame to all the site-facing interfaces in the VXLAN. To avoid loops, the destination VTEPs do not flood the frame to VXLAN tunnels.

Figure 5 Forwarding of flood traffic



Centralized EVPN gateway deployment

IMPORTANT:

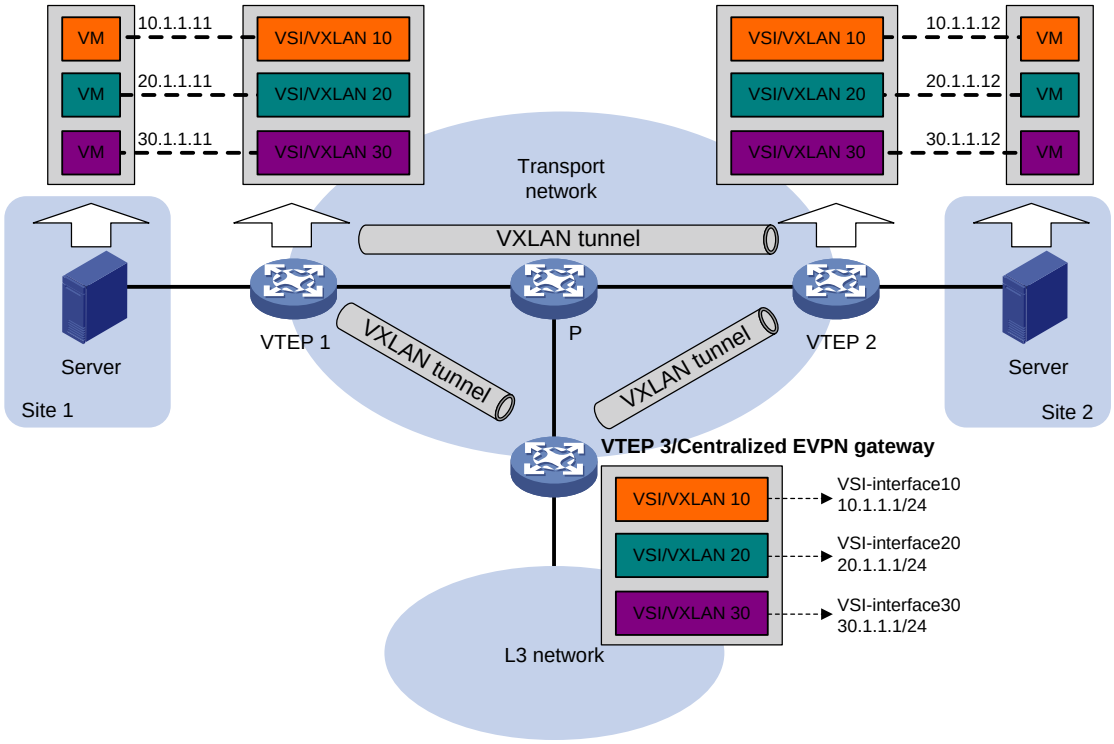
This section uses IPv4 sites as examples to describe the Layer 3 forwarding process of EVPN networks. The Layer 3 forwarding process does not differ between IPv4 and IPv6 sites.

Centralized EVPN gateway deployment uses one VTEP to provide Layer 3 forwarding for VXLANs. The VTEP uses virtual Layer 3 VSI interfaces as gateway interfaces for VXLANs. Typically, the gateway-collocated VTEP connects to other VTEPs and the external network. To use this design, make sure the gateway has sufficient bandwidth and processing capability.

As shown in Figure 6, a VTEP acts as a gateway for VMs in the VXLANs. The VTEP both terminates the VXLANs and performs Layer 3 forwarding for the VMs. The network uses the following process to forward Layer 3 traffic from a VM to the destination:

1. The VM sends an ARP request to obtain the MAC address of the VSI interface that acts as the gateway, and then sends the Layer 3 traffic to the centralized EVPN gateway.
2. The local VTEP looks up the matching VSI's MAC address table and forwards the traffic to the centralized EVPN gateway through a VXLAN tunnel.
3. The centralized EVPN gateway removes the VXLAN encapsulation and forwards the traffic at Layer 3.
4. The centralized EVPN gateway forwards the replies sent by the destination node to the VM based on the ARP entry for the VM.

Figure 6 Example of centralized EVPN gateway deployment



Distributed EVPN gateway deployment

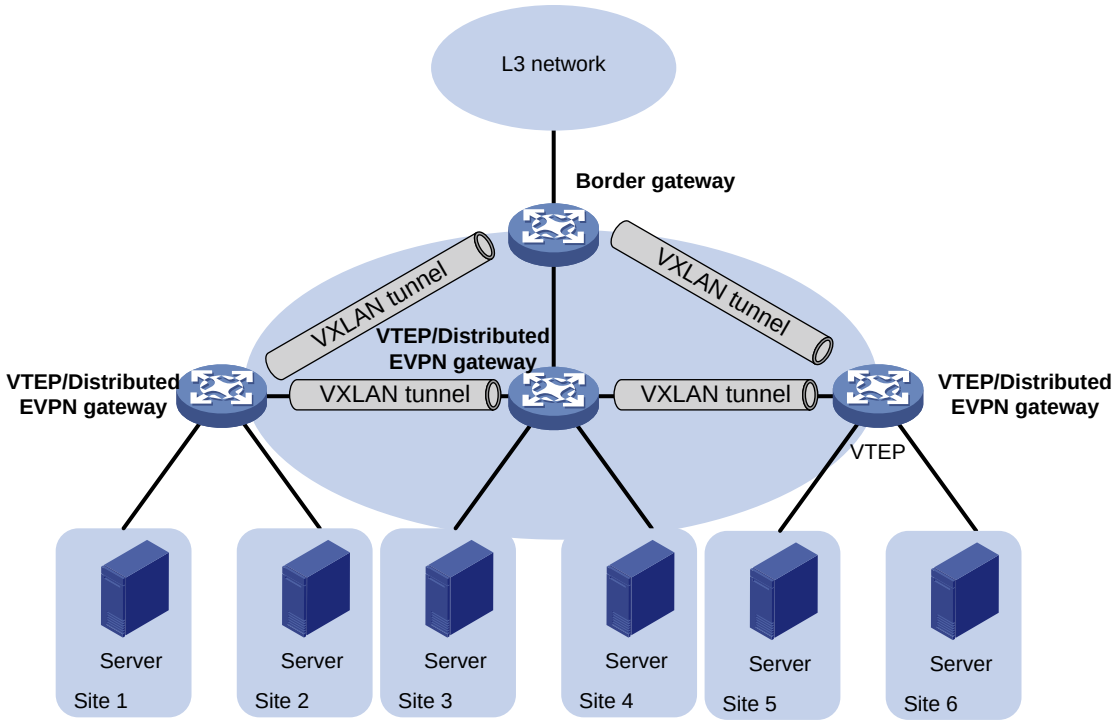
IMPORTANT:

This section uses IPv4 sites as examples to describe the Layer 3 forwarding process of EVPN networks. The Layer 3 forwarding process does not differ between IPv4 and IPv6 sites.

About distributed EVPN gateway deployment

As shown in Figure 7, each site's VTEP acts as a gateway to perform Layer 3 forwarding for the VXLANs of the local site. A VTEP acts as a border gateway to the Layer 3 network for the VXLANs.

Figure 7 Distributed EVPN gateway placement design



A distributed EVPN gateway supports the following traffic forwarding modes:

- **Asymmetric IRB**—The ingress gateway performs Layer 2 and Layer 3 lookups and the egress gateway performs only Layer 2 forwarding.
- **Symmetric IRB**—Both the ingress and egress gateways perform Layer 2 and Layer 3 lookups.

Symmetric IRB

Basic concepts

Symmetric IRB introduces the following concepts:

- **L3 VXLAN ID**—Also called L3 VNI. An L3 VXLAN ID identifies the traffic of a routing domain where devices have Layer 3 reachability. An L3 VXLAN ID is associated with one VPN instance. Distributed EVPN gateways use VPN instances to isolate traffic of different services on VXLAN tunnel interfaces.
- **Router MAC address**—Each distributed EVPN gateway has a unique router MAC address used for inter-gateway forwarding. The MAC addresses in the inner Ethernet header of VXLAN packets are router MAC addresses of distributed EVPN gateways.

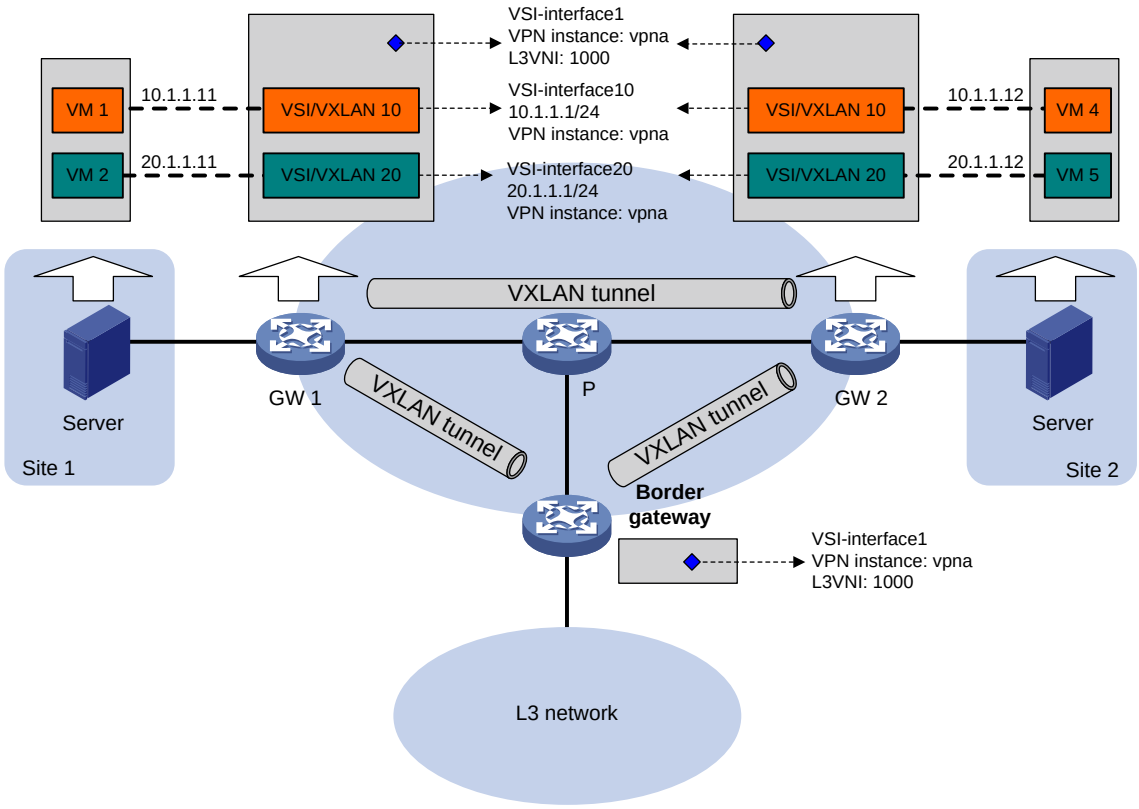
VSI interfaces

As shown in Figure 8, each distributed EVPN gateway has the following types of VSI interfaces:

- **VSI interface as a gateway interface of a VXLAN**—The VSI interface acts as the gateway interface for VMs in a VXLAN. The VSI interface is associated with a VSI and a VPN instance. On different distributed EVPN gateways, the VSI interface of a VXLAN use the same IP address to provide services.
- **VSI interface associated with an L3 VXLAN ID**—The VSI interface is associated with a VPN instance and assigned an L3 VXLAN ID. VSI interfaces associated with the same VPN instance share an L3 VXLAN ID.

A border gateway only has VSI interfaces that are associated with an L3 VXLAN ID.

Figure 8 Example of distributed EVPN gateway deployment



Layer 3 forwarding entry learning

A distributed EVPN gateway forwards Layer 3 traffic based on FIB entries generated from BGP EVPN routes and ARP information.

A VTEP advertises an external route imported in the EVPN address family through MP-BGP. A remote VTEP adds the route to the FIB table of a VPN instance based on the L3 VXLAN ID carried in the route. In the FIB entry, the outgoing interface is a VXLAN tunnel interface, and the next hop is the peer VTEP address in the NEXT_HOP attribute of the route.

A VTEP has the following types of ARP information:

- **Local ARP information**—ARP information of VMs in the local site. The VTEP snoops GARP packets, RARP packets, and ARP requests for the gateway MAC address to learn the ARP information of the senders and generates ARP entries and FIB entries. In an ARP or FIB entry, the outgoing interface is the site-facing interface where the packet is received, and the VPN instance is the instance associated with the corresponding VSI interface.
- **Remote ARP information**—ARP information of VMs in remote sites. Each VTEP uses MP-BGP to advertise its local ARP information with L3 VXLAN IDs in routes to remote sites. A VTEP generates only FIB entries for the remote ARP information. A FIB entry contains the following information:
  - Outgoing interface: VSI interface associated with the L3 VXLAN ID.
  - Next hop: Peer VTEP address in the NEXT_HOP attribute of the route.
  - VPN instance: VPN instance associated with the L3 VXLAN ID.

  The VTEP then creates an ARP entry for the next hop in the FIB entry.

Traffic forwarding

A distributed EVPN gateway can work in one of the following mode:

- **Switching and routing mode**—Forwards Layer 2 traffic based on the MAC address table and forwards Layer 3 traffic based on the FIB table. In this mode, you need to enable ARP flood suppression on the distributed EVPN gateway to reduce flooding.
- **Routing mode**— Forwards both Layer 2 and Layer 3 traffic based on the FIB table. In this mode, you need to enable local proxy ARP on the distributed EVPN gateway.

For more information about MAC address table-based Layer 2 forwarding, see "Unicast."

Figure 9 shows the intra-site Layer 3 forwarding process.

1. The source VM sends an ARP request to obtain the MAC address of the destination VM.
2. The gateway replies to the source VM with the MAC address of the VSI interface associated with the source VM's VSI.
3. The source VM sends a Layer 3 packet to the gateway.
4. The gateway looks up the FIB table of the VPN instance associated with the source VM's VSI and finds the matching outgoing site-facing interface.
5. The gateway processes the Ethernet header of the Layer 3 packet as follows:
   - Replaces the destination MAC address with the destination VM's MAC address.
   - Replaces the source MAC address with the VSI interface's MAC address.
6. The gateway forwards the Layer 3 packet to the destination VM.

Figure 9 Intra-site Layer 3 forwarding

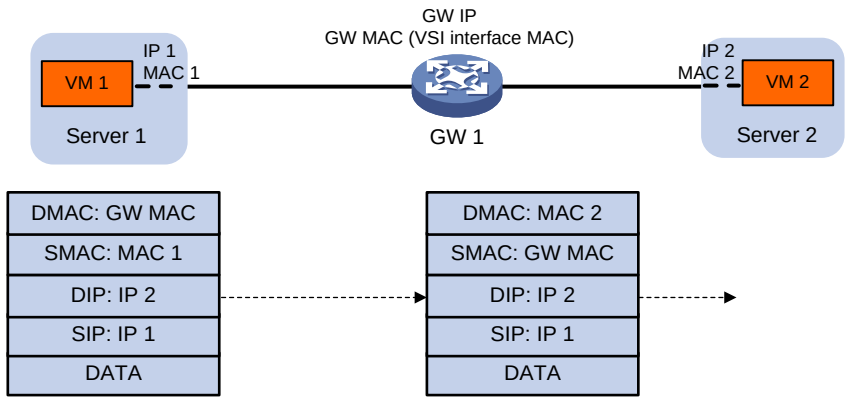
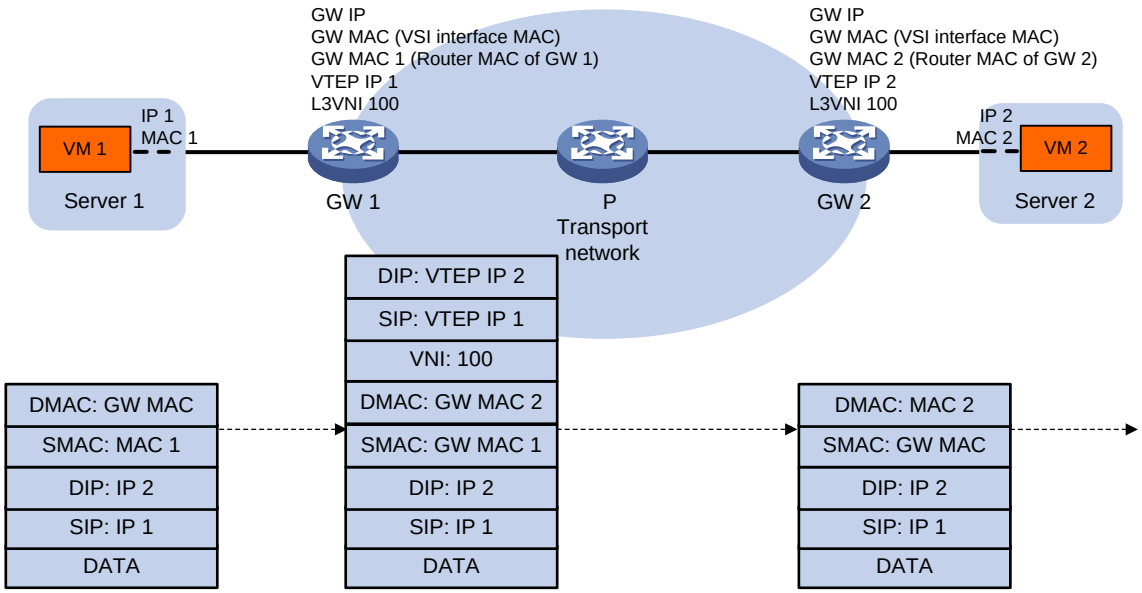


Figure 10 shows the inter-site Layer 3 forwarding process.

7. The source VM sends an ARP request to obtain the MAC address of the destination VM.
8. The gateway replies to the source VM with the MAC address of the VSI interface associated with the source VM's VSI.
9. The source VM sends a Layer 3 packet to the gateway.
10. The gateway looks up the FIB table of the VPN instance associated with the source VM's VSI and finds the matching outgoing VSI interface.
11. The gateway processes the Ethernet header of the Layer 3 packet as follows:
    - Replaces the destination MAC address with the destination gateway's router MAC address.
    - Replaces the source MAC address with its own router MAC address.
12. The gateway adds VXLAN encapsulation to the Layer 3 packet and forwards the packet to the destination gateway. The encapsulated VXLAN ID is the L3 VXLAN ID of the corresponding VPN instance.
13. The destination gateway identifies the VPN instance of the packet based on the L3 VXLAN ID and removes the VXLAN encapsulation. Then the gateway forwards the packet based on the matching ARP entry.

Figure 10 Inter-site Layer 3 forwarding



Communication between private and public networks

A distributed EVPN gateway uses the public instance to perform Layer 3 forwarding for the public network and to enable communication between private and public networks. The public instance is similar to a VPN instance. A distributed EVPN gateway processes traffic of the public instance in the same way it does for a VPN instance. For the public instance to work correctly, you must configure an RD, an L3 VXLAN ID, and route targets for it. If a VSI interface is not associated with any VPN instance, the VSI interface belongs to the public instance.

Asymmetric IRB

VSI interfaces

Asymmetric IRB uses the same distributed EVPN gateway deployment as symmetric IRB.

As shown in Figure 11, each distributed EVPN gateway has the following types of VSI interfaces:

- **VSI interface as a gateway interface of a VXLAN**—The VSI interface is associated with a VSI and a VPN instance. On different distributed EVPN gateways, the VSI interface of a VXLAN must use different IP addresses to provide services.
- **VSI interface associated with an L3 VXLAN ID**—The VSI interface acts as the gateway for VMs in a VXLAN to communicate with the external network through the border gateway. The VSI interface is associated with a VPN instance and assigned an L3 VXLAN ID. VSI interfaces associated with the same VPN instance share an L3 VXLAN ID.

A border gateway only has VSI interfaces that are associated with an L3 VXLAN ID.

Figure 11 Example of distributed EVPN gateway deployment

Layer 3 forwarding

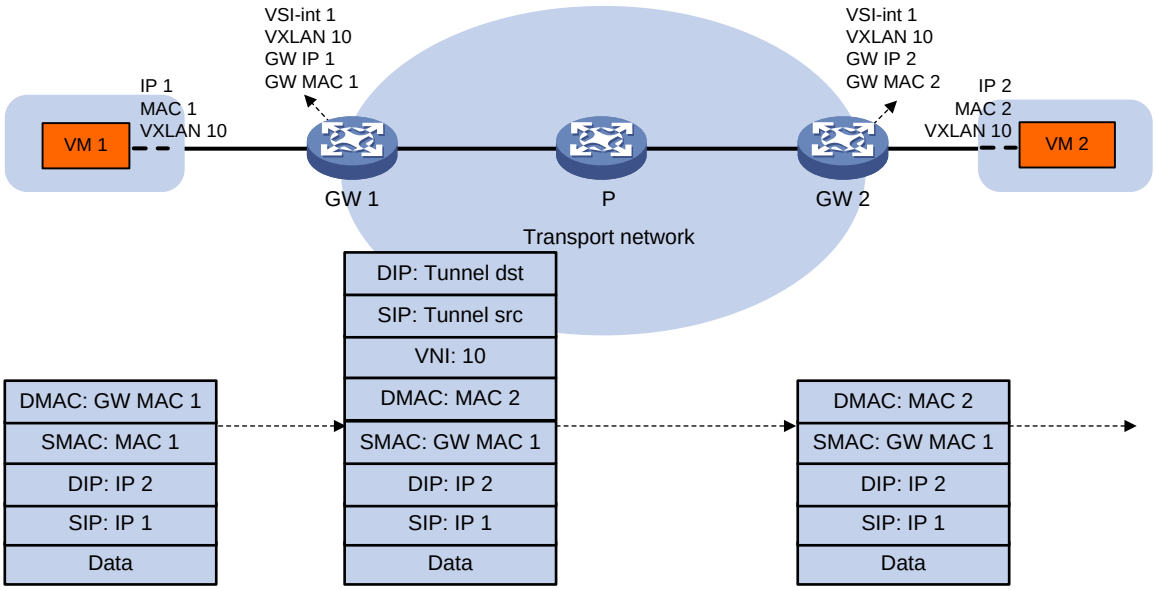
Asymmetric IRB supports only Layer 3 forwarding in the same VXLAN on distributed EVPN gateways.

After a distributed EVPN gateway learns ARP information about local VMs, it advertises the information to other distributed EVPN gateways through MAC/IP advertisement routes. Other distributed EVPN gateways generate FIB entries based on the advertised ARP information.

As shown in Figure 12, VM 1 and VM 2 belong to VXLAN 10 and they can reach each other at Layer 3 through the distributed EVPN gateways. The distributed EVPN gateways use the following process to perform Layer 3 forwarding in asymmetric IRB mode when VM 1 sends a packet to VM 2:

1. After GW 1 receives the packet from VM 1, it finds that the destination MAC address is itself. Then, GW 1 removes the Layer 2 frame header and looks up the FIB table for the destination IP address.
2. GW 1 matches the packet to the FIB entry generated based on the ARP information of VM 2.
3. GW 1 encapsulates the packet source and destination MAC addresses as the MAC addresses of GW 1 and VM 2, respectively. Then, GW 1 adds VXLAN encapsulation to the packet and forwards the packet to GW 2 through a VXLAN tunnel.
4. GW 2 removes the VXLAN encapsulation from the packet, and performs Layer 2 forwarding in VXLAN 10 by looking up the MAC address table for the destination MAC address.
5. GW 2 forwards the packet to VM 2 based on the MAC address table lookup result.

Figure 12 Layer 3 forwarding in the same VXLAN (asymmetric IRB)



EVPN VXLAN multihoming

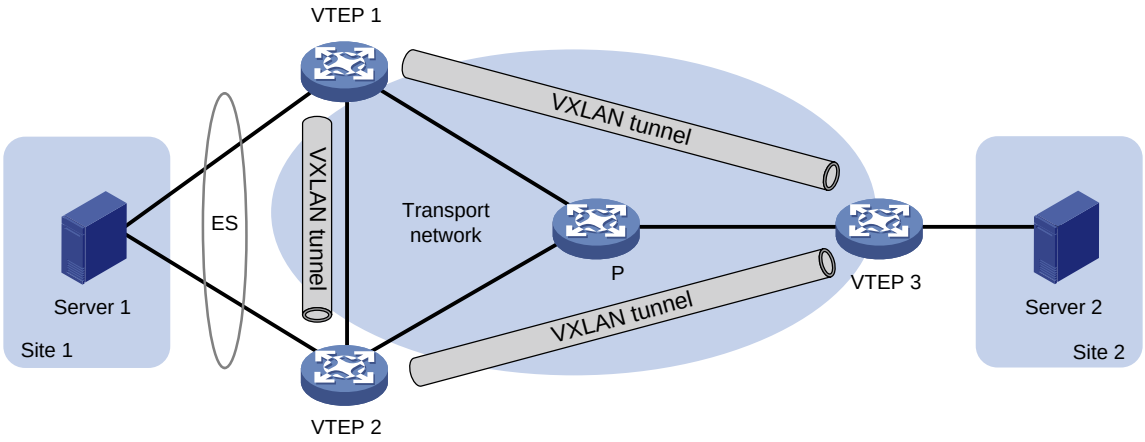
IMPORTANT:

EVPN multihoming supports only IPv4 over IPv4 overlay and IPv6 over IPv6 overlay.

About EVPN multihoming

As shown in Figure 13, EVPN supports deploying multiple VTEPs at a site for redundancy and high availability. On the redundant VTEPs, Ethernet links connected to the site form an Ethernet segment (ES) that is uniquely identified by an Ethernet segment identifier (ESI).

Figure 13 EVPN multihoming

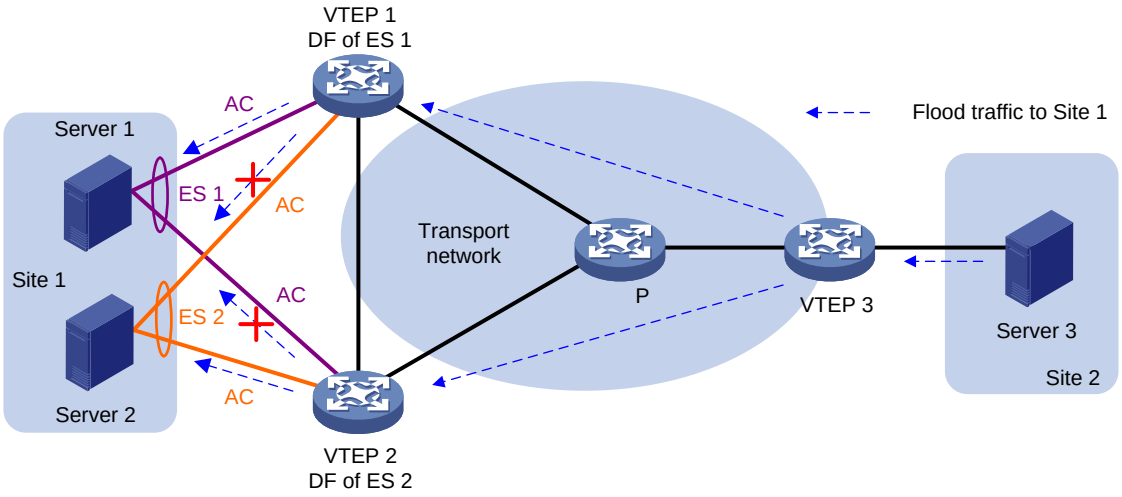


DF election

To prevent redundant VTEPs from sending duplicate flood traffic to a multihomed site, a designated forwarder (DF) is elected from the VTEPs to forward flood traffic to the site. VTEPs that fail the election are assigned the backup designated forwarder (BDF) role. BDFs do not forward flood traffic to the site.

Redundant VTEPs at a site send Ethernet segment routes to one another to advertise ES and VTEP IP mappings. A VTEP accepts the Ethernet segment routes only when it is configured with an ESI. Then, the VTEPs select a DF for each AC based on the ES and VTEP IP mappings. DF election can be performed by using a VLAN tag-based algorithm or preference-based algorithm.

Figure 14 DF election



VLAN tag-based DF election

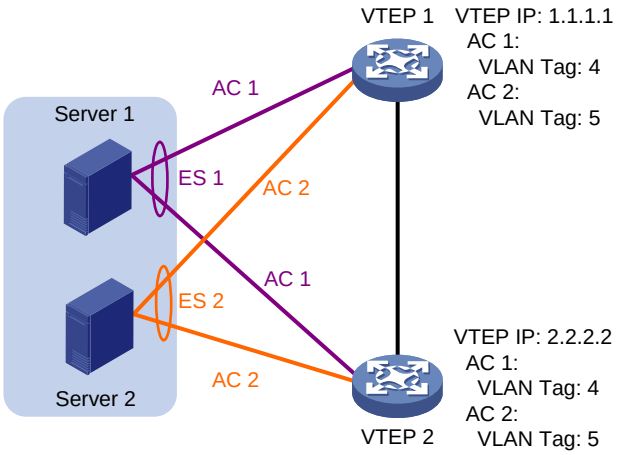
VTEPs select a DF for each AC based on the VLAN tag and VTEP IP address as follows:

1. Arrange source IP addresses in Ethernet segment routes with the same ESI in ascending order and assign a sequence number to each IP address, starting from 0.
2. Divide the lowest VLAN ID permitted on an AC by the number of the redundant VTEPs, and match the reminder to the sequence numbers of IP addresses.
3. Assign the DF role to the VTEP that uses the IP address with the matching sequence number.

The following uses AC 1 in Figure 15 as an example to explain the DF election procedure:

4. VTEP 1 and VTEP 2 send Ethernet segment routes to each other.
5. The VTEPs assign sequence numbers 0 and 1 to IP addresses 1.1.1.1 and 2.2.2.2 in the Ethernet segment routes, respectively.
6. The VTEPs divide 4 (the lowest VLAN ID permitted by AC 1) by 2 (the number of redundant VTEPs), and match the reminder 0 to the sequence numbers of the IP addresses.
7. The DF role is assigned to VTEP 1 at 1.1.1.1.

Figure 15 VLAN tag-based DF election



Preference-based DF election

VTEPs select a DF for each ES based on the DF election preference, the Don't Preempt Me (DP) bit in Ethernet segment routes, and VTEP IP address. The DP bit can be set to one of the following values:

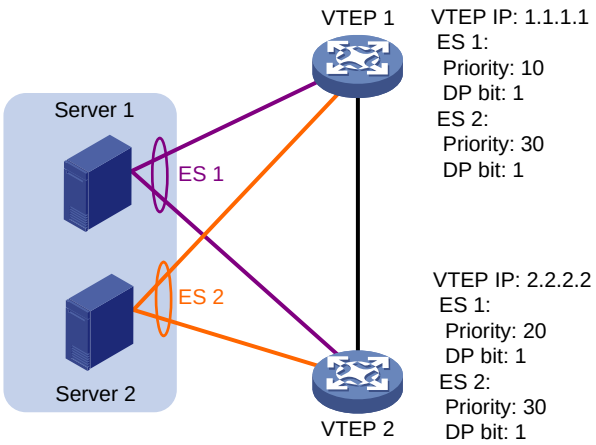
- **1**—DF preemption is disabled. A DF retains its role when a new DF is elected.
- **0**—DF preemption is enabled.

Preference-based DF election uses the following rules to select a DF for an ES:

- The VTEP with higher preference becomes the DF.
- If two VTEPs have the same preference, the VTEP with the DP bit set to 1 becomes the DF. If both of the VTEPs have the DP bit set to 1, the VTEP with a lower IP address becomes the DF.

As shown in Figure 16, VTEP 2 is the DF for ES 1, and VTEP 1 is the DF for ES 2.

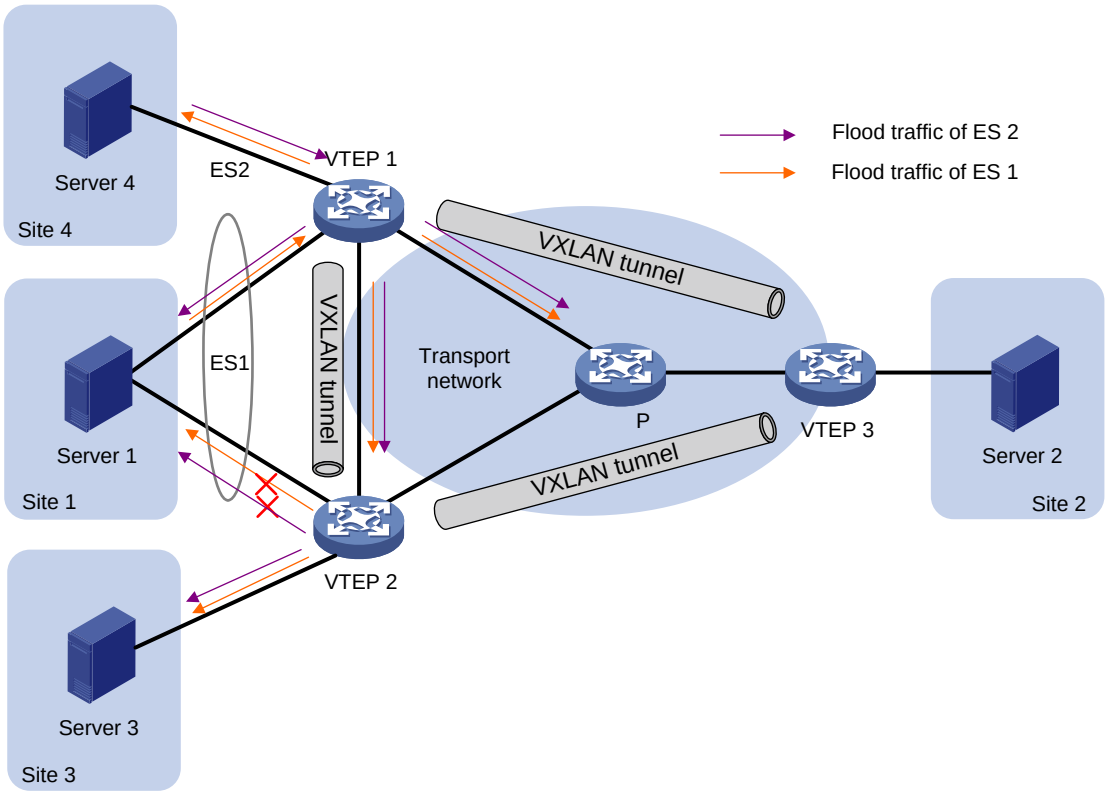
Figure 16 Preference-based DF election



Split horizon

In a multihomed site, a VTEP forwards multicast, broadcast, and unknown unicast frames received from ACs out of all site-facing interfaces and VXLAN tunnels in the corresponding VXLAN, except for the incoming interface. As a result, the other VTEPs at the site receive these flood frames and forward them to site-facing interfaces, which causes duplicate floods and loops. EVPN introduces split horizon to resolve this issue. Split horizon disables a VTEP from forwarding flood traffic received from another local VTEP to site-facing interfaces if an ES on that local VTEP has the same ESI as these interfaces. As shown in Figure 17, both VTEP 1 and VTEP 2 have ES 1. When receiving flood traffic from VTEP 1, VTEP 2 does not forward the traffic to interfaces with ESI 1.

Figure 17 Split horizon



Redundancy mode

The device supports the all-active redundancy mode of EVPN multihoming. This mode allows all redundant VTEPs at a multihomed site to forward broadcast, multicast, and unknown unicast traffic.

- For flood frames received from remotes sites, a VTEP forwards them to the ACs of which it is the DF.
- For flood frames received from the local site, a VTEP forwards them out of all site-facing interfaces and VXLAN tunnels in the corresponding VXLAN, except for the incoming interfaces. For flood frames to be sent out of a VXLAN tunnel interface, a VTEP replicates each flood frame and sends one replica to all the other VTEPs in the corresponding VXLAN.

IP aliasing

In all-active redundancy mode, all redundant VTEPs of an ES advertise the ES to remote VTEPs through MP-BGP. IP aliasing allows a remote VTEP to add the IP addresses of all the redundant

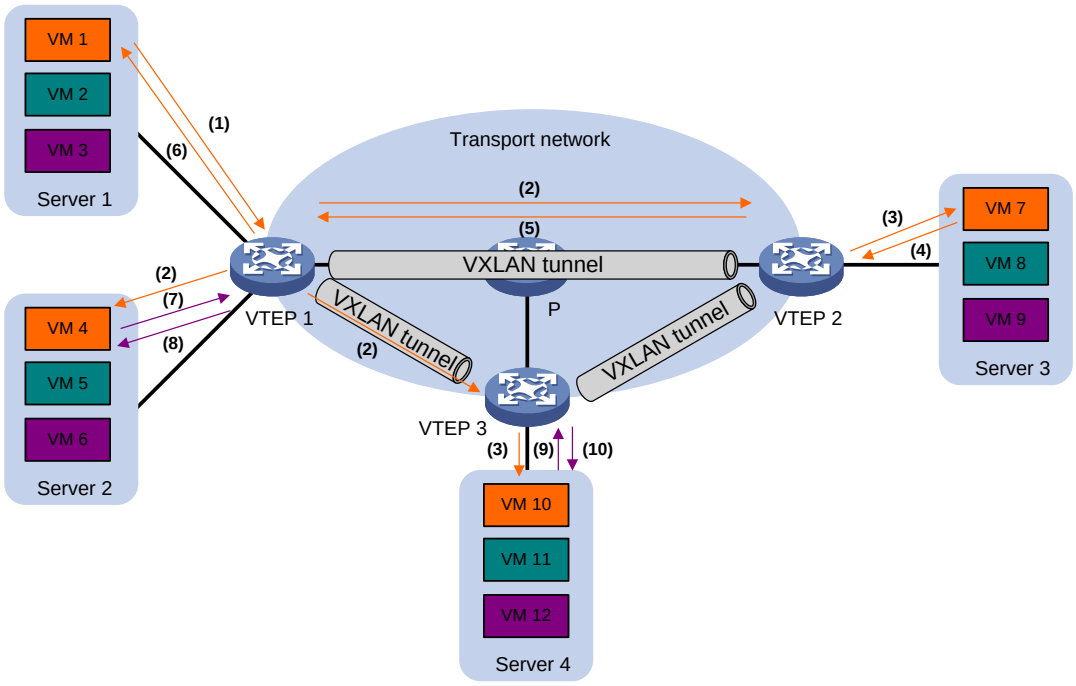
VTEPs as the next hops for the MAC or ARP information received from one of these VTEPs. This mechanism creates ECMP routes between the remote VTEP and the redundant VTEPs.

ARP and ND flood suppression

ARP or ND flood suppression reduces ARP request broadcasts or ND request multicasts by enabling the VTEP to reply to ARP or ND requests on behalf of VMs.

As shown in Figure 18, this feature snoops ARP or ND requests, ARP or ND responses, and BGP EVPN routes to populate the ARP or ND flood suppression table with local and remote MAC addresses. If an ARP or ND request has a matching entry, the VTEP replies to the request on behalf of the VM. If no match is found, the VTEP floods the request to both local and remote sites.

Figure 18 ARP and ND flood suppression



The following uses ARP flood suppression as an example to explain the flood suppression workflow:

1. VM 1 sends an ARP request to obtain the MAC address of VM 7.
2. VTEP 1 creates a suppression entry for VM 1, floods the ARP request in the VXLAN, and sends the suppression entry to VTEP 2 and VTEP 3 through BGP EVPN.
3. VTEP 2 and VTEP 3 de-encapsulate the ARP request and broadcast the request in the local site.
4. VM 7 sends an ARP reply.
5. VTEP 2 creates a suppression entry for VM 7, forwards the ARP reply to VTEP 1, and sends the suppression entry to VTEP 1 and VTEP 3 through BGP EVPN.
6. VTEP 1 de-encapsulates the ARP reply and forwards the ARP reply to VM 1.
7. VM 4 sends an ARP request to obtain the MAC address of VM 1.
8. VTEP 1 creates a suppression entry for VM 4 and replies to the ARP request.
9. VM 10 sends an ARP request to obtain the MAC address of VM 1.
10. VTEP 3 creates a suppression entry for VM 10 and replies to the ARP request.

MAC mobility

MAC mobility refers to that a VM or host moves from one ES to another. The source VTEP is unaware of the MAC move event. To notify other VTEPs of the change, the destination VTEP advertises a MAC/IP advertisement route for the MAC address. The source VTEP withdraws the old route for the MAC address after receiving the new route. The MAC/IP advertisement route has a sequence number that increases when the MAC address moves. The sequence number identifies the most recent move if the MAC address moves multiple times.

EVPN M-LAG

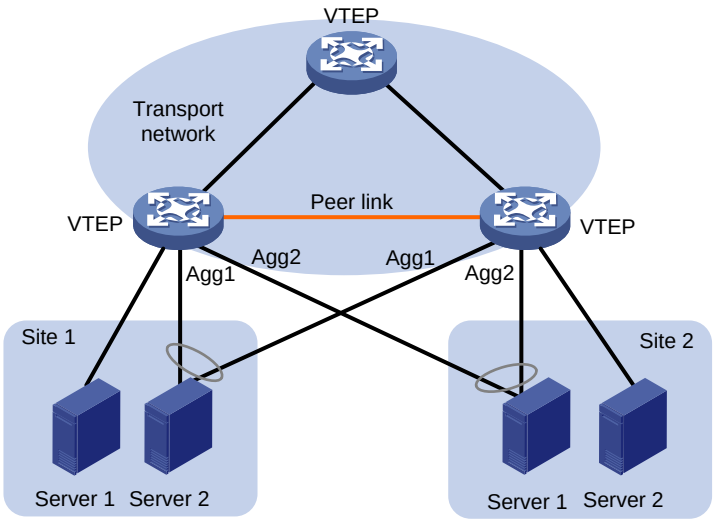
NOTE:

You can use EVPN distributed relay only on IPv4 sites extended by IPv4 underlay networks.

About EVPN M-LAG

As shown in Figure 19, EVPN multichassis link aggregation (M-LAG) virtualizes two VTEPs or EVPN gateways into one M-LAG system through M-LAG to avoid single points of failure. The VTEPs or EVPN gateways are called M-LAG member devices. For more information about M-LAG, see *Layer 2—LAN Switching Configuration Guide*.

Figure 19 EVPN M-LAG



VM reachability information synchronization

To ensure VM reachability information consistency in the M-LAG system, the M-LAG member devices synchronize MAC address entries and ARP or ND information with each other through a peer link. When the peer link is an Ethernet aggregate link, it is referred to as a direct peer link. When the peer link is a VXLAN tunnel, it is referred to as a tunnel peer link.

IMPORTANT:

- The VXLAN tunnel that acts as the peer link is automatically associated with all VXLANs on each M-LAG member device.

Virtual VTEP address

The M-LAG member devices use a virtual VTEP address to set up VXLAN tunnels with remote VTEPs or EVPN gateways.

Independent BGP neighbor relationship establishment

The M-LAG member devices use different BGP peer addresses to establish neighbor relationships with remote devices. For load sharing and link redundancy, a neighbor sends traffic destined for the virtual VTEP address to both of the M-LAG member devices through ECMP routes of the underlay network.

Site-facing link redundancy

IMPORTANT:

This mechanism ensures service continuity in case of site-facing AC failure.

As shown in Figure 19, a VM accesses the EVPN network through multiple Ethernet links that connect to the VTEPs. On each VTEP, all site-facing Ethernet links are assigned to a Layer 2 aggregation group for high availability. On the corresponding Layer 2 aggregate interfaces, Ethernet service instances are configured as ACs of VXLANs to match customer traffic.

Link redundancy mechanism for a direct peer link

If the peer link is an Ethernet aggregate link, VTEPs in the M-LAG system transmit data traffic between them over the peer link or a VXLAN tunnel when a site-facing AC fails.

- **Data traffic transmission over a VXLAN tunnel**—The VTEPs automatically set up a VXLAN tunnel between them and assign it to all VXLANs. When a site-facing AC on one M-LAG member device fails, the device forwards the remote packets destined for the AC to the other M-LAG member device over the VXLAN tunnel. The remote packets are encapsulated with the VXLAN ID of the failed site-facing AC. When the other M-LAG member device receives the packets, it decapsulates them and forwards them in the VXLAN where they belong.
- **Data traffic transmission over the peer link**—Each VTEP in the M-LAG system creates dynamic ACs on the peer-link interface by using one of the following methods:
  - **Creation based on site-facing ACs**—When a site-facing AC is created, a VTEP automatically creates an AC on the peer-link interface. The automatically created AC uses the same traffic match criterion as the site-facing AC and is mapped to the same VSI as the site-facing AC.
  - **Creation based on VXLAN IDs**—When a VXLAN is created, a VTEP automatically creates an AC on the peer-link interface. The automatically created AC uses a frame match criterion generated based on the VXLAN ID and is mapped to the VSI of the VXLAN.

  When a site-facing AC goes down, traffic that a remote device sends to the AC is forwarded to the other M-LAG member device through the peer link. The other M-LAG member device identifies the VSI of the traffic and forwards the traffic to the destination.

Link redundancy mechanism for a tunnel peer link

If a site-facing AC on an M-LAG member device is down, traffic received from a VXLAN tunnel and destined for the AC will be encapsulated into VXLAN packets. The VXLAN ID belongs to the VXLAN that is associated with the VSI of the site-facing AC. The M-LAG member device forwards the VXLAN packets through the tunnel peer link to the peer M-LAG member device. The peer M-LAG member device assigns the traffic to the correct VSI based on the VXLAN ID in the received packets.

Configuring EVPN VXLAN

EVPN VXLAN **tasks at a glance**

To configure EVPN VXLAN, perform the following tasks:

1.
2. Configuring a VXLAN on a VSI
   a. Creating a VXLAN on a VSI
   b. (Optional.) Configuring VSI parameters
3. Configuring an EVPN instance
4. (Optional.) Configuring EVPN multihoming
   a. Assigning an ESI to an interface
   b. (Optional.) Configuring the DF election algorithm
   c. (Optional.) Enabling AC state-based DF election
   d. (Optional.) Setting the DF election delay
   e. (Optional.) Enabling non-revertive mode for preference-based DF election
   f. (Optional.) Configuring ES route advertisement delay
   g. Disabling advertisement of EVPN multihoming routes
5. Configuring BGP to advertise BGP EVPN routes
   a. Enabling BGP to advertise BGP EVPN routes
   b. (Optional.) Configuring BGP EVPN route settings
   c. (Optional.) Enabling the device to ignore default routes in route recursion
   d. (Optional.) Maintaining BGP sessions
6. Mapping ACs to a VSI
7. Configuring an EVPN gateway

   Choose one of the following tasks:
   - Configuring a centralized EVPN gateway
   - Configuring a distributed EVPN gateway
8. (Optional.) Managing remote MAC address entries and remote ARP or ND learning
   - Disabling remote MAC address learning and remote ARP or ND learning
   - Disabling MAC address advertisement
   - Enabling MAC mobility event suppression
   - Disabling learning of MAC addresses from ARP or ND information
   - Disabling ARP information advertisement
   - Enabling ND information advertisement
   - Disabling the VSI interface on a centralized EVPN gateway from learning ARP or ND information across subnets
   - Enabling ARP mobility event suppression
   - Enabling ND mobility event suppression
   - Enabling ARP request proxy
   - Enabling ND request proxy
9. (Optional.) Configuring BGP EVPN route redistribution and advertisement

- Redistributing MAC/IP advertisement routes into BGP unicast routing tables
- Setting the metric of BGP EVPN routes added to a VPN instance's routing table
- Setting the preference of MAC/IP advertisement routes redistributed to the BGP-VPN unicast address families
- Enabling BGP EVPN route advertisement to the local site

10. (Optional.) Maintaining and optimizing an EVPN network
    - Disabling flooding for a VSI
    - Enabling ARP or ND flood suppression
    - Enabling packet statistics for VXLAN tunnels
    - Testing the connectivity of a VXLAN tunnel
    - Enabling SNMP notifications for EVPN
11. (Optional.) Configuring EVPN M-LAG

    Perform this task to virtualize two VTEPs or EVPN gateways into one M-LAG system to avoid single points of failure.
12. (Optional.) Configuring EVPN VXLAN tunnels to learn dynamic entries

Restrictions and guidelines: EVPN VXLAN configuration

Make sure the following VXLAN tunnels are not associated with the same VXLAN when they have the same tunnel destination IP address:

- A VXLAN tunnel automatically created by EVPN.
- A manually created VXLAN tunnel.

For more information about manual tunnel configuration, see *VXLAN Configuration Guide*.

As a best practice to ensure correct traffic forwarding, configure the same MAC address for all VSI interfaces on an EVPN gateway.

Restrictions: System operating mode

To use EVPN VXLAN, first set the system operating mode to expert by using the **`system-working-mode expert`** command in system view, save the running configuration, and then reboot the device. For more information about the system operating mode, see device management in *Fundamentals Configuration Guide*.

Configuring a VXLAN on a VSI

Restrictions and guidelines for VXLAN configuration on a VSI

For more information about the VXLAN commands in this task, see *VXLAN Command Reference*.

Creating a VXLAN on a VSI

1. Enter system view.

   `system-view`
2. Enable L2VPN.

l2vpn enable

By default, L2VPN is disabled.

3. Create a VSI and enter VSI view.

   vsi *vsi-name*

4. Enable the VSI.

   undo shutdown

   By default, a VSI is enabled.

5. Create a VXLAN and enter VXLAN view.

   vxlan *vxlan-id*

   You can create only one VXLAN on a VSI. The VXLAN ID must be unique for each VSI.

Configuring VSI parameters

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Configure a VSI description.

   description *text*

   By default, a VSI does not have a description.

4. Set the MTU for the VSI.

   mtu *size*

   The default MTU is 1500 bytes for a VSI.

5. Enable MAC address learning for the VSI.

   mac-learning enable

   By default, MAC address learning is enabled for a VSI.

Configuring an EVPN instance

About EVPN instance configuration

If a VXLAN requires only Layer 2 connectivity, you do not need to associate a VPN instance with it. The BGP EVPN routes advertised by a VTEP carry the RD and route targets configured for the EVPN instance associated with the VXLAN.

Restrictions and guidelines for EVPN instance configuration

You can bind a VSI only to one EVPN instance.

Configuring an EVPN instance created in VSI view

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Create an EVPN instance and enter EVPN instance view.

   evpn encapsulation vxlan

4. Configure an RD for the EVPN instance.

   route-distinguisher { *route-distinguisher* | **auto** [**router-id**] }

   By default, no RD is configured for an EVPN instance.

5. Configure route targets for the EVPN instance.

   vpn-target { *vpn-target*&<1-8> | **auto** } * [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, an EVPN instance does not have route targets.

   Make sure the following requirements are met:

   - The import targets of the EVPN instance do not match the export targets of the VPN instance associated with the VXLAN or the public instance.
   - The export targets of the EVPN instance do not match the import targets of the VPN instance associated with the VXLAN or the public instance.

   For more information about VPN instance configuration and public instance configuration, see "Configuring an L3 VXLAN ID for a VSI interface."

Configuring EVPN multihoming

Restrictions and guidelines for EVPN multihoming

In a multihomed site, AC configuration and VXLAN IDs must be consistent on redundant VTEPs of the same ES. For each VXLAN ID, you must configure unique RDs for the EVPN instance of VSIs on the redundant VTEPs. You must configure different RDs for the VPN instances and the public instance that use the same VXLAN IP gateway.

IRF and EVPN multihoming are mutually exclusive. Do not configure them together on the device.

You can assign ESIs to a main interface and its subinterfaces.

- For a subinterface, the ESI assigned to the subinterface takes precedence over that assigned to its main interface.
- If you assign an ESI only to a main interface, its subinterfaces inherit the ESI of the main interface.

Assigning an ESI to an interface

About this task

An ESI uniquely identifies an ES. The links on interfaces with the same ESI belong to the same ES. Traffic of the ES can be distributed among the links for load sharing.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
   - Enter Layer 3 Ethernet interface view.

interface *interface-type interface-number*

- Enter Layer 3 aggregate interface view.

  interface route-aggregation *interface-number*

3. Assign an ESI to the interface.

   esi *esi-id*

   By default, no ESI is assigned to an interface.

Configuring the DF election algorithm

About this task

At a multihomed EVPN network site, you can modify the DF election algorithm to control the DF election result.

Restrictions and guidelines

You can configure the DF election algorithm in system view and in interface view. The global DF election algorithm takes effect on all ESs, and the interface-specific DF election algorithm takes effect only on the ESs on an interface. The interface-specific DF election algorithm takes precedence over the global DF election algorithm.

Configuring the DF election algorithm globally

1. Enter system view.

   system-view

2. Configure the DF election algorithm.

   evpn df-election algorithm *algorithm*

   By default, the VLAN tag-based algorithm is used for DF election.

Configuring the DF election algorithm on an interface

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*

   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*

   - Enter Layer 3 Ethernet interface view.

     interface *interface-type interface-number*

   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number*

3. Configure the DF election algorithm.

   evpn df-election algorithm *algorithm*

   By default, the DF election algorithm specified in system view takes effect.

Configuring preference-based DF election

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

interface *interface-type interface-number*

- Enter Layer 2 aggregate interface view.
  interface bridge-aggregation *interface-number*
- Enter Layer 3 Ethernet interface view.
  interface *interface-type interface-number*
- Enter Layer 3 aggregate interface view.
  interface route-aggregation *interface-number*

3. Set the DF election preference.
   evpn df-election preference *preference*
   By default, the DF election preference is 32767.
   The larger the value, the higher the preference.

Enabling AC state-based DF election

About this task

At a multihomed EVPN network site, all VTEPs in an ES participate in DF election by default, regardless of their AC states. If the AC on the VTEP elected as the DF is down, traffic forwarding will fail. You can configure this feature to resolve this issue.

After you configure this feature, a VTEP participates in DF election only when the AC on the VTEP is up. A VTEP determines that the AC on the remote VTEP is up only when it receives AD per ES and AD per EVI routes from the remote VTEP.

Restrictions and guidelines

This feature takes effect only when you enable this feature on all VTEPs in an ES.

Procedure

1. Enter system view.
   system-view
2. Enable AC state-based DF election.
   evpn df-election ac-influence enable
   By default, AC state-based DF election is disabled.

Setting the DF election delay

About this task

The DF election can be triggered by site-facing interface status changes, redundant VTEP membership changes, and interface ESI changes. To prevent frequent DF elections from degrading network performance, set the DF election delay. The DF election delay defines the minimum interval allowed between two DF elections.

Procedure

1. Enter system view.
   system-view
2. Set the DF election delay.
   evpn multihoming timer df-delay *delay-value*
   By default, the DF election delay is 3 seconds.

Enabling non-revertive mode for preference-based DF election

About this task

At a multihomed EVPN network site, configuration changes, device restarts, or process restarts will trigger DF re-election, and DF switchover might cause packet loss. You can configure this feature to resolve this issue.

For example, VTEP 1 and VTEP 2 are in the same ES, and VTEP 1 is the DF. When VTEP 1 restarts and VTEP 2 becomes the new DF, non-revertive mode for preference-based DF election works as follows:

- If VTEP 1 receives Ethernet segment routes from VTEP 2 after the restart, VTEP 1 compares the DF election preference in the Ethernet segment routes with the local DF election preference.
  - If the local DF election preference is smaller than that in the received Ethernet segment routes, VTEP 1 uses the local DF election preference in the advertised Ethernet segment routes.
  - If the local DF election preference is equivalent to that in the received Ethernet segment routes, VTEP 1 compares the IP addresses as follows:
    - If VTEP 1's IP address is lower than VTEP 2's IP address, VTEP 1 uses the local DF election preference and the DP value of 0 in the advertised Ethernet segment routes.
    - If VTEP 1's IP address is higher than VTEP 2's IP address, VTEP 1 uses the local DF election preference and the DP value of 1 in the advertised Ethernet segment routes.
  - If the local DF election preference is larger than that in the received Ethernet segment routes, VTEP 1 uses the DF election preference in the received Ethernet segment routes and the DP value of 0 when advertising Ethernet segment routes.
- If VTEP 1 does not receive Ethernet segment routes from VTEP 2 after the restart, VTEP 1 uses the local DF election preference in the advertised Ethernet segment routes.

Restrictions and guidelines

As a best practice, configure this feature when the all-active redundancy mode is used, and do not configure this feature when the single-active redundancy mode is used.

When the DF restarts or the process restarts and a new DF is elected, the original DF will immediately advertise Ethernet segment routes after the restart to become the DF again. However, access-side connections restore slowly and packet loss might occur after DF switchover. To resolve this issue, configure this feature together with the Ethernet segment route advertisement delay feature on all VTEPs in an ES.

Procedure

1. Enter system view.

   system-view
2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
   - Enter Layer 3 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number*

3. Enable non-revertive mode for preference-based DF election.

   evpn df-election preference non-revertive

   By default, non-revertive mode is disabled for preference-based DF election.

Configuring ES route advertisement delay

About this task

When the DF restarts or the process restarts and a new DF is elected, the original DF will immediately advertise ES routes after the restart to become the DF again. However, access-side connections restore slowly and packet loss might occur after DF switchover. To resolve this issue, perform this task to configure a VTEP to delay the advertisement of ES routes.

Restrictions and guidelines

As a best practice, configure this feature when the all-active redundancy mode is used, and do not configure this feature when the single-active redundancy mode is used.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*
   - Enter Layer 3 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 3 aggregate interface view.

     interface route-aggregation *interface-number*
3. Configure ES route advertisement delay when preference-based DF election is used.

   evpn timer es-recovery *recovery-time*

   By default, ES route advertisement delay is disabled when preference-based DF election is used.

Disabling advertisement of EVPN multihoming routes

About this task

EVPN multihoming routes include Ethernet auto-discovery routes and Ethernet segment routes.

In a multihomed EVPN network, perform this task on a redundant VTEP before you reboot it. This operation allows other VTEPs to refresh their EVPN routing table to prevent traffic interruption caused by the reboot.

Restrictions and guidelines

Do not perform this task on VTEPs at a multihomed EVPN VXLAN site if EVPN forwards multicast traffic based on SMET, IGMP join sync, and IGMP leave sync routes. Violation of this restriction might cause multicast forwarding errors.

Procedure

1. Enter system view.

   system-view

2. Disable advertisement of EVPN multihoming routes and withdraw the EVPN multihoming routes that have been advertised to remote sites.

   evpn multihoming advertise disable

   By default, the device advertises EVPN multihoming routes.

Configuring BGP to advertise BGP EVPN routes

Restrictions and guidelines for BGP EVPN route advertisement

For more information about BGP commands in this task, see *Layer 3—IP Routing Command Reference*.

Enabling BGP to advertise BGP EVPN routes

1. Enter system view.

   system-view

2. Configure a global router ID.

   router id *router-id*

   By default, no global router ID is configured.

3. Enable a BGP instance and enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

   By default, BGP is disabled and no BGP instances exist.

4. Specify remote VTEPs as BGP peers.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **as-number** *as-number*

5. Create the BGP EVPN address family and enter BGP EVPN address family view.

   address-family l2vpn evpn

6. Enable BGP to exchange BGP EVPN routes with a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **enable**

   By default, BGP does not exchange BGP EVPN routes with peers.

Configuring BGP EVPN route settings

Configuring BGP EVPN to advertise default routes

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP EVPN address family view.

   address-family l2vpn evpn

4. Advertise a default route to a peer or peer group.

peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **default-route-advertise** { **ipv4** | **ipv6** } **vpn-instance** *vpn-instance-name*

By default, no default route is advertised to any peers or peer groups.

Configuring attributes of BGP EVPN routes

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP EVPN address family view.

   address-family l2vpn evpn

4. Permit the local AS number to appear in routes from a peer or peer group and set the number of appearances.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **allow-as-loop** [*number*]

   By default, the local AS number is not allowed in routes from peers.

5. Advertise the COMMUNITY attribute to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise-community**

   By default, the device does not advertise the COMMUNITY attribute to peers or peer groups.

6. Remove the default-gateway extended community attribute from the EVPN gateway routes advertised to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **default-gateway no-advertise**

   By default, EVPN gateway routes advertised to peers and peer groups contain the default-gateway extended community attribute.

7. Configure the device to advertise BGP routes with the lowest priority to peers. Choose one of the following options:

   - Configure the device to advertise BGP routes with the lowest priority to a peer within a specified period of time after the peer goes down.

     advertise lowest-priority on-peer-up duration *seconds*

   - Configure the device to advertise BGP routes with the lowest priority to peers within a specified period of time after the BGP process restarts because of a device reboot.

     advertise lowest-priority on-startup duration *seconds*

   By default, the device does not change the priority of the BGP routes advertised to peers.

   To set the priority of BGP routes to the lowest, the device sets the local preference and MED value of the BGP routes to 0 and 4294967295, respectively. To restore the original priority of the BGP routes during the specified duration, execute the **reset bgp advertise lowest-priority** command in user view.

Configuring optimal BGP EVPN route selection

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP EVPN address family view.

   address-family l2vpn evpn

4. Configure BGP to prefer routes with an IPv6 next hop during optimal route selection.
   bestroute ipv6-nexthop
   By default, BGP prefer routes with an IPv4 next hop during optimal route selection.
5. (Optional.) Set the optimal route selection delay timer. Choose one of the following tasks or both:
   - Set the optimal route selection delay timer for all BGP routes in the current address family.
     route-select delay *delay-value*
   - Delay optimal route selection for a peer in the specified duration after the peer comes up.
     route-select suppress on-peer-up *milliseconds*

   By default, optimal route selection is not delayed.

Configuring BGP route reflection

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP EVPN address family view.
   address-family l2vpn evpn
4. Configure the device as an RR and specify a peer or peer group as its client.
   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **reflect-client**
   By default, no RR or client is configured.
5. (Optional.) Enable BGP EVPN route reflection between clients.
   reflect between-clients
   By default, BGP EVPN route reflection between clients is enabled.
6. (Optional.) Configure the cluster ID of the RR.
   reflector cluster-id { *cluster-id* | *ipv4-address* }
   By default, an RR uses its own router ID as the cluster ID.
7. (Optional.) Create a reflection policy for the RR to filter reflected BGP EVPN routes.
   rr-filter *ext-comm-list-number*
   By default, an RR does not filter reflected BGP EVPN routes.
8. (Optional.) Enable the RR to change the attributes of routes to be reflected.
   reflect change-path-attribute
   By default, an RR cannot change the attributes of routes to be reflected.

Configuring the route server feature

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP EVPN address family view.
   address-family l2vpn evpn
4. Configure the device as a route server and specify a peer or peer group as a client.
   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-server-client** [**external**]
   By default, neither the route server nor the client is configured.

The packets advertised to the clients does not carry the AS number of the route server. To avoid peer establishment failures, execute the **peer ignore-first-as** command on the clients.

Filtering BGP EVPN routes

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP EVPN address family view.

   address-family l2vpn evpn

4. Apply a routing policy to routes received from or advertised to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **route-policy** *route-policy-name* { **export** | **import** }

   By default, no routing policies are applied to routes received from or advertised to peers or peer groups.

5. Configure Layer 2 ACL-based route filtering.

   - Filter routes for a peer or peer group by using a Layer 2 ACL.

     peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **filter-policy** { *mac-acl-number* | **name** *mac-acl-name* } { **export** | **import** }

   - Filter all advertised routes.

     filter-policy { *mac-acl-number* | **name** *mac-acl-name* } **export**

   - Filter all received routes.

     filter-policy { *mac-acl-number* | **name** *mac-acl-name* } **import**

   By default, Layer 2 ACL-based route filtering is not configured for a peer or peer group.

   In a Layer 2 ACL, only the **rule** [*rule-id*] { **deny** | **permit** } **dest-mac** *dest-address dest-mask* rule is used to filter MAC/IP advertisement routes that carry the specified MAC addresses. The other rules in a Layer 2 ACL do not take effect in routing filtering.

6. Enable route target filtering for BGP EVPN routes.

   policy vpn-target

   By default, route target filtering is enabled for BGP EVPN routes.

7. Enable the device to filter advertised objects based on the first AS number in EBGP routes.

   peer-as-check enable

   By default, the device advertises a received EBGP route to all BGP peers except the peer that sends the EBGP route.

   After you execute this command, the device checks the first AS number in the AS_Path attribute of an EBGP route when advertising the EBGP route to EBGP peers. The device will not advertise the EBGP route to the EBGP peers in that AS.

Configuring the BGP Additional Paths feature

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP EVPN address family view.

   address-family l2vpn evpn

4. Configure the BGP Additional Paths capabilities.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **additional-paths** { **receive** | **send** } *

   By default, no BGP Additional Paths capabilities are configured.

5. Set the maximum number of Add-Path optimal routes that can be advertised to a peer or peer group.

   peer { *group-name* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] } **advertise additional-paths best** *number*

   By default, a maximum of one Add-Path optimal route can be advertised to a peer or peer group.

6. Set the maximum number of Add-Path optimal routes that can be advertised to all peers.

   additional-paths select-best *best-number*

   By default, a maximum of one Add-Path optimal route can be advertised to all peers.

Enabling the device to ignore default routes in route recursion

About this task

By default, the device selects a default route to forward traffic if only the default route is obtained after BGP route recursion. If the default route does not point to the desired next hop, traffic forwarding will fail.

To resolve this issue, enable the device to ignore default routes in route recursion. If only the default route is obtained after route recursion is performed for a BGP route, that BGP route becomes invalid, and other BGP routes with the same prefix are selected for forwarding.

Enable this feature if multiple links exist between the device and a destination IP address. If one of the links fail, traffic will be switched to the other available links instead of being incorrectly forwarded based on a default route.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP EVPN address family view.

   address-family l2vpn evpn

4. Enable the device to ignore default routes in route recursion.

   nexthop recursive-lookup default-route ignore [**route-policy** *route-policy-name*]

   By default, the device can select a default route for forwarding after performing route recursion.

CAUTION:

After you execute this command, VXLAN tunnels might be reestablished, and transient VXLAN traffic loss might occur. As a best practice, enable BGP EVPN route reception and advertisement again after you execute this command.

Maintaining BGP sessions

Perform the following tasks in user view:

- Reset BGP sessions of the BGP EVPN address family.

  reset bgp [**instance** *instance-name*] { *as-number* | *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } **l2vpn evpn**

- Soft-reset BGP sessions of the BGP EVPN address family.

  refresh bgp [**instance** *instance-name*] { *ipv4-address* [*mask-length*] | *ipv6-address* [*prefix-length*] | **all** | **external** | **group** *group-name* | **internal** } { **export** | **import** } **l2vpn evpn**

Mapping ACs to a VSI

Mapping a static Ethernet service instance to a VSI

About this task

A static Ethernet service instance matches a list of VLANs on a site-facing interface by using a frame match criterion. The VTEP assigns traffic from the VLANs to a VXLAN by mapping the Ethernet service instance to a VSI. The VSI performs Layer 2 forwarding for the VLANs based on its MAC address table.

For more information about the VXLAN commands in this task, see *VXLAN Command Reference*.

Restrictions and guidelines

Link aggregation group membership is mutually exclusive with Ethernet service instance-to-VSI mappings on a Layer 2 interface. Do not map a VSI to an Ethernet service instance on a Layer 2 interface if the interface is in a Layer 2 aggregation group.

If an Ethernet service instance matches both inner and outer VLAN IDs on an interface, do not configure any other Ethernet service instances to match the same outer VLAN ID on that interface.

For information about the frame match criterion configuration restrictions and guidelines of Ethernet service instances, see *VXLAN Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Enter interface view.
   - Enter Layer 2 Ethernet interface view.

     interface *interface-type interface-number*
   - Enter Layer 2 aggregate interface view.

     interface bridge-aggregation *interface-number*

3. Create an Ethernet service instance and enter Ethernet service instance view.

   service-instance *instance-id*

4. Choose one option to configure a frame match criterion.
   - Match frames with the specified inner VLAN tags.

     encapsulation c-vid { *vlan-id* | *vlan-id-list* }
   - Match frames with the specified outer VLAN tags.

     encapsulation s-vid { *vlan-id* | *vlan-id-list* } [**only-tagged**]
   - Match frames with the specified inner and outer VLAN tags.

     encapsulation s-vid { *vlan-id* | *vlan-id-list* } **c-vid** { *vlan-id-list* | **all** }

- Match any VLAN tagged or untagged frames.

  encapsulation { **tagged** | **untagged** }
- Match frames that do not match any other service instance on the interface.

  encapsulation default

  An interface can contain only one Ethernet service instance that uses the **encapsulation default** criterion.

  An Ethernet service instance that uses the **encapsulation default** criterion matches any frames if it is the only instance on the interface.

By default, an Ethernet service instance does not contain a frame match criterion.

5. Map the Ethernet service instance to a VSI.

   xconnect vsi *vsi-name* [**access-mode** { **ethernet** | **vlan** }] [**track** *track-entry-number*&<1-3>]

   By default, an Ethernet service instance is not mapped to any VSI.

Configuring a centralized EVPN gateway

Restrictions and guidelines for centralized EVPN gateway configuration

If an EVPN network contains a centralized EVPN gateway, you must enable ARP or ND flood suppression on VTEPs. Typically, remote ARP or ND learning is disabled in an EVPN network. When ARP or ND requests for the gateway MAC address are sent to the centralized EVPN gateway through VXLAN tunnels, the gateway does not learn ARP or ND entries, which might cause traffic blockage on VXLAN tunnels.

Configuring a centralized gateway interface

1. Enter system view.

   system-view
2. Create a VSI interface and enter VSI interface view.

   interface vsi-interface *vsi-interface-id*

   For more information about this command, see *VXLAN Command Reference*.
3. Assign an IP address to the VSI interface.

   IPv4:

   ip address *ip-address* { *mask* | *mask-length* } [**sub**]

   IPv6:

   See IPv6 basics in *Layer 3—IP Services Configuration Guide*.

   By default, no IP address is assigned to a VSI interface.
4. Return to system view.

   quit
5. Enter VSI view.

   vsi *vsi-name*
6. Specify the VSI interface as the gateway interface for the VSI.

   gateway vsi-interface *vsi-interface-id*

By default, no gateway interface is specified for a VSI.

For more information about this command, see *VXLAN Command Reference*.

Setting the static flag for the MAC addresses of centralized gateway interfaces

About this task

In a network with a centralized EVPN gateway deployed, a VTEP considers a MAC address move occurs if an endpoint uses a MAC address identical to that of a centralized gateway interface. As a result, the VTEP overwrites the MAC address entry created for the centralized gateway interface with that created for the endpoint, and errors will occur in traffic forwarding.

To resolve this issue, set the static flag for the MAC addresses of centralized gateway interfaces on the centralized EVPN gateway. When advertising those MAC addresses through MAC/IP advertisement routes, the centralized EVPN gateway will set the static flag bit to 1 in the MAC mobility extended community. If an endpoint accesses the network with a MAC address identical to that of a centralized gateway interface, the endpoint's MAC address entry will not overwrite the entry for the centralized gateway interface.

Procedure

1. Enter system view.

   `system-view`
2. Set the static flag for the MAC addresses of centralized gateway interfaces.

   `evpn route gateway-mac unmovable`

   By default, the static flag is not set for the MAC addresses of centralized gateway interfaces.

Configuring a distributed EVPN gateway

Restrictions and guidelines for distributed EVPN gateway configuration

When you configure VSI interfaces on a distributed EVPN gateway, follow these restrictions and guidelines:

- Do not assign reserved MAC addresses to VSI interfaces.
- You must assign the same MAC address to the VSI interfaces with L3 VXLAN IDs associated.
- On distributed EVPN gateways, you must assign the same MAC address and IP address to the VSI interfaces that act as the gateway for the same VXLAN.
- If a distributed EVPN gateway is connected to an IPv6 site, make sure the VSI interfaces with L3 VXLAN IDs associated use the same link-local address.

As a best practice, do not use ARP flood suppression and local proxy ARP or ND flood suppression and local ND proxy together on distributed EVPN gateways. If both ARP flood suppression and local proxy ARP are enabled on a distributed EVPN gateway, only local proxy ARP takes effect. If both ND flood suppression and local ND proxy are enabled on a distributed EVPN gateway, only local ND proxy takes effect.

On a distributed EVPN gateway, make sure the VSI interfaces configured with L3 VXLAN IDs use the same MAC address. To modify the MAC address of a VSI interface, use the **`mac-address`** command.

Prerequisites for distributed EVPN gateway configuration

For a VXLAN to access the external network, specify the VXLAN's VSI interface on the border gateway as the next hop on distributed EVPN gateways by using one of the following methods:

- Configure a static route.
- Configure a routing policy, and apply the policy by using the **apply default-next-hop** or **apply next-hop** command. For more information about configuring routing policies, see routing policy configuration or IPv6 routing policy configuration in *Layer 3—IP Routing Configuration Guide*.

Configuring the traffic forwarding mode for EVPN VXLAN

Restrictions and guidelines

The asymmetric IRB mode is supported only on distributed EVPN gateways. The mode takes effect only on Layer 3 traffic forwarded in the same VXLAN. In addition, the same VSI interface on different distributed EVPN gateways must have different IP addresses.

Procedure

1. Enter system view.

   system-view
2. Configure the traffic forwarding mode for EVPN VXLAN. Choose one of the following options:
   - Enable asymmetric IRB mode.

     evpn irb asymmetric [**route-policy** *route-policy-name*]
   - Enable symmetric IRB mode.

     undo evpn irb asymmetric

     By default, a distributed EVPN gateway forwards EVPN VXLAN traffic in symmetric IRB mode.

Configuring a VSI interface

About this task

To save Layer 3 interface resources on a distributed EVPN gateway, multiple VSIs can share one VSI interface. You can assign multiple IP addresses to the VSI interface for the VSIs to use as gateway addresses.

When VSIs share a VSI interface, you must specify the subnet of each VSI for the VSI interface to identify the VSI of a packet. The subnets must be unique.

Procedure

1. Enter system view.

   system-view
2. Create a VSI interface and enter VSI interface view.

   interface vsi-interface *vsi-interface-id*

   For more information about this command, see *VXLAN Command Reference*.
3. Assign an IP address to the VSI interface.

   IPv4:

   ip address *ip-address* { *mask* | *mask-length* } [**sub**]

   IPv6:

See IPv6 basics in *Layer 3—IP Services Configuration Guide*.

By default, no IP address is assigned to a VSI interface.

4. Assign a MAC address to the VSI interface.

   mac-address *mac-address*

   The default MAC address of VSI interfaces is the same as the default MAC address of Layer 3 Ethernet interfaces.

   To ensure correct forwarding after VM migration, you must assign the same MAC address to the VSI interfaces of a VXLAN on all distributed gateways.

5. Specify the VSI interface as a distributed gateway.

   distributed-gateway local

   By default, a VSI interface is not a distributed gateway.

   For more information about this command, see *VXLAN Command Reference*.

6. (Optional.) Enable local proxy ARP or local ND proxy.

   IPv4:

   local-proxy-arp enable [**ip-range** *startIP* **to** *endIP*]

   By default, local proxy ARP is disabled.

   For more information about the command, see proxy ARP commands in *Layer 3—IP Services Command Reference*.

   IPv6:

   local-proxy-nd enable

   By default, local ND proxy is disabled.

   For more information about the commands, see IPv6 basic commands *Layer 3—IP Services Command Reference*.

7. Return to system view.

   quit

8. Enter VSI view.

   vsi *vsi-name*

9. Specify the VSI interface as the gateway interface for the VSI.

   gateway vsi-interface *vsi-interface-id*

   By default, no gateway interface is specified for a VSI.

   For more information about this command, see *VXLAN Command Reference*.

10. Assign a subnet to the VSI.

    gateway subnet { *ipv4-address wildcard-mask* | *ipv6-address prefix-length* }

    By default, no subnet exists on a VSI.

    For more information about this command, see *VXLAN Command Reference*.

Configuring an L3 VXLAN ID for a VSI interface

Configuring an L3 VXLAN ID for the VSI interface of a VPN instance

1. Enter system view.

   system-view

2. Configure a VPN instance:

   a. Create a VPN instance and enter VPN instance view.

      ip vpn-instance *vpn-instance-name*

b. Configure an RD for the VPN instance.

route-distinguisher *route-distinguisher*

By default, no RD is configured for a VPN instance.

c. Configure route targets for the VPN instance.

vpn-target { *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

By default, a VPN instance does not have route targets.

d. (Optional.) Apply an export routing policy to the VPN instance.

export route-policy *route-policy*

By default, no export routing policy is applied to a VPN instance.

The specified export routing policy is used to filter the BGP EVPN routes advertised for the VPN instance.

e. (Optional.) Apply an import routing policy to the VPN instance.

import route-policy *route-policy*

By default, no import routing policy is applied to a VPN instance. The VPN instance accepts a route when the export route targets of the route match local import route targets.

The specified import routing policy is used to filter the BGP EVPN routes redistributed to the VPN instance.

3. Configure EVPN on the VPN instance:

a. Enter VPN instance EVPN view.

address-family evpn

b. Configure route targets for EVPN on the VPN instance.

vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

By default, EVPN does not have route targets on a VPN instance.

Make sure the following requirements are met:

- The import targets of EVPN do not match the export targets of the VPN instance.
- The export targets of EVPN do not match the import targets of the VPN instance.

c. (Optional.) Apply an export routing policy to EVPN on the VPN instance.

export route-policy *route-policy*

By default, no export routing policy is applied to EVPN on a VPN instance.

d. (Optional.) Apply an import routing policy to EVPN on the VPN instance.

import route-policy *route-policy*

By default, no import routing policy is applied to EVPN on a VPN instance. The VPN instance accepts a route when the route targets of the route match local import route targets.

4. Execute the following commands in sequence to return to system view.

a. **quit**

b. **quit**

5. Create a VSI interface and enter VSI interface view.

interface vsi-interface *vsi-interface-id*

6. Associate the VSI interface with the VPN instance.

ip binding vpn-instance *vpn-instance-name*

By default, a VSI interface is not associated with a VPN instance. The interface is on the public network.

7. Configure an L3 VXLAN ID for the VSI interface.

   l3-vni *vxlan-id* [**vsi** *vsi-name*]

   By default, no L3 VXLAN ID is configured for a VSI interface.

   A VPN instance can have only one L3 VXLAN ID. If multiple L3 VXLAN IDs are configured for a VPN instance, the VPN instance uses the lowest one. To view the L3 VXLAN ID of a VPN instance, use the **display evpn routing-table** command.

Configuring an L3 VXLAN ID for the VSI interface of the public instance

1. Enter system view.

   system-view

2. Create the public instance and enter its view.

   ip public-instance

3. Configure an RD for the public instance.

   route-distinguisher *route-distinguisher*

   By default, no RD is configured for the public instance.

4. Configure an L3 VXLAN ID for the public instance.

   l3-vni *vxlan-id*

   By default, the public instance does not have an L3 VXLAN ID.

   The public instance can have only one L3 VXLAN ID. To modify the L3 VXLAN ID for the public instance, you must first delete the original L3 VXLAN ID.

5. (Optional.) Configure route targets for the public instance.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, the public instance does not have route targets.

6. Enter IPv4 address family view, IPv6 address family view, or EVPN view.
   - Enter IPv4 address family view.

     address-family ipv4
   - Enter IPv6 address family view.

     address-family ipv6
   - Enter EVPN view.

     address-family evpn

7. Configure route targets for the IPv4 address family, IPv6 address family, or EVPN.

   vpn-target *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   By default, the IPv4 address family, IPv6 address family, and EVPN do not have route targets on the public instance.

   Make sure the following requirements are met:
   - The import targets of an EVPN instance do not match the export targets of the public instance.
   - The export targets of an EVPN instance do not match the import targets of the public instance.

8. Execute the following commands in sequence to return to system view.

   a. **quit**

   b. **quit**

9. Enter VSI interface view.

   interface vsi-interface *vsi-interface-id*

10. Configure an L3 VXLAN ID for the VSI interface.

    l3-vni *vxlan-id*

    By default, no L3 VXLAN ID is configured for a VSI interface.

    Of the VSI interfaces associated with the public instance, a minimum of one VSI interface must use the same L3 VXLAN ID as the public instance.

Configuring IP prefix route advertisement

About this task

If IGP routes are imported to the BGP-VPN IPv4 or IPv6 unicast address family and the corresponding VPN instance has an L3 VXLAN ID, the device advertises the imported routes as IP prefix advertisement routes.

If IGP routes are imported to the BGP IPv4 or IPv6 unicast address family and the public instance has an L3 VXLAN ID, the device advertises the imported routes as IP prefix advertisement routes.

A VTEP compares the export route targets of received IP prefix advertisement routes with the import route targets configured for the IPv4 address family or IPv6 address family on a VPN instance or the public instance. If the route targets match, the VTEP accepts the routes and adds the routes to the routing table of the VPN instance or public instance.

Restrictions and guidelines

This feature is supported only by distributed EVPN gateway deployment.

For more information about the BGP commands in this task, see *Layer 3—IP Routing Command Reference*.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP address family view.
   - Enter BGP IPv4 unicast address family view.

     address-family ipv4 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv4 unicast address family view.

     ip vpn-instance *vpn-instance-name*

     address-family ipv4 [**unicast**]
   - Enter BGP IPv6 unicast address family view.

     address-family ipv6 [**unicast**]
   - Execute the following commands in sequence to enter BGP-VPN IPv6 unicast address family view.

     ip vpn-instance *vpn-instance-name*

     address-family ipv6 [**unicast**]

4. Enable BGP to redistribute routes from an IGP protocol.

   import-route *protocol* [{ *process-id* | **all-processes** } [**allow-direct** | **med** *med-value* | **route-policy** *route-policy-name*] *]

   By default, BGP does not redistribute IGP routes.

5. (Optional.) Enable default route redistribution into the BGP routing table.

default-route imported

By default, default route redistribution into the BGP routing table is disabled.

6. (Optional.) Configure ECMP VPN route redistribution:
   a. Return to BGP instance view.
      quit
   b. Enter BGP EVPN address family view.
      address-family l2vpn evpn
   c. Enable ECMP VPN route redistribution.
      vpn-route cross multipath

      By default, ECMP VPN route redistribution is disabled. If multiple routes have the same prefix and RD, BGP only imports the optimal route into the EVPN routing table.

      ECMP VPN route redistribution enables BGP to import all routes that have the same prefix and RD into the EVPN routing table.

Configuring the EVPN global MAC address

About this task

The EVPN global MAC address is used only by VSI interfaces associated with an L3 VXLAN ID. For such a VSI interface, the MAC address assigned to it by using the **mac-address** command takes precedence over the EVPN global MAC address.

A distributed EVPN gateway selects the lowest-numbered VSI interface that is associated with an L3 VXLAN ID as its router MAC address. In an M-LAG system, distributed EVPN gateways that act as M-LAG member devices might use different router MAC addresses, which causes forwarding errors. To resolve this problem, you can configure the same EVPN global MAC address on the gateways.

Restrictions and guidelines

Do not use a reserved MAC address as the EVPN global MAC address.

As a best practice, configure the EVPN global MAC address in either of the following methods:

- Use the default MAC address of a Layer 3 Ethernet interface (an Ethernet interface in route mode) on a member device as the EVPN global MAC for all member devices in the DR system.
- Specify a unicast MAC address that meets the requirements as the EVPN global MAC for all member devices in the DR system. As a best practice, specify an unused MAC address in the range of 0001-0001-0001 to 0001-0001-FFFE.

Procedure

1. Enter system view.
   system-view
2. Configure the EVPN global MAC address.
   evpn global-mac *mac-address*

   By default, no EVPN global MAC address is configured.

Disabling generation of IP prefix advertisement routes for the subnets of a VSI interface

About this task

A distributed VXLAN IP gateway by default generates IP prefix advertisement routes for the subnets of VSI interfaces and advertises these routes to remote VTEPs. The remote VTEPs advertise these

routes to their local sites. To disable advertisement of these routes to remote sites, you can disable generation of IP prefix advertisement routes for the subnets of VSI interfaces.

Restrictions and guidelines

This feature takes effect only on a VSI interface that provides distributed VXLAN IP gateway service (configured by using the **distributed-gateway local** command). It does not take effect on VSI interfaces that provide centralized VXLAN IP gateway service. The device only generates MAC/IP advertisement routes for VSI interfaces that provide centralized VXLAN IP gateway service.

Procedure

1. Enter system view.
   system-view
2. Enter VSI interface view.
   interface vsi-interface *vsi-interface-id*
3. Disable generation of IP prefix advertisement routes for the subnets of the VSI interface.
   ip-prefix-route generate disable
   By default, the device generates IP prefix advertisement routes for the subnets of a VSI interface that provides distributed VXLAN IP gateway service.

Enabling a distributed EVPN gateway to send RA messages over VXLAN tunnels

About this task

By default, a distributed EVPN gateway drops the RS messages received from VXLAN tunnels and periodically advertises RA messages only to the local site. As a result, a distributed EVPN gateway does not send RA messages over VXLAN tunnels, and remote gateways cannot update information about the gateway based on RA messages. To resolve the issue, perform this task to enable distributed EVPN gateways to reply to remote RS messages with RA messages and periodically advertise RA messages over VXLAN tunnels.

Restrictions and guidelines

You can configure RA message tunneling for VSI interfaces globally or on a per-VSI interface basis. The global configuration takes effect on all VSI interfaces. The interface-specific configuration takes precedence over the global configuration on a VSI interface.

Globally enabling VSI interfaces to send RA messages over VXLAN tunnels

1. Enter system view.
   system-view
2. Globally enable VSI interfaces to send RA messages over VXLAN tunnels.
   ipv6 nd ra tunnel-broadcast global enable
   By default, VSI interfaces do not send RA messages over VXLAN tunnels.

Enabling a VSI interface to send RA messages over VXLAN tunnels

1. Enter system view.
   system-view
2. Enter VSI interface view.
   interface vsi-interface *vsi-interface-id*
3. Enable the VSI interface to send RA messages over VXLAN tunnels.
   ipv6 nd ra tunnel-broadcast enable
   By default, a VSI interface uses the global RA message tunneling configuration.

Enabling traffic statistics for the VSIs automatically created for L3 VXLAN IDs

About this task

If you configure an L3 VXLAN ID on a distributed EVPN gateway, the gateway automatically creates a VSI for the L3 VXLAN ID. You cannot enter the view of such a VSI to configure settings on it.

This task enables the device to collect incoming and outgoing traffic statistics for the automatically created VSIs. You can use the **`display l2vpn vsi verbose`** command to view the traffic statistics and use the **`reset l2vpn statistics vsi`** command to clear the traffic statistics.

Procedure

1. Enter system view.

   `system-view`

2. Enable traffic statistics for the VSIs that are automatically created for L3 VXLAN IDs and all VXLAN tunnels of VSIs.

   `l2vpn statistics vsi l3-vni`

   By default, the traffic statistics feature is disabled for the VSIs that are automatically created for L3 VXLAN IDs and all VXLAN tunnels of VSIs.

Enabling the device to advertise ARP information for the distributed EVPN gateway interfaces through MAC/IP advertisement routes

About this task

If a distributed EVPN gateway has downstream VTEPs attached, the gateway advertises ARP information for gateway interfaces through IP prefix advertisement routes. Because the VTEPs do not have gateway configuration, they cannot learn the ARP information for the gateway interfaces or forward traffic to the gateway. For the VTEPs to learn ARP information for the gateway interfaces, enable the distributed EVPN gateway to advertise ARP information for the gateway interfaces through MAC/IP advertisement routes.

Procedure

1. Enter system view.

   `system-view`

2. Enable the device to advertise ARP information for the distributed EVPN gateway interfaces through MAC/IP advertisement routes.

   `evpn mac-ip advertise distributed-gateway`

   By default, the device does not advertise ARP information for the distributed EVPN gateway interfaces through MAC/IP advertisement routes.

Managing remote MAC address entries and remote ARP or ND learning

Disabling remote MAC address learning and remote ARP or ND learning

About this task

By default, the device learns MAC information, ARP information, and ND information of remote user terminals from packets received on VXLAN tunnel interfaces. The automatically learned remote MAC, ARP, and ND information might conflict with the remote MAC, ARP, and ND information advertised through BGP. As a best practice to avoid the conflicts, disable remote MAC address learning and remote ARP or ND learning on the device.

For more information about the VXLAN commands in this task, see *VXLAN Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Disable remote MAC address learning.
   vxlan tunnel mac-learning disable
   By default, remote MAC address learning is enabled.
3. Disable remote ARP learning.
   vxlan tunnel arp-learning disable
   By default, remote ARP learning is enabled.
4. Disable remote ND learning.
   vxlan tunnel nd-learning disable
   By default, remote ND learning is enabled.

Disabling MAC address advertisement

About this task

The MAC information and ARP or ND information advertised by the VTEP overlap. To avoid duplication, disable MAC address advertisement and withdraw the MAC addresses advertised to remote VTEPs.

Procedure (VSI EVPN instance view)

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Enter VSI EVPN instance view.
   evpn encapsulation vxlan
4. Disable MAC address advertisement and withdraw advertised MAC addresses.
   mac-advertising disable
   By default, MAC address advertisement is enabled.

Enabling MAC mobility event suppression

About this task

On an EVPN VXLAN network, misconfiguration of MAC addresses might cause two sites to contain the same MAC address. In this condition, VTEPs at the two sites constantly synchronize and update EVPN MAC entries and determine that MAC mobility events occur. As a result, an inter-site loop might occur, and the bandwidth is occupied by MAC entry synchronization traffic. To eliminate loops and suppress those MAC mobility events, enable MAC mobility event suppression on the VTEPs. This feature allows a MAC address to move a specified number of times (the MAC mobility suppression threshold) from a site within a MAC mobility detection cycle. If a MAC address moves more than the MAC mobility suppression threshold, the VTEP at the site will suppress the last MAC move to the local site and will not advertise information about the MAC address.

Restrictions and guidelines

After you execute the **undo evpn route mac-mobility suppression** command or the suppression time expires, a VTEP acts as follows:

- Advertises MAC address entries immediately for the suppressed MAC address entries that have not aged out.
- Relearns the MAC addresses for the suppressed MAC address entries that have aged out and advertises the MAC address entries.

If both MAC address entry conflicts and ARP entry conflicts exist for a MAC address, you must enable both MAC mobility event suppression and ARP mobility event suppression. If you enable only MAC mobility event suppression, the system cannot suppress MAC mobility events for the MAC address.

Procedure

1. Enter system view.

   system-view

2. Enable MAC mobility event suppression.

   evpn route mac-mobility suppression [**detect-cycle** *detect-time* | **detect-threshold** *move-times* | **suppression-time** [*suppression-time* | **permanent**]] *

   By default, MAC mobility event suppression is disabled.

Disabling learning of MAC addresses from ARP or ND information

About this task

The MAC information and ARP or ND information advertised by a remote VTEP overlap. To avoid duplication, disable the learning of MAC addresses from ARP or ND information. EVPN will learn remote MAC addresses only from the MAC information advertised from remote sites.

Procedure (VSI EVPN instance view)

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Enter VSI EVPN instance view.

   evpn encapsulation vxlan

4. Disable the EVPN instance from learning MAC addresses from ARP information.

arp mac-learning disable

By default, an EVPN instance learns MAC addresses from ARP information.

5. Disable the EVPN instance from learning MAC addresses from ND information.

   nd mac-learning disable

   By default, an EVPN instance learns MAC addresses from ND information.

Disabling ARP information advertisement

About this task

In an EVPN network with distributed gateways, you can disable ARP information advertisement for a VXLAN to save resources if all its user terminals use the same EVPN gateway device. The EVPN instance of the VXLAN will stop advertising ARP information through MAC/IP advertisement routes and withdraw advertised ARP information. When ARP information advertisement is disabled, user terminals in other VXLANs still can communicate with that VXLAN through IP prefix advertisement routes.

Procedure (VSI EVPN instance view)

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Enter VSI EVPN instance view.

   evpn encapsulation vxlan

4. Disable ARP information advertisement for the EVPN instance.

   arp-advertising disable

   By default, ARP information advertisement is enabled for an EVPN instance.

Enabling ND information advertisement

About this task

In an EVPN network with distributed gateways, you can disable ND information advertisement for a VXLAN to save resources if all its user terminals use the same EVPN gateway device. The EVPN instance of the VXLAN will stop advertising ND information through MAC/IP advertisement routes and withdraw advertised ND information. When ND information advertisement is disabled, user terminals in other VXLANs still can communicate with that VXLAN through IP prefix advertisement routes.

Procedure (VSI EVPN instance view)

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Enter VSI EVPN instance view.

   evpn encapsulation vxlan

4. Enable ND information advertisement for the EVPN instance.

   nd-advertising disable

   By default, ND information advertisement is enabled for an EVPN instance.

Disabling the VSI interface on a centralized EVPN gateway from learning ARP or ND information across subnets

About this task

On an EVPN VXLAN network deployed with a centralized EVPN gateway, VM 1 and VM 2 belong to the same VXLAN in subnet 10.1.1.0/24. The gateway interface is VSI-interface 1 and the gateway is connected to external Layer 3 network 10.1.2.0/24. The VTEP to which VM 2 is attached is configured with ARP or ND flood suppression. The IP address of VM 2 is mistakenly configured as an IP address in subnet 10.1.2.0/24 (for example, 10.1.2.2). In this situation, the VTEP connected to VM 2 advertises MAC/IP advertisement routes that contain ARP or ND information to the gateway. The IP address and MAC address in the routes are the IP address and MAC address of VM 2, respectively. The gateway learns the ARP or ND information and issues the information to the forwarding table. When VM 1 visits 10.1.2.2 in the external network, the gateway will forward the traffic to VM 2. As a result, VM 1 cannot visit 10.1.2.2.

To resolve the above issue, perform this task on the VSI interface to disable the VSI interface from learning ARP or ND information across subnets from MAC/IP advertisement routes.

Procedure

1. Enter system view.

   system-view

2. Enter VSI interface view on a centralized EVPN gateway.

   interface vsi-interface *vsi-interface-id*

   For more information about this command, see *VXLAN Command Reference*.

3. Disable the VSI interface from learning ARP or ND information that does not belong to its subnet from MAC/IP advertisement routes.

   evpn span-segment { **arp-learning** | **nd-learning** } **disable**

   By default, the VSI interface on a centralized EVPN gateway learns ARP or ND information that does not belong to its subnet from MAC/IP advertisement routes.

Enabling ARP mobility event suppression

About this task

On an EVPN VXLAN network, misconfiguration of IP addresses might cause two sites to contain the same IP address. In this condition, VTEPs at the two sites constantly synchronize and update EVPN ARP entries and determine that ARP mobility events occur. As a result, an inter-site loop might occur, and the bandwidth is occupied by ARP entry synchronization traffic. To eliminate loops and suppress those ARP mobility events, enable ARP mobility event suppression on the VTEPs. This feature allows an IP address to move a specified number of times (the ARP mobility suppression threshold) from a site within an ARP mobility detection cycle. If an IP address moves more than the ARP mobility suppression threshold, the VTEP at the site will suppress the last ARP move to the local site and will not advertise ARP information for the IP address.

Restrictions and guidelines

ARP mobility event suppression takes effect only on an EVPN VXLAN network configured with distributed VXLAN IP gateways.

After you execute the **undo evpn route arp-mobility suppression** command or the suppression time expires, a VTEP acts as follows:

- Advertises ARP information immediately for the suppressed ARP entries that have not aged out.

- Relearns ARP information for the suppressed ARP entries that have aged out and advertises the ARP information.

If both MAC address entry conflicts and ARP entry conflicts exist for a MAC address, you must enable both MAC mobility event suppression and ARP mobility event suppression. If you enable only MAC mobility event suppression, the system cannot suppress MAC mobility events for the MAC address.

Procedure

1. Enter system view.

   system-view

2. Enable ARP mobility event suppression.

   evpn route arp-mobility suppression [**detect-cycle** *detect-time* | **detect-threshold** *move-times* | **suppression-time** [*suppression-time* | **permanent**]] *

   By default, ARP mobility event suppression is disabled.

Enabling ND mobility event suppression

About this task

On an EVPN VXLAN network, misconfiguration of IP addresses might cause two sites to contain the same IP address. In this condition, VTEPs at the two sites constantly synchronize and update EVPN ND entries and determine that ND mobility events occur. As a result, an inter-site loop might occur, and the bandwidth is occupied by ND entry synchronization traffic. To eliminate loops and suppress those ND mobility events, enable ND mobility event suppression on the VTEPs. This feature allows an IP address to move a specified number of times (the ND mobility suppression threshold) from a site within an ND mobility detection cycle. If an IP address moves more than the ND mobility suppression threshold, the VTEP at the site will suppress the last ND move to the local site and will not advertise ND information for the IP address.

Restrictions and guidelines

After you execute the **undo evpn route nd-mobility suppression** command or the suppression time expires, a VTEP acts as follows:

- Advertises ND information immediately for the suppressed ND entries that have not aged out.
- Relearns ND information for the suppressed ND entries that have aged out and advertises the ND information.

ND mobility event suppression takes effect only on the following EVPN VXLAN networks:

- EVPN VXLAN network enabled with ND flood suppression.
- EVPN VXLAN network configured with distributed VXLAN IP gateways.

If both MAC address entry conflicts and ND entry conflicts exist for a MAC address, you must enable both MAC mobility event suppression and ND mobility event suppression. If you enable only MAC mobility event suppression, the system cannot suppress MAC mobility events for the MAC address.

Procedure

1. Enter system view.

   system-view

2. Enable ND mobility event suppression.

   evpn route nd-mobility suppression [**detect-cycle** *detect-time* | **detect-threshold** *move-times* | **suppression-time** [*suppression-time* | **permanent**]] *

   By default, ND mobility event suppression is disabled.

Enabling ARP request proxy

About this task

ARP request proxy allows a VSI interface to send an ARP request sourced from itself when the VTEP forwards an ARP request. This feature helps resolve certain communication issues.

In an EVPN VXLAN network, VM 1 and VM 2 are attached to VTEP 1 and VTEP 2, respectively, and the VMs are in the same subnet. The gateway interfaces of VM 1 and VM 2 are VSI-interface 1 on VTEP 1 and VSI-interface 2 on VTEP 2, respectively. The following conditions exist on the VTEPs:

- The VTEPs have established BGP EVPN neighbor relationships.
- EVPN is disabled from learning MAC addresses from ARP information.
- MAC address advertisement is disabled, and advertised MAC addresses are withdrawn.
- Remote-MAC address learning is disabled.
- Local proxy ARP is enabled on the VSI interfaces.
- The VSI interfaces use different IP addresses and MAC addresses.

In this network, when VM 1 attempts to communicate with VM 2, the following procedure occurs:

1. VM 1 sends an ARP request.
2. VTEP 1 learns the MAC address of VM 1 from the ARP request, replies to VM 1 on behalf of VM 2, and sends an ARP request to obtain the MAC address of VM 2.
3. VTEP 2 forwards the ARP request, and VM 2 replies to VTEP 1.
4. VTEP 2 forwards the ARP reply sent by VM 2 without learning the MAC address of VM 2 because EVPN is disabled from learning MAC addresses from ARP information.
5. VTEP 1 does not learn the MAC address of VM 2 because remote-MAC address learning is disabled.

As a result, VM 1 fails to communicate with VM 2.

For VM 1 to communicate with VM 2, enable ARP request proxy on VSI-interface 2 of VTEP 2. When receiving the ARP request sent by VTEP 1, VTEP 2 forwards it and sends an ARP request sourced from VSI-interface 2 simultaneously, and VM 2 replies to both ARP requests. Then, VTEP 2 learns the MAC address of VM 2 from the ARP reply destined from VSI-interface 2 and advertises the MAC address to VTEP 1 through BGP EVPN routes. In this way, VTEP 1 obtains the MAC address of VM 2, and VM 1 and VM 2 can communicate.

Procedure

1. Enter system view.

   `system-view`
2. Enter VSI interface view.

   `interface vsi-interface` *`vsi-interface-id`*
3. Enable ARP request proxy.

   `arp proxy-send enable`

   By default, ARP request proxy is disabled on VSI interfaces.

Enabling ND request proxy

About this task

ND request proxy allows a VSI interface to send an ND request sourced from itself when the VTEP forwards an ND request. This feature helps resolve certain communication issues.

In an EVPN VXLAN network, VM 1 and VM 2 are attached to VTEP 1 and VTEP 2, respectively, and the VMs are in the same subnet. The gateway interfaces of VM 1 and VM 2 are VSI-interface 1 on VTEP 1 and VSI-interface 2 on VTEP 2, respectively. The following conditions exist on the VTEPs:

- The VTEPs have established BGP EVPN neighbor relationships.
- EVPN is disabled from learning MAC addresses from ND information.
- MAC address advertisement is disabled, and advertised MAC addresses are withdrawn.
- Remote-MAC address learning is disabled.
- Local proxy ND is enabled on the VSI interfaces.
- The VSI interfaces use different IP addresses and MAC addresses.

In this network, when VM 1 attempts to communicate with VM 2, the following process occurs:

1. VM 1 sends an NS packet.
2. VTEP 1 learns the MAC address of VM 1 from the NS packet, replies to VM 1 on behalf of VM 2, and sends an NS packet to obtain the MAC address of VM 2.
3. VTEP 2 forwards the NS packet, and VM 2 replies to VTEP 1.
4. VTEP 2 forwards the NA packet sent by VM 2 without learning the MAC address of VM 2 because EVPN is disabled from learning MAC addresses from ND information.
5. VTEP 1 does not learn the MAC address of VM 2 because remote-MAC address learning is disabled.

As a result, VM 1 fails to communicate with VM 2.

For VM 1 to communicate with VM 2, enable NS packet proxy on VSI-interface 2 of VTEP 2. When receiving the NS packet sent by VTEP 1, VTEP 2 forwards it and sends an NS packet sourced from VSI-interface 2 simultaneously, and VM 2 replies to both NS packets. Then, VTEP 2 learns the MAC address of VM 2 from the NA packet sent to VSI-interface 2 and advertises the MAC address to VTEP 1 through BGP EVPN routes. In this way, VTEP 1 obtains the MAC address of VM 2, and VM 1 and VM 2 can communicate.

Procedure

1. Enter system view.
   `system-view`
2. Enter VSI interface view.
   `interface vsi-interface` *`vsi-interface-id`*
3. Enable ND request proxy.
   `ipv6 nd proxy-send enable`
   By default, ND request proxy is disabled on VSI interfaces.

Configuring BGP EVPN route redistribution and advertisement

Redistributing MAC/IP advertisement routes into BGP unicast routing tables

About this task

This task enables the device to redistribute received MAC/IP advertisement routes that contain ARP or ND information into a BGP unicast routing table.

- If you perform this task for the BGP IPv4 or IPv6 unicast address family, the device will redistribute the routes into the BGP IPv4 or IPv6 unicast routing table. In addition, the device will advertise the routes to the local site.
- If you perform this task for the BGP-VPN IPv4 or IPv6 unicast address family, the device will redistribute the routes into the BGP-VPN IPv4 or IPv6 unicast routing table of the corresponding VPN instance. To advertise the routes to the local site, you must configure the **advertise l2vpn evpn** command.

Procedure (BGP instance view)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP IPv4 or IPv6 unicast address family view.

   address-family { **ipv4** | **ipv6** }
4. Redistribute MAC/IP advertisement routes that contain ARP or ND information into the BGP IPv4 or IPv6 unicast routing table.

   import evpn mac-ip

   By default, MAC/IP advertisement routes that contain ARP or ND information are not redistributed into the BGP IPv4 or IPv6 unicast routing table.

Procedure (BGP-VPN instance view)

1. Enter system view.

   system-view
2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*
4. Enter BGP-VPN IPv4 or IPv6 unicast address family view.

   address-family { **ipv4** | **ipv6** }
5. Redistribute MAC/IP advertisement routes that contain ARP or ND information into the BGP-VPN IPv4 or IPv6 unicast routing table.

   import evpn mac-ip

   By default, MAC/IP advertisement routes that contain ARP or ND information are not redistributed into the BGP-VPN IPv4 or IPv6 unicast routing table.

Setting the metric of BGP EVPN routes added to a VPN instance's routing table

About this task

After you perform this task, the device sets the metric of a BGP EVPN route added to a VPN instance's routing table to the metric of the IGP route pointing to the next hop in the original BGP EVPN route.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP EVPN address family view.

   address-family l2vpn evpn

4. Set the metric of a BGP EVPN route added to a VPN instance's routing table to the metric of the IGP route pointing to the next hop in the original BGP EVPN route.

   igp-metric inherit

   By default, the device sets the metric to 0 when adding BGP EVPN routes a VPN instance's routing table.

Setting the preference of MAC/IP advertisement routes redistributed to the BGP-VPN unicast address families

About this task

Different routing protocols, direct routes, and static routes might discover different routes for the same destination in the IP routing table, but not all of these routes are optimal. To determine the optimal route, each routing protocol, direct route, and static route is assigned a preference, and the route discovered by the routing protocol with higher preference will become the optimal route.

Set the preference of MAC/IP advertisement routes redistributed to the BGP-VPN unicast address families for the routes to become the optimal routes in the IP routing table of a VPN instance.

Restrictions and guidelines

To use a routing policy for preference configuration, first use the **apply preference** command to set the preference of matching routes for the routing policy. If you do not set the preference of matching routes, the matching routes will use the default preference.

Procedure

1. Enter system view.

   system-view

2. Enter BGP instance view.

   bgp *as-number* [**instance** *instance-name*]

3. Enter BGP-VPN instance view.

   ip vpn-instance *vpn-instance-name*

4. Enter BGP-VPN IPv4 or IPv6 unicast address family view.
   - Enter BGP-VPN IPv4 unicast address family view.

     address-family ipv4 [**unicast**]
   - Enter BGP-VPN IPv6 unicast address family view.

     address-family ipv6 [**unicast**]

5. Set the preference of MAC/IP advertisement routes redistributed to the BGP-VPN unicast address family.

   evpn mac-ip preference { *external-preference internal-preference* | **route-policy** *route-policy-name* }

   By default, the preference of both EBGP and IBGP routes is 255.

Enabling BGP EVPN route advertisement to the local site

About this task

This feature enables the device to advertise BGP EVPN routes to the local site after the device adds the routes to the routing table of a VPN instance. The BGP EVPN routes here are IP prefix advertisement routes and MAC/IP advertisement routes that contain ARP or ND information.

Procedure (IPv4)

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
4. Enter BGP-VPN IPv4 unicast address family view.
   address-family ipv4 [**unicast**]
5. Enable BGP EVPN route advertisement to the local site.
   advertise l2vpn evpn
   By default, BGP EVPN route advertisement to the local site is enabled.

Procedure (IPv6)

1. Enter system view.
   system-view
2. Enter BGP instance view.
   bgp *as-number* [**instance** *instance-name*]
3. Enter BGP-VPN instance view.
   ip vpn-instance *vpn-instance-name*
4. Enter BGP-VPN IPv6 unicast address family view.
   address-family ipv6 [**unicast**]
5. Enable BGP EVPN route advertisement to the local site.
   advertise l2vpn evpn
   By default, BGP EVPN route advertisement to the local site is enabled.

Disabling flooding for a VSI

About this task

By default, the VTEP floods broadcast, unknown unicast, and unknown multicast frames received from the local site to the following interfaces in the frame's VXLAN:

- All site-facing interfaces except for the incoming interface.
- All VXLAN tunnel interfaces.

When receiving broadcast, unknown unicast, and unknown multicast frames on VXLAN tunnel interfaces, the device floods the frames to all site-facing interfaces in the frames' VXLAN.

To confine a kind of flood traffic, disable flooding for that kind of flood traffic on the VSI bound to the VXLAN.

You can use selective flood to exclude a remote MAC address from the remote flood suppression done by using the **flooding disable** command. The VTEP will flood the frames destined for the specified MAC address to remote sites when floods are confined to the local site.

For more information about the VXLAN commands in this task, see *VXLAN Command Reference*.

Procedure

1. Enter system view.
   system-view
2. Enter VSI view.
   vsi *vsi-name*
3. Disable flooding for the VSI.
   flooding disable { **all** | { **broadcast** | **unknown-multicast** | **unknown-unicast** } * } [**all-direction**]
   By default, flooding is enabled for a VSI.
4. (Optional.) Enable selective flood for a MAC address.
   selective-flooding mac-address *mac-address*

Enabling ARP or ND flood suppression

Configuring ARP/ND flood suppression on a VSI

About this task

Use ARP or ND flood suppression to reduce ARP request broadcasts or ND request multicasts.

This feature is suitable for AC-VSI mapping and VLAN-based VXLAN assignment scenarios. In these scenarios, a VSI must be mapped through an AC and you must configure flood suppression in VSI view.

For more information about VLAN-based VXLAN assignment, see *VXLAN Configuration Guide*.

Restrictions and guidelines

The aging timer is fixed at 25 minutes for ARP or ND flood suppression entries. If the **flooding disable** command is configured, set the MAC aging timer to a higher value than the aging timer for ARP or ND flood suppression entries on all VTEPs. This setting prevents the traffic blackhole that occurs when a MAC address entry ages out before its ARP or ND flood suppression entry ages out. To set the MAC aging timer, use the **mac-address timer** command.

When remote ARP or ND learning is disabled for VXLANs, the device does not use ARP or ND flood suppression entries to respond to ARP or ND requests received on VXLAN tunnels.

To delete ARP flood suppression entries, use the **reset arp suppression vsi** command instead of the **reset arp** command. For more information about the **reset arp suppression vsi** command, see *VXLAN Command Reference*. For more information about the **reset arp** command, see ARP commands in *Layer 3—IP Services Command Reference*.

To delete ND flood suppression entries, use the **reset ipv6 nd suppression vsi** command instead of the **reset ipv6 neighbors** command. For more information about the **reset ipv6 nd suppression vsi** command, see *VXLAN Command Reference*. For more information about the **reset ipv6 neighbors** command, see IPv6 basics commands in *Layer 3—IP Services Command Reference*.

Enabling ARP flood suppression

1. Enter system view.

system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Enable ARP flood suppression.

   arp suppression enable

   By default, ARP flood suppression is disabled.

   For more information about this command, see *VXLAN Command Reference*.

Enabling ND flood suppression

1. Enter system view.

   system-view

2. Enter VSI view.

   vsi *vsi-name*

3. Enable ND flood suppression.

   ipv6 nd suppression enable

   By default, ND flood suppression is disabled.

   For more information about this command, see *VXLAN Command Reference*.

Configuring ARP flood suppression for a VLAN

About this task

This feature applies to the VLAN-VXLAN mapping scenario. In this scenario, a VLAN is directly mapped to a VXLAN, and you must configure the flood suppression feature in VLAN view.

Procedure

1. Enter system view.

   system-view

2. Enter VLAN view.

   vlan *vlan-id*

3. Enable ARP flood suppression for the VLAN.

   arp suppression enable

   By default, ARP flood suppression is disabled.

   For more information about this command, see VXLAN commands in *VXLAN Command Reference*.

Enabling packet statistics for VXLAN tunnels

About this task

Perform this task to enable packet statistics globally for automatically created VXLAN tunnels.

To display the packet statistics for a VXLAN tunnel, use the **display interface tunnel** command in any view.

To clear the packet statistics for a VXLAN tunnel, use the **reset counters interface tunnel** command in user view.

Procedure

1. Enter system view.

system-view

2. Enable packet statistics for automatically created VXLAN tunnels.

   tunnel statistics vxlan auto

   By default, the packet statistics feature is disabled for automatically created VXLAN tunnels.

   For more information about this command, see *VXLAN Command Reference*.

Testing the connectivity of a VXLAN tunnel

Enabling overlay OAM

About this task

You must enable overlay OAM on the tunnel destination device for a VXLAN tunnel before you can use the **ping vxlan** or **tracert vxlan** command to test reachability of the VXLAN tunnel on the tunnel source device.

Restrictions and guidelines

To specify the **-r 3** parameter in the **ping vxlan** or **tracert vxlan** command on the tunnel source device, you must also enable overlay OAM on the tunnel source device.

Procedure

1. Enter system view.

   system-view

2. Enable overlay OAM.

   overlay oam enable

   By default, overlay OAM is disabled.

Pinging a VXLAN tunnel destination

About this task

Perform this task to test the connectivity of a VXLAN tunnel in an EVPN VXLAN network when the tunnel has traffic loss or interruption issues. The process of a ping VXLAN operation is as follows:

1. The tunnel source VTEP sends VXLAN-encapsulated VXLAN echo requests to the tunnel destination VTEP.
2. The tunnel destination VTEP responds with VXLAN echo replies.
3. The tunnel source VTEP outputs packet statistics and the test result based on the received VXLAN echo replies.

Restrictions and guidelines

Before you perform this task on the tunnel source device, you must enable overlay OAM on the tunnel destination device by using the **overlay oam enable** command.

The VTEP can distribute VXLAN echo requests among multiple paths to the destination based on the source UDP port. When a VXLAN tunnel has multiple paths on the transport network, you can configure load sharing parameters to ensure accuracy of the test result. You can use one of the following methods to configure source UDP ports for VXLAN echo requests:

- Specify a source UDP port range. The device will send VXLAN echo requests sourced from each UDP port in the UDP port range. You need to execute the **ping vxlan** command only once.

- Specify load balancing parameters such as source and destination MAC addresses, source and destination IP addresses, and protocol for the VTEP to calculate a source UDP port number. You need to execute the **ping vxlan** command multiple times to test connectivity of all paths.

The load balancing parameters change only the source UDP port number of VXLAN echo requests. Other fields of the requests will not be changed.

If you specify the **vxlan-source-udpport** *vxlan-source-udpport* [*end-vxlan-src-udpport*] parameters, the number of VXLAN echo requests sourced from each UDP port in the UDP port range is determined by the **-c** *count* parameter.

Procedure

Execute the following command in any view.

ping vxlan [**-a** *inner-src-address* | **-c** *count* | **-m** *interval* | **-r** *reply-mode* | **-t** *timeout* | **-tos** *tos-value*] * **vxlan-id** *vxlan-id* **tunnel-source** *source-address* **tunnel-destination** *dest-address* [**destination-udpport** *dest-port*] [**vxlan-source-address** *vxlan-source-address*] [**load-balance** { **vxlan-source-udpport** *vxlan-source-udpport* [*end-vxlan-src-udpport*] | **source-address** *lb-src-address* **destination-address** *lb-dest-address* **protocol** { **udp** | *lb-protocol-id* } **source-port** *lb-src-port* **destination-port** *lb-dest-port* **source-mac** *lb-source-mac* **destination-mac** *lb-destination-mac* }]

For more information about this command, see *VXLAN Command Reference*.

Tracing the path to a VXLAN tunnel destination

About this task

Perform this task to locate failed nodes on the path for a VXLAN tunnel that has traffic loss or interruption issues in an EVPN VXLAN network. The process of a tracert VXLAN operation is as follows:

1. The tunnel source VTEP sends VXLAN-encapsulated VXLAN echo requests to the tunnel destination VTEP. The TTL in the IP header of the requests is set to 1.
2. The first hop on the path responds to the tunnel source VTEP with a TTL-expired ICMP error message.
3. The tunnel source VTEP sends VXLAN echo requests with the TTL set to 2.
4. The second hop responds with a TTL-expired ICMP error message.
5. This process continues until a VXLAN echo request reaches the tunnel destination VTEP or the maximum TTL value is reached. If a VXLAN echo request reaches the tunnel destination VTEP, the tunnel destination VTEP sends a VXLAN echo reply to the tunnel source VTEP.
6. The tunnel source VTEP outputs packet statistics and the test result based on the received ICMP error messages and whether a VXLAN echo reply is received.

Restrictions and guidelines

Before you perform this task on the tunnel source device, you must enable overlay OAM on the tunnel destination device by using the **overlay oam enable** command.

The VTEP can distribute VXLAN echo requests among multiple paths to the destination based on the source UDP port. When a VXLAN tunnel has multiple paths on the transport network, you can configure load sharing parameters to ensure accuracy of the test result. You can use one of the following methods to configure source UDP ports for VXLAN echo requests:

- Specify a source UDP port range. The device will send VXLAN echo requests sourced from each UDP port in the UDP port range. You need to execute the **ping vxlan** command only once.

- Specify load balancing parameters such as source and destination MAC addresses, source and destination IP addresses, and protocol for the VTEP to calculate a source UDP port number. You need to execute the **ping vxlan** command multiple times to test connectivity of all paths.

The load balancing parameters change only the source UDP port number of VXLAN echo requests. Other fields of the requests will not be changed.

Procedure

Execute the following command in any view.

tracert vxlan [**-a** *inner-src-address* | **-h** *ttl-value* | **-r** *reply-mode* | **-t** *timeout*] * **vxlan-id** *vxlan-id* **tunnel-source** *source-address* **tunnel-destination** *dest-address* [**destination-udpport** *dest-port*] [**vxlan-source-address** *vxlan-source-address*] [**load-balance** { **vxlan-source-udpport** *vxlan-source-udpport* | **source-address** *lb-src-address* **destination-address** *lb-dest-address* **protocol** { **udp** | *lb-protocol-id* } **source-port** *lb-src-port* **destination-port** *lb-dest-port* **source-mac** *lb-source-mac* **destination-mac** *lb-destination-mac* }]

For more information about this command, see *VXLAN Command Reference*.

Enabling SNMP notifications for EVPN

About this task

If SNMP notifications are enabled for EVPN, a MAC mobility suppression notification is sent to SNMP module after the MAC mobility suppression threshold is reached. For SNMP notifications to be sent correctly, you must also configure SNMP on the device. For more information about SNMP configuration, see *Network Management and Monitoring Configuration Guide*.

Procedure

1. Enter system view.

   system-view

2. Enable SNMP notifications for EVPN.

   snmp-agent trap enable evpn [**mac-mobility-suppression**]

   By default, SNMP notifications are disabled for EVPN.

Configuring EVPN M-LAG

About this task

EVPN M-LAG virtualizes two VTEPs or EVPN gateways into one M-LAG system to avoid single points of failure. The VTEPs or EVPN gateways use a virtual VTEP address to establish VXLAN tunnels to remote devices.

An AC that is attached to only one of the VTEPs in an M-LAG system is called a single-homed AC. To ensure that the traffic of a single-homed AC is forwarded to its attached VTEP, specify the IP addresses of the VTEPs in the M-LAG system by using the **evpn m-lag local** command. After you configure this command, each VTEP in an M-LAG system changes the next hop of the routes for single-homed ACs to its local VTEP IP address when advertising the routes. When a VTEP receives BGP EVPN routes from the peer VTEP IP address specified by using this command, it does not set up a VXLAN tunnel to the peer VTEP.

You must execute the **evpn m-lag local** command if single-homed ACs are attached to an M-LAG system that uses a direct peer link. You do not need to execute this command on an M-LAG

system that uses a tunnel peer link. In such an M-LAG system, a VTEP uses the source IP address of the peer link as the next hop of routes for single-homed ACs to ensure correct traffic forwarding.

Restrictions and guidelines

EVPN M-LAG restrictions and guidelines

When you configure EVPN M-LAG, follow these restrictions and guidelines:

- For an M-LAG member device to re-establish VXLAN tunnels, you must execute the **address-family l2vpn evpn** command in BGP instance view after you enable or disable EVPN M-LAG.
- You cannot specify a secondary IP address of an interface as the virtual VTEP address.
- Make sure the virtual VTEP address and the underlay network use the same IP version. Otherwise, the VTEPs in an M-LAG group cannot set up VXLAN tunnels with remote VTEPs.

If an Ethernet aggregate link is used as the peer link, follow these restrictions:

- If the frame match criteria of dynamic ACs on the peer link are created based on site-facing Ethernet service instances, you can configure only the following criteria for site-facing Ethernet service instances:
  - **encapsulation s-vid** { *vlan-id* | *vlan-id-list* }
  - **encapsulation untagged**

  In addition, you must set the access mode to VLAN for the site-facing Ethernet service instances.
- You must configure VLAN access mode for the site-facing Ethernet service instances when the frame match criteria of dynamic ACs on the IPL are created based on VXLAN IDs.
- Make sure the following settings are consistent on the M-LAG member devices:
  - Ethernet service instances and their match criterion on the M-LAG interfaces in the same M-LAG group or single-homed site-facing interfaces.
  - VXLAN IDs of VSIs.

  In addition, the Ethernet service instances must be created manually.
- As a best practice, do not redistribute external routes on the M-LAG member devices.

If a VXLAN tunnel is used as the peer link, make sure the following settings are consistent on the M-LAG member devices:

- Ethernet service instances and their match criterion on the M-LAG interfaces in the same M-LAG group.
- VXLAN IDs of VSIs.

In addition, the Ethernet service instances must be created manually.

On an EVPN M-LAG network, to forward VXLAN-encapsulated overlay Layer 3 packets through the peer link, you must create VLAN-interface 1 on M-LAG member devices.

Forwarding entry configuration restrictions and guidelines

The VTEPs in an M-LAG system synchronize local and remote MAC address entries with each other over the peer link. However, they do not synchronize MAC address entry deletions. When you delete a MAC address entry from one VTEP, the other VTEP retains the entry that contains the same MAC address until the entry ages out.

At an IPv6 site, if you enable ND flood suppression on the VTEPs in an M-LAG system, both VTEPs reply with NA packets when one of the VTEPs receives an NS packet on an M-LAG interface.

If a route reflector reflects routes between the VTEPs in an M-LAG system, after you execute the **evpn m-lag local** command on both VTEPs or execute the **undo evpn m-lag local** command on one of the VTEPs, also execute the following commands on the VTEPs:

- **reset arp**.
- **reset arp suppression vsi**.
- **reset ipv6 neighbors**.
- **reset ipv6 nd suppression vsi**.

These commands clear ARP- and ND-related entries on the VTEPs to ensure correct forwarding.

Prerequisites

In addition to EVPN M-LAG configuration, you must configure the following settings:

- Configure other M-LAG and EVPN settings depending on your network. For information about M-LAG configuration, see *Layer 2—LAN Switching Configuration Guide*.
- Use the **m-lag mad exclude interface** command to exclude the following interfaces:
  - Set the default M-LAG MAD action to NONE by using the **m-lag mad default-action none** command.
  - Do not configure the M-LAG MAD action on the VLAN interfaces of the VLANs to which the M-LAG interfaces and peer-link interfaces belong. These interfaces will not be shut down by M-LAG MAD.
  - If you use a direct peer link, add the uplink Layer 3 interfaces, VLAN interfaces, and physical interfaces to the list of included interfaces by using the **m-lag mad include interface** command. These interfaces will be shut down by M-LAG MAD. For a tunnel peer link, you must exclude the physical outgoing interfaces for the tunnel from the shutdown action by M-LAG MAD.
  - Do not configure the M-LAG MAD action on the interfaces used by EVPN, including the VSI interfaces, interfaces that provide BGP peer addresses, and interfaces used for setting up the keepalive link. These interfaces will not be shut down by M-LAG MAD.
  - Do not configure the M-LAG MAD action on the interface that provides the IP address specified by using the **evpn m-lag group** command. These interfaces will not be shut down by M-LAG MAD.
- Execute the **m-lag restore-delay** command to set the data restoration interval to a value equal to or larger than 300 seconds.
- You must disable spanning tree on the Layer 2 Ethernet interface that acts as the physical traffic outgoing interface of a VXLAN tunnel. If you enable spanning tree on that interface, the upstream device will falsely block the interfaces connected to the M-LAG member devices.
- After you execute the **l2vpn m-lag peer-link ac-match-rule vxlan-mapping** command, the PVIDs of the peer-link interfaces are automatically set to 4094. Therefore, to ensure normal traffic forwarding when the **l2vpn m-lag peer-link ac-match-rule vxlan-mapping** command is executed, set the PVIDs of the peer-link interfaces are automatically to 4094 in advance.

If you use a tunnel peer link, you must also complete the following tasks:

- Manually create the VXLAN tunnel interface and configure it as the peer-link interface. An automatically created VXLAN tunnel cannot be used as a peer link.
- Use the **m-lag mad exclude interface** command to exclude VXLAN tunnel interfaces and their traffic outgoing interfaces from the MAD shutdown action by M-LAG before you configure them as peer-link interfaces. If you have configured the VXLAN tunnel interfaces as peer-link interfaces before excluding them and their traffic outgoing interfaces from the MAD shutdown action, you must first remove the peer-link interface configuration. After the VXLAN tunnel interfaces and their traffic outgoing interfaces come up, exclude the interfaces from the MAD shutdown action by M-LAG. Then, configure the VXLAN tunnel interfaces as peer-link interfaces.

- The source address of the peer-link VXLAN tunnel must be the address used by the device to establish BGP peer relationships with other devices.
- To prioritize transmission of M-LAG protocol packets on the peer link, use the **tunnel tos** command on the peer-link VXLAN tunnel interface to set a high ToS value for tunneled packets.
- Specify the virtual VTEP address and the source address of the tunnel peer link as the IP addresses of different loopback interfaces. Configure a routing protocol to advertise the IP addresses.
- Use the **reserved vxlan** command to specify a reserved VXLAN to forward M-LAG protocol packets. The M-LAG member devices in an M-LAG system must have the same reserved VXLAN.

Procedure (IPv4)

1. Enter system view.

   system-view
2. Enable EVPN M-LAG and specify the virtual VTEP address.

   evpn m-lag group *virtual-vtep-ipv4*

   By default, EVPN M-LAG is disabled.

   To modify the virtual VTEP address, you must first delete the original virtual VTEP address.
3. Specify the IP addresses of the VTEPs in the M-LAG system.

   evpn m-lag local *local-ipv4-address* **remote** *remote-ipv4-address* [**mac-ip**]

   By default, the IP addresses of the VTEPs in an M-LAG system are not specified.

   Make sure the IP address of the local VTEP belongs to a local interface. Make sure the local VTEP IP address and peer VTEP IP address are reversed on the VTEPs in the M-LAG system.
4. Allow only MAC/IP advertisement routes to carry the local VTEP address.

   evpn m-lag local mac-ip

   By default, IMET routes, MAC/IP advertisement routes, and IP prefix advertisement routes carry the local VTEP address.
5. (Optional.) Enable the device to create frame match criteria based on VXLAN IDs for the dynamic ACs on the peer link when the M-LAG system uses a direct peer link.

   l2vpn m-lag peer-link ac-match-rule vxlan-mapping

   By default, on an EVPN M-LAG system that uses a direct peer link, dynamic ACs on the peer link use frame match criteria that are identical to those of site-facing ACs.

   If you do not execute this command, do not configure overlapping outer VLAN IDs for Ethernet service instances of different VSIs.

   If you execute this command, do not create VXLANs with IDs larger than 16000000.

Procedure (IPv6)

1. Enter system view.

   system-view
2. Enable EVPN M-LAG and specify the virtual VTEP address.

   evpn m-lag group *virtual-vtep-ipv6*

   By default, EVPN M-LAG is disabled.

   To modify the virtual VTEP address, you must first delete the original virtual VTEP address.
3. Specify the IP addresses of the VTEPs in the M-LAG system.

evpn m-lag local *local-ipv6-address* **remote** *remote-ipv6-address* [**mac-ip**]

By default, the IP addresses of the VTEPs in an M-LAG system are not specified.

Make sure the IP address of the local VTEP belongs to a local interface. Make sure the local VTEP IP address and peer VTEP IP address are reversed on the VTEPs in the M-LAG system.

4. Allow only MAC/IP advertisement routes to carry the local VTEP address.

   evpn m-lag local mac-ip

   By default, IMET routes, MAC/IP advertisement routes, and IP prefix advertisement routes carry the local VTEP address.

5. (Optional.) Enable the device to create frame match criteria based on VXLAN IDs for the dynamic ACs on the peer link.

   l2vpn m-lag peer-link ac-match-rule vxlan-mapping

   By default, on an EVPN M-LAG system that uses a direct peer link, dynamic ACs on the peer link use frame match criteria that are identical to those of site-facing ACs.

   If you do not execute this command, do not configure overlapping outer VLAN IDs for Ethernet service instances of different VSIs.

   If you execute this command, do not create VXLANs with IDs larger than 16000000.

Configuring EVPN VXLAN tunnels to learn dynamic entries

About this task

In an EVPN VXLAN network, if certain special reasons cause EVPN routes to be missing, the device cannot learn MAC, ARP, and ND entries because MAC, ARP, and ND entries are transmitted through EVPN routes. This leads to packet forwarding failures. To resolve the preceding issues, execute this command to allow the EVPN VXLAN tunnels to dynamically learn MAC, ARP, and ND entries. This enables the device to learn these entries and ensures proper packet forwarding.

For example, in an EVPN VXLAN centralized gateway network, the EVPN VXLAN tunnels do not learn MAC, ARP, and ND entries by default. When the leaf device is not configured with the learning feature, the gateway device might fail to learn the relevant entries, affecting packet forwarding. After you execute this command, the device will allow the EVPN VXLAN tunnels to learn dynamic entries, ensuring normal packet forwarding.

Procedure

1. Enter system view.

   system-view

2. Allow EVPN VXLAN tunnels to learn ARP, MAC, and ND entries.

   evpn vxlan tunnel learning { **arp** | **mac** | **nd** } *

   By default, EVPN VXLAN tunnels do not learn dynamic entries.

Display and maintenance commands for EVPN

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display BGP peer group information.	**display bgp** [**instance** *instance-name*] **group l2vpn evpn** [**group-name** *group-name*]

	display bgp [**instance** *instance-name*] **l2vpn evpn** [**peer** { *ipv4-address* | *ipv6-address* } { **advertised-routes** | **received-routes** } [**statistics**] | [**route-distinguisher** *route-distinguisher* | **route-type** { **auto-discovery** | **es** | **igmp-ls** | **igmp-js** | **imet** | **ip-prefix** | **mac-ip** | **s-pmsi** | **smet** }] * [{ *evpn-route route-length* | *evpn-prefix* } [**advertise-info** | **as-path** | **cluster-list** | **community** | **ext-community**] | { *ipv4-address* | *ipv6-address* | *mac-address* } [**verbose**]] | **statistics**]
Display BGP EVPN routes.	**display bgp** [**instance** *instance-name*] **l2vpn evpn** [**route-distinguisher** *route-distinguisher*] [**statistics**] **community** [*community-number*&<1-32> | *aa:nn*&<1-32>] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] [**whole-match**]
	display bgp [**instance** *instance-name*] **l2vpn evpn** [**route-distinguisher** *route-distinguisher*] [**statistics**] **community-list** { *basic-community-list-number* | *adv-community-list-number* | *comm-list-name* } [**whole-match**]
	display bgp [**instance** *instance-name*] **l2vpn evpn** [**route-distinguisher** *route-distinguisher*] [**statistics**] **ext-community** [**bandwidth** *link-bandwidth-value* | **rt** *route-target* | **soo** *site-of-origin*]&<1-32> [**whole-match**]
Display BGP peer or peer group information.	**display bgp** [**instance** *instance-name*] **peer l2vpn evpn** [*ipv4-address mask-length* | *ipv6-address prefix-length* | { *ipv4-address* | *ipv6-address* | **group-name** *group-name* } **log-info** | [*ipv4-address*] **verbose**]
Display information about BGP update groups.	**display bgp** [**instance** *instance-name*] **update-group l2vpn evpn** [*ipv4-address* | *ipv6-address*]
Display information about IPv4 peers that are automatically discovered through BGP.	**display evpn auto-discovery** { { **imet** | **mac-ip** } [**peer** *ip-address*] [**vsi** *vsi-name*] | **macip-prefix** [**nexthop** *next-hop*] [**count**] }
Display M-LAG-synchronized MAC address entries.	**display evpn m-lag synchronized-mac** [**vsi** *vsi-name*] [**count**]
Display EVPN ES information.	**display evpn es** { **local** [**vsi** *vsi-name* | **xconnect-group** *group-name*] [**esi** *esi-id*] [**verbose**] | **remote** [**vsi** *vsi-name* | **xconnect-group** *group-name*] [**esi** *esi-id*] [**nexthop** *next-hop*] }
Display DF election information.	**display evpn df-election** [**vsi** *vsi-name*] [**esi** *esi-id*]

Display EVPN instance information.	**display evpn instance** [**name** *instance-name* | **vsi** *vsi-name*] | **xconnect-group** *group-name*] [**vxlan**]
Display information about IPv6 peers that are automatically discovered through BGP.	**display evpn ipv6 auto-discovery** { { **imet** | **mac-ip** } [**peer** *ipv6-address*] [**vsi** *vsi-name*] | **mac-ip** | **macip-prefix** [**nexthop** *next-hop*] [**count**] }
Display IPv6 EVPN MAC address entries.	**display evpn ipv6 route mac** [**local** | **remote** | **nexthop** *ipv6-address*] [**vsi** *vsi-name*] [**count**]
Display EVPN ARP entries.	**display evpn route arp** [**local** | **remote**] [**public-instance** | **vpn-instance** *vpn-instance-name*] [**count**]
Display ARP flood suppression entries.	**display evpn route arp suppression** [**local** | **remote**] [**vsi** *vsi-name*] [**count**]
Display EVPN ARP mobility information.	**display evpn route arp-mobility** [**public-instance** | **vpn-instance** *vpn-instance-name*] [**ip** *ip-address*]
Display EVPN MAC address entries.	**display evpn route mac** [**local** | **remote** | **nexthop** *ipv4-address*] [**vsi** *vsi-name*] [**count**]
Display EVPN MAC mobility information.	**display evpn** [**ipv6**] **route mac-mobility** [**vsi** *vsi-name*] [**mac-address** *mac-address*]
Display EVPN ND entries.	**display evpn route nd** [**local** | **remote**] [**public-instance** | **vpn-instance** *vpn-instance-name*] [**count**]
Display EVPN ND flood suppression entries.	**display evpn route nd suppression** [**local** | **remote**] [**vsi** *vsi-name*] [**count**]
Display EVPN ND mobility information.	**display evpn route nd-mobility** [**public-instance** | **vpn-instance** *vpn-instance-name*] [**ip** *ipv6-address*]
Display the routing table for a VPN instance.	**display evpn routing-table** { **public-instance** | **vpn-instance** *vpn-instance-name* } [**nexthop** *ipv4-address*] [**count**] **display evpn routing-table ipv6** { **public-instance** | **vpn-instance** *vpn-instance-name* } [**nexthop** *ipv6-address*] [**count**]
Cancel ARP mobility event suppression.	**reset evpn route arp-mobility suppression** [**public-instance** | **vpn-instance** *vpn-instance-name* [**ip** *ip-address*]]
Cancel MAC mobility event suppression.	**reset evpn route mac-mobility suppression** [**vsi** *vsi-name* [**mac** *mac-address*]]
Cancel ND mobility event suppression.	**reset evpn route nd-mobility suppression** [**public-instance** | **vpn-instance** *vpn-instance-name* [**ip** *ipv6-address*]]
Display ARP gateway protection configuration for Ethernet service	

instances.	**display arp filter source service-instance** [**interface** *interface-type interface-number* [**service-instance** *instance-id*]] [**slot** *slot-number*]

NOTE:

For more information about the **display bgp group**, **display bgp peer**, and **display bgp update-group** commands, see BGP commands in *Layer 3—IP Routing Command Reference*.

EVPN VXLAN configuration examples

Example: Configuring a centralized EVPN gateway

Network configuration

As shown in Figure 20:

- Configure VXLAN 10 and VXLAN 20 on Switch A, Switch B, and Switch C to provide connectivity for the VMs in the VXLANs across the network sites.
- Configure Switch C as a centralized IPv4 EVPN gateway to provide gateway services and access to the connected Layer 3 network.
- Configure Switch D as an RR to reflect BGP EVPN routes between Switch A, Switch B, and Switch C.

NOTE:

This example provides configuration of IPv4 sites over an IPv4 underlay network. The configuration procedure does not differ between IPv4 and IPv6 sites or underlay networks.

Figure 20 Network diagram

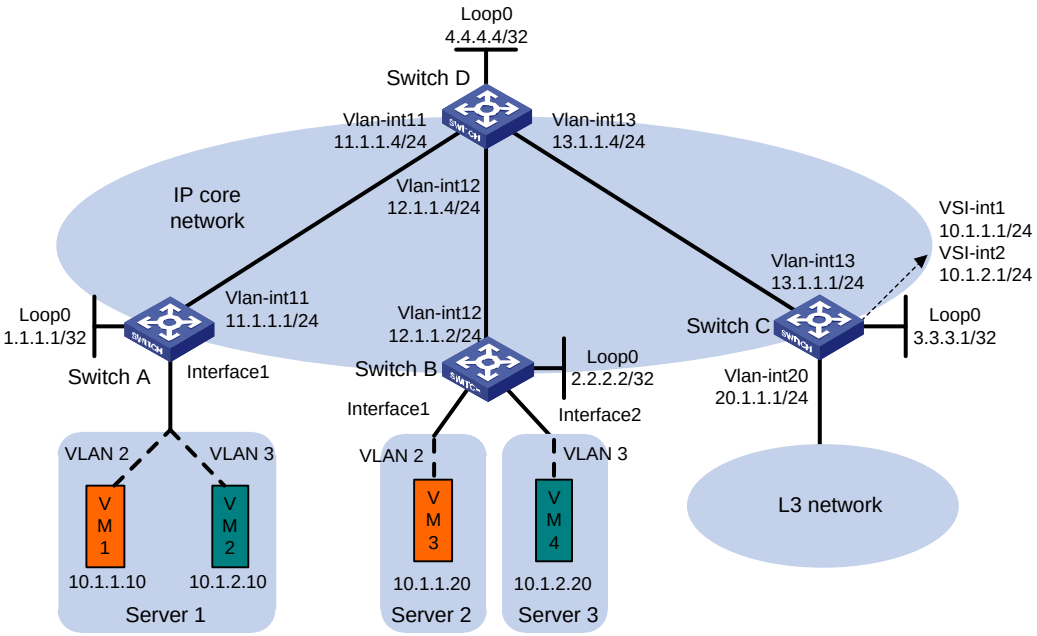


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. On VM 1 and VM 3, specify 10.1.1.1 as the gateway address. On VM 2 and VM 4, specify 10.1.2.1 as the gateway address. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 20. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC address and ARP learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   [SwitchA] vxlan tunnel arp-learning disable
   ```

   # Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] arp suppression enable
   [SwitchA-vsi-vpna] evpn encapsulation vxlan
   [SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto
   [SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
   [SwitchA-vsi-vpna-evpn-vxlan] quit
   ```

   # Create VXLAN 10.

   ```
   [SwitchA-vsi-vpna] vxlan 10
   [SwitchA-vsi-vpna-vxlan-10] quit
   [SwitchA-vsi-vpna] quit
   ```

   # Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

   ```
   [SwitchA] vsi vpnb
   [SwitchA-vsi-vpnb] arp suppression enable
   [SwitchA-vsi-vpnb] evpn encapsulation vxlan
   [SwitchA-vsi-vpnb-evpn-vxlan] route-distinguisher auto
   [SwitchA-vsi-vpnb-evpn-vxlan] vpn-target auto
   [SwitchA-vsi-vpnb-evpn-vxlan] quit
   ```

   # Create VXLAN 20.

   ```
   [SwitchA-vsi-vpnb] vxlan 20
   [SwitchA-vsi-vpnb-vxlan-20] quit
   [SwitchA-vsi-vpnb] quit
   ```

   # Configure BGP to advertise BGP EVPN routes.

   ```
   [SwitchA] bgp 200
   [SwitchA-bgp-default] peer 4.4.4.4 as-number 200
   ```

```
[SwitchA-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchA-GigabitEthernet1/0/1] service-instance 2000
[SwitchA-GigabitEthernet1/0/1-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchA-GigabitEthernet1/0/1-srv2000] xconnect vsi vpnb
[SwitchA-GigabitEthernet1/0/1-srv2000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC address and ARP learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] arp suppression enable
[SwitchB-vsi-vpna] evpn encapsulation vxlan
[SwitchB-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] arp suppression enable
[SwitchB-vsi-vpnb] evpn encapsulation vxlan
[SwitchB-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchB-vsi-vpnb] vxlan 20
```

```
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 4.4.4.4 as-number 200
[SwitchB-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

On GigabitEthernet 1/0/2, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] service-instance 2000
[SwitchB-GigabitEthernet1/0/2-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchB-GigabitEthernet1/0/2-srv2000] xconnect vsi vpnb
[SwitchB-GigabitEthernet1/0/2-srv2000] quit
[SwitchB-GigabitEthernet1/0/2] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address and ARP learning.

```
[SwitchC] vxlan tunnel mac-learning disable
[SwitchC] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] evpn encapsulation vxlan
[SwitchC-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchC-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] evpn encapsulation vxlan
```

```
[SwitchC-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchC-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 4.4.4.4 as-number 200
[SwitchC-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Create VSI-interface 1 and assign the interface an IP address. The IP address will be used as the gateway address for VXLAN 10.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchC-Vsi-interface1] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] gateway vsi-interface 1
[SwitchC-vsi-vpna] quit
```

Create VSI-interface 2 and assign the interface an IP address. The IP address will be used as the gateway address for VXLAN 20.

```
[SwitchC] interface vsi-interface 2
[SwitchC-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchC-Vsi-interface2] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] gateway vsi-interface 2
[SwitchC-vsi-vpnb] quit
```

6. Configure Switch D:

Establish BGP connections with other transport network switches.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] group evpn
[SwitchD-bgp-default] peer 1.1.1.1 group evpn
[SwitchD-bgp-default] peer 2.2.2.2 group evpn
[SwitchD-bgp-default] peer 3.3.3.3 group evpn
[SwitchD-bgp-default] peer evpn as-number 200
[SwitchD-bgp-default] peer evpn connect-interface loopback 0
```

Configure BGP to advertise BGP EVPN routes, and disable route target filtering for BGP EVPN routes.

```
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer evpn enable
[SwitchD-bgp-default-evpn] undo policy vpn-target
```

Configure Switch D as an RR.

```
[SwitchD-bgp-default-evpn] peer evpn reflect-client
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

1. Verify the EVPN gateway settings on Switch C:

Verify that Switch C has advertised MAC/IP advertisement routes and IMET routes for the gateways and received MAC/IP advertisement routes and IMET routes from Switch A and Switch B. (Details not shown.)

Verify that the VXLAN tunnel interfaces are up on Switch C.

```
[SwitchC] display interface tunnel
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 3.3.3.3, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 7 bytes/sec, 56 bits/sec, 0 packets/sec
Input: 10 packets, 980 bytes, 0 drops
Output: 85 packets, 6758 bytes, 0 drops

Tunnel2
Current state: UP
Line protocol state: UP
Description: Tunnel2 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 3.3.3.3, destination 1.1.1.1
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 1 bytes/sec, 8 bits/sec, 0 packets/sec
Last 300 seconds output rate: 9 bytes/sec, 72 bits/sec, 0 packets/sec
Input: 277 packets, 20306 bytes, 0 drops
Output: 1099 packets, 85962 bytes, 0 drops
```

Verify that the VSI interfaces are up on Switch C.

```
[SwitchC] display interface vsi-interface brief
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP        Description
Vsi1                 UP   UP       10.1.1.1
Vsi2                 UP   UP       10.1.2.1
```

Verify that the VXLAN tunnels have been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

```
[SwitchC] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel              0x5000000  UP       Auto        Disabled
    Tunnel2              0x5000001  UP       Auto        Disabled

VSI Name: vpnb
  VSI Index               : 1
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 2
  VXLAN ID                : 20
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel              0x5000000  UP       Auto        Disabled
    Tunnel2              0x5000001  UP       Auto        Disabled
```

Verify that Switch C has created EVPN ARP entries for the VMs.

```
[SwitchC] display evpn route arp
Flags: D - Dynamic   B - BGP      L - Local active
       G - Gateway   S - Static   M - Mapping        I - Invalid
```

```
Public instance                                 Interface: Vsi-interface1
IP address      MAC address     Router MAC      VSI index   Flags
10.1.1.1        0003-0003-0003  -               0           GL
10.1.1.10       0000-1234-0001  -               0           B
10.1.1.20       0000-1234-0003  -               0           B

Public instance                                 Interface: Vsi-interface2
IP address      MAC address     Router MAC      VSI index   Flags
10.1.2.1        0005-0005-0005  -               1           GL
10.1.2.10       0000-1234-0002  -               1           B
10.1.2.20       0000-1234-0004  -               1           B
```

Verify that Switch C has created FIB entries for the VMs.

```
[SwitchC] display fib 10.1.1.10
Destination count: 1 FIB entry count: 1
Flag:
  U:Usable    G:Gateway   H:Host   B:Blackhole   D:Dynamic   S:Static
  R:Relay     F:FRR
Destination/Mask   Nexthop         Flag     OutInterface/Token       Label
10.1.1.10/32       10.1.1.10       UH       Vsi1                     Null
```

2. Verify that VM 1, VM 2, VM 3, and VM 4 can communicate with one another.

Example: Configuring distributed EVPN gateways in symmetric IRB mode (IPv4 underlay network)

Network configuration

As shown in Figure 21:

- Configure VXLAN 10 and VXLAN 20 on Switch A and Switch B to provide connectivity for the VMs in the VXLANs across the network sites.
- Configure Switch A and Switch B as distributed EVPN gateways to provide gateway services in symmetric IRB mode. Configure Switch C as a border gateway to provide access to the connected Layer 3 network.
- Configure Switch D as an RR to reflect BGP EVPN routes between Switch A, Switch B, and Switch C.

NOTE:

This example provides configuration of IPv4 sites. The configuration procedure does not differ between IPv4 and IPv6 sites.

Figure 21 Network diagram

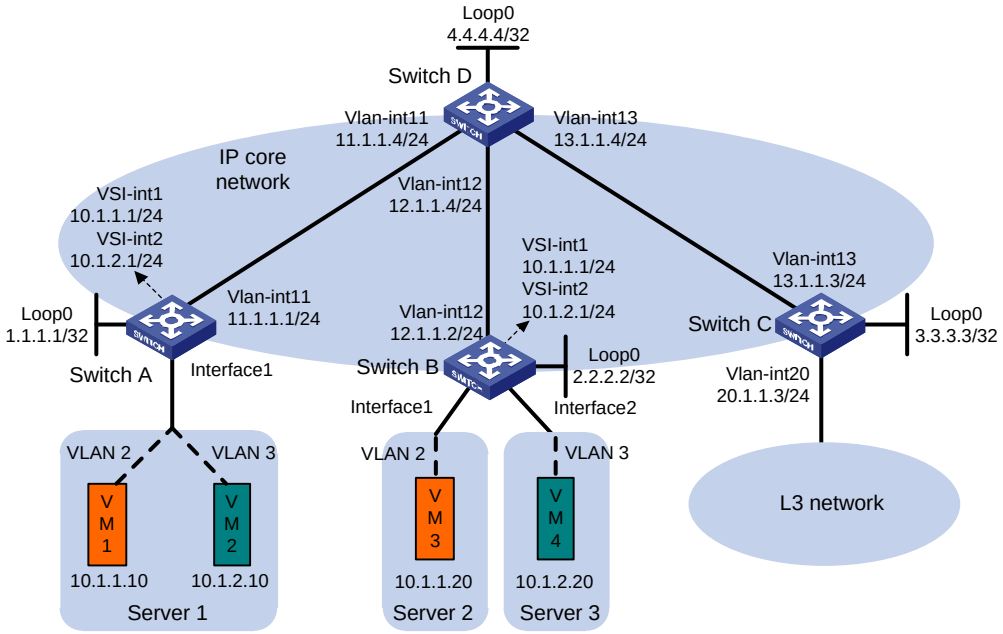


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. On VM 1 and VM 3, specify 10.1.1.1 as the gateway address. On VM 2 and VM 4, specify 10.1.2.1 as the gateway address. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 21. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC address learning and remote ARP learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   [SwitchA] vxlan tunnel arp-learning disable
   ```

   # Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] evpn encapsulation vxlan
   [SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto
   [SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
   [SwitchA-vsi-vpna-evpn-vxlan] quit
   ```

   # Create VXLAN 10.

```
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] evpn encapsulation vxlan
[SwitchA-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchA-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchA-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] peer 4.4.4.4 as-number 200
[SwitchA-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchA-GigabitEthernet1/0/1] service-instance 2000
[SwitchA-GigabitEthernet1/0/1-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchA-GigabitEthernet1/0/1-srv2000] xconnect vsi vpnb
[SwitchA-GigabitEthernet1/0/1-srv2000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchA] ip vpn-instance l3vpna
[SwitchA-vpn-instance-l3vpna] route-distinguisher 1:1
[SwitchA-vpn-instance-l3vpna] address-family ipv4
[SwitchA-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchA-vpn-ipv4-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] address-family evpn
[SwitchA-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchA-vpn-evpn-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchA] interface vsi-interface 1
```

```
[SwitchA-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchA-Vsi-interface1] mac-address 1-1-1
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-arp enable
[SwitchA-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[SwitchA] interface vsi-interface 2
[SwitchA-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchA-Vsi-interface2] mac-address 2-2-2
[SwitchA-Vsi-interface2] distributed-gateway local
[SwitchA-Vsi-interface2] local-proxy-arp enable
[SwitchA-Vsi-interface2] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchA] interface vsi-interface 3
[SwitchA-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface3] l3-vni 1000
[SwitchA-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] gateway vsi-interface 2
[SwitchA-vsi-vpnb] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] evpn encapsulation vxlan
[SwitchB-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] evpn encapsulation vxlan
[SwitchB-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 4.4.4.4 as-number 200
[SwitchB-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

On GigabitEthernet 1/0/2, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] service-instance 2000
[SwitchB-GigabitEthernet1/0/2-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchB-GigabitEthernet1/0/2-srv2000] xconnect vsi vpnb
[SwitchB-GigabitEthernet1/0/2-srv2000] quit
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchB] ip vpn-instance l3vpna
[SwitchB-vpn-instance-l3vpna] route-distinguisher 1:2
[SwitchB-vpn-instance-l3vpna] address-family ipv4
[SwitchB-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchB-vpn-ipv4-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] address-family evpn
[SwitchB-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchB-vpn-evpn-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchB-Vsi-interface1] mac-address 1-1-1
```

```
[SwitchB-Vsi-interface1] distributed-gateway local
[SwitchB-Vsi-interface1] local-proxy-arp enable
[SwitchB-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[SwitchB] interface vsi-interface 2
[SwitchB-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchB-Vsi-interface2] mac-address 2-2-2
[SwitchB-Vsi-interface2] distributed-gateway local
[SwitchB-Vsi-interface2] local-proxy-arp enable
[SwitchB-Vsi-interface2] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchB] interface vsi-interface 3
[SwitchB-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface3] l3-vni 1000
[SwitchB-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
[SwitchB-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] gateway vsi-interface 2
[SwitchB-vsi-vpnb] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchC] vxlan tunnel mac-learning disable
[SwitchC] vxlan tunnel arp-learning disable
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 4.4.4.4 as-number 200
[SwitchC-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchC] ip vpn-instance l3vpna
[SwitchC-vpn-instance-l3vpna] route-distinguisher 1:3
[SwitchC-vpn-instance-l3vpna] address-family ipv4
[SwitchC-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchC-vpn-ipv4-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] address-family evpn
[SwitchC-vpn-evpn-l3vpna] vpn-target 1:1
```

```
[SwitchC-vpn-evpn-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchC] interface vsi-interface 3
[SwitchC-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchC-Vsi-interface3] l3-vni 1000
[SwitchC-Vsi-interface3] quit
```

Configure a default route. The next hop is the IP address of a device in the Layer 3 network.

```
[SwitchC] ip route-static vpn-instance l3vpna 0.0.0.0 0 20.1.1.100
```

Import the default route to the BGP IPv4 unicast routing table of VPN instance **l3vpna**.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] ip vpn-instance l3vpna
[SwitchC-bgp-default-l3vpna] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4-l3vpna] default-route imported
[SwitchC-bgp-default-ipv4-l3vpna] import-route static
[SwitchC-bgp-default-ipv4-l3vpna] quit
[SwitchC-bgp-default-l3vpna] quit
[SwitchC-bgp-default] quit
```

Associate VLAN-interface 20 with VPN instance **l3vpna**.

```
[SwitchC] interface vlan-interface 20
[SwitchC-Vlan-interface20] ip binding vpn-instance l3vpna
[SwitchC-Vlan-interface20] ip address 20.1.1.3 24
[SwitchC-Vlan-interface20] quit
```

6. Configure Switch D:

Establish BGP connections with other transport network switches.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] group evpn
[SwitchD-bgp-default] peer 1.1.1.1 group evpn
[SwitchD-bgp-default] peer 2.2.2.2 group evpn
[SwitchD-bgp-default] peer 3.3.3.3 group evpn
[SwitchD-bgp-default] peer evpn as-number 200
[SwitchD-bgp-default] peer evpn connect-interface loopback 0
```

Configure BGP to advertise BGP EVPN routes, and disable route target filtering for BGP EVPN routes.

```
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer evpn enable
[SwitchD-bgp-default-evpn] undo policy vpn-target
```

Configure Switch D as an RR.

```
[SwitchD-bgp-default-evpn] peer evpn reflect-client
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

1. Verify the distributed EVPN gateway settings on Switch A:

Verify that Switch A has advertised the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI. Verify that Switch A has

received the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI from Switch B. (Details not shown.)

Verify that the VXLAN tunnel interfaces are up on Switch A. (This example uses Tunnel 1.)

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 9 packets, 882 bytes, 0 drops
Output: 9 packets, 882 bytes, 0 drops
```

Verify that the VSI interfaces are up on Switch A. (This example uses VSI-interface 1.)

```
[SwitchA] display interface vsi-interface brief
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP        Description
Vsi1                 UP   UP       10.1.1.1
Vsi2                 UP   UP       10.1.2.1
```

Verify that the VXLAN tunnels have been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: Auto_L3VNI1000_3
  VSI Index               : 1
  VSI State               : Down
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 3
  VXLAN ID                : 1000

VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
```

```
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type      Flood proxy
    Tunnel1              0x5000001  Up     Auto      Disabled
    Tunnel2              0x5000002  Up     Auto      Disabled
  ACs:
    AC                               Link ID    State    Type
    GE1/0/1 srv1000                  0          Up       Manual

VSI Name: vpnb
  VSI Index               : 2
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 2
  VXLAN ID                : 20
  Tunnels:
    Tunnel Name          Link ID    State  Type      Flood proxy
    Tunnel1              0x5000001  Up     Auto      Disabled
    Tunnel2              0x5000002  Up     Auto      Disabled
  ACs:
    AC                               Link ID    State    Type
    GE1/0/1 srv2000                  0          Up       Manual
```

Verify that Switch A has created ARP entries for the VMs. (Details not shown.)

Verify that Switch A has created EVPN ARP entries for the local VMs.

```
[SwitchA] display evpn route arp
Flags: D - Dynamic   B - BGP      L - Local active
```

```
      G - Gateway   S - Static   M - Mapping       I - Invalid

VPN instance: l3vpna                          Interface: Vsi-interface1
IP address      MAC address     Router MAC      VSI Index   Flags
10.1.1.1        0001-0001-0001  a0ce-7e40-0400  0           GL
10.1.1.10       0000-1234-0001  a0ce-7e40-0400  0           DL
10.1.2.10       0000-1234-0002  a0ce-7e40-0400  0           DL
10.1.1.20       0000-1234-0003  a0ce-7e40-0400  0           B
10.1.2.20       0000-1234-0004  a0ce-7e40-0400  0           B
```

2. Verify that VM 1, VM 2, VM 3, and VM 4 can communicate with one another. (Details not shown.)

Example: Configuring distributed EVPN gateways in symmetric IRB mode (IPv6 underlay network)

Network configuration

As shown in Figure 22:

- Configure VXLAN 10 and VXLAN 20 on Switch A and Switch B to provide connectivity for the VMs in the VXLANs across the network sites.
- Configure Switch A and Switch B as distributed EVPN gateways to provide gateway services in symmetric IRB mode. Configure Switch C as a border gateway to provide access to the connected Layer 3 network.
- Configure Switch D as an RR to reflect BGP EVPN routes between Switch A, Switch B, and Switch C.

NOTE:

This example provides configuration of IPv6 sites. The configuration procedure does not differ between IPv4 and IPv6 sites.

Figure 22 Network diagram

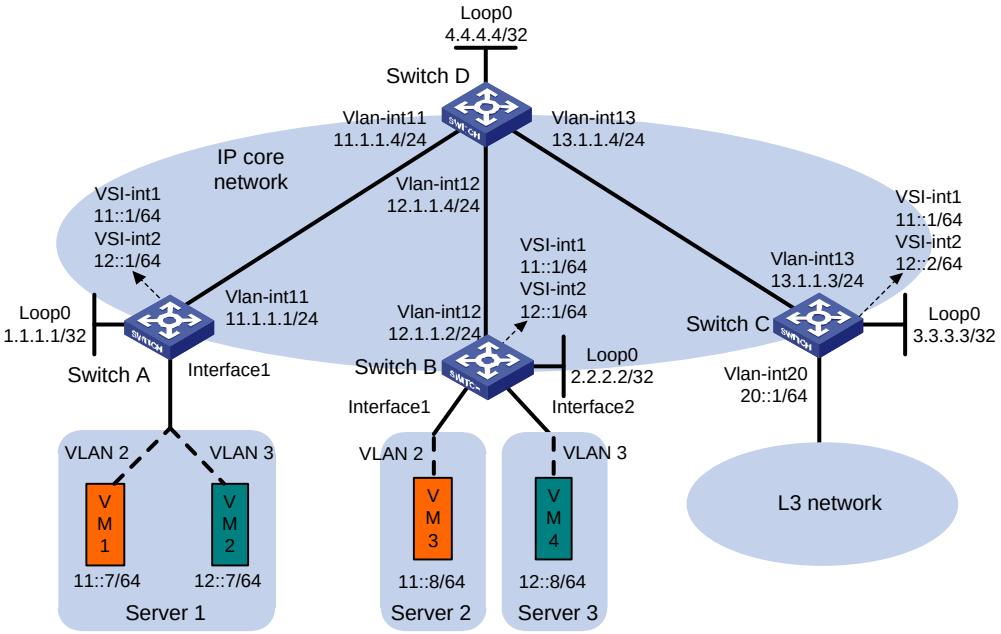


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. On VM 1 and VM 3, specify 11::1 as the gateway address. On VM 2 and VM 4, specify 12::1 as the gateway address. (Details not shown.)
2. Configure IPv6 addresses and unicast routing settings:

   # Assign IPv6 addresses to interfaces, as shown in Figure 22. (Details not shown.)

   # Configure OSPFv3 on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC address learning and remote ND learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   [SwitchA] vxlan tunnel nd-learning disable
   ```

   # Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] evpn encapsulation vxlan
   [SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto
   [SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
   [SwitchA-vsi-vpna-evpn-vxlan] quit
   ```

   # Create VXLAN 10.

```
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] evpn encapsulation vxlan
[SwitchA-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchA-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchA-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] router-id 1.1.1.1
[SwitchA-bgp-default] peer 4::4 as-number 200
[SwitchA-bgp-default] peer 4::4 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 4::4 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchA-GigabitEthernet1/0/1] service-instance 2000
[SwitchA-GigabitEthernet1/0/1-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchA-GigabitEthernet1/0/1-srv2000] xconnect vsi vpnb
[SwitchA-GigabitEthernet1/0/1-srv2000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchA] ip vpn-instance l3vpna
[SwitchA-vpn-instance-l3vpna] route-distinguisher 1:1
[SwitchA-vpn-instance-l3vpna] address-family ipv6
[SwitchA-vpn-ipv6-l3vpna] vpn-target 2:2
[SwitchA-vpn-ipv6-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] address-family evpn
[SwitchA-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchA-vpn-evpn-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface1] ipv6 address 11::1 64
[SwitchA-Vsi-interface1] mac-address 1-1-1
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-nd enable
[SwitchA-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[SwitchA] interface vsi-interface 2
[SwitchA-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface2] ipv6 address 12::1 64
[SwitchA-Vsi-interface2] mac-address 2-2-2
[SwitchA-Vsi-interface2] distributed-gateway local
[SwitchA-Vsi-interface2] local-proxy-nd enable
[SwitchA-Vsi-interface2] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchA] interface vsi-interface 3
[SwitchA-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface3] ipv6 address auto link-local
[SwitchA-Vsi-interface3] l3-vni 1000
[SwitchA-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] gateway vsi-interface 2
[SwitchA-vsi-vpnb] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC address learning and remote ND learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel nd-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] evpn encapsulation vxlan
[SwitchB-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] evpn encapsulation vxlan
[SwitchB-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] router-id 2.2.2.2
[SwitchB-bgp-default] peer 4::4 as-number 200
[SwitchB-bgp-default] peer 4::4 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 4::4 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

On GigabitEthernet 1/0/2, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] service-instance 2000
[SwitchB-GigabitEthernet1/0/2-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchB-GigabitEthernet1/0/2-srv2000] xconnect vsi vpnb
[SwitchB-GigabitEthernet1/0/2-srv2000] quit
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchB] ip vpn-instance l3vpna
[SwitchB-vpn-instance-l3vpna] route-distinguisher 1:2
[SwitchB-vpn-instance-l3vpna] address-family ipv6
[SwitchB-vpn-ipv6-l3vpna] vpn-target 2:2
[SwitchB-vpn-ipv6-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] address-family evpn
[SwitchB-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchB-vpn-evpn-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchB] interface vsi-interface 1
```

```
[SwitchB-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface1] ipv6 address 11::1 64
[SwitchB-Vsi-interface1] mac-address 1-1-1
[SwitchB-Vsi-interface1] distributed-gateway local
[SwitchB-Vsi-interface1] local-proxy-nd enable
[SwitchB-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[SwitchB] interface vsi-interface 2
[SwitchB-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface2] ipv6 address 12::1 64
[SwitchB-Vsi-interface2] mac-address 2-2-2
[SwitchB-Vsi-interface2] distributed-gateway local
[SwitchB-Vsi-interface2] local-proxy-nd enable
[SwitchB-Vsi-interface2] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchB] interface vsi-interface 3
[SwitchB-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface3] ipv6 address auto link-local
[SwitchB-Vsi-interface3] l3-vni 1000
[SwitchB-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
[SwitchB-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] gateway vsi-interface 2
[SwitchB-vsi-vpnb] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address learning and remote ND learning.

```
[SwitchC] vxlan tunnel mac-learning disable
[SwitchC] vxlan tunnel nd-learning disable
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] router-id 3.3.3.3
[SwitchC-bgp-default] peer 4::4 as-number 200
[SwitchC-bgp-default] peer 4::4 connect-interface loopback 0
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer 4::4 enable
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchC] ip vpn-instance l3vpna
[SwitchC-vpn-instance-l3vpna] route-distinguisher 1:3
```

```
[SwitchC-vpn-instance-l3vpna] address-family ipv6
[SwitchC-vpn-ipv6-l3vpna] vpn-target 2:2
[SwitchC-vpn-ipv6-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] address-family evpn
[SwitchC-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchC-vpn-evpn-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchC] interface vsi-interface 3
[SwitchC-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchC-Vsi-interface3] ipv6 address auto link-local
[SwitchC-Vsi-interface3] l3-vni 1000
[SwitchC-Vsi-interface3] quit
```

Configure a default route. The next hop is the IP address of a device in the Layer 3 network.

```
[SwitchC] ipv6 route-static vpn-instance l3vpna :: 0 20::100
```

Import the default route to the BGP IPv6 unicast routing table of VPN instance **l3vpna**.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] ip vpn-instance l3vpna
[SwitchC-bgp-default-l3vpna] address-family ipv6 unicast
[SwitchC-bgp-default-ipv6-l3vpna] default-route imported
[SwitchC-bgp-default-ipv6-l3vpna] import-route static
[SwitchC-bgp-default-ipv6-l3vpna] quit
[SwitchC-bgp-default-l3vpna] quit
[SwitchC-bgp-default] quit
```

Associate VLAN-interface 20 with VPN instance **l3vpna** and assign an IPv6 address to the interface.

```
[SwitchC] interface vlan-interface 20
[SwitchC-Vlan-interface20] ip binding vpn-instance l3vpna
[SwitchC-Vlan-interface20] ipv6 address 20::1 64
[SwitchC-Vlan-interface20] quit
```

6. Configure Switch D:

Establish BGP connections with other transport network switches.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] router-id 4.4.4.4
[SwitchD-bgp-default] group evpn
[SwitchD-bgp-default] peer 1::1 group evpn
[SwitchD-bgp-default] peer 2::2 group evpn
[SwitchD-bgp-default] peer 3::3 group evpn
[SwitchD-bgp-default] peer evpn as-number 200
[SwitchD-bgp-default] peer evpn connect-interface loopback 0
```

Configure BGP to advertise BGP EVPN routes, and disable route target filtering for BGP EVPN routes.

```
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer evpn enable
[SwitchD-bgp-default-evpn] undo policy vpn-target
```

Configure Switch D as an RR.

```
[SwitchD-bgp-default-evpn] peer evpn reflect-client
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

1. Verify the distributed EVPN gateway settings on Switch A:

   # Verify that Switch A has advertised the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI. Verify that Switch A has received the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI from Switch B. (Details not shown.)

   # Verify that the VXLAN tunnel interfaces are up on Switch A. (This example uses Tunnel 1.)

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1::1, destination 2::2
Tunnel protocol/transport UDP_VXLAN/IPv6
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

   # Verify that the VSI interfaces are up on Switch A. (This example uses VSI-interface 1.)

```
[SwitchA] display ipv6 interface vsi-interface brief
*down: administratively down
(s): spoofing
Interface                                Physical Protocol IPv6 Address
Vsi1                                     UP       UP       11::1
Vsi2                                     UP       UP       12::1
```

   # Verify that the VXLAN tunnels have been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: Auto_L3VNI1000_3
  VSI Index               : 1
  VSI State               : Down
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
```

```
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 3
  VXLAN ID                : 1000

VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type      Flood proxy
    Tunnel1              0x5000001  Up     Auto      Disabled
    Tunnel2              0x5000002  Up     Auto      Disabled
  ACs:
    AC                               Link ID    State    Type
    GE1/0/1 srv1000                  0          Up       Manual

VSI Name: vpnb
  VSI Index               : 2
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : Unlimited
  Broadcast Restrain      : Unlimited
  Multicast Restrain      : Unlimited
  Unknown Unicast Restrain: Unlimited
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 2
  VXLAN ID                : 20
  Tunnels:
    Tunnel Name          Link ID    State  Type      Flood proxy
    Tunnel1              0x5000001  Up     Auto      Disabled
    Tunnel2              0x5000002  Up     Auto      Disabled
```

```
  ACs:
    AC                              Link ID    State    Type
    GE1/0/1 srv2000                 0          Up       Manual
```

Verify that Switch A has created EVPN ND entries for the local VMs.

```
[SwitchA] display evpn route nd
Flags: D - Dynamic   B - BGP      L - Local active
       G - Gateway   S - Static   M - Mapping        I - Invalid

VPN instance: l3vpna                          Interface: Vsi-interface1
IPv6 address :   11::1
MAC address  :   0001-0001-0001       Router MAC   :   06dc-93de-0100
VSI index    :   0                    Flags        :   GL

IPv6 address :   11::7
MAC address  :   06dc-98ca-0206       Router MAC   :   06dc-93de-0100
VSI index    :   0                    Flags        :   DL

IPv6 address :   11::8
MAC address  :   06dc-a8dd-0506       Router MAC   :   06dc-a235-0400
VSI index    :   0                    Flags        :   B

VPN instance: l3vpna                          Interface: Vsi-interface2
IPv6 address :   12::1
MAC address  :   0002-0002-0002       Router MAC   :   06dc-93de-0100
VSI index    :   1                    Flags        :   GL

IPv6 address :   12::7
MAC address  :   06dc-9ca0-0306       Router MAC   :   06dc-93de-0100
VSI index    :   1                    Flags        :   DL

IPv6 address :   12::8
MAC address  :   06dc-ad91-0606       Router MAC   :   06dc-a235-0400
VSI index    :   1                    Flags        :   B
```

2. Verify that VM 1, VM 2, VM 3, and VM 4 can communicate with one another. (Details not shown.)

Example: Configuring distributed IPv4 EVPN gateways in asymmetric IRB mode

Network configuration

As shown in Figure 23:

- Configure VXLAN 10 and VXLAN 20 on Switch A and Switch B to provide connectivity for the VMs in the VXLANs across the network sites.
- Configure Switch A and Switch B as distributed EVPN gateways to provide gateway services in asymmetric IRB mode. Configure Switch C as a border gateway to provide access to the connected Layer 3 network.
- Configure Switch D as an RR to reflect BGP EVPN routes between Switch A, Switch B, and Switch C.

Figure 23 Network diagram

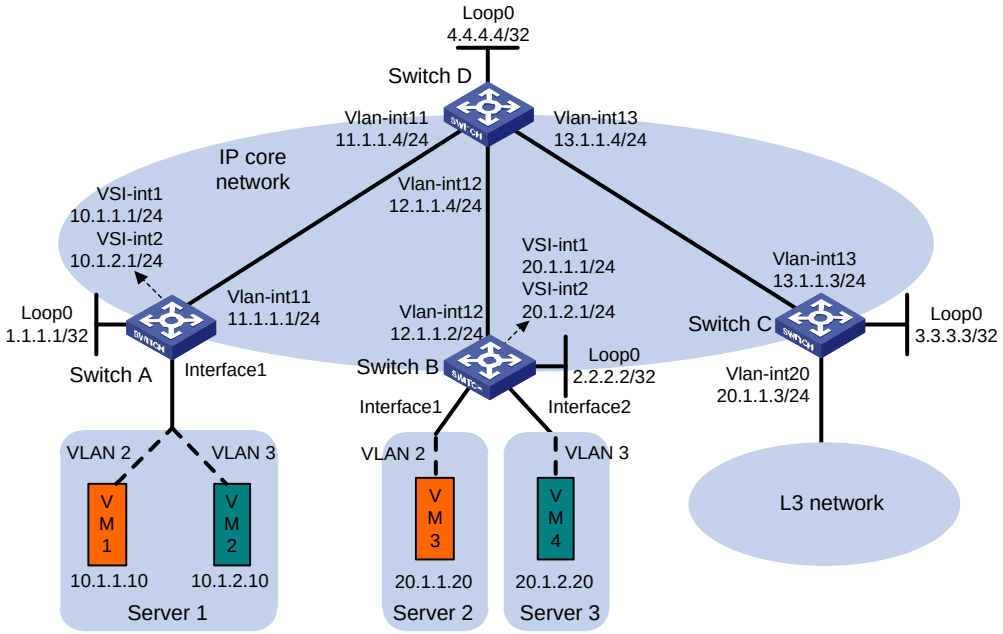


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. Specify 10.1.1.1, 10.1.2.1, 20.1.1.1, and 20.1.2.1 as the gateway addresses on VM 1, VM 2, VM 3, and VM 4, respectively. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 23. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC address learning and remote ARP learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   [SwitchA] vxlan tunnel arp-learning disable
   ```

   # Enable asymmetric IRB mode for EVPN VXLAN.

   ```
   [SwitchA] evpn irb asymmetric
   ```

   # Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] evpn encapsulation vxlan
   [SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto
   [SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
   ```

```
[SwitchA-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] evpn encapsulation vxlan
[SwitchA-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchA-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchA-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchA-vsi-vpnb] vxlan 20
[SwitchA-vsi-vpnb-vxlan-20] quit
[SwitchA-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] peer 4.4.4.4 as-number 200
[SwitchA-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchA-GigabitEthernet1/0/1] service-instance 2000
[SwitchA-GigabitEthernet1/0/1-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchA-GigabitEthernet1/0/1-srv2000] xconnect vsi vpnb
[SwitchA-GigabitEthernet1/0/1-srv2000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchA] ip vpn-instance l3vpna
[SwitchA-vpn-instance-l3vpna] route-distinguisher 1:1
[SwitchA-vpn-instance-l3vpna] address-family ipv4
[SwitchA-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchA-vpn-ipv4-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] address-family evpn
[SwitchA-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchA-vpn-evpn-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchA-Vsi-interface1] mac-address 1-1-1
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-arp enable
[SwitchA-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[SwitchA] interface vsi-interface 2
[SwitchA-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchA-Vsi-interface2] mac-address 2-2-2
[SwitchA-Vsi-interface2] distributed-gateway local
[SwitchA-Vsi-interface2] local-proxy-arp enable
[SwitchA-Vsi-interface2] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchA] interface vsi-interface 3
[SwitchA-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface3] l3-vni 1000
[SwitchA-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchA] vsi vpnb
[SwitchA-vsi-vpnb] gateway vsi-interface 2
[SwitchA-vsi-vpnb] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel arp-learning disable
```

Enable asymmetric IRB mode for EVPN VXLAN.

```
[SwitchA] evpn irb asymmetric
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] evpn encapsulation vxlan
[SwitchB-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchB-vsi-vpna] vxlan 10
```

```
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] evpn encapsulation vxlan
[SwitchB-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 4.4.4.4 as-number 200
[SwitchB-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

On GigabitEthernet 1/0/2, create Ethernet service instance 2000 to match VLAN 3.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] service-instance 2000
[SwitchB-GigabitEthernet1/0/2-srv2000] encapsulation s-vid 3
```

Map Ethernet service instance 2000 to VSI **vpnb**.

```
[SwitchB-GigabitEthernet1/0/2-srv2000] xconnect vsi vpnb
[SwitchB-GigabitEthernet1/0/2-srv2000] quit
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchB] ip vpn-instance l3vpna
[SwitchB-vpn-instance-l3vpna] route-distinguisher 1:2
[SwitchB-vpn-instance-l3vpna] address-family ipv4
[SwitchB-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchB-vpn-ipv4-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] address-family evpn
[SwitchB-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchB-vpn-evpn-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface1] ip address 20.1.1.1 255.255.255.0
[SwitchB-Vsi-interface1] mac-address 1-1-1
[SwitchB-Vsi-interface1] distributed-gateway local
[SwitchB-Vsi-interface1] local-proxy-arp enable
[SwitchB-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[SwitchB] interface vsi-interface 2
[SwitchB-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface2] ip address 20.1.2.1 255.255.255.0
[SwitchB-Vsi-interface2] mac-address 2-2-2
[SwitchB-Vsi-interface2] distributed-gateway local
[SwitchB-Vsi-interface2] local-proxy-arp enable
[SwitchB-Vsi-interface2] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchB] interface vsi-interface 3
[SwitchB-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface3] l3-vni 1000
[SwitchB-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
[SwitchB-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] gateway vsi-interface 2
[SwitchB-vsi-vpnb] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchC] vxlan tunnel mac-learning disable
[SwitchC] vxlan tunnel arp-learning disable
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 4.4.4.4 as-number 200
[SwitchC-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchC] ip vpn-instance l3vpna
[SwitchC-vpn-instance-l3vpna] route-distinguisher 1:3
[SwitchC-vpn-instance-l3vpna] address-family ipv4
```

```
[SwitchC-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchC-vpn-ipv4-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] address-family evpn
[SwitchC-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchC-vpn-evpn-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchC] interface vsi-interface 3
[SwitchC-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchC-Vsi-interface3] l3-vni 1000
[SwitchC-Vsi-interface3] quit
```

Configure a default route. The next hop is the IP address of a device in the Layer 3 network.

```
[SwitchC] ip route-static vpn-instance l3vpna 0.0.0.0 0 20.1.1.100
```

Import the default route to the BGP IPv4 unicast routing table of VPN instance **l3vpna**.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] ip vpn-instance l3vpna
[SwitchC-bgp-default-l3vpna] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4-l3vpna] default-route imported
[SwitchC-bgp-default-ipv4-l3vpna] import-route static
[SwitchC-bgp-default-ipv4-l3vpna] quit
[SwitchC-bgp-default-l3vpna] quit
[SwitchC-bgp-default] quit
```

Associate VLAN-interface 20 with VPN instance **l3vpna**. VLAN-interface 20 provides access to the Layer 3 network connected to Switch C.

```
[SwitchC] interface vlan-interface 20
[SwitchC-Vlan-interface20] ip binding vpn-instance l3vpna
[SwitchC-Vlan-interface20] ip address 20.1.1.3 24
[SwitchC-Vlan-interface20] quit
```

6. Configure Switch D:

Establish BGP connections with other transport network switches.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] group evpn
[SwitchD-bgp-default] peer 1.1.1.1 group evpn
[SwitchD-bgp-default] peer 2.2.2.2 group evpn
[SwitchD-bgp-default] peer 3.3.3.3 group evpn
[SwitchD-bgp-default] peer evpn as-number 200
[SwitchD-bgp-default] peer evpn connect-interface loopback 0
```

Configure BGP to advertise BGP EVPN routes, and disable route target filtering for BGP EVPN routes.

```
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer evpn enable
[SwitchD-bgp-default-evpn] undo policy vpn-target
```

Configure Switch D as an RR.

```
[SwitchD-bgp-default-evpn] peer evpn reflect-client
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

1. Verify the distributed EVPN gateway settings on Switch A:

   # Verify that Switch A has advertised the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI. Verify that Switch A has received the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI from Switch B. (Details not shown.)

   # Verify that the VXLAN tunnel interfaces are up on Switch A. (This example uses Tunnel 1.)

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

   # Verify that the VSI interfaces are up on Switch A. (This example uses VSI-interface 1.)

```
[SwitchA] display interface vsi-interface 1
Vsi-interface1
Current state: UP
Line protocol state: UP
Description: Vsi-interface1 Interface
Bandwidth: 1000000 kbps
Maximum transmission unit: 1500
Internet address: 10.1.1.1/24 (primary)
IP packet frame type: Ethernet II, hardware address: 0003-0003-0003
IPv6 packet frame type: Ethernet II, hardware address: 0003-0003-0003
Physical: Unknown, baudrate: 1000000 kbps
Last clearing of counters: Never
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

   # Verify that the VXLAN tunnels have been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: Auto_L3VNI1000_3
  VSI Index               : 1
  VSI State               : Down
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
```

```
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 3
  VXLAN ID                : 1000
  Tunnel Statistics       : Disabled

VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type      Flood proxy
    Tunnel1              0x5000000  UP     Auto      Disabled
  ACs:
    AC                               Link ID    State    Type
    GE1/0/1 srv1000                  0x0        Up       Manual


VSI Name: vpnb
  VSI Index               : 2
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
```

```
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 2
  VXLAN ID                : 20
  Tunnels:
    Tunnel Name           Link ID    State  Type      Flood proxy
    Tunnel1               0x5000000  UP     Auto      Disabled
  ACs:
    AC                                Link ID    State    Type
    GE1/0/1 srv2000                   0x0        Up       Manual
```

Verify that Switch A has created ARP entries for the VMs. (Details not shown.)

Verify that Switch A has created EVPN ARP entries for the VMs.

```
[SwitchA] display evpn route arp
Flags: D - Dynamic   B - BGP      L - Local active
       G - Gateway   S - Static   M - Mapping        I - Invalid
       E - Multihoming ES sync    F - Leaf

VPN instance: l3vpna                          Interface: Vsi-interface1
IP address      MAC address     Router MAC      VSI index   Flags
10.1.1.1        0001-0001-0001  522b-3413-0200  0           GL
10.1.1.10       521f-b814-0106  522b-3413-0200  0           DL
20.1.1.20       522b-3c6a-0406  522b-38cd-0300  0           B
```

2. Verify that VM 1, VM 2, VM 3, and VM 4 can communicate with one another. (Details not shown.)

Example: Configuring communication between EVPN networks and the public network

Network configuration

As shown in Figure 24:

- Configure VXLAN 10, VXLAN 20, and VXLAN 30 on Switch A, Switch B, and Switch C to meet the following requirements:
  - VXLAN 10 and VXLAN 20 are on the private network, and VXLAN 30 is on the public network.
  - VXLAN 10 can communicate with VXLAN 20 and VXLAN 30, and VXLAN 20 is isolated from VXLAN 30.
- Configure Switch A, Switch B, and Switch C as distributed EVPN gateways to provide gateway services for the VXLANs.
- Configure Switch D as an RR to reflect BGP EVPN routes between Switch A, Switch B, and Switch C.

NOTE:

This example provides configuration of IPv4 sites over an IPv4 underlay network. The configuration procedure does not differ between IPv4 and IPv6 sites or underlay networks.

Figure 24 Network diagram

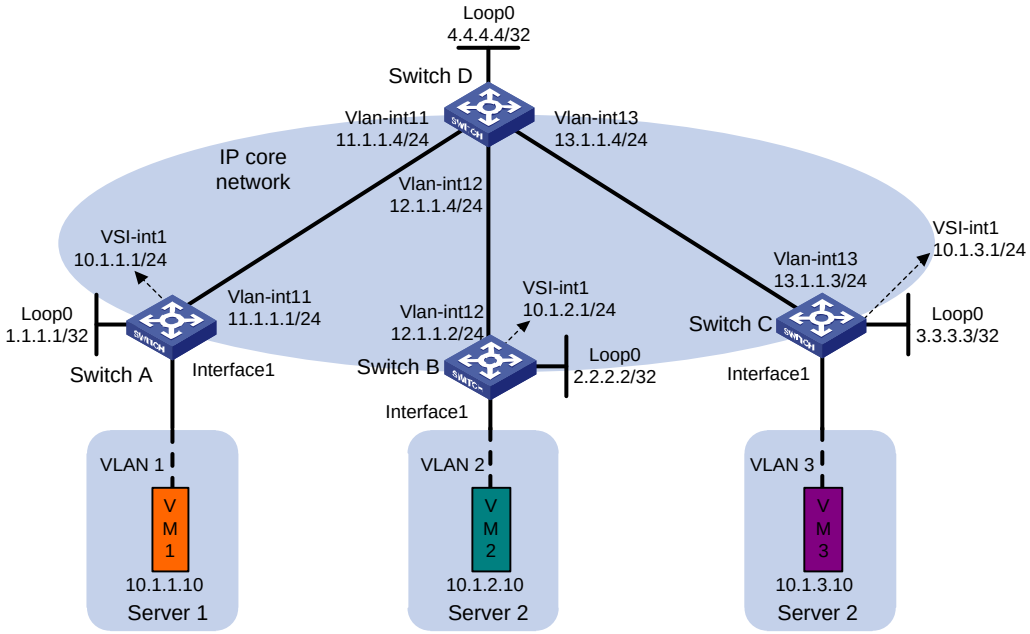


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1

Procedure

1. On VM 1, VM 2, and VM 3, specify 10.1.1.1, 10.1.2.1, and 10.1.3.1 as the gateway address, respectively. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces, as shown in Figure 24. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC address learning and remote ARP learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   [SwitchA] vxlan tunnel arp-learning disable
   ```

   # Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

   ```
   [SwitchA] vsi vpna
   [SwitchA-vsi-vpna] evpn encapsulation vxlan
   [SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto
   [SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
   [SwitchA-vsi-vpna-evpn-vxlan] quit
   ```

   # Create VXLAN 10.

   ```
   [SwitchA-vsi-vpna] vxlan 10
   ```

```
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] peer 4.4.4.4 as-number 200
[SwitchA-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 1.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 1
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchA] ip vpn-instance l3vpna
[SwitchA-vpn-instance-l3vpna] route-distinguisher 1:1
[SwitchA-vpn-instance-l3vpna] address-family ipv4
[SwitchA-vpn-ipv4-l3vpna] vpn-target 1:1
[SwitchA-vpn-ipv4-l3vpna] vpn-target 2:2 import-extcommunity
[SwitchA-vpn-ipv4-l3vpna] vpn-target 3:3 import-extcommunity
[SwitchA-vpn-ipv4-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] address-family evpn
[SwitchA-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchA-vpn-evpn-l3vpna] vpn-target 2:2 import-extcommunity
[SwitchA-vpn-evpn-l3vpna] vpn-target 3:3 import-extcommunity
[SwitchA-vpn-evpn-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-arp enable
[SwitchA-Vsi-interface1] quit
```

Associate VSI-interface 2 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchA] interface vsi-interface 2
[SwitchA-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface2] l3-vni 1000
[SwitchA-Vsi-interface2] quit
```

Create VSI-interface 3 and configure its L3 VXLAN ID as 2000 for matching routes from Switch B.

```
[SwitchA] interface vsi-interface 3
[SwitchA-Vsi-interface3] l3-vni 2000
```

```
[SwitchA-Vsi-interface3] quit
```

Create VSI-interface 4 and configure its L3 VXLAN ID as 3000 for matching routes from Switch C.

```
[SwitchA] interface vsi-interface 4
[SwitchA-Vsi-interface4] l3-vni 3000
[SwitchA-Vsi-interface4] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] evpn encapsulation vxlan
[SwitchB-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchB-vsi-vpnb] vxlan 20
[SwitchB-vsi-vpnb-vxlan-20] quit
[SwitchB-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 4.4.4.4 as-number 200
[SwitchB-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpnb**.

```
[SwitchB-GigabitEthernet1/0/1-srv1000] xconnect vsi vpnb
[SwitchB-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure RD and route target settings for VPN instance **l3vpnb**.

```
[SwitchB] ip vpn-instance l3vpnb
[SwitchB-vpn-instance-l3vpnb] route-distinguisher 2:2
[SwitchB-vpn-instance-l3vpnb] address-family ipv4
```

```
[SwitchB-vpn-ipv4-l3vpnb] vpn-target 2:2
[SwitchB-vpn-ipv4-l3vpnb] vpn-target 1:1 import-extcommunity
[SwitchB-vpn-ipv4-l3vpnb] quit
[SwitchB-vpn-instance-l3vpnb] address-family evpn
[SwitchB-vpn-evpn-l3vpnb] vpn-target 2:2
[SwitchB-vpn-evpn-l3vpnb] vpn-target 1:1 import-extcommunity
[SwitchB-vpn-evpn-l3vpnb] quit
[SwitchB-vpn-instance-l3vpnb] quit
```

Configure VSI-interface 1.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ip binding vpn-instance l3vpnb
[SwitchB-Vsi-interface1] ip address 10.1.2.1 255.255.255.0
[SwitchB-Vsi-interface1] distributed-gateway local
[SwitchB-Vsi-interface1] local-proxy-arp enable
[SwitchB-Vsi-interface1] quit
```

Create VSI-interface 2, and configure its L3 VXLAN ID as 1000 for matching routes from Switch A.

```
[SwitchB] interface vsi-interface 2
[SwitchB-Vsi-interface2] l3-vni 1000
[SwitchB-Vsi-interface2] qui
```

Associate VSI-interface 3 with VPN instance **l3vpnb**, and configure the L3 VXLAN ID as 2000 for the VPN instance.

```
[SwitchB] interface vsi-interface 3
[SwitchB-Vsi-interface3] ip binding vpn-instance l3vpnb
[SwitchB-Vsi-interface3] l3-vni 2000
[SwitchB-Vsi-interface3] quit
```

Create VSI-interface 4, and configure its L3 VXLAN ID as 3000 for matching routes from Switch C.

```
[SwitchB] interface vsi-interface 4
[SwitchB-Vsi-interface4] l3-vni 3000
[SwitchB-Vsi-interface4] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpnb**.

```
[SwitchB] vsi vpnb
[SwitchB-vsi-vpnb] gateway vsi-interface 1
[SwitchB-vsi-vpnb] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchC] vxlan tunnel mac-learning disable
[SwitchC] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpnc**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpnc
[SwitchC-vsi-vpnc] evpn encapsulation vxlan
[SwitchC-vsi-vpnc-evpn-vxlan] route-distinguisher auto
[SwitchC-vsi-vpnc-evpn-vxlan] vpn-target auto
```

```
[SwitchC-vsi-vpnc-evpn-vxlan] quit
```

Create VXLAN 30.

```
[SwitchC-vsi-vpnc] vxlan 30
[SwitchC-vsi-vpnc-vxlan-30] quit
[SwitchC-vsi-vpnc] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 4.4.4.4 as-number 200
[SwitchC-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchC-bgp-default] address-family ipv4 unicast
[SwitchC-bgp-default-ipv4] quit
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Configure RD, route target, and L3 VXLAN ID settings for the public instance.

```
[SwitchC] ip public-instance
[SwitchC-public-instance] route-distinguisher 3:3
[SwitchC-public-instance] l3-vni 3000
[SwitchC-public-instance] address-family ipv4
[SwitchC-public-instance-ipv4] vpn-target 3:3
[SwitchC-public-instance-ipv4] vpn-target 1:1 import-extcommunity
[SwitchC-public-instance-ipv4] quit
[SwitchC-public-instance] address-family evpn
[SwitchC-public-instance-evpn]vpn-target 3:3
[SwitchC-public-instance-evpn] vpn-target 1:1 import-extcommunity
[SwitchC-public-instance-evpn] quit
[SwitchC-public-instance] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 3.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 3
```

Map Ethernet service instance 1000 to VSI **vpnc**.

```
[SwitchC-GigabitEthernet1/0/1-srv1000] xconnect vsi vpnc
[SwitchC-GigabitEthernet1/0/1-srv1000] quit
[SwitchC-GigabitEthernet1/0/1] quit
```

Configure VSI-interface 1.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip address 10.1.3.1 255.255.255.0
[SwitchC-Vsi-interface1] distributed-gateway local
[SwitchC-Vsi-interface1] local-proxy-arp enable
[SwitchC-Vsi-interface1] quit
```

Create VSI-interface 2, and configure its L3 VXLAN ID as 1000 for matching routes from Switch A.

```
[SwitchC] interface vsi-interface 2
[SwitchC-Vsi-interface2] l3-vni 1000
[SwitchC-Vsi-interface2] quit
```

Create VSI-interface 3, and configure its L3 VXLAN ID as 2000 for matching routes from Switch B.

```
[SwitchC] interface vsi-interface 3
[SwitchC-Vsi-interface3] l3-vni 2000
[SwitchC-Vsi-interface3] quit
```

Create VSI-interface 4 for the public instance, and configure the L3 VXLAN ID as 3000 for the VSI interface.

```
[SwitchC] interface vsi-interface 4
[SwitchC-Vsi-interface4] l3-vni 3000
[SwitchC-Vsi-interface4] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpnc**.

```
[SwitchC] vsi vpnc
[SwitchC-vsi-vpnc] gateway vsi-interface 1
[SwitchC-vsi-vpnc] quit
```

6. Configure Switch D:

Establish BGP connections with other transport network switches.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] group evpn
[SwitchD-bgp-default] peer 1.1.1.1 group evpn
[SwitchD-bgp-default] peer 2.2.2.2 group evpn
[SwitchD-bgp-default] peer 3.3.3.3 group evpn
[SwitchD-bgp-default] peer evpn as-number 200
[SwitchD-bgp-default] peer evpn connect-interface loopback 0
```

Configure BGP to advertise BGP EVPN routes, and disable route target filtering for BGP EVPN routes.

```
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer evpn enable
[SwitchD-bgp-default-evpn] undo policy vpn-target
```

Configure Switch D as an RR.

```
[SwitchD-bgp-default-evpn] peer evpn reflect-client
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

1. Verify the distributed EVPN gateway settings on Switch A:

Verify that Switch A has advertised the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI. Verify that Switch A has received the IP prefix advertisement routes for the gateways and the MAC/IP advertisement routes and IMET routes for each VSI from Switch B and Switch C. (Details not shown.)

Verify that the VXLAN tunnel interfaces are up on Switch A.

```
[SwitchA] display interface tunnel
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
```

```
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 15 packets, 1470 bytes, 0 drops
Output: 15 packets, 1470 bytes, 0 drops

Tunnel2
Current state: UP
Line protocol state: UP
Description: Tunnel2 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 3.3.3.3
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 22 packets, 2156 bytes, 0 drops
Output: 23 packets, 2254 bytes, 0 drops
```

Verify that the VSI interfaces are up on Switch A.

```
[SwitchA] display interface vsi-interface brief
Brief information on interfaces in route mode:
Link: ADM - administratively down; Stby - standby
Protocol: (s) - spoofing
Interface            Link Protocol Primary IP        Description
Vsi1                 UP   UP       10.1.1.1
Vsi2                 UP   UP       --
Vsi3                 UP   UP       --
Vsi4                 UP   UP       --
```

Verify that the VXLAN tunnels have been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: Auto_L3VNI1000_2
  VSI Index               : 1
  VSI State               : Down
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
```

```
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 2
  VXLAN ID                : 1000

VSI Name: Auto_L3VNI2000_3
  VSI Index               : 2
  VSI State               : Down
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 3
  VXLAN ID                : 2000

VSI Name: Auto_L3VNI3000_4
  VSI Index               : 3
  VSI State               : Down
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 4
  VXLAN ID                : 3000

VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
```

```
MAC Learning rate       : -
Drop Unknown            : -
Flooding                : Enabled
Statistics              : Disabled
Gateway Interface       : VSI-interface 1
VXLAN ID                : 10
ACs:
  AC                              Link ID  State       Type
  GE1/0/1 srv1000                 0        Up          Manual
```

Verify that Switch A has created ARP entries for the VMs. (Details not shown.)

2. Verify that VM 1 can communicate with VM 2 and VM 3, and VM 2 cannot communicate with VM 3. (Details not shown.)

Example: Configuring IPv4 EVPN M-LAG with a direct peer link

Network configuration

As shown in Figure 25, perform the following tasks to make sure the VMs can communicate with one another:

- Configure VXLAN 10 on Switch A and Switch B, and configure VXLAN 20 on Switch D.
- Configure EVPN M-LAG on Switch A and Switch B to virtualize them into one VTEP. The switches use a direct peer link.
- Configure Switch C as a centralized EVPN gateway and RR.

NOTE:

This example provides configuration of IPv4 sites extended by an IPv4 underlay network. The configuration procedure does not differ between site or underlay network types.

Figure 25 Network diagram

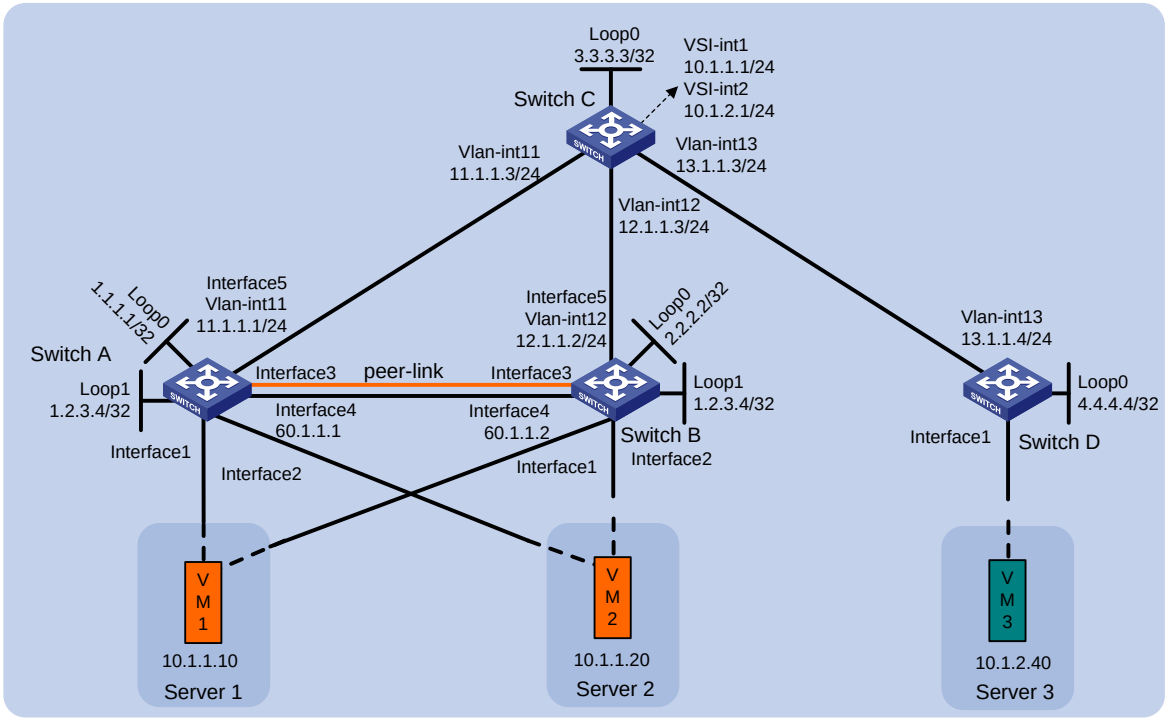


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

1. On VM 1 and VM 2, specify 10.1.1.1 as the gateway address. On VM 3, specify 10.1.2.1 as the gateway address. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces (including loopback interfaces), as shown in Figure 25. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)

   # Execute the **`ospf peer hold-max-cost duration`** command on the interfaces used for setting up OSPF neighbor relationships. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC learning and remote ARP learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   ```

```
[SwitchA] vxlan tunnel arp-learning disable
```
Enable EVPN M-LAG, and specify the virtual VTEP address as 1.2.3.4.
```
[SwitchA] evpn m-lag group 1.2.3.4
```
Configure M-LAG system parameters.
```
[SwitchA] m-lag system-mac 0001-0001-0001
[SwitchA] m-lag system-number 1
[SwitchA] m-lag system-priority 10
[SwitchA] m-lag keepalive ip destination 60.1.1.2 source 60.1.1.1
[SwitchA] m-lag restore-delay 180
```
Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3.
```
[SwitchA] interface bridge-aggregation 3
[SwitchA-Bridge-Aggregation3] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation3] quit
```
Assign GigabitEthernet 1/0/3 to link aggregation group 3.
```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] port link-aggregation group 3
[SwitchA-GigabitEthernet1/0/3] quit
```
Specify Bridge-Aggregation 3 as the peer-link interface.
```
[SwitchA] interface bridge-aggregation 3
[SwitchA-Bridge-Aggregation3] port m-lag peer-link 1
[SwitchA-Bridge-Aggregation3] quit
```
Disable spanning tree on GigabitEthernet 1/0/5.
```
[SwitchA] interface gigabitethernet 1/0/5
[SwitchA-GigabitEthernet1/0/5] undo stp enable
[SwitchA-GigabitEthernet1/0/5] quit
```
Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.
```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation4] quit
```
Assign GigabitEthernet 1/0/1 to link aggregation group 4.
```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 4
[SwitchA-GigabitEthernet1/0/1] quit
```
Assign Bridge-Aggregation 4 to M-LAG group 4.
```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] port m-lag group 4
[SwitchA-Bridge-Aggregation4] quit
```
Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5.
```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation5] quit
```
Assign GigabitEthernet 1/0/2 to link aggregation group 5.
```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-aggregation group 5
[SwitchA-GigabitEthernet1/0/2] quit
```
Assign Bridge-Aggregation 5 to M-LAG group 5.
```
[SwitchA] interface bridge-aggregation 5
```

```
[SwitchA-Bridge-Aggregation5] port m-lag group 5
[SwitchA-Bridge-Aggregation5] quit
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] arp suppression enable
[SwitchA-vsi-vpna] evpn encapsulation vxlan
[SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchA-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] peer 3.3.3.3 as-number 200
[SwitchA-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 3.3.3.3 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On Bridge-Aggregation 4, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] service-instance 1000
[SwitchA-Bridge-Aggregation4-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-Bridge-Aggregation4-srv1000] xconnect vsi vpna
[SwitchA-Bridge-Aggregation4-srv1000] quit
```

On Bridge-Aggregation 5, create Ethernet service instance 1000 to match VLAN 3.

```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] service-instance 1000
[SwitchA-Bridge-Aggregation5-srv1000] encapsulation s-vid 3
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-Bridge-Aggregation5-srv1000] xconnect vsi vpna
[SwitchA-Bridge-Aggregation5-srv1000] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC learning and remote ARP learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel arp-learning disable
```

Enable EVPN M-LAG, and specify the virtual VTEP address as 1.2.3.4.

```
[SwitchB] evpn m-lag group 1.2.3.4
```

Configure M-LAG system parameters.

```
[SwitchB] m-lag system-mac 0001-0001-0001
```

```
[SwitchB] m-lag system-number 2
[SwitchB] m-lag system-priority 10
[SwitchB] m-lag keepalive ip destination 60.1.1.1 source 60.1.1.2
[SwitchB] m-lag restore-delay 180
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 3.

```
[SwitchB] interface bridge-aggregation 3
[SwitchB-Bridge-Aggregation3] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation3] quit
```

Assign GigabitEthernet 1/0/3 to aggregation group 3.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port link-aggregation group 3
[SwitchB-GigabitEthernet1/0/3] quit
```

Specify Bridge-Aggregation 3 as the peer-link interface.

```
[SwitchB] interface bridge-aggregation 3
[SwitchB-Bridge-Aggregation3] port m-lag peer-link 1
[SwitchB-Bridge-Aggregation3] quit
```

Disable spanning tree on GigabitEthernet 1/0/5.

```
[SwitchB] interface gigabitethernet 1/0/5
[SwitchB-GigabitEthernet1/0/5] undo stp enable
[SwitchB-GigabitEthernet1/0/5] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 to aggregation group 4.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-aggregation group 4
[SwitchB-GigabitEthernet1/0/1] quit
```

Assign Bridge-Aggregation 4 to M-LAG group 4.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] port m-lag group 4
[SwitchB-Bridge-Aggregation4] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/2 to aggregation group 5.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-aggregation group 5
[SwitchB-GigabitEthernet1/0/2] quit
```

Assign Bridge-Aggregation 5 to M-LAG group 5.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] port m-lag group 5
[SwitchB-Bridge-Aggregation5] quit
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpna
```

```
[SwitchB-vsi-vpna] arp suppression enable
[SwitchB-vsi-vpna] evpn encapsulation vxlan
[SwitchB-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 3.3.3.3 as-number 200
[SwitchB-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 3.3.3.3 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

On Bridge-Aggregation 4, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] service-instance 1000
[SwitchB-Bridge-Aggregation4-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-Bridge-Aggregation4-srv1000] xconnect vsi vpna
[SwitchB-Bridge-Aggregation4-srv1000] quit
```

On Bridge-Aggregation 5, create Ethernet service instance 1000 to match VLAN 3.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] service-instance 1000
[SwitchB-Bridge-Aggregation5-srv1000] encapsulation s-vid 3
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-Bridge-Aggregation5-srv1000] xconnect vsi vpna
[SwitchB-Bridge-Aggregation5-srv1000] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address learning.

```
[SwitchC] vxlan tunnel mac-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] evpn encapsulation vxlan
[SwitchC-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchC-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
```

```
[SwitchC-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] evpn encapsulation vxlan
[SwitchC-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchC-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes, and configure Switch C as an RR.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] group evpn
[SwitchC-bgp-default] peer 1.1.1.1 group evpn
[SwitchC-bgp-default] peer 2.2.2.2 group evpn
[SwitchC-bgp-default] peer 4.4.4.4 group evpn
[SwitchC-bgp-default] peer evpn as-number 200
[SwitchC-bgp-default] peer evpn connect-interface loopback 0
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer evpn enable
[SwitchC-bgp-default-evpn] undo policy vpn-target
[SwitchC-bgp-default-evpn] peer evpn reflect-client
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Create VSI-interface 1 and assign it an IP address. The IP address is the gateway address of VXLAN 10.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchC-Vsi-interface1] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] gateway vsi-interface 1
[SwitchC-vsi-vpna] quit
```

Create VSI-interface 2 and assign it an IP address. The IP address is the gateway address of VXLAN 20.

```
[SwitchC] interface vsi-interface 2
[SwitchC-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchC-Vsi-interface2] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] gateway vsi-interface 2
[SwitchC-vsi-vpnb] quit
```

6. Configure Switch D:

Enable L2VPN.

```
<SwitchD> system-view
[SwitchD] l2vpn enable
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchD] vsi vpnb
[SwitchD-vsi-vpnb] arp suppression enable
[SwitchD-vsi-vpnb] evpn encapsulation vxlan
[SwitchD-vsi-vpnb-evpn-vxlan] route-distinguisher auto
[SwitchD-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchD-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchD-vsi-vpnb] vxlan 20
[SwitchD-vsi-vpnb-vxlan-20] quit
[SwitchD-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchD] bgp 200
[SwitchD-bgp-default] peer 3.3.3.3 as-number 200
[SwitchD-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer 3.3.3.3 enable
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 4.

```
[SwitchD] interface gigabitethernet 1/0/1
[SwitchD-GigabitEthernet1/0/1] service-instance 1000
[SwitchD-GigabitEthernet1/0/1] encapsulation s-vid 4
```

Map Ethernet service instance 1000 to VSI **vpnb**.

```
[SwitchD-GigabitEthernet1/0/1] xconnect vsi vpnb
[SwitchD-GigabitEthernet1/0/1] quit
```

Verifying the configuration

1. Verify the centralized EVPN gateway settings on Switch C:

Verify that Switch C has advertised MAC/IP advertisement routes and IMET routes of the gateway to other devices. Verify that Switch C has received MAC/IP advertisement routes and IMET routes from Switch A, Switch B, and Switch D. (Details not shown.)

Verify that the VXLAN tunnel to Switch A and Switch B is up, and the tunnel destination address is the virtual VTEP address.

```
[SwitchC] display interface tunnel
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 3.3.3.3, destination 1.2.3.4
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 4 bytes/sec, 32 bits/sec, 0 packets/sec
Input: 2 packets, 340 bytes, 0 drops
```

```
Output: 16 packets, 2793 bytes, 0 drops
```

Verify that the VXLAN tunnels have been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

- For dynamic ACs whose frame match criteria are generated based on VXLAN IDs:

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 1
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : Unlimited
  Broadcast Restrain      : Unlimited
  Multicast Restrain      : Unlimited
  Unknown Unicast Restrain: Unlimited
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel1              0x5000000  UP       Auto        Disabled
  ACs:
    AC                              Link ID  State       Type
    BAGG3 srv1                      0        Up          Dynamic (M-LAG)
    BAGG4 srv1000                   1        Up          Manual
    BAGG5 srv1000                   2        Up          Manual
```

- For dynamic ACs whose frame match criteria are identical to those of site-facing ACs:

```
[SwitchC] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
```

```
      Tunnel1              0x5000000  UP       Auto        Disabled

   VSI Name: vpnb
     VSI Index               : 1
     VSI State               : Up
     MTU                     : 1500
     Bandwidth               : -
     Broadcast Restrain      : -
     Multicast Restrain      : -
     Unknown Unicast Restrain: -
     MAC Learning            : Enabled
     MAC Table Limit         : -
     MAC Learning rate       : -
     Drop Unknown            : -
     Flooding                : Enabled
     Statistics              : Disabled
     Gateway Interface       : VSI-interface 2
     VXLAN ID                : 20
     Tunnels:
       Tunnel Name          Link ID    State    Type        Flood proxy
       Tunnel2              0x5000001  UP       Auto        Disabled
```

2. Verify the M-LAG settings on Switch A:

Verify that Switch A has BGP EVPN routes.

```
[SwitchA] display bgp l2vpn evpn

 BGP local router ID is 1.2.3.4
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
               Origin: i - IGP, e - EGP, ? - incomplete

 Total number of routes from all PEs: 5

 Route distinguisher: 1:100
 Total number of routes: 5

* >  Network : [2][0][48][0800-2700-400e][0][0.0.0.0]/104
     NextHop : 1.2.3.4                                LocPrf    : 100
     PrefVal : 32768                                  OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >i Network : [2][0][48][46b2-aea0-0101][0][0.0.0.0]/104
     NextHop : 3.3.3.3                                LocPrf    : 100
     PrefVal : 0                                      OutLabel  : NULL
     MED     : 0
     Path/Ogn: i
```

```
* >  Network : [2][0][48][ac1e-24e3-0201][0][0.0.0.0]/104
     NextHop : 3.3.3.3                                  LocPrf    : 100
     PrefVal : 0                                        OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >i Network : [3][0][32][1.2.3.4]/80
     NextHop : 1.2.3.4                                  LocPrf    : 100
     PrefVal : 32768                                    OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >i Network : [3][0][32][3.3.3.3]/80
     NextHop : 3.3.3.3                                  LocPrf    : 100
     PrefVal : 0                                        OutLabel  : NULL
     MED     : 0
     Path/Ogn: i
```

Verify that the VXLAN tunnel to Switch C is up, and the tunnel source address is the virtual VTEP address.

```
[SwitchA] display interface tunnel
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.2.3.4, destination 3.3.3.3
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 1 bytes/sec, 8 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 33 packets, 6121 bytes, 0 drops
```

Verify that ACs are automatically created on the peer link and assigned to VSIs.

- For dynamic ACs whose frame match criteria are generated based on VXLAN IDs:

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index              : 1
  VSI State              : Up
  MTU                    : 1500
  Bandwidth              : -
  Broadcast Restrain     : -
  Multicast Restrain     : -
  Unknown Unicast Restrain: -
  MAC Learning           : Enabled
  MAC Table Limit        : -
  MAC Learning rate      : -
```

```
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel1              0x5000000  UP       Auto        Disabled
  ACs:
    AC                              Link ID  State       Type
    BAGG3 srv1                      0        Up          Dynamic (M-LAG)
    BAGG4 srv1000                   1        Up          Manual
    BAGG5 srv1000                   2        Up          Manual
```

- For dynamic ACs whose frame match criteria are identical to those of site-facing ACs:

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 1
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel1              0x5000000  UP       Auto        Disabled
  ACs:
    AC                              Link ID  State       Type
    BAGG4 srv1000                   0        Up          Manual
    BAGG3 srv2                      1        Up          Dynamic (M-LAG)
    BAGG5 srv1000                   2        Up          Manual
    BAGG3 srv3                      3        Up          Dynamic (M-LAG)
```

3. Verify network connectivity for the VMs:

Verify that VM 1, VM 2, and VM 3 can communicate when both Switch A and Switch B are operating correctly. (Details not shown.)

Verify that VM 1, VM 2, and VM 3 can communicate when Switch A's or Switch B's links to the local site are disconnected. (Details not shown.)

Example: Configuring IPv4 EVPN M-LAG with a tunnel peer link

Network configuration

As shown in Figure 26, perform the following tasks to make sure the VMs can communicate with one another:

- Configure VXLAN 10 on Switch A, Switch B, and Switch C, and configure VXLAN 20 on Switch C and Switch D.
- Configure EVPN M-LAG on Switch A and Switch B to virtualize them into one VTEP. The switches use a tunnel peer link.
- Create a monitor link group on Switch A and Switch B. Configure the transport-facing interfaces of Switch A and Switch B as uplink interfaces for the monitor link group, and member interfaces of M-LAG interfaces as downlink interfaces.
- Configure Switch C as a centralized EVPN gateway and RR.

NOTE:

This example provides configuration of IPv4 sites extended by an IPv4 underlay network. The configuration procedure does not differ between site or underlay network types.

Figure 26 Network diagram

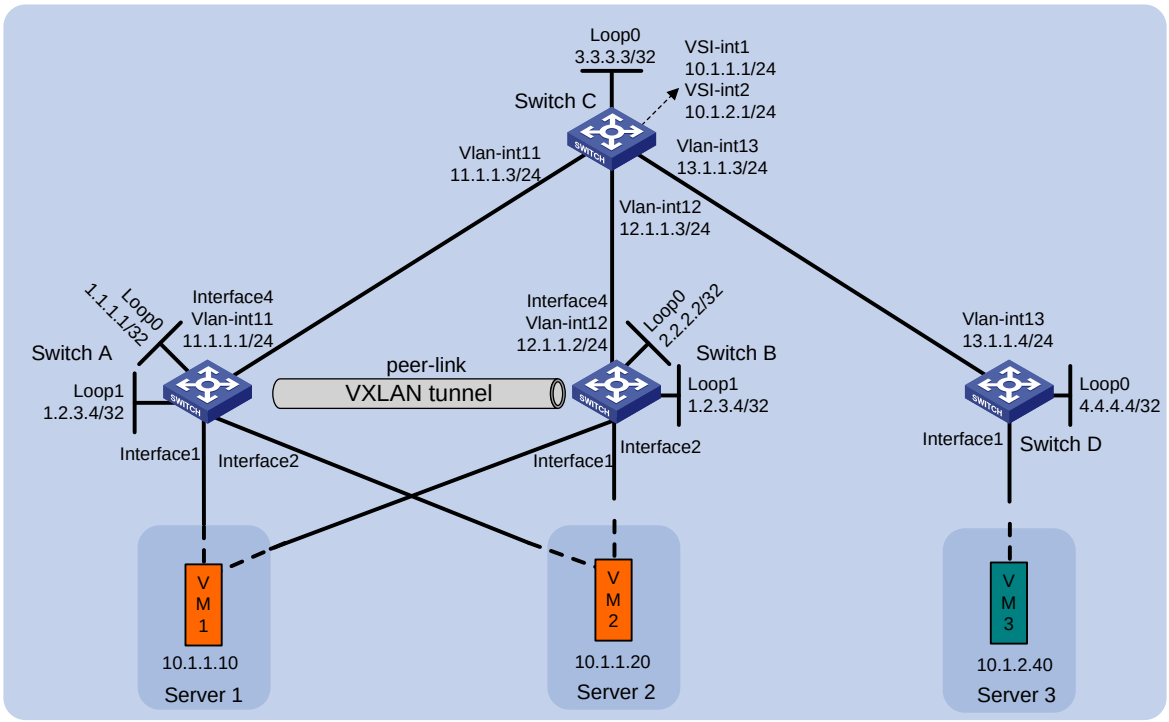


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Interface4	GigabitEthernet1/0/4

Procedure

1. On VM 1 and VM 2, specify 10.1.1.1 as the gateway address. On VM 3, specify 10.1.2.1 as the gateway address. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to interfaces (including loopback interfaces), as shown in Figure 26. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)

   # Execute the **`ospf peer hold-max-cost duration`** command on the interfaces used for setting up OSPF neighbor relationships. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC learning and remote ARP learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   [SwitchA] vxlan tunnel arp-learning disable
   ```

   # Enable EVPN M-LAG, and specify the virtual VTEP address as 1.2.3.4.

   ```
   [SwitchA] evpn m-lag group 1.2.3.4
   ```

   # Specify the reserved VXLAN as VXLAN 1234.

   ```
   [SwitchA] reserved vxlan 1234
   ```

   # Configure M-LAG system parameters.

   ```
   [SwitchA] m-lag system-mac 0001-0001-0001
   [SwitchA] m-lag system-number 1
   [SwitchA] m-lag system-priority 10
   [SwitchA] m-lag restore-delay 180
   ```

   # Create a tunnel to Switch B, and set the ToS of tunneled packets to 100.

   ```
   [SwitchA] interface tunnel 1 mode vxlan
   [SwitchA-Tunnel1] source 1.1.1.1
   [SwitchA-Tunnel1] destination 2.2.2.2
   [SwitchA-Tunnel1] tunnel tos 100
   [SwitchA-Tunnel1] quit
   ```

   # Exclude Tunnel 1 from the shutdown action by M-LAG MAD.

   ```
   [SwitchA] m-lag mad exclude interface tunnel 1
   ```

   # Specify Tunnel 1 as the peer-link interface.

   ```
   [SwitchA] interface tunnel 1
   [SwitchA-Tunnel1] port m-lag peer-link 1
   [SwitchA-Tunnel1] quit
   ```

   # Disable spanning tree on GigabitEthernet 1/0/4.

   ```
   [SwitchA] interface gigabitethernet 1/0/4
   [SwitchA-GigabitEthernet1/0/4] undo stp enable
   [SwitchA-GigabitEthernet1/0/4] quit
   ```

   # Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

   ```
   [SwitchA] interface bridge-aggregation 4
   [SwitchA-Bridge-Aggregation4] link-aggregation mode dynamic
   ```

```
[SwitchA-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 to link aggregation group 4.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 4
[SwitchA-GigabitEthernet1/0/1] quit
```

Assign Bridge-Aggregation 4 to M-LAG group 4.

```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] port m-lag group 4
[SwitchA-Bridge-Aggregation4] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5.

```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] link-aggregation mode dynamic
[SwitchA-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/2 to link aggregation group 5.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-aggregation group 5
[SwitchA-GigabitEthernet1/0/2] quit
```

Assign Bridge-Aggregation 5 to M-LAG group 5.

```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] port m-lag group 5
[SwitchA-Bridge-Aggregation5] quit
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] arp suppression enable
[SwitchA-vsi-vpna] evpn encapsulation vxlan
[SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchA-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] peer 3.3.3.3 as-number 200
[SwitchA-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 3.3.3.3 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On Bridge-Aggregation 4, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface bridge-aggregation 4
[SwitchA-Bridge-Aggregation4] service-instance 1000
[SwitchA-Bridge-Aggregation4-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-Bridge-Aggregation4-srv1000] xconnect vsi vpna
[SwitchA-Bridge-Aggregation4-srv1000] quit
```

On Bridge-Aggregation 5, create Ethernet service instance 1000 to match VLAN 3.

```
[SwitchA] interface bridge-aggregation 5
[SwitchA-Bridge-Aggregation5] service-instance 1000
[SwitchA-Bridge-Aggregation5-srv1000] encapsulation s-vid 3
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-Bridge-Aggregation5-srv1000] xconnect vsi vpna
[SwitchA-Bridge-Aggregation5-srv1000] quit
```

Create monitor link group 1 and assign uplink and downlink interfaces to it.

```
[SwitchA] monitor-link group 1
[SwitchA-mtlk-group1] port gigabitethernet 1/0/1 downlink
[SwitchA-mtlk-group1] port gigabitethernet 1/0/2 downlink
[SwitchA-mtlk-group1] port gigabitethernet 1/0/4 uplink
[SwitchA-mtlk-group1] quit
```

Exclude the interfaces used by EVPN from the shutdown action by M-LAG MAD, except the aggregation member ports.

```
[SwitchA] m-lag mad exclude interface loopback 0
[SwitchA] m-lag mad exclude interface loopback 1
[SwitchA] m-lag mad exclude interface gigabitethernet 1/0/4
[SwitchA] m-lag mad exclude interface vlan-interface 11
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC learning and remote ARP learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel arp-learning disable
```

Enable EVPN M-LAG, and specify the virtual VTEP address as 1.2.3.4.

```
[SwitchB] evpn m-lag group 1.2.3.4
```

Specify the reserved VXLAN as VXLAN 1234.

```
[SwitchB] reserved vxlan 1234
```

Configure M-LAG system parameters.

```
[SwitchB] m-lag system-mac 0001-0001-0001
[SwitchB] m-lag system-number 2
[SwitchB] m-lag system-priority 10
[SwitchB] m-lag restore-delay 180
```

Create a tunnel to Switch A, and set the ToS of tunneled packets to 100.

```
[SwitchB] interface tunnel 1 mode vxlan
[SwitchB-Tunnel1] source 2.2.2.2
[SwitchB-Tunnel1] destination 1.1.1.1
[SwitchB-Tunnel1] tunnel tos 100
[SwitchB-Tunnel1] quit
```

Exclude Tunnel 1 from the shutdown action by M-LAG MAD.

```
[SwitchB] m-lag mad exclude interface tunnel 1
```

Specify Tunnel 1 as the peer-link interface.

```
[SwitchB] interface tunnel 1
[SwitchB-Tunnel1] port m-lag peer-link 1
[SwitchB-Tunnel1] quit
```

Disable spanning tree on GigabitEthernet 1/0/4.

```
[SwitchB] interface gigabitethernet 1/0/4
[SwitchB-GigabitEthernet1/0/4] undo stp enable
[SwitchB-GigabitEthernet1/0/4] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 4.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation4] quit
```

Assign GigabitEthernet 1/0/1 to aggregation group 4.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-aggregation group 4
[SwitchB-GigabitEthernet1/0/1] quit
```

Assign Bridge-Aggregation 4 to M-LAG group 4.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] port m-lag group 4
[SwitchB-Bridge-Aggregation4] quit
```

Create Layer 2 dynamic aggregate interface Bridge-Aggregation 5.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] link-aggregation mode dynamic
[SwitchB-Bridge-Aggregation5] quit
```

Assign GigabitEthernet 1/0/2 to aggregation group 5.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-aggregation group 5
[SwitchB-GigabitEthernet1/0/2] quit
```

Assign Bridge-Aggregation 5 to M-LAG group 5.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] port m-lag group 5
[SwitchB-Bridge-Aggregation5] quit
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] arp suppression enable
[SwitchB-vsi-vpna] evpn encapsulation vxlan
[SwitchB-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchB-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 3.3.3.3 as-number 200
[SwitchB-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 3.3.3.3 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

On Bridge-Aggregation 4, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchB] interface bridge-aggregation 4
[SwitchB-Bridge-Aggregation4] service-instance 1000
[SwitchB-Bridge-Aggregation4-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-Bridge-Aggregation4-srv1000] xconnect vsi vpna
[SwitchB-Bridge-Aggregation4-srv1000] quit
```

On Bridge-Aggregation 5, create Ethernet service instance 1000 to match VLAN 3.

```
[SwitchB] interface bridge-aggregation 5
[SwitchB-Bridge-Aggregation5] service-instance 1000
[SwitchB-Bridge-Aggregation5-srv1000] encapsulation s-vid 3
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchB-Bridge-Aggregation5-srv1000] xconnect vsi vpna
[SwitchB-Bridge-Aggregation5-srv1000] quit
```

Create monitor link group 1 and assign uplink and downlink interfaces to it.

```
[SwitchB] monitor-link group 1
[SwitchB-mtlk-group1] port gigabitethernet 1/0/1 downlink
[SwitchB-mtlk-group1] port gigabitethernet 1/0/2 downlink
[SwitchB-mtlk-group1] port gigabitethernet 1/0/4 uplink
[SwitchB-mtlk-group1] quit
```

Exclude the interfaces used by EVPN from the shutdown action by M-LAG MAD, except the aggregation member ports.

```
[SwitchB] m-lag mad exclude interface loopback 0
[SwitchB] m-lag mad exclude interface loopback 1
[SwitchB] m-lag mad exclude interface gigabitethernet 1/0/4
[SwitchB] m-lag mad exclude interface vlan-interface 12
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address learning.

```
[SwitchC] vxlan tunnel mac-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] evpn encapsulation vxlan
[SwitchC-vsi-vpna-evpn-vxlan] route-distinguisher auto
[SwitchC-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] evpn encapsulation vxlan
[SwitchC-vsi-vpnb-evpn-vxlan] route-distinguisher auto
```

```
[SwitchC-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes, and configure Switch C as an RR.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] group evpn
[SwitchC-bgp-default] peer 1.1.1.1 group evpn
[SwitchC-bgp-default] peer 2.2.2.2 group evpn
[SwitchC-bgp-default] peer 4.4.4.4 group evpn
[SwitchC-bgp-default] peer evpn as-number 200
[SwitchC-bgp-default] peer evpn connect-interface loopback 0
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer evpn enable
[SwitchC-bgp-default-evpn] undo policy vpn-target
[SwitchC-bgp-default-evpn] peer evpn reflect-client
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Create VSI-interface 1 and assign it an IP address. The IP address is the gateway address of VXLAN 10.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchC-Vsi-interface1] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] gateway vsi-interface 1
[SwitchC-vsi-vpna] quit
```

Create VSI-interface 2 and assign it an IP address. The IP address is the gateway address of VXLAN 20.

```
[SwitchC] interface vsi-interface 2
[SwitchC-Vsi-interface2] ip address 10.1.2.1 255.255.255.0
[SwitchC-Vsi-interface2] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] gateway vsi-interface 2
[SwitchC-vsi-vpnb] quit
```

6. Configure Switch D:

Enable L2VPN.

```
<SwitchD> system-view
[SwitchD] l2vpn enable
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchD] vsi vpnb
[SwitchD-vsi-vpnb] arp suppression enable
[SwitchD-vsi-vpnb] evpn encapsulation vxlan
[SwitchD-vsi-vpnb-evpn-vxlan] route-distinguisher auto
```

```
[SwitchD-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchD-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchD-vsi-vpnb] vxlan 20
[SwitchD-vsi-vpnb-vxlan-20] quit
[SwitchD-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchD] bgp 200
[SwitchD-bgp-default] peer 3.3.3.3 as-number 200
[SwitchD-bgp-default] peer 3.3.3.3 connect-interface loopback 0
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer 3.3.3.3 enable
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 4.

```
[SwitchD] interface gigabitethernet 1/0/1
[SwitchD-GigabitEthernet1/0/1] service-instance 1000
[SwitchD-GigabitEthernet1/0/1] encapsulation s-vid 4
```

Map Ethernet service instance 1000 to VSI **vpnb**.

```
[SwitchD-GigabitEthernet1/0/1] xconnect vsi vpnb
[SwitchD-GigabitEthernet1/0/1] quit
```

Verifying the configuration

1. Verify the centralized EVPN gateway settings on Switch C:

   # Verify that Switch C has advertised MAC/IP advertisement routes and IMET routes of the gateway to other devices. Verify that Switch C has received MAC/IP advertisement routes and IMET routes from Switch A, Switch B, and Switch D. (Details not shown.)

   # Verify that the VXLAN tunnels to Switch A and Switch B are up, and the device has established a VXLAN tunnel to Switch A and Switch B with the destination address as the virtual VTEP address.

```
[SwitchC] display interface tunnel
Tunnel11
Current state: UP
Line protocol state: UP
Description: Tunnel11 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 3.3.3.3, destination 1.1.1.1
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 2 packets, 84 bytes, 0 drops

Tunnel12
Current state: UP
Line protocol state: UP
```

```
Description: Tunnel12 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 3.3.3.3, destination 1.2.3.4
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 1 packets, 42 bytes, 0 drops

Tunnel13
Current state: UP
Line protocol state: UP
Description: Tunnel13 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 3.3.3.3, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 1 packets, 42 bytes, 0 drops

Tunnel14
Current state: UP
Line protocol state: UP
Description: Tunnel14 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Output queue - Urgent queuing: Size/Length/Discards 0/100/0
Output queue - Protocol queuing: Size/Length/Discards 0/500/0
Output queue - FIFO queuing: Size/Length/Discards 0/75/0
Last clearing of counters: Never
Tunnel source 3.3.3.3, destination 4.4.4.4
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops
```

Verify that the VXLAN tunnels have been assigned to the VXLANs, and that the VSI interfaces are the gateway interfaces of their respective VXLANs.

```
[SwitchC] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
```

```
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 1
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel11             0x5000001  UP       Auto        Disabled
    Tunnel12             0x5000002  UP       Auto        Disabled
    Tunnel13             0x5000003  UP       Auto        Disabled

VSI Name: vpnb
  VSI Index               : 1
  VSI State               : Up
  MTU                     : 1500
...
  Flooding                : Enabled
  Statistics              : Disabled
  Gateway Interface       : VSI-interface 2
  VXLAN ID                : 20
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel10              0x5000000  UP       Auto        Disabled
```

2. Verify the M-LAG settings on Switch A:

Verify that Switch A has BGP EVPN routes.

```
[SwitchA] display bgp l2vpn evpn

 BGP local router ID is 1.2.3.4
 Status codes: * - valid, > - best, d - dampened, h - history,
               s - suppressed, S - stale, i - internal, e - external
               a - additional-path
               Origin: i - IGP, e - EGP, ? - incomplete

 Total number of routes from all PEs: 3

 Route distinguisher: 1:10
 Total number of routes: 5

* >i Network : [2][0][48][7e9a-48e9-0100][32][10.1.1.1]/136
```

```
     NextHop : 3.3.3.3                                  LocPrf    : 100
     PrefVal : 0                                        OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >  Network : [3][0][32][1.1.1.1]/80
     NextHop : 1.1.1.1                                  LocPrf    : 100
     PrefVal : 32768                                    OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >  Network : [3][0][32][1.2.3.4]/80
     NextHop : 1.2.3.4                                  LocPrf    : 100
     PrefVal : 32768                                    OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >i Network : [3][0][32][3.3.3.3]/80
     NextHop : 3.3.3.3                                  LocPrf    : 100
     PrefVal : 0                                        OutLabel  : NULL
     MED     : 0
     Path/Ogn: i

* >i Network : [3][0][32][2.2.2.2]/80
     NextHop : 2.2.2.2                                  LocPrf    : 100
     PrefVal : 0                                        OutLabel  : NULL
     MED     : 0
     Path/Ogn: i
```

Verify that the VXLAN tunnel to Switch C is up, and the tunnel source address is the virtual VTEP address.

```
[SwitchA] display interface tunnel
Tunnel10
Current state: UP
Line protocol state: UP
Description: Tunnel10 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.2.3.4, destination 3.3.3.3
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Last 300 seconds output rate: 0 bytes/sec, 0 bits/sec, 0 packets/sec
Input: 0 packets, 0 bytes, 0 drops
Output: 0 packets, 0 bytes, 0 drops

Tunnel11
Current state: UP
Line protocol state: UP
```

```
Description: Tunnel11 Interface
Bandwidth: 64 kbps
Maximum transmission unit: 1464
Internet protocol processing: Disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
Last 300 seconds input rate: 12 bytes/sec, 96 bits/sec, 0 packets/sec
Last 300 seconds output rate: 12 bytes/sec, 96 bits/sec, 0 packets/sec
Input: 239 packets, 25558 bytes, 0 drops
Output: 1241 packets, 109811 bytes, 0 drops
```

Verify that ACs are automatically created on the peer link and assigned to VSIs.

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  MAC Learning rate       : -
  Drop Unknown            : -
  Flooding                : Enabled
  Statistics              : Disabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State    Type        Flood proxy
    Tunnel10              0x5000000  UP       Auto        Disabled
    Tunnel11              0x5000001  UP       Manual      Disabled
  ACs:
    AC                               Link ID  State       Type
    BAGG4 srv1000                    0        Down        Manual
    BAGG5 srv1000                    1        Down        Manual
```

3. Verify network connectivity for the VMs:

   # Verify that VM 1, VM 2, and VM 3 can communicate when both Switch A and Switch B are operating correctly. (Details not shown.)

   # Verify that VM 1, VM 2, and VM 3 can communicate when Switch A's or Switch B's links to the local site are disconnected. (Details not shown.)

Example: Configuring IPv4 EVPN multihoming

Network configuration

As shown in Figure 27:

- Configure VXLANs as follows:

- Configure VXLAN 10 on Switch A, Switch B, and Switch C. Configure Switch A and Switch B as redundant VTEPs for Server 2, and configure Switch B and Switch C as redundant VTEPs for Server 3.
- Configure VXLAN 20 on Switch C.

- Configure Switch A, Switch B, and Switch C as distributed EVPN gateways.
- Configure Switch D as an RR to reflect BGP EVPN routes between Switch A, Switch B, and Switch C.

Figure 27 Network diagram

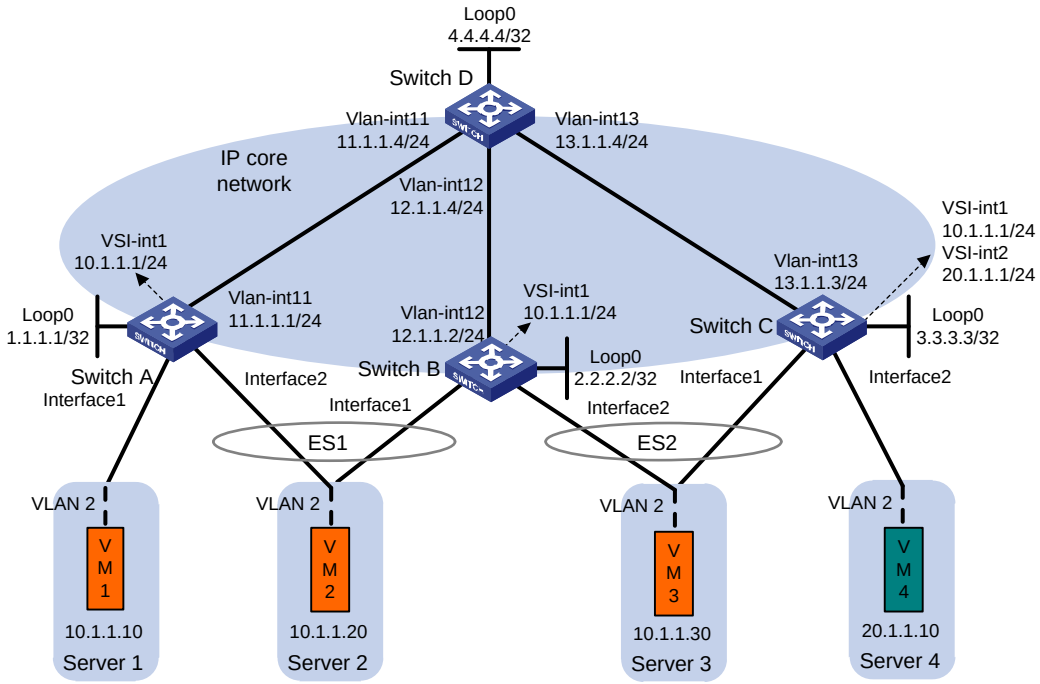


Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

1. On VM 1, VM 2, and VM 3, specify 10.1.1.1 as the gateway address. On VM 4, specify 20.1.1.1 as the gateway address. (Details not shown.)
2. Configure IP addresses and unicast routing settings:

   # Assign IP addresses to the interfaces, as shown in Figure 27. (Details not shown.)

   # Configure OSPF on all transport network switches (Switches A through D) for them to reach one another. (Details not shown.)
3. Configure Switch A:

   # Enable L2VPN.

   ```
   <SwitchA> system-view
   [SwitchA] l2vpn enable
   ```

   # Disable remote MAC address learning and remote ARP learning.

   ```
   [SwitchA] vxlan tunnel mac-learning disable
   ```

```
[SwitchA] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] evpn encapsulation vxlan
[SwitchA-vsi-vpna-evpn-vxlan] route-distinguisher auto router-id
[SwitchA-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchA-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan-10] quit
[SwitchA-vsi-vpna] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchA] bgp 200
[SwitchA-bgp-default] peer 4.4.4.4 as-number 200
[SwitchA-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchA-bgp-default] address-family l2vpn evpn
[SwitchA-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchA-bgp-default-evpn] quit
[SwitchA-bgp-default] quit
```

On GigabitEthernet 1/0/1, create Ethernet service instance 1000 to match VLAN 2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

Map Ethernet service instance 1000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-GigabitEthernet1/0/1] quit
```

Assign an ESI to GigabitEthernet 1/0/2.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] esi 0.0.0.0.1
```

On GigabitEthernet 1/0/2, create Ethernet service instance 2000 to match VLAN 2.

```
[SwitchA-GigabitEthernet1/0/2] service-instance 2000
[SwitchA-GigabitEthernet1/0/2-srv2000] encapsulation s-vid 2
```

Map Ethernet service instance 2000 to VSI **vpna**.

```
[SwitchA-GigabitEthernet1/0/2-srv2000] xconnect vsi vpna
[SwitchA-GigabitEthernet1/0/2-srv2000] quit
[SwitchA-GigabitEthernet1/0/2] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchA] ip vpn-instance l3vpna
[SwitchA-vpn-instance-l3vpna] route-distinguisher 1:1
[SwitchA-vpn-instance-l3vpna] address-family ipv4
[SwitchA-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchA-vpn-ipv4-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] address-family evpn
[SwitchA-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchA-vpn-evpn-l3vpna] quit
[SwitchA-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchA] interface vsi-interface 1
[SwitchA-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchA-Vsi-interface1] mac-address 1-1-1
[SwitchA-Vsi-interface1] distributed-gateway local
[SwitchA-Vsi-interface1] local-proxy-arp enable
[SwitchA-Vsi-interface1] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchA] interface vsi-interface 3
[SwitchA-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchA-Vsi-interface3] l3-vni 1000
[SwitchA-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] gateway vsi-interface 1
[SwitchA-vsi-vpna] quit
```

4. Configure Switch B:

Enable L2VPN.

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchB] vxlan tunnel mac-learning disable
[SwitchB] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] evpn encapsulation vxlan
[SwitchB-vsi-vpna-evpn-vxlan] route-distinguisher auto router-id
[SwitchB-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchB-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan-10] quit
[SwitchB-vsi-vpna] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchB] bgp 200
[SwitchB-bgp-default] peer 4.4.4.4 as-number 200
[SwitchB-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchB-bgp-default] address-family l2vpn evpn
[SwitchB-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchB-bgp-default-evpn] quit
[SwitchB-bgp-default] quit
```

Assign an ESI to GigabitEthernet 1/0/1.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] esi 0.0.0.0.1
```

On GigabitEthernet 1/0/1, create Ethernet service instance 2000 to match VLAN 2.

```
[SwitchB-GigabitEthernet1/0/1] service-instance 2000
[SwitchB-GigabitEthernet1/0/1-srv2000] encapsulation s-vid 2
```

Map Ethernet service instance 2000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/1-srv2000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/1-srv2000] quit
[SwitchB-GigabitEthernet1/0/1] quit
```

Assign an ESI to GigabitEthernet 1/0/2.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] esi 0.0.0.0.2
```

On GigabitEthernet 1/0/2, create Ethernet service instance 3000 to match VLAN 2.

```
[SwitchB-GigabitEthernet1/0/2] service-instance 3000
[SwitchB-GigabitEthernet1/0/2-srv3000] encapsulation s-vid 2
```

Map Ethernet service instance 3000 to VSI **vpna**.

```
[SwitchB-GigabitEthernet1/0/2-srv3000] xconnect vsi vpna
[SwitchB-GigabitEthernet1/0/2-srv3000] quit
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchB] ip vpn-instance l3vpna
[SwitchB-vpn-instance-l3vpna] route-distinguisher 2:2
[SwitchB-vpn-instance-l3vpna] address-family ipv4
[SwitchB-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchB-vpn-ipv4-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] address-family evpn
[SwitchB-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchB-vpn-evpn-l3vpna] quit
[SwitchB-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchB] interface vsi-interface 1
[SwitchB-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchB-Vsi-interface1] mac-address 1-1-1
[SwitchB-Vsi-interface1] distributed-gateway local
[SwitchB-Vsi-interface1] local-proxy-arp enable
[SwitchB-Vsi-interface1] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchB] interface vsi-interface 3
[SwitchB-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchB-Vsi-interface3] l3-vni 1000
[SwitchB-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] gateway vsi-interface 1
[SwitchB-vsi-vpna] quit
```

5. Configure Switch C:

Enable L2VPN.

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```

Disable remote MAC address learning and remote ARP learning.

```
[SwitchC] vxlan tunnel mac-learning disable
[SwitchC] vxlan tunnel arp-learning disable
```

Create an EVPN instance on VSI **vpna**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] evpn encapsulation vxlan
[SwitchC-vsi-vpna-evpn-vxlan] route-distinguisher auto router-id
[SwitchC-vsi-vpna-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpna-evpn-vxlan] quit
```

Create VXLAN 10.

```
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan-10] quit
[SwitchC-vsi-vpna] quit
```

Create an EVPN instance on VSI **vpnb**, and configure the switch to automatically generate an RD and a route target for the EVPN instance.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] evpn encapsulation vxlan
[SwitchC-vsi-vpnb-evpn-vxlan] route-distinguisher auto router-id
[SwitchC-vsi-vpnb-evpn-vxlan] vpn-target auto
[SwitchC-vsi-vpnb-evpn-vxlan] quit
```

Create VXLAN 20.

```
[SwitchC-vsi-vpnb] vxlan 20
[SwitchC-vsi-vpnb-vxlan-20] quit
[SwitchC-vsi-vpnb] quit
```

Configure BGP to advertise BGP EVPN routes.

```
[SwitchC] bgp 200
[SwitchC-bgp-default] peer 4.4.4.4 as-number 200
[SwitchC-bgp-default] peer 4.4.4.4 connect-interface loopback 0
[SwitchC-bgp-default] address-family l2vpn evpn
[SwitchC-bgp-default-evpn] peer 4.4.4.4 enable
[SwitchC-bgp-default-evpn] quit
[SwitchC-bgp-default] quit
```

Assign an ESI to GigabitEthernet 1/0/1.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] esi 0.0.0.0.2
```

On GigabitEthernet 1/0/1, create Ethernet service instance 3000 to match VLAN 2.

```
[SwitchC-GigabitEthernet1/0/1] service-instance 3000
[SwitchC-GigabitEthernet1/0/1-srv3000] encapsulation s-vid 2
```

Map Ethernet service instance 3000 to VSI **vpna**.

```
[SwitchC-GigabitEthernet1/0/1-srv3000] xconnect vsi vpna
[SwitchC-GigabitEthernet1/0/1-srv3000] quit
[SwitchC-GigabitEthernet1/0/1] quit
```

On GigabitEthernet 1/0/2, create Ethernet service instance 4000 to match VLAN 3.

```
[SwitchC] interface gigabitethernet 1/0/2
[SwitchC-GigabitEthernet1/0/2] service-instance 4000
[SwitchC-GigabitEthernet1/0/2-srv4000] encapsulation s-vid 3
```

Map Ethernet service instance 4000 to VSI **vpnb**.

```
[SwitchC-GigabitEthernet1/0/2-srv4000] xconnect vsi vpnb
[SwitchC-GigabitEthernet1/0/2-srv4000] quit
[SwitchC-GigabitEthernet1/0/2] quit
```

Configure RD and route target settings for VPN instance **l3vpna**.

```
[SwitchC] ip vpn-instance l3vpna
[SwitchC-vpn-instance-l3vpna] route-distinguisher 3:3
[SwitchC-vpn-instance-l3vpna] address-family ipv4
[SwitchC-vpn-ipv4-l3vpna] vpn-target 2:2
[SwitchC-vpn-ipv4-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] address-family evpn
[SwitchC-vpn-evpn-l3vpna] vpn-target 1:1
[SwitchC-vpn-evpn-l3vpna] quit
[SwitchC-vpn-instance-l3vpna] quit
```

Configure VSI-interface 1.

```
[SwitchC] interface vsi-interface 1
[SwitchC-Vsi-interface1] ip binding vpn-instance l3vpna
[SwitchC-Vsi-interface1] ip address 10.1.1.1 255.255.255.0
[SwitchC-Vsi-interface1] mac-address 1-1-1
[SwitchC-Vsi-interface1] distributed-gateway local
[SwitchC-Vsi-interface1] local-proxy-arp enable
[SwitchC-Vsi-interface1] quit
```

Configure VSI-interface 2.

```
[SwitchC] interface vsi-interface 2
[SwitchC-Vsi-interface2] ip binding vpn-instance l3vpna
[SwitchC-Vsi-interface2] ip address 20.1.1.1 255.255.255.0
[SwitchC-Vsi-interface2] mac-address 2-2-2
[SwitchC-Vsi-interface2] distributed-gateway local
[SwitchC-Vsi-interface2] local-proxy-arp enable
[SwitchC-Vsi-interface2] quit
```

Associate VSI-interface 3 with VPN instance **l3vpna**, and configure the L3 VXLAN ID as 1000 for the VPN instance.

```
[SwitchC] interface vsi-interface 3
[SwitchC-Vsi-interface3] ip binding vpn-instance l3vpna
[SwitchC-Vsi-interface3] l3-vni 1000
[SwitchC-Vsi-interface3] quit
```

Specify VSI-interface 1 as the gateway interface for VSI **vpna**.

```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] gateway vsi-interface 1
[SwitchC-vsi-vpna] quit
```

Specify VSI-interface 2 as the gateway interface for VSI **vpnb**.

```
[SwitchC] vsi vpnb
[SwitchC-vsi-vpnb] gateway vsi-interface 2
[SwitchC-vsi-vpnb] quit
```

6. Configure Switch D:

Establish BGP connections with other transport network switches.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp-default] group evpn
```

```
[SwitchD-bgp-default] peer 1.1.1.1 group evpn
[SwitchD-bgp-default] peer 2.2.2.2 group evpn
[SwitchD-bgp-default] peer 3.3.3.3 group evpn
[SwitchD-bgp-default] peer evpn as-number 200
[SwitchD-bgp-default] peer evpn connect-interface loopback 0
```

Configure BGP to advertise BGP EVPN routes, and disable route target filtering for BGP EVPN routes.

```
[SwitchD-bgp-default] address-family l2vpn evpn
[SwitchD-bgp-default-evpn] peer evpn enable
[SwitchD-bgp-default-evpn] undo policy vpn-target
```

Configure Switch D as an RR.

```
[SwitchD-bgp-default-evpn] peer evpn reflect-client
[SwitchD-bgp-default-evpn] quit
[SwitchD-bgp-default] quit
```

Verifying the configuration

1. Verify the EVPN multihoming configuration on Switch C.

Verify that Switch C has advertised and received the following BGP EVPN routes (details not shown):

- IP prefix advertisement routes for the gateways.
- IMET routes for VSIs.
- MAC/IP advertisement routes.
- Ethernet auto-discovery routes and Ethernet segment routes.

Verify that Switch C has ECMP routes to VM 2.

```
<SwitchC> display evpn routing-table vpn-instance l3vpna
Flags: E - with valid ESI   A - AD ready   L - Local ES exists

VPN instance:l3vpna                          Local L3VNI:1000
IP address       Next hop         Outgoing interface    NibID       Flags
10.1.1.10        1.1.1.1          Vsi-interface3        0x18000001  -
10.1.1.20        2.2.2.2          Vsi-interface3        0x18000000  EA
                 1.1.1.1          Vsi-interface3        0x18000001  EA
```

Verify that Switch C has equal-cost L2VPN MAC address entries for VM 2.

```
<SwitchC> display l2vpn mac-address
* - The output interface is issued to another VSI
MAC Address    State     VSI Name                       Link ID/Name   Aging
0001-0001-0010 EVPN      vpna                           Tunnel1        NotAging
0001-0001-0020 EVPN      vpna                           Tunnel1        NotAging
                                                        Tunnel2        NotAging
0001-0001-0030 Dynamic   vpna                           GE1/0/1        Aging
0002-0001-0010 Dynamic   vpnb                           GE1/0/2        Aging
--- 4 mac address(es) found  ---
```

Verify that Switch C has EVPN MAC address entries for VM 2.

```
<SwitchC> display evpn route mac
Flags: D - Dynamic   B - BGP      L - Local active
       G - Gateway   S - Static   M - Mapping        I - Invalid
       A - MAC-authentication     P – Multiport/Multicast
```

```
VSI name: vpna
EVPN instance: -
MAC address     Link ID/Name   Flags   Encap           Next hop
0001-0001-0030  0              DL      VXLAN           -
0001-0001-0010  Tunnel1        B       VXLAN           1.1.1.1
0001-0001-0020  Tunnel1        B       VXLAN           1.1.1.1
                Tunnel2        B       VXLAN           2.2.2.2

VSI name: vpnb
EVPN instance: -
MAC address     Link ID/Name   Flags   Encap           Next hop
0002-0001-0010  0              DL      VXLAN           -
```

Verify that Switch C has information about local and remote ESs.

```
<SwitchC> display evpn es local
Redundancy mode: A - All active, S - Single active

VSI name : vpna
EVPN instance: -
ESI                         Tag ID      DF address      Mode  State ESI label
0000.0000.0000.0000.0002    -           2.2.2.2         A     Up    -
<SwitchC> display evpn es remote
Control Flags: P – Primary, B – Backup, C - Control word
Status codes : * - valid
Flags: D - Delay

VSI name : vpna
EVPN instance: -
  ESI                     : 0000.0000.0000.0000.0001
  A-D per ES routes       :
    Peer IP             Remote Redundancy mode
    1.1.1.1             All active
    2.2.2.2             All active
  A-D per EVI routes      :
    Tag ID    Flags   Peer IP
    -         -       1.1.1.1
    -         -       2.2.2.2

  ESI                     : 0000.0000.0000.0000.0002
  Ethernet segment routes :
    2.2.2.2
  A-D per ES routes       :
    Peer IP             Remote Redundancy mode
    2.2.2.2             All active
  A-D per EVI routes      :
    Tag ID    Flags   Peer IP
    -         -       2.2.2.2
```

2. Verify that the VMs can communicate with one another. (Details not shown.)

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

TAP Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the TAP fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring TAP 1

About TAP 1
How TAP works 1
TAP deployment mode 2
Restrictions and guidelines: TAP configuration 3
TAP tasks at a glance 3
Prerequisites for TAP configuration 4
Enabling TAP globally 4
Configuring a traffic class 4
Configuring a traffic behavior 4
Configuring an M:N copying action 4
Configuring a VLAN ID marking action 5
Configuring an outer VLAN tag deleting action 6
Configuring an outer VLAN tag adding action 6
Configuring an IP or MAC address marking action 7
Configuring a timestamp and Ethernet header adding action 8
Configuring a packet truncation action 9
Configuring a tunnel header stripping action 9
Configuring a TAP policy 11
Applying a TAP policy to the inbound direction of an interface 11
Display and maintenance commands for TAP 11
TAP configuration examples 12
Example: Configuring basic M:N copying 12
Example: Configuring M:N copying for aggregate interfaces 13
Example: Configuring VLAN ID marking 16
Example: Configuring outer VLAN tag deleting 17
Example: Configuring outer VLAN tag adding 19
Example: Configuring destination MAC address marking 20
Example: Configuring timestamp and Ethernet header adding 21
Example: Configuring packet truncation 23
Example: Configuring NVGRE header stripping 24

Configuring TAP

About TAP

The growing scale of data center networks places more and more stringent requirements on security and performance. Monitoring network traffic and extracting data from the traffic is an important aspect to be addressed.

The test access point (TAP) feature can copy and forward traffic in real time without interrupting the traffic. Specifically, it copies traffic and sends a copy of the traffic to each of multiple monitoring devices for Internet access behavior analysis, abnormal traffic monitoring, and application monitoring.

How TAP works

The TAP feature is implemented through a QoS policy. Such a QoS policy is also called a TAP policy. In a TAP policy, you can use match criteria to mark packets and use actions to edit the packets and redirecting them to a monitoring group. A monitoring group can contain multiple interfaces, from which copies of traffic are sent.

Packet editing action

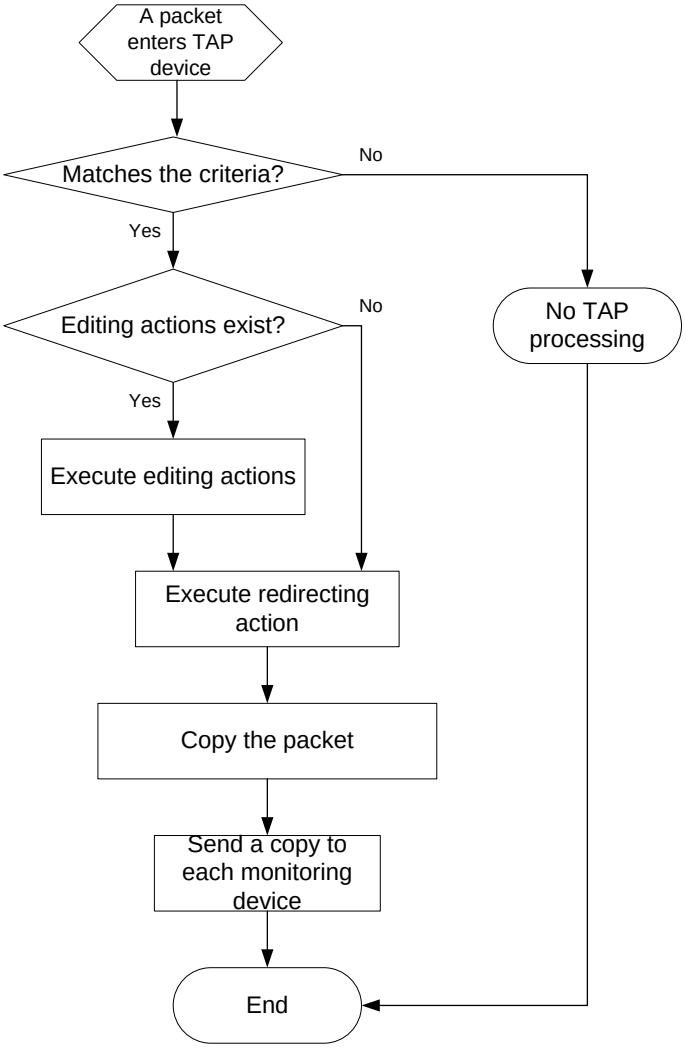
Packet editing actions include the following:

- Mark the IP address or MAC address of packets.
- Add, delete, or mark the VLAN tag of packets.
- Truncate packets.
- Add the timestamp and Ethernet header.
- Strip the tunnel header.

TAP workflow

Figure 1 shows the TAP workflow.

Figure 1 TAP workflow

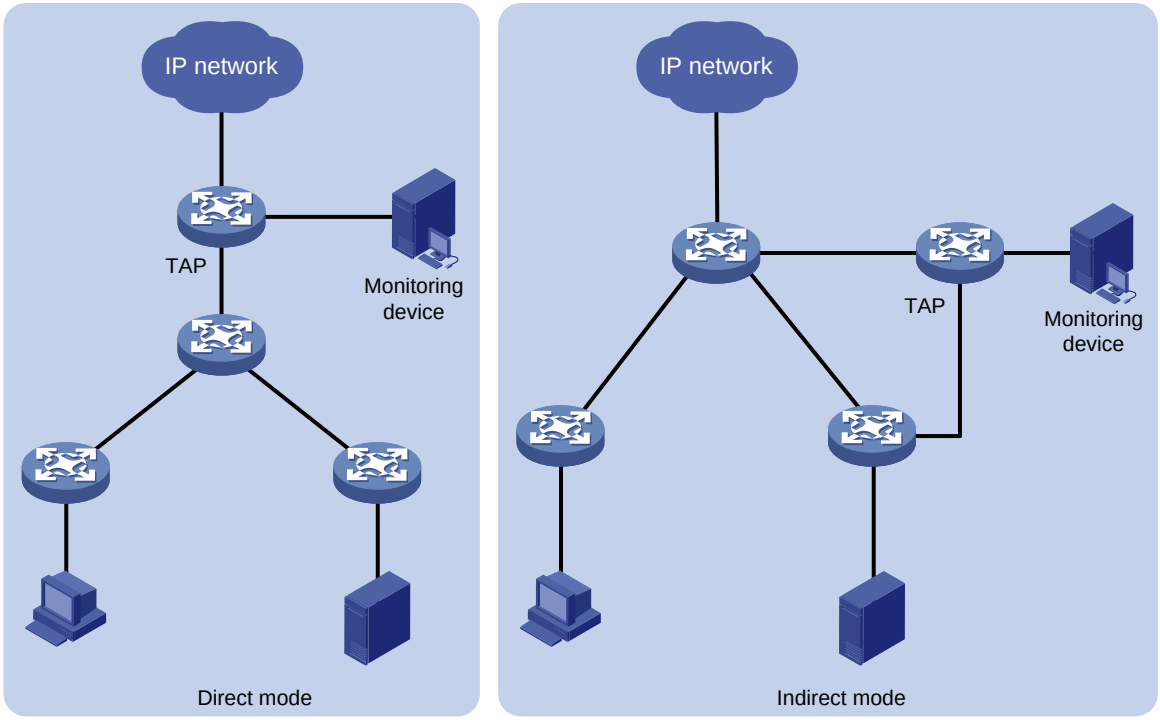


TAP deployment mode

A TAP device can be deployed in the following modes:

- **Direct**—The TAP device is deployed in the user network and can directly copy user traffic without affecting traffic forwarding and send copied traffic to monitoring devices. In this mode, packet editing actions cannot be configured.
- **Indirect**—The TAP device is deployed beside the user network. The user network device sends monitored traffic to the TAP device, which then sends the received traffic to monitoring devices. In this mode, packet editing actions can be configured as needed.

Figure 2 Deployment modes



Restrictions and guidelines: TAP configuration

In a TAP policy, you must configure the action of redirecting packets to a monitoring group, and you can configure packet editing actions as needed. If both the redirecting action and a packet editing action are configured, the packet editing action is executed before the redirecting action.

A TAP policy cannot take effect if the monitoring group in a redirecting action contains unsupported configurations. For such a TAP policy to take effect, perform the following tasks:

1. Delete the unsupported configurations in the monitoring group.
2. Delete the redirecting action and then reconfigure it, or remove the TAP policy and reapply it.

If you configure a packet header stripping action by executing the **`strip-header position`** command and specify an odd value for the *`offset-value`* argument, the deleted part is the odd value minus 1.

A TAP policy can be applied to only a Layer 2 or Layer 3 Ethernet interface.

A TAP policy is implemented through a QoS policy. For more information about related commands, see *ACL and QoS Command Reference*.

TAP tasks at a glance

To configure TAP, perform the following tasks:

3. Enabling TAP globally
4. Configuring a traffic class
5. Configuring a traffic behavior
   - Configuring an M:N copying action
   - Configuring a VLAN ID marking action

- Configuring an outer VLAN tag deleting action
- Configuring an outer VLAN tag adding action
- Configuring an IP or MAC address marking action

6. Configuring a TAP policy
7. Applying a TAP policy to the inbound direction of an interface

Prerequisites for TAP configuration

Before configuring the redirecting action, configure a monitoring group (see flow monitoring configuration in *Network Management and Monitoring Configuration Guide*.

Enabling TAP globally

About this task

Perform this task if the device acts as a pure TAP device between a production network and a monitoring device. After TAP is enabled globally, the device disables unnecessary packet forwarding functions to prevent loops and broadcast storms.

Procedure

8. Enter system view.
   system-view
9. Enable TAP globally.
   tap enable
   By default, TAP is disabled globally.

Configuring a traffic class

10. Enter system view.
    system-view
11. Create a traffic class and enter traffic class view.
    traffic classifier *classifier-name* [**operator** { **and** | **or** }]
12. Configure match criteria.
    if-match *match-criteria*
    By default, no match criterion is configured in a traffic class.
    For more information, see *ACL and QoS Command Reference*.

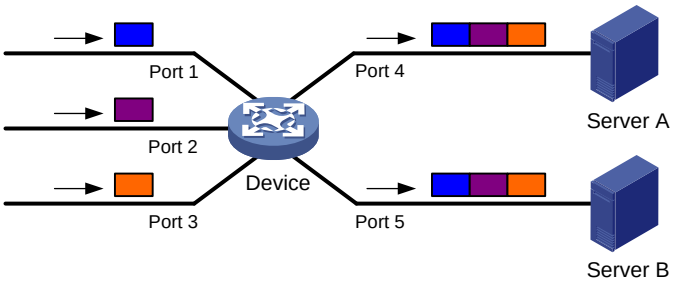
Configuring a traffic behavior

Configuring an M:N copying action

About this task

This feature allows the device to forward packets it receives from multiple (M) interfaces to multiple (N) data monitoring devices (Server A and Server B), as shown in Figure 3.

Figure 3 Network diagram



Restrictions and guidelines

You can implement this feature by performing the following steps:

13. Configure a monitoring group with N member interfaces.
14. Create M TAP policies, and configure an action of redirecting packets to the monitoring group with N member interfaces for each QoS policy.
15. Apply the M TAP policies to M interfaces.

Procedure

16. Enter system view.

    system-view
17. Create a traffic behavior and enter traffic behavior view.

    traffic behavior *behavior-name*
18. Configure an action of redirecting packets to a monitoring group.

    redirect monitoring-group *group-id*

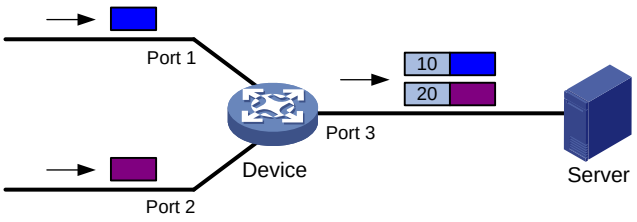
    By default, no action of redirecting packets to a monitoring group is configured.

Configuring a VLAN ID marking action

About this task

This feature allows the device to mark packets received on different interfaces with different VLAN IDs. Then, the data monitoring device (the server) can identify the input interfaces of packets according to the marked VLAN IDs, as shown in Figure 4.

Figure 4 Network diagram



Procedure

19. Enter system view.

    system-view
20. Create a traffic behavior and enter traffic behavior view.

    traffic behavior *behavior-name*
21. Configure a VLAN ID marking action. Choose one option as needed:

- Configure a CVLAN ID marking action.
  remark customer-vlan-id *vlan-id*
  By default, no CVLAN ID marking action is configured.
- Configure an SVLAN ID marking action.
  remark service-vlan-id *vlan-id*
  By default, no SVLAN ID marking action is configured.

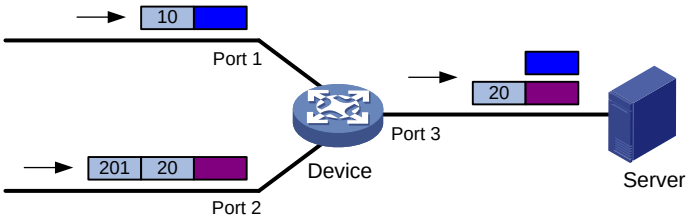
22. Configure an action of redirecting packets to a monitoring group.
    redirect monitoring-group *group-id*
    By default, no action of redirecting packets to a monitoring group is configured.

Configuring an outer VLAN tag deleting action

About this task

This feature is used in scenarios where the data monitoring device (the server) can only receive packets without VLAN tags or with one VLAN tag, as shown in Figure 5.

Figure 5 Network diagram



Procedure

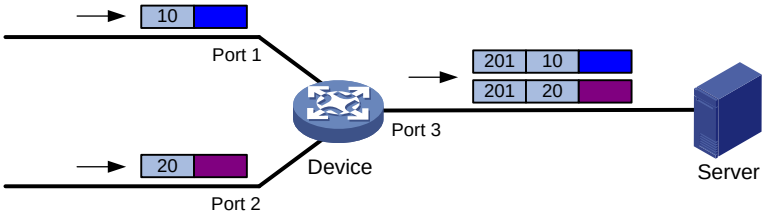
23. Enter system view.
    system-view
24. Create a traffic behavior and enter traffic behavior view.
    traffic behavior *behavior-name*
25. Configure an outer VLAN tag deleting action.
    strip-header top-most-vlan
    By default, no outer VLAN tag deleting action is configured.
26. Configure an action of redirecting packets to a monitoring group.
    redirect monitoring-group *group-id*
    By default, no action of redirecting packets to a monitoring group is configured.

Configuring an outer VLAN tag adding action

About this task

In some scenarios, the device receives packets with different CVLAN IDs on different interfaces. However, the data monitoring device (the server) must receive packets with the same SVLAN ID, as shown in Figure 6. This feature is used in such scenarios.

Figure 6 Network diagram



Procedure

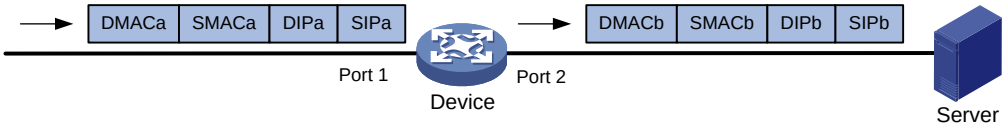
27. Enter system view.
    system-view
28. Create a traffic behavior and enter traffic behavior view.
    traffic behavior *behavior-name*
29. Configure an outer VLAN tag adding action.
    nest top-most vlan *vlan-id* [**dot1p** *802.1p*]
    By default, no outer VLAN tag adding action is configured.
30. Configure an action of redirecting packets to a monitoring group.
    redirect monitoring-group *group-id*
    By default, no action of redirecting packets to a monitoring group is configured.

Configuring an IP or MAC address marking action

About this task

In some scenarios, the data monitoring device (the server) must receive packets with its own address as the destination address or identify the input interfaces of packets based on source addresses. This feature allows the device to modify the IP address and MAC address of packets, as shown in Figure 7.

Figure 7 Network diagram



Procedure

31. Enter system view.
    system-view
32. Create a traffic behavior and enter traffic behavior view.
    traffic behavior *behavior-name*
33. Configure an IP or MAC address marking action. Choose one option as needed:
    - Configure a destination IPv4 address marking action.
      remark destination-ip *ipv4-address*
      By default, no destination IPv4 address marking action is configured.
    - Configure a source IPv4 address marking action.
      remark source-ip *ipv4-address*
      By default, no source IPv4 address marking action is configured.

- Configure a destination IPv6 address marking action.
  remark destination-ipv6 *ipv6-address*
  By default, no destination IPv6 address marking action is configured.
- Configure a source IPv6 address marking action.
  remark source-ipv6 *ipv6-address*
  By default, no source IPv6 address marking action is configured.
- Configure a destination MAC address marking action.
  remark destination-mac *mac-address*
  By default, no destination MAC address marking action is configured.
- Configure a source MAC address marking action.
  remark source-mac *mac-address*
  By default, no source MAC address marking action is configured.

34. Configure an action of redirecting packets to a monitoring group.
    redirect monitoring-group *group-id*
    By default, no action of redirecting packets to a monitoring group is configured.

Configuring a timestamp and Ethernet header adding action

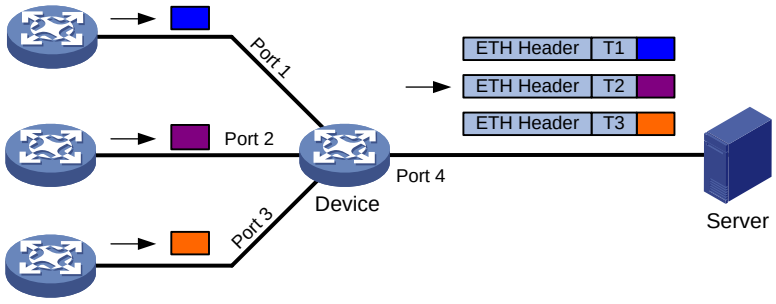
About this task

This feature allows a network to monitor the traffic of its internal devices.

Traditional TAP applications monitor the traffic on the egress device of a network. As the network grows and the performance requirements become higher, monitoring the traffic of the internal network devices is required. This feature is used in such scenarios.

As shown in Figure 8, interfaces on a TAP-enabled device connect to different devices. The device adds a timestamp, an Ethernet header, and a user-defined Ethernet protocol number to packets received on an interface, and redirects the packets to a monitoring group. Then, the device sends the packets to the data monitoring device (server). The server identifies the packets according to the Ethernet protocol number in the **ETH Header** field. The server calculates the network delay according to the timestamps TI, T2, and T3.

Figure 8 Network diagram



Procedure

35. Enter system view.
    system-view
36. Create a traffic behavior and enter traffic behavior view.
    traffic behavior *behavior-name*
37. Configure a timestamp and Ethernet header adding action.

timestamp-over-ether destination-mac *mac-address* **source-mac** *mac-address* **ethtype-id** *ethtype-id*

By default, no timestamp and Ethernet header adding action is configured.

38. Configure an action of redirecting packets to a monitoring group.

    redirect monitoring-group *group-id*

    By default, no action of redirecting packets to a monitoring group is configured.

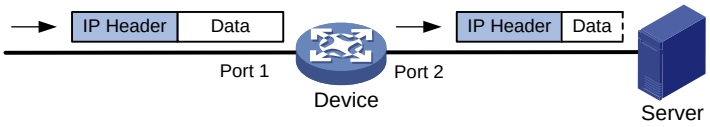
Configuring a packet truncation action

About this task

This feature reduces the packet processing pressure on the data monitoring device (the server) or prevents the device from sending packet payloads to the server.

This feature truncates packets beginning from the Layer 2 header. The CRC is recalculated for the truncated part. The packet length after truncation is the truncated part plus the CRC length.

Figure 9 Network diagram



Procedure

39. Enter system view.

    system-view

40. Create a traffic behavior and enter traffic behavior view.

    traffic behavior *behavior-name*

41. Enable packet truncation.

    truncation enable

    By default, packet truncation is disabled.

42. Configure an action of redirecting packets to a monitoring group.

    redirect monitoring-group *group-id*

    By default, no action of redirecting packets to a monitoring group is configured.

43. Return to system view.

    quit

44. Set the packet length after truncation.

    qos truncation length *length*

    The default setting is 128 bytes.

Configuring a tunnel header stripping action

About this task

In some scenarios, the data monitoring device (the server) cannot parse GRE, NVGRE, or VXLAN packets.

This feature allows the device to strip the tunnel header from the start of the outer Layer 2 header to the specified position.

Figure 10 GRE packet header stripping

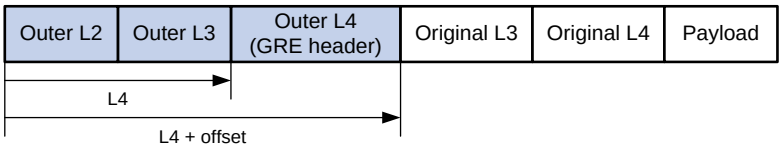


Figure 11 NVGRE packet header stripping

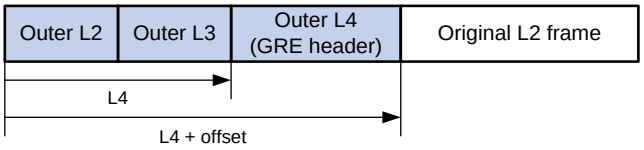


Figure 12 VXLAN packet header stripping

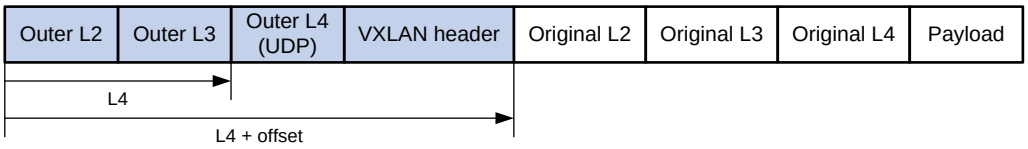
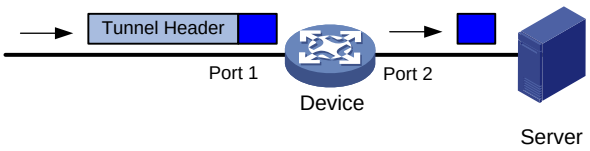


Figure 13 Network diagram for tunnel packet header stripping



Procedure

45. Enter system view.

system-view

46. Create a traffic behavior and enter traffic behavior view.

traffic behavior *behavior-name*

47. Configure a tunnel header stripping action.

- Delete a tunnel header by using the user-defined method.

  strip-header position { **l2** | **l3** | **l4** } [**offset** *offset-value*]

- Delete the GRE header.

  strip-header gre header-length *header-length* **encap-eth-header destination-mac** *mac-address* **source-mac** *mac-address* [**vlan** *vlan-id* [**dot1p** *dot1p-value*]] **ethtype-id** *ethtype-id*

- Delete the NVGRE header.

  strip-header nvgre

- Delete the VXLAN header.

  strip-header vxlan

By default, no tunnel header stripping action is configured.

48. Configure an action of redirecting packets to a monitoring group.

redirect monitoring-group *group-id*

By default, no action of redirecting packets to a monitoring group is configured.

Configuring a TAP policy

49. Enter system view.
    system-view
50. Create a TAP policy and enter TAP policy view.
    qos tap policy *policy-name*
51. Associate a traffic class with a traffic behavior in the TAP policy.
    classifier *classifier-name* **behavior** *behavior-name*
    By default, no traffic behavior is associated with a traffic class.

Applying a TAP policy to the inbound direction of an interface

52. Enter system view.
    system-view
53. Enter interface view.
    interface *interface-type interface-number*
54. Apply a TAP policy to the interface.
    qos apply tap policy *policy-name* **inbound** [**inner-match**]
    By default, no TAP policy is applied to an interface.

Display and maintenance commands for TAP

Execute **display** commands in any view.

Task	Command
Display traffic class configuration.	**display traffic classifier user-defined** [*classifier-name*] [**slot** *slot-number*] For more information about this command, see flow mirroring commands in *ACL and QoS Command Reference*.
Display traffic behavior configuration.	**display traffic behavior user-defined** [*behavior-name*] [**slot** *slot-number*] For more information about this command, see flow mirroring commands in *ACL and QoS Command Reference*.
Display monitoring group information.	**display monitoring-group** { *group-id* | **all** } For more information about this command, see flow mirroring commands in *Network Management and Monitoring Command Reference*.
Display TAP policy configuration.	**display qos policy user-defined tap** [*policy-name* [**classifier** *classifier-name*]] [**slot** *slot-number*]
Display the configuration and statistics of TAP policies applied to interfaces.	**display qos tap policy interface** [*interface-type interface-number* [**inbound**] [**slot** *slot-number*]

TAP configuration examples

Example: Configuring basic M:N copying

Network configuration

As shown in Figure 14, Device D is a TAP device. Configure M:N copying on Device D to forward all packets received on GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 to Server A and Server B through GigabitEthernet 1/0/4 and GigabitEthernet 1/0/5.

Figure 14 Network diagram

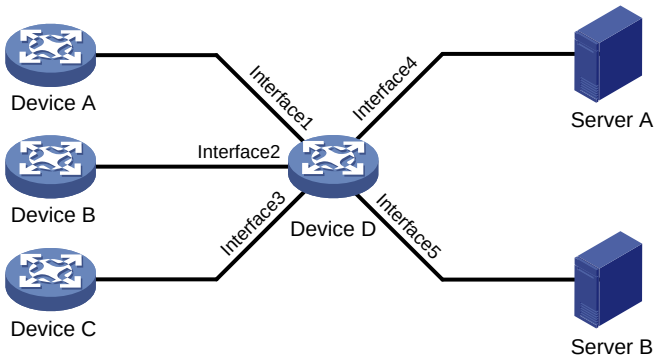


Table 1 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

55. Enable TAP globally.

```
<DeviceD> system-view
[DeviceD] tap enable
```

56. Create monitoring group 1, and assign GigabitEthernet 1/0/4 and GigabitEthernet 1/0/5 to the monitoring group.

```
[DeviceD] monitoring-group 1
[DeviceD-monitoring-group-1] monitoring-port gigabitethernet 1/0/4 to
gigabitethernet 1/0/5
[DeviceD-monitoring-group-1] quit
```

57. Create a traffic class named **classifier_tap**, and configure the traffic class to match all packets.

```
[DeviceD] traffic classifier classifier_tap
[DeviceD-classifier-classifier_tap] if-match any
[DeviceD-classifier-classifier_tap] quit
```

58. Create a traffic behavior named **behavior_tap**, and configure an action of redirecting packets to monitoring group 1.

```
[DeviceD] traffic behavior behavior_tap
```

```
[DeviceD-behavior-behavior_tap] redirect monitoring-group 1
[DeviceD-behavior-behavior_tap] quit
```

59. Create a TAP policy named **policy_tap**, and associate the traffic class with the traffic behavior in the TAP policy.

```
[DeviceD] qos tap policy policy_tap
[DeviceD-qospolicy-policy_tap] classifier classifier_tap behavior behavior_tap
[DeviceD-qospolicy-policy_tap] quit
```

60. Applying the TAP policy.

Applying the TAP policy to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] qos apply tap policy policy_tap inbound
[DeviceD-GigabitEthernet1/0/1] quit
```

Applying the TAP policy to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] qos apply tap policy policy_tap inbound
[DeviceD-GigabitEthernet1/0/2] quit
```

Applying the TAP policy to the inbound direction of GigabitEthernet 1/0/3.

```
[DeviceD] interface gigabitethernet 1/0/3
[DeviceD-GigabitEthernet1/0/3] qos apply tap policy policy_tap inbound
[DeviceD-GigabitEthernet1/0/3] quit
```

Example: Configuring M:N copying for aggregate interfaces

Network configuration

As shown in Figure 15, Device B is a TAP device. Packets from Device A to Device C enter Device B on GigabitEthernet 1/0/1 and are sent to Device C on GigabitEthernet 1/0/2. Packets from Device C to Device A enter Device B on GigabitEthernet 1/0/2 and are sent to Device A on GigabitEthernet 1/0/1.

Configure M:N copying and link aggregation load sharing on Device B to meet the following requirements:

- Incoming packets on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are redirected to the aggregate interface.
- The aggregate interface loads share packets based on source IP addresses and destination IP addresses.
- The aggregate interface sends incoming packets on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to Device D over the same link by using algorithm 1 as the link aggregation load sharing algorithm.

Figure 15 Network diagram

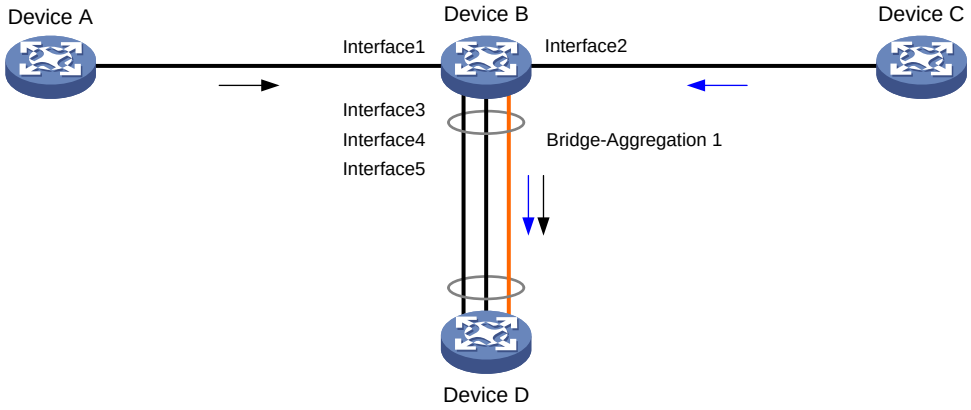


Table 2 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4
Interface5	GigabitEthernet1/0/5

Procedure

61. Enable TAP globally.

```
<DeviceB> system-view
[DeviceB] tap enable
```

62. Configure an aggregate interface:

Create a Layer 2 aggregate interface named Bridge-Aggregation 1.

```
[DeviceB] interface bridge-aggregation 1
[DeviceB-Bridge-Aggregation1] quit
```

Assign GigabitEthernet 1/0/3, GigabitEthernet 1/0/4, and GigabitEthernet 1/0/5 to the aggregation group.

```
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] port link-aggregation group 1
[DeviceB-GigabitEthernet1/0/3] quit
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] port link-aggregation group 1
[DeviceB-GigabitEthernet1/0/4] quit
[DeviceB] interface gigabitethernet 1/0/5
[DeviceB-GigabitEthernet1/0/5] port link-aggregation group 1
[DeviceB-GigabitEthernet1/0/5] quit
```

NOTE:

You can use the **port link-mode** command to configure an Ethernet port as a Layer 2 or Layer 3 interface (see *Layer 2—LAN Switching Configuration Guide*).

63. Configure a monitoring group:

Create monitoring group 1, and assign the Layer 2 aggregate interface to the monitoring group.

```
[DeviceB] monitoring-group 1
[DeviceB-monitoring-group-1] monitoring-port bridge-aggregation 1
[DeviceB-monitoring-group-1] quit
```

64. Configure traffic classes:

Create IPv4 advanced ACL 3001, and configure a rule to match packets with source IP address 192.168.1.1 and destination IP address 192.168.2.1.

```
[DeviceB] acl advanced 3001
[DeviceB-acl-ipv4-adv-3001] rule permit ip source 192.168.1.1 0 destination
192.168.2.1 0
[DeviceB-acl-ipv4-adv-3001] quit
```

Create a traffic class named **classifier1**, and use ACL 3001 as the match criterion.

```
[DeviceB] traffic classifier classifier1
[DeviceB-classifier-classifier1] if-match acl 3001
[DeviceB-classifier-classifier1] quit
```

Create IPv4 advanced ACL 3002, and configure a rule to match packets with source IP address 192.168.2.1 and destination IP address 192.168.1.1.

```
[DeviceB] acl advanced 3002
[DeviceB-acl-ipv4-adv-3002] rule permit ip source 192.168.2.1 0 destination
192.168.1.1 0
[DeviceB-acl-ipv4-adv-3002] quit
```

Create a traffic class named **classifier2**, and use ACL 3002 as the match criterion.

```
[DeviceB] traffic classifier classifier2
[DeviceB-classifier-classifier2] if-match acl 3002
[DeviceB-classifier-classifier2] quit
```

65. Configure a traffic behavior:

Create a traffic behavior named **behavior1**, and configure an action of redirecting packets to monitoring group 1.

```
[DeviceD] traffic behavior behavior_tap
[DeviceD-behavior-behavior_tap] redirect monitoring-group 1
[DeviceD-behavior-behavior_tap] quit
```

66. Configure TAP policies:

Create a TAP policy named **policy1**, and associate traffic class **classifier1** with traffic behavior **behavior1** in the TAP policy.

```
[DeviceB] qos tap policy policy1
[DeviceB-qospolicy-policy1] classifier classifier1 behavior behavior1
[DeviceB-qospolicy-policy1] quit
```

Create a TAP policy named **policy2**, and associate traffic class **classifier2** with traffic behavior **behavior1** in the TAP policy.

```
[DeviceB] qos tap policy policy2
[DeviceB-qospolicy-policy2] classifier classifier2 behavior behavior1
[DeviceB-qospolicy-policy2] quit
```

67. Applying TAP policies:

Applying TAP policy **policy1** to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceD] interface gigabitethernet 1/0/1
```

```
[DeviceD-GigabitEthernet1/0/1] qos apply tap policy policy1 inbound
[DeviceD-GigabitEthernet1/0/1] quit
```

Applying TAP policy **policy2** to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] qos apply tap policy policy2 inbound
[DeviceD-GigabitEthernet1/0/2] quit
```

68. Set the global load sharing mode to load share packets based on source IP addresses and destination IP addresses.

```
[DeviceB] link-aggregation global load-sharing mode source-ip destination-ip
```

69. Specify algorithm 1 as the link aggregation load sharing algorithm.

```
[DeviceB] link-aggregation global load-sharing algorithm 1
```

Example: Configuring VLAN ID marking

Network configuration

As shown in Figure 16, Device C is a TAP device. Configure VLAN ID marking on Device C to meet the following requirements:

- Mark CVLAN 10 for packets received on GigabitEthernet 1/0/1.
- Mark CVLAN 20 for packets received on GigabitEthernet 1/0/2.
- Forward packets with CVLAN 10 and packets with CVLAN 20 to the server through GigabitEthernet 1/0/3, enabling the server to identify the input interfaces of packets according to VLAN IDs.

Figure 16 Network diagram

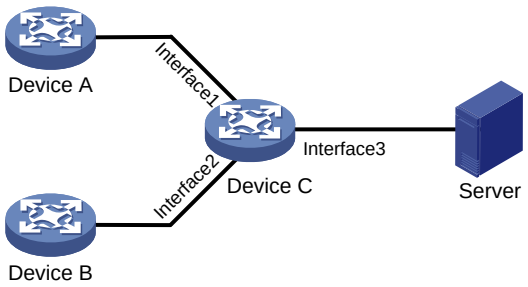


Table 3 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

70. Enable TAP globally.

```
<DeviceC> system-view
[DeviceC] tap enable
```

71. Create monitoring group 1, and assign GigabitEthernet 1/0/3 to the monitoring group.

```
[DeviceC] monitoring-group 1
[DeviceC-monitoring-group-1] monitoring-port gigabitethernet 1/0/3
```

```
[DeviceC-monitoring-group-1] quit
```

72. Create a traffic class named **classifier_tap**, and configure the traffic class to match all packets.

```
[DeviceC] traffic classifier classifier_tap
[DeviceC-classifier-classifier_tap] if-match any
[DeviceC-classifier-classifier_tap] quit
```

73. Configure traffic behaviors:

Create a traffic behavior named **behavior_tap1**, and configure an action of marking CVLAN 10 and an action of redirecting packets to monitoring group 1.

```
[DeviceC] traffic behavior behavior_tap1
[DeviceC-behavior-behavior_tap1] remark customer-vlan-id 10
[DeviceC-behavior-behavior_tap1] redirect monitoring-group 1
[DeviceC-behavior-behavior_tap1] quit
```

Create a traffic behavior named **behavior_tap2**, and configure an action of marking CVLAN 20 and an action of redirecting packets to monitoring group 1.

```
[DeviceC] traffic behavior behavior_tap2
[DeviceC-behavior-behavior_tap2] remark customer-vlan-id 20
[DeviceC-behavior-behavior_tap2] redirect monitoring-group 1
[DeviceC-behavior-behavior_tap2] quit
```

74. Configure TAP policies:

Create a TAP policy named **policy_tap1**, and associate traffic class **classifier_tap** with traffic behavior **behavior_tap1** in the TAP policy.

```
[DeviceC] qos tap policy policy_tap1
[DeviceC-qospolicy-policy_tap1] classifier classifier_tap behavior behavior_tap1
[DeviceC-qospolicy-policy_tap1] quit
```

Create a TAP policy named **policy_tap2**, and associate traffic class **classifier_tap** with traffic behavior **behavior_tap2** in the TAP policy.

```
[DeviceC] qos tap policy policy_tap2
[DeviceC-qospolicy-policy_tap2] classifier classifier_tap behavior behavior_tap2
[DeviceC-qospolicy-policy_tap2] quit
```

75. Apply the TAP policies.

Apply TAP policy **policy_tap1** to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] qos apply tap policy policy_tap1 inbound
[DeviceC-GigabitEthernet1/0/1] quit
```

Apply TAP policy **policy_tap2** to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] qos apply tap policy policy_tap2 inbound
[DeviceC-GigabitEthernet1/0/2] quit
```

Example: Configuring outer VLAN tag deleting

Network configuration

As shown in Figure 17, Device C is a TAP device, and GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 receive double-tagged packets.

Configure Device C to meet the following requirements:

- Delete the outer VLAN tag of the packets received on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

- Forward the packets to the server through GigabitEthernet 1/0/3.

Figure 17 Network diagram

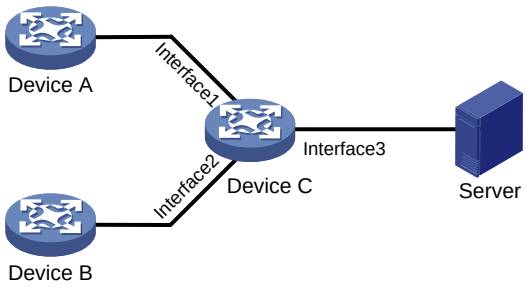


Table 4 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

76. Enable TAP globally.

```
<DeviceC> system-view
[DeviceC] tap enable
```

77. Create monitoring group 1, and assign GigabitEthernet 1/0/3 to the monitoring group.

```
[DeviceC] monitoring-group 1
[DeviceC-monitoring-group-1] monitoring-port gigabitethernet 1/0/3
[DeviceC-monitoring-group-1] quit
```

78. Create a traffic class named **classifier_tap**, and configure the traffic class to match all packets.

```
[DeviceC] traffic classifier classifier_tap
[DeviceC-classifier-classifier_tap] if-match any
[DeviceC-classifier-classifier_tap] quit
```

79. Create a traffic behavior named **behavior_tap**, and configure an outer VLAN tag deleting action and an action of redirecting packets to monitoring group 1.

```
[DeviceC] traffic behavior behavior_tap
[DeviceC-behavior-behavior_tap] strip-header top-most-vlan
[DeviceC-behavior-behavior_tap] redirect monitoring-group 1
[DeviceC-behavior-behavior_tap] quit
```

80. Create a TAP policy named **policy_tap**, and associate the traffic class with the traffic behavior in the TAP policy.

```
[DeviceC] qos tap policy policy_tap
[DeviceC-qospolicy-policy_tap] classifier classifier_tap behavior behavior_tap
[DeviceC-qospolicy-policy_tap] quit
```

81. Apply the TAP policy.

Apply TAP policy **policy_tap** to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] qos apply tap policy policy_tap inbound
[DeviceC-GigabitEthernet1/0/1] quit
```

Apply TAP policy **policy_tap** to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] qos apply tap policy policy_tap inbound
[DeviceC-GigabitEthernet1/0/2] quit
```

Example: Configuring outer VLAN tag adding

Network configuration

As shown in Figure 18, Device C is a TAP device, and GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 receive single-tagged packets.

Configure Device C to meet the following requirements:

- Add an outer VLAN tag for the packets received on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.
- Forward the packets to the server through GigabitEthernet 1/0/3.

Figure 18 Network diagram

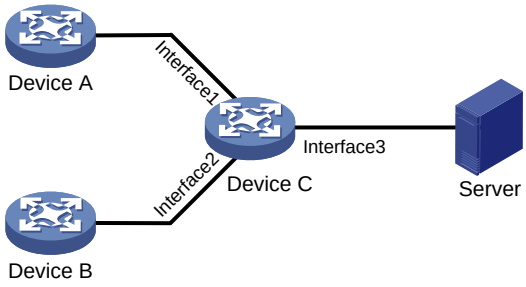


Table 5 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2
Interface3	GigabitEthernet1/0/3

Procedure

82. Enable TAP globally.

```
<DeviceC> system-view
[DeviceC] tap enable
```

83. Create monitoring group 1, and assign GigabitEthernet 1/0/3 to the monitoring group.

```
[DeviceC] monitoring-group 1
[DeviceC-monitoring-group-1] monitoring-port gigabitethernet 1/0/3
[DeviceC-monitoring-group-1] quit
```

84. Create a traffic class named **classifier_tap**, and configure the traffic class to match all packets.

```
[DeviceC] traffic classifier classifier_tap
[DeviceC-classifier-classifier_tap] if-match any
[DeviceC-classifier-classifier_tap] quit
```

85. Create a traffic behavior named **behavior_tap**, and configure an outer VLAN tag adding action and an action of redirecting packets to monitoring group 1.

```
[DeviceC] traffic behavior behavior_tap
[DeviceC-behavior-behavior_tap] nest top-most vlan 10
```

```
[DeviceC-behavior-behavior_tap] redirect monitoring-group 1
[DeviceC-behavior-behavior_tap] quit
```

86. Create a TAP policy named **policy_tap**, and associate the traffic class with the traffic behavior in the TAP policy.

```
[DeviceC] qos tap policy policy_tap
[DeviceC-qospolicy-policy_tap] classifier classifier_tap behavior behavior_tap
[DeviceC-qospolicy-policy_tap] quit
```

87. Apply the TAP policy.

Apply TAP policy **policy_tap** to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] qos apply tap policy policy_tap inbound
[DeviceC-GigabitEthernet1/0/1] quit
```

Apply TAP policy **policy_tap** to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] qos apply tap policy policy_tap inbound
[DeviceC-GigabitEthernet1/0/2] quit
```

Example: Configuring destination MAC address marking

Network configuration

As shown in Figure 19, Device B is a TAP device. Configure Device B to meet the following requirements:

- Modify the destination MAC address of the packets received on GigabitEthernet 1/0/1 as the MAC address of the server.
- Forward the packets to the server through GigabitEthernet 1/0/2.

Figure 19 Network diagram

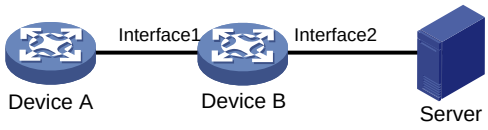


Table 6 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

88. Enable TAP globally.

```
<DeviceB> system-view
[DeviceB] tap enable
```

89. Create monitoring group 1, and assign GigabitEthernet 1/0/2 to the monitoring group.

```
[DeviceB] monitoring-group 1
[DeviceB-monitoring-group-1] monitoring-port gigabitethernet 1/0/2
[DeviceB-monitoring-group-1] quit
```

90. Create a traffic class named **classifier_tap**, and configure the traffic class to match all packets.

```
[DeviceB] traffic classifier classifier_tap
```

```
[DeviceB-classifier-classifier_tap] if-match any
[DeviceB-classifier-classifier_tap] quit
```

91. Create a traffic behavior named **behavior_tap**, and configure a destination MAC address marking action and an action of redirecting packets to monitoring group 1.

```
[DeviceB] traffic behavior behavior_tap
[DeviceB-behavior-behavior_tap] remark destination-mac 0050-ba27-bed3
[DeviceB-behavior-behavior_tap] redirect monitoring-group 1
[DeviceB-behavior-behavior_tap] quit
```

92. Create a TAP policy named **policy_tap**, and associate the traffic class with the traffic behavior in the TAP policy.

```
[DeviceB] qos tap policy policy_tap
[DeviceB-qospolicy-policy_tap] classifier classifier_tap behavior behavior_tap
[DeviceB-qospolicy-policy_tap] quit
```

93. Apply TAP policy **policy_tap** to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] qos apply tap policy policy_tap inbound
[DeviceB-GigabitEthernet1/0/1] quit
```

Example: Configuring timestamp and Ethernet header adding

Network configuration

As shown in Figure 20, Device D is a TAP device. Configure timestamp and Ethernet header adding on Device D to meet the following requirements:

- Add destination MAC address 0050-ba27-bed4, source MAC address 0050-ba27-bed1, and Ethernet protocol number FF for packets received on GigabitEthernet 1/0/1.
- Add destination MAC address 0050-ba27-bed4, source MAC address 0050-ba27-bed2, and Ethernet protocol number FF for packets received on GigabitEthernet 1/0/2.
- Add destination MAC address 0050-ba27-bed4, source MAC address 0050-ba27-bed3, and Ethernet protocol number FF for packets received on GigabitEthernet 1/0/3.
- Forward the packets to the server through GigabitEthernet 1/0/4.

Figure 20 Network diagram

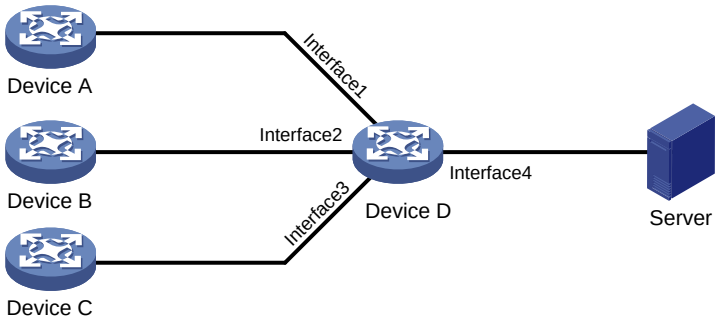


Table 7 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Interface3	GigabitEthernet1/0/3
Interface4	GigabitEthernet1/0/4

Procedure

94. Enable TAP globally.

```
<DeviceD> system-view
[DeviceD] tap enable
```

95. Create monitoring group 1, and assign GigabitEthernet 1/0/4 to the monitoring group.

```
[DeviceD] monitoring-group 1
[DeviceD-monitoring-group-1] monitoring-port gigabitethernet 1/0/4
[DeviceD-monitoring-group-1] quit
```

96. Configure a traffic class.

Create IPv4 advanced ACL 3001, and configure a rule to match all TCP packets.

```
[DeviceD] acl advanced 3001
[DeviceD-acl-ipv4-adv-3001] rule permit tcp
[DeviceD-acl-ipv4-adv-3001] quit
```

Create a traffic class named **classifier_tap**, and use ACL 3001 as the match criterion.

```
[DeviceD] traffic classifier classifier_tap
[DeviceD-classifier-classifier_tap] if-match acl 3001
[DeviceD-classifier-classifier_tap] quit
```

97. Configure traffic behaviors:

Create a traffic behavior named **behavior_tap1**, and configure an action of add destination MAC address 0050-ba27-bed4, source MAC address 0050-ba27-bed1, and Ethernet protocol number FF, and an action of redirecting packets to monitoring group 1.

```
[DeviceD] traffic behavior behavior_tap1
[DeviceD-behavior-behavior_tap1] timestamp-over-ether destination-mac
0050-ba27-bed4 source-mac 0050-ba27-bed1 ethtype-id ff
[DeviceD-behavior-behavior_tap1] redirect monitoring-group 1
[DeviceD-behavior-behavior_tap1] quit
```

Create a traffic behavior named **behavior_tap2**, and configure an action of adding destination MAC address 0050-ba27-bed4, source MAC address 0050-ba27-bed2, and Ethernet protocol number FF, and an action of redirecting packets to monitoring group 1.

```
[DeviceD] traffic behavior behavior_tap2
[DeviceD-behavior-behavior_tap2] timestamp-over-ether destination-mac
0050-ba27-bed4 source-mac 0050-ba27-bed2 ethtype-id ff
[DeviceD-behavior-behavior_tap2] redirect monitoring-group 1
[DeviceD-behavior-behavior_tap2] quit
```

Create a traffic behavior named **behavior_tap3**, and configure an action of adding destination MAC address 0050-ba27-bed4, source MAC address 0050-ba27-bed3, and Ethernet protocol number FF, and an action of redirecting packets to monitoring group 1.

```
[DeviceD] traffic behavior behavior_tap3
[DeviceD-behavior-behavior_tap3] timestamp-over-ether destination-mac
0050-ba27-bed4 source-mac 0050-ba27-bed3 ethtype-id ff
[DeviceD-behavior-behavior_tap3] redirect monitoring-group 1
[DeviceD-behavior-behavior_tap2] quit
```

98. Configure TAP policies:

Create a TAP policy named **policy_tap1**, and associate traffic class **classifier_tap** with traffic behavior **behavior_tap1** in the TAP policy.

```
[DeviceD] qos tap policy policy_tap1
[DeviceD-qospolicy-policy_tap1] classifier classifier_tap behavior behavior_tap1
[DeviceD-qospolicy-policy_tap1] quit
```

Create a TAP policy named **policy_tap2**, and associate traffic class **classifier_tap** with traffic behavior **behavior_tap2** in the TAP policy.

```
[DeviceD] qos tap policy policy_tap2
[DeviceD-qospolicy-policy_tap2] classifier classifier_tap behavior behavior_tap2
[DeviceD-qospolicy-policy_tap2] quit
```

Create a TAP policy named **policy_tap3**, and associate traffic class **classifier_tap** with traffic behavior **behavior_tap3** in the TAP policy.

```
[DeviceD] qos tap policy policy_tap3
[DeviceD-qospolicy-policy_tap3] classifier classifier_tap behavior behavior_tap3
[DeviceD-qospolicy-policy_tap3] quit
```

99. Apply the TAP policies:

Apply TAP policy **policy_tap1** to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] qos apply tap policy policy_tap1 inbound
[DeviceD-GigabitEthernet1/0/1] quit
```

Apply TAP policy **policy_tap2** to the inbound direction of GigabitEthernet 1/0/2.

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] qos apply tap policy policy_tap2 inbound
[DeviceD-GigabitEthernet1/0/2] quit
```

Apply TAP policy **policy_tap3** to the inbound direction of GigabitEthernet 1/0/3.

```
[DeviceD] interface gigabitethernet 1/0/3
[DeviceD-GigabitEthernet1/0/3] qos apply tap policy policy_tap2 inbound
[DeviceD-GigabitEthernet1/0/3] quit
```

Example: Configuring packet truncation

Network configuration

As shown in Figure 21, Device B is a TAP device. Configure Device B to meet the following requirements:

- Truncate the packets received on GigabitEthernet 1/0/1.
- Forward the truncated packets to the server through GigabitEthernet 1/0/2.

Figure 21 Network diagram

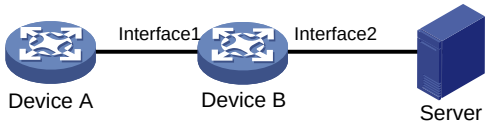


Table 8 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

100. Enable TAP globally.

```
<DeviceB> system-view
[DeviceB] tap enable
```

101. Create monitoring group 1, and assign GigabitEthernet 1/0/2 to the monitoring group.

```
[DeviceB] monitoring-group 1
[DeviceB-monitoring-group-1] monitoring-port gigabitethernet 1/0/2
[DeviceB-monitoring-group-1] quit
```

102. Configure a traffic class.

Create IPv4 advanced ACL 3001, and configure a rule to match TCP packets with source IP address 10.0.0.1 and mask 255.255.255.0.

```
[DeviceB] acl advanced 3001
[DeviceB-acl-ipv4-adv-3001] rule permit tcp source 10.0.0.1 255.255.255.0
[DeviceB-acl-ipv4-adv-3001] quit
```

Create a traffic class named **classifier_tap**, and use ACL 3001 as the match criterion.

```
[DeviceB] traffic classifier classifier_tap
[DeviceB-classifier-classifier_tap] if-match acl 3001
[DeviceB-classifier-classifier_tap] quit
```

103. Create a traffic behavior named **behavior_tap**, and configure a packet truncation enabling action and an action of redirecting packets to monitoring group 1.

```
[DeviceB] traffic behavior behavior_tap
[DeviceB-behavior-behavior_tap] truncation enable
[DeviceB-behavior-behavior_tap] redirect monitoring-group 1
[DeviceB-behavior-behavior_tap] quit
```

104. Configure packet truncation.

Set the packet truncation length to 100 bytes.

```
[DeviceB] qos truncation length 100
```

105. Create a TAP policy named **policy_tap**, and associate the traffic class with the traffic behavior in the TAP policy.

```
[DeviceB] qos tap policy policy_tap
[DeviceB-qospolicy-policy_tap] classifier classifier_tap behavior behavior_tap
[DeviceB-qospolicy-policy_tap] quit
```

106. Apply TAP policy **policy_tap** to the inbound direction of GigabitEthernet 1/0/1.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] qos apply tap policy policy_tap inbound
[DeviceB-GigabitEthernet1/0/1] quit
```

Example: Configuring NVGRE header stripping

Network configuration

As shown in Figure 22, Device B is a TAP device, and the server does not support parsing NVGRE packets.

Configure NVGRE header stripping on GigabitEthernet 1/0/1 and traffic redirecting to GigabitEthernet 1/0/2.

Figure 22 Network diagram

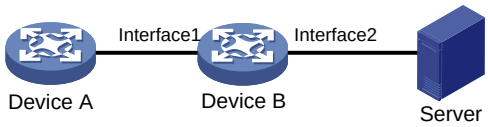


Table 9 Interface label and interface name mappings

Interface label	Interface name
Interface1	GigabitEthernet1/0/1
Interface2	GigabitEthernet1/0/2

Procedure

107. Enable TAP globally.

```
<DeviceB> system-view
[DeviceB] tap enable
```

108. Create monitoring group 1, and assign GigabitEthernet 1/0/2 to the monitoring group.

```
[DeviceB] monitoring-group 1
[DeviceB-monitoring-group-1] monitoring-port gigabitethernet 1/0/2
[DeviceB-monitoring-group-1] quit
```

109. Create a traffic class named **classifier_tap**, and configure the traffic class to match NVGRE packets with VSID FFFF28.

```
[DeviceB] traffic classifier classifier_tap
[DeviceB-classifier-classifier_tap] if-match nvgre-vsid ffff28
[DeviceB-classifier-classifier_tap] quit
```

110. Create a traffic behavior named **behavior_tap**, and configure an action of stripping the outer Layer 2 header, outer Layer 3 header, and the GRE header of NVGRE packets, and an action of redirecting packets to monitoring group 1.

```
[DeviceB] traffic behavior behavior_tap
[DeviceB-behavior-behavior_tap] strip-header position l4 offset 4
[DeviceB-behavior-behavior_tap] redirect monitoring-group 1
[DeviceB-behavior-behavior_tap] quit
```

111. Create a TAP policy named **policy_tap**, and associate the traffic class with the traffic behavior in the TAP policy.

```
[DeviceB] qos tap policy policy_tap
[DeviceB-qospolicy-policy_tap] classifier classifier_tap behavior behavior_tap
[DeviceB-qospolicy-policy_tap] quit
```

112. Apply TAP policy **policy_tap** to the inbound direction of GigabitEthernet 1/0/1.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] qos apply tap policy policy_tap inbound
[DeviceB-GigabitEthernet1/0/1] quit
```

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Network Slicing Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the Network Slicing fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring network slicing 1

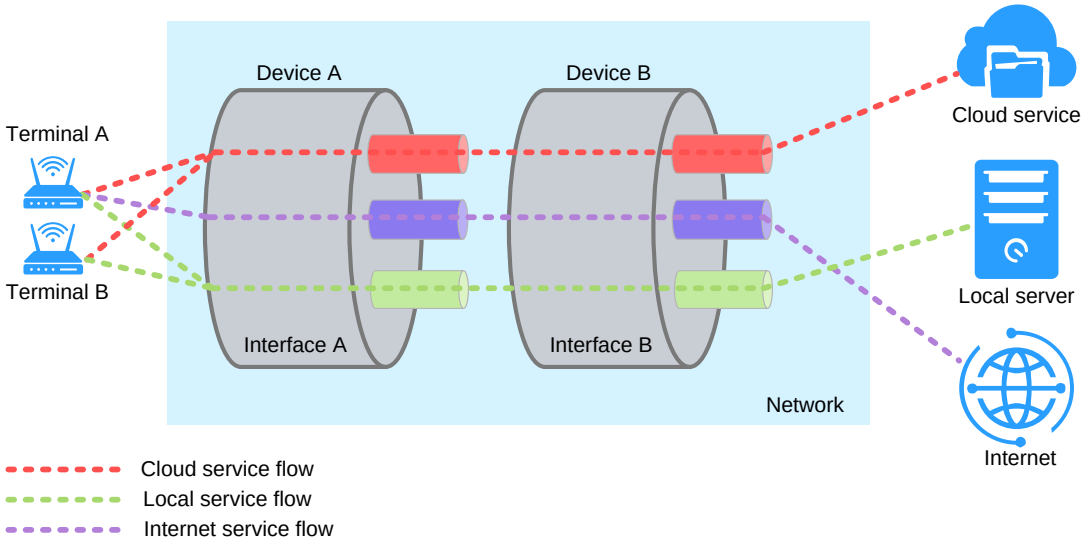
About network slicing 1
Basic concepts 1
Operating mechanism 2
Network slicing tasks at a glance 3
Creating an NSI and configuring parameters 3
Configuring a network slice channel on an interface 4
Configuring bindings between packets and NSI IDs 5
Including NSI IDs in the reserved field of VXLAN packet headers 6
Configuring network slice packet statistics 7
Display and maintenance commands for network slicing 8

Configuring network slicing

About network slicing

In the era of IoT, with a massive amount of devices and various applications connected to the network, home and industry users have diversified demands. A single network or unified policy can no longer meet the evolving demands of new services and scenarios. Building separate networks for each scenario is costly and time-consuming. The network slicing technology allows partitioning a physical network into multiple independent virtual networks. It enables on-demand resource allocation for different services and provides differentiated queue scheduling capabilities, meeting various business demands without affecting existing networks.

Figure 1 Network slicing architecture



Basic concepts

Network slicing has the following concepts:

- Network slice instance (NSI)—An independent virtual network in a network with network slicing deployed. A virtual network is identified by its unique NSI ID (also called slice ID).
- Network slice packet—Based on DSCP/VLAN/VXLAN information, you can bind common packets to specific NSI IDs for transmission. The packets can then be forwarded through the associated NSIs, which are known as network slice packets.
- Network slice channel—A logical channel on an interface, used for forwarding network slice packets. A network slice channel is associated with an NSI by the NSI ID. You can create multiple network slice channels on an interface and assign an independent scheduling queue to

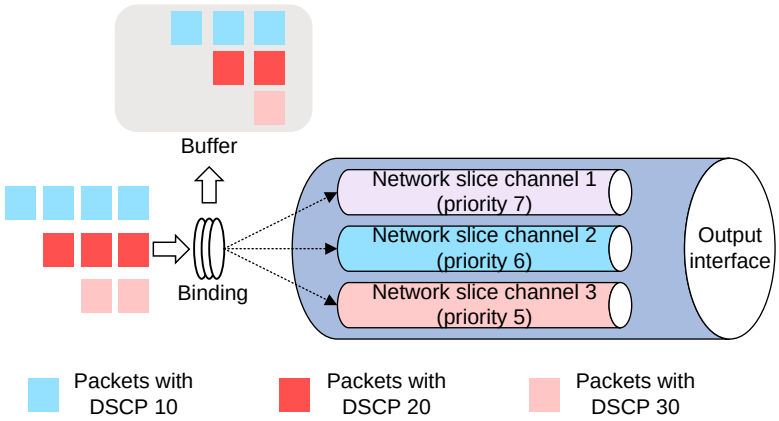
each network slice channel. In this way, the scheduling queues for network slice channels do not affect each other.

Operating mechanism

Figure 2 shows the specific operating mechanism for network slicing as follows:

1. The device maps packets to a network slice channel on the output interface based on the configured bindings between packets and NSI IDs. In the current software version, you can bind packets to NSI IDs based on only the DSCP values, VLAN IDs, VXLAN IDs, input interfaces, and the VSI information of packets. In this example, packets with DSCP 10 are mapped to network slice channel 2. Packets with DSCP 20 are mapped to network slice channel 3. Packets with DSCP 30 are mapped to network slice channel 1.
2. Before being forwarded through a network slice channel on the output interface, the packets first passes through the buffer in the outbound direction. If network congestion occurs in the outbound direction, the buffer helps prevent packet loss. For packets bound to different NSI IDs, you can adjust their buffer classes. A higher buffer class allows the packets in the network slice channel to occupy more buffer space.
3. You can specify guaranteed bandwidth for each network slice channel on the interface. The network slice channels can use different queues for scheduling without affecting each other, ensuring service bandwidth resources.
4. When forwarding packets through network slice channels, the device performs strict priority-based scheduling according to the priority of the network slice channels. If network congestion occurs on the output interface, packets in higher-priority network slice channels are forwarded first. Packets in lower-priority network slice channels can be scheduled in sequence only after all packets from the high-priority network slice channels are forwarded out of the queue.

Figure 2 Network slice packet forwarding process



Network slicing tasks at a glance

To configure network slicing, perform the following tasks:

1. Creating an NSI and configuring parameters
2. Configuring a network slice channel on an interface
3. Configuring bindings between packets and NSI IDs
4. (Optional.) Including NSI IDs in the reserved field of VXLAN packet headers
5. (Optional.) Configuring network slice packet statistics

Creating an NSI and configuring parameters

Restrictions and guidelines

If an NSI is associated with an interface through the **slice-id flex-channel** command, use the **undo slice-id flex-channel** command to cancel the association before you delete the NSI.

When multiple network slice channels exist on the output interface, the device performs strict priority-based scheduling according to the scheduling priority of the network slice channels. If network congestion occurs on the output interface, packets in higher-priority network slice channels are forwarded first. Packets in lower-priority network slice channels can be scheduled in sequence only after all packets from the high-priority network slice channels are forwarded out of the queue.

Procedure

1. Enter system view.

   system-view

2. Enable network slicing and enter network slice view.

   network-slice

   By default, network slicing is disabled.

3. Create an NSI and enter its view.

   instance *slice-instance-id*

4. (Optional.) Configure a description for the NSI.

   description *text*

   By default, no description is configured for an NSI.

5. (Optional.) Configure the queue buffer class for the network slice channel.

   queue-buffer-class *number*

   By default, the queue buffer class for the network slice channel is 7.

Configuring a network slice channel on an interface

About this task

After enabling network slicing on an interface, you can create network slice channels on the interface. The device allocates independent scheduling queue resources for network slice channels on the interface. Packets matching an NSI ID are forwarded through the associated network slice channel. The bandwidth specified for the network slice channel created on the interface serves as both the guaranteed bandwidth and the upper limit for the channel. Traffic exceeding this bandwidth is processed by traffic shaping for buffering.

The device supports the following types of network slice channels:

- **Network slice channels in exclusive mode**—When you configure a network slice channel on an interface, specify the **flex-channel** keyword without the **shared** keyword. The device will allocate exclusive-mode scheduling queue resources to this network slice channel. Network slice packets from other network slice channels and non-network slice packets cannot preempt the scheduling queue resources of this network slice channel.
- **Network slice channels in shared mode**—When you configure a network slice channel on an interface, specify the **flex-channel** keyword with the **shared** keyword. The device will assign shared-mode scheduling queue resources to this network slice channel. Non-network slice packets and network slice packets from other network slice channels can compete for the scheduling queue resources of this network slice channel during network congestion. Network slice channels in shared mode enable more flexible traffic scheduling.

Restrictions and guidelines

Before enabling network slicing on an interface, you must use the **instance** command to create an NSI.

You can configure multiple network slice channels for an interface.

If you configure the **network-slice enable** command to enable network slicing for the interface, you cannot add the interface to a link aggregate group.

This feature is supported only on physical interfaces and Layer 2 aggregate interfaces. In the current software version, this feature is not supported by Layer 3 aggregate interfaces and their member ports (including the associated Layer 3 aggregate subinterfaces).

Procedure

1. Enter system view.

   system-view

2. Enter interface view.

   interface *interface-type interface-number*

3. Enable network slicing for the interface and enter network slice view of the interface.

   network-slice enable

   By default, network slicing is disabled on the interface.

4. Create a network slice channel for the interface, and specify the bandwidth for the network slice channel.

   slice-id *slice-instance-id* **flex-channel** *flex-channel-value* [**shared**]

   By default, no network slice channels are configured on the interface.

Configuring bindings between packets and NSI IDs

About this task

With this feature configured, the device binds the specified packets to NSI IDs. When forwarding the packets through an output interface, the device searches for the network slice channel configured on the output interface. The device then schedules and forwards the packets through the network slice channel identified by the bound NSI ID. If no network slice channel identified by the bound NSI ID exists on the output interface, the device forwards the packets according to routing table lookup.

You can repeatedly execute the **index binding-type** command to configure multiple bindings between packets and NSI IDs identified by different *index-value* values. Upon receiving a packet, the system compares it with packet match criteria in ascending order of index values specified with the **index binding-type** command. If a match is found, the system does not continue to compare the packet with other bindings.

Restrictions and guidelines

If you repeatedly execute the **index binding-type** command and specify the same index value, the most recent configuration takes effect.

If you bind a DSCP/VLAN/VXLAN value or interface/VSI to an NSI ID, you cannot bind the DSCP/VLAN/VXLAN value or interface/VSI to any other NSI IDs.

In the current software version, only Layer 2 Ethernet interfaces, Layer 2 aggregate interfaces, and Layer 3 Ethernet interfaces support scheduling and forwarding the packets through the network slice channel identified by the bound NSI ID.

For a queue-mode network slice, if a queue on the interface has been configured as a network slice channel, a binding rule exists between packets with that slice ID and the NSI ID. If a packet in that queue does not match any slice binding rule, the local priority for that packet is changed to 0, and the packet is forwarded through queue 0.

Procedure

1. Enter system view.

system-view

2. Enable network slicing and enter network slice view.

   network-slice

   By default, network slicing is disabled.

3. Create and enter NSI-to-packet mapping view.

   nsi-mapping

4. Configure bindings between packets and NSI IDs.

   index *index-value* **binding-type** { **dscp** { *dscp-startvalue* [**to** *dscp-endvalue*] } &<1-32> | **interface** *interface-type interface-number* | **vlan** { *vlan-startvalue* [**to** *vlan-endvalue*] } &<1-32> | **vsi-name** *vsi-name* | **vxlan-list** { *vxlan-startvalue* [**to** *vxlan-endvalue*] } &<1-32> } **slice** *slice-id* [**replace dscp** *dscp-value*]

   By default, the bindings between packets and NSI IDs are not configured. That is, no packets are scheduled or forwarded through the network slice channels identified by NSI IDs on an interface.

Including NSI IDs in the reserved field of VXLAN packet headers

About this task

You can enable this feature on VTEPs and non-VTEPs in a VXLAN network to carry different service packets through network slices and achieve differentiated scheduling for various services.

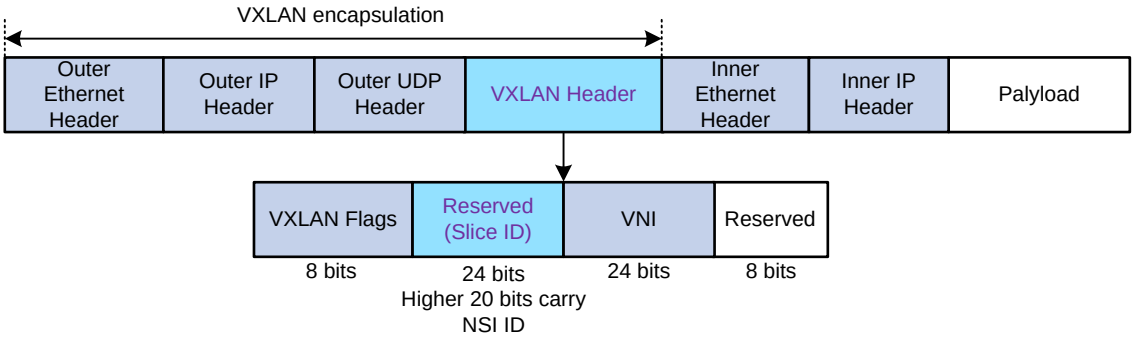
- Enable this feature on a VTEP:

  When the VTEP receives the original packet, it first matches the packet's characteristics with the bindings configured with the **index binding-type** command to associate the original packet with the NSI ID. Then, it encapsulates the packet with a VXLAN header. During VXLAN packet encapsulation, the VTEP uses the higher 20 bits of the Reserved field in the VXLAN header to carry the NSI ID. On the output interface, the device schedules and forwards the packet based on the NSI ID through the associated network slice channel.

- Enable this feature on a non-VTEP in a VXLAN network:

  The non-VTEP can read the NSI ID from the VXLAN header in a received packet. On the output interface, the device schedules and forwards the packet based on the NSI ID through the associated network slice channel.

Figure 3 VXLAN header that carries an NSI ID



Restrictions and guidelines

If you configure both the **index binding-type** and **vxlan reserved enable** commands on a non-VTEP in a VXLAN network, follow these restrictions and guidelines: The device first compares the packet with the bindings configured with the **index binding-type** command. If a matching NSI ID is found, the device schedules the packet on the output interface based on the NSI ID. If no matching NSI ID is found, the device reads the NSI ID from the VXLAN header of the packet, and then schedules the packet on the output interface based on the NSI ID.

If the **index binding-type** command is not configured on the non-VTEP in a VXLAN network, you can use this feature as a supplementary measure to obtain the NSI ID from packets.

For the same VXLAN packet, all devices in the VXLAN network will obtain the same NSI ID for configuring unified scheduling mechanism.

Procedure

1. Enter system view.

   system-view

2. Enter network slice view.

   network-slice

3. Include NSI IDs in the reserved field of VXLAN packet headers.

   vxlan reserved enable

   By default, no NSI IDs are not included in the reserved field of VXLAN packet headers.

Configuring network slice packet statistics

About this task

With network slice packet statistics enabled, you can use the **display network-slice statistics** command to display statistics of packets forwarded through the specified network slice.

Restrictions and guidelines

After you create a mapping rule by executing the **index binding-type** command in NSI-to-packet mapping view, packets can be assigned to the configured slice. At this point, you must execute the **slice-id flex-channel** command for the interface to configure the network slice channel for the interface, so that packets can be counted for the corresponding network slice channel. Without the configuration, no slice channel is configured, and the network slice traffic data cannot be collected.

Procedure

1. Enter system view.

   system-view

2. Enter network slice view.

   network-slice

3. Enable network slice packet statistics.

   statistics enable

   By default, network slice packet statistics is disabled.

4. (Optional.) configure the interval for collecting network slice packet statistics.

   statistics interval *time*

   By default, the device collects network slice packet statistics at 30-second intervals.

Display and maintenance commands for network slicing

Execute **display** commands in any view and **reset** commands in user view.

Task	Command
Display network slice configuration information.	**display network-slice instance** [**slice-id** *slice-id*]
Display the bandwidth usage of network slices on interfaces.	**display network-slice bandwidth usage statistics** [**interface** *interface-type interface-number*]
Display the associations between NSIs and interfaces.	**display network-slice binding-list** [**slice-id** *slice-instance-id*] [**interface** *interface-type interface-number*]
Display network slice packet statistics.	**display network-slice statistics** [**slice-id** *slice-instance-id*] [**interface** *interface-type interface-number*] [**verbose**]
Clear network slice packet statistics.	**reset network-slice statistics** [**slice-id** *slice-instance-id*] [**interface** *interface-type interface-number*]

H3C S5590-HI[S5590-EI][S5500V3-HI] Switch Series

Intelligent Lossless Network Configuration Guide

New H3C Technologies Co., Ltd.
http://www.h3c.com

Software version: Release 8336Pxx
Document version: 6W101-20260803

Copyright © 2026, New H3C Technologies Co., Ltd. and its licensors

All rights reserved

No part of this manual may be reproduced or transmitted in any form or by any means without prior written consent of New H3C Technologies Co., Ltd.

Trademarks

Except for the trademarks of New H3C Technologies Co., Ltd., any trademarks that may be mentioned in this document are the property of their respective owners.

Notice

The information in this document is subject to change without notice. All contents in this document, including statements, information, and recommendations, are believed to be accurate, but they are presented without warranty of any kind, express or implied. H3C shall not be liable for technical or editorial errors or omissions contained herein.

Preface

This configuration guide describes the Intelligent Lossless Network fundamentals and configuration procedures.

This preface includes the following topics about the documentation:

- Audience.
- Conventions.
- Documentation feedback.

Audience

This documentation is intended for:

- Network planners.
- Field technical support and servicing engineers.
- Network administrators working with the S5570S-EI and S5500V3-SI switch series.

Conventions

The following information describes the conventions used in the documentation.

Command conventions

Convention	Description
Boldface	**Bold** text represents commands and keywords that you enter literally as shown.
Italic	*Italic* text represents arguments that you replace with actual values.
[ ]	Square brackets enclose syntax choices (keywords or arguments) that are optional.
{ x | y | ... }	Braces enclose a set of required syntax choices separated by vertical bars, from which you select one.
[x | y | ...]	Square brackets enclose a set of optional syntax choices separated by vertical bars, from which you select one or none.
{ x | y | ... } *	Asterisk marked braces enclose a set of required syntax choices separated by vertical bars, from which you select a minimum of one.
[x | y | ...] *	Asterisk marked square brackets enclose optional syntax choices separated by vertical bars, from which you select one choice, multiple choices, or none.
&<1-n>	The argument or keyword and argument combination before the ampersand (&) sign can be entered 1 to n times.
#	A line that starts with a pound (#) sign is comments.

GUI conventions

Convention	Description
Boldface	Window names, button names, field names, and menu items are in Boldface. For example, the **New User** window opens; click **OK**.
>	Multi-level menus are separated by angle brackets. For example, **File** > **Create** > **Folder**.

Symbols

Convention	Description
WARNING!	An alert that calls attention to important information that if not understood or followed can result in personal injury.
CAUTION:	An alert that calls attention to important information that if not understood or followed can result in data loss, data corruption, or damage to hardware or software.
IMPORTANT:	An alert that calls attention to essential information.
NOTE:	An alert that contains additional or supplementary information.
TIP:	An alert that provides helpful information.

Network topology icons

Convention	Description
	Represents a generic network device, such as a router, switch, or firewall.
	Represents a routing-capable device, such as a router or Layer 3 switch.
	Represents a generic switch, such as a Layer 2 or Layer 3 switch, or a router that supports Layer 2 forwarding and other Layer 2 features.
	Represents an access controller, a unified wired-WLAN module, or the access controller engine on a unified wired-WLAN switch.
	Represents an access point.
	Represents a wireless terminator unit.
	Represents a wireless terminator.
	Represents a mesh access point.
	Represents omnidirectional signals.
	Represents directional signals.
	Represents a security product, such as a firewall, UTM, multiservice security gateway, or load balancing device.
	Represents a security module, such as a firewall, load balancing, NetStream, SSL VPN, IPS, or ACG module.

Examples provided in this document

Examples in this document might use devices that differ from your device in hardware model, configuration, or software version. It is normal that the port numbers, sample output, screenshots, and other information in the examples differ from what you have on your device.

Documentation feedback

You can e-mail your comments about product documentation to info@h3c.com.

We appreciate your comments.

Contents

Configuring PFC 1
About PFC 1
Restrictions and guidelines 1
Configuring PFC on interfaces 2
Configuring PFC on an Ethernet interface 2
Configuring the early warning thresholds for PFC packets 2
Verifying and maintaining PFC 3

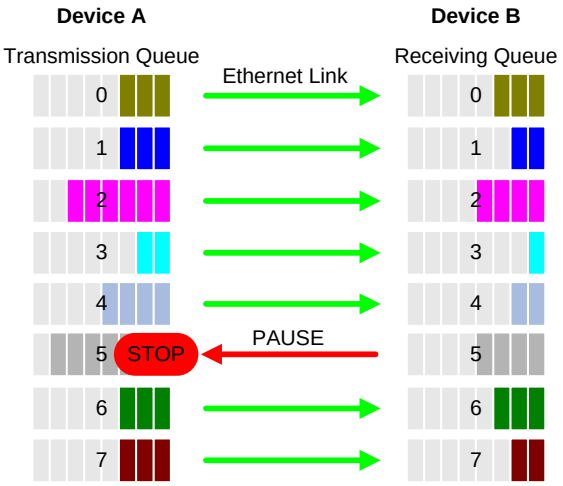
Configuring PFC

About PFC

Priority-based flow control (PFC) provides a finer flow control mechanism to implement lossless packet transmission on Ethernet.

PFC performs flow control for packets based on the 802.1p priorities carried in packets. As shown in Figure 1, PFC establishes eight virtual channels over an Ethernet link, each corresponding to an 802.1p priority. Any virtual channel can be paused or restarted independent of the other channels. This mechanism allows multiple types of traffic to coexist on and share an Ethernet link.

Figure 1 How PFC works



When congestion occurs on the local end, the device determines how to process received packets based on the 802.1p priorities carried in packets as follows:

- If PFC is enabled for the 802.1p priority carried in a packet, the local end accepts the packet and sends PFC pause frames to notify the remote end to stop sending packets carrying the 802.1p priority. The remote end stops sending packets carrying the 802.1p priority after receiving the PFC pause frames. This process is repeated until congestion is eliminated.
- If PFC is not enabled for the 802.1p priority carried in a packet, the local end drops the packet.

Restrictions and guidelines

You can configure PFC in system view or Ethernet interface view. When you configure PFC in system view and Ethernet interface view multiple times, the most recent configuration takes effect.

If you do not enable PFC on an interface, the interface can receive but cannot process PFC pause frames. To make PFC take effect, you must enable PFC on both ends.

To avoid packet loss, apply the same PFC configuration to all interfaces that the packets pass through.

Generic flow control and PFC are mutually exclusive on an interface. For more information about generic flow control, see Ethernet interface configuration in *Interface Configuration Guide*.

In an IRF network, follow these restrictions and guidelines:

- For IRF and other protocols to operate correctly, as a best practice, do not enable PFC for 802.1p priority 0, 6, or 7.

- To perform PFC on an IRF port, configure PFC on the IRF port and the IRF physical ports that are bound to the IRF port.

For information about IRF, see *IRF configuration Guide*.

For PFC to take effect in an overlay network, execute the **qos trust tunnel-dscp** command. For information about the overlay network, see *VXLAN Configuration Guide*. For information about the **qos trust tunnel-dscp** command, see *ACL and QoS Command Reference.*

Configuring PFC on interfaces

Configuring PFC on an Ethernet interface

1. Enter system view.
   system-view
2. Enable PFC on all Ethernet interfaces.
   priority-flow-control { **auto** | **enable** }
   By default, PFC is disabled on all Ethernet interfaces.
3. Enable PFC for 802.1p priorities on all Ethernet interfaces.
   priority-flow-control no-drop dot1p *dot1p-list*
   By default, PFC is disabled for all 802.1p priorities on all Ethernet interfaces.
4. Enter Ethernet interface view.
   interface *interface-type interface-number*
5. Enable PFC on the Ethernet interface.
   priority-flow-control { **auto** | **enable** }
   By default, PFC is disabled.
6. Enable PFC for 802.1p priorities.
   priority-flow-control no-drop dot1p *dot1p-list*
   By default, PFC is disabled for all 802.1p priorities.

Configuring the early warning thresholds for PFC packets

About this task

You can configure the early warning threshold for incoming or outgoing PFC packets of an interface as needed. The early warning threshold notifies a situation where the PFC packet transmission rate is still within a normal range but needs attention.

When the rate of PFC packets that an interface sends or receives reaches the early warning threshold, the system generates traps and logs to notify the user. According to the traps and logs, the user can discover some exceptions in the network, for example:

- The NIC of the peer device fails and continuously sends PFC packets at a high speed. In this case, you can set the early warning threshold for incoming PFC packets.
- The device fails and continuously sends PFC frames. In this case, you can set the early warning threshold for outgoing PFC packets.

To monitor bidirectional PFC packets, you can set the early warning thresholds for incoming packets and outgoing packets separately.

Restrictions and guidelines

The number of PFC pause frames that an interface sends or receives is counted and the early warning threshold configuration takes effect only when PFC is enabled.

Procedure

1. Enter system view.
   system-view
2. Enter Ethernet interface view.
   interface *interface-type interface-number*
3. Configure the early warning threshold for incoming PFC packets.
   priority-flow-control early-warning inpps *pps-value*
   By default, no early warning threshold is configured for incoming PFC packets.
4. Configure the early warning threshold for outgoing PFC packets.
   priority-flow-control early-warning outpps *pps-value*
   By default, no early warning threshold is configured for outgoing PFC packets.

Verifying and maintaining PFC

Execute **display** commands in any view.

Task	Command
Display the PFC information of interfaces.	**display priority-flow-control interface** [*interface-type* [*interface-number*]]

H3C Network Technology Acronyms

A B C D E F G H I K L M N O P Q R S T U V W X Z

10GE	Ten-GigabitEthernet
1DM	One-way Delay Measurement
3DES	Triple Data Encryption Standard
6PE	IPv6 Provider Edge
A	Return
AAA	Authentication, Authorization and Accounting
AAL	ATM Adaptation Layer
AAL5	ATM Adaptation Layer 5
ABR	Area Border Router
AC	Access Controller
AC	Access Circuit
AC	Attachment Circuit
AC	Access Category
AC Name	Access Concentrator Name
ACA	Acquire Change Authorization
ACCM	Async-Control-Character-Map
ACFC	Address-and-Control-Field-Compression
ACFP	Application Control Forwarding Protocol
Ack	Acknowledgment
ACL	Access Control List
ACR	Adaptive Clock Recovery
ACS	Auto-Configuration Server
ACS	AS Control Server
ACSEI	ACFP Client and Server Exchange Information
Active-Stateful PCC	Active-Stateful Path Computation Client
Active-Stateful PCE	Active-Stateful Path Computation Element
AD	Active Directory
ADM	ATM Direct Mapping
ADP	Add Diagnostic Parameter
ADPCM	Adaptive Differential Pulse Code Modulation
ADR	Adaptive Date Rate
ADSL	Asymmetric Digital Subscriber Line
ADVPN	Auto Discovery Virtual Private Network
AER	AS Edge Router
AES	Advanced Encryption Standard

AF	Assured Forwarding
AFI	Address Family Identifier
AFI	Authority and Format Identifier
AFT	Address Family Translation
AFTR	Address Family Transition Router
AH	Authentication Header
AIFSN	Arbitration Inter Frame Spacing Number
AIGP	Accumulated Interior Gateway Protocol Metric
AIS	Alarm Indication Signal
AIW	APPN Implementers Workshop
AKM	Authentication and Key Management
ALG	Application Level Gateway
AM	Analog Modem
AMB	Active Main Board
AMI	Alternate Mark Inversion
A-MPDU	Aggregate MAC Protocol Data Unit
A-MSDU	Aggregate MAC Service Data Unit
ANCP	Access Node Control Protocol
ANI	Adaptive Noise Immunity
ANSI	American National Standards Institute
AP	Access Point
AP	Agreement Protocol
APD	Avalanche Photo Diode
APDB	Access Point Information Database
APoT	AP of things
APP	Application Protocol
APPN	Advanced Peer-to-Peer Networking
APR	Application Recognition
ARP	Address Resolution Protocol
ARQ	Automatic Repeat Request
AS	Autonomous System
ASBR	Autonomous System Boundary Router
ASM	Any-Source Multicast
ASPF	Application Specific Packet Filter
ASPF	Advanced Stateful Packet Filter
AT	AppleTalk
ATD	Attribute Discovery
ATM	Asynchronous Transfer Mode
AU	Administration Unit
AUG	Administration Unit Group

AU-PTR	Administration Unit Pointer
AVF	Active Virtual Forwarder
AVF	Appointed VLAN-x Forwarder
AVP	Attribute Value Pair
B	Return
B4	Basic Bridging BroadBand
B8ZS	Bipolar 8-zero substitution
BA	Block Acknowledgment
BAS	Broadband Access Server
BB_Credit	Buffer-to-Buffer Credit
BC	Boundary Clock
BC	Bandwidth Constraint
BCB	Backbone Core Bridge
BCMP	Broadband-access-network Cluster Management Protocol
BDF	Backup DF
BDI	Backward Defect Indication
BDR	Backup Designated Router
BE	Best Effort
BEB	Backbone Edge Bridge
BECN	Backward Explicit Congestion Notification
BF	Build Fabric
BFD	Bidirectional Forwarding Detection
BGP	Border Gateway Protocol
BGP-EPE	BGP Egress Peer Engineering
BIDIR-PIM	Bidirectional Protocol Independent Multicast
BIMS	Branch Intelligent Management System
BIP	Bit-Interleaved Parity
BIP	Broadcast Integrity Protocol
BITS	Building Integrated Timing Supply System
BLE	Bluetooth Low Energy
B-MAC	Backbone MAC
BMC	Best Master Clock
BMP	BGP Monitoring Protocol
BOOTP	Bootstrap Protocol
BPDU	Bridge Protocol Data Unit
BQ	Bandwidth Queuing
BRAS	Broadband Remote Access Server
BRI	Basic Rate Interface
BRPC	Backward-Recursive PCE-Based Computation

BSB	Blockade State Block
BSM	Bootstrap Message
BSMF	Bootstrap Message Fragment
BSR	Bootstrap Router
BSS	Basic Service Set
BSV	BRI S/T Voice
BT	BitTorrent
B-VLAN	Backbone VLAN
C	Return
CA	Certificate Authority
CA	Connectivity Association
CAC	Connect Admission Control
CAMS	Comprehensive Access Management Server
CAPWAP	Controlling and Provisioning of Wireless Access Point
CAR	Committed Access Rate
CARP	Cache Array Routing Protocol
CB	Controlling Bridge
CB	Customer Bridge
CBQ	Class Based Queuing
CBR	Constant Bit Rate
CBS	Committed Burst Size
C-BSR	Candidate-BSR
CBT	Core-Based Tree
CBTS	Class-of-service Based Tunnel Selection
CBWFQ	Class Based Weighted Fair Queuing
CC	Continuity Check
CC	Call Control
CC	Control Channel
CC	Challenge Collapsar
CC	Common Criteria
CCA	Clear Channel Assessment
CCC	Circuit Cross Connect
CCF	Congestion Controlled Flow
CCM	Continuity Check Message
CD	Computed Distance
C-DCC	Circular DCC
CDCP	S-Channel Discovery and Configuration Protocol
CDP	Cisco Discovery Protocol
CDR	Call Detail Record

CDV	Cell Delay Variation
CE	Customer Edge
CED	Called Station Identifier
CEDCNG	Called Station Identifier
CEE	Converged Enhanced Ethernet
CEM	Circuit Emulation
CEM Class	Circuit Emulation Class
CESoPSN	Circuit Emulation Services over PSN
CFD	Connectivity Fault Detection
CFI	Canonical Format Indicator
CFM	Connectivity Fault Management
CGN	Carrier Grade NAT
CHAP	Challenge Handshake Authentication Protocol
CID	Calling Identity Delivery
CIDR	Classless Inter-Domain Routing
CIR	Committed Information Rate
CIR ALLOW	Committed Information Rate ALLOW
CIST	Common and Internal Spanning Tree
CLC	Channel Link Control
CLI	Command Line Interface
CLI	Calling Line Identification
CLNP	Connection-Less Network Protocol
CLP	Cell Loss Priority
CMC	Call Management Center
CNAME	Canonical Name
CND	Congestion Notification Domain
CNG	Comfortable Noise Generation
CNG	Calling Tone
CNGCED	Calling Tone
CNM	Congestion Notification Message
CNPV	Congestion Notification Priority Value
CO	Central Office
COPS	Common Open Policy Service
CoS	Class of Service
CP	Congestion Point
CP	Control Plane
CPE	Customer Premises Equipment
CPOS	Channelized POS
CPS	Certification Practice Statement
CPTone	Call Progress Tone

CQ	Custom Queuing
CR	Constraint-based Routing
CR	Core Router
CRC	Cyclic Redundancy Check
CRC	Cyclic Redundancy Code
CRC4	Cyclic Redundancy Check 4
CRL	Certificate Revocation List
CRLDP	Constraint-based Routed Label Distribution Protocol
CRLSP	Constraint-based Routed Label Switched Paths
C-RP	Candidate-RP
CS	Class Selector
CSMA/CD	Carrier Sense Multiple Access/Collision Detect
CSNP	Complete Sequence Number PDU
CSP	Port Extender Control and Status Protocol
CSPF	Constraint-based Shortest Path First
CST	Common Spanning Tree
CSV	Comma Separated Value
CT	Class Type
CTR	Counter mode
CTS	Clear to Send
CUG	Closed User Group
CV	Connectivity Verification
CWMP	CPE WAN Management Protocol
D	Return
DAC	Data Analysis Center
DAC	Digital to Analog Converter
DAD	Duplicate Address Detection
DAE	Dynamic Authorization Extensions
DAR	Deeper Application Recognition
Data-MDT	Data-Multicast Distribution Tree
DBA	Dynamic Bandwidth Allocation
DCBX	Data Center Bridging Exchange Protocol
DCC	Dial Control Center
DCD	Data Carrier Detection
DCE	Data Center Ethernet
DCE	Data Circuit-terminating Equipment
DCF	Distributed Coordination Function
DCN	Data Communication Network
DCR	Differential Clock Recovery

DD	Database Description
DDN	Digital Data Network
DDNS	Dynamic Domain Name System
DDoS	Distributed Denial of Service
DDR	Dial-on-Demand Routing
DDT	Delegated Database Tree
DE	Discard Eligibility
DED	Designated Edge Device
Default-MDT	Default-Multicast Distribution Tree
DER	Distinguished Encoding Rules
DF	Designated Forwarder
DGAF	Downstream Group-Addressed Forwarding
DH	Diffie-Hellman
DHCP	Dynamic Host Configuration Protocol
DHCPv6	Dynamic Host Configuration Protocol for IPv6
DID	Direct Inward Dialing
DID	Device ID
DiffServ	Differentiated Service
DIS	Designated IS
DLCI	Data Link Connection Identifier
DLDP	Device Link Detection Protocol
DLSw	Data Link Switching
DM	Diagnostic and Monitoring
DM	Delay Measurement
DMC	Device Management Controller
DMM	Delay Measurement Message
DMR	Delay Measurement Reply
DMZ	Demilitarized Zone
DN	Distinguished Name
DN	Directory Number
DNAT	Destination Network Address Translation
DND	Do-not Disturb
DNS	Domain Name System
DNSSL	DNS Search List
DNU	Do Not Use for synchronization
DoD	Downstream On Demand
DoS	Denial of Service
DP	Drop Profile
DP	Data Plane
DPD	Dead Peer Detection

DPI	Deep Packet Inspection
DR	Designated Router
DR device	Distributed Relay device
DRB	Designated Routing Bridge
DRCP	Distributed Relay Control Protocol
DRCPDU	Distributed Relay Control Protocol Data Unit
DRE	Data Redundancy Elimination
DRNI	Distributed Resilient Network Interconnect
DRS	DPI Report Statistics
DS	Differentiated Services
DSA	Digital Signature Algorithm
DSCP	Differentiated Services Code Point
DSL	Digital Subscriber Line
DS-Lite	Dual Stack Lite
DSU	Data Service Unit
DTD	Document Type Definition
DTE	Data Terminal Equipment
DTIM	Delivery Traffic Indication Message
DTLS	Datagram Transport Layer Security
DTMF	Dual Tone Multi-Frequency
DU	Downstream Unsolicited
DUAL	Diffusing Update Algorithm
DUID	DHCP Unique Identifier
DUID-LL	DUID Based on Link-layer Address
D-V	Distance Vector
DVMRP	Distance Vector Multicast Routing Protocol
DVPN	Dynamic Virtual Private Network
DWDM	Dense Wavelength Division Multiplexing
E	Return
E2ETC	End-to-End Transparent Clock
EAA	Embedded Automation Architecture
EACL	Enhanced ACL
EAD	Endpoint Admission Defense
EAIS	Ethernet Alarm Indication Signal
EAN	Ethernet Access Node
EAP	Extensible Authentication Protocol
EAPOL	Extensible Authentication Protocol over LAN
EAPOL KCK	EAPOL-Key Confirmation Key
EAPOL KEK	EAPOL-Key Encryption Key

EAP-TLS	Extensible Authentication Protocol -Transport Layer Security
EBS	Excess Burst Size
ECDSA	Elliptic Curve Digital Signature Algorithm
ECID	E-channel Identifier
ECM	Encapsulated Control Message
ECM	Error Correction Mode
ECN	Explicit Congestion Notification
ECP	Edge Control Protocol
ECT	Equal Cost Tree
ECWmax	Exponent form of CWmax
ECWmin	Exponent form of CWmin
ED	Edge Device
EDCA	Enhanced Distributed Channel Access
EDSG	Enhanced Dynamic Service Gateway
EEE	Energy Efficient Ethernet
EF	Expedited Forwarding
EFM	Ethernet First Mile
EGP	Exterior Gateway Protocol
EID	Endpoint Identifier
EIGRP	Enhanced Interior Gateway Routing Protocol
EIRP	Effective Isotropic Radiated Power
EM	Event Monitor
EMO	Endpoint Mobile Office
EMR	Ensure Minimum Rate
ENDC	EVI Neighbor Discovery Client
ENDC	Enhanced Neighbor Discovery Client
End-Of-RIB	End of Routing-Information-Base
ENDP	Enhanced Neighbor Discovery Protocol
ENDP	EVI Neighbor Discovery Protocol
ENDS	Enhanced Neighbor Discovery Server
ENDS	EVI Neighbor Discovery Server
eNodeB	evolved Node B
ENUM	Telephone Number Mapping
EOAM	Ethernet Operation, Administration and Maintenance
EoPW	Ethernet over Pseudo Wire
EPC	External Path Computation
EPC	Evolved Packet Core
EPON	Ethernet Passive Optical Network
EPS	Endpoints Profiling System
ER	Edge Relay

ERPS	Ethernet Ring Protection Switching
ERSPAN	Encapsulated Remote Switch Port Analyzer
ERT	Export Route-target
ES	End System
ES	Ethernet Segment
ESF	Extended Super Frame
ESI	Ethernet Segment Identifier
ESMC	Ethernet Synchronous Message Channel
ESN	Extended Sequence Number
ESP	Encapsulating Security Payload
ESS	Extended Service Set
ESS	Exchange Switch Support
ETR	Egress Tunnel Router
ETS	Enhanced Transmission Selection
ETSI	European Telecommunications Standards Institute
EUI	Extended Unique Identifier
EUI-64	64-bit Extended Unique Identifier
EVB	Edge Virtual Bridging
Event MIB	Event Management Information Base
EVFP	Exchange Virtual Fabrics Parameter
EVI	Ethernet Virtual Interconnect
EVPN	Ethernet Virtual Private Network
Extended-MAM	Extended Maximum Allocation Model
F	Return
FA	Foreign Agent
FA	Foreign-AC
FAC	Feature Access Code
FAS	Frame Alignment Signal
FC	Forwarding Class
FC	Fibre Channel
FC	Feasibility Condition
FCF	FCoE Forwarder
FCM	Fast Connect Modem
FCoE	Fibre Channel over Ethernet
FCPS	FC Port Security
FCS	Fabric Configuration Server
FCS	Frame Check Sequence
FD	Feasible Distance
FDB	Forwarding Database

FDMI	Fabric Device Management Interface
FE	Fast Ethernet
FEAC	Far End Alarm and Control signal
FEC	Forward Error Correction
FEC	Forwarding Equivalence Class
FECN	Forward Explicit Congestion Notification
FEP	Front End Processor
FF	Fixed-Filter
FFD	Fast Failure Detection
FFT	Fast Fourier Transform
FG	Forwarding Group
FHR	First Hop Router
FIB	Forwarding Information Base
FIFO	First In First Out
FIFO	First In First Out Queuing
FIP	FCoE Initialization Protocol
FIP Snooping	FCoE Initialization Protocol Snooping
FIPS	Federal Information Processing Standards
Flowspec	Flow Specification
FNG	Fault Notification Generator
FP	Forwarding Profile
FPGA	Field Programmable Gate Array
FPMA	Fabric Provided MAC Address
FQ	Fair Queuing
FQDN	Fully Qualified Domain Name
FR	Frame Relay
FRR	Fast Reroute
FRTP	Frame Relay Traffic Policing
FRTS	Frame Relay Traffic Shaping
FRU	Field Replaceable Unit
FS	Forced Switch
FS	Feasible Successor
FSK	Frequency Shift Keying
FSPF	Fabric Shortest Path First
FT	Fault Tolerance
FT	Fast BSS Transition
FTN	FEC to NHLFE map
FTP	File Transfer Protocol
FXO	Foreign Exchange Office
FXS	Foreign Exchange Station

G	Return
GAL	G-ACh Label
GARP	Generic Attribute Registration Protocol
GAS	Generic Advertisement Service
GDOI	Group Domain of Interpretation
GE	Gigabit Ethernet
GI	Guard Interval
GK	GateKeeper
GLB	Global Load Balance
GM	Group Member
GM	Grandmaster Clock
GMK	Group Master Key
GNE	Gateway Network Element
GNS	Get Nearest Server
GOLD	Generic OnLine Diagnostics
GPS	Global Positioning System
GR	Graceful Restart
GRE	Generic Routing Encapsulation
Group Domain VPN	Group Domain Virtual Private Network
gRPC	Google Remote Procedure Call
GSMPv3	General Switch Management Protocol Version 3
GTC	Generic Token Card
GTK	Group Temporal Key
GTP	GPRS Tunneling Protocol
GTP-U	GPRS Tunneling Protocol User
GTS	Generic Traffic Shaping
GTSM	Generalized TTL Security Mechanism
GVRP	GARP VLAN Registration Protocol
GW	Gateway
GWMP	Gateway Message Protocol
H	Return
HA	Home-AC
HA	High Availability
HA	Home Agent
HABP	HW Bypass Protocol
HBA	Host Bus Adapter
HDB3	High-density bipolar 3
HDLC	High-level Data Link Control
HDSL	High-speed Digital Subscriber Line

HESSID	Homogenous Extended Service Set IDentifier
HGMP	HW Group Management Protocol
HGMPv2	HW Group Management Protocol version 2
HMAC	Hash-based Message Authentication Code
HoPE	Hierarchy of PE
HoVPN	Hierarchy of VPN
HQoS	Hierarchical Quality of Service
HS	High-performance Switch
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
H-VPLS	Hierarchical VPLS
HWTACACS	HW Terminal Access Controller Access Control System
I	Return
IA	Identity Association
IACTP	Inter Access Controller Tunneling Protocol
IAD	Integrated Access Device
IANA	Internet Assigned Numbers Authority
IBGP	Internal Border Gateway Protocol
ICANN	Internet Corporation for Assigned Names and Numbers
ICMP	Internet Control Message Protocol
ICMPv6	Internet Control Message Protocol for IPv6
ICP	IMA Control Protocol
ICPIF	Calculated Planning Impairment Factor
ICRQ	Incoming Call Request
IDN	Integrated Digital Network
IDS	Intrusion Detection System
IE	Information Element
IE	Interconnect Element
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IGMP Snooping	Internet Group Management Protocol Snooping
IGP	Interior Gateway Protocol
IGTK	Integrity Group Temporal Key
IIH	Intermediate System-to-Intermediate System Hello
IKE	Internet Key Exchange
IKEv2	Internet Key Exchange Version 2
ILM	Incoming Label Map
ILS	Internet Locator Service

IM	Instant Messaging
IMA	Inverse Multiplexing for ATM
IMAP	Internet Mail Access Protocol
iMC	Intelligent Management Center
IMSI	International Mobile Subscriber Identification Number
InARP	Inverse Address Resolution Protocol
IND	Inverse Neighbor Discovery
INT	In-band Network Telemetry
IntServ	Integrated Service
IPC	Inter-process Communication
IPE	Image Package Envelope
IPHC	IP Header Compression
IPL	Intra-Portal Link
IPN	IGTK Packet Number
IPng	IP Next Generation
IPoA	IP over ATM
IPoEoA	IP over Ethernet over ATM
IPP	Intra-Portal Port
IPS	Intrusion Prevention System
IPsec	IP Security
IPTA	IP Terminal Access
IPv6	Internet Protocol version 6
IPv6 MBGP	IPv6 Multicast BGP
IPv6 ND	IPv6 Neighbor Discovery
IPv6 PIM	IPv6 Protocol Independent Multicast
IPv6 PIM Snooping	IPv6 Protocol Independent Multicast Snooping
IPX	Internetwork Packet Exchange
IRB	Integrated Bridging and Routing
IRB	Integrated Routing and Bridging
IRB	Integrated Routing Bridging and Bridging Routing
IRDP	ICMP Router Discovery Protocol
IRF	Intelligent Resilient Framework
IS	Intermediate System
ISAKMP	Internet Security Association and Key Management Protocol
ISATAP	Intra-Site Automatic Tunnel Addressing Protocol
ISDN	Integrated Services Digital Network
I-SID	Backbone Service Instance Identifier
IS-IS	Intermediate System-to-Intermediate System
ISO	International Organization for Standardization
ISP	Internet Service Provider

ISPF	Incremental Shortest Path First
ISSU	In-Service Software Upgrade
IST	Internal Spanning Tree
ITA	Intelligent Target Accounting
ITR	Ingress Tunnel Router
ITSP	Internet Telephone Service Provider
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union Telecommunication
IV	Initialization Vector
IVR	Interactive Voice Response
K	Return
KAT	Known-Answer Test
KCK	EAPOL-Key Confirmation Key
KEK	Key Encrytion Key
KEK	EAPOL-Key Encryption Key
KS	Key Server
L	Return
L2F	Layer 2 Forwarding
L2PT	Layer 2 Protocol Tunneling
L2TP	Layer 2 Tunneling Protocol
L2VPN	Layer 2 Virtual Private Network
L3VNI	Layer 3 VNI
L3VPN	Layer 3 Virtual Private Network
LAC	L2TP Access Concentrator
LACP	Link Aggregation Control Protocol
LACPDU	Link Aggregation Control Protocol Data Unit
LAN	Local Area Network
LB	Loopback
LB	Label Base
LB	Load Balance
LBM	Loopback Message
LBR	Loopback Reply
LC	Logic Channel
LCFO	Loop Current Feed Open
LCI	Logic Channel Identifier
LCP	Link Control Protocol
LDAP	Lightweight Directory Access Protocol
LDN	Local Dialing Number
LDP	Label Distribution Protocol

LDP ID	LDP Identifier
LDPC	Low-Density Parity Check
LER	Label Edge Router
LFI	Link Fragmentation and Interleaving
LFIB	Label Forwarding Information Base
LGS	Loop-start & Ground-start Signaling
LHR	Last-Hop Router
LI	Leap Indicator
LLC	Logical Link Control
LLDP	Link Layer Discovery Protocol
LLDPDU	Link Layer Discovery Protocol Data Unit
LLDP-MED	Link Layer Discovery Protocol Media Endpoint Discovery
LLID	Logical Link ID
LLQ	Low Latency Queuing
LM	Loss Measurement
LMI	Local Management Interface
LMM	Loss Measurement Message
LMR	Loss Measurement Reply
LNS	L2TP Network Server
LO	Label-block Offset
LOCV	Loss of Connectivity Verification defect
LOID	Logical ONU Identifier
LoRaWAN	Long Range Wide Area Network
LOS	Loss of Signal
LPU	Line Processing Unit
LPU Port	Line Processing Unit Port
LQR	Link Quality Reports
LR	Line Rate
LR	Label Range
LS	Link State
LSA	Link State Acknowledgment
LSA	Link State Advertisement
LSAck	Link State Acknowledgment
LSDB	Link State Database
LSDB	Link State DataBase
LSN	Large-scale NAT
LSP	Label Switched Path
LSP	Link State PDU
LSPDU	Link State Protocol Data Unit
LSR	Link State Request

LSR	Link State Record
LSR	Label Switching Router
LSU	Link State Update
LT	Linktrace
LTM	Linktrace Message
LTR	Linktrace Reply
LTS	L2TP Tunnel Switch
LVF	Listening Virtual Forwarder
LZ	Lempel-Ziv compression
M	Return
M flag	Managed Address Configuration Flag
MA	Mobility Agent
MA	Maintenance Association
MAC	Message Authentication Code
MAC	Media Access Control
MACsec	Media Access Control Security
MAD	Multi-Active Detection
MAM	Maximum Allocation Model
MAN	Metropolitan Area Network
MAP	Mesh Access Point
MBGP	Multicast BGP
MBS	Maximum Burst Size
MC	Mobility Controller
MCC	Mobile Country Code
MCE	Multi-VPN-Instance Customer Edge
MCS	Modulation and Coding Scheme
MCU	Multipoint Control Unit
MD	Maintenance Domain
MD	Multicast Domain
MDC	Multitenant Device Context
MDI	Medium Dependent Interface
MDL	Maintenance Data Link
mDNS	Multicast DNS
MDT	Multicast Distribution Tree
ME	Maintenance Entity
MEC	Multi-access Edge Computing
MED	Multi-Exit Discriminator
MEG	Maintenance Entity Group
MEP	Maintenance association End Point

MEP	MEG End Point
MFAS	Multiframe FAS
MFC	Multi-Frequency Compelled
MFF	MAC-Forced Forwarding
MFR	Multilink Frame Relay
MGCP	Media Gateway Control Protocol
mGRE	Multipoint Generic Routing Encapsulation
MIB	Management Information Base
MIC	Message Integrity Check
MIMO	Multiple Input, Multiple Output
MIP	Maintenance association Intermediate Point
MIP	Mobile IP
MKA	MACsec Key Agreement
MKD	Mesh Key Distributor
MKI	Master Key Identifier
MLD	Multicast Listener Discovery Protocol
MLD Snooping	Multicast Listener Discovery Snooping
MLSP	Mobile Link Switch Protocol
MME	Management MIC IE
MNC	Mobile Network Code
MoH	Music on Hold
MOS	Mean Opinion Scores
MOTD	Message of The Day
MP	Maintenance Point
MP	Mesh Point
MP	Merge Point
MP_REACH_NLRI	Multiprotocol Reachable NLRI
MP_UNREACH_NLRI	Multiprotocol Unreachable NLRI
MP-BGP	Multiprotocol Border Gateway Protocol
MPCP	Multipoint Control Protocol
MPDU	MAC Protocol Data Unit
MPE	Middle-Level PE
MPLS	Multiprotocol Label Switching
MPLS SR	Segment Routing with MPLS
MPLS TE	Multiprotocol Label Switching Traffic Engineering
MPLS VPN	MPLS Virtual Private Network
MPLS-TP	MPLS Transport Profile
MPP	Mesh Portal Point
MQC	Model QoS Command
MR	Map Resolver

MR	Merge Request
MRIB	Multicast Routing Information Base
MRP	Multiple Registration Protocol
MRPDU	MRP Protocol Data Unit
MRRA	Merge Request Resource Allocation
MRU	Maximum-Receive-Unit
MS	Manual Switch
MS	Map Server
MSCHAP	Microsoft Challenge Handshake Authentication Protocol
MSDP	Multicast Source Discovery Protocol
MSDU	MAC Service Data Unit
MSK	Multicast Session Key
MSOH	Multiplex Section Overhead
MSS	Maximum Segment Size
MSTI	Multiple Spanning Tree Instance
MSTP	Multiple Spanning Tree Protocol
MT	Multicast Tunnel
MTI	Multicast Tunnel Interface
MTP	Maintain Probe
MTR	Multi-Topology Routing
MTU	Maximum Transmission Unit
MTU-s	Multi-Tenant Unit switch
MU	Mobile Unit
MVPN	Multicast Virtual Private Network
MVRP	Multiple VLAN Registration Protocol
MVXLAN	Multicast VXLAN
MWI	Message Waiting Indication
MX	Mail Exchanger
N	Return
NA	Neighbor Advertisement
NAI	Network Access Identifier
NAPT	Network Address Port Translation
NAPT-PT	Network Address Port Translation-Protocol Translation
NAS	Network Access Service
NAS	Network Access Server
NAS-ID	Network Access Server Identifier
NAS-Port-ID	Network Access Server Port Identifier
NAS-VLAN-ID	Network Access Server VLAN Identifier
NAT	Network Address Translation

NAT-PT	Network Address Translation-Protocol Translation
NBAR	Network Based Application Recognition
NBMA	Non-Broadcast Multi-Access
NCP	Network Control Protocol
ND	Neighbor Discovery
NDA	NetStream Data Analyzer
NDE	NetStream Data Exporter
NDP	Neighbor Discovery Protocol
NDPP	Network Device Protection Profile
NE	Network Element
NEMO	Network Mobility
NET	Network Entity Title
NetBIOS	Network Basic Input/Output System
NETCONF	Network Configuration Protocol
NFAS	non–FAS
NFV	Network Function Virtualisation
NHLFE	Next Hop Label Forwarding Entry
NHRP	Next Hop Resolution Protocol
NI	National ISDN
NIB	Nexthop Information Base
NII	Network International Identifier
NIST	National Institute of Standards and Technology
NIT	Not Initial Terminal
NLB	Network Load Balancing
NLBS	Network Load Balancing Service
NLPID	Network Layer Protocol Identifier
NLRI	Network Layer Reachability Information
NM	Network Management
NMFAS	Non-Multiframe FAS
NMS	Network Management Station
NMS	Network Management System
NMS	Network Management Server
NNI	Network-to-Network Interface
NO-PAT	Not Port Address Translation
NPDU	Network Protocol Data Unit
NPE	Network Provider Edge
NPTv6	IPv6-to-IPv6 Network Prefix Translation
NPV	N_Port Virtualization
NQA	Network Quality Analyzer
NR	No Request

nrt-VBR	non-real-time Variable Bit Rate
NS	Neighbor Solicitation
NS	Name Server
NSAP	Network Service Access Point
NSC	NetStream Collector
NSF	Non-Standard Facilities
NSR	Nonstop Routing
NSS	Number of Spatial Streams
NTDP	Neighbor Topology Discover Protocol
NTE	Named Telephone Event
NTP	Network Time Protocol
NTT	Nippon Telegraph and Telephone Corpration
NVC	Network Virtualization Controller
NVE	Network Virtualization Edge
NVGRE	Network Virtualization using Generic Routing Encapsulation
O	Return
O Flag	Other Stateful Configuration Flag
OAA	Open Application Architecture
OAMPDU	OAM Protocol Data Unit
OAP	Open Application Platform
OC	Optical Carrier
OC	Ordinary Clock
ODAP	On-Demand Address Pool
ODF	Optical Distribution Frame
ODN	Optical Distribution Network
ODU	Optical Data Unit
OEM	Original Equipment Manufacturer
OFDM	Orthogonal Frequency Division Multiplexing
OI	Organization Identifier
OID	Object Identifier
OLT	Optical Line Terminal
ONU	Optical Network Unit
OOB	Out-of-Band
OPU	Optical Payload Unit
ORF	Outbound Route Filtering
OSI	Open System Interconnection
OSPF	Open Shortest Path First
OSU	Online Sign Up
OTN	Optical Transport Network

OTU	Optical Transmit Unit
OUI	Organizationally Unique Identifier
OVSDB	Open vSwitch Database
P	Return
P2MP	Point-to-MultiPoint
P2P	PW to PW
P2P	Peer-to-Peer
P2P	Point-to-Point
P2PTC	Peer-to-Peer Transparent Clock
PA	Provider Allocated
PADT	PPPoE Active Discovery Terminate
PAGP	Port Aggregation Protocol
PAM	Port to Application Map
PAM	Pulse Amplitude Modulation
PAN ID	Personal Area Network Identifier
PAP	Password Authentication Protocol
Passive-Stateful PCC	Passive-Stateful Path Computation Client
Passive-Stateful PCE	Passive-Stateful Path Computation Element
PAT	Port Address Translation
PBAR	Port Based Application Recognition
PBB	Provider Backbone Bridge
PBBN	Provider Backbone Bridge Network
PBN	Provider Bridge Network
PBR	Policy-based Routing
PBX	Private Branch Exchange
PCB	Protocol Control Block
PCC	Path Computation Client
PCE	Path Computation Element
PCEP	Path Computation Element Communication Protocol
PCM	Pulse Code Modulation
PCR	Peak Cell Rate
PD	Powered Device
PD	Prefix Delegation
PDH	Plesiochronous Digital Hierarchy
PDP	Policy Decision Point
PDU	Protocol Data Unit
PE	Provider Edge
PE-agg	PE Aggregation
PEAP	Protected Extensible Authentication Protocol

PEAP-GTC	Protected Extensible Authentication Protocol-Microsoft Generic Token Card
PEAP-MSCHAPv2	Protected Extensible Authentication Protocol-Microsoft Challenge Handshake Authentication Protocol v2
PEM	Privacy Enhanced Mail
PEM	Power Entry Module
PEP	Policy Enforcement Point
PFC	Power Free Connector
PFC	Priority-based Flow Control
PFC	Protocol-Field-Compression
PFS	Perfect Forward Secrecy
PHB	Per-hop Behavior
PHP	Penultimate Hop Popping
PI	Power Interface
PI	Provider Independent
PI	Progress Indicator
PIC	Prefix Independent Convergence
PIM	Protocol Independent Multicast
PIM Snooping	Protocol Independent Multicast Snooping
PIM-DM	Protocol Independent Multicast-Dense Mode
PIM-SM	Protocol Independent Multicast-Sparse Mode
PIM-SSM	Protocol Independent Multicast Source-Specific Multicast
PIN	Personal Identification Number
PIR	Peak Information Rate
PKCS	Public Key Cryptography Standards
PKI	Public Key Infrastructure
PLCP	Physical Layer Convergence Procedure
PLMN	Public Land Mobile Network
PLR	Point of Local Repair
PMK	Pairwise Master Key
PMTU	Path MTU
PoE	Power over Ethernet
POH	Path Overhead
POP3	Post Office Protocol - Version 3
POS	Passive Optical Splitter
POS	Point of Sale
PPP	Point-to-Point Protocol
PPPoA	PPP over ATM
PPPoE	Point-to-Point Protocol over Ethernet
PPPoEoA	PPPoE over ATM

PPTP	Point-to-Point Tunneling Protocol
PQ	Priority Queuing
PRBS	Pseudo Random Bit Sequence
PRC	Primary Reference Clock
PRI	Primary Rate Interface
PRL	Preferred Roaming List
PS	Protection Switching
PSB	Path State Block
PSC	Protection State Coordination
PSD	Power Spectral Density
PSE	Power Sourcing Equipment
PSK	Pre-Shared Key
PSN	Packet Switched Network
PSNP	Partial Sequence Number PDU
PSST	Principal Switch Selection Timer
PSTN	Public Switched Telephone Network
PTK	Pairwise Transient Key
PTP	Precision Time Protocol
PTR	Pointer Record
PTY	Pseudo-Terminals
PUK	PIN Unlocking Key
PVC	Permanent Virtual Circuit
PVC PQ	PVC Priority Queuing
PVID	Port VLAN ID
PVP	Permanent Virtual Path
PVST	Per-VLAN Spanning Tree
PW	Pseudowire
PW-ACH	PW Associated Channel Header
PWCT	Pairwise Conditional Test
PWE3	Pseudo Wire Emulation Edge-to-Edge
PXE	Preboot eXecution Environment
Q	Return
QCN	Quantized Congestion Notification
QL	Quality Level
QoS	Quality of Service
QPPB	QoS Policy Propagation Through the Border Gateway Protocol
R	Return
RA	Router Alert
RA	Router Advertisement

RA	Registration Authority
RADIUS	Remote Authentication Dial-In User Service
RAI	Remote Alarm Indication
RALM	RADIUS Authenticated Login Using MAC-address
R-APS	Ring Automatic Protection Switching
RAS	Registration, Admission, and Status
RB	Routing Bridge
RBAC	Role Based Access Control
RBM	Remote Backup Management
RCA	Release Change Authorization
RCF	Reconfigure Fabric
RCPI	Received Channel Power Indicator
RD	Routing Domain
RD	Reported Distance
RD	Route Distinguisher
RDI	Remote Defect Indication
RDM	Russian Dolls Model
RDP	Read Diagnostic Parameters
RED	Random Early Detection
RedisDBM	Redis Database Manager
REG	Registration Center
RFID	Radio Frequency Identification
RIB	Routing Information Base
RIP	Routing Information Protocol
RIP-2	Routing Information Protocol version 2
RIPng	RIP next generation
RIR	Resilient Intelligent Route
RLOC	Routing Locator
RLSN	Remote Link Status Notification
RMON	Remote Network Monitoring
ROA	Recognized operating Agency
Roam OI	Roam Organization Identifier
RP	Rendezvous Point
RP	Reflective Relay
RP	Reaction Point
RPA	Register Port Attributes
RPA	Rendezvous Point Address
RPC	Remote Procedure Call
RPF	Reverse Path Forwarding
RPI	Receive Power Indication

RPKI	Resource Public Key Infrastructure
RPL	Rendezvous Point Link
RPL	Ring Protection Link
RPR	Resilient Packet Ring
RPRT	Register Port
RPS	Redundant Power System
RP-Set	Rendezvous Point Set
RPT	Rendezvous Point Tree
RR	Route Reflector
RR	Reflective Relay
RR	Router Renumber
RRI	Reverse Route Injection
RRM	Radio Resource Management
RRPP	Rapid Ring Protection Protocol
RS	Router Solicitation
RSB	Reservation State Block
RSCN	Registered State Change Notification
RS-DCC	Resource-Shared DCC
RSH	Remote Shell
RSN	Robust Security Network
RSNA	Robust Security Network Association
RSNI	Received Signal to Noise Indicator
RSOH	Regenerator Section Overhead
RSSI	Received Signal Strength Indicator
RSSI	Received Signal Strength Indication
RSTP	Rapid Spanning Tree Protocol
RSVP	Resource Reservation Protocol
RT	Route Target
RTC	Remote Terminal Connection
RTCP	Real-Time Transport Control Protocol
RTM	Real-Time Event Manager
RTP	Reliable Transport Protocol
RTP	Real-Time Transport Protocol
RTPQ	Real-Time Transport Protocol Priority Queuing
RTS	Request to Send
RTSP	Real Time Streaming Protocol
RTT	Round-Trip Time
rt-VBR	Real-time Variable Bit Rate
RU	Rack Unit

S	Return
S/MIME	Secure/Multipurpose Internet Mail Extensions
SA	Security Association
SA	Source Active
SA Query	Security Association Query
SABM	Set Asynchronous Balanced Mode
SACK	Selective Acknowledgment
SAE	Simultaneous authentication of equals
SAFI	Subsequent Address Family Identifier
SAII	Source Attachment Individual Identifier
SAM	Security Accounting Management
SAN	Storage Area Networks
SAP	Service Advertising Protocol
SAToP	Structure-Agnostic TDM over Packet
SAVI	Source Address Validation Improvement
SB	Service Bridge
SBFD	Seamless BFD
SBM	Subnetwork Bandwidth Management
SCCP	Skinny Client Control Protocol
SCEP	Simple Certificate Enrollment Protocol
SCFF	Single Choke Fairness Frame
SCR	Sustainable Cell Rate
SCR	State Change Registration
SD	Signal Degrade
SDH	Synchronous Digital Hierarchy
SDLC	Synchronous Data Link Control
SDN	Software Defined Network
SDP	Session Description Protocol
SDSL	Symmetric Digital Subscriber Line
SDU	Service Data Unit
SE	Shared-Explicit
SEC	SDH Equipment Clock
SF	Super Frame
SF	Signal Fail
SFC	Stage Fabric Configuration Update
sFlow	Sampled Flow
SFM	Source-Filtered Multicast
SFP	Small Form-factor Pluggable
SFTP	Secure FTP

Short GI	Short Guard Interval
SIA	Stuck In Active
SID	Segment Identifier
SIP	Session Initiation Protocol
SIPS	SIP Secure
SLA	Service Level Agreement
SLAAC	Stateless Address Autoconfiguration
SLB	Server Load Balance
SLC	Service Logic Control
SLIP	Serial Line Internet Protocol
SLM	Single-ended Loss Measurement
SMA	State Machine based Anti-spoofing
SmartMC	Smart Management Center
SMET	Selective Multicast Ethernet Tag Route
S-MLAG	Simple Multichassis Link Aggregation
SMTP	Simple Mail Transfer Protocol
SN	Serial Number
SNA	System Network Architecture
SNAP	Subnet Access Protocol
SNAT	Source Network Address Translation
SNMP	Simple Network Management Protocol
SNMP-DCA	SNMP Data Collection Agent
SNP	Sequence Number PDU
SNPA	Subnetwork Point of Attachment
SNR	Signal-to-Noise Ratio
SNTP	Simple NTP
SOA	Start of Authority
SOAP	Simple Object Access Protocol
SOH	Section Overhead
SONET	Synchronous Optical Network
SoO	Site of Origin
SP	Strict Priority
SP	Service Provider
SP	Scheduler Policy
SPB	Shortest Path Bridging
SPBM	Shortest Path Bridging MAC Mode
SPBM	Shortest Path Bridging MAC
SPBN	Shortest Path Bridging Network
SPCS	Stored Program Control Switching System
SPE	Superstratum PE or Service Provider-end PE

SPF	Shortest Path First
SPI	Security Parameter Index
SPID	Service Profile Identification
SPM	Smart PoE Manager
SPSource ID	Shortest Path Source Identifier
SPT	Shortest Path Tree
SQL	Structured Query Language
SR	Segment Routing
Srefresh	Summary Refresh
SRGB	Segment Routing Global Block
SRLB	Segment Routing Local Block
SRLSP	Segment Routing Label Switched Path
SRMC	Segment Routing Mapping Client
SRMS	Segment Routing Mapping Server
SRST	Survivable Remote Site Telephony
sr-TCM	single rate TCM
SRTP	Secure Real-Time Transport Protocol
SRV	Service
SSA	SIP Stack Adapt
SSH	Secure Shell
SSID	Service Set Identifier
SSL	Secure Sockets Layer
SSM	Synchronization Status Message
SSM	Source-Specific Multicast
SSP	Switch-to-Switch Protocol
SSU-A	primary level SSU
SSU-B	second level SSU
ST	Shared Tree
Stateful PCE	Stateful Path Computation Element
Stateless PCC	Stateless Path Computation Client
Stateless PCE	Stateless Path Computation Element
STBC	Space-Time Block Coding
STP	Spanning Tree Protocol
STR	Switch Trace Route
SVC	Static Virtual Circuit
SVC	Switched Virtual Circuit
SVP	SpectraLink Voice Priority
T	Return
TA	Terminal Adapter

TACACS	Terminal Access Controller Access Control System
TACL	Temporary Access Control List
TAI	International Atomic Time
TAII	Target Attachment Individual Identifier
T-BC	Telecom Boundary Clock
TC	Topology Checksum
TC	Topology Change
TC	Topology client
TC	Traffic Class
TC	Transparent Clock
TCA	Topology Change Acknowledge
TCAM	Ternary Content Addressable Memory
TCI	Tag Control Information
Tcl	Tool Command Language
TCN BPDU	Topology Change Notification BPDU
TCP	Transmission Control Protocol
TCT-BC	Transparent Telecom Boundary Clock
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TE	Traffic Engineering
TEDB	TE DataBase
TEI	Terminal Endpoint Identifier
TEID	Tunnel Endpoint Identifier
TEK	Tranfic Encrytion Key
TFO	Transport Flow Optimization
TFTP	Trivial File Transfer Protocol
TI-LFA FRR	Topology-Independent Loop-free Alternate Fast Reroute
TIM	Traffic Indication Map
TK	Temporal Key
TKIP	Temporal Key Integrity Protocol
TLS	Transport Layer Security
TLS	Transparent LAN Service
TLV	Type/Length/Value
TM	Topology Master
ToS	Type of Service
TP	Traffic Policing
TP	Topology Protection
TPA	Third-part Application
TPC	Transmit Power Control
TPDU	Transport Protocol Data Unit

TPID	Tag Protocol Identifier
TPMR	Two-Port MAC Relay
TQ	Time Quantum
TRILL	TRansparent Interconnection of Lots of Links
TRIP	Triggered RIP
tr-TCM	two-rate TCM
TS	Traffic Shaping
TSA	Tunnel Switching Aggregator
TST	Test
T-TC	Telecom Transparent Clock
TTI	Trail Trace Identifier
TTL	Time to Live
TTLS	Tunneled Transport Layer Security
TTLS-GTC	Tunneled Transport Layer Security-Microsoft Generic Token Card
TTLS-MSCHAPv2	Tunneled Transport Layer Security-Microsoft Challenge Handshake Authentication Protocol v2
T-TSC	Telecom Time Slave Clock
TTY	True Type Terminal
TU	Tributary Unit
TUG	Tributary Unit Group
TWAMP	Two-Way Active Measurement Protocol
TxBF	Tx Beamforming
TXOP Limit	Transmission Opportunity Limit
TXT	Text
U	Return
UA	Unnumbered Acknowledge
UA	User Agent
UAC	User Agent Client
U-APSD	Unscheduled Automatic Power-Save Delivery
UAS	User Agent Server
UBR	Unspecified Bit Rate
UCM	User Connection Management
UDLD	Uni-directional Link Direction
UDP	User Data Protocol
UE	User Equipment
UFC	Update Fabric Configuration
UIM	User Identity Module
UNI	User-to-Network Interface
UNI	User Network Interface

UPE	User facing-Provider Edge
UPE	Underlayer PE or User-end PE
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
uRPF	Unicast Reverse Path Forwarding
USK	Unicast Session Key
USM	User-Based Security Model
UTC	Coordinated Universal Time
UTM	Unified Threat Management
V	Return
V2V	Video to Video
VA	Virtual Access
VACM	View-based Access Control Model
VAD	Voice Activity Detection
VAM	VPN Address Management
VBR-NRT	Variable Bit Rate-Non Real Time
VBR-RT	Variable Bit Rate-Real Time
VC	Virtual Circuit
VC	Virtual Container
VCC	Virtual Circuit Connection
VCCV	Virtual Circuit Connectivity Verification
VCF	Vertical Converged Framework
VCFC	VCF Controller
VCI	Virtual Channel Identifier
VCN	Virtual Circuit Number
VCPM	Voice Co-processing Module
VD	Virtual Device
VDP	VSI Discovery and Configuration Protocol
VDSL	Very high-speed Digital Subscriber Line
VE	Virtual Ethernet
VE-Bridge	Virtual Ethernet Bridge
VEth	Virtual Ethernet
VF Owner	Virtual Forwarder Owner
VFW	Virtual Firewall
VLAN	Virtual Local Area Network
VM	Virtual Machine
VMM	Virtual Machine Management
VN	Version Number
VNC	Virtual Network Computing

VNI	VXLAN Network Identifier
VoD	Video on Demand
VoIP	Voice over IP
VP	Virtual Path
VPC	Virtual Path Connection
VPC	Virtual Private Cloud
VPDN	Virtual Private Dial-up Network
VPI	Virtual Path Identifier
VPLS	Virtual Private LAN Service
VPM	Voice Processing Module
VPN	Virtual Private Network
VPWS	Virtual Private Wire Service
VR	Virtual Router
VRF	Virtual Routing and Forwarding
VRRP	Virtual Router Redundancy Protocol
VSAN	Virtual Storage Area Network
VSD	Virtual Security Domain
VSI	Virtual Switch Instance
VSI	Virtual Station Interface
VSID	Virtual Subnet Identifier
VSIP	Virtual Service IP
VSRP	Virtual Service Redundancy Protocol
vSwitch	Virtual Switch
VT	Virtual Template
VTEP	VXLAN Tunnel End Point
VTP	VLAN Trunking Protocol
VTY	Virtual Type Terminal
VXLAN	Virtual eXtensible LAN
VXLAN-DCI	VXLAN Data Center Interconnection
W	Return
WAAS	Wide Area Application Services
WAF	Web Application Firewall
WAI	WLAN Authentication Infrastructure
WAN	Wide Area Network
WAP	Wireless Application Protocol
WAPI	WLAN Authentication and Privacy Infrastructure
WASS	Wide Area Application Services
WDS	Wireless Distribution System
WEP	Wired Equivalent Privacy

WFQ	Weighted Fair Queuing
WHA	WLAN High Availability
WIDS	Wireless Intrusion Detection System
WiNet	Wisdom Network
WINS	Windows Internet Naming Service
WIPS	Wireless Intrusion Prevention System
WLAN	Wireless Local Area Network
WLAN RRM	Radio Resource Management
WMM	Wi-Fi Multimedia
WPA	Wi-Fi Protected Access
WPAD	Web Proxy Auto-Discovery
WPI	WLAN Privacy Infrastructure
WRED	Weighted Random Early Detection
WRR	Weighted Round Robin
WSA	Wireless Spectrum Analysis
WT	Wireless Terminator
WTU	Wireless Terminator Unit
WWN	World Wide Name
X	Return
XFP	10-Gigabit Small Form-factor Pluggable
XML	Extensible Markup Language
XSD	XML Schema Document
XSD	XML Schema Definition
Z	Return
ZBR	Zone Border Router